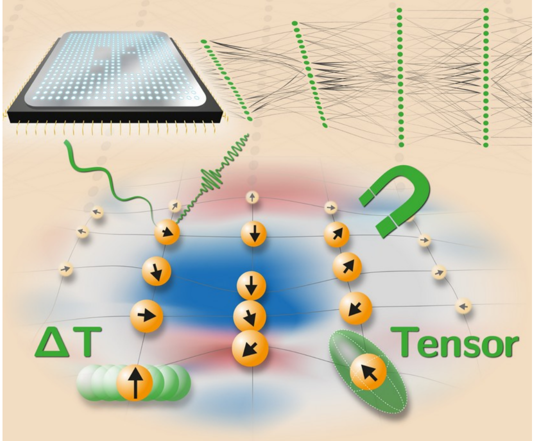
Machine Learning of First Principles Observables



July 8, 2024 - July 12, 2024
Zuse Institute Berlin, Germany

Simone Köcher
IET-1, Forschungszentrum Jülich GmbH, Germany

Angela Harper
Fritz-Haber Institut der Max Planck Gesellschaft, Germany

Hanna Türk
École Polytechnique Fédérale de Lausanne, Switzerland

Elena Gelžinytė
Fritz-Haber Institut der Max Planck Gesellschaft, Germany

Giulia Glorani
Fritz-Haber Institut der Max Planck Gesellschaft, Germany

With the support of:

Psi-k

DFG Deutsche
Forschungsgemeinschaft
German Research Foundation

German Research Foundation (DFG)

FRITZ-HABER-INSTITUT
MAX-PLANCK-GESELLSCHAFT

Fritz-Haber-Institut der Max-Planck-Gesellschaft

Table of Contents

General Information

Contact & Important Information

Website

https://www.cecam.org/workshop-details/1322

Data Protection Information

according to Art. 13, 14 and 21 of the General Data Protection Regulation (GDPR)

Creation of photos

We will be taking photographs at this workshop, and this data will be shared according to the GDPR compliant guidelines of the Fritz Haber Institute of the Max Planck Society. Please see the QR code for further information about data protection.

Data protection is an important concern for us. Below we inform you how we process your data and which rights you have.

1. Processing purposes and legal basis

Your personal data will be processed in accordance with the provisions of the General Data Protection Regulation (GDPR), the Federal Data Protection Act (BDSG) and other relevant data protection regulations. In the case of photographs, the Art Copyright Act will apply in the event of possible publication.

1.1 Consent (Art. 6 para. 1 lit a GDPR)

By signing in at registration, you give us consent to the processing of personal data, which is the legal basis for the processing mentioned here. You can revoke your consent at any time with effect for the future. If you don’t want to give your consent, please affix one of the **“No Photos” stickers** (available at registration) onto your nametag.

1.2 Press and public relations (Art. 6 para. 1 lit f GDPR)

Photographs are processed for press and public relations work and reporting to the funding agencies.

2. Categories of personal data processed by us

The following data is collected/created and further processed:

- Photographs

3. Who receives the data?

We pass on your personal data within our company to the divisions, who need this data to fulfil their contractual and legal obligations or to implement our legitimate interest. Internet presences of the funding agencies and the Fritz Haber Institute of the Max Planck Society also record presences in social networks such as Mastodon. Information on their terms of use can be found in the data protection declaration (see QR code above).

4. How long do we store your data?

If necessary, we process your personal data for the duration of our business relationship or the duration of publication. The specified periods for storage or documentation are up to ten years after the end of the business relationship or the pre-contractual legal relationship. Ultimately, the storage period is also assessed according to the statutory limitation periods and it is regularly checked whether data can be deleted.

5. Your privacy rights

You have the right of Information pursuant to Art. 15 GDPR, the right of rectification pursuant to Art. 16 GDPR, the right of deletion pursuant to Art. 17 GDPR, the right of restriction of processing pursuant to Art. 18 GDPR and the right of data transfer pursuant to Art. 20 GDPR. In addition, there is a right of appeal to a data protection supervisory authority (Art. 77 GDPR).

6. Your right to appeal to the competent supervisory authority

You have the right to appeal to the data protection supervisory authority (Art. 77 GDPR). The supervisory authority responsible is:

Heidi Schuster
Max Planck Society for the Advancement of Science e.V. (MPG)
Hofgartenstrasse 8
D-80539 Munich
Telephone: +49 (89) 2108-1554
datenschutz@mpg.de

Workshop Description

Recently, Machine Learning (ML) methods have penetrated almost all research areas in materials modelling and high-throughput materials screening. And yet the ML triumph has so far mainly focused on developing surrogate models for the potential energy surface (PES) with superior computational efficiency while retaining first principles accuracy. The approach to learn observable properties directly is just emerging and is challenged by several issues, which we intend to address.

The event is meant to support the development of a new collaborative, international network connecting different fields of research and integrating the young researchers community with the help of a scientifically diverse, interactive workshop.

Topics:
- *ML of electron density and Hamiltonians*
- *ML of electronic observables*
- *ML of mechanical & magnetic observables*
- *ML of spectroscopic observables*
- *ML of reaction networks*
- *Theoretical and experimental databases*

Objectives

The majority of materials modelling with ML methods represents the (PES) of a material based on the assumption, that the total energy of the system can be decomposed into atomic contributions, which in a first approximation are described as a function of the local atomic environment [1,2]. However, several observables such as charge transfer, dipoles, and the material's properties in an applied electric field are inherently non-local properties. First approaches [3,4] are able to include long range interactions in models for interatomic potentials. It remains an open question if or how these or other methods can be adapted for ML models of non-local observables.

In contrast to the scalar potential energy of a system, many properties are either vectors, like dipoles, or high rank tensors, like electric field gradients, or more complex properties, like density of states. First implementations encode tensorial properties in either rotationally invariant or equivariant representations [5-7]. Other approaches aim to learn the entire electron density in order to derive the observables [8-10]. However, the practical application to physical observables is still very limited. An open discussion of the concepts is necessary and will provide an essential contribution to the dissemination of the methods within the community.

The majority of materials and their observables are unambiguously described by the atomic structure features. Some properties however also depend on spin states, magnetic arrangement or atomic charges. Including information like atomic magnetic vectors or atomic charges in the feature vector is challenging and not many approaches exist [11,12] to solve the problem, which needs to be addressed to model key properties like charge transfer and spin waves.

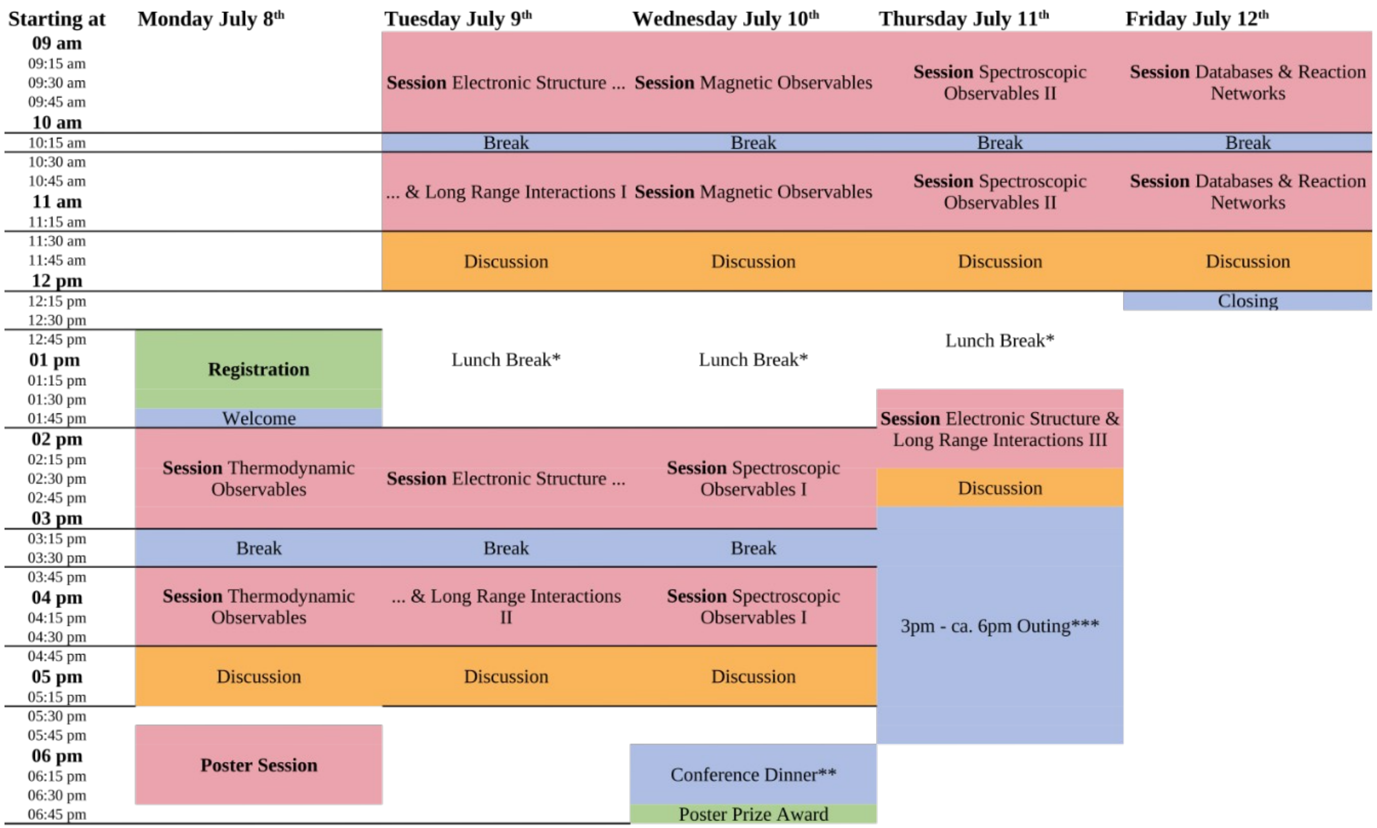
The accuracy and performance of any ML model depends critically on the extent and diversity of its training dataset. The different available databases [13-16] of 'synthetic' first principle material properties provide a valuable wealth of information, but approaches to combine data have yet to be developed. Other databases [17-19] complement the available information by experimental data. The question remains, whether or how to combine experimental and theoretical data on equal footing in order to pool the available resources.The workshop aims to tackle the challenges in the fledgling field of ML of observables covering spectroscopic, electronic, thermodynamic, magnetic, and mechanical properties as well as ML approaches to predict the electron density.

We gratefully acknowledge the support by CECAM, the Psi-k Charity, Deutsche Forschungsgemeinschaft, and the Max-Planck-Gesellschaft.

Key References

*[1] M. Langer, A. Goeßmann, M. Rupp, npj. Comput. Mater., **8**, 41 (2022)*
*[2] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. Shapeev, A. Thompson, M. Wood, S. Ong, J. Phys. Chem. A, **124**, 731-745 (2020)*
*[3] T. Ko, J. Finkler, S. Goedecker, J. Behler, Nat. Commun., **12**, 398 (2021)*
*[4] C. Staacke, H. Heenen, C. Scheurer, G. Csányi, K. Reuter, J. Margraf, ACS Appl. Energy Mater., **4**, 12562-12569 (2021)*
*[5] A. Grisafi, D. Wilkins, G. Csányi, M. Ceriotti, Phys. Rev. Lett., **120**, 036002 (2018)*
*[6] V. Nguyen, A. Lunghi, Phys. Rev. B, **105**, 165131 (2022)*
*[7] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. Mailoa, M. Kornbluth, N. Molinari, T. Smidt, B. Kozinsky, Nat. Commun., **13**, 2453 (2022)*
*[8] M. Gastegger, A. McSloy, M. Luya, K. Schütt, R. Maurer, J. Chem. Phys., **153**, 044123 (2020)*
*[9] B. Focassio, M. Domina, U. Patil, A. Fazzio, S. Sanvito, npj. Comput. Mater., **9**, 87 (2023)*
*[10] J. Nigam, M. Willatt, M. Ceriotti, J. Chem. Phys., **156**, 014115 (2022)*
*[11] M. Domina, M. Cobelli, S. Sanvito, Phys. Rev. B, **105**, 214439 (2022)*
*[12] M. Eckhoff, J. Behler, npj. Comput. Mater., **7**, 170 (2021)*
*[13] S. Curtarolo, W. Setyawan, G. Hart, M. Jahnatek, R. Chepulskii, R. Taylor, S. Wang, J. Xue, K. Yang, O. Levy, M. Mehl, H. Stokes, D. Demchenko, D. Morgan, Computational Materials Science, **58**, 218-226 (2012)*
*[14] C. Draxl, M. Scheffler, J. Phys. Mater., **2**, 036001 (2019)*
*[15] A. Jain, S. Ong, G. Hautier, W. Chen, W. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. Persson, APL Materials, **1**, (2013)*
*[16] S. Kirklin, J. Saal, B. Meredig, A. Thompson, J. Doak, M. Aykol, S. Rühl, C. Wolverton, npj. Comput. Mater., **1**, 15010 (2015)*
*[17] C. Groom, I. Bruno, M. Lightfoot, S. Ward, Acta. Crystallogr. Sect. B., **72**, 171-179 (2016)*
*[18] D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, S. Rehme, J. Appl. Cryst., **52**, 918-925 (2019)*
*[19] H. Sun, S. Dwaraknath, H. Ling, X. Qu, P. Huck, K. Persson, S. Hayes, npj. Comput. Mater., **6**, 53 (2020)*

Program Overview



Starting at	Monday July 8th	Tuesday July 9th	Wednesday July 10th	Thursday July 11th	Friday July 12th
09 am		**Session** Electronic Structure ...	**Session** Magnetic Observables	**Session** Spectroscopic Observables II	**Session** Databases & Reaction Networks
09:15 am					
09:30 am					
09:45 am					
10 am					
10:15 am		Break	Break	Break	Break
10:30 am		... & Long Range Interactions I	**Session** Magnetic Observables	**Session** Spectroscopic Observables II	**Session** Databases & Reaction Networks
10:45 am					
11 am					
11:15 am					
11:30 am		Discussion	Discussion	Discussion	Discussion
11:45 am					
12 pm					
12:15 pm		Lunch Break*	Lunch Break*	Lunch Break*	Closing
12:30 pm					
12:45 pm	**Registration**				
01 pm					
01:15 pm					
01:30 pm				**Session** Electronic Structure & Long Range Interactions III	
01:45 pm	Welcome				
02 pm	**Session** Thermodynamic Observables	**Session** Electronic Structure ...	**Session** Spectroscopic Observables I		
02:15 pm					
02:30 pm				Discussion	
02:45 pm					
03 pm				3pm - ca. 6pm Outing***	
03:15 pm	Break	Break	Break		
03:30 pm					
03:45 pm	**Session** Thermodynamic Observables	... & Long Range Interactions II	**Session** Spectroscopic Observables I		
04 pm					
04:15 pm					
04:30 pm					
04:45 pm	Discussion	Discussion	Discussion		
05 pm					
05:15 pm					
05:30 pm					
05:45 pm	**Poster Session**				
06 pm			Conference Dinner**		
06:15 pm					
06:30 pm					
06:45 pm			Poster Prize Award		

Detailed Program

Monday July 8th

12:45 – 13:45 Registration *

13:45 – 14:00 Welcome

14:00 – 16:45 Session **Thermodynamic Observables**

Chair: Dr. Simone Köcher

14:00 – 14:45 **Prof. Dr. Karsten Reuter** (FHI Berlin, DE)

Out of the Crystalline Comfort Zone:
Tackling Working Interfaces with Machine Learning

14:45 – 15:15 **Dr. Michele Simoncelli** (University of Cambridge, UK)

Machine learning opens a wonderland for looking through glasses

15:15 – 15:45 Break

15:45 – 16:15 **Dr. Christian Carbogno** (FHI Berlin, DE)

Accelerating Transport Coefficient
Predictions via Machine Learning

16:15 – 16:45 **Prof. Dr. Nong Artrith** (Utrecht University, NL)

Harnessing Machine Learning for Advancing Amorphous Battery Materials

16:45 – 17:30 Discussion

17:30 – 19:00 Poster Session

* **Lunch** is not provided. You'll find an ample offer of restaurants and take-away in the vicinity.

Tuesday July 9th

09:00 – 11:30 Session **Electronic Structure & Long Range Interactions I**

Chair: Dr. Elena Gelzinyte

09:00 – 09:45 **Prof. Dr. Gábor Csányi** (University of Cambridge, UK)

A foundational atomistic model for materials

09:45 – 10:15 **Prof. Dr. Janine George** (BAM, DE)

High-throughput Approaches for Materials Understanding and Design

10:15 – 10:30 Break

10:30 – 11:00 **Sergey Pozdnyakov** (EPFL, CH)

Challenging the dogma of rotational equivariance in atomistic ML

11:00 – 11:30 **Prof. Dr. Kulbir Ghuman** (INRS, CAN)

Leveraging Computational Advances to Design and Optimize Energy Materials: From Traditional Methods to Machine Learning

11:30 – 12:15 Discussion

12:15 – 14:00 Lunch Break*

14:00 – 16:45 Session **Electronic Structure & Long Range Interactions II**

Chair: Dr. Hanna Türk

14:00 – 14:45 **Prof. Dr. Michele Ceriotti** (EPFL, CH)

Machine-learning for electronic structure

14:45 – 15:15 **Alexander Knoll** (Ruhr-Universität Bochum, DE)

Advanced Software Frameworks for Describing Local and Non-Local Interactions in High-Dimensional Neural Network Potentials

15:15 – 15:45 Break (& Group Picture)

15:45 – 16:15 **Prof. Dr. Reinhard Maurer** (University of Warwick, UK)

Electronic Structure Surrogate Learning for Quantum Dynamics and Inverse Design

16:15 – 16:45 **William Baldwin** (University of Cambridge, UK)

ML Electrostatics Models in Relevant Test Systems

16:45 – 17:30 Discussion

* **Lunch** is not provided.

Wednesday July 10th

09:00 – 11:30 Session **Magnetic Observables**

Chair: Dr. Simone Köcher

09:00 – 09:45 **Prof. Dr. Stefano Sanvito** (Trinity College Dublin, IRL)

The Jacobi-Legendre framework for materials discovery

09:45 – 10:15 **Johannes Wasmer** (IAS-1 Forschungszentrum Jülich, DE)

Prediction of magnetic exchange interaction in doped topological insulators

10:15 – 10:30 Break

10:30 – 11:00 **Prof. Dr. Alessandro Lunghi** (Trinity College Dublin, IRL)

Machine Learning for Molecular Magnetism

11:00 – 11:30 **Shuping Guo** (IFW Dresden, DE)

Machine learning facilitated by microscopic features for discovery of novel magnetic double perovskites

11:30 – 12:15 Discussion

12:15 – 14:00 Lunch Break*

14:00 – 16:45 Session **Spectroscopic Observables I**

Chair: Prof. Dr. Josef Granwehr

14:00 – 14:45 **Prof. Dr. Patrick Rinke** (Aalto University, FIN)

Machine Learning for Spectroscopy – Concepts, Successes, and Challenges

14:45 – 15:15 **Dr. Tigany Zarrouk** (Aalto University, FIN)

Experiment-driven atomistic materials modeling: Combining XPS and MLPs to infer the structure of a-CO_x

15:15 – 15:45 Break

15:45 – 16:15 **Prof. Dr. Rose Cersonsky** (University of Wisconsin, USA)

Categorizing three-dimensional photonic crystals: open challenges in scale-covariant problems

16:15 – 16:45 **Clelia Middleton** (Newcastle University, UK)

p-DOS: a descriptor with electronic wisdom for learning X-Ray spectroscopy

16:45 – 17:30 Discussion

18:00 Conference Dinner

Thursday July 11th

09:00 – 11:30 Session **Spectroscopic Observables II**

Chair: Dr. Angela Harper

09:00 – 09:45 **Prof. Dr. Rebecca Nicholls** (University of Oxford, UK)

Interpreting core-loss spectroscopy

09:45 – 10:15 **Prof. Dr. Josef Granwehr** (IEK-9 FZJ, DE)

Predicting electron paramagnetic resonance parameters and their sensitivity to structural configuration

10:15 – 10:30 Break

10:30 – 11:00 **Prof. Dr. Claudia Draxl** (Humboldt-Universität zu Berlin, DE)

Assessing spectroscopic features: from fingerprinting to predictions

11:00 – 11:30 **Prof. Dr. Stefan Sandfeld** (IAS-9 FZJ, DE)

Scientific Machine Learning and Explainable AI Approaches for the Physical Sciences

11:30 – 12:15 Discussion

12:15 – 13:30 Lunch Break*

13:30 – 14:30 Session **Electronic Structure & Long Range Interactions III**

Chair: Prof. Dr. Janine George

13:30 – 14:00 **Luca Leoni** (University of Bologna, IT)

Machine learned small polaron dynamics

14:00 – 14:30 **Bartosz Brzoza** (CASUS/HZDR, DE)

Applying SE(3)-Equivariant Attentional Graph Neural Networks for the purpose of predicting the electronic structure of molecular hydrogen

14:30 – 15:00 Discussion

15:00 – ca. 18:00 Outing

Friday July 12th

09:00 – 11:30 Session **Databases & Reaction Networks**

Chair: Dr. Christoph Scheurer

09:00 – 09:45 **Prof. Dr. Johannes Margraf** (Universität Bayreuth, DE)

Machine Learning in Chemical Reaction Space

09:45 – 10:15 **Dr. Jonathan Schmidt** (ETH Zürich, CH)

Alexandria database: All you need is more data in material science?

10:15 – 10:30 Break

10:30 – 11:00 **Prof. Dr. Olexandr Isayev** (Carnegie Mellon University, USA)

Scaling Molecular Modeling to Millions of Reactions
with Neural Network Potentials

11:00 – 11:30 **Dr. Pierre-Paul De Breuck** (UC Louvain, BE)

Property predictions from limited and multi-fidelity datasets

11:30 – 12:15 Discussion

12:15 – 12:30 Closing

List of Posters

Within a topic, the abstracts are sorted alphabetically based on the first author's family name.

Thermodynamic Observables

1.	**Chen, Lin**	Synthesis of triangulene made simple by single-atom alloys
2.	**Demiroglu, Ilker**	Gaussian approximation potential development for bimetallic Pt-Cu nanoparticles

Electronic Structure and Long Range Interactions

3.	**Alibakhshi, A.**	tba
4.	**Aryal, N.**	Efficient ensemble averaging methods to study electronic structure at finite temperature from first principles calculations using neural network
5.	withdrawn	
6.	**Febrer Calabozo, P.**	Learning the density matrix, a symmetry rich encoding of the electronic density
7.	**Friede, M.**	dxtb - an efficient and fully differentiable framework for extended tight-binding
8.	**Froitzheim, T.**	GP3-xTB: a general purpose self-consistent tight-binding quantum chemical method
9.	**Lou, Z.**	Predicting the electronic densities and nuclear dynamics of 2d materials
10.	**Vinod, V.**	Multifidelity machine learning for quantum chemical properties
11.	**Vondrák, M.**	Charge equilibration in machine learning potentials
12.	**Zhao, H.**	Deep learning functionals based on Møller-Plesset adiabatic connection for non-covalent interactions

Magnetic Observables

13.	**Hele, T.**	Machine learning spin-pure excited states of organic radicals
14.	**Polak, E.**	Applying a well-defined energy density for machine-learned density functionals

Spectroscopic Observables

15.	**Chen, R.**	Uncovering the potential of applying principal component analysis (PCA) on Raman spectra of biochar
16.	**Daniel, D. T.**	Machine learning isotropic g values of organic battery materials
17.	**Das, B.**	Theoretical nonlinear vibrational spectroscopy of water in

	slit pores
18. **Grunert, M.**	Predicting the spectra of semiconductors and insulators
19. **Jähnigen, S.**	Vibrational circular dichroism
20. **Jindal, A.**	Ion effects on terahertz spectra of microsolvated clusters
21. **König, P.**	Complexions at the electrolyte/electrode interface in solid oxide cells
22. **Walker, M.**	Can machine learning accelerate the discovery of new photovoltaics?
23. **Willimetz, D.**	Modelling tensorial properties of zeolites using machine learning
24. **Yaman, M.**	Decoding the fucose migration products in blood group epitopes

Databases & Reaction Networks

25. withdrawn	
26. **Ketkaew, R.**	Unsupervised Collective Variables for Exploring Expanded Configurational Space
27. **Van Veerdeghem, J.**	Modeling the potential energy surfaces of the $[H_2O - Kr]^+$ system with gaussian process regression

Mechanical Observables

28. **Clovin, N.**	Designing the phase stability in metal-organic frame works and metal halide perovskites: Two sides of the same strain coin?

Abstracts

Abstracts for Talks

Monday July 8th

Out of the crystalline comfort zone: tackling working interfaces with machine learning

Karsten Reuter
Fritz-Haber-Institut der MPG, Germany

Machine learning (ML) promises a significant enhancement of multi-scale modeling capabilities in the context of energy conversion and storage (ECS). In particular, ML interatomic potentials (MLIPs) trained with first-principles data already offer orders of magnitude speed-ups in the computation of predictive-quality energies and forces in atomic-scale simulations. This new efficiency finally allows to heads-on tackle the highly dynamic evolution of working interfaces in ECS systems, where the targeted functionality like catalytic activity or ion mobility both inherently drives and results from ongoing substantial structural, compositional and morphological changes. Unable to fully capture such operando evolution, direct first-principles based multiscale modeling focused hitherto on model (single-)crystalline surfaces or interfaces, where the system dynamics was typically restricted to select reacting or diffusing species that were considered central for a targeted primary function. The MLIP-enabled enhanced sampling capabilities instead allow to assess the thermodynamic stability of complex, possibly amorphous configurations and thereby establish reliable structural models for the working interfaces. Automated process exploration in turn provides more systematic access to the elementary steps that drive the operando evolution, paving the way for microkinetic simulations that analyze the entanglement of this evolution with the primary function. Obviously, these advances in modeling capabilities now call for ever more ML methodology, to efficiently analyze (trajectory) data from simulation cells containing thousands of atoms, to compute spectroscopic fingerprints or to extract refined descriptors for operando-evolution aware high-throughput virtual screening.

Machine learning opens a wonderland for looking through glasses

Michele Simoncelli
Theory of Condensed Matter Group of the Cavendish Laboratory, University of Cambridge (UK), United Kingdom

The relationship between atomistic structure and macroscopic properties of solids has been intriguing scientists since the development of quantum theory, and plays a critical role in many and diverse technologies. This talk will discuss how advancements in machine learning atomic force fields, and the recently developed unified theory of thermal transport in crystals and glasses [1,2], allow us to predict with quantum accuracy the vibrational and thermal properties of glassy solids with arbitrary composition and structural disorder [3]. I will outline some of our efforts on this topic: elucidating how controlling disorder in the atomistic bond topology of network solids allows us to engineer their macroscopic thermomechanical properties [4], describing defects and hot spots in perovskites solid solutions, and predicting the structure and heat conductivity of hybrid materials with crystal-like bond topology and glass-like bond geometry [5].

*[1] M. Simoncelli, N. Marzari, F. Mauri, Nat. Phys., **15**, 809-813 (2019)*
*[2] M. Simoncelli, N. Marzari, F. Mauri, Phys. Rev. X, **12**, 041011 (2022)*
*[3] M. Simoncelli, F. Mauri, N. Marzari, npj. Comput. Mater., **9**, 106 (2023)*
*[4] A. Harper, K. Iwanowski, W. Witt, M. Payne, M. Simoncelli, Phys. Rev. Materials, **8**, 043601 (2024)*
[5] M. Simoncelli, D. Fournier, M. Marangolo, E. Balan, K. Béneut, B. Baptiste, B. Doisneau, N. Marzari, and F. Mauri, 10.48550/arXiv.2405.13161 (2024).

Accelerating transport coefficients predictions via machine learning

Christian Carbogno
Fritz-Haber-Institut der Max-Planck-Gesellschaft, Germany

Heat and charge transport are ubiquitous phenomena in material science that play a pivotal role in a myriad of scientific and industrial applications. These macroscopic, thermodynamic transport phenomena are determined by subtle details in the microscopic nuclear dynamics and, on top of that, by the coupling of electronic and nuclear degrees of freedom. Especially at high temperatures and/or for system determined by non-trivial interactions, higher-order couplings –so called strongly anharmonic and vibronic effects– can become decisive for accurately describing macroscopic transport [1].
In this talk, we utilize heat transport as an example and first discuss how well-balanced data-sets that cover all relevant strongly anharmonic effects can be built from fully anharmonic *ab initio* molecular dynamics [2]. Second, we discuss how these improves and streamlines the training of a machine-learning interatomic potential (MLIP). Such an MLIP can help reaching the time and length scales necessary to describe transport phenomena; however, it requires specific adaptations for the evaluation of transport coefficients [3]. Third, we analyze how symbolic regression in combination with sensitivity analysis can accelerate and guide *in silico* materials' space exploration for optimal transport coefficients [4]. Eventually, we will discuss how the proposed strategies can be generalized to cover even more complex transport phenomena, including charge transport.

*[1] M. Simoncelli, N. Marzari, F. Mauri, Phys. Rev. X, **12**, 041011 (2022)*
*[2] F. Knoop, T. Purcell, M. Scheffler, C. Carbogno, Phys. Rev. Lett., **130**, 236301 (2023)*
*[3] M. Langer, F. Knoop, C. Carbogno, M. Scheffler, M. Rupp, Phys. Rev. B, **108**, L100302 (2023)*
*[4] T. Purcell, M. Scheffler, L. Ghiringhelli, C. Carbogno, npj. Comput. Mater., **9**, 112 (2023)*

Harnessing machine learning for advancing amorphous battery materials

Nong Artrith[1], H. Guo[2], C. Cao[3], M. R. Carbone[3], S. Yoo[3], F. Wang[4], D. Lu[3], A. Urban[2]
[1]Debye Institute for Nanomaterials Science, Netherlands
[2]Department of Chemical Engineering, Columbia University, New York,, United States
[3]Brookhaven National Laboratory, Upton, New York, United States
[4]Applied Materials Division, Argonne National Laboratory, Lemont, United States

Many materials with applications in energy, e.g., batteries, are non-crystalline, exhibiting amorphous structures, chemical disorder, and complex compositions. This complexity makes direct modeling with first principles methods challenging. To address this challenge, we developed accelerated sampling strategies based on machine learning interatomic potentials, genetic algorithms, and molecular-dynamics simulations [1]. Here, I will discuss the methodology and its applications to amorphous battery materials. We constructed the phase diagram of amorphous LiSi alloys, which are prospective anode materials for lithium-ion batteries [2]. Additionally, we mapped the composition and structure space of amorphous lithium thiophosphate (LPS) solid electrolytes [3-5]. The thermodynamic stability and ionic conductivity of the non-crystalline phases were correlated with local structural motifs, leading to the identification of structure-composition-conductivity relationships that can be used for materials optimization and design. X-ray absorption spectroscopy (XAS) characterizes materials, revealing details of the absorber atom's local chemistry. Our work created an S/P K-edge XAS spectra database for LPS materials using structures from [3]. This study presents the initial atomic-scale insights into the oxidative degradation of LPS electrolytes, guiding macroscopic reactions via microstructural engineering and enhancing sulfide electrolyte design.

*[1] N. Artrith, A. Urban, Computational Materials Science, **114**, 135-150 (2016)*
*[2] N. Artrith, A. Urban, G. Ceder, The Journal of Chemical Physics, **148**, (2018)*
*[3] H. Guo, Q. Wang, A. Urban, N. Artrith, Chem. Mater., **34**, 6702-6712 (2022)*
*[4] H. Guo, M. Carbone, C. Cao, J. Qu, Y. Du, S. Bak, C. Weiland, F. Wang, S. Yoo, N. Artrith, A. Urban, D. Lu, Sci. Data., **10**, 349 (2023)*
*[5] C. Cao, M. Carbone, C. Komurcuoglu, J. Shekhawat, K. Sun, H. Guo, S. Liu, K. Chen, S. Bak, Y. Du, C. Weiland, X. Tong, D. Steingart, S. Yoo, N. Artrith, A. Urban, D. Lu, F. Wang, Cell Reports Physical Science, **5**, 101909 (2024)*

Tuesday July 9th

A foundational atomistic model for materials

Gábor Csányi

University of Cambridge, UK

A new computational task has been defined and solved over the past 15 years for extended material systems: the analytic fitting of the Born-Oppenheimer potential energy surface as a function of nuclear coordinates. The resulting potentials ("force fields") are reactive, many-body, with evaluation costs that are currently on the order of 0.1-10 ms/atom/cpu core (or about 1-10ms on a powerful GPU), and reach accuracies of a few meV/atom when trained specifically for a given system using iterative or active learning methods. The latest and most successful architectures leverage many-body symmetric descriptions of local geometry and equivariant message passing networks. Perhaps the most surprising recent result is the stability of models trained on very diverse training sets across the whole periodic table. Our recently discovery is that the MACE-MP-0 model that was trained on just ~150,000 real and hypothetical small inorganic crystals (90% of training set < 70 atoms), is capable of stable molecular dynamics on any system tested so far - this includes crystals, liquids, surfaces, clusters, molecules, and combinations of all of these. The astounding generalisation performance of such foundation models open the possibility to creating a universally applicable interatomic potential with useful accuracy (especially when fine-tuned with a little bit of domain-specific data), and democratise quantum-accurate large scale molecular simulations by lowering the barrier to entry into the field. Furthermore, there is a wide ranging efforts to train models that can predict observables on top of the MD trajectories, encompassing a wide range of spectroscopy (certainly in terms of wavelength!), from X-ray to infrared absorption and NMR.

High-throughput approaches for materials understanding and design

Janine George

Federal Intitute for Materials Research and Testing in Germany, Germany

Bonds and local atomic environments are crucial descriptors of material properties. They have been used to create design rules and heuristics and as features in machine learning of materials properties.[1] Implementations and algorithms (e.g., ChemEnv and LobsterEnv) for identifying local atomic environments based on geometrical characteristics and quantum-chemical bonding analysis are nowadays available.[2,3] Fully automatic workflows and analysis tools have been developed to use quantum-chemical bonding analysis on a large scale.[3,4] The lecture will demonstrate how our tools, that assess local atomic environments and perform automatic bonding analysis, help to develop new machine learning models and a new intuitive understanding of materials.[5,6] Furthermore, the general trend toward automation in computational materials science and some of our recent contributions will be discussed.[7–10]

*[1] J. George, G. Hautier, Trends in Chemistry, **3**, 86-95 (2021)*
*[2] D. Waroquiers, J. George, M. Horton, S. Schenk, K. Persson, G. Rignanese, X. Gonze, G. Hautier, Acta. Crystallogr. Sect. B., **76**, 683-695 (2020)*
*[3] J. George, G. Petretto, A. Naik, M. Esters, A. Jackson, R. Nelson, R. Dronskowski, G. Rignanese, G. Hautier, ChemPlusChem, **87**, (2022)*
*[4] A. Naik, K. Ueltzen, C. Ertural, A. Jackson, J. George, JOSS., **9**, 6286 (2024)*
*[5] A. Naik, C. Ertural, N. Dhamrait, P. Benner, J. George, Sci. Data., **10**, 610 (2023)*
[6] K. Ueltzen, A. Naik, C. Ertural, P. Benner, J. George, in preparation (2024)
*[7] J. George, Trends in Chemistry, **3**, 697-699 (2021)*
[8] A. Ganose, et al., “atomate2,” (2023)
*[9] A. Rosen, M. Gallant, J. George, J. Riebesell, H. Sahasrabuddhe, J. Shen, M. Wen, M. Evans, G. Petretto, D. Waroquiers, G. Rignanese, K. Persson, A. Jain, A. Ganose, JOSS., **9**, 5995 (2024)*
[10] I. Batatia, et al., 10.48550/arXiv.2401.00096 (2023)

Challenging the dogma of rotational equivariance in atomistic machine learning

Sergey Pozdnyakov

École polytechnique fédérale de Lausanne, Switzerland

In the last decade, methods rigorously incorporating rotational symmetry into their functional form have dominated the field of atomistic machine learning. Two main considerations have driven this design choice. First, it is widely believed that rotational symmetry, when intrinsically built into a machine learning architecture, serves as a necessary inductive bias or physical preconditioning, thus being crucial for a model's performance and generalizability. Second, rigorous rotational equivariance per se is believed to be essential for atomistic simulations such as molecular dynamics to avoid subtle artifacts.

In this talk, I will challenge the first belief by presenting an unconstrained model, Point Edge Transformer (PET), which is not rotationally equivariant but achieves state-of-the-art performance on multiple benchmark datasets of molecules and solids, covering both invariant targets such as potential energy and covariant ones such as vectorial dipole moments. To address the second concern, we introduce a general symmetrization method that enforces rigorous rotational equivariance a posteriori for any backbone architecture.

However, our exact symmetrization scheme incurs additional computational costs and might be superfluous, as PET inherently learns approximate rotational equivariance with typical discrepancies as small as of order 1e-4 eV/atom. In the final part of my talk, I aim to foster an interactive discussion on whether such rotational discrepancies can drive malignant collective behavior in practical simulations and what tests are needed to evaluate this.

Leveraging computational advances to design and optimize energy materials: from traditional methods to machine learning

Kulbir Ghuman
Institut National de la Recherché, Canada

The functionality of the materials used for energy applications is critically determined by the physical properties of small active regions such as dopants, dislocations, interfaces, grain boundaries, etc. The capability to manipulate and utilize the inevitable disorder in materials, whether due to the finite-dimensional defects (such as vacancies, dopants, grain boundaries) or due to the complete atomic randomness (as in amorphous materials), can bring innovation in designing energy materials. With the increase in computational material science capabilities, it is now possible to understand the complexity present in materials due to various defects resulting in pathways required for optimizing their efficiencies. In this talk, I will provide a critical overview of such computational advancements specifically for designing materials for sustainable ammonia synthesis and CO_2 capture technologies. I'll provide a comprehensive review of our recent research efforts, which involve employing traditional methods like density functional theory (DFT), alongside leveraging the data they generate to implement machine learning techniques, thereby accelerating the materials discovery for these vital applications.

Machine-learning for electronic structure

Michele Ceriotti
EPFL, Switzerland

Atomic-scale simulations of materials and condensed-matter systems have been transformed by the application of machine learning potentials, that facilitate and greatly reduce the computational cost for predicting their structure and (thermo)dynamics. In order to also determine functional properties, and more in general to extend further the scope of these simulations, it is desirable to develop machine-learning models that target quantities that are more intimately connected with the electronic structure -- the charge density, the electron density of states, the electronic excitations.

I will present a few examples of this kind of models, and discuss in particular how to construct "hybrid" frameworks that combine data-driven elements with physically-motivated components. For example, I will demonstrate the use of a model of the ground-state electronic density of states to perform simulations at finite electron temperature, and the use of a minimal-basis Hamiltonian as an intermediate step in a model architecture targeting excited-state properties.

Advanced software frameworks for describing local and non-local interactions in high-dimensional neural network potentials

Alexander Knoll, Jörg Behler
Ruhr Universität Bochum, Germany

Machine learning potentials (MLPs) have emerged as powerful tools for the accurate description of realistic systems containing large numbers of atoms. A well established example for MLPs are high-dimensional neural network potentials (HDNNPs), which leverage the fitting abilities of neural networks to express the potential energy surface (PES) of a system. Recently, advanced models like fourth-generation HDNNPs (4G-

HDNNP)[1] have demonstrated the ability to capture interactions extending beyond the short-range environment of an atom, including dispersion, electrostatics and non-local charge transfer. This talk will emphasise the importance of these interactions, highlighting the advancements achieved across all four generations of HDNNPs.[2] The increasing sophistication of these advanced MLPs brings about a corresponding rise in complexity when managing training data, controlling workflows and assessing results. This, in turn, necessitates robust tools which automate the training and validation process and ensure the reproducibility of the final results. Addressing this need, we present modern software frameworks, RuNNer 2.0 and runnerase, as solutions to these growing challenges. RuNNer 2.0 is an open-source, stand-alone software package for the construction and evaluation of second-, third-, and fourth-generation HDNNPs. It unifies the entire workflow in a fully MPI-parallel program: from the evaluation of atomistic descriptors, through the training of diverse properties like atomic spins, charges, energies and forces, to their final application in molecular dynamics simulations. We demonstrate how RuNNer 2.0's incorporation of modern fitting strategies can significantly improve training quality and performance. RuNNer 2.0 is conveniently accessible via Python through the ASE[3]-based framework runnerase. This tool automates the construction of descriptors and hyperparameter searches, streamlines simulation control, and provides active learning workflows. Moreover, runnerase extends the use of RuNNer to all applications covered by the ASE ecosystem, such as its integration into the materials science IDE pyiron.[4] Together, these tools enable the efficient construction of HDNNPs that consider both local and non-local atomic interactions in complex systems.

[1] T. Ko, J. Finkler, S. Goedecker, J. Behler, Nat. Commun., ***12****, 398 (2021)*
[2] J. Behler, Chem. Rev., ***121****, 10037-10072 (2021)*
[3] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. Castelli, R. Christensen, M. Dułak, J. Friis, M. Groves, B. Hammer, C. Hargus, E. Hermes, P. Jennings, P. Bjerre Jensen, J. Kermode, J. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, K. Jacobsen, J. Phys.: Condens. Matter, ***29****, 273002 (2017)*
[4] J. Janssen, S. Surendralal, Y. Lysogorskiy, M. Todorova, T. Hickel, R. Drautz, J. Neugebauer, Computational Materials Science, ***163****, 24-36 (2019)*

Electronic structure surrogate learning for quantum dynamics and inverse design

Reinhard Maurer
University of Warwick, United Kingdom

Atomistic simulation based on quantum mechanics (QM) is currently being revolutionized by machine-learning (ML) methods. Many existing approaches use ML to predict molecular properties from quantum chemical calculations. This has enabled molecular property prediction within vast chemical compound spaces and high-dimensional parametrization of energy landscapes for the efficient simulation of measurable observables. However, as all properties derive from the QM wave function, an ML model that can predict the wave function also has the potential to predict other properties. In this talk, I will explore ML approaches that deliver surrogate models of the electronic structure based on the QM Hamiltonian with the aim to develop methods that use ML and QM in synergy. Using example systems from heterogeneous catalysis and organic electronics, I will discuss the challenges associated with encoding physical symmetries and invariance properties into machine learning models of electronic structure. Upon overcoming these challenges, integrated ML-QM methods within modern, modular software frameworks offer the combined benefits of data-driven parametrization and first-principles-based methods. I will discuss several opportunities associated with building ML-augmented quantum chemical methods, including Inverse Chemical Design based on ML-predicted wave functions and the development of efficient and accurate surrogate models to study materials chemistry.

ML Electrostatics Models in Relevant Test Systems

William Baldwin
University of Cambridge, United Kingdom

Incorporating electrostatics into machine learning will be essential for modelling electronically active systems such as electrode-electrolyte interfaces or electrocatalysis. Several models have therefore been proposed which include some description of electrostatic effects, or which can predict observables related to charge density. Despite this interest, there is a lack of test cases on which to benchmark electrostatic models. In this case, a 'good' benchmark is one which (i) cannot be properly be tackled with purely short range models, (ii) is relevant to real modelling questions and (iii) separates the performance of different approaches. In this work, we introduce new test systems which fulfill the above requirements. In particular, we look at systems relevant to electrode/electrolyte interfaces. By implementing several previously proposed ideas within in the

the MACE MLIP framework, we analyse how well different approaches can capture the necessary physics. Finally, we present new a architecture which performs well on these examples, suggesting that modelling electronically active systems is within reach.

Wednesday July 10th

The Jacobi-Legendre framework for materials discovery

Stefano Sanvito
Trinity College Dublin, Ireland

Computational methods based on density functional theory (DFT) have served as the workhorse of most electronic structure calculations for decades. The success of DFT is generally rooted in its strong theoretical foundation, in the availability of accurate approximations and on the multitude of numerical implementations constructed over the years. Yet, DFT becomes increasingly computational expensive when extremely accurate energies are required, when one has to deal with large systems or when it is combined with an open-boundary-condition problem (e.g. in quantum transport). Here, I will demonstrate that the integration of machine-learning techniques in DFT can solve some of the issues and effectively extends the DFT range of applications.
I will introduce a compact cluster expansion framework, rooted in Jacobi and Legendre polynomials, to construct highly accurate, efficient and interpretable machine-learning surrogate models [1]. This can be then deployed to the construction of ML-based interatomic potentials (IAPs) as well as for the prediction of the ground-state electronic density, the key variable in all DFT calculations [2]. With this toolkit at hand, I will demonstrate as one can accelerate materials discovery, for example the construction of phase diagrams [3]. Furthermore, the same models can be employed within quantum transport to enormously increase the calculation throughput.

*[1] M. Domina, U. Patil, M. Cobelli, S. Sanvito, Phys. Rev. B, **108**, 094102 (2023)*
*[2] B. Focassio, M. Domina, U. Patil, A. Fazzio, S. Sanvito, npj. Comput. Mater., **9**, 87 (2023)*
*[3] H. Rossignol, M. Minotakis, M. Cobelli, S. Sanvito, J. Chem. Inf. Model., **64**, 1828-1840 (2024)*

Prediction of the magnetic exchange interaction in doped topological insulators

Johannes Wasmer[1], Rubel Mozumder[1], David Antognini Silva[1], Stefan Blügel[1], Philipp Rüßmann[2]
[1]Forschungszentrum Jülich, Germany
[2]Institute of Theoretical Physics and Astrophysics, University of Würzburg, Germany

We present a benchmark study of surrogate models for impurities embedded into crystalline solids. Using the Korringa-Kohn-Rostoker Green Function method and the AiiDA workflow engine [1], we have built a database of magnetic transition metal impurity dimers embedded in the topological insulator Bi_2Te_3. We predict isotropic exchange interaction of the impurity dimer in the classical Heisenberg model with machine learning and then use these surrogates as input for spin dynamics calculations to find the magnetic ground state of the material [2]. The study compares various recent E(3)-equivariant models such as ACE and MACE [3] in terms of performance and reproducible end-to-end workflows.

*[1] P. Rüßmann, F. Bertoldo, S. Blügel, npj. Comput. Mater., **7**, 13 (2021)*
*[2] P. Rüßmann, J. Ribas Sobreviela, M. Sallermann, M. Hoffmann, F. Rhiem, S. Blügel, Front. Mater., **9**, (2022)*
[3] Batatia, I., Kovács, D. P., Simm, G. N. C., Ortner, C. & Csányi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. Preprint (2022).

Machine learning for molecular magnetism

Alessandro Lunghi
Trinity College Dublin, Ireland

Magnetic molecules have been proposed as key elements of novel technologies, ranging from spintronics, quantum sensing and high-density information storage, but their integration with complex environments at room temperature stands out as a main road block. Among the many challenges in this area, one of the most crucial ones is the stabilization of the coherence of magnetic molecules' spin at high temperature. Ab initio simulations of this processes have recently become available [1], but at very high computational costs, making it very hard to thoroughly explore the phenomenology of spin relaxation. In addition to current uncertainties on the nature of spin-lattice interactions, the goal of actually translating new design principles into chemical strategies poses some serious challenges to the field.
In this talk I will discuss how machine learning can be leveraged to speed up the evaluation of those molecular properties required to estimate spin relaxation, thus making it possible increase the throughput of these

simulations [2]. In particular, I will show how molecular normal modes can be efficiently computed through linear machine learning methods combined with active learning [3] and how magnetic properties can be predicted through equivariant models [4].

*[1] A. Lunghi, Sci. Adv., **8**, (2022)*
*[2] A. Lunghi, S. Sanvito, Nat. Rev. Chem., **6**, 761-781 (2022)*
*[3] V. Briganti, A. Lunghi, Mach. Learn.: Sci. Technol., **4**, 035005 (2023)*
*[4] V. Nguyen, A. Lunghi, Phys. Rev. B, **105**, 165131 (2022)*

Machine learning facilitated by microscopic features for discovery of novel magnetic double perovskites

Shuping Guo, Ryan Morrow, Jeroen van den Brink, Oleg Janson
IFW Dresden, Germany

Double perovskites are a growing class of compounds with prospects for realization of novel magnetic behaviors. The rich chemistry of double perovskites calls for high-throughput computational screening that can be followed by or combined with machine-learning techniques. Yet, most approaches neglect the bulk of microscopic information implicitly provided by first-principles calculations, severely reducing the predictive power. In this work, we remedy this drawback by including onsite energies and transfer integrals between the *d* states of magnetic atoms. These quantities were computed by Wannierization of the relevant energy bands. By combining them with the experimental information on the magnetism of studied materials and applying machine learning, we constructed a model capable of predicting the magnetic properties of the remaining materials whose magnetism has not been addressed experimentally. Our approach combines classification learning to distinguish between double perovskites with dominant ferromagnetic or antiferromagnetic interactions and regression employed to estimate magnetic transition temperatures. In this way, we identified one antiferromagnet and three ferromagnets with a high transition temperature. Another 28 antiferromagnetic candidates were identified as magnetically frustrated compounds. Among them, cubic Ba_2LaReO_6 shows the highest frustration parameter, which is further validated by a direct first-principles calculation. Our methodology holds promise for eliminating the need for resource-demanding calculations.

Machine learning for spectroscopy – concepts, successes, and challenges

Patrick Rinke
Aalto University, Finland

Spectroscopy is a fundamental tool in molecular and materials research, characterization, and discovery. It has therefore become a major objective of machine-learning (ML) development but has not received as much attention in the latest artificial intelligence (AI) boom as other experimental or computational techniques. Machine learning for spectroscopy pursues two parallel goals [1]: spectra prediction (typical in computational studies) and property inference (typical in experimental approaches). Successful ML spectra predictions allow us to cut down on the time and resources behind computational or experimental spectroscopy. Trained on available input (e.g., atomic structure or materials attributes) and output (e.g., spectra or spectroscopic quantities) pairs, the ML model can make output predictions for new input instantaneously, without further resource requirements. In property inference tasks, data input and output are reversed to echo spectroscopic applications. ML models predict materials structure and properties from spectral input or classify the inputs into different categories.

In this presentation, I will use the spectroscopy paradigm to present different ways to implement ML for spectral prediction and inference tasks, show successful examples and address remaining challenges. For computational data [2], I will demonstrate that ML models can predict ionization energies and spectra of organic molecules from their atomic structure alone [3,4]. We trained kernel ridge regression and neural network models on quantum mechanically computed molecule-property pairs. The molecules are represented by simple, easily attainable numerical descriptors based on nuclear charges and cartesian coordinates. The complexity of the molecular descriptor and the diversity of the data sets turn out to be crucial for the learning success [3]. Our best AI models predict ionization energies within 0.2 eV and excitation spectra with 97% accuracy surpassing typical measurement uncertainties.

For inference, I will illustrate how we correlate the properties of biopolymers with nuclear magnetic resonance (NMR) spectroscopy. We extracted lignin from birch wood and characterized it with 2D NRM spectroscopy [5]. Using the 2D NRM spectra as input, our AI model infers lignin properties like the antioxidant activity directly from the spectra. We trace the predictions back to NRM peaks to link the antioxidant activity to structural features of lignin encoded in the spectra.

*[1] H. Kulik, T. Hammerschmidt, J. Schmidt, S. Botti, M. Marques, M. Boley, M. Scheffler, M. Todorović, P. Rinke, C. Oses, A. Smolyanyuk, S. Curtarolo, A. Tkatchenko, A. Bartók, S. Manzhos, M. Ihara, T. Carrington, J. Behler, O. Isayev, M. Veit, A. Grisafi, J. Nigam, M. Ceriotti, K. Schütt, J. Westermayr, M. Gastegger, R. Maurer, B. Kalita, K. Burke, R. Nagai, R. Akashi, O. Sugino, J. Hermann, F. Noé, S. Pilati, C. Draxl, M. Kuban, S. Rigamonti, M. Scheidgen, M. Esters, D. Hicks, C. Toher, P. Balachandran, I. Tamblyn, S. Whitelam, C. Bellinger, L. Ghiringhelli, Electron. Struct., **4**, 023004 (2022)*
*[2] A. Stuke, C. Kunkel, D. Golze, M. Todorović, J. Margraf, K. Reuter, P. Rinke, H. Oberhofer, Sci. Data., **7**, 58 (2020)*
*[3] A. Stuke, M. Todorović, M. Rupp, C. Kunkel, K. Ghosh, L. Himanen, P. Rinke, The Journal of Chemical Physics, **150**, (2019)*
*[4] K. Ghosh, A. Stuke, M. Todorović, P. Jørgensen, M. Schmidt, A. Vehtari, P. Rinke, Advanced Science, **6**, (2019)*
*[5] J. Löfgren, D. Tarasov, T. Koitto, P. Rinke, M. Balakshin, M. Todorović, ACS Sustainable Chem. Eng., **10**, 9469-9479 (2022)*

Experiment-driven atomistic materials modeling: combining XPS and MLPS to infer the structure of a-CO$_x$

Tigany Zarrouk
Aalto University, Finland

One of the most important, and most challenging, aspects of atomistic materials modeling is to reconcile experimental and computational results. This requires an effective strategy for identifying atomistic structures whose predicted observables align with the experimentally available data. Conventional approaches involve generating numerous configurations through molecular dynamics or Monte Carlo structure optimization and selecting the one with the closest match to experiment. However, this is an inefficient process which is not guaranteed to succeed. In this paper, we introduce a general method to combine atomistic machine learning (ML) with experimental observables that produces atomistic structures compatible with the experimental observables by design. We use this approach in combination with grand-canonical Monte Carlo within a modified Hamiltonian formalism, to generate configurations that agree with both experimental data and are chemically sound (i.e., low in energy). We apply our approach to understand the atomistic structure of oxygenated amorphous carbon (a-CO$_x$), an intriguing carbon-based material, to answer the question of how much oxygen can be added to carbon before it fully decomposes into CO and CO2. Utilizing an ML-based X-ray photoelectron spectroscopy (XPS) model trained from GW and density functional theory (DFT) data, in conjunction with an ML interatomic potential for the C-O system, we identify a-CO$_x$ structures compliant with experimental XPS predictions that are also energetically favorable with respect to DFT. Employing a network analysis, we obtain an accurate deconvolution of the XPS spectrum into motif contributions, both revealing the inaccuracies inherent to experimental XPS interpretation and granting us atomistic insight into the structure of a-CO$_x$. This method generalizes to multiple sets of experimental data and allows for the elucidation of the atomistic structure of materials directly from experimental results, thereby enabling experiment-driven materials modeling with a degree of realism previously out of reach

Categorizing three-dimensional photonic crystals: open challenges in scale-covariant problems

Rose Cersonsky
University of Wisconsin, United States

Many butterflies, birds, beetles, and chameleons owe their spectacular colors to the microscopic pattern of dielectric materials within their wings, feathers, or skin. When these patterns, or photonic crystals, result in the omnidirectional reflection of commensurate wavelengths of light, it is due to a complete photonic band gap (PBG). **So, what such patterns will lead to a photonic band gap?** While this appears to be a straightforward question, it is, in truth, difficult to answer. Ordered photonic band gap crystals were first theorized in the early 1990s, and the efficient computation of photonic band structures for three-dimensional patterns occurred much later. Thus, early studies aimed to define design principles through analyses of phenomena in two-dimensional systems or from limited three-dimensional calculations. While these general principles of design hold for many cases, the space of potential photonics targets is far wider than they suggest, and thus, we look to machine learning to derive new insight into this structure-property landscape. This talk will focus on the challenges of deriving surrogate models for photonic structures, largely due to the scale-covariance of their governing equations.

p-DOS: a descriptor with electronic wisdom for learning X-Ray spectroscopy

Clelia Middleton
Newcastle University, United Kingdom

Subsequent to developing a successful methodology for the prediction of X-Ray spectra using machine learning, we introduce p-DOS, a novel descriptor to represent the electronic properties of species. Within applied machine learning, several approaches to the prediction of spectral characteristics have been developed. Penfold et al. introduced the machine learning architecture XANESNET, where success has been demonstrated when predicting the K- and L-edges of various transition metals from a purely geometric representation. [1]

When moving to domains where the transitions have significant electronic character, such as with XES, or where the absorption edges of light non-metals are investigated, a representation which correspondingly encodes electronic character is required. To produce accurate electronic representations while maintaining the speed and efficiency with which machine learning is able to produce results, we produce a descriptor based on a partial Density-of-States representation [2] from a guess wavefunction, and apply this descriptor using the previously developed XANESNET architecture. This p-DOS representation is able to accurately predict the K-edge of sulphur complexes, as demonstrated with tests executed on a dataset extracted from the GDB-13 database.

We further demonstrate the utility of our novel representational method by applying it to an interesting test case; the photostimulated ring-opening of the simple ring system thiophenone. The final population distribution of the photoproducts of this system is presently an open question. [3] Machine learning predictions using the p-DOS descriptor recommend that near-edge X-Ray absorption spectroscopy would reveal an informative peak, enabling the populations of previously indifferentiable species to be quantified. Our algorithm acts here as a “pocket oracle”, enhancing the arsenal of experimentalists and aiding their capacity to make informed decisions about the potential value of dedicating valuable beamtime to a particular line of investigation at minimal computational cost when compared to first-principles calculations.

Provisional results applying p-DOS to other light non-metals and investigations of the utility of applying this generalist (type II) model in the time-resolved domain will also be shown.

[1] C. Rankine, T. Penfold, The Journal of Chemical Physics, ***156****, (2022)*
[2] C. Middleton, B. Curchod, T. Penfold, Partial Density of States Representation for Accurate Deep Neural Network Predictions of X-ray Spectra, 2024
[3] D. Murdock, S. Harris, J. Luke, M. Grubb, A. Orr-Ewing, M. Ashfold, Phys. Chem. Chem. Phys., ***16****, 21271-21279 (2014)*

Thursday July 11th

Interpreting core-loss spectroscopy

Rebecca Nicholls
University of Oxford, UK

Core-loss spectroscopy, produced by electron or X-ray beams, has the potential to tell us about the bonding and structure of a material. Experimental spectra can be difficult to interpret and simulation potentially allows us to make a link between spectral features and bonding environment. Just using simulation to replicate experimental data does not always help us and I will discuss the importance and challenge of keeping the scientific question at the centre of the work when combining simulation and experiment.

Predicting electron paramagnetic resonance parameters and their sensitivity to structural configuration

Josef Granwehr
Forschungszentrum Jülich GmbH and RWTH Aachen University, Germany

Organic radical batteries (ORBs) represent a more sustainable and resource-friendly alternative to conventional Li–ion batteries, since their cathodes consist of an organic radical polymer rather than transition metal oxides. As ions are transported through the polymer during charge/discharge cycling of a battery, radical polymers experience recurring operational stress that requires tedious material optimisation with long turnaround times. Support from computational methods would be highly desirable, yet computationally capturing redox processes, which are intrinsic for the operation of batteries, requires non-standard methods, customised for this particular use case. Moreover, for testing the suitability and accuracy of a computational approach, benchmarking using experimental data is necessary.

Electron paramagnetic resonance (EPR) is a spectroscopic technique to investigate the interaction of electron spin from unpaired electrons, such as radicals or transition metal ions, with magnetic fields. The obtained coupling parameters provide structural and electronic information regarding the paramagnetic centre. Spectro-electrochemical *in operando* EPR showed a good contrast of the radical g value as a function of the state-of-charge (SoC) of an ORB. However, in EPR simple relations between spectroscopically determined coupling parameter variations and sample properties are generally less straightforward than for the more widespread nuclear magnetic resonance spcetroscopy. Therefore, a common task for computational chemistry is the model confirmation by *ab initio* calculation of coupling parameters and comparison with experiment. While density functional theory (DFT) methods have been developed to accurately predict electron g values or hyperfine coupling constants for crystalline solids or in solution, non-crystalline solids such as polymers without recurring conformational motifs are more challenging to compute. Large systems would be required that are beyond the capability of common DFT codes, and EPR observables are not directly accessible from MD simulations. Nevertheless, for organic radical polymers, the combination of MD simulations with DFT computations showed good predictions of g values as a function of the radical density, which itself depends on the battery SoC.[1]

g is generally a tensorial quantity. However, ORBs require a high local electron spin concentration to achieve competitive capacity, where Heisenberg spin exchange between spatially close radical centres causes an averaging of anisotropic interactions and results in an isotropic g value. Experimentally it was found that g is concentration dependent, hence not only the global radical concentration can be determined via the EPR signal intensity, but also heterogeneities of local concentrations during cycling of an ORB can be monitored. In addition, the obtained contrast in g may be employed for benchmarking of numerical simulations. An MD model can be assessed whether it relaxes towards a structure that provides the correct experimental g value, using MD snapshots as input structure for DFT calculations.[1]

A computationally more efficient approach is provided by machine learning (ML) of isotropic g values. After validating an MD simulation approach, its trajectory provided an abundance of structures as input for DFT computations. These then served as data for training and testing of an ML model based on regression trees.[2] Different radical densities or, analogously, different SoC were obtained by varying the number of monomers that were radicals. The difference between DFT-derived and ML-predicted g values was found to be on the same order of magnitude or below the uncertainty of experimentally derived g values obtained from *in operando* EPR measurements of a PTMA-based organic radical battery. The ML model was sensitive to the radical density and learnt to predict g values also for radical densities that were not part of the training data set. Furthermore, different molecular descriptors were compared, showing variations in performance.

Once a coupling parameter has been identified experimentally that shows good contrast upon sample variation, a subsequent step may be the identification of fundamental relationships with atomic configuration. This typically requires experimental empiricism or systematic theoretical studies. As an alternative, a statistical approach is discussed to gauge the importance of structure parameters, such as bond lengths, bond angles or dihedral angles, using hyperfine coupling constants as an example.[3] In EPR, hyperfine coupling constants describe measurable electron–nuclear interactions. For some organic radicals, importance quantifiers were computed with the neighborhood components analysis ML algorithm, also using molecular dynamics trajectory snapshots. Atomic–electronic structure relationships are visualized in matrices correlating structure parameters with coupling constants of all magnetic nuclei in a molecule. It is shown that qualitatively, the results adequately reproduce common hyperfine coupling models.

[1] D. Daniel, S. Oevermann, S. Mitra, K. Rudolf, A. Heuer, R. Eichel, M. Winter, D. Diddens, G. Brunklaus, J. Granwehr, Sci. Rep., ***13****, 10934 (2023)*
[2] D. Daniel, S. Mitra, R. Eichel, D. Diddens, J. Granwehr, J. Chem. Theory Comput., ***20****, 2592-2604 (2024)*
[3] C. Szczuka, R. Eichel, J. Granwehr, RSC Adv., ***13****, 14565-14574 (2023)*

Assessing spectroscopic features: from fingerprinting to predictions

Claudia Draxl
Humboldt-Universität zu Berlin, Germany

Comparing spectra can be viewed either as a standard, everyday procedure in laboratories or as a challenging task with the potential for scientific discovery. Differences in spectra can arise from many sources, such as experimental conditions (temperature, pressure, UHV, etc.), defects or structural parameters, or even a chemical reaction. Likewise, uncertainties in computed spectra can come from the underlying methodology or approximations, computational parameters, or structural input. Interpreting such results requires tools beyond pure chemical intuition. A first step in this direction is to define what we mean by similarity of spectra and metrics that allow to “measure” it. By developing a spectral fingerprint [1], we could demonstrate the impact of various parameters on spectral features of computed and measured data [1,2]. At the same time, our approach allowed us to assess data quality on a quantitative level. Combined with (un)supervised learning, spectral fingerprints can also be used to explore large data spaces for patterns or trends, leading to rather unexpected findings [3].

[1] M. Kuban, S. Rigamonti, C. Draxl, 10.48550/arXiv.2403.10470 (2024)
[2] M. Kuban, S. Rigamonti, M. Scheidgen, C. Draxl, Sci. Data., ***9****, 646 (2022)*
[3] M. Kuban, Š. Gabaj, W. Aggoune, C. Vona, S. Rigamonti, C. Draxl, MRS Bulletin, ***47****, 991-999 (2022)*

Scientific machine learning and explainable AI approaches for the physical sciences

Stefan Sandfeld
Forschungszentrum Jülich, Germany

Scientific applications of machine learning and artificial intelligence often pose entirely different challenges than applications typically considered in computer science. Data for scientific problems can extremely be noisy or sparse due to a high (computational) cost of simulations or experiments. In other situations, it might be easy to obtain large amounts of data but creating accurate models that generalize well and that additionally are computationally even more efficient than classical methods is still very challenging.

In this presentation we discuss this challenge from the perspective of physical sciences and introduce solutions for chosen problems. An additional focus is on the development of domain-agnostic ML methods to avoid the construction of highly specialized solutions: (i) a way to circumvent the necessity to hand-label microscopy images, (ii) an explainable AI approach for creating high-accuracy number encodings for generative models, and (iii) an outlook to the next generation of scientific AI models, i.e., foundation models -- and how they even might be able to mitigate the problem of sparse data.

Machine learned small polaron dynamics

Luca Leoni[1], Cesare Franchini[1], Viktor C. Birschitzky[2], Michele Reticcioli[2]
[1]University of Bologna, Italy
[2]University of Vienna, Austria

Polarons are charged quasiparticles that form in polarizable materials. These particles play a significant role in important phenomena such as charge transfer, electron-hole recombination, and catalytic processes [1].

Different methods have been used to study the equilibrium properties of these states [2]. However, their dynamics can only be studied through standard ab initio molecular dynamics (AIMD) [3], which limits both the time scales and system sizes of the simulation. In recent years, machine learning potentials (MLPs) have helped bridge the gap between time- and length-scales in AIMD. In this study, we develop a MLPs architecture using an equivariant graph neural network that successfully accounts for small polaron hopping dynamics at the nanosecond timescale. We apply the proposed ML framework to study hole and electron polaron dynamics in MgO and (F-doped) TiO_2, which enables us to estimate the polaron mobility at different temperatures and study the influence of dopants on the formation and dynamics of polarons.

*[1] M. Reticcioli, I. Sokolović, M. Schmid, U. Diebold, M. Setvin, C. Franchini, Phys. Rev. Lett., **122**, 016805 (2019)*
*[2] C. Franchini, M. Reticcioli, M. Setvin, U. Diebold, Nat. Rev. Mater., **6**, 560-586 (2021)*
*[3] P. Kowalski, M. Camellone, N. Nair, B. Meyer, D. Marx, Phys. Rev. Lett., **105**, 146405 (2010)*

Applying SE(3)-Equivariant Attentional Graph Neural Networks for the purpose of predicting the electronic structure of molecular hydrogen

Bartosz Brzoza
CASUS/HZDR, Germany

In this work, we demonstrate the efficacy of a neural network model implemented as the Materials Learning Algorithms (MALA) package in predicting the electronic structure of a system of hydrogen molecules under various pressure and temperature conditions across the molecular liquid-solid phase boundary, demonstrating the potential of our methods for molecular systems. Additionally, we investigate the use of SE(3)-Transformer Graph Neural Networks to improve the generalizability and extrapolation capabilities of our models. Our results indicate that the MALA framework provides a powerful and efficient tool for accelerating Kohn-Sham density functional theory calculations in molecular systems. This work paves the way for future research in developing advanced machine-learning algorithms for accelerating electronic structure calculations both accurately and efficiently.

Friday July 12th

Machine Learning in Chemical Reaction Space

Johannes T. Margraf
University of Bayreuth, Germany

In this talk, we discuss the challenges associated with developing ML models for chemical reaction space.[1] The special properties of chemical reaction networks have significant implications for data collection and model training. This even applies to the relatively simple task of predicting reaction energies. Here, the influence of important ‘hub’ molecules can strongly drive prediction errors.[2] Furthermore, the conformational space of the reactants needs to be adequately explored using global optimization techniques. Kinetic properties (e.g. rate constants) are even more challenging to predict, as they are not simple molecular properties.[3] Here, active learning strategies must incorporate rare-event sampling, and convergence of the target observable must be checked carefully.[4] To accelerate this process, foundation models for chemical applications (trained on large reference databases) have recently been shown to be remarkably transferable across different domains.[5] This opens the door towards finetuning based strategies for data-efficient chemical machine learning.

*[1] J. Margraf, H. Jung, C. Scheurer, K. Reuter, Nat. Catal., **6**, 112-121 (2023)*
*[2] S. Stocker, G. Csányi, K. Reuter, J. Margraf, Nat. Commun., **11**, 5505 (2020)*
*[3] H. Jung, L. Sauerland, S. Stocker, K. Reuter, J. Margraf, npj. Comput. Mater., **9**, 114 (2023)*
*[4] S. Stocker, H. Jung, G. Csányi, C. Goldsmith, K. Reuter, J. Margraf, J. Chem. Theory Comput., **19**, 6796-6804 (2023)*
[5] I. Batatia et al., 10.48550/arXiv.2401.00096 (2024)

Alexandria database: all you need is more data in material science?

Jonathan Schmidt[1], Noah Hoffman[2], Hai-Chen Wang[3], Aldo Romero[4], Ludger Witz[5], Pedro Borlido[6], Pedro JMA Carriço[6], Tiago FT Cerqueira[6], Silvana Botti[3], Miguel Marques[3]
[1]ETH Zurich, Switzerland
[2]Department of Physics, Humboldt University Berlin, Berlin, Germany
[3]Interdisciplinary Centre for Advanced Materials Simulation, Ruhr University Bochum, Bochum, Germany
[4]Department of Physics and Astronomy West Virginia University, United States
[5]Department of Physics, University of Luxembourg, Luxembourg
[6]Department of Physics, University of Coimbra, Coimbra, Portugal

We develop and apply crystal graph attention networks along with other universal machine learning models to scan billions of materials for thermodynamic stability. By validating these predictions with density functional theory (DFT), we have curated a new database comprising 4.5 million materials, including a convex hull of over 115,000 materials. We observe that each iteration of the database and its associated ML models enhances the efficiency of our high-throughput searches. To improve the accuracy of DFT calculations for thermodynamic stability, we have added SCAN and PBEsol databases to the Alexandria database, which focus on stable and metastable materials. Additionally, we examine the dependency of prediction errors when applying transfer learning from PBE to these datasets. Transfer learning has also enabled us to expand our high-throughput searches to include two-dimensional and one-dimensional materials, significantly broadening the scope of the database.
The database now serves as a benchmark for evaluating the performance of existing universal force fields and other universal machine learning models, and facilitates the development of new models.

*[1] J. Schmidt, N. Hoffmann, H. Wang, P. Borlido, P. Carriço, T. Cerqueira, S. Botti, M. Marques, Advanced Materials, **35**, (2023)*
*[2] H. Wang, J. Schmidt, M. Marques, L. Wirtz, A. Romero, 2D Mater., **10**, 035007 (2023)*
*[3] N. Hoffmann, J. Schmidt, S. Botti, M. Marques, Digital Discovery, **2**, 1368-1379 (2023)*
*[4] J. Schmidt, L. Pettersson, C. Verdozzi, S. Botti, M. Marques, Sci. Adv., **7**, (2021)*
[5] https://alexandria.icams.rub.de/

Scaling molecular modeling to millions of reactions with neural network potentials

Olexandr Isayev, Dylan M. Anstine, Roman Zubatiuk
Carnegie Mellon University, United States

Developments in high-throughput experimentation, automated chemistry platforms, and chemical generative models have created an urgent need to rapidly predict reaction outcomes so that synthetic planning and evaluations of degradability can match the emerging pace of molecule/material discovery. Machine-learned interatomic potentials (MLIPs), whereby potential energy surface representations are learned from datasets of first-principles calculations, present an attractive opportunity to overcome the expensive and/or time-consuming tasks required to characterize and refine reactions experimentally or with quantum chemistry models. In this talk, I will introduce two models, RxnAIMNet and AIMNet2-Pd, that illustrate how MLIPs can be used to predict the thermodynamics and kinetics of general organic reactions and carbon-carbon cross-couplings that are catalyzed by Pd organometallic complexes. By learning from a newly constructed and exhaustive dataset of ~10 million molecular conformers, RxnAIMNet is shown to reliably perform minimum energy pathway searches, transition state optimization, and intrinsic reaction coordinate calculations, leading to predicted activation barriers within ~2 kcal/mol of reference range-separated hybrid density functional theory (DFT) calculations. To meet the needs of high-throughput reaction characterization, we introduce a method for batched nudged elastic band calculations, which allows RxnAIMNet to identify ~350,000 minimum energy pathways daily on a single medium-end GPU.

Property predictions from limited and multi-fidelity datasets

Pierre-Paul De Breuck
Université catholique de Louvain, Belgium

The available materials data are often rather limited. Unfortunately, typical machine-learning approaches generally require large amounts of data to make accurate predictions. To tackle this limitation, I will introduce MODNet, an all-round framework which relies on a feedforward neural network, the selection of physically meaningful features, and when applicable, joint-learning. Next to being faster in terms of training time, this approach is shown to outperform current graph-network models on small datasets.

Furthermore, materials data often present various levels of accuracy, with typically much less high- than low-fidelity data. I will therefore present approaches to extract as much information as possible from all available data.

Finally, I will discuss how active learning can be used to optimally perform data acquisition, with two practical applications.

Poster Abstracts

Within a topic, the abstracts are sorted alphabetically based on the first author's family name.

Thermodynamic Observables

Synthesis of triangulene made simple by single-atom alloys

Lin Chen, Jonas Björk, Johanna Rosen
[1]Linköping University, Sweden

Triangulene, the smallest triplet-ground-state polybenzenoid, has previously been successfully synthesized by tedious tip-manipulation experiments, employing scanning probe techniques to dehydrogenate precursor molecules under ultra-high vacuum and low-temperature conditions. Here, we propose an alternative approach to form triangulene on single-atom alloys, predicted using density functional theory calculations combined with microkinetic analysis. Our findings indicate that isolated Pd and Pt atoms on Au(111) and Ag(111) surfaces reduce the activation energies for the dehydrogenation of dihydrotriangulene (the precursor molecule) and make the reaction energies less endothermic compared to pristine Au(111) and Ag(111). Microkinetic analysis based on the proposed reaction mechanism reveals that alloying Au and Ag with single Pd and Pt atoms increases the yield of triangulene at moderate temperatures. Our theoretical work suggests a thermal synthesis pathway for triangulenes, offering an alternative to the prevalent tip-manipulation techniques that could facilitate manufacturing by enabling synthesis through simple heating.

Gaussian approximation potential development for bimetallic Pt-Cu nanoparticles

Ilker Demiroglu
[1]Eskisehir Technical University, Turkey

Metallic nanoparticles such as Pt have an ever-increasing scientific attraction due to their high catalytic activity on a wide range of technological applications from fuel cells to dental alloys. However, some of the most important problems in the field of metallic nanoparticle catalysts are high cost, mining difficulties and insufficient raw material resources. Therefore, one of the basic solutions is that alloying them with cheaper metals which will not reduce their catalytic activity. In some cases, it is even possible to increase the catalytic activity by alloying due to the synergic effects depending on the metal selected.

Theoretical studies in metallic catalysis concentrate on the crystal structures and stabilities of catalyst nanoparticles and the reaction dynamics on them. In addition to the accuracy and precision limitations due to the selected method in theoretical studies, there exist also limitations such as temperature and entropic effects, which are not generally considered especially in quantum mechanical methods. Although such calculations can be made on device operating temperatures by using techniques based on molecular dynamics simulations, they have limitations especially for quantum mechanics level studies as these techniques require much larger system sizes and need very long simulation times. Therefore, systems in simulations for catalysis cannot reach experimental sizes and they are limited to ideal and ordered crystal structures. However, actual catalyst materials in experiments usually have defects and mixed phase structures. Although calculations for large systems can be reachable by techniques based on classical mechanics using force fields, the accuracy of these calculations are quite limited because they do not include electronic interactions. At this point, machine learning potentials such as GAP started to gain importance in literature due to their potential to generate the accuracy level of quantum mechanics-based techniques.

In this respect, the main goal is set to generate Gaussian Approximation Potential that has quantum mechanical accuracy using machine learning techniques for actual size nanoparticles of Pt-Cu alloys. Thus, technologically important Pt-Cu alloy nanoparticles are investigated first by quantum mechanical level molecular dynamics simulations and then the results are used as training set to develop machine learning type GAP potentials. The one-to-one energy and force comparison of DFT and GAP potentials gives good linear relationship both for training and validation sets. Furthermore, the structural optimization results obtained from GAP potentials show very promising accuracy.

Electronic Structure and Long Range Interactions

Efficient ensemble averaging methods to study electronic structure at finite temperature from first principles calculations using neural network

Niraj Aryal[1], Sheng Zhang[2], Weiguo Yin[1], Gia-Wei Chern[2]
[1]Brookhaven National Laboratory, United States
[2]University of Virginia, Charlottesville, Virginia, United States

Calculating electronic and lattice properties of materials at finite temperature in the presence of disorder and defects is important for data-driven design and discovery of materials. While various perturbative and non-perturbative methods exist to calculate finite temperature properties within first-principles calculations, they are not without limitations. Here, we propose an alternative group theory-based supervised machine learning (ML) non-perturbative approach which can reduce the computational cost of such finite temperature calculations. We demonstrate that our Density functional theory+ML based approach, after appropriate training and neural network optimization, can (i) reduce the number of DFT calculations necessary to perform ensemble average for a given temperature and (ii) efficiently predict the temperature dependence of electronic band gap thereby making finite temperature electronic structure calculations computationally tractable.

Learning the density matrix, a symmetry rich encoding of the electronic density.

Pol Febrer Calabozo[1], Peter B. Jørgensen[2], Miguel Pruneda[3], Alberto García[4], Arghya Bhowmik[2]
[1]ICN2, Spain
[2]DTU, Denmark
[3]CINN, Spain
[4]ICMAB, Spain

In the most recent years, the electronic density has been getting attention as a target for machine learning (ML) models due to the amount of information it contains. In fact, density functional theory (DFT) proposes that all ground state electronic properties of an atomic system should be inferable from it. The last advances in ML interatomic potentials have shown that taking into account the equivariance of the data (e.g. forces should rotate when the system is rotated) greatly enhances the learning capacity while needing less data to train. In this context, equivariant models that predict the electronic density have quickly appeared. These models predict scalar values on a real space grid or coefficients for a density fitting expansion. DFT codes with atom-centered basis sets, however, compute the electronic density by products of orbitals. The coefficients of these products follow the equivariance of products of spherical harmonics, which is of higher order than the target values for the previous approaches. In our work, we target the density matrix, which contains these coefficients. By doing so, we force the model to learn more meaningful details about the atomic interactions. The computation of the density matrix scales linearly with system size and the representation is more compressed than that of a 3D grid, resulting in better data efficiency. In this talk, we present the architecture of our models and the results of predicting electronic densities, which show similar performance as state-of-the-art grid based models. We also show the results of computing properties directly from the predicted densities (energy, dipole moments). Finally, we feed the predicted densities into DFT as initial guesses, which results in a reduced number of self consistent field (SCF) iterations to reach convergence.

dxtb - an efficient and fully differentiable framework for extended tight-binding

Marvin Friede[1], Christian Hölzer[1], Sebastin Ehlert (2)[2], Stefan Grimme[1]
[1]University of Bonn, Germany
[2](2) AI4Science, Microsoft Research, Netherlands

Automatic differentiation[1] (AD) emerged as an integral part of machine learning, enabling rapid prototyping and model development by allowing gradient-based optimization without involved analytical derivatives. Recently, the capabilities of AD were also recognized in the field of quantum chemistry.[2,3] In this work, we present *dxtb* - an open-source, fully differentiable framework for semi-empirical extended tight-binding[4] (xTB) methods. Developed entirely in Python and leveraging PyTorch for array operations, *dxtb* focuses on extensibility and user-friendliness, while remaining computationally efficient. Through comprehensive code vectorization and optimization, we closely rival the speed of compiled xTB programs for high-throughput calculations of small molecules. The excellent performance also scales to large systems, and batch operability yields additional benefits for execution on massively parallel hardware. We showcase the simplicity of obtain-

ing derivatives with AD by calculating various molecular and spectroscopic properties. *dxtb* streamlines optimization tasks, for instance, in the development and parametrization of novel tight-binding methods. Furthermore, the framework offers seamless integration of semi-empirical quantum chemistry in machine learning models, paving the way for physics-inspired end-to-end differentiable models. Exploiting the inherent synergies of such tandem approaches,[5] we envision *dxtb* as an extensible foundation for hybrid machine learning approaches.

[1] A. Griewank, A. Walther, Evaluating Derivatives, 2008
*[2] A. McSloy, G. Fan, W. Sun, C. Hölzer, M. Friede, S. Ehlert, N. Schütte, S. Grimme, T. Frauenheim, B. Aradi, The Journal of Chemical Physics, **158**, (2023)*
*[3] X. Zhang, G. Chan, The Journal of Chemical Physics, **157**, (2022)*
*[4] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, WIREs. Comput. Mol. Sci., **11**, (2020)*
*[5] N. Fedik, B. Nebgen, N. Lubbers, K. Barros, M. Kulichenko, Y. Li, R. Zubatyuk, R. Messerly, O. Isayev, S. Tretiak, The Journal of Chemical Physics, **159**, (2023)*

GP3-xTB: a general purpose self-consistent tight-binding quantum chemical method

Thomas Froitzheim, Marcel Müller, Andreas Hansen, Stefan Grimme
[1]University of Bonn, Germany

We present our third-generation extended tight-binding (xTB) model named GP3-xTB. It aims at general-purpose (GP) applicability in chemistry, going beyond the special-purpose parametrization for geometries, frequencies, and non-covalent interactions (GFN) of previous xTB-versions [1,2]. The centerpiece of the new method is the molecule-optimized adaptive minimal valence basis set q-vSZP [3], which allows for a different radial expansion of basis functions on symmetry-distinct atoms in a molecule. Beyond improvements to the basis set, the GP3-xTB Hamiltonian is augmented by three main features: (i) approximate non-local Fock exchange to emulate the behavior of range-separated hybrid DFT, (ii) atomic correction potentials (ACP [4]) to introduce anisotropy into the minimal basis, and (iii) a complete range-dependent expansion of the electrostatic interactions up to third-order, including a novel first-order term to improve the description of charged species. Together with a substantially enlarged and diverse training set for elements up to Z=103, GP3-xTB approaches DFT accuracy more closely and for a wider range of chemical properties such as barrier heights, ionization potentials, or thermochemistry, than previous semi-empirical methods at only slightly increased computational cost. In conjunction with the current developments of dxtb, a fully differentiable implementation of the xTB methods, GP3-xTB offers an ideal platform for machine-learning applications and feature generation.

*[1] C. Bannwarth, S. Ehlert, S. Grimme, J. Chem. Theory Comput., **15**, 1652-1671 (2019)*
*[2] S. Grimme, C. Bannwarth, P. Shushkov, J. Chem. Theory Comput., **13**, 1989-2009 (2017)*
*[3] S. Grimme, M. Müller, A. Hansen, The Journal of Chemical Physics, **158**, (2023)*
*[4] V. Prasad, A. Otero-de-la-Roza, G. DiLabio, J. Chem. Theory Comput., **18**, 2913-2930 (2022)*

Predicting the electronic densities and nuclear dynamics of 2d materials

Zekun Lou[1], Alan Lewis[2], Andrea Grisafi[3], Mariana Rossi[4]
[1]Max Planck Institute for the Structure and Dynamics of Matter, Germany
[2]York University, United Kingdom
[3]Institut Sciences du Calcul et des Données, France
[4]MPI for the Structure and Dynamics of Matter, Germany

Machine Learning (ML) methods have significantly advanced the field of materials modeling. A comprehensive ML framework at the nanoscale needs to target both nuclear and electronic subsystems. The electronic density is a fundamental quantity that allows accessing several material properties, in particular within the framework of density-functional theory. However, obtaining a ground-state electronic density of large materials is prohibitive due to the cost of converging the self-consistent procedure and simulating nuclear motion with quantum-mechanical accuracy adds an even larger overhead. For 2D materials and 2D twisted bilayer materials, large unit-cell sizes are necessary to ensure commensurability and to capture the real-space dynamics of quantum phases such as charge-density-waves [1].

Here we show improvements to the symmetry-adapted learning of three-dimensional electron densities (SALTED) method which allows one to bypass the cost of performing a DFT calculation and instead predict the ground-state electronic density based only on atomic positions and composition [2]. These improvements allow us to obtain accurate predictions of the density for a diverse set of twisted bilayer systems, when training on small unit cells and predicting on large unit cells with a low twisting angle. We also show that these densities lead to good band-structures and thus the prediction of several electronic properties of large structures. In particular, we show results for twisted bilayer graphene, hBN, ZrS2 and MoS2, and discuss the current limitations on accuracy of this method. Finally, we couple SALTED predictions of twisted-bilayer systems to thermally displaced structures obtained from machine-learned potentials trained on smaller-scale DFT data from GGA and hybrid exchange-correlation functionals [3].

[1] https://scipost.org/SciPostPhys.16.2.046
[2] https://pubs.acs.org/doi/full/10.1021/acs.jctc.1c00576
[3] https://arxiv.org/abs/2403.10343

Multifidelity machine learning for quantum chemical properties

Vivin Vinod
[1]University of Wuppertal, Germany

The accurate but fast calculation of quantum chemical (QC) properties such as molecular excited states is still computationally expensive. For many applications, a detailed knowledge of such properties is of key importance thereby demanding high accuracy calculations. Machine learning (ML) methods have reduced the compute cost of such properties in the recent past. However, these bring along a new bottle neck: the cost of generating high accuracy training data. This still remains a challenging issue in the application of ML for QC. Multifidelity machine learning (MFML) methods provide a respite to this hurdle by requiring very little training data from high accuracies which is combined with cheaper and less accurate data to achieve the accuracy of the costlier level. In this work, some key results which show the reduction of the cost of training data are shown for excitation energies at DFT accuracy. In addition, a novel methodological improvement, optimized MFML (o-MFML) is shown. Some further results are also discussed which establish MFML as a superior method to conventional ML methods employing single fidelity training data sets.

*[1] V. Vinod, U. Kleinekathöfer, P. Zaspel, Mach. Learn.: Sci. Technol., **5**, 015054 (2024)*
*[2] V. Vinod, S. Maity, P. Zaspel, U. Kleinekathöfer, J. Chem. Theory Comput., **19**, 7658-7670 (2023)*

Charge equilibration in machine learning potentials

Martin Vondrák
[1]Fritz Haber Institute of the Max Planck Society, Germany

Machine learning (ML) techniques have recently been shown to bridge the gap between accurate first-principles methods and computationally cheap empirical potentials. This is achieved by learning a mapping between a systems structure and its physical properties. State-of-the-art models typically represent chemical structures in terms of local atomic environments to this end. This inevitably leads to the neglect of long-range interactions (most prominently electrostatics) and non-local phenomena (e.g. charge transfer), resulting in significant errors in the description of polar molecules and materials (particularly in non-isotropic environments). To overcome these issues, we are developing ML frameworks for predicting charge distributions in molecules based on Charge Equilibration (QEq). Here, atomic charges are derived from a physical model using environment-dependent atomic electronegativities. In this contribution, we will demonstrate the strategies for creating long-range interatomic potential termed Kernel Charge Equilibration (kQEq) combining local Gaussian Approximation Potential (GAP), and an approach of incorporating QEq into the equivariant MACE neural network scheme.

Deep learning functionals based on Møller-Plesset adiabatic connection for non-covalent interactions

Heng Zhao
[1]University of Fribourg, Switzerland

The modeling of non-covalent interactions (NCIs) is crucial in many areas of chemistry and material science, as these interactions often govern the structure, stability, and function of complex molecular systems. To improve pure quantum chemical simulations of NCIs, we propose an interpolation method along the Møller–

Plesset adiabatic connection (MP AC), which approximates the correlation energy by combining MP2 at small coupling strengths and the strong-coupling limit of the MP AC. By leveraging deep learning techniques, we obtain models ensuring size-consistency and the accurate capture of NCIs, which particularly shines for pi-pi stacking dominated systems. While our models have the same cost as double hybrids, they offer major improvements over double hybrids for noncovalent interactions.

Magnetic Observables

Machine learning spin-pure excited states of organic radicals

Tim Hele, James Green, Jingkun Shen, Keith Butler
[1]University College London, United Kingdom

Recent years have seen an explosion of interest in organic radicals due to their promise for organic light-emitting diodes and molecular qubits [1]. However, simulating their electronic structure is very challenging due to the large size of the molecules and their open-shell nature, which can lead to spin-contaminated results. Very recently two of us have combined extended configuration interaction singles (XCIS) and Pariser-Parr-Pople (PPP) theory to create a fast, accurate and spin-pure method for radical excited state calculation which we call ExROPPP [2]. The original proof-of-concept demonstration only considered hydrocarbon radicals and extending this to heteroatoms would greatly extend its applicability. However, as far as we are aware there is no consistent published heteroatom parameter set. We are therefore using machine learning to optimize the parameters of ExROPPP for carbon and heteroatoms by fitting its predictions to experimental measurements and high-level calculations. We believe this is the first application of machine learning to the spin-pure excited states of radicals.

*[1] X. Ai, E. Evans, S. Dong, A. Gillett, H. Guo, Y. Chen, T. Hele, R. Friend, F. Li, Nature, **563**, 536-540 (2018)*
*[2] J. Green, T. Hele, The Journal of Chemical Physics, **160**, (2024)*

Applying a well-defined energy density for machine-learned density functionals

Elias Polak, Stefan Vuckovic, Heng Zhao
[1]University of Fribourg, Switzerland

The recent integration of machine learning techniques in density functional theory (DFT) has established a powerful framework for developing next generation density functionals. While robust modelling of the exchange-correlation requires a well-defined energy density, conventional training sets usually rely on global quantities. We propose the application of the local slope in the non-interacting limit of the adiabatic connection approach in DFT [1]. The presentation will elucidate the methods for an efficient implementation of this quantity, with a focus on its spin-resolved components and its regularized version. Furthermore, we will highlight the potential of this strategy in paving the way for the next generation of machine-learned local dynamic hybrid functionals. Our results show a marked improvement in the prediction of observables while also maintaining computational efficiency.

*[1] S. Vuckovic, T. Irons, A. Savin, A. Teale, P. Gori-Giorgi, J. Chem. Theory Comput., **12**, 2598-2610 (2016)*

Spectroscopic Observables

Uncovering the potential of applying principal component analysis (PCA) on Raman spectra of biochar

Raymond Chen, Ewa J. Marek
[1]University of Cambridge, United Kingdom

Raman spectroscopy is a powerful tool, which can be used for analysing nanostructure features of biochar, but the plethora of methods for band fitting and analysis makes finding the right approach for uncovering all relevant characteristics difficult, and questions what information remains undiscovered in Raman spectra [1]. To address both problems, we applied principal component analysis (PCA) on Raman spectra from biochar produced by pyrolysing dried *Sargassum* algae and pelletised *Sargassum* algae powder in N_2 at 260-460°C. The first principal component (PC) clearly correlates with the pyrolysis temperature and the third PC showed a difference between Raman spectra from dried *Sargassum* and *Sargassum* pellets. Furthermore, the pyrolysis temperature of the biochar samples could be distinguished by solely considering the values of the first PC, even if PCA is only applied on a commonly neglected region of biochar Raman spectra (500-1000 cm^{-1}), indicating unexplored information. Further analysis revealed differences in the contributions of Raman shift regions to a PC value between the typically analysed region of 1200-1700 cm^{-1} and commonly neglected region of 500-1000 cm^{-1}, possibly indicating that information is contained differently in the region 500-1000 cm^{-1}. These results demonstrate the capabilities of PCA in analysing Raman spectra of biochar samples with different characteristics and revealing hidden information in Raman spectra of biochar.

[1] J. Xu, Q. He, Z. Xiong, Y. Yu, S. Zhang, X. Hu, L. Jiang, S. Su, S. Hu, Y. Wang, J. Xiang, Energy. Fuels., ***35****, 2870-2913 (2020)*

Machine learning isotropic g values of organic battery materials

Davis Thomas Daniel[1], Souvik Mitra[2], Rüdiger-A. Eichel[1], Diddo Diddens[1], Josef Granwehr[1]
[1]Forschungszentrum Jülich, Germany
[2]University of Münster, Germany

Organic radical polymers with pendant radical centers are used as active materials in organic radical batteries (ORBs) and can be investigated using electron paramagnetic resonance spectroscopy owing to their paramagnetic nature. EPR spectra of such species with closely-packed spin centers are primarily characterised by an isotropic g value which describes the interaction of the electron spin with the applied magnetic field. The g-shift is analogous to chemical shifts in nuclear magnetic resonance spectroscopy and is characteristic of the chemical identity of the paramagnetic system and its environment. EPR g tensors can be calculated at the DFT level which allows for the use of g values as a parameter which can be experimentally verified and theoretically computed to validate simulated paramagnetic systems, such as organic cathodes. A severe drawback of applying DFT to obtain g values of large, disordered systems like polymers, is its unfavourable scalability with system size and high computational expense. Machine learning methods offer a viable alternative to predict spectroscopic observables with comparable accuracy with low computational demands. Herein, a machine learning workflow which explores the prediction of g values of from structural features of radical polymers is presented. Polymer structures for generating the training data was sourced from molecular dynamics simulations of an organic cathode at different states of charge. The model is then trained on DFT-derived g values which are benchmarked against experimental g-values obtained from *in operando* EPR measurments of an ORB. To represent the polymers, smooth overlap of atomic positions and many body tensor representation were used along with structural properties of the polymer as features. The model is evaluated with respect to its ability to predict the evolution of g along MD trajectories for polymer structures with varying radical density. Mean deviations of ML-predicted g values and DFT-derived g values were found to be on the order of 0.0001 which is less than the variation in g observed experimentally using EPR and calculated using DFT. Furthermore, the ML model was found to be sensitive to small structural variations corresponding to small g-shifts and important structural features of the polymer related to g-value predictions were identified. As the computational cost of g-value predictions using the ML model is smaller than DFT, the protocol represents a viable alternative to predict g values for simulated systems with a larger system size.[1] As the method has more favourable scalability than DFT to larger systems, state of charge dependent g-shifts for large-scale MD simualtions of ORBs may be computationally accessible, allowing for comparison with experimental ORB set-ups and further optimisation of simulated environments.

*[1] D. Daniel, S. Mitra, R. Eichel, D. Diddens, J. Granwehr, J. Chem. Theory Comput., **20**, 2592-2604 (2024)*

Theoretical nonlinear vibrational spectroscopy of water in slit pores

Banshi Das, Sergi Ruiz-Barragan, Dominik Marx
[1]Ruhr-Universität Bochum, Germany

Water exhibits surprisingly different properties when it is confined between two parallel walls at nanometer distances with reference to bare interfaces, where the interfacial region is connected with a bulk like region. Although water under confinement has great importance in the biological as well as technological realms, its properties are being disclosed only recently due to breathtaking advances in the fabrication of ultranarrow slit pore confinement systems with different environments, geometries, and sizes. Unlike interfacial water, nanoconfined water has not yet been probed using vibrational sum frequency generation (VSFG) spectroscopy – a technique that can efficiently probe interfacial water. This might be due to the fact that most nanoconfined systems used in experiments, such as graphene (GRA)-based nanodevices, are centrosymmetric in nature and hence VSFG inactive. In an effort to explore the VSFG technique to study the confinement effects of aqueous systems, we have modelled an asymmetrically nanoconfined system in a slit pore geometry where water is being sandwiched between GRA and boron nitride sheets, which can be considered as a slightly perturbed form of GRA-GRA slit pores that breaks their centrosymmetry. Using an electronic structure-based parameterized approach within force field-based molecular dynamics simulations, VSFG spectra of such confined water have been computed from the monolayer limit up to thick water lamellae where the confining walls are largely decoupled.

Predicting the spectra of semiconductors and insulators

Malte Grunert, Max Großmann, Erich Runge
[1]TU Ilmenau, Germany

Based on ab-initio calculations for thousands of crystalline semiconductors and insulators at varying levels of theory, we employ deep learning techniques to predict the optical spectra of semiconductors and insulators in the visible and ultraviolet region.

Comparing our results to state-of-the-art approaches for related static properties like the static refractive index, we present our first promising findings. Furthermore, we also present various related aspects, e.g., how learning the color of a material compares to the full frequency dependence of its relative permittivity and whether learning additions to lower theory levels performs better than learning the higher theory levels directly.

Vibrational circular dichroism

Sascha Jähnigen
[1]Freie Universität Berlin, Germany

Chiroptical spectroscopy provides an increasingly important, cost-effective methodology for the study of chiral substances in the condensed phase. In recent years, vibrational circular dichroism (VCD) – the chiral form of IR absorption spectroscopy – has come into focus as a very sensitive probe of molecular conformation and environment. It has been applied to a wide range of molecules including natural products, host-guest systems, proteins, nanoparticles, or catalysts, as well as the formation of chiral phases from achiral subunits. VCD differs from electronic circular dichroism in that it relates directly to vibrational transitions in the supramolecular chiral framework, such as functional groups connected by covalent or non-covalent interactions.

VCD therefore delivers a wealth of stereochemical information, yet the spectra cannot straightforwardly be interpreted. It requires both accurate calculations on the basis of ab initio theory (at least DFT level), and an accurate sampling of the conformational phase space of the molecular system.

This is why ab initio molecular dynamics (AIMD), which can deliver both accuracy and sampling, provide an attractive means to calculate VCD spectra. However, their tremendous computational cost represents a true bottleneck when harvesting the potential of VCD. In particular the calculation of the spectroscopic properties for individual time frames exacerbates the already high costs of the DFT calculation.

While the calculation of IR absorption only requires the electric dipolar resonse of the electrons to a vibrational transition, VCD in addition relies on the magnetic dipolar response. This can be done in the realms of quantum perturbation theory, which has become a standard feature of many quantum chemical codes and has been pushed forward by us in the realms of AIMD simulations using Nuclear Velocity Perturbation Theory (NVPT).

We present here our recent results and efforts to understand the nature of the chiroptical response within VCD spectroscopy. It is our aim to invest our gained knowledge about the structure of the dipolar response properties, both electric and magnetic, in order to implement efficient algorithms that are apt to obtain these properties through machine learning (ML). In particular, we devote our attention to the atomic polar tensor (APT) and the atomic axial tensor (AAT) as molecular properties that only depend on the positions, but not the (MD-)velocities of the atoms. We present our examinations of the locality and transferability of these tensors, and recent developments to extract effective tensors from AIMD simulations and hope for a broad feedback from the ML community on this workshop.

*[1] S. Jähnigen, Angew. Chem. Int. Ed., **62**, (2023)*
*[2] S. Jähnigen, K. Le Barbu-Debus, R. Guillot, R. Vuilleumier, A. Zehnacker, Angew. Chem. Int. Ed., **62**, (2022)*
*[3] K. Le Barbu-Debus, J. Bowles, S. Jähnigen, C. Clavaguéra, F. Calvo, R. Vuilleumier, A. Zehnacker, Phys. Chem. Chem. Phys., **22**, 26047-26068 (2020)*
*[4] A. Scherrer, R. Vuilleumier, D. Sebastiani, The Journal of Chemical Physics, **145**, (2016)*

Ion effects on terahertz spectra of microsolvated clusters

Aman Jindal[1], Philipp Schienbein[1], Prashant K Gupta (2)[2], Dominik Marx[1]
[1]Ruhr-Universitaet Bochum, Germany
[2](2) Department of Chemistry, Central University of Kashmir, India

Terahertz (THz) spectroscopy deals with radiations of frequencies from 0.1 to 10 THz (1 THz = 10^{12} Hz = 1 ps^{-1}). The fuctuations in water hydrogen-bonding network also occur at the sub-ps to ps timescale, making THz spectroscopy a sensitive tool to probe these changes. Here, we have studied these changes on addition of simple ions in water clusters from THz spectral features. The THz spectra of finite size $Cl^-(H_2O)_n$ clusters (n = 4, 7, 20 and 64) are computed from their ab-initio molecular dynamics trajectories, in terms of only a few auto- and cross-correlations (including the second solvation shell). The focus is to find out if ion hydration can be understood by studying the THz regime of these small ion-water clusters in the gas phase. To do so, we compare the spectra of clusters with their Cl^-(aq) bulk solutions, and seek answers to the following questions: What is the effect of ion on water molecules in its neighboring shells? What happens when the cluster size is gradually increased? Is there an optimum cluster size that might just be sufficient to reproduce the bulk THz responses?

Complexions at the electrolyte/electrode interface in solid oxide cells

Patricia König, Hanna Türk, Xuan Quy Tran, Axel Knop-Gericke, Thomas Lunkenbein, Christoph Scheurer
[1]Fritz-Haber-Institut of the Max Planck Society, Germany

Degradation of the air electrode in electrolysis mode during the oxygen evolution reaction (OER) severely limits the commercial adoption of solid oxide cells (SOCs). Up to now, the atomistic structure of this active catalyst region is essentially unknown, which prevents a detailed analysis of the actual degradation mechanisms.

In prior research, we identified a complexion at the solid/solid interface of the sintered anode, featuring partial amorphization and varying elemental distributions deviating from the confining bulk phases. Located around the complexion area, we propose deactivation mechanisms driven by strong cation inter-diffusion. To assess if these cation migrations impede active sites and cell functionality, we conduct a large-scale study on possible active site structural motifs. We sample structures with polarons near oxygen defects, altering ion dopant concentrations to simulate ion migration effects. By employing DFT for electron structure calculations and computing EELS spectra, we link oxygen defects and their proximity to polarons to distinctive peaks in experimentally recorded EELS spectra. Ultimately, this approach aids in identifying active site structures and deactivation mechanisms, enhancing future cell design with improved efficiency and durability.

Can machine learning accelerate the discovery of new photovoltaics?

Matthew Walker, Keith Butler
[1]UCL, United Kingdom

The discovery of photovoltaic (PV) materials has historically largely consisted of trial-and-error: using chemical trends to suggest materials with similar performances. A material's band-gap is a key determinant of its PV efficiency due to the trade-off of high voltage (current) at large (small) band-gaps. The Shockley-Queisser limit was calculated in the 1960s, suggesting a maximal efficiency of 30% would require a material with a band-gap of 1.1 eV. More recently, spectroscopic limited maximum efficiency (SLME) theory was developed to establish a more rigorous approach, including the use of a material's absorption spectrum rather than assuming a step function at the band-gap. However, calculating these spectra at high accuracy is a computationally intensive process, implicitly requiring an accurate band-gap which is notoriously difficult with cheap DFT. With the increased availability of large datasets and powerful GPUs, machine learning becomes an attractive option to 'fill in the gaps' in the SLME workflow, replacing *ab initio* calculations by predicting various scalar and vector quantities. If successful, this would facilitate rapid screening of chemical space to identify high-efficiency, previously unexplored PV materials. In this work we use graph neural networks (ALIGNN) to establish which parts of this workflow can be accelerated by machine learning and how the accuracy of efficiency estimates is affected.

*[1] W. Shockley, H. Queisser, Journal of Applied Physics, **32**, 510-519 (1961)*
*[2] L. Yu, A. Zunger, Phys. Rev. Lett., **108**, 068701 (2012)*
*[3] K. Choudhary, B. DeCost, npj. Comput. Mater., **7**, 185 (2021)*
*[4] R. Woods-Robinson, Y. Xiong, J. Shen, N. Winner, M. Horton, M. Asta, A. Ganose, G. Hautier, K. Persson, Matter, **6**, 3021-3039 (2023)*
*[5] S. Kim, M. Lee, C. Hong, Y. Yoon, H. An, D. Lee, W. Jeong, D. Yoo, Y. Kang, Y. Youn, S. Han, Sci. Data., **7**, 387 (2020)*
*[6] D. Fabini, M. Koerner, R. Seshadri, Chem. Mater., **31**, 1561-1574 (2019)*

Modelling tensorial properties of zeolites using machine learning

Daniel Willimetz, Lukáš Grajciar, Andreas Erlebach, Christopher J. Heard, Carlos Bornes
[1]Charles University, Czech Republic

Zeolites are microporous aluminosilicates crucial in various industries, which often require theoretical calculations for deeper understanding of their behaviour. It is essential to investigate the zeolites at *operando* conditions, modelling the chemical composition (including molecules present in the zeolite channels) and temperature as close to the corresponding experimental situation as possible.[1] However, it is computationally prohibitive to do such calculations at the *ab initio* level. Instead, we have developed highly accurate reactive neural network potentials (NNPs) that lead to more than a 10^3 computational speedup.[2, 3] The accuracy of our NNPs has been further validated through collaboration with Prof. Sauer’s group from Humboldt University in Berlin, who is carrying out high-level QM/QM benchmarking of the NNP predictions in our undergoing joint study of water adsorption in zeolites.

By employing NNPs, we can conduct extensive sampling of the configurational space via ns-long molecular dynamics (MD) simulations, carried out for the experimentally relevant model system (chemical composition including explicit water solvation). Subsequently, the MD simulation data can be sampled and analyzed using another machine learning (ML)-based method. This 'double machine learning' approach has been applied to predict NMR chemical shifts using kernel ridge regression predictor as the NMR shift predictor.[4] Modelling NMR chemical shifts can provide insights into the distribution of aluminium in the zeolite framework, which has been shown to modulate the catalytic properties of zeolites, making it an attractive area of research.[5] Through collaboration with Dr. Teresa Blasco from the Instituto de Tecnología Química, we have demonstrated excellent agreement between experimental data and calculations, providing essential information about the aluminium distribution that cannot be obtained by any experimental method alone. To our knowledge, the previous studies have only calculated chemical shifts for a single structure corresponding to a minimum on the potential energy surface (PES), making ours the first study to consider *operando* conditions for interpreting experimental NMR spectra in zeolitic materials.

In our future research, our objective is to develop and train neural network-based models capable of predicting entire NMR shielding tensors, considering and benchmarking currently two avenues to achieve this goal, i.e., via using equivariant message passing NNs to learn tensors directly [6] or/and leveraging NNs predicting energies explicitly dependent on the external fields.[7] This will enable us to simulate NMR spectra under *operando* conditions, allowing us to further interpret experimental data.

In conclusion, our research underscores the importance of utilizing ML to comprehend zeolite dynamics under *operando* conditions. Through the application of the double ML approach, we investigated NMR shifts in experimentally relevant scenarios. These ML advancements hold the potential to provide deeper insights into zeolite properties within realistic environments and we hope to develop more ML-based applications in further work.

*[1] C. Lei, A. Erlebach, F. Brivio, L. Grajciar, Z. Tošner, C. Heard, P. Nachtigall, Chem. Sci., **14**, 9101-9113 (2023)*
[2] A. Erlebach, M. Šípka, I. Saha, P. Nachtigall, C.J. Heard and L. Grajciar, arXiv:2307.00911 (2023)
*[3] A. Erlebach, P. Nachtigall, L. Grajciar, npj. Comput. Mater., **8**, 174 (2022)*
[4] D. Willimetz, Theoretical investigation of 27Al chemical shifts dependence on water amount and temperature in zeo-

lite MFI, bachelor thesis (2023)
*[5] T. Liang, J. Chen, Z. Qin, J. Li, P. Wang, S. Wang, G. Wang, M. Dong, W. Fan, J. Wang, ACS Catal., **6**, 7311-7325 (2016)*
[6] K. Schütt, O. Unke and M. Gastegger, PMLR 139:9377-9388 (2021)
*[7] M. Gastegger, K. Schütt, K. Müller, Chem. Sci., **12**, 11473-11483 (2021)*

Decoding the fucose migration products in blood group epitopes

Murat Yaman, Vasilis Kontodimas
[1]City University of New York, United States

Fucose is a monosaccharide in the L-conformation that lacks a hydroxyl group on C6 and it is involved in various biological processes such as signaling, selectin-dependent leukocyte adhesion, and intercellular communication. However, during the mass spectrometry (MS) analysis, fucose can migrate along a glycan sequence leading to incorrect glycan structure assignments and this has significant implications for MS of carbohydrate analysis, as fucose migration can readily occur following in-source activation of intact ions during the nanoESI process. In this study, we explored the potential internal rearrangement reaction of fucosylated blood group antigens. These trisaccharide structures do not differ in composition, but rather in connectivity and configuration. We compared the density functional theory (DFT) calculations with the available experimental collision cross section (CCS) and cryogenic IR spectroscopy data. The DFT results of the protonated Lewis x and BG-H2 antigens, which both have a T2 core, did not agree with the experimental cryo-IR and CCS values. Instead, extensive sampling of all possible isomers revealed that another structure, where the fucose moiety is linked to the galactose of the T2 core via an α 1-6 glycosidic linkage, provided the best match with the experimental findings. Similar analysis of protonated ions of Lewis a and BG-H1 antigens, which contain on the T1 core, resulted in the assignment of the experimental spectra to the structures of the respective parent ions. The discrepancy in the results raises the question and necessitate an exploration into why fucose migration is evident in the Lewis x and BG-H2 pair, while the Lewis a and BG-H1 pair inhibits the migration. The internal rearrangement reaction is triggered by the proximity of the glycosidic bond of the fucose-T2 core to the presence of a mobile proton. In contrast, the Lewis a and BG-H1 structures lacked the mobile proton to initiate the migration, as the glycosidic bond of fucose moiety to the T1 core was not sufficiently close. These findings show that the migration of fucose moiety is a sequence-dependent phenomenon.

Databases & Reaction Networks

Unsupervised collective variables for exploring expanded configurational space

Rangsiman Ketkaew, Sandra Luber
[1]University of Zurich, Switzerland

We develop a machine learning algorithm so-called Deep Learning AutoEncoder Neural Network (DAENN) algorithm for generalizing collective variables (CVs) for enhanced sampling. The essential novelties of the DAENN are a self-learning engine that applies a specific-purpose loss function using the min-max technique and eXtended Social PeRmutation INvarianT (xSPRINT) representation. To train a model, the DAENN requires only a trajectory of reactant conformers as input and no intermediate or product structures. Carefully manifested by primitive organic reactions such as the Diels-Alder reaction, DAENN-generated CVs are capable of identifying transitions between rare-event metastable states on the expanded configurational space. The results show that the adopted CVs give accurate computed free energy that is in line with those obtained by previously developed CVs and experimental results. In addition, the proposed method is implemented in Deep Learning for Collective Variables (DeepCV), a computer code that provides an efficient and customizable implementation of the DAENN, which can communicate with other molecular dynamics software, namely CP2K and PLUMED.

*[1] R. Ketkaew, S. Luber, J. Chem. Inf. Model., **62**, 6352-6364 (2022)*
*[2] R. Ketkaew, F. Creazzo, S. Luber, J. Phys. Chem. Lett., **13**, 1797-1805 (2022)*

Modeling the potential energy surfaces of the [H_2O - Kr]$^+$ system with gaussian process regression

Jenne Van Veerdeghem, Jérôme Loreau
[1]KU Leuven, Belgium

The detection of several hundreds of organic molecules in the interstellar medium indicates a rich chemistry,[1,2] and is notably influenced by the presence of molecular and atomic ions. Interactions with charged particles are assumed to exhibit barrierless energy landscapes, making them especially important in low temperature environments such as interstellar molecular clouds and star-forming regions. The chemical evolution and composition of these interstellar regions is determined through kinetic models which rely on the availability of rate constants over a wide range of temperatures. One pertinent issue is the lack of rate constants determined at low temperatures (< 100 K).

Ion-molecule reactions are often modeled through capture theory, which is not usually able to include the influence of short-range effects.[3] This can lead to severe under- or overestimation of computed rate constants compared to experimentally determined rate constants.[4,5,6]

Our work aims to include short-range effects in ion-molecule interactions to improve the accuracy of rate constants at low temperatures by designing machine learned representations of adiabatic potential energy surfaces. Our machine learning method of choice is Gaussian process regression as it is able to achieve high accuracy on limited data sets.

As an illustration, we have modeled the first three electronic coupled states of the charge-transfer reaction between H_2O and Kr^+, which has recently been studied experimentally.[7] Since H_2O and H_2O^+ have similar geometries, the dimensionality of the potential energy surface is currently limited to the three coordinates required to position the Kr^+ atom around a frozen H_2O molecule.

Ab initio structure calculations (ic-MRCI+Q/aug-cc-pVDZ) serve as the input of the model. The effect of spin-orbit coupling of Kr^+ was included through the state interaction method. The potential energy surfaces fits, trained on around 2000 data points, achieve a root mean square error of 6 meV in the area around the potential well when tested against an independent validation set. Proper physical long-range behavior is conserved in regions with a low density of training data by using a custom mean function, removing the need to switch to an analytical function. The training set was designed through biased Latin hypercube sampling, favoring the low energy region.

Final validation of our models will be through comparison of calculated rate constants with state-of-the-art cold collision experiments. To this end, quasi-classical trajectory code coupled to a surface hopping algorithm is currently being developed in-house to interface with our machine learning models.

*[1] E. Herbst, E. van Dishoeck, Annu. Rev. Astron. Astrophys., **47**, 427-480 (2009)*

[2] B. McGuire, R. Loomis, A. Burkhardt, K. Lee, C. Shingledecker, S. Charnley, I. Cooke, M. Cordiner, E. Herbst, S. Kalenskii, M. Siebert, E. Willis, C. Xue, A. Remijan, M. McCarthy, Science, ***371****, 1265-1269 (2021)*
[3] J. Loreau, F. Lique, A. Faure, ApJL., ***853****, L5 (2018)*
[4] A. Gingell, M. Bell, J. Oldham, T. Softley, J. Harvey, The Journal of Chemical Physics, ***133****, (2010)*
[5] L. Petralia, A. Tsikritea, J. Loreau, T. Softley, B. Heazlewood, Nat. Commun., ***11****, 173 (2020)*
[6] A. Tsikritea, K. Park, P. Bertier, J. Loreau, T. Softley, B. Heazlewood, Chem. Sci., ***12****, 10005-10013 (2021)*
[7] A. Tsikritea, J. Diprose, J. Loreau, B. Heazlewood, ACS Phys. Chem Au, ***2****, 199-205 (2022)*

Mechanical Observables

Designing the phase stability in metal-organic frameworks and metal halide perovskites: Two sides of the same strain coin?

Nils Clovin, Sven M.J. Rogge
Center for Molecular Modeling (CMM), Ghent University, Belgium

Disorder and strain introduced by dopants and defects have become increasingly essential to rationalise the design of nanostructured materials like metal-organic frameworks (MOFs) and metal halide perovskites (MHPs) [1], [2]. Strain engineering puts this strain at the heart of the design process by mapping the strained topology of the underlying net to its macroscopic properties, such as the material's phase stability [3]. Recently, we showed that such strain fields in MHPs are additive, and while their decay was fully correlated with the position of the strain-inducing defect within the lattice, it was found to be independent of its chemical nature [4]. This raises the question: can strain engineering constitute a material-transcending design tool unifying the impact disorder has in both MHPs and MOFs alike?

To explore this hypothesis, we herein trained ab initio-based machine learning potentials [5] to contrast point defects and their strain fields between two classes of MOFs and MHPs sharing the pcu topology: the M-IRMOF-L series [M={Zn, Be, Mg, Cd, Ca, Sr, Ba}, L={0-8, 10, 12, 14, 16, 18, 20}] and CsBX3 perovskites [B={Ba, Pb, Ca, Mn, Zn, Ni}, X={Cl, Br, I}]. Besides investigating the generality of the strain decay rate within each material class, we compare how strain fields induced in one class inform the design of the other. To this end, we compare (i) strain fields induced by varying the metal ions in MOFs with those induced by changing B-site cation in perovskites and (ii) strain fields induced by changing the IRMOF linker with those induced by a change in the perovskite's X-site halogen. This, in turn, informs us of how we can impact the octahedral tilt and, hence, phase stability in MHPs and MOFs.

[1] A. K. Cheetham, T.D. Bennett, F.-X. Coudert, A.L. Goodwin, Dalton Trans. 2016, 45, 4113-4126.
[2] J.A. Steele et al., J. Am. Chem. Soc. 2021, 143, 10500-10508.
[3] S.M.J. Rogge, S. Borgmans, and V. Van Speybroeck, Matter 2023, 6, 1435-1462.
[4] J. L. Teunissen et al., J. Phys. Chem. C, 2023, 127, 23400-23411.
[5] S. Vandenhaute et al., npj Comput Mater. 2023, 9, 19.