

Belgian-German WE-Heraeus Seminar: "Machine Learning for Spectroscopy"

26.05.2025 – 28.05.2025 Brussels, Belgium

Book of Abstracts

WILHELM UND ELSE HERAEUS-STIFTUNG









We are delighted to welcome you to the **Belgian-German WE-Heraeus Seminar: "Machine Learning for Spectroscopy" (ML4SPEC2025).** This event brings together experts in spectroscopy, data science, and machine learning to explore the latest methodologies that enhance the analysis and interpretation of complex spectral data.

The conference focuses on Al-driven spectral analysis, advanced classification techniques, and data-driven prediction models applied to various spectroscopic methods, including Fourier Transform Infrared (FTIR), Raman, near-infrared (NIR). fluorescence, and core-electron techniques like Xray absorption (XAS), photoelectron (XPS), and fluorescence (XRF) spectroscopy. Discussions will delve into deep learning approaches for also analyzing vibrational properties, light-driven processes, and excited state dynamics.

We would like to thank you all for such a wide interest in participation to ML4Spec2025.

We extend our sincere gratitude to our sponsors for their generous support, which has been instrumental in making this seminar possible. We warmly thank the *Wilhelm and Else Heraeus Foundation*, the *EUSpecLab Doctoral Network*, and the *New Technologies Research Centre at the University of West Bohemia* for their commitment to fostering interdisciplinary research and innovation.

We wish you that this seminar provide a stimulating environment for collaboration and the exchange of ideas.

The ML4Spec2025 Organizing Committee



Committee

The committee comprises 11 PhD students of the EUSpecLab Doctoral Network. The aim of the project is to form the next generation of theoretical material scientists (physicists and chemists).



Benedini Giulio

Vrije Universiteit

Benaissa Mohammed University of Rennes



Eddhib Ridha University of West Bohemia



Janusz Cezary University of Camerino



Bhatia Nitik Technische Universität München



Chandrashekhar Prathibha Uppsala University



Loew Antoine Ruhr University Bochum



Longo Samuel University of Liège





Simperl Florian TU Wien



Sharma Karthikeya Technical University of Denmark

Event venue, how to get there?

ML4Spec2025













Program

Day one (26.05)

- 8:15-8:45 Registration
- 8:45-9:00 Welcome address
- **9:00-9:45** Giorgia Fugallo: Machine Learning for Phonon Engineering: Toward Predictive Thermal and Spectroscopic Modeling
- **9:45-10:30** Toon Verstraelen: *Explicit Electrons in Machine Learning Potentials*
- 10:30-11:00 Coffee break
- **11:00-12:00** Xavier Gonze: *High-throughput calculations of vibrational properties and their analysis*
- 12:00-14:00 Lunch break
- **14:00-14:45** Pierre-Paul De Breuck: *Conditioning on Spectra for Generative Crystal Design*
- **14:45-15:30** Michiel J. van Setten: Spectroscopy and AI in semiconductor technology research
- 15:30-15:45 Coffee break
- **15:45-17:00** Contributed talks: Malte Grunert, Nathalie Brun, Dario Baum, Arno Annys, Sabana Shabnam
- 17:00-17:35 Flash talks: Sourour Ayari, Scheffler Johannes, Gorfer Alexander, Adhyatma Abdurrahman, Karimi Nejad Sara, Hakeem Luqman, Marco Stecca
- 17:35:18:30 Poster session



Program

ML4Spec2025

Day two (27.05)

- 8:45-9:00 WE-Heraeus Foundation Introduction by Dr. Stefan • Jorda
- 9:00-9:45 Annika Bande: Spectra Prediction and Peak • Assignment using Graph Neural Networks
- 9:45-10:30 Dorothea Golze: Accurate XPS predictions of amorphous materials: A machine-learning model combining DFT and GW
- 10:30-11:00 Coffee break
- 11:00-12:00 Venkat Kapil: Vibrational spectroscopy of molecules, condensed phases & interfaces using machine learning and quantum statistical mechanics
- 12:00-14:00 Lunch break
- 14:00-14:45 Aloïs Castellano: Phonons and anharmonicity in the time of machine-learning interatomic potentials
- 14:45-15:30 Ke Chen: Enhancing Vibrational Spectroscopy Predictions Using Machine Learning
- 15:30-15:45 Coffee break ٠
- 15:45-17:00 Contributed talks: Andrea Della Valle, Laurens de • Boer, Max Großmann, Daria M. Tomecka, Joël Eymery
- 17:00-18:30 • Poster session
- 19:30 Social dinner •



Program

ML4Spec2025

Day three (28.05)

- **9:00-9:45** Matthew L. Evans: Decentralized materials research data management, curation and dissemination for accelerated materials discovery
- **9:45-10:30** Lukas Pielsticker: Automatic Quantification of Transition Metal X-ray Photoelectron Spectra using Convolutional Neural Networks
- 10:30-11:00 Coffee break
- **11:00-12:00** Maria K. Y. Chan: *Data- and theory-informed approaches for accelerating the capture and interpretation of core-level spectra*
- 12:00-13:00 Lunch break
- **13:00-13:45** Vaclav Smidl: *Deep Generative Models for Tractable Probabilistic Inference over Molecular Graphs*
- **13:45-15:00** Contributed talks: Mandira Das, Thomas P. van Waas, Trung-Phuc Vo, Strocov Vladimir, Frederico Zecchi
- 15:00-15:15
 Best poster award, closing remarks



Invited speakers

Day One: 26.05.2025

•9:00-9:45 Giorgia Fugallo: Machine Learning for Phonon Engineering: Toward Predictive Thermal and Spectroscopic Modeling
•9:45-10:30 Toon Verstraelen: Explicit Electrons in Machine Learning Potentials
•11:00-12:00 Xavier Gonze: High-throughput calculations of vibrational properties and their analysis

•14:00-14:45 Pierre-Paul De Breuck: Conditioning on Spectra for Generative Crystal Design

•14:45-15:30 Michiel J. van Setten: Spectroscopy and AI in semiconductor technology research

Day Two: 27.05.2025

•9:00-9:45 Annika Bande: *Spectra Prediction and Peak Assignment using Graph Neural Networks*

•9:45-10:30 Dorothea Golze: Accurate XPS predictions of amorphous materials: A machine-learning model combining DFT and GW

•11:00-12:00 Venkat Kapil: Vibrational spectroscopy of molecules, condensed phases & interfaces using machine learning and quantum statistical mechanics
•14:00-14:45 Aloïs Castellano: Phonons and anharmonicity in the time of machine-learning interatomic potentials

•14:45-15:30 Ke Chen: *Enhancing Vibrational Spectroscopy Predictions Using Machine Learning*

Day Three: 28.05.2025

•9:00-9:45 Matthew L. Evans: Decentralized materials research data management, curation and dissemination for accelerated materials discovery
•9:45-10:30 Maria K. Y. Chan: Data- and theory-informed approaches for accelerating the capture and interpretation of core-level spectra
•11:00-12:00 Lukas Pielsticker: Automatic Quantification of Transition Metal Xray Photoelectron Spectra using Convolutional Neural Networks
•13:00-13:45 Vaclav Smidl: Deep Generative Models for Tractable Probabilistic Inference over Molecular Graphs

Machine Learning for Phonon Engineering: Toward Predictive Thermal and Spectroscopic Modeling

Giorgia Fugallo,

Laboratoire de Physique, École Normale Supérieure, Paris, France

giorgia.fugallo@ens.fr

Understanding and controlling thermal properties at the nanoscale is central to the development of nextgeneration electronic, photonic, and energy devices. In this context, phonons play a pivotal role, governing heat transport, energy dissipation, and spectral signatures in both 3D and low-dimensional materials.

Recent advances in machine learning (ML) open up new opportunities to complement and extend this approach. From accelerating the construction of interatomic force fields and predicting thermal conductivities, to enabling high-throughput screening of thermal materials and interpreting vibrational spectroscopies, ML offers a powerful set of tools to address limitations of conventional first-principles simulations.

In this talk, I will present how phonon engineering, thermal transport, and vibrational spectroscopy can intersect with machine learning and how physically-informed models can help bridge the gap between atomistic insight and data-driven discovery[1-9].

- [1] C. Chen et al. Nature Computational Science 2 (2022).
- [2] B. Deng, et al., Nature Machine Intelligence 5, 1031 (2023)
- [3] J. Riebesell, et al. arXiv:2308.14920 (2024).
- [4] H. Yu et al. Materials Genome Engineering Advances, 2 (2024).
- [5] B. Deng, et al. npj Computational Materials, 11 (2025).
- [6] H. Lee et al. arXiv:2407.09674 (2024).
- [7] J. L. Lansford, et al. Nat Commun 11(2020).
- [8] Y. Luo, et al. npj Comput Mater 9, 4 (2023).
- [9] B. Póta et al. arXiv:2408.00755 (2024).

Explicit Electrons in Machine Learning Potentials

Maarten Cools-Ceuppens¹, Joni Dambre², **Toon** Verstraelen¹

¹Center for Molecular Modeling (CMM), Ghent University – Technologiepark-Zwijnaarde 46, B-9052 Gent, Belgium ²IDLab, Electronics and Information Systems Department, Ghent University – imec, Technologiepark-Zwijnaarde 126, B-9052 Gent, Belgium

toon.verstraelen@ugent.be

A suitable model for the potential energy surface (PES) is essential for any molecular simulation. Depending on the application of interest and the physics and chemistry at hand, different requirements such as accuracy, ability to describe relevant processes, and computational efficiency must be balanced. Machine-learning potentials have reached unprecedented trade-offs between these requirements: they can be trained to mimic density functional theory (or even better) training data with arbitrary precision, while being computationally much more efficient than electronic structure methods.

Despite their popularity and promise. machine-learning potentials have an important limitation: they are inherently short-ranged. The restriction to short ranges has two origins. First, it is common to use short real-space cutoffs to characterize the local environment of an atom. (To some extent, message-passing networks circumvent this limitation: multiple message-passing iterations take into account information from beyond the cutoff distance.) Second, the number of possible configurations of atoms within a larger cutoff sphere is enormous, making it impossible to generate relevant examples for all possibilities. Whenever long-range interactions are important, a physical model is unavoidable.

Many physically motivated models for long-range electrostatics and polarization have been developed for normal force fields and were later combined with machine-learning potentials, which handle the short-range interactions. Advantages and disadvantages of different models for long-range interactions will be discussed, and a new framework, the electron machine-learning potential (eMLP), will be presented. [1] eMLP belongs to the class of explicit electron models, [2] which approximate the electron distribution in more detail than a conventional force field with fixed charges, or even a conventional polarizable force field with atomic dipoles. Electrons or electron pairs are introduced as mobile particles with their formal charge and the freedom to polarize due to changes in geometry or the application of an external field. This is illustrated for the case of water in Figure 1. A fundamental advantage of eMLP over variable charge models, such as Qeq [3], is that all sites carry a fixed net charge, allowing the polarization of a solid to be uniquely defined. [4]

By construction, eMLP is well equipped to describe the geometry dependence of the electronic charge distribution: electron (pair) particles are located between nuclei and exhibit anistropic response due to alignment with chemical bonds. As fragments rotate or vibrate, their static and response properties naturally follow these changes in geometry. In eMLP, this enabled accurate predictions of IR activity, even for molecules on which eMLP had never been trained.

References

- M. Cools-Ceuppens, J. Dambre and T. Verstraelen, J. Chem. Theory and Comput., 18 (2022) 1672. DOI:10.1021/acs.jctc.1c00978
- [2] C. Bai, S. Kale, J. Herzfeld, Chem. Sci. 8 (2017) 4203–4210. DOI:10.1039/C7SC01181D
- [3] A. K. Rappe, W. A. Goddard, J. Phys. Chem. 95 (1991), 3358. DOI:10.1021/j100161a070
- [4] N. A. Spaldin, J. Solid State Chem. 195 (2012) 2 DOI:10.1016/j.jssc.2012.05.010



Figure 1. Illustration of the charge centers in the eMLP model.

High-throughput calculations of vibrational properties and their analysis

Xavier Gonze¹

¹Université catholique de Louvain (IMCN/MODL), 8 Chemin des étoiles B-1348 Belgium.

xavier.gonze@uclouvain.be

The knowledge of the vibrational properties of a material is of key importance to understand physical phenomena such as infra-red and Raman spectroscopy, thermal conductivity, ferroelectricity, as well as the temperature dependence of many properties, including thermodynamic ones. However, detailed experimental vibrational (phonon) spectra are available only for a limited number of materials, which hinders the large-scale analysis of vibrational properties and their derived quantities.

I will first present ab initio calculations of the phonon dispersion and vibrational density of states for 1521 semiconductor compounds [1] based on density functional perturbation theory. The data is collected along with derived dielectric and thermodynamic properties. The procedure used to obtain the results will be described, as well as the details of the provided database and a validation based on the comparison with experimental data, see Fig.1.

Then, I will describe the usage of such database to examine electron-phonon interaction [2], that is also central to condensed matter, e.g. through electrical resistance, superconductivity or the formation of polarons, and that has a strong impact on observables such as band gaps or optical spectra. The most common framework for band energy corrections and polaron formation is the Fröhlich model, which often agrees qualitatively with experiments in polar materials, but has limits for complex cases. A generalized version includes anisotropic and degenerate electron bands, and multiple phonons. Trends are examined for the Fröhlich model on 1260 materials. The limitations of the Fröhlich model and its perturbative treatment is tested, in particular the large polaron hypothesis. Among our extended dataset most materials host perturbative large polarons, but there are many instances that are non-perturbative and/or localize on distances of a few bond lengths, see Fig.2.

References

- [1] G. Petretto, S. Dwaraknath, H.P.C. Miranda, D. Winston, M. Giantomassi, M. J. van Setten, X. Gonze, K.A. Persson, G. Hautier and Gian-Marco Rignanese. Scientific data, 5 (2018) 180065
- [2] P.M.M.C de Melo, J.C. de Abreu, B. Guster, M. Giantomassi, Z. Zanolli, X. Gonze, and M. Verstraete. npj Computational Materials, 9 (2023) 147



Figure 1. Relative error of selected calculated phonon frequencies with respect to experimental data. The test set consists of the frequencies at the Γ point for which experimental values are available in literature. Each of the 53 considered materials is identified by a unique combination of symbol and color. The bar plot shows the distribution of the errors [1].



Figure 2. Upper panel : distribution of electron-phonon coupling strength (α) and hole polaron radii (a_P) for hole polarons in 1260 materials [2]. The color denotes materials that contain chemical elements from groups 15 (pnictides - green) to 17 (halides - blue) of the periodic table. When no such element is present, brown is used. Lower panel : histogram and cumulative frequency for the electron-phonon coupling strength. α bigger than 6 correspond to the strong coupling regime, while a_P lower than 10 Bohr corresponds to small polarons.

Conditioning on Spectra for Generative Crystal Design

Pierre-Paul De Breuck¹, Hashim Piracha¹, Gian-Marco Rignanese², Miguel A.L. Marques¹ ¹Research Center Future Energy Materials and Systems of the University Alliance Ruhr and Interdisciplinary Centre for Advanced Materials Simulation, Ruhr University Bochum, Universitätsstraße 150, D-44801 Bochum, Germany

²UCLouvain, Institute of Condensed Matter and Nanosciences (IMCN), Chemin des Étoiles 8, Louvain-la-Neuve 1348, Belgium

pierre-paul.debreuck@rub.de

Generative models can significantly accelerate the discovery and design of crystal structures. In this talk, I will introduce Matra-Genoa [1], an autoregressive transformer model designed to generate stable crystals. By leveraging invertible tokenized representations of symmetrized crystals, our approach enables sampling from a hybrid action space while conditioning on key properties such as the distance to the convex hull. Beyond stability, I will discuss early efforts to use generative modeling for discovering materials with tailored optical properties. Optical spectra in crystalline materials remain an underexplored frontier, despite their photovoltaics, photocatalysis, potential for epsilon-near-zero applications, and more. Bv conditioning Matra-Genoa on spectral features-such as the dielectric function-we aim to design materials optimized for specific optical applications. opening new possibilities for spectroscopy-driven materials discovery.

References

 De Breuck, P.-P., Piracha, H. A., Rignanese, G.-M. & Marques, M. A. L., ArXiv preprint (2025), A generative material transformer using Wyckoff representation.

Figures



Figure 1. Schematic representation of conditioned generation of crystal structures.

Spectroscopy and AI in semiconductor technology research

Michiel J. van Setten

Imec, kapeldreef 75, 3001 Heverlee, Belgium

Michiel.vansetten@imec.be

The semiconductor industry is arguably one of the, if not the, most technologically advanced industries. To keep innovation moving along, many types of spectroscopies are continuously used to gain more understanding in, and ultimately, improve the involved materials and processes. Also, the ideas and influences of the development of artificial intelligence enter this industry at many dimensions.

In this talk I will introduce some of the key processes involved in the production of semiconductor devices and show where the different spectroscopies provide crucial insights. The main example will focus on the research to understand and improve the photo lithography process, see Fig. 1. Under the exposure to of EUV light of 92 eV, the photo active resist material undergoes a chemical change that transfers the pattern of light onto the wafer. By combing measured spectra at increasing exposure doses to computed spectra we can pinpoint which atomistic changes are taking place, see Fig. 2.^{1,2}

Next, we will visit two other topics, related to the deposition of the materials layers. The first is the estimation of reaction energy barriers³, the second the estimation of vapor pressure⁴. Both are quantities that are important to improve the chemistries of atomic layer deposition and etch. Both are also very difficult to compute directly from first principles. But, by using descriptors that contain 'simple' first principles results, machine learning models can be trained that are transferable and have predictive power and can be used to optimize the processes.

References

- [1] L. Galleni, A. Meulemans, F.S. Sajjadian, D.P. Singh, S. Arvind, K. Dorney, T. Conard, G. D'Avino, G. Pourtois, D. Escudero, and M.J. van Setten, J. Phys. Chem. Lett. 15 (2024) 834
- [2] F.S. Sajjadian, L. Galleni, M.J. van Setten, S. De Gendt, K.M. Dorney, D.P. Singh, F. Holzmeier, and T. Conard, J. Vac. Sci. A 41 (2023) 053206
- [3] E. Marques, S. De Gendt, G. Pourtois, and M.J. van Setten, J. Chem. Inf. Model. 63 (2023) 1454–1461
- [4] C. Pashartis, B. Motmans, G. Pourtois, and M.J. van Setten, in preparation



Figure 1. Schematic steps of the photolithographic process. PAB = post apply bake. PEB = post-exposure bake.



Figure 2. Computed spectrum of the unexposed resist highlighting the individual contributions of different (a) atomic orbitals and (b) molecular components. The spectra were computed assuming the nominal molar composition. Numbers 1–6 identify the most relevant peaks. Arrows highlight the peaks 1, 4 and 5 that are observed to be degrading in experiment.

Spectra Prediction and Peak Assignment using Graph Neural Networks

Annika Bande^{1,2}, Amir Kotobi³, Kanishka Singh^{2,4}, Daniel Höche³, Sadia Bari^{5,6}, Robert H. Meißner^{3,7} ¹Leibniz Hannover University, Institute of Inorganic Chemistry, Callinstr. 9, 30167 Hannover, Germany ²Helmholtz-Zentrum Berlin für Materialien und Energie GmbH, Hahn-Meitner-Platz 1, 10409 Berlin, Germany ³Helmholtz-Zentrum Hereon, Institute of Surface Science, Max-Planck-Straße 1, 21502 Geesthacht, Germany ⁴Institute of Chemistry and Biochemistry, Freie Universität Berlin, Arnimallee 22, 14195 Berlin, Germany ⁵Deutsches Elektronen-Synchrotron (DESY), Notkestraße 85, 22607 Hamburg, Germany ⁶Zernike Institute for Advanced Materials, University of

Groningen, Broerstraat 5, 9712 Groningen, Netherlands ⁷Hamburg University of Technology, Institute of Polymers and Composites, Denickestr. 15, 21073 Hamburg, Germany

annika.bande@aci.uni-hannover.de

The use of sophisticated machine learning (ML) models, such as graph neural networks (GNNs) to predict complex molecular properties or all kinds of spectra has grown rapidly [1,2]. However, other than known from quantum chemistry calculations of spectra, the peak assignment is no integral part of ML models. Explainable black-box artificial intelligence offers tools to open the box: Feature attribution serves to determine the contributions of various atoms in the molecules (nodes in the GNN) to the peaks observed in the spectrum (see figure). By numerically comparing this peak assignment to the core and virtual orbitals from the underlying quantum-chemical calculations the exemplary X-ray absorption spectra data set of small organic molecules, we demonstrate how the atomic contributions deliver spectra interpretation [3]. Robustness tests further demonstrate the reliability of the predictions.

References

- K. Singh, J. Münchmeyer, L. Weber, U. Leser, A. Bande, J. Chem. Theory Comput. 18 (2022) 4408.
- [2] A. Kotobi, L. Schwob, G. B. Vonbun-Feldbauer, M. Rossi, P. Gasparotto, C. Feiler, G. Berden, J. Oomens, B. Oostenrijk, D. Scuderi, S. Bari, R. H. Meißner, Commun. Chem. 6 (2023) 46.
- [3] A. Kotobi, K. Singh, D. Höche, S. Bari, R. H. Meißner, A. Bande, J. Am. Chem. Soc. 145 (2023) 22584.



Figure 1. Workflow of creating a molecular graph, training and executing the graph neural network to return X-ray absorption spectra. Peak assignment is done through the analysis of the accumulated weights of the nodes at every energy.

Accurate XPS predictions of amorphous materials: A machine-learning model combining DFT and GW

Dorothea Golze

TU Dresden, Bergstrasse 66c, Dresden, Germany

dorothea.golze@tu-dresden.de

The *GW* approach has become the method of choice for the computation of addition and removal energies of valence electrons. However, core-level spectroscopy was widely uncharted in *GW*. In the last years, we advanced the *GW* methodology to deep core excitations as measured by X-ray photoelectron spectroscopy (XPS) combining exact numeric algorithms in the real frequency domain with partial self-consistency and relativistic corrections [1-4]. We benchmarked our core-level *GW* implementation for 65 1s core excitations, for which we find that *GW* reproduces absolute molecular 1s excitations within 0.3 eV of experiment and relative binding energies with average deviations smaller than 0.2 eV [2,4].

The application of our core-level *GW* method to disordered structures is computationally challenging due to the high computational cost of the method and the need for extensive structural sampling. To enable the prediction of XPS for amorphous materials, we developed a kernel ridge regression machine learning (ML) model based on a comprehensive database of density functional theory (DFT) and *GW* data. We applied our combined DFT-*GW*-ML approach to materials containing carbon, hydrogen and oxygen and showed that we obtain qualitative and quantitative agreement with experiment, resolving spectral features within 0.1 eV of reference experimental spectra [5].

Generating GW training data is the primary computational bottleneck in our DFT-ML-GWworkflow. I will present our recent algorithmic advances that reduce the system-size scaling of core-level GW calculations, cutting the cost of training-data generation by approximately an order of magnitude [6].

References

- [1] D. Golze, J. Wilhelm, M. van Setten, P. Rinke, J. Chem. Theory Comput., 14 (2018), 4856
- [2] D. Golze, L. Keller, P. Rinke, *J. Phys. Chem. Lett.*, 11 (2020), 1840
- [3] L. Keller, V. Blum, P. Rinke, D. Golze, J. Chem. Phys., 153 (2020), 114110
- [4] J. Li, Y. Jin, P. Rinke, W. Yang, D. Golze, *J. Chem. Theory Comput.*, 18 (2022), 7570
- [5] D. Golze, M. Hirvensalo, P. Hernández-León, A. Aarva, J. Etula, T. Susi, P. Rinke, T. Laurila, M. A. Caro, *Chem. Mater.*, *34* (2022), 6240
- [6] R. L. Panadés-Barrueta, D. Golze, J. Chem. Theory Comput., 19 (2023), 5450



Figure 1. XPS prediction of amorphous carbon structures with GW accuracy using a DFT-GW-ML model. Figure from Ref. [5].

Vibrational spectroscopy of molecules, condensed phases & interfaces using machine learning and quantum statistical mechanics

Venkat Kapil^{1,2}

¹ University College London, UK ² London Center for Nanotechnology, UK

v.kapil@ucl.ac.uk

Vibrational spectroscopy is a cornerstone technique for characterizing molecules, materials, and atomicscale phenomena. However, mapping vibrational spectra to local structure and dynamics remains challenging, requiring first-principles simulations or calculations. A truly predictive approach to vibrational spectra must accurately describe electronic structure for interatomic interactions and transition dielectric response, while also incorporating quantum nuclear motion to capture zero-point fluctuations, tunneling, and non-Condon effects.

In this talk, I will discuss our progress in modeling vibrational spectra with a fully quantum description by employing a combination of machine learning interatomic potentials [1,2,3], machine learning models of the dielectric response tensor [3,4], and a novel technique known as PIGS to incorporate quantum corrections to classical trajectories [5]. These approaches achieve high accuracy and computational efficiency while accounting for the aforementioned effects. I will present showcase examples ranging from the spectroscopy of small drug-like molecules [6] to the air-water interface [4].

References

- Kaur, H., Della Pia, F., Batatia, I., Advincula, X. R., Shi, B. X., Lan, J., Csányi, G., Michaelides, A., & Kapil, V. (2024). Data-efficient fine-tuning of foundational models for first-principles quality sublimation enthalpies. <u>https://doi.org/10.48550/</u> <u>ARXIV.2405.20217</u>
- Batatia, I., Benner, P., Chiang, Y., Elena, A. M., Kovács, D. P., Riebesell, J., Advincula, X. R., Asta, M., Avaylon, M., Baldwin, W. J., Berger, F., Bernstein, N., Bhowmik, A., Blau, S. M., Cărare, V., Darby, J. P., De, S., Della Pia, F., Deringer, V. L., ... Csányi, G. (2024). A foundation model for atomistic materials chemistry. arXiv. <u>https://doi.org/10.48550/</u> <u>ARXIV.2401.00096</u>
- Kovács, D. P., Moore, J. H., Browning, N. J., Batatia, I., Horton, J. T., Kapil, V., Witt, W. C., Magdău, I.-B., Cole, D. J., & Csányi, G. (2023). Mace-off23: Transferable machine learning force fields for organic molecules. <u>https:// doi.org/10.48550/ARXIV.2312.15211</u>

- Kapil, V., Kovács, D. P., Csányi, G., & Michaelides, A. (2023). First-principles spectroscopy of aqueous interfaces using machine-learned electronic and quantum nuclear effects. Faraday Discussions, 10.1039.D3FD00113J. <u>https://doi.org/10.1039/</u> D3FD00113J
- Musil, F., Zaporozhets, I., Noé, F., Clementi, C., & Kapil, V. (2022). Quantum dynamics using path integral coarse-graining. The Journal of Chemical Physics, 157(18), 181102. <u>https:// doi.org/10.1063/5.0120386</u>
- Pracht, P., Pillai, Y., Kapil, V., Csányi, G., Gönnheimer, N., Vondrák, M., Margraf, J. T., & Wales, D. J. (2024). Efficient composite infrared spectroscopy: Combining the double-harmonic approximation with machine learning potentials. Journal of Chemical Theory and Computation, 20(24), 10986–11004. <u>https://doi.org/10.1021/ acs.jctc.4c01157</u>

Figures



Figure. A schematic demonstrating quantum mechanical accuracy in predicting the O—H region of the IR spectrum of hexagonal ice at classical computational cost.

Phonons and anharmonicity in the time of machine-learning interatomic potentials

Aloïs Castellano¹, José Pedro Alvarinhas Batista¹ Matthieu Jean Verstraete^{1,2}

¹Nanomat, Q-mat research center and European Theoretical Spectroscopy Facility, Université de Liège, allée du 6 août, B-4000 Liège, Belgium ²ITP, Physics Department, Utrecht University 3508 TA Utrecht, The Netherlands

alois.castellano@uliege.be

Atomic vibrations are central to many physical properties of materials, from thermal conductivity and thermodynamic behavior to electrical transport and spectroscopic signatures. The workhorse to describe atomic vibrations in solids is the harmonic approximation, which reformulates the atomic of term phonons: quantized dynamics in quasiparticles of lattice vibrations. However, the simplicity of the harmonic approximation limits its accuracy, and contributions beyond the harmonic ones, named anharmonic, are often needed to describe the complex motion of atom in solids. To include anharmonicity, the most complete solution is to sample the canonical ensemble of the system. This is usually done with ab initio molecular dynamics where atoms positions evolve following forces computed from density functional theory, allowing then to compute properties through averages. However, the computational cost of the method limits the time and length scale accessible In recent years, through such simulations. machine-learning interatomic potentials (MLIPs) have significantly reduced this computational cost by replacing expensive DFT force evaluations with fast and accurate surrogate models trained on ab initio data. However, once a trajectory has been obtained, a key challenge remains: how to extract physically meaningful vibrational properties from raw atomic motion. This is where a phonon picture can come back to play a critical role. In this talk, I will introduce theories of anharmonic lattice dynamics [1, 2.3.4] that allows to interpret finite-temperature molecular dynamics in terms of renormalized phonon quasiparticles, thus providing direct access to spectroscopic observables or transport properties. In particular, I will show how combining these methods with MLIPs offers an efficient and accurate framework to study vibrational properties in materials well beyond the harmonic approximation.

References

- [1] A Castellano, JP Batista, MJ Verstraete, J. Chem. Phys., 159 234501 (2023)
- [2] A Castellano, JPA Batista, O Hellman, MJ Verstraete, Phys. Rev. B, 111, 094306 (2025)
- [3] O Hellman, IA Abrikosov, SI Simak, Physical Review B, 84, 180301 (2011)
- [4] O Hellman, P Steneteg, IA Abrikosov, SI Simak, Physical Review B, 87, 104111 (2013)



Figure 1. Phonon spectral function of FCC helium using anharmonic lattice dynamics (top) and perturbation theory (bottom). The green dot are experimental measurements.

Enhancing Vibrational Spectroscopy Predictions Using Machine Learning

Ke Chen¹, Sandra Luber¹

¹Department of Chemistry, University of Zurich, Zurich, Switzerland

ke.chen2@uzh.ch

Vibrational spectroscopy, including infrared (IR) and Raman spectroscopy, is a powerful tool for investigating molecular structure, bonding, and dynamics. However, accurately and efficiently predicting spectra remains challenging due to the complexity of vibrational interactions, anharmonic effects. Traditional computational methods, such as ones based on density functional theory (DFT), can offer accurate predictions but are often limited by their high computational cost, especially for large and complex systems^[1]. Machine learning (ML) has emerged as a powerful alternative, enabling efficient and scalable spectra prediction^[2]. In this work, we introduce an ML-based approach, which can accurately predict IR spectra and extends to systems subjected to external electric fields. Furthermore, we explore Raman spectra prediction under external electric fields, demonstrating the model's ability to capture field-dependent vibrational responses. The developed approach offers a fast and accurate solution for vibrational spectra prediction, bridging the gap between efficiency and precision in vibrational spectroscopic simulations.

- [1] Edward Ditler, Sandra Luber. *WIREs Comput Mol Sci.* 12, (2022), e1605.
- [2] Ruocheng Han, Rangsiman Ketkaew and Sandra Luber. *J. Phys. Chem. A*, 126, 6 (2022), 801-812.

Decentralized materials research data management, curation and dissemination for accelerated materials discovery

Dr Matthew L. Evans^{1,2,3}

¹ MODL-IMCN, UCLouvain, Louvain-la-Neuve, Belgium
 ² Matgenix SRL., Charleroi, Belgium
 ³ datalab industries Itd, King's Lynn, United Kingdom

matthew.evans@uclouvain.be

The primary barrier to widespread adoption of Al-accelerated materials science is the availability and quality of data. Researchers lack frictionless tooling and have limited incentive to record their data in such a way that is immediately amenable for machine learning, whether by them or by others. This talk introduces two data projects in the materials space that aim to lower the barrier to data access and curation by both humans and machines: the OPTIMADE federation of materials databases, and the open-source datalab materials data management platform.

OPTIMADE consists of an international consortium of databases that have designed, over many years, a common application programming interface (API) format [1], which now allows for 30+ databases across 20+ providers to be seamlessly queried. Such federated data unification enables decentralized data-driven workflows in materials informatics and beyond, from materials selection up to materials discovery [2]. OPTIMADE is supported by several community-oriented tools that allow others to easily contribute their data to this growing ecosystem. This talk will introduce the OPTIMADE ecosystem, discuss the process of consensus-forming amongst providers, and outline how OPTIMADE could be extended to other domains.

The second project primarily concerns experimental data; datalab [3] is a open-source data management platform that can be customized and adopted by materials research groups to allow for straightforward provenance tracking of samples, devices and raw data. It integrates with the broad open-source community of file format parsers (from the datatractor initiative and other popular packages) to allow for data normalization and simple analysis in the browser for many characterisation techniques (XRD, NMR, Raman, electrochemistry, etc). This platform provides the traditional benefits of having a digital system of record (e.g., an electronic lab notebook), whilst also enabling programmatic re-use of data across a research group via its API, with the aim to allow end user programming. By providing labs with control over their data platform, they can develop their own Al-driven developments, as well as selectively sharing and collaborating with others on shared workflows and samples. This talk will summarize the ongoing developments of datalab,

including the integration of AI-based agents, and motivate future use cases of a federation of such datalab deployments.

References

- [1] Evans, M. *et al.*, Digital Discovery, 3 (2024) 1509-1533. DOI: <u>10.1039/D4DD00039K</u>
- [2] Trinquet, V. *et al.*, Faraday Discussions, 256 (2025) 459-482. DOI: <u>10.1039/D4FD00092G</u>
- [3] Evans, M. and Bocarsly, J., <u>https://github.com/datalab-org/datalab</u> DOI: <u>10.5281/zenodo.13692122</u>



Figure 1. A network of interconnected experimental samples, devices and measurements captured using *datalab*.

Data- and theory-informed approaches for accelerating the capture and interpretation of corelevel spectra

Maria K. Y. Chan¹, ¹Center for Nanoscale Materials, Argonne National Laboratory

The use of advanced, including in situ and operando, core-level spectroscopy techniques has enabled detailed characterization of materials during synthesis and operations. Data-driven approaches, especially guided by computational spectroscopy, enable robust and accelerated information extraction from such spectroscopy data. In this talk, we discuss the featurization of x-ray absorption near edge structure (XANES) for extraction of local structure and electronic properties in battery cathode materials [1], and the extension thereof towards the detection of antisite defect and oxygen vacancy from multimodal EELS/XANES [2]. We also discuss the use of datadriven approaches to accelerate the acquisition [2] and analysis [3] of XANES mapping data. Finally, we discuss the use of AI to extract relevant labeled microscopy and spectroscopy data from scientific literature [4], and recent extensions involving the use of large language models and multi-model contrastive learning.

- [1] Y. Chen, C. Chen, I. Hwang, M. J. Davis, W. Yang, C.J. Sun, G. Lee, D. McReynolds, D. Allan, J. M. Arias, S. P. Ong, and M. K. Y. Chan, "Robust Machine Learning Inference from X-ray Absorption Near Edge Spectra through Featurization," Chemistry of Materials, 36, 5, 2304–2313 (2024).
- [2] H. Jia, Y. Chen, G. H. Lee, J. Smith, M. Chi, W. Yang, M. K. Y. Chan, "Revealing Local Structures through Machine-Learning-Fused Multimodal Spectroscopy," arXiv preprint arXiv:2501.08919.
- [3] S. Tetef, A. Pattammattel, Y. S. Chu, M. K. Y. Chan#, G. T. Seidler, "Accelerating nano-XANES imaging via feature selection," Digital Discovery 3(1), 201-209 (2024).
- [4] S. Tetef, A. Pattammattel, Y. S. Chu, M. K. Y. Chan#, G. T. Seidler, "Manifold Projection Image Segmentation for Nano-XANES Imaging," APL Machine Learning 1, 046119 (2023).
- [5] E. Schwenker, W. Jiang, T. Spreadbury, N. Ferrier, O. Cossairt, M. K. Y. Chan, "EXSCLAIM! -- Harnessing materials science literature for labeled microscopy datasets," Patterns 4, 100843 (2023).Corresponding author information

Automatic Quantification of Transition Metal X-ray Photoelectron Spectra using Convolutional Neural Networks

Lukas Pielsticker¹, Rachel L. Nicholls¹, Serena DeBeer¹, Walid Hetaba¹, Mark Greiner¹ ¹ Max Planck Institute for Chemical Energy Conversion, Mülheim an der Ruhr, Germany

lukas.pielsticker@cec.mpg.de

X-ray photoelectron spectroscopy (XPS) is a powerful tool for studying the electronic structure and chemical composition of solid surfaces. Quantitative analysis of XP spectra typically relies on manual curve fitting by expert spectroscopists. However, recent advancements in the ease of use and reliability of XPS instruments have led to a growing number of (novice) users generating large datasets that are becoming difficult to analyze manually. Additionally, the expansion of publicly available XPS databases further increases the volume of data requiring efficient analysis. Reflecting these developments, more automated techniques are desirable to assist users in processing large XPS datasets.

Here we present a scalable framework for automated XPS quantification using convolutional neural networks (CNNs). By training CNN models on artificially generated XP spectra with known quantifications (i.e., for each spectrum, the concentration of each chemical species is known), it is possible to obtain models for auto-quantification of transition metal XP spectra [1]. CNNs are shown to be capable of quantitatively determining the presence of metallic and oxide phases, achieving accuracy comparable to or exceeding that of conventional data analysis methods. The models are flexible enough to handle spectra containing multiple chemical elements and acquired under varying experimental conditions. The use of Bayesian CNNs and dropout variational inference for the determination of quantification uncertainty is discussed. Finally, we demonstrate how these network models can be integrated into research data management systems, enabling real-time analysis of newly generated data.

References

 Pielsticker, L.; Nicholls, R.; DeBeer, S.; Greiner, M., Analytica Chimica Acta, 1271 (2023) 341433.

Deep Generative Models for Tractable Probabilistic Inference over Molecular Graphs

Milan Papež², Martin Rektoris², **Václav Šmídl**^{1,2}, Tomáš Pevný²

¹University of West Bohemia, Univerzitni 8, Pilsen, Czech Republic

²OCzech Technical University, Karlovo namesti 13, Prague, Czech Republic

vsmidl@fel.zcu.cz

Various methods of deep learning are able to represent probability distributions over complex spaces. However, only a fraction of them is tractable, i.e. able to provide accurate answers to arbitrary probabilistic questions that are formulated using conditioning or marginalization. For example, a model that was trained as generative cannot be easily used as regressive. Probabilistic circuits are one class of models that are tractable, i.e. they are trained only once and answer arbitrary questions in a single pass through the model.

The aim of this talk is to introduce an extension of this framework that is capable of generating graphs representing molecules [1]. Due to inherent tractability, the same model can provide answers to many other tasks such as classification of the molecules, generation of molecules conditioned on their properties or subgraphs, selecting the most important part of the molecule, etc.

This property makes the model distinct from other deep generative models (DGMs) that have recently demonstrated remarkable success in capturing complex probability distribution over graphs [2]. Their excellent performance is attributed to powerful and scalable deep neural networks, it is, at the same time, exactly the presence of these highly non-linear transformations that makes DGMs intractable. That means that tasks for which the model was not trained can be answered only after retraining, sampling or other types of potentially expensive postprocessing.

We propose probabilistic graph circuits (PGCs), a framework of tractable DGMs that provide exact and efficient probabilistic inference over (arbitrary parts of) graphs. Nonetheless, achieving both exact and efficient inference is challenging for the following reasons:

- C1. Graphs live in large and complex combinatorial spaces. Indeed, estimated numbers of possible graphs in the molecular domain are enormous. This poses considerable requirements on the expressivity of DGMs.
- C2. Graphs are not random only in values of node and edge features but also in the number of these nodes and edges. Therefore, specific architectures accounting for this variable-size character of graphs are required.

- C3. Graphs are permutation invariant, i.e., there is a factorial number of possible configurations of a single graph. The key property of graph DGMs should be to recognize all the configurations as the same graph.
- C4. Graphs respect domain-specific semantic validity. For example, not all molecular graphs are chemically valid but must adhere to chemical valency constraints.

In this talk, we will outline the principles of solution of these challenges such as marginalization padding (C2) and canonical graph ordering (C3). However, the main focus will be on demonstration of the model capabilities of tractable inference. Specifically, we will demonstrate generation of novel molecules that are consistent with trainig datasets QM9 and ZINC250k for completely novel molecules or completion of molecules with predefined subgraph, Figure 1.

References

- [1] Milan Papež, Martin Rektoris, Václav Šmídl, Tomáš Pevný, Probabilistic Graph Circuits: Deep Generative Models for Tractable Probabilistic Inference over Graphs. UAI 2025. under review.
- [2] Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. Digress: Discrete denoising diffusion for graph generation. In 11th International Conference on Learning Representations, 2023

Figures \checkmark <t

Figure 1. Generated samples from a model trained on QM9 dataset for a given part of the graph. The given part is highlighted by yellow color.



Contributed talks

ML4Spec2025

Day 1 **(26.05)**

•15:45-16:00 Malte Grunert: *OptiMate: ML for UV-VIS spectra of crystalline solids*

•16:00-16:15 Nathalie Brun: *EELS hyperspectral images unmixing using autoencoders*

•16:15-16:30 Dario Baum: *Fast and Simple Complete Basis Set qsGW Calculations*

•16:30-16:45 Arno Annys: *Nanoscale spectroscopy inside the electron microscope*

•16:45-17:00 Sabana Shabnam: *Coupling molecular dynamics (MD) with multiple scattering calculations for the understanding of time dependent surface phenomena*

Day 2 (27.05)

•15:45-16:00 Della Valle Andrea: *Machine learning recognition of Volatile Organic Compounds absorption spectra based on experimental and synthetic data*

•16:00-16:15 Laurens de Boer: *Carotenoids in algal biomass: coupling quantum chemistry to AI to solve the conformer puzzle*

•16:15-16:30 Max Großmann: *Towards machine learning of dielectric functions of metals: Dataset creation, challenges, and opportunities*

•16:30-16:45 Daria M. Tomecka: *Generating, sorting and standardizing microscopy data with Artificial Intelligence*

•16:45-17:00 Joël Eymery: *Non-negative Matrix Factorization for Excited X-Ray Luminescence*

Day 3 **(28.05)**

•14:30-14:45 Mandrira Das: *Decoding Aerosol Surface Chemistry: Insights from XPS Spectra via DFT and Machine Learning*

•14:45-15:00 Thomas P. van Waas: *Learning the many-body properties from angle-resolved photoemission spectroscopy*

•15:00–15:15 Trung-Phuc Vo: *Theoretical and experimental HARPES study of Weyl-semimetal TaAs: The application of machine-learning*

•15:15-15:30 Strocov Vladimir: *Photoemission Final States Beyond the Free-Electron Approximation*

•15:30–15:45 Federico Zecchi: *Exploring the Limits of Machine Learning Algorithms for the Classification of X-ray Absorption Spectra*

OptiMate: ML for UV-VIS spectra of crystalline solids

Malte Grunert, Max Großmann, Erich Runge Technische Universität Ilmenau, Ilmenau, Germany

malte.grunert@tu-ilmenau.de

We present the OptiMate family of state-of-the-art Graph Attention Networks [1], capable of predicting the optical spectra of crystalline materials directly from their crystal structure in a matter of milliseconds.

Optical properties in the visible and UV region are crucial for many technologically relevant applications, such as photovoltaics, optoelectronics, etc. In recent years, large databases of crystal structures have become available [1, 2], which are available to be scanned for various functional properties with the intent of discovering promising new materials for diverse applications.

However, compared to many other properties, the optical properties of materials are extremely costly to calculate using ab-initio methods, and screening large databases using them is out of the question. Machine Learning (ML) therefore is a promising alternative.

The high computational cost also complicates machine learning efforts, as all major materials databases, like the Materials Project [1] or the Alexandria database [2], lack high-quality optical data. Apart from the limited training data available, optical properties also pose additional challenges from the machine-learning perspective, being inherently frequency-dependent (i.e., high-dimensional) and tensorial in nature.

We have developed in-house ab-initio workflows to efficiently calculate the optical properties of crystalline solids in the UV-VIS spectral range at various levels of theory, going from PBE-IPA [3] to QSGŴ-BSE [4]. Based on the thus generated data, we have trained state-of-the-art Graph Attention Networks which can predict the optical properties in a wide frequency range.

In this talk, we will present this so-called OptiMate family of models [3-5]. The spectra generated by the models are quantitatively accurate (see Fig. 1) and show a surprising amount of "physical insight": Without explicit constraints in either the model architecture or the training process, the generated spectra are smooth, artifact-free and respect the Kramers-Kronig relation. In addition, even for challenging materials, where peak intensities are not incorrect, peak positions are often predicted well.

Focusing on the OptiMate models for semiconductors and insulators, we will outline the model architecture and the training data, and show various performance metrics. In addition, we show results for transfer learning to higher levels of theory [5] and what kind of datasets we believe are necessary to reach experimental accuracy (see Fig. 2).

References

- [1] A. Jain et al., APL Mater. 1 (2013), 011002
- [2] J. Schmidt *et al.*, Mater. Today Phys. **48** (2024) 101560
- [3] M. Grunert, M. Großmann, E. Runge, Phys. Rev. Mater. 8 (2024) L122201
- [4] M. Grunert, M. Großmann, E. Runge, in preparation
- [5] M. Grunert, M. Großmann, E. Runge, submitted

Figures



Figure 1. Comparison between ab-initio calculated spectra (blue) and spectra predicted by OptiMate (orange). Materials are chosen from the test set based on the quantile of similarity between ab-initio and ML selected spectra as indicated on the right, i.e., the bottommost row are some of the worst performing materials. Image reproduced from [3].



Figure 3. Performance of transfer learning (orange) versus direct learning (green) on the test set when going from PBE-IPA to PBE-RPA, i.e., including local-field effects. In general, an order of magnitude less training data is necessary when transfer learning from a model pretrained on data from a lower theory level to achieve similar error metrics. Shown in blue are results for transfer learning only on small cells, with the maximum number of sites per unit cell indicated by the small number next to each data point.

Image reproduced from [5].

EELS hyperspectral images unmixing using autoencoders

Nathalie Brun¹, Laura Bocher¹

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

nathalie.brun@universite-paris-saclay.fr

Recent advances in analytical electron microscopy allow the acquisition of hyperspectral images (HSI), with two spatial and one spectral dimension. In Scanning Transmission Electron Microscopy (STEM), a focused electron beam scans the sample, and for each position, a spectrum is recorded. This spectrum, obtained through Electron Energy Loss Spectroscopy (EELS), provides information on the material's composition and electronic structure by measuring how much energy the electrons lose when interacting with the sample [1].

Classical EELS data processing techniques can be applied to each spectrum individually. However, since the HSI dataset contains redundant information, it is more efficient to analyze it as a whole. Statistical methods such as Principal Component Analysis (PCA) are commonly used to reduce noise in HSI. But the ultimate goal of data processing is to transform the three-dimensional dataset (position x position x spectrum) into a set of two-dimensional maps, each associated with characteristic spectra representative of the sample's different components.

A direct approach to obtaining these physically meaningful components and their spatial distribution is spectral unmixing. This technique assumes that each recorded spectrum can be expressed as a combination of a few characteristic spectra corresponding to the materials present in the sample. However, spectral unmixing is a challenging problem, and various methods have been developed in the hyperspectral imaging community to address it [2].

In this study, we assess the potential of neural networks, specifically Autoencoders (AE) (Fig. 1), for spectral unmixing. We generate synthetic HSI to quantitatively evaluate the accuracy of the unmixing process (Fig. 2) and then apply these methods to experimental data acquired from a Pt/Co/Ru/Pt multilayer. We initially tested relatively simple AE models, which produced interesting results but did not significantly outperform classical approaches [3]. To further improve performance, we explored more advanced architectures, incorporating Transformers, which leverage attention mechanisms to selectively focus on the most relevant spectral features. Unlike traditional neural networks, Transformers dynamically weigh the importance of different parts of the input data, enabling more effective spectral unmixing. This choice was motivated by the rapid

progress in the field, as new models are introduced almost every month, pushing the boundaries of hyperspectral data analysis. Our results show that AE-based methods hold promise for STEM-EELS spectral unmixing and highlight the potential of deep learning techniques for extracting meaningful physical information from complex spectral datasets.

References

- [1] Kociak, M., et al. "Spatially resolved EELS: The spectrum-imaging technique and its applications." in *Scanning Transmission Electron Microscopy: Imaging and Analysis.* New York, NY: Springer New York (2010) 163-205.
- [2] Dobigeon, N., and Brun, N., Ultramicroscopy 120 (2012) 25-34.
- [3] Brun, N., Lambert G., and Bocher.L., The European Physical Journal Applied Physics 99 (2024): 28.
- [4] Palsson, B., et al., *IEEE Access* 6 (2018): 25646-25656.







Figure 2. Comparison of unmixing algorithm performance for endmember extraction. Neural network-based methods are shown in orange, while traditional methods are in green. The evaluation metric used is the Spectral Angle Distance (SAD).

Fast and Simple Complete Basis Set qsGW Calculations

Dario Baum¹, Arno Förster¹, Luuk Visscher¹ ¹Vrije Universiteit Amsterdam, De Boelelaan 1105 1081HV, Amsterdam, Netherlands

d.d.baum@vu.nl

The GW approximation is a widely used method for computing quasiparticle energies of crystalline as well as molecular systems, providing accurate predictions of ionization potentials and electron affinities with higher accuracy than DFT while being more efficient than wavefunction methods. However, its practical applicability is often limited by the slow convergence of computed energies with respect to the basis set size. In this talk, I demonstrate that a recently proposed and simple complete basis set (CBS) limit extrapolation method [1] can be generalized to quasiparticle self-consistent GW (qsGW) calculations. Further, I demonstrate that commonly used basis set extrapolation methods underestimate the CBS limit of GW. Finally, and based on that extrapolation scheme I showcase a large dataset of highly accurate qsGW quasi-particle energies and GW-BSE excitation energies for Machine Learning and benchmarking purposes.

References

[1] Bruneval, Fabien, et al. "Extrapolating unconverged GW energies up to the complete basis set limit with linear regression." *Journal of Chemical Theory and Computation* 16.7 (2020): 4399-4407



Figure 1. Fast and Simple qsGW

Nanoscale spectroscopy inside the electron microscope

Arno Annys^{1,2}, Daen Jannis^{1,2}, Jo Verbeeck^{1,2} ¹Electron Microscopy for Materials Science (EMAT), University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium

²NANOlight Center of Excellence, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium

arno.annys@uantwerpen.be

The electron microscope is not just a powerful tool for the imaging of materials at resolutions routinely reaching the atomic scale, it is a versatile characterization tool providing structural and chemical information through imaging, diffraction and spectroscopy. The two main spectroscopic techniques in the transmission electron microscope (TEM) are electron energy loss spectroscopy (EELS) and energy-dispersive X-ray spectroscopy (EDX).

EELS provides insight into a broad range of physical properties through inelastic scattering events in the meV range (phonon, exciton), the eV range (plasmon) and the 100 eV-keV range (core electron transitions), with nanoscale spatial resolution. The most common way of performing EELS and EDX is in the scanning TEM mode (STEM) where an ångström-sized probe is scanned over the specimen, providing large and information-rich hyperspectral images. In this work, we will focus on core-loss EELS, which provides for example chemical information like oxidation state and bonding. EXD provides related chemical information, and we demonstrate how coincidence detection of EELS and EDX can improve the information retrieval [1].

For over two decades, model-based quantification using the maximum likelihood method for the estimation of physical parameters describing an EELS spectrum has been the golden standard [2]. Recently, the parametric model has been refined further [3,4] and an open-source python project has been released compiling the state-of-the-art modelbased quantification methods [5].

Inversion of the established parametric models for core-loss spectra provides an ideal simulation approach for large amounts of synthetic EELS data, proven to accurately represent experimental data, that can be used to train deep learning models such as neural networks. These neural networks can fill the gaps in processing workflows where modelbased methods do not provide the answer, in this case automatic identification of chemical content [6]. By applying the output of the deep learning model as input to the model-based approach, we can form entirely unsupervised quantification workflows which are urgently needed to cope with the ever-increasing amounts of data generated in modern STEM-EELS experiments. At the same time, they offer the potential to remove the dependency on tuning parameters that inevitably lead to experimenters' bias and reproducibility issues that can plague EELS quantification methods.

References

- D. Jannis et al. Coincidence Detection of EELS and EDX Spectral Events in the Electron Microscope. Appl. Sci. 11 (2021) 10.3390/app11199058
- J. Verbeeck & S. Van Aert, Model based quantification of EELS spectra. Ultramicroscopy 101 (2004) 10.1016/j.ultramic.2004.06.004.
- [3] W. Van den Broek, D. Jannis & J. Verbeeck, Convexity constraints on linear background models for electron energy-loss spectra. Ultramicroscopy 254 (2023) 10.1016/j.ultramic.2023.113830.
- [4] D. Jannis et al. Improved precision and accuracy of electron energy-loss spectroscopy quantification via fine structure fitting with constrained optimization. Ultramicroscopy 269 (2025) 10.1016/j.ultramic.2024.114084.
- [5] D. Jannis et al. pyEELSMODEL. github.com/joverbee/pyEELSMODEL
- [6] A. Annys, D. Jannis & J. Verbeeck, Deep learning for automated materials characterisation in core-loss electron energy loss spectroscopy. Scientific Reports 13 (2023) 10.1038/s41598-023-40943-7

This work received funding from the Horizon Europe framework program for research and innovation under grant agreement n. 101094299 (IMPRESS)



Figure 1. Example of an unsupervised quantification workflow of a STEM-EELS dataset using neural network element identification as input for model-based quantification.

Coupling molecular dynamics (MD) with multiple scattering calculations for the understanding of time dependent surface phenomena

S. Shabnam, M. Vassaux, P. Schieffer, D. Sébilleau, S. Tricot

Univ Rennes, CNRS, IPR(Institut de Physique de Rennes)- UMR 6251, F-35000, Rennes, France

sabana.shabnam@univ-rennes.fr

Recently, the time resolution of some experimental techniques has improved to the point where nonequilibrium microscopic processes can be observed. These new instruments are particularly interesting for studying materials and their surfaces. Timeresolved X-ray photoelectron diffraction (TR-PED) is one of the emerging fields among these. It is capable of giving access to chemically resolved local geometric properties with a high sensitivity to the extreme surface conditions (adsorption site of a molecule, its orientation, relaxation of the surface lattice parameter etc.). [1,2]

However, experimental results need to be compared with a theoretical model in order to obtain maximum information and accuracy using a complex numerical simulation of the multiple scattering process at work. [3]

Our work focuses on the study of TR-PED at finite temperature using molecular dynamics (MD) simulation for getting the positions of the atoms of the system with time. These temperaturedependent atomic motions are used to compute photoelectron diffraction patterns within the multiple scattering approach. Once integrated in time, the simulations give a more accurate picture of the temperature effect on the shape of photo diffraction peaks, and allow us to compare these results with a non time-resolved calculation in which temperature is modelled by statistically adding decoherence to photoelectrons.

- [1] Multiple Scattering Theory for Spectroscopies, Didier Sébilleau, Keisuke Hatada, Hubert Ebert, Springer Proceedings in Physics (vol. 204).
- [2] Ultrafast Spectroscopy: State of the Art and Open Challenges, M. Maiuri, M. Garavelli, G. Cerullo, J. Am. Chem. Soc. 142(2020) 3-15.
- [3] Energy dependence of interference phenomena in the forward-scattering regime of photoelectron diffraction, S. Tricot, T. Jaouen, D. Sébilleau, P. Schieffer, Journal of Electron Spectroscopy and Related Phenomena 256 (2022) 147176.

Machine learning recognition of Volatile Organic Compounds absorption spectra based on experimental and synthetic data

Della Valle, A.¹, D'Arco, A.^{2,3}, Lupi, S.^{3,4}, Pilati, S.^{5,6}, Perali A.¹

¹School of Pharmacy, University of Camerino, Camerino, Italy
 ²INFN-LNF, Frascati, Italy
 ³Department of Physics, University of Rome 'La Sapienza', Rome, Italy
 ⁴INFN-Sezione di Roma, Rome, Italy
 ⁵School of Science and Technology, University of Camerino, Camerino, Italy
 ⁶INFN-Sezione di Perugia, Perugia, Italy

andrea.dellavalle@unicam.it

Volatile Organic Compounds (VOCs) are characterized by a low boiling point, and they easily evaporate in air. They can be emitted from different sources, both natural and anthropic, contributing to air pollution with dangerous effects on the environment and on human health [1]. Machine learning is increasingly used for the analysis of spectral data [2], suggesting that it might be suitable for the identification of VOCs.

The first part of this work focuses on the realization of a deep learning model for the classification and the quantitative analysis of VOCs' absorption spectra. We generate the training datasets with a multipass gas cell in a Vertex 70v Michelson interferometer, measuring low-concentration VOCs [3].

For each spectrum, the dataset provides the molecular identity, and the concentration of such VOC measured in *ppm*. The dataset is composed of 970 spectra divided in acetone, benzene, ethanol, isopropanol, m-xylene, o-xylene, p-xylene, styrene, and toluene.

The training phase of a deep neural network requires a sufficiently large dataset of examples, which is not always available in this context. For this reason, the second part of this work investigates the creation of synthetic data using a conditional variational autoencoder (cVAE) [4]. The decoder of the cVAE is used to generate synthetic data similar to the experimental spectra. This model extends the generation of new spectra to conditions not present in the original dataset (Fig. 1), e.g., to different concentrations.

This work is divided into three steps:

i) Training of the *master* model on the original experimental dataset to predict molecular identity and concentration;

ii) Training of the cVAE conditioned on the molecular identity and the concentration of the VOC;
iii) Training of the *slave* model on a dataset augmented with synthetic spectra.

The *master* and *slave* models use the same architecture: the model is a deep convolutional neural network, with a sequence of convolutional layers and two heads: a classifier-head recognising the molecular identity, and a multi-regression-head predicting the concentration of the VOC.

In each epoch of the *slave* models' training, the trained conditional decoder is used to generate the synthetic data, with randomly sampled concentrations. These synthetic data are then added to the original training dataset.

The amount of added data is changed to train different *slave* models, and the performance of the model associated with the minimum mean square error is reported in Fig. 2.

The use of cVAE to generate synthetic data not only allows augmentation to train new models but also permits to generate new spectra with concentrations not present in the original training dataset.

References

- [1] McDonald, *et al.* Science, 6377 (2018) 760-764
- [2] Mishra, et al. TrAC, 2 (2022) 116804
- [3] D'Arco, et al. Sensors, 15 (2022) 5624
- [4] Harvey, *et al.* arXiv:2102.12037 (2022)

Figures



Figure 1. Examples of generated (red, and yellow lines), and original experimental spectra (blue lines) of Isopropanol. Red lines indicate spectra with concentrations available in the dataset, the yellow lines refer to concentrations not present in the dataset.



Figure 2. Comparison of performance of *master* and *slave* models divided per class on test set.

Carotenoids in algal biomass: coupling quantum chemistry to Al to solve the conformer puzzle

Laurens de Boer¹, Héla Friha^{1,2}, Roberto Linguerri¹, Tsveta Miteva³, Majdi Hochlaf¹,

¹University Gustave Éiffel, COSYS/IMSE, 5 Bd Descartes 77454, Champs-sur-Marne, France. ²Institut Polytechnique des Sciences Avancées, 63 Bd de Brandebourg Bis 94200, Ivry-sur-Seine, France. ³Laboratoire de Chimie Physique-Matière et Rayonnement (LCPMR), UMR 7614, CNRS, Sorbonne Université, 4 Place Jussieu 75005 Paris, France

laurens.de-boer@univ-eiffel.fr

Biofuels derived from biomass are a viable and sustainable alternative to fossil fuels, but their largescale implementation is hampered by high production costs [1]. The optimization of (photo)bioreactors for biofuel production relies on the precise characterization of algal biomass, particularly its organic constituents [2]. RAMAN spectroscopy is a highly effective technique for this analysis [3]. However, fully characterizing these compounds in biomass is challenging due to the vast number of possible conformers and isomers (>10⁴) present in the mixture. A key challenge lies thus in differentiating the spectral contributions of individual compounds and accurately identifying their conformers within complex mixtures [2].

At present, we treat the case of (α , β , γ)-carotenoids as a case study. First, an experimental investigation was carried out to study these compounds in a culture of *Chlamydomonas reinhardtii* microalgae. Specifically, a micro-Raman spectroscopy analysis was carried out using a Horiba X-plora confocal microscopy Raman spectroscopy system, equipped with 10× and 50× long-working-distance objectives and a charge-coupled device (CCD) detector (see Figure 1). The Raman excitation source is provided by a 638 nm LED laser beam, with a beam power of 6 mW, focused on the sample with a spot size of approximately 1 µm in diameter at 532 nm.

The next step was to conduct a theoretical study combining first-principles techniques with various statistical tools to clarify the contributions of the different molecular species and relative conformers to the measured spectra. First, we predict the structure of the various conformers using the GFN-FF method [4,5,6]. Then, we compute their RAMAN spectra using the Density Functional Theory (DFT). A subsequent Machine Learning (Linear regression) coupled to clustering (k-Means) treatment allows to minimize the redundancy in the generated conformer dataset. Afterwards, by combining the Spectral Angle Mapper (SAM) analysis and Principal Component Analysis (PCA), we were able to successfully identify the spectral contribution and key regions for these conformers (see Figure 1). In fact, the comparison of the computed RAMAN spectra to that of microalgae (*Chlamydomonas reinhardtii*) allowed the identification of the few (α , β , γ)-carotenoids conformers present in such biological media.

In conclusion, our approach offers a viable framework for the rapid analysis of the contribution of individual conformers of (α , β and γ)-carotenoids and more generally of precursors in algal biomass.

References

- [1] C. S. Jones et al., *Current Opinion in Biotechnology*, **2012**, 23, 346–351.
- [2] Miteva, T. et al., Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, **2025**, Under review.
- [3] H. Wu, et al., *Proceedings of the National Academy of Sciences*, **2011**, *108*, 3809–3814.
- [4] Spicher S. et al., *Angewandte Chemie* International Edition, **2020**, 59, 15665–15673
- [5] Ehlert, S., et al. , **2022**, *The Journal of Physical Chemistry A*, 22, 3521-3535.
- [6] P. Pracht, et al., *Phys. Chem. Chem. Phys.*, **2020**, 22, 7169–7192.



Figure 1. A) cluster of α -carotenoid conformers generated by the GFN-FF method. B) overlay of the generated RAMAN spectra of the theoretically calculated α carotenoid conformers, and the experimental RAMAN spectra of *Chlamydomonas reinhardtii* microalgae. C) SAM analysis of the RAMAN spectra obtained by DFT of the theoretically generated α -carotenoid conformers. D) PCA of all generated RAMAN spectra of the theoretically calculated α -carotenoid conformers, and the experimental RAMAN spectra of *Chlamydomonas reinhardtii* microalgae.

Towards machine learning of dielectric functions of metals: Dataset creation, challenges, and opportunities

Max Großmann, Malte Grunert, Erich Runge Technische Universität Ilmenau, Ilmenau, Germany

max.grossmann@tu-ilmenau.de

The dielectric functions of metals provide critical insight into plasmonic, photonic, and electronic phenomena. However, no comprehensive dataset exists for these properties, in part due to the computational challenge of converging the dielectric function with respect to Brillouin zone sampling, particularly the Drude tail. Here, we address this gap by performing high-throughput *ab initio* calculations on 200,000 intermetallic compounds from the Alexandria database [1]. These results represent the presumably first and largest dataset of metal-dielectric functions assembled to date.

Central to our approach is the Prandini et al. [2] SIMPLE code, which implements the optimal basis method originally proposed by E. L. Shirley [3], allowing us to easily interpolate k-point grids. This method allowed reliable convergence of the dielectric function and the Drude tail, facilitated by our in-house high-throughput workflow. To verify the accuracy of our data, we compared calculated dielectric spectra for simple metals with experimental measurements. The agreement is already quite good, which highlights the robust nature of our workflow, see for example Figure 1. Nevertheless, our results confirm a well-known limitation of density functional theory in predicting the position of *d*-bands and their associated absorption peaks, as is the case for elemental Cu, Ag, and Au.

We present first results on learning the dielectric function from our newly generated dataset. In doing so, we face several challenges: the handling of the Drude tail in machine learning models, in particular its divergence in the static limit, and the accurate joint prediction of real and imaginary parts. Finally, we outline prospects for transferring a model to quasiparticle self-consistent GW (QSGW) calculations performed in QUESTAAL [4], where the *d*-band position is significantly improved, as illustrated for Au in Figure 2.

Our work paves the way for rapid and reliable prediction of dielectric properties in a wide range of metallic systems, likely enabling and supporting future discoveries in plasmonics, photonics, and materials design.

References

[1] Schmidt, J. et al., Improving machine-learning models in materials science through large

datasets, *Mater. Today Phys.* **48**, 101560 (2024).

- [2] Prandini, G. et al., SIMPLE code: Optical properties with optimal basis functions, *Comput. Phys. Commun.* 240, 106–119 (2019).
- [3] Shirley, E. L., Optimal basis sets for detailed Brillouin-zone integrations, *Phys. Rev. B* 54, 16464–16469 (1996).
- [4] Pashov, D. *et al.*, Questaal: a package of electronic structure methods based on the linear muffin-tin orbital technique, *Comput. Phys. Commun.* **249**, 107065 (2020).
- [5] Palik, E. D., Handbook of Optical Constants of Solids, New York: Academic (1985).
- [6] Babar, S. & Weaver, J. H., Optical constants of Cu, Ag, and Au revisited, *Applied Optics* 54, 477 (2015).



Figure 1. Comparison of the imaginary (solid lines) and real (dashed lines) parts of the dielectric function of Ta obtained from experiment (black) [5] and our *ab initio* workflow (orange).



Figure 2. Comparison of the imaginary (solid lines) and real (dashed lines) parts of the dielectric function of Au obtained from experiment (black) [6] and advanced *ab initio* calculation. In particular, we compare all-electron LDA (orange, similar to PBE) with all-electron QSGW calculations (blue) performed in QUESTAAL [4].

Generating, sorting and standardizing microscopy data with Artificial Intelligence

Daria M. Tomecka¹, S. Khalatabad², R. Jaeken¹, T. Nguyen Minh², L. Kestens², S. Cottenier², M. Sluydts¹

¹ePotentia, Frans van Dijckstraat 59, B-2100 Deurne Antwerpen, Belgium,

²Ghent University, Department of Electromechanical, Systems and Metal Engineering, Tech Lane Ghent Science Park - Campus A, Technologiepark 131, B-9052 Gent, Belgium

daria.tomecka@epotentia.com

Abstract

Modern Artificial Intelligence (AI) methods offer the potential to revolutionize the prediction of material properties and the optimization of processes. However, creating robust AI models requires access to large standardized databases [1,2], which are not only rare but also difficult to combine due to licensing restrictions and confidentiality concerns. Available data is often unstandardized and poorly labeled, making it challenging to use effectively.

Within the AID4GREENEST project [3], we develop self-supervised methods to automatically sort and curate large scanning electron microscopy datasets of steels, organizing them into a structured microstructural space [4]. Analyzing this space allows us to identify imperfections and imbalances within datasets. We further demonstrate synthetic data techniques that interpolate within this space to generate additional image data, helping models better account for expected variations in real data. These approaches lay the groundwork for improved AI-driven materials characterization and, in the future, can support optimal data production through CHADA documentation [5].

References

[1] DeCost, B. L., Hecht, M. D., Francis, T., Webler, B. A., Picard, Y. N., & Holm, E. A. (2017). UHCSDB: UltraHigh Carbon Steel Micrograph DataBase: Tools for Exploring Large Heterogeneous Microstructure Datasets. Integrating Materials and Manufacturing Innovation, 6(2), 197–205. https://doi.org/10.1002/040102.047.0007.0

https://doi.org/10.1007/s40192-017-0097-0

- [2] MicrostructureDB. (2025). [Online]. Available: https://www.microstructuredb.com
- [3] AID4GREENEST. (2025). [Online]. Available: https://www.aid4greenest.eu

- [4] Larmuseau, M., Sluydts, M., Theuwissen, K., Duprez, L., Dhaene, T., & Cottenier, S. (2020). Compact representations of microstructure images using triplet networks. Npj Computational Materials 2020 6:1, 6(1), 1–11. https://doi.org/10.1038/s41524-020-00423-2
- [5] "Materials characterization Terminology and structured documentation," CEN Workshop Agreement 17815:2025, 2025.

Figures



Figure 1. Mapping the microstructure space: Each point represents a single micrograph. On the right, a zoomed-in micrograph is shown; can you determine if it is real or synthetic?

Non-negative Matrix Factorization for Excited X-Ray Luminescence

Joël Eymery, B. De Goes Foschiani, O. Ulrich Univ. Grenoble Alpes, CEA, IRIG, MEM, NRX, 38000 Grenoble Cedex, France

joel.eymery@cea.fr

The light management in nitride materials is directly related to the growth and technology process controls, but the development of advanced characterization techniques with high spatial resolutions also plays a major role. Noticeably, focused X-ray beams provide an efficient probe to analyse quantitatively and correlatively the strain and light emission by combining µLaue diffraction (µLaue) and X-ray excited optical luminescence (XEOL), the two signals being recorded for the same time during mappings. This work will illustrate some recent experimental and analysis breakthroughs obtained at the BM32 beamline of the European Synchrotron Radiation Facility. A focus will be done on the XEOL analysis and illustrated with mappings of nitride materials in terms of structural analysis (epitaxial relationships, strain, orientation) and light emission. The hyperspectral analysis by Nonnegative Matrix Factorization (NMF) can be used to decompose the light emission in components of signals coming from the different parts of the sample and with real physical interpretations. It can be compared to other techniques. e.g., cathodoluminescence with electron probes and photoluminescence with laser light excitation.

The light emission of GaN µwires [1-3] and µLEDs obtained by etching GaN/InGaN Multiple Quantum Well (MQW) of commercial MOVPE UV heterostructures are studied by XEOL hyperspectral analysis, and the local strain variation and lattice rotation is obtained from µLaue analysis. A complete mapping (e.g., 40000 spectra of 1024 channels) of µLED takes benefits from the small beam size (~ 250 nm), short counting time (~1 s) and from the polychromatic diffraction Laue pattern method that can record many Bragg reflections without rocking the sample.

The XEOL data are analysed in detail with the Alrelated method of Non-negative Matrix Factorization. The definition of the number of components, the use of GPU, the number of trials to find the solution, as well as the count statistics to describe the counting will be discussed [4]. It is shown that the three main emissions of the samples (MQW, near band edge peaks and defects band) can be directly retrieved in a fast and "ab initio" way with the NMF method.

The results of the combination of both methods enable the correlation of the visible emission and the crystalline structure of the materials, and therefore to improve manufacturing techniques. It will be also demonstrated that fast scans allow for obtaining a statistical description of the samples opening the way to production control and a fast and systematic screening of optoelectronic materials and microstructures. New potential applications for different MOVPE materials will be given in this communication.

References

- [1] R. Koester R et al., Nano Lett. 11 (2011) 4839.
- [2] D. Salomon et al., Nano Lett. 17 (2017) 946.
- [3] J. Segura-Ruiz et al., Nano Letters 21 (2021) 9494.
- [4] M. Shiga, S. Muto, e-Journal of Surface Science and Nanotechnology 17 (2019) 148.

Figures



Figure 1. Left: XEOL/ μ Laue measurement schematics. SEM-view of μ LED samples and schematics. Right: XEOL hyperspectral mappings of the μ LEDs showing the panchromatic spectrum and image, and the individual mappings of near-band edge, blue and defect-band emission.



Figure 2. Finding two low-rank non-negative matrices $W \in \mathbb{R}^{(M^{\times}K)}$ and $H \in \mathbb{R}^{(N^{\times}K)}$ whose product can well approximate the matrix X, $(X = W H^{T})$, W is called the basis matrix, H the encoding or mixture coefficient matrix, K decomposition rank. Top: decomposition with 4 components. Bottom: the 4 mixture coefficients mappings for the different physical signals. Residual at the bottom right.

Acknowledgements: Work granted by the two French projects Magnifix (PIA3 Equipex+ ANR-21-ESRE-0011) and DIADEM-ESRF(PIA4 22-PEXD-0011).

Decoding Aerosol Surface Chemistry: Insights from XPS Spectra via DFT and Machine Learning

Mandira Das^{1*}, Abdurrahman Adhyatma¹, Jack Lin², Nønne Prisle², Milica Todorović¹

¹Materials Informatics Laboratory, University of Turku, Vesilinnantie 5, Turku 20500, Finland

²Center for Atmospheric Research (ATMOS), University of Oulu, Pentti Kaiteran katu 1, 90570 Oulu, Finland

mandda@utu.fi

Aerosols, which are nano- to microscale particles suspended in the air, have a significant influence on climate, weather, health, and ecology. The size and composition of aerosol particles determine their interactions with atmospheric compounds. Among these, Sodium Chloride (NaCl) is the most abundant aerosol particle. Surface-sensitive Ambient Pressure X-ray Photoelectron Spectroscopy (APXPS) has revealed that NaCl aerosol particles undergo structural transformations depending on atmospheric humidity levels [1]. However, understanding the surface atomic arrangement that governs the interaction of NaCl aerosols with the atmosphere under humid conditions remains an ongoing challenge.

To address this challenge, we employ Bayesian Optimization Structure Search (BOSS) [2] code combined with DFT to model the adsorption geometry of atmospheric water on the NaCl aerosol surface. BOSS code samples various configurations of atmospheric water on aerosol surfaces, enabling the learning of adsorption energy landscapes. Once the adsorption geometry is optimized, we apply the Δ self-consistent field (Δ SCF) [3] approach to compute the core electron binding energy of the Na 1s binding energy before and after atmospheric water adsorption, we can interpret changes in experimental XPS spectra under humid conditions.

This study leverages ML-driven DFT to reveal atomic-scale interactions between aerosols and atmospheric water, providing insights into APXPS data. Additionally, it serves as a framework for exploring interactions between aerosols and more complex atmospheric compounds.

References

[1] Lin, J. J et. al., Pre - deliquescent water uptake in deposited nanoparticles observed with in situ ambient pressure X-ray photoelectron spectroscopy, Atmos. Chem. Phys., 21, 4709–4727

- [2] Järvi, J. et. al., Detecting stable adsorbates of (1S)-camphor on Cu (111) with Bayesian optimization, Beilstein J. Nanotechnol. 2020, 11, 1577–1589
- [3] Kahk, J. M. *et. al.*, Core Electron Binding Energies in Solids from Periodic All-Electron Δ- Self-Consistent-Field Calculations, J. Phys. Chem. Lett. 2021, 12, 9353–9359





Learning the many-body properties from angle-resolved photoemission spectroscopy with Bayesian inference

Thomas P. van Waas¹, Christophe Berthod², Jan Berges³, Nicola Marzari^{3,4,5}, J. Hugo Dil^{6,7}, and Samuel Poncé^{1,8}

¹ European Theoretical Spectroscopy Facility, Institute of Condensed Matter and Nanosciences, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium ² Department of Quantum Matter Physics, University of Geneva, 1211 Geneva, Switzerland

Geneva, 1211 Geneva, Switzerland ³ U Bremen Excellence Chair, Bremen Center for Computational Materials Science, and MAPEX Center for Materials and Processes, University of Bremen, 28359 Bremen, Germany

 ⁴ Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
 ⁵ Laboratory for Materials Simulations, Paul Scherrer Institut (PSI), 5232 Villigen, Switzerland
 ⁶ Institut de Physique, École Polytechnique Fédérale de

^o Institut de Physique, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

⁷ Photon Science Division, Paul Scherrer Institut (PSI), 5232 Villigen, Switzerland

⁸ WEL Research Institute, 1300 Wavre, Belgium

thomas.vanwaas@uclouvain.be

spectroscopy Angle-resolved photoemission (ARPES) gives access to the non-interacting dispersion and the electron self-energy $\Sigma_n(E)$ for electron energy E through the electron spectral function [1]. According to Matthiessen's rule, $\Sigma_n(E)$ combines the interaction of electrons with phonons $(\Sigma^{\mathrm{ph}}_{n}(E))$, impurities $(\Sigma^{\mathrm{imp}}_{n}(E))$, and other electrons $(\Sigma_n^{\rm el}(E))$ [2]. After this decomposition, the maximumentropy method (MEM) [3] can be used to extract from $\Sigma_n^{\text{ph}}(E)$ the Eliashberg function $\alpha^2 F_n(\omega)$ [4], which quantifies the magnitude of electron-phonon coupling for given phonon frequencies ω . However, quantification of the parameters describing $\Sigma_n^{\rm el}(E)$, $\Sigma_n^{imp}(E)$, and the non-interacting dispersion is obstructed by the energy resolution, and has so far relied on a time-consuming visual inspection of the fit quality [5]. We obtain a probabilistic description for these parameters by extending the MEM with Bayesian inference, resulting in an automated and objective quantification, and we release it as the xARPES code (https://xarpes.github.io). We apply our approach to two high-quality band maps, finding signatures of longitudinal-optical phonons in the 2dimensional electron liquid on the surface of TiO₂terminated SrTiO₃ as well as rich and reproducible features in two separately extracted Eliashberg functions in Li-doped graphene. Finally, we outline a fitting approach that accounts for the energy resolution through bypassing the self-energy extraction step.

References

- [1] [1] H. Zhang, T. Pincelli, C. Jozwiak, T. Kondo, R. Ernstorfer, T. Sato, and S. Zhou, Nat. Rev. Methods Primers 2, 54 (2022).
- [2] [2] K. Wandelt, ed., Surface and Interface Science (Wiley-VCH, Weinheim, 2012).
- [3] [3] J. Skilling, ed., Maximum Entropy and Bayesian Methods: Cambridge, England, 1988 (Springer Netherlands, Dordrecht, 1989).
- [4] G. Grimvall, The Electron–Phonon Interaction in Metals (North-Holland Publishing Company, Amsterdam, 1981).
- [5] X. J. Zhou, J. Shi, T. Yoshida, T. Cuk, W. L. Yang, V. Brouet, J. Nakamura, N. Mannella, S. Komiya, Y. Ando, F. Zhou, W. X. Ti, J. W. Xiong, Z. X. Zhao, T. Sasagawa, T. Kakeshita, H. Eisaki, S. Uchida, A. Fujimori, Z. Zhang, E. W. Plummer, R. B. Laughlin, Z. Hussain, and Z.-X. Shen, Phys. Rev. Lett. **95**, 117001 (2005).



Figure 1. Flow chart for the quantification of many-body properties for an ARPES band map acquired with angle-resolved photoemission spectroscopy. After the Fermiedge and momentum-distribution-curve fits, Bayesian inference is used in the optimization loop to quantify all of the involved model parameters.

Theoretical and experimental HARPES study of Weyl-semimetal TaAs: The application of machinelearning

Trung-Phuc Vo^{1,2}, Arian Arab³, Aki Pulkkinen¹, Laurent Nicolai¹, Sunil Wilfred D'Souza¹, Tien-Lin Lee⁴, Nitesh Kumar⁵, Claudia Felser⁵, Alexander Gray^{3,*}, Jan Minar^{1,*} ¹New Technologies-Research Centre, University of West Bohemia, 30100 Pilsen, Czech Republic. ²Institute of Physics, Czech Academy of Sciences, Cukrovarnicka 10, 162 00 Praha 6, Czech Republic. ³Department of Physics, Temple University, Philadelphia, Pennsylvania 19122, USA. ⁴Diamond Light Source Ltd., Didcot, Oxfordshire OX11 0DE, United Kingdom. ⁵Max Planck Institute for Chemical Physics of Solids, Nothnitzer Str. 40, 01187 Dresden, Germany.

votrung@ntc.zcu.cz

Abstract

A Weyl semimetal is a new matter state possessing Weyl fermions near the Fermi level with several unique physical properties and it is confirmed by the existence of Fermi arc surface states [1]. In this work we study tantalum arsenide (TaAs) which is a prototypical Weyl semimetal compound. The electronic structure properties have been studied by soft and hard X-ray angle-resolved photoemission spectroscopy (ARPES) at energies of 440 eV and 2150 eV, respectively. For the first time, TaAs is experimentally investigated by the bulk sensitive photoemission in the hard X ray regime. In order to interpret experimental data we performed one-step model of photoemission calculations which include [2-4] all matrix elements and final state effects. Due to the strong photon momentum effects and uncertainty in the tilt of experimental geometry we used a so-called machine learning algorithm combined with a free-electron final state model to find best possible experimental parameters. Our findings re-emphasize the overwhelming accuracy of hard X-ray ARPES (HARPES) compared to the traditional ultraviolet and soft X-ray one in case of bulk electronic structure, motivating further material discoveries.

- [1] S. Xu et al., Science 349, (2015) 6248.
- [2] H. Ebert et al., Rep. Prog. Phys. 74, (2011) 96501.
- [3] H Ebert et al. The munich spr-kkr package, version 8.5, (2022).
- [4] J. Braun et al., Phys. Rep. 740, (2018) 1.

Photoemission Final States Beyond the Free-Electron Approximation

Vladimir N. Strocov

Swiss Light Source, Paul Scherrer Institute, 5232 Villigen-PSI, Switzerland

vladimir.strocov@psi.ch

The k-space resolution angle-resolved in photoemission spectroscopy (ARPES) relies on the conservation of the full three-dimensional (3D) electron momentum **k** during the photoexcitation event. During the photoelectron's escape to vacuum only the in-plane momentum $\mathbf{k}_{//}$ is conserved, however, whereas the out-of-plane one k_{\perp} is distorted as the photoelectron crosses the surface. The original k_{\perp} can only be recovered if the out-ofplane dispersion $E(k_{\perp})$ of the ARPES final states back in the bulk is known. Furthermore, the finite mean free path of photoelectrons results in intrinsic broadening of their k_{\perp} that can shift ARPES peaks from their positions dictated by strict momentum conservation.

There are two common approaches to describe the ARPES final states. The simplest approach is to approximate them by free-electron-like dispersions. Although this approach may reasonably describe overall behavior of the ARPES spectra, finer details indicate deviations of the final states from the free-electron-like model. The second approach uses the theoretical bandstructure of the final states, described as scattering states [1,2]. However, this approach typically produces a dense multitude of bands, whose connection to the ARPES spectra is obscure, especially at high excitation energies [3,4].

Within one-step theory, the photoemission final states are time-reversed LEED states. At excitation energies below ~50 eV typical of VUV-range ARPES, this connection enables direct access to the states using Very-Low-Energy final Electron Diffraction (VLEED). Specifically, the energies of the VLEED spectral structures reflect critical points in the final-state $E(k_1)$ (such as those situated at the symmetry lines of the Brillouin zone) while their energy broadening reflects the corresponding lifetimes [3,4]. The VLEED data analysis should engage reference calculations of the final-state bandstructure to identify, in a multitude of all bands available for given photoelectron energy and \mathbf{k}_{ll} , the few that effectively couple to the outgoing photoelectron plane wave in vacuum and thereby effective in the photoemission process.

Extensive VLEED experimental data show that complexity of the final states can go far beyond the naive free-electron-like picture. Not only can these states deviate dramatically from parabolic dispersions, but they can also exhibit a multiband composition where the final state for given photoelectron energy and $\mathbf{k}_{//}$ includes several Bloch waves with different k_{\perp} values. Such phenomena

have been identified for a wide range of materials, from metals (Cu, Ni, Ag) to van-der-Waals materials (graphite, VSe₂, TiS₂, TiTe₂, Bi2212, etc) [6]. Another key finding is that self-energy can vary in a highly non-trivial way as a function of **k** and the orbital composition of the final-states [5]. VLEEDderived final states with controlled intrinsic k_{\perp} broadening enable reliable determination of 3D bandstructures from ARPES data for materials with complex non-free-electron final states.

At high excitation energies in the soft- and hard-Xray range, the photoelectron energy far exceeds the crystal potential modulations. One might expect that the final states there become truly free-electron-like. Surprisingly, recent soft-X-ray ARPES experiments in the energy range around 1 keV [4] have revealed distinct multiband final states in GaN and even in Ag, where the crystal potential is well screened (Fig. 1). Similar effects have also been observed in vander-Waals materials and even in AI, the classical free-electron metal. Qualitatively, these effects can be understood through the momentum-transfer dependence of the crystal pseudopotential [4]. A key benefit of high-energy ARPES in exploring 3D bandstructures is the sharp intrinsic k_{\perp} broadening due to the increasing photoelectron mean free path.

References

- H. Ebert, D. Ködderitzsch & J. Minár. Rep. Prog. Phys. **74**, 096501 (2011).
- [2] E. E. Krasovskii & W. Schattke. Phys. Rev. Lett. 93, 027601 (2004).
- [3] V. N. Strocov *et al.* Phys. Rev. B **74** (2006) 195125
- [4] V. N. Strocov *et al.* Nature Comm. **14** (2023) 4827
- [5] V. N. Strocov *et al.* Phys. Rev. B **66** (2002) 195104



Figure 1. Normal-emission ARPES data from Ag(100), theoretical with free-electron-like final states (*left*) and experimental (*right*). Splitting and broadening of the ARPES dispersions (most clear in $E_{\rm F}$ -MDC) manifests multiband final states.

Exploring the Limits of Machine Learning Algorithms for the Classification of X-ray Absorption Spectra

Federico Zecchi¹, Pieter Glatzel¹, Marius Retegan¹ ¹ESRF, 71 Av. des Martyrs, Grenoble 38000, France

federico.zecchi@esrf.fr

X-ray absorption spectroscopy provides a wealth of information regarding the local structure and electronic properties of materials. However, data analysis is significantly more time-consuming than acquisition and initial data reduction. Decoding the information relies on comparing it with similar compounds for which the spectrum–property mapping is already established, a task that is very often performed by visual inspection.

Machine learning (ML) is revolutionizing many fields with its ability to extract and learn patterns in big data without providing additional prior information other than the data itself. ML models give access to instantaneous predictions of properties and observables, which makes them particularly attractive for performing real-time analysis of the measured data or autonomous experimental acquisitions. Therefore, ML has quickly become an important tool for analyzing X-ray spectroscopy data, speeding up and improving the accuracy with which materials' structural parameters are obtained^{1,2}.

In this contribution, we present the application of the random forests algorithm and convolutional neural networks to identify the coordination environment of iron in a given compound from the corresponding Kedge X-ray absorption spectrum. Since we train our models on theoretical data and then use them to infer properties from measured spectra, we carefully quantify various error sources that can limit prediction quality. These include spectral shift, normalization issues, overabsorption, and the level of Poisson noise. In addition, we present the use of oversampling techniques to tackle class imbalance, a common issue in such datasets, as most materials in nature tend to adopt a small set of specific environments. Finally, coordination we will showcase the implementation of the models for the online analysis acquired at the ID26 beamline of ESRF.

References

- [1] S.B. Torrisi et al., npj Computational Materials, 1 (2020) 109.
- [2] M.R. Carbone et al., Physical Review Materials, 3 (2019) 033604.
- [3] Zecchi et al. in preparation.







Two poster sessions are planned – on 26.05 at 17:00-18:30 including the introductory flash talks and on 27.05 at 17:00-18:30.

- Sourour Ayari: *Investigating the origins of spin polarization in Au (111) : Experiments Vs Theory* (flash talk 17:00-17:05)
- Johannes Scheffler: *Application of machine learning for calculating Raman spectra of layered two dimensional materials* (flash talk 17:05-17:10)
- Alexander Gorfer: *Superresolution for Real-Time TDDFT, enabling spectral predictions of large molecules* (flash talk 17:10-17:15)
- Abdurrahman Adhyatma: *Exploiting Materials Symmetry to Accelerate Bayesian Optimization Structure Search in Computational Atomistic Studies (*flash talk 17:15-17:20)
- Sara Karimi Nejad: *Refining the XAS-3dtm Dataset using TDDFT and Graph Neural Network* (flash talk 17:20-17:25)
- Luqman Hakeem: Advanced Characterization of Lipids from Algal Biomass using Artificial Intelligence (flash talk 17:25-17:30)
- Marco Stecca: *EstraPy: A Flexible and Efficient Tool for X-ray Spectroscopy Data Processing* (flash talk 17:30-17:35)
- Dario Baum: Fast and Simple Complete Basis Set qsGW Calculations
- Gusein Bedirkhanov: *Optimization of Active Learning Strategies for Infra-red Spectra Predictions in Catalysis*
- Laurens de Boer: *Carotenoids in algal biomass: coupling quantum chemistry to AI to solve the conformer puzzle*
- Marc Briant: *Al algorithms for LIBS spectroscopy of nanoparticles*
- H. Elsayed: *Ab-initio study of the electronic and thermoelectric properties of nanoplates and monolayer InSe and GaSe*
- Cezary Janusz: Structure of oxide glasses by x-ray absorption spectroscopy and RMC: the case of amorphous GeO2 under high pressure
- Shanmugasundaram Kamalakannan: *Towards the formation of biofuel precursors by heterogeneous catalysis: The case of aldol condensation on Ru metal.*
- Samuel Longo: Anharmonic vibrational properties of Molybdenum Sulphides from Machine Learning-driven canonical space sampling
- Yashasvi Mehra: *Resonant Photoemission Studies of Transition Metal Sulfides and Selenides*
- Dedy L. Nadeak: *Non-Destructive Screening of Deoxynivalenol in Wheat Kernel Using Near Infrared Spectroscopy (NIRS) and Machine Learning*
- Raphaël Salazar: *Cassiopy: a python suite for ARPES analysis*
- Sarath Sasi: Disorder effects in the Band structure of Transition Metal Dichalcogenide alloys - AxB1- xSe2 (A, B= Cr, Mo, W)
- Daria M. Tomecka: *Real, Fake, or Just Useful? Generation and Evaluation of Synthetic Steel Micrographs with Physical Constraints*
- Shin Yasuda: *Simulating X-ray photoelectron diffraction accelerating convergence efficiency*

Investigating the origins of spin – polarization in Au (111) : Experiments Vs Theory

Sourour Ayari¹, Laurent Nicolaï¹, Aki Pulkkinen¹, Ridha Eddhib¹, Saleem Khan¹, Trung Phuc¹, Jan Minar¹, and Mauro Fanciulli^{2,3} ¹University of West Bohemia/New Technologies – Research Centre (NTC), Plzeň, Czech Republic ²CY Cergy Paris University) ³CEA Paris Saclay

sourour@ntc.zcu.cz

Abstract :

This study explores the origins of spin polarization in semi-infinite Au(111), examining whether the observed spin polarization arises primarily from the system's initial states or solely through the Photoemission process. To address this, we will integrate both experimental and theoretical results to provide an understanding of this spin-polarization origins, we calculate the electronic band structure with and without the influence of Mott-Scattering, while isolating additional factors, such as the Rashba effect[1]. On the theoretical side the calculations are performed using the SPRKKR method [2,3] which is based on DFT calculations, which will account for fundamental effects, while the additional one-step

model will account for the photoemission process.

References

- E. E. Krasovskii and E. V. Chulkov Phys. Rev. B 83, 155401 (2011) , [2] H. Ebert, D.
- [2] Ködderitzsch and J. Minár, Rep. on Prog. in Phys. 74, 096501 (2011) , [3] J. Braun et al.,
- [3] Phys. Rev. B 88, 205409 (2013)



Application of machine learning for calculating Raman spectra of layered two dimensional materials

Johannes Scheffler¹, Dorothea Golze¹ ¹Technische Universität Dresden, 01062, Dresden, Germany

johannes.scheffler@tu-dresden.de

Raman spectroscopy is widely used for the characterization layered two-dimensional of materials (2DMs); however, interpreting experimental data without theoretical reference can be challenging. To address this, we employ a machine learning driven molecular dynamics (MLMD)-based approach to calculate Raman spectra. Compared to conventional normal mode analysis, MLMD offers the advantage that anharmonic effects, such as solvent interactions, can be included. In addition, low-frequency modes $(< 100 \text{ cm}^{-1})$ can be described, which can be used as indicators for the layer thickness and stacking order in 2DMs. To perform MLMD-based Raman calculations, we train the underlying ML model on relevant reference quantities, namely density functional theory (DFT) energies and forces, as well as density functional perturbation theory (DFPT) polarizabilities.

In layered 2DM, layer-layer interactions are dominated by long-range van der Waals (vdW) forces. Accurately describing these interactions, as emphasized in Ref. [1], is essential for stable MD trajectories and must be accurately represented by both the selected vdW and the ML model. In our current work, we investigate two ML strategies for MLMD simulations of layered hexagonal boron nitride (hBN). First, we employ Gaussian Approximation Potentials (GAP) [2], which separately learn PBE energies, forces, and vdW model parameters. Second, we apply the equivariant message passing neural network 'PaiNN' [3] which directly learns total energies and forces.

The preliminary results are promising; we find that both our GAP and PaiNN models successfully learn energies and forces for the hBN system, showing good agreement between peak positions in the MLMD-based power spectra and an ab initio MDbased reference.

- [1] E. Bas, K. Alvarez, A. Schneemann, T. Heine, and D. Golze, Journal of Chemical Theory and Computation, 20, (2024), 9547-9561.
- [2] A. Bartok, M. Payne, R. Kondor, and G. Csanyi. Physical review letters, 104, (2010), 136403.
- [3] K. Schütt, O. Unke, and M. Gastegger. International Conference on Machine Learning, 38, (2021), 9377-9388.

Superresolution for Real-Time TDDFT, enabling spectral predictions of large molecules

Alexander Gorfer¹, Karsten Reuter¹, Matthias Kick¹ ¹Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, 14195 Berlin, Germany

gorfer@fhi-berlin.mpg.de

Real-time time-dependent density functional theory (RT-TDDFT) directly simulates the photoresponse observed in spectroscopy experiments, but for large molecules its computational cost restricts us to short simulation times which limits achievable spectral resolution. A recent superresolution method [1] addressed this limitation by augmenting RT-TDDFT dynamics with approximate results from linear response TDDFT. This hybrid approach reduced the required number of RT-TDDFT timesteps by up to a factor of 20 compared to standard Fourier analysis. However, this drastic improvement remains somewhat insufficient for hybrid DFT calculations of very large systems, such as nanosized quantum dots or plasmonic nanoparticles, which exhibit strong size-dependent optical properties, requiring RT-TDDFT with tens of thousands of electrons [2].

To address this, we propose further enhancements to the superresolution methodology. These include: (1) incorporating higher-order response information beyond the dipole, (2) optimizing beam and signal parameters, (3) dynamically adjusting timesteps, (4) implementing noise reduction techniques and (5) integrating machine learning algorithms for spectral reconstruction. These advancements aim to drastically reduce the number of required RT-TDDFT timesteps, enabling accurate and efficient spectral prediction for large molecules. Furthermore, we will also discuss broadening the scope of the existing framework to extract additional observables such as electronic couplings.

- M. Kick, A. Ezra, A. Beiersdorfer and T. Van Vorhoois, Super-resolution techniques to simulate electronic spectra of large molecular systems, Nature Com., 1 (2024) 8001.
- [2] J. Jakowski, W. Lu, E. Briggs, D. Lingerfelt, B.G. Sumpter, P. Ganesh and J. Bernholc, Simulation of 24,000 Electron Dynamics: Real-Time Time-Dependent Density Functional Theory (TDDFT) with the Real-Space Multigrids (RMG), JCTC, 3 (2025) 1322-1399

Exploiting Materials Symmetry to Accelerate Bayesian Optimization Structure Search in Computational Atomistic Studies

Abdurrahman Adhyatma, Mandira Das, Milica Todorović*

Materials Informatics Laboratory, Department of Mechanical and Materials Engineering, University of Turku 20014, Finland

milica.todorovic@utu.fi

The interface between organic and inorganic materials plays a crucial role in many technologically relevant materials, where comprehensive understanding of the adsorption phenomena is essential. Core electron spectroscopy methods are powerful tools for probing the chemistry of these interfaces. However, it is challenging to deduce the interfacial structure from the resulting spectra. Computational methods, such as density functional theory calculations, can complement spectral data by searching for energetically favorable adsorption geometries to obtain atomic-scale insights. The Bayesian Optimization Structure Search (BOSS)[1] algorithm accelerates such computational studies by modeling the adsorption energy landscape in an active learning framework. Its performance can be further enhanced by exploiting the inherent symmetries of inorganic substrates and organic adsorbates. In this work, we take into account substrate and adsorbate symmetries to identify analogous adsorption configurations in each sampling iteration. Using the spglib^[2] Python package, we detect crystalline surface symmetry to determine equivalent adsorbate positions and orientations. Additionally, the rotational symmetry of the adsorbate molecule-when present-is leveraged to identify further corresponding orientations. The symmetry-aware approach facilitated a more effective exploration of adsorption geometries by improving both the efficiency and accuracy of the search. Furthermore, the approach has the potential to be generalized towards any symmetry-rich structure search problem.

- [1] Todorović, M., Gutmann, M.U., Corander, J. et al. Bayesian inference of atomistic structure in functional materials. npj Comput Mater 5, 35 (2019).
- [2] Togo, A., Shinohara, K., Tanaka, I. Spglib: a software library for crystal symmetry search. Science and Technology of Advanced Materials: Methods, 4 (2024)

Refining the XAS-3dtm Dataset using TDDFT and Graph Neural Network

Sara Karimi Nejad¹, Dacong Wei¹, Hendrik Junkawitsch^{2,3}, Annika Bande^{1,2}

¹Leibniz University Hannover, Institute of Inorganic Chemistry, Callinstr. 9, 30167 Hannover, Germany ²Helmholtz-Zentrum Berlin für Materialien und Energie GmbH, Hahn-Meitner-Platz 1, 10409 Berlin, Germany ³Humboldt-Universität zu Berlin,Unter den Linden 6, 10117 Berlin, Germany

sara.karimi.nejad@aci.uni-hannover.de

The remarkable structural and electronic diversity of transition metal complexes (TMCs) poses a persistent challenge for accurate spectral prediction, limiting the effectiveness of data-driven models. The XAS-3dtm dataset [1] provides K-edge X-ray absorption spectra for first-row TMCs (Ti - Zn), simulated using FDMNES and multiple scattering theory. While this dataset is valuable for machine learning, it lacks orbital-level information necessary for detailed spectral interpretation. The XAS-3dtm dataset complements existing comprehensive datasets, such as tmQM. designed specifically for quantum mechanical modeling of transition metal complexes [2].

In this work, we aim to enrich XAS-3dtm by incorporating orbital-resolved features obtained from linear-response time-dependent density functional theory (TDDFT) calculations, as similarly done in our previous work [3]. We are currently focusing on a subset of Ni-containing complexes, for which TDDFT spectra will serve as the basis for a transfer-learning framework, allowing us to extend the dataset by orbital-level predictions.

For the spectra prediction we are training graphbased neural networks including GCN, GATv2, and Graph SAGE on molecular structure data. Preliminary, results show that model performance improves with more accurate graph representations (e.g., including ligand identity, conjugation, topology) This ongoing work aims to bridge the gap between low-cost simulations and orbital-resolved models.

References

[1] C. D. Rankine, T. J. Penfold, J.Chem.Phys. 156 (2022) 164102.

[2] D. Balcells, B. B. Skjelstad, J.Chem.Inf.Model. 60 (2020)6135-6146.

[3] A. Kotobi, L. Schwob, G. B. Vonbun-Feldbauer, M. Rossi, P. Gasparotto, C. Feiler, G. Berden, J. Oomens, B. Oostenrijk, D. Scuderi, S. Bari, R. H. Meißner, Commun. Chem. 6 (2023) 46.



Advanced Characterization of Lipids from Algal Biomass using Artificial Intelligence

Luqman Hakeem¹, Hela Friha^{2,3}, Majdi Hochlaf³, Tsveta Miteva^{1,*}

¹ Sorbonne Universite, CNRS, Laboratoire de Chimie Physique Matiere et Rayonnement, UMR 7614, F-75005 Paris, France

² Institute of Polytechnic Science and Aeronautics (IPSA), Paris, France

³ Université Gustave Eiffel, COSYS/IMSE, 5 Bd

Descartes, 77454 Champs sur Marne, France

*tsveta.miteva@sorbonne-universite.fr

Abstract

Reducing greenhouse gas emissions, particularly CO₂, is critical, as it accounts for over 75% of the anthropogenic greenhouse effect¹. Among the various available mitigation strategies, capturing CO₂ from the atmosphere and converting it into value-added products (VAPs) is a promising approach. Microalgae outperform terrestrial plants in CO₂ bio-fixation by assimilating CO₂ under light exposure and converting it into algal biomass. This biomass is rich in lipids, sugars, proteins, and pigments², making it a valuable resource for bioenergy³ renewable and other industries⁴. However, the high cost of biofuel production limits its competitiveness with fossil fuels⁵. Optimizina experimental conditions in photobioreactors in order to maximize lipid yield, requires a detailed characterization of algal biomass components. Infrared (IR) and Raman spectroscopies are effective in analyzing biofuel composition⁶, but a major challenge lies in identifying individual organic compounds within complex mixtures. This study focuses on fatty acids, key biofuel precursors, by employing a combined computational chemistry and machine learning approach.

We generated conformers of octanoic acid, a representative fatty acid, using the state-of-the-art Conformer-Rotamer Ensemble Sampling Tool (CREST)⁷. These conformers were then represented using molecular descriptors (e.g., torsion angles, chain lengths) and clustered by geometry via machine learning techniques. Representative Raman spectra were computed for each cluster using quantum chemical methods, such as Density Functional Theory, DFT, and it will be compared with experimental data to identify conformers of this fatty acid present in algal biomass. This methodology will be applied to other biofuel-relevant fatty acids, including decanoic, dodecanoic, myristic, palmitic, and stearic acids. Similar strategy can be extended to other chemicals, paints, etc. The computational approach will help in optimizing the reaction conditions for experimental photobioreactors thus increasing the yield of biofuel.

- (1) https://www.c2es.org/content/internationalemissions/ (accessed 2025-03-31).
- (2) Daneshvar, E.; Wicker, R. J.; Show, P.; Bhatnagar, A, Chem. Eng. J. 2022, 427, 130884.
- (3) Liu, W.; Fu, D.; Pan, T.; Sing, R. P., Water 2021, 13 (22), 3190.
- (4) Enamala, M. K.; Enamala, S.; Chavali, M.; Donepudi, J.; Yadavalli, R.; Kolapalli, B.; Aradhyula, T. V.; Velpuri, J.; Kuppam, C., Renew. Sustain. Energy Rev. 2018, 94, 49-68.
- (5) Rafa, N.; Ahmed, S. F.; Badruddin. I. R.; Mofijur, M.; Kamangar, S., Front. Energy Res. 2021, 9, 749968.
- Chun, H. J.; Waqued, S.; Thapa, H. R.; Han,
 A.; Yakovlev, V. V.; Laane, J.; Devarenne, T.
 P., J. Mol. Struct. 2017, 1129, 216–221.
- Pracht, P.; Grimme, S.; Bannwarth, C.;
 Bohle, F.; Ehlert, S.; Feldmann, G.; Gorges, J.; Müller, M.; Neudecker, T.; Plett, C.;
 Spicher, S.; Steinbach, P.; Wesołowski, P. A.; Zeller, F. Journal of Chemical Physics 2024, 160 (11).



Figure 1. Clustering of octanoic acid conformers (star represents the centre of cluster while cluster representative molecular structure is available at the bottom or right side of cluster)

EstraPy: A Flexible and Efficient Tool for X-ray Spectroscopy Data Processing

Marco Stecca¹, Paolo Centomo¹, Carlo Meneghini² ¹ Dipartimento di Scienze Chimiche, Università degli Studi di Padova, via Marzolo 1, 35131 Padua, Italy ² Dipartimento di Scienze, Università degli Studi Roma Tre, Via della Vasca Navale, 84, 00146 Rome, Italy

marco.stecca.5@phd.unipd.it

X-ray Absorption Spectroscopy (XAS) is a powerful technique for probing the local atomic structure of materials. However, the analysis of XAS data often involves complex and time-consuming processing steps, especially during the extraction of the EXAFS signal from the raw data. EstraPy is a Python program designed to streamline this process by providing a highly adaptable and efficient platform for XAS data analysis. Unlike its predecessor, ESTRA^[1], which is primarily tailored for EXAFS, EstraPy adopts a sequential, instruction-driven architecture, enabling users to execute a series of operations with precise control over each step. This flexibility allows for the analysis of diverse XAS datasets, including both XANES and EXAFS spectra in both fluorescence and transmission modes, read from standard ASCII tables.

A key feature of EstraPy is its enhanced flexibility in data handling, with particular attention to preliminary data treatment procedures, including energy scale (E_0) finding algorithms, data alignment, abscissa (*E* or *k*) binning and interpolation, glitches and spurious point removal. It has been designed to specifically treat large data files as those output from Q-XAFS set-up routinely providing up to thousands of data points. The program supports batch processing of multiple files, with options for spectral averaging or independent analysis, significantly improving throughput.

The script/command-line interface, flexible architecture and high portability are characteristics making EstraPy a valuable tool for researchers seeking to optimize their XAS data analysis workflows.

The current implementation focuses on direct sequential execution of instructions; the integration of scripting capabilities, multivariate analysis and machine learning algorithms for automated data processing and feature extraction are under development. Moreover, the functionality of FitEXA are being integrated into EstraPy, toward a one-step procedure from raw absorption data to EXAFS data refinements, analogous to Rietveld method^[2] or Fitheo module^[3] in GnXAS suite.

EstraPy is accessible on GitHub at the following URL: <u>https://github.com/ramsteak/EstraPy</u>

References

- C. Meneghini, F. Bardelli, S. Mobilio, Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. At. 2012, 285, 153–157.
- [2] H. M. Rietveld, J. of App. Cryst. 1969, 2, 65–71
- [3] A. Filipponi and A. Di Cicco TASK QUARTERLY 4 No 4 (2000), 575–669



Figure 1. Batch processing of XANES spectra during an oxygen treatment of palladium-doped zeolites.



Figure 2. Batch processing of EXAFS spectra of palladium foil measurements



Figure 3. Batch processing of EXAFS spectra of palladium foil measurements

Fast and Simple Complete Basis Set qsGW Calculations

Dario Baum¹, Arno Förster¹, Luuk Visscher¹ ¹Vrije Universiteit Amsterdam, De Boelelaan 1105 1081HV, Amsterdam, Netherlands

d.d.baum@vu.nl

The GW approximation is a widely used method for computing quasiparticle energies of crystalline as well as molecular systems, providing accurate predictions of ionization potentials and electron affinities with higher accuracy than DFT while being more efficient than wavefunction methods. However, its practical applicability is often limited by the slow convergence of computed energies with respect to the basis set size. In this talk, I demonstrate that a recently proposed and simple complete basis set (CBS) limit extrapolation method [1] can be generalized to quasiparticle self-consistent GW (qsGW) calculations. Further, I demonstrate that commonly used basis set extrapolation methods underestimate the CBS limit of GW. Finally, and based on that extrapolation scheme I showcase a large dataset of highly accurate qsGW quasi-particle energies and GW-BSE excitation energies for Machine Learning and benchmarking purposes.

References

[1] Bruneval, Fabien, et al. "Extrapolating unconverged GW energies up to the complete basis set limit with linear regression." *Journal of Chemical Theory and Computation* 16.7 (2020): 4399-4407



Figure 1. Fast and Simple qsGW

Optimization of Active Learning Strategies for Infra-red Spectra Predictions in Catalysis

Gusein Bedirkhanov¹, Nitik Bhatia¹, Ondrej Krejci^{2,3}, Patrick Rinke^{1,2,4,5}

 ¹Department of Physics, Technical University of Munich, Lichtenbergstr. 4, 85748 Garching, Germany
 ²Department of Applied Physics, Aalto University, Otakaari 1, 02150 Espoo, Finland
 ³Department of Mechanical and Materials Engineering, University of Turku, Vesilinnantie 5, 20014 Turku, Finland
 ⁴Atomistic Modelling Center, Munich Data Science

Institute, Technical University of Munich, Walther-Von-

Dyck Str. 10, 85748 Garching, Germany.

⁵Munich Center for Machine Learning (MCML), Munich, Germany.

gusein.bedirkhanov@tum.de

Infrared spectroscopy is an important tool for understanding catalytic processes. By probing the vibrational properties of molecules, it offers detailed insights into molecular structure transformations during catalytic reactions. However, interpreting experimental spectra is challenging due to the strong influence of the species environment. *Ab-initio* molecular dynamics (AIMD), which goes beyond the harmonic approximation of the molecular potential energy surface, offers valuable support for this task by providing accurate theoretical spectra. But this approach is computationally intensive.

Machine-learned interatomic potentials (MLIPs), such as MACE [1], offer a promising alternative to AIMD for simulating molecular systems. Once trained, MLIPs enable accurate and fast IR spectra predictions. However, MLIPs require extensive training data to achieve accurate spectra.

To address this challenge, we have established an active learning (AL) [2] method PALIRS [3] (a Pythonbased active learning code for infrared spectroscopy). In our preliminary work, PALIRS was applied to predict IR spectra of organic molecules comprising C, H, N, and O with carbon count of \leq 2, which are essential in catalytic processes. PALIRS has demonstrated the ability to reduce the required data by a factor of 100 and offered highly accurate IR spectra, with peak positions matching experiment within 20 cm⁻¹.

In this work, PALIRS is extended on more complex organic molecules, comprising C, H, N, and O, with carbon count of \leq 5. The key objective of the work is to compare two MLIP training strategies in AL, aiming to further improve the efficiency of PALIRS. The first approach is the conventional training from-scratch, where a new MLIP is generated from scratch at each AL iteration using the entire accumulated training data. The second strategy is the transfer learning, where at each AL iteration the MLIP is fine-tuned on new data acquired at the current AL iteration.

Our results demonstrate that transfer learning maintains high final MLIP accuracy (< 11 meV for total energies and < 10 meV/Å for forces), while reducing total training time by a factor of ~8 (Fig. 1). This significant reduction in computational cost makes transfer learning a promising approach for MLIP training in AL.

Building on this foundation, the updated PALIRS offers an even more efficient workflow for developing the MLIP capable of predicting IR spectra for a wide range of organic molecules. The approach will advance the field of catalysis by providing an efficient and accurate method for predicting and interpreting IR spectra, facilitating better understanding of catalytic processes and the design of better catalysts.

References

- [1] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, G. Csanyi, Adv. Neural Inf. Process. Syst. , (2022).
- [2] Z. Tang, S. T. Bromley, B. Hammer, J. Chem. Phys., (2023) 224108.
- [3] N. Bhatia, Ondrej Krejci, Patrick Rinke, GitHub (2024), gitlab.com/cest-group/PALIRS



Figure 1. Comparison of Transfer Learning (TL) and From-Scratch (FS) training strategies for MLIP in the AL. (a) Mean absolute error (MAE) of total energies (in meV) across AL iterations, evaluated on a separate test set generated using molecular dynamics simulations. (b) MAE of atomic forces (in meV/Å) at each AL iteration on the same test set. (c) MLIP training time (in hours) across AL iterations and cumulative training time on 8 AMD MI250X GPUs.

Carotenoids in algal biomass: coupling quantum chemistry to Al to solve the conformer puzzle

Laurens de Boer¹, Héla Friha^{1,2}, Roberto Linguerri¹, Tsveta Miteva³, Majdi Hochlaf¹,

¹University Gustave Éiffel, COSYS/IMSE, 5 Bd Descartes 77454, Champs-sur-Marne, France. ²Institut Polytechnique des Sciences Avancées, 63 Bd de Brandebourg Bis 94200, Ivry-sur-Seine, France. ³Laboratoire de Chimie Physique-Matière et Rayonnement (LCPMR), UMR 7614, CNRS, Sorbonne Université, 4 Place Jussieu 75005 Paris, France

laurens.de-boer@univ-eiffel.fr

Biofuels derived from biomass are a viable and sustainable alternative to fossil fuels, but their largescale implementation is hampered by high production costs [1]. The optimization of (photo)bioreactors for biofuel production relies on the precise characterization of algal biomass, particularly its organic constituents [2]. RAMAN spectroscopy is a highly effective technique for this analysis [3]. However, fully characterizing these compounds in biomass is challenging due to the vast number of possible conformers and isomers (>10⁴) present in the mixture. A key challenge lies thus in differentiating the spectral contributions of individual compounds and accurately identifying their conformers within complex mixtures [2].

At present, we treat the case of (α , β , γ)-carotenoids as a case study. First, an experimental investigation was carried out to study these compounds in a culture of *Chlamydomonas reinhardtii* microalgae. Specifically, a micro-Raman spectroscopy analysis was carried out using a Horiba X-plora confocal microscopy Raman spectroscopy system, equipped with 10× and 50× long-working-distance objectives and a charge-coupled device (CCD) detector (see Figure 1). The Raman excitation source is provided by a 638 nm LED laser beam, with a beam power of 6 mW, focused on the sample with a spot size of approximately 1 µm in diameter at 532 nm.

The next step was to conduct a theoretical study combining first-principles techniques with various statistical tools to clarify the contributions of the different molecular species and relative conformers to the measured spectra. First, we predict the structure of the various conformers using the GFN-FF method [4,5,6]. Then, we compute their RAMAN spectra using the Density Functional Theory (DFT). A subsequent Machine Learning (Linear regression) coupled to clustering (k-Means) treatment allows to minimize the redundancy in the generated conformer dataset. Afterwards, by combining the Spectral Angle Mapper (SAM) analysis and Principal Component Analysis (PCA), we were able to successfully identify the spectral contribution and key regions for these conformers (see Figure 1). In fact, the comparison of the computed RAMAN spectra to that of microalgae (*Chlamydomonas reinhardtii*) allowed the identification of the few (α , β , γ)-carotenoids conformers present in such biological media.

In conclusion, our approach offers a viable framework for the rapid analysis of the contribution of individual conformers of (α , β and γ)-carotenoids and more generally of precursors in algal biomass.

References

- [1] C. S. Jones et al., *Current Opinion in Biotechnology*, **2012**, 23, 346–351.
- [2] Miteva, T. et al., Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, **2025**, Under review.
- [3] H. Wu, et al., *Proceedings of the National Academy of Sciences*, **2011**, *108*, 3809–3814.
- [4] Spicher S. et al., *Angewandte Chemie* International Edition, **2020**, 59, 15665–15673
- [5] Ehlert, S., et al. , **2022**, *The Journal of Physical Chemistry A*, 22, 3521-3535.
- [6] P. Pracht, et al., *Phys. Chem. Chem. Phys.*, **2020**, 22, 7169–7192.



Figure 1. A) cluster of α -carotenoid conformers generated by the GFN-FF method. B) overlay of the generated RAMAN spectra of the theoretically calculated α carotenoid conformers, and the experimental RAMAN spectra of *Chlamydomonas reinhardtii* microalgae. C) SAM analysis of the RAMAN spectra obtained by DFT of the theoretically generated α -carotenoid conformers. D) PCA of all generated RAMAN spectra of the theoretically calculated α -carotenoid conformers, and the experimental RAMAN spectra of *Chlamydomonas reinhardtii* microalgae.

Al algorithms for LIBS spectroscopy of nanoparticles

Marc Briant¹, Olivier Sublemontier¹, Mathieu Thévenin²

¹Université Paris-Saclay, CEA, CNRS, NIMBE UMR 3685, 91191, Gif-sur-Yvette, France ²Université Paris-Saclay, CEA, CNRS, SPEC UMR 3680,

91191, Gif-sur-Yvette, France.

marc.briant@cea.fr

Our researches focuses on the custom synthesis and characterisation of various nano-objects. One of our key manufacturing processes is laser pyrolysis ^[1, 2], which uses a high-power CO₂ laser (2 to 3 kW) to synthesize nanoparticles in the gas phase. This process involves the interaction of the laser with a stream of precursors in a controlled atmosphere, forming nanoparticles within the laser pyrolysis flame. The properties of the resulting nanoparticles are determined by the nature and quantities of the reactants injected into the reactor.

Currently, nanoparticles are characterized after synthesis. If their properties are not suitable, the process must be repeated with modified precursor injections, wasting considerable time. To address this, real-time characterization during synthesis is being explored. This involves qualitative analysis (identifying elements) and quantitative analysis (determining elemental composition) to adjust synthesis parameters in real time fly.

Laser-Induced Breakdown Spectroscopy (LIBS)^[3, 4] can be used to identify and quantify many elements at the cost of spectral analysis. However, analysing a large number of spectra in real-time requires advanced data processing. As this is a classification problem, related works show that it is interesting to use approaches based on Artificial Intelligence (AI)^[5].

More specifically, the approach considered in this initially developed for work was infrared spectroscopy^[6]. It uses a Peak Correlation Classifier (PCC) algorithm that includes a Support Vector Machine (SVM) classifier. The PCC algorithm extracts a vector of correlations between an annotated reference spectrum (top spectrum of the top left panel of Figure 1) and an unknown spectrum (bottom spectrum of the top left panel of Figure 1), characterizing the unknown spectrum relative to the reference. Several reference spectra, corresponding to different atomic elements or concentrations, are used to create a database for training the SVM.

Our objective is to adapt the PCC and train the SVM classifier to determine the optimal transformation to separate data based on whether a spectrum contains the atomic or ionic lines of the desired element. Then, in real-time, during the synthesis process, the SVM algorithm will identify if the acquired spectrum contains the desired elements as

represented in Figure 1. Finally, the algorithm will be adapted for quantitative analysis.

References

- [1] Y. Q. He, X. G. Li and M. T. Swihart, Chem. Mater., 17 (2005) 1017–1026.
- [2] O. Sublemontier, H. Kintz, F. Lacour, X. Paquez, V. Maurice, Y. Leconte, D. Porterat, N. Herlin-Boime and C. Reynaud, Kona Powder And Particule Journal, (2011) 236–250.
- [3] F. A. Barreda, C. Nicolas, J. B. Sirven, F. X. Ouf, J. L. Lacour, E. Robert, S. Benkoula, J. Yon, C. Miron and O. Sublemontier, Sci. Rep., 5 (2015) 15696.
- [4] P. Liu, P. J. Ziemann, D. B. Kittelson and P. H. McMurry, Aerosol Sci. Tech., 22 (1995) 293– 313.
- [5] P. Malfrait, J. Bobin and A. de Vismes Ott, Appl. Radiat. Isot., 201 (2023) 111011.
- [6] R. M. French, V. Simic and M. Thevenin, let Signal Process., 14 (2020) 737–744.



Figure 1. Schematic diagram of the adapted algorithm envisaged for this project

Ab-initio study of the electronic and thermoelectric properties of nanoplates and monolayer InSe and GaSe

H. Elsayed^{1,2}, D. Olguín³, and A. Cantarero⁴

¹Physics department, University of Rome Tor Vergata, Rome 00133, Italy.

² Programa de Nanociencias y Nanotecnología, Centro de Investigacion y de Estudios Avanzados del Instituto Politécnico Nacional, Ciudad de Mexico 07360, Mexico
³ Centro de Investigación y de Estudios Avanzados del I.P.N. Unidad Querétaro, Libramiento Norponiente No.
2000, Fracc. Real de Juriquilla, Querétaro, Qro C.P.
76230, Mexico

⁴ Molecular Science Institute, University of Valencia, Valencia 46071, Spain.

Hannan.sadek@uv.es

There is a global crisis due to the vast consumption of non-renewable energy resources and the continuous growth of the world economy and industrial development. At the same time, our climate is affected by the burning of fossil fuels due to the emissions of carbon dioxide that have passed the record level in the last decade, the so-called Global Warming. To solve these problems, there is an urgent need to find a new effective, mechanically and thermally stable, and non-toxic thermoelectric material that can be used in future in thermoelectric devices for power generation.

One of the most influential concepts in thermodynamics is the dimensional reduction as it can simultaneously improve the power factor and reduce the lattice thermal conductivity. This started with the approach, proposed by Hicks and Dresselhaus [1], of the possibility to increase the figure of merit Z of certain materials by preparing them in quantum-well superlattice structures. Their calculations showed that layering has the potential to increase significantly the figure of merit of a highly anisotropic material such as Bi2Te3 that has ascertained their early assumption. This stimulated subsequent work on nanostructured thermoelectrics to increase the figure of merit [2,3] especially, in the layered compounds [4,5].

A comprehensive study of the thermoelectric coefficients of the ε - polytype of InSe and GaSe is reported. Three structures have been studied in each case; the bulk, nanoplates (7 quadruple layer), and monolayer (one quadruple layer). The calculations have been done within the framework of the density functional theory (DFT) [6,7], where the electronic properties were calculated using the full potential linearized augmented plane wave method as implemented in the Wien2k code [8]. Based on

these calculations, the evolution of the transport coefficient as a function of the chemical potential is evaluated with the use of the Boltzmann transport theory as implemented in the BoltzTraP code [9].

References

[1] L. D. Hicks and M. S. Dresselhaus, Phys. Rev. B 47 (1993), pp. 12727-12731.

[2] R. D. Delgado W. Liang E. C. Garnett M.
Najarian A. Majumdar A. I. Hochbaum R. Chen and P. Yang, Nature 451 (2008), 163–167.
[3] G. Chen, Phys. Rev. B 57 (1998), p. 14958.

[4] J. Kang2W. Fan Ch. Ko A. V. Luce K. X.Wang J. Suh K. D. Patel. Appl. Phys. Lett. 104 (2014), p. 012101.

[5] G. Ding G. Li and G. Gao, Journal of Physics: Condensed Matter 29 (2017), p. 015001.

[6] P. Hohenberg and W. Kohn. Phys. Rev.136 (1964), B864–B871.

[7] W. Kohn and L. J. Sham. Phys. Rev.140 (1965), A1133–A1138.

[8] P.Blaha, K.Schwarz, G.K.H.Madsen,

D.Kvasnicka, J.Luitz, Vienna University of

Technology, Austria (2001).

[9] G.K.H. Madsen, D.J.Singh, Comput.Phys. Commun.175 (2006) 67.

Structure of oxide glasses by x-ray absorption spectroscopy and RMC: the case of amorphous GeO₂ under high pressure

Cezary Janusz¹, Andrea Di Cicco¹, Fabio Iesari², Nodoka Hara¹, Marco Minicucci¹, Giuliana Aquilanti³, Angela Trapananti¹ ¹University of Camerino, School of Science and Technology, Physics Division, via Madonna Delle Carceri 9, CAP: 62032 (MC), Camerino, Italy ²Aichi Synchrotron Radiation Center, Seto, 489-0965, Japan ³Elettra Sincrotrone, Trieste, 34149, Italy

³Elettra Sincrotrone, Trieste, 34149, Italy cezary.janusz@unicam.it

Reverse Monte Carlo is a method originally proposed by Pusztai and McGreevy [1,2] for building atomic models consistent with one or more sets of structure sensitive experimental data from either xray or neutron scattering.

By means of randomized atomic movements and a selection procedure based on the standard Metropolis algorithm, configurations that minimize a residual function χ^2 between experimental signals and those calculated from the atomic model are generated. The final result is a structural model that best matches the whole experimental dataset.

The approach is particularly useful for highlydisordered materials, e.g. glasses or liquids, for which pair distribution functions – directly calculated from the experimental S(q) by Fourier transform – are insufficient to describe the structure. Additional descriptors (e.g. bond-angle distribution, coordination, analysis of atom chains) can be easily obtained from an atomic model, providing a more comprehensive structural characterization.

The RMC approach was also integrated into the RMC-GnXAS code [3] (a component package developed of GnXAS at the University of Camerino for X-ray absorption data analysis), which serves the purpose of obtaining an atomic model consistent with an experimental EXAFS signal. Throughout two decades of its existence, the code, originally written for the simplest monoatomic systems, has been constantly adapted to accommodate more complex applications.

In this contribution, total static, neutron and X-ray scattering structure factor refinement was introduced as an alternative to previously available pair distribution function which constrains the long-range order. That approach eliminates the necessity for deriving partial pair distribution functions (only simulations accessible via or expensive experimental techniques) for multiatomic systems and expands the possibilities of structure analysis under extreme conditions (i.e. high pressure or Additionally, temperature). constraint on а coordination number was introduced as another component within the residual function to improve the accuracy of structural representation for materials where that factor is already wellestablished from other studies.

New features within the RMC-GnXAS code were validated through the GeO_2 glass analysis under pressures of 0.0 and 4.3GPa, using experimental EXAFS spectra collected at Elettra synchrotron radiation facility in Trieste, neutron scattering structure factor data from the literature [4,5] and local coordination numbers to constrain the RMC refinement process. The resulting structures are compared with the available literature references [4,5,6,7,8].

- [1] R. L. McGreevy, L. Pusztai, Molecular Simulation 1:6 (1988) 359-367
- [2] R. L. McGreevy, J. Phys.: Condens. Matter 13 R877 (2001)
- [3] F. lesari et al., "GNXAS: Advances in the Suite of Programs for Multiple-Scattering Analysis of X-ray Absorption Data", Springer International Publishing AG (2018)
- [4] Salmon et al., J. Phys.: Condens. Matter 19 415110 (2007)
- [5] Salmon et al., J. Phys.: Condens. Matter 24 415102 (2012)
- [6] Peralta et al., J. Phys.: Condens. Matter 20 145215 (2008)
- [7] Anh et al., Computational Materials Science 177 109597 (2020)
- [8] Hong et al., J. Phys.: Condens. Matter, 35 164001 (2023)

Towards the formation of biofuel precursors by heterogeneous catalysis: The case of aldol condensation on Ru metal.

Shanmugasundaram Kamalakannan^{1,*}, Roberto Linguerri¹, Majdi Hochlaf¹ ¹COSYS/IMSE, Université Gustave Eiffel, 5 Bd

Descartes, 77454 Champs sur Marne, France

k.shanmugasundram@univ-eiffel.fr

Abstract

The challenges of depleting fossil fuel reserves and their impacts on the environment have inspired substantial research into the efficient production of renewable biomass fuels and chemicals over the past decade. An emerging strategy for biomass production is the formation of C-C bond by heterogeneous catalysis of aldol condensation of bio-based target molecules such as acetone, furfural 5-hydroxymethylfurfural and (HMF).[1] This approach can produce larger organic compounds such as the C13 furanic adduct 1,5-di-2-furanyl-1,4pentadien-3-one (referred to as: FAF), which can be transformed to high-quality diesel fuels by further hydrogenation.[2] However, the molecular level understanding of the reaction mechanisms at the catalysts for the selective production of diesel is still scarce.[3] For remediation, we treat the reaction of formation of FAF through aldol condensation on the most stable phase (0001) of Ru-based catalysts, thus understanding the targeting reaction mechanism for the large hydrocarbon production.[4] Gas phase computations showed that there are 30 different possible conformations of FAF molecules. However, when these molecules were impregnated on to the Ru(0001) surface, they are found to stabilize through FAF O-Ru (Ru(0001)), and FAF C-Ru (Ru(0001)) bonding with very few varietry of conformations (i.e., parallel, tilted, parallel+vertical and diagonal conformers) in contrast to the gas phase data. We show that this is due to the chemical adsorption of these FAF conformers on the Ru active sites of the Ru(0001) slab and the subsequent electron and charge transfer induced on adsorbed FAFs.

The calculated lowest interaction distances between the FAF O-Ru (Ru(0001)) and FAF C-Ru (Ru(0001)) vary from 2.01 Å to 2.54 Å and 2.10 Å to 2.22 Å, respectively. In addition, the total energies of these 30 different FAF at the Ru-slab clearly show that several FAF confromers exhibit similar patterns and also strongly stabilized via the above-mentioned chemical bonding with the metal slab. This screening allowed us to identify the most stable configuration of the FAF@Ru slab among the 30 possible conformers (identified from the gas phase). This screening not only reduces the computational costs but also significantly drop the activation energies, i.e. energy requirements, for the production of longer-chain hydrocarbons.

Besides, the Neuged elastic-band calculations will be carried out to examine the in-depth reaction

mechanism. In particular, we are looking to determine the kinetically determining step. Also, machine learning methods will be used to find the activity relationship between the major intermediate and the Ru catalyst by using the accuracy plot of the regularized Model. Obviously, this goes through the identification of the corresponding transition states and reaction barriers on Ru metal and products. Our work will suggest the most prominent catalyst for the efficient bio-fuel production with less energy barrier and cost.

References

- [1] X. Kong, X. Wei, L. Li, Z. Fang, H. Lei, Catalysis Communications, 149, (2021), 106207.
- [2] V. Korolova, O. Kikhtyanin, E. Grechman, V. Russo, J. Warna, D. Y. Murzin. D. Kubicka, Catalysis Today, 423, (2023), 114272.
- [3] E. F. William, B. K. Tomas, B. Jonas, Journal of Catalysis, 346, (2017), 78-91.
- [4] M. M. Kauppinen, E. N. Słap, J. L. G. Escobedo, R. L. Puurunen, K. Honkala, 751, (2025), 122624.



Figure 1. Biomass production at the Ru metal surface.

Anharmonic vibrational properties of Molybdenum Sulphides from Machine Learning-driven canonical space sampling

Samuel Longo¹, Matthieu J. Verstraete¹ ¹Université de Liège, Physiscs Dep., Nanomat group, Allée du six Août 19, 4000 Liège

samuel.longo@uliege.be

Molybdenum Sulphides (MSs) represent one of the most promising non-toxic and inexpensive alternatives to the Pt-based catalysts for the Hydrogen Evolution Reaction (HER). The design of such catalysts is still mainly limited by the difficulty in determining their structure [1]. Vibrational spectroscopy is a powerful tool to unveil their local structural features and can assist in the study of the catalytic mechanisms. We investigate the temperature-dependent vibrational properties of MoS₂, such as phonons and Raman activity, by combining state-of-the-art Machine Learning-Interatomic Potentials (MLIPs) with Molecular-Dynamics (MD) simulations. More specifically, we use the Temperature Dependent Effective Potential [2] (TDEP) to extract the anharmonic interatomic force constants, the phonon lifetimes and the Raman activity from canonical sampling at a given temperature. To attain an accurate sampling, it is performed via Machine Learning-Molecular Dynamics (MLMD) simulations, where the MLIP (we use the Moment Tensor Potential, MTP [3]) replaces the computationally demanding ab initio calculation of the atomic forces. The MLIP itself is trained on a MoS₂ DFT dataset generated via a self-consistent iterative scheme of phase space exploration called Machine Learning Assisted Canonical Sampling (MLACS) [4]. Accordingly, the training datapoints are obtained as snapshots of MLMD, by retraining the MLIP at each augmentation of the dataset, to improve its accuracy, until the error converges. The accuracy of the training set (ab initio) and the enforcement of a temperature during the MLMD runs, ensure that the resulting interatomic force constants account for anharmonicity effects. The Raman spectra we obtained and their temperature dependence are in good agreement with experimental values, confirming the effectiveness of both the MTP model and the TDEP method.

References

- [1] Phong D. Tran et al., Nature Materials, 6 (2016) 640-646.
- [2] Olle Hellman et al., Physical Review B, 10 (2013) 104111.
- [3] Ivan S. Novikov et al., Machine Learning: Science and Technology, 2 (2020).
- [4] Aloïs Castellano et al., Physical Review B, 16 (2020) L161110.



Figure 1. Raman spectra of MoS₂ vs. Temperature



Figure 2. Tdep model truncated to the 3rd order and scheme of optimization

Resonant Photoemission Studies of Transition Metal Sulfides and Selenides

Yashasvi Mehra^{1,2,3}, Samuel Beaulieu⁴, Mauro Faniculli^{1,2}, Aki I. O. Pulkkinen³,Olivier Heckmann^{1,2}, Karol Hricovini^{1,2}, Marchin Rosmos⁵, Natalia Olszowska⁵, Tomasz Sobol⁵, Edyta Beyer⁵, Jan Minar³, Maria Christine Richter^{1,2}

1Université Paris-Saclay, CEA, LIDYL,Gif-sur-Yvette, France 2CY Cergy Paris Université, CEA, LIDYL, Gif-sur-Yvette, France 3University of West Bohemia, NTC, Pilsen, Czech Republic 4University of Bordeaux, CNRS, CEA, CELIA, UMR 5107, Talence, France 5SOLARIS National Synchrotron Radiation Centre, Jagiellonian University, Krakow, Poland

E-mail:yashasvi@ntc.zcu.cz

By performing resonant ARPES measurements and SPR-KKR photoemission calculations on Transition Metal Selenide, Sulfide we study the interplay between different decay mechanisms in resonant conditions, radiation-less Raman Auger and Classical Auger emissions. Through a method proposed by Cini[1] and Sawatzky[2], [3] we can determine the on-site Coulomb interaction per element in some cases. In this method, one compares the energy of the correlation satellite associated with the two-valence-hole (VV) Auger final state in resonant photoemission with the twovalence-hole energies without correlations obtained from a self-convolution of the single-hole states obtained from a non-resonant photoemission spectrum. The energy separation between the main peaks of the resonantly enhanced spectrum and the two-hole spectrum without correlations gives a measure of the Coulomb energy [4]. On the theoretical front the calculations are performed using the SPR-KKR method, which is based on one-step model, that incorporates the effect of all matrix elements which accounts for the photoemission process. In Figure 1 I have showed the experimental results of NbS₂, for 2p S excitations, to measure U_{pp} for the case. Furthermore, we analyze calculated ARPES, XAS, element and orbital resolved band structure and DOS underlining agreement with results and helping experimental with its interpretation.

References

- [1] Cini M., Phys. Rev. B 17, 2788 (1978).
- [2] Sawatzky G. A., Phys. Rev. Lett. 39, 504 (1977).
- [3] Sawatzky G.A., et al Phys.Rev. B. 21,1790 (1980).

[4] Chainani A., et al., Phys. Rev. Lett. 119, 057001 (2017).



Figure 1. Element-specific X-ray absorption spectroscopy (XAS) and resonant photoemission spectroscopy (ResPES) for the NbS₂. The figure shows XAS $L_{2,3}$ -edge spectra (right side of panel) and corresponding ResPES energy distribution curves (EDCs) for S. The EDCs are plotted on a binding energy (EB) scale, with the photon energy as the vertical axis.

Non-Destructive Screening of Deoxynivalenol in Wheat Kernel Using Near Infrared Spectroscopy (NIRS) and Machine Learning

Dedy L. Nadeak¹, Rudolf Krska^{1,2}, Stephan Freitag¹ ¹Department of Agricultural Sciences, Institute of Bioanalytics and Agro-Metabolomics, BOKU University, Konrad-Lorenz-Strasse 20, Tulln an der Donau, Austria ²Institute for Global Food Security, School of Biological Sciences, Queens University Belfast, University Road, Belfast, BT7 1NN, Northern Ireland, United Kingdom

dedy.nadeak@boku.ac.at

Wheat, one of the most important staple foods globally, faces significant challenges to crop yield quality and food safety due to contamination by Fusarium fungi. Infection of grains with Fusarium not only alters the composition of major biochemical components, such as carbohydrates, proteins, and fats, but also leads to the accumulation of the mycotoxin Deoxynivalenol (DON) [1]. As a nondestructive and rapid analytical method, Near Infrared Spectroscopy (NIRS) has been employed to screen for the presence of DON in various crops by detecting changes in these key biocomponents [2]. In this study, we aimed to utilize NIRS to classify DON contamination in both unbalanced and balanced wheat kernel samples, according to the maximum permitted limit set by the European Commission (EC). To achieve this, machine learning models, such as partial least square discriminant analysis (PLS-DA), random forests, support vector machines, and extreme gradient boost, were developed and combined with variable selection techniques to enhance their classification performance and interpretability. The outcomes of this approach, including model performance and variable relevance, will be presented.

- [1] Boyacioğlu, D., and Hettiarachchy, N. S., Journal of Cereal Science, 21 (1995), 57-62.
- [2] Freitag, S., *et al.*, Comprehensive Reviews in Food Science and Food Safety, 21 (2022), 5199-5224.

Cassiopy : a python suite for ARPES analysis

Raphaël Salazar¹

¹New Technologies Research Centre, University of West Bohemia, Pilsen, Czech Republic ¹Cassiopee Beamline, Synchrotron SOLEIL, L'Orme des Meurisiers, Saint-Aubain, France

rsalazar@ntc.zcu.cz

Abstract

While angle resolved photoemission spectroscopy (ARPES) has established itself as the go-to technique for electronic structure determination and beyond, there is not to this day a consensus for the data treatment pipeline when it comes to the analysis beyond the simplest axis conversion. With the variety of detectors that exist in the market and the fact that each experimental station has its own specificities, there is also no standard also in the data format. This usually means that each and every photoemission experimental station has its own conventions for the storing and exploitation of the data. This reflects a situation in which the experimental setups are increasingly more complex setups: polarization dependent momentum microscopy [1], spin-resolved ARPES [2], timeresolved momentum microscopy [3] and micro-ARPES [4]. On the other hand, Igor-pro is the software that the community adopted for its flexibility and scripting option but it comes with several drawbacks: Igor language's difficult syntax and conventions, usage of proprietary software, very limited cross-version compatibility, MS Windows only. This means in the context of the very time constrained conditions of a synchrotron beamtime, quick analysis of the data is nearly impossible, hampering a quality decision making process. Cassiopy is an answer to these problematics. It is a python-based, open-source, cross-platform, solution thay allows for most common data analysis tasks beyond data viewing. It involves a dedicated architecture to accomodate new parsers/loaders to enable cross-electron analyzer compatibility. The analysis methods range from basic image analysis with standard filters to more complex processing schemes so that one can get a quality on-the-fly data treatment. Used as a python module it makes for efficient scripting of new functions. Ultimately experimentalist will be able to use the principal clustering techniques to analyse ARPES data quickly. The aim of this presentation will be to familiarise the audience with this solution.

References

- [1] Schüler, M. et al. Phys. Rev. X 12 (2022), 011019.
- [2] Rongione, E. et al. Advanced Science 10 (2023), 2301124.
- [3] Reutzel, M., Jansen, G. S. M. & Mathias, S. Advances in Physics: X 9 (2024), 2378722.

[4] Iwasawa, H., Ueno, T., Masui, T. & Tajima, S. npj Quantum Mater. 7 (2022), 1–9.



Figure 1. Example of principal component analysis applied to a micro-ARPES dataset

Disorder effects in the Band structure of Transition Metal Dichalcogenide alloys - A_xB₁. _xSe₂ (A, B= Cr, Mo, W)

<u>Sarath Sasi</u>¹, Aki Pulkkinen¹, Laurent Nicolaï¹, Christine Richter^{2,3}, Karol Hricovini^{2,3}, Ján Minár¹

¹New Technologies Research Centre, University of West Bohemia, Pilsen, Czech Republicrganization, Address, City, Country (Arial 9)

³Université Paris-Saclay, CEA, CNRS, LIDYL, Gif-sur-Yvette, France

saraths@ntc.zcu.cz

Among the many two-dimensional (2D) materials that have gained attention since the discovery of graphene, transition metal dichalcogenides (TMDCs) stand out as promising candidates for electronic and optoelectronic applications. TMDCs, with the general formula MX2 (where M = Mo, W, and X = S, Se, Te), typically exhibit a distinct bandgap and spinpolarized bands. Numerous artificial methods have been proposed to engineer these properties, including chemical doping, strain induction, and external electric fields. Recent advancements in the synthesis of TMDCs have led to the emergence of 2D TMDC alloys [1]. This study focuses on the band structure of AxB1-xSe2 (where A, B = Cr, Mo, W) alloys with varying composition fractions (x).

The coherent potential approximation (CPA) [2] effectively models the average scattering properties in homogeneous random alloys and, within the KKR formalism, ensures no additional scattering when embedding an alloy component. Theoretical investigations using the SPR-KKR [3] package, employing CPA, reveal novel families of TMDC allovs that do not exhibit disorder effects in their structure across different composition band fractions. Experimentally, we examined the band structure of MoxW1-x Se2 alloys (x = 0 to 1) using Angle-Resolved Photoemission Spectroscopy (ARPES). The experimental results align well with the one-step model photoemission calculations within the SPR-KKR framework. Investigating these disordered systems provides fundamental insights, enhancing their potential applications.

References

Zhou, J., Lin, J., Huang, X., et al. Nature, 556, 355-359 (2018).
 Soven, P., Phys. Rev., 156, 809(1967).
 Braun, J., Minar, J., Ebert, H. Physics Reports, 740 (2018).

²LPMS, CY Cergy Paris Université, Neuville-sur-Oise, France

Real, Fake, or Just Useful? Generation and Evaluation of Synthetic Steel Micrographs with **Physical Constraints**

Daria M. Tomecka¹, S. Khalatabad², R. Jaeken¹, T. Nguyen Minh², L. Kestens², S. Cottenier², M. Sluvdts¹

¹ePotentia, Frans van Dijckstraat 59, B-2100 Deurne

Antwerpen, Belgium, ²Ghent University, Department of Electromechanical, Systems and Metal Engineering, Tech Lane Ghent Science Park - Campus A, Technologiepark 131, B-9052 Gent, Belgium

daria.tomecka@epotentia.com

Abstract

The growing adoption of AI in materials demands increasingly characterization large, diverse, and unbiased datasets. However, acquiring comprehensive steel microstructure data through electron microscopy remains resource-intensive and often leads to incomplete or imbalanced coverage of processing conditions. While generative AI offers a promising solution through synthetic data creation. validating these artificial microstructures raises fundamental questions about authenticity and scientific utility.

Within the AID4GREENEST project [1], we use the Ultra High Carbon Steel Database (UHCSDB) [2] to explore the entire synthetic data pipeline: from generation to validation. Synthetic micrographs are produced using a novel, phase- and attribute-aware sampling strategy based on real data distributions. We assess whether these microstructures correspond to realistic phase compositions and reflect variations driven by processing conditions such as annealing temperature, time, cooling method, and magnification. Our study critically examines standard AI-based image similarity metrics in both pixel and Fourier space, questioning their capacity to preserve physical and structural properties. These computational validations are compared with expert metallurgists' assessments to examine whether current approaches align with domain expertise in distinguishing real from synthetic microstructures.

Although our focus is on scanning electron microscopy, the core challenges of validating synthetic data with domain-specific metrics and expert alignment are broadly relevant across datadriven materials science, including spectroscopy. This comprehensive analysis seeks to establish robust criteria for evaluating synthetic microstructures, ensuring that generated data not only resembles real steel samples visually, but also retains scientific utility for downstream materials analysis.

References

- [1] DeCost, B. L., Hecht, M. D., Francis, T., Webler, B. A., Picard, Y. N., & Holm, E. A. (2017). UHCSDB: UltraHigh Carbon Steel Micrograph DataBase: Tools for Exploring Large Heterogeneous Microstructure Datasets. Integrating Materials and Manufacturing Innovation, 6(2), 197–205. https://doi.org/10.1007/s40192-017-0097-0
- [2] AID4GREENEST (2025). [Online]. Available: https://www.aid4greenest.eu



Figure 1. Mapping the microstructure space: Each point represents a single micrograph. On the right, a zoomed-in micrograph is shown-can you determine if it is real or synthetic?

Simulating X-ray photoelectron diffraction accelerating convergence efficiency

Shin Yasuda^{1,2}, Mariko Terao-Dunseath², Kevin Dunseath², Philippe Schieffer², Didier Sébilleau² and Sylvain Tricot²

¹Department of Physics, University of Toyama, 3190 Gofuku, Toyama 930-8555, Toyama, Japan

²Univ Rennes, CNRS, IPR (Institut de Physique de Rennes) - UMR 6251, F-35000, Rennes, France

d24c8201@ems.u-toyama.ac.jp

Perovskites (ABO3) play a significant role in various technologies, including spintronics, solar energy systems, and energy-efficient devices. They are particularly remarkable their distinctive for characteristics, such as ferroelectricity and the presence of two- dimensional electron gases (2DEGs) at certain interfaces such as LaAlO3/SrTiO3 [1] or Fe/SrTiO3 [2, 3]. These interesting properties come from tiny changes in the structure either close to the surface (e.g. oxygen vacancies) or deeper in the bulk (e.g. octohedral rotations), highlighting the importance of accurately probing such a structure. X-ray photoelectron diffraction (XPD) is regarded as an appropriate method for this analysis.

However, extracting the most information from experimental results needed for a structural interpretation requires a complex numerical simulation based on the multiple scattering theory.

Such a calculation can be performed using the MsSpec code [4, 5]. However, when simulating large clusters of perovskites (approximately 1000 atoms), the resulting spectra exhibit divergence (see Fig. 1 (a)). To tackle this issue, previous studies proposed "Simple renormalization schemes," which were effective in avoiding divergence in smaller clusters (for instance, a 50-atom Cu (111) system) [6]. Nevertheless, this approach proved to be timeconsuming for large clusters. In this work, we address this problem by showing that atomic chains are responsible for the poor convergence, and that we can greatly improve the simulation by using the G1 renormalization scheme on one-dimensional atomic chains that exist in the cluster. This approach allowed the renormalization procedure to be applied more effectively to large clusters, and improved the calculated XPD spectra for SrTiO3 (STO) when compared to the experiment (Fig. 1 (b)).

- [1] Y.-Q. Gao et al., Physica E, 80 (2016) 195–201
- [2] P. Catrou et al., Phys. Rev. B 98 (2018) 115402
- [3] R. Arras et al., Phys. Rev. B 102 (2020) 205307
- [4] D. Sébilleau et al., Comput. Phys. Commun., 182 (2001) 2567–2579
- [5] S. Tricot et al., J Electron Spectrosc 256 (2022) 147176
- [6] A. Takatsu et al., Phys. Chem. Chem. Phys., 24 (2022) 5658–5668





Figure 1. Experiment and simulation for a cluster of 1920 atoms. (a) The renormalization (G₁) is not used. (b) G₁ is implemented using an atomic chain in (001) direction. In both cases, we employed a scattering path filtering technique referred to as "Forward Filtering (FF)" to obtain results within an acceptable timeframe.