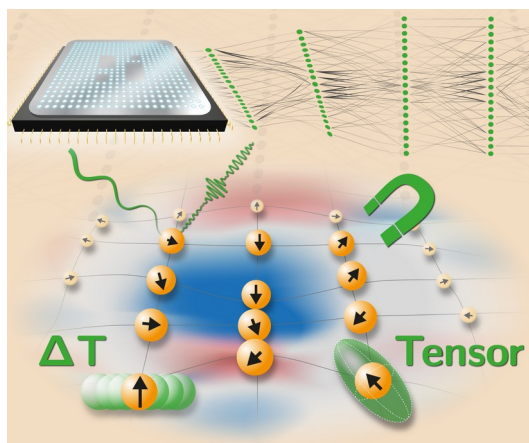


Machine Learning of First Principles Observables



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

Simone Köcher

IET-1, Forschungszentrum Jülich GmbH, Germany

Angela Harper

Fritz-Haber Institut der Max Planck Gesellschaft, Germany

Hanna Türk

École Polytechnique Fédérale de Lausanne, Switzerland

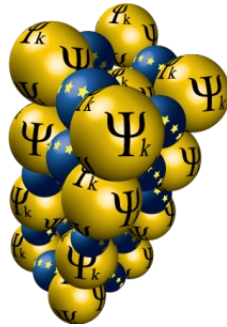
Elena Gelžinytė

Fritz-Haber Institut der Max Planck Gesellschaft, Germany

Giulia Glorani

Fritz-Haber Institut der Max Planck Gesellschaft, Germany

With the support of:



Psi-k

DFG Deutsche
Forschungsgemeinschaft
German Research Foundation

German Research Foundation (DFG)

FRITZ-HABER-INSTITUT
MAX-PLANCK-GESELLSCHAFT



Fritz-Haber-Institut der Max-Planck-Gesellschaft

Session Thermodynamic Observables

The opening talk was given by Prof. Karsten Reuter. He outlined the context, the potential, and the challenges of machine learning (ML) methods in material sciences from the well established machine learned interatomic potentials (MLIPs) as efficient surrogate models for the potential energy surface (PES) to ML models predicting observable properties directly. His talk emphasized the importance of moving on from the simplified, idealized simulation models to more realistic systems in particular with respect to interfaces and surfaces. ML methods are indispensable for embracing complexity in atomistic simulations and approaching the real-life experimental samples.

Moreover, Prof. Reuter stressed the significance of multi-scale modelling for approaching experimental length and timescales. Automated process exploration techniques can bridge the gap between first principles modelling and microkinetic simulations to derive thermodynamic properties as well as spectroscopic observables of complex, dynamic systems and empowering a higher level of complexity in high-throughput virtual screening.

Dr. Michele Simoncelli presented how the unified Wigner theory of thermal transport can be combined with MLIPs to predict the thermal and vibrational properties of solid materials with variable degree of disorder in the atomic bonds, ranging from perfect crystals to amorphous glasses. He showed that the macroscopic thermomechanical properties can be engineered by modifying the degree of atomistic disorder.

Dr. Christian Carbogno continued on the topic of thermal transport focusing on the impact of strong anharmonic and vibronic effects on heat transport. He demonstrated that MLIPs trained on anharmonic training data sampling the PES beyond the ground state minimum regions can model heat transport in a variety of materials. With a combination of symbolic regression and sensitivity analysis, material screening for optimal transport properties can be facilitated.

Prof. Nong Artrith reiterated the significance of considering the complexity of structural space when studying phase diagrams and ionic conductivity of battery materials. MLIPs are an invaluable tool for sampling amorphous phase space and elucidating structure-composition-property relationships. To bridge the gap between theory and experiment, an X-ray absorption spectra (XAS) database facilitates the correlation between theoretically modelled structural motifs and experimental samples.

After discussing some specific questions concerning the individual talks, the discussion soon focused on the data for training ML models. Several databases such as materials project, Aflow, open catalyst project, and different data repositories, where datasets of publications or entire trajectory data are published, were pointed out. However, concerns were raised with respect to comparability, reliability, and completeness of the data in particular of uncurated databases and of data based on higher level calculations such as spin-polarized. It is not always guaranteed that the calculations in those datasets were conducted correctly, are sufficiently converged and metadata about the settings might be missing. The same applies to experimental databases, where metadata might be incomplete and the accuracy and reproducibility of the experiment in doubt. Publishing workflows instead of or in addition to datasets makes the datasets reproducible, allows users to repeat the calculations and extend existing datasets by more data points. It was pointed out that

especially for areas, where calculations beyond the ground state or higher level or extremely accurate calculations are required, such as thermodynamics, spectroscopy, and magnetism, the available data is rather sparse both experimentally and theoretically.

The question of how to climb up the multi-scale ladder and to achieve experimentally relevant properties with modelling was raised. The gap between the idealized, simplified models in simulation and real samples was still considered to be significant. Following the statements by Prof. Reuter and Prof. Artrith, including a higher level of complexity in theoretical simulations was deemed essential. For more complex systems such as multi-phase boundaries, ML models that are transferable, universal, and linear scaling with number of species, are a powerful tool. So-called foundation models such as MACE-MP-0 are already available but might not have sufficient accuracy for predicting certain properties. However, foundation models are a great and effective starting point for active learning approaches, which customize the ML model to the system of interest and increase the accuracy for a specific problem. On the other hand, both theoreticians and experimentalists need to move towards each other, the theoreticians modelling more complex systems as well as the experimentalists studying more simplified samples to understand fundamental relationships and processes. A lot still needs to be done in the area of modelling properties determined by the materials' nano- or microstructure, to bridge first principles calculations to larger length and timescales, where ML promises a large potential for efficient upscaling.

Structural complexity was also discussed in the context of heat transport not only in amorphous materials as presented by Dr. Simoncelli but also in polycrystalline systems with grain boundaries and interfaces. A number of approximate models for simulating heat transport in these systems exist, but they all rely on phenomenological assumptions. Overcoming the phenomenological approximations employed in these models is an open problem which might benefit from recent developments in MLIPs.

Session Electronic Structure & Long Range Interactions I

The session started with the overview talk given by Prof. Gábor Csányi, who introduced the recently developed concept of foundation atomistic models for MLIPs based on the MACE architecture, which combines equivariant message passing neural networks with higher body-order local descriptors derived from atomic cluster expansion. Foundation models, such as MACE-MP-0 for inorganic materials and MACE-OFF for organic molecules, are trained on highly diverse training sets of small systems yet show excellent transferability, stability, generalisability, and universality with qualitative to quantitative accuracy for a broad range of applications. In particular, MACE-MP-0, which encompasses the entire materials project database, is a powerful model for pre-screening or as a starting point for fine-tuning the MLIPs for a specific application with iterative or active learning methods. By combining the MLIPs with spectroscopic properties, the bridge to experimental observables is built.

Prof. Janine George picked up on the topic of descriptors of local atomic environments and introduced the quantum-mechanical bond analysis to understand and model magnetic materials beyond the Goodenough-Kanamori rules. Based on MAGNDATA, a database of magnetic structures, Prof. George discussed the impact of different geometrical and atomistic features on magnetic properties and introduced standardised workflow and analysis tools for reproducible materials simulations.

Sergey Pozdnyakov challenged the necessity of explicit rotational equivariance in atomistic machine learning models. Based on knowledge from 3D point clouds, the Point Edge Transformer model based on an ensemble of local coordination systems and message passing networks of higher order inherently learns approximate rotational equivariance for both rotational invariant as well as equivariant targets. The inaccuracy due to lack of explicit rotational symmetry of the model is deemed negligible for most applications and does not impact the stability of molecular dynamics (MD) trajectories. Alternatively, an a posteriori exact symmetrization scheme can be applied, which however adds to the computational cost.

Prof. Kulbir Ghuman focused on the importance of incorporating defects and amorphism in theoretical structure models in particular in the areas of energy materials, electrochemistry, and catalysis. Since the properties and functionality of materials crucially depend on these aperiodicities, structure models need to be as realistic as possible embracing complexity to enable simulation of realistic properties. Prof. Ghuman demonstrated the potential of combining first principles calculations with ML methods first for the examples of modelling Fe-Ru alloy catalysts for NH_3 synthesis balancing NH_3 poisoning and catalytic activity and second for N_2 adsorption on doped Cu-Ni alloys.

The panel discussion started with the potential and challenges of foundation models as introduced by Prof. Csányi. Generalized models such as MACE-MP-0 are computationally much faster and scale better than density functional theory (DFT) calculations with a certain loss in accuracy. On the other hand, they are computationally still less efficient than classical force fields although they are more accurate. The trade-off between efficiency and accuracy as well as transferability of MLIPs was targeted. In order to compare models and analyse their performance, a Pareto front analysis of computational cost versus accuracy was recommended. As potential areas of application, which would justify the transferable, yet computationally expensive models, qualitative pre-screening methods for materials

screening as well as MLIP/molecular mechanics (MLIP/MM) embedding replacing quantum mechanics/molecular mechanics (QM/MM) methods were suggested. Additionally, several applications have already proven the benefit of using foundation models as starting point for active learning schemes tailoring the universal foundation models to fit a specific group of materials, while reducing the required amount of training data by orders of magnitude. The initial MACE-MP-0 proved too soft for describing phonons, but additional learning, e.g. with workflows such as introduced by Prof. George, showed improved accuracy for phonon predictions. Prof. George also gave a short outlook on more complex bond analysis methods for more efficient force constant and phonon predictions.

The extrapolation properties of the MACE method, which contains few non-linearities and tends to yield smoother models in comparison to highly non-linear models with more parameters, were discussed. However, a direct comparison of total number of parameters of different ML models was discouraged, since parameters in different model architectures such as linear and non-linear are rather different and not comparable. The question whether to improve models with more data or more physics, e.g. by multi-fidelity models, was raised. In contrast, Pozdnyakov's talk prompted the question, whether reducing the amount of explicit physical constraints, such as rotational symmetry, would not be an option to simplify models and increase efficiency while not reducing the accuracy significantly.

The models unconstrained by rotational symmetry were discussed next with a particular focus on where they would fit in the cost-accuracy Pareto front analysis. The improvement in efficiency and the reduction in complexity of MLIPs naturally results in an additional error by releasing the rotational equivariance constrain. But this additional uncertainty can be quantified and was demonstrated to be marginal in comparison to the overall uncertainty of the ML model for several examples. The potential impact of the lack of explicit rotational symmetry on the extrapolation properties of ML models was discussed controversially.

Possible areas of application of foundation models and of MLIPs in general were debated in the context of Prof. Ghuman's talk on the importance of realistic, aperiodic, yet complex model systems. Point defect might still be manageable to simulate with small, static systems and first principles. Surfaces and interfaces such as grain-grain boundaries based on DFT bulk structures however are entirely unrealistic due to enormous strain and require equilibration, which in most cases is beyond the scope of ab initio MD but necessitates MLIPs. Especially when simultaneously considering impurities or multi-phase boundaries of different materials, foundation models which cover the entire periodic system are a versatile and efficient tool for streamlining simulations.

Session Electronic Structure & Long Range Interactions II

In his overview talk, Prof. Michele Ceriotti picked up on the topic of MLIPs from the previous session. Based on the common features of different local atomic descriptors, he elucidated their common problems. First, degenerate, non-distinguishable structures, which can be resolved with message passing networks or three-centre ensembles. Second, the exponentially growing number of coefficients, which can be circumvented with Wigner kernels. Finally, the finite range problem, which can be addressed by higher order descriptors in reciprocal space. Alternatively, Prof. Ceriotti proposed to develop ML models that instead of the PES directly target more complex properties connected to the electronic structure. He demonstrated, that a ML model trained on the ground state electronic density of states can simulate free energies at a finite electron temperature. When developing a ML model to predict the Hamiltonian, Prof. Ceriotti stressed that it is advantageous to optimize the model with respect to basis-independent observables of interest rather than the matrix elements and to include physically-motivated components, such as the orbital structure, to teach the ML model MO theory. Independent on which property the ML for Hamiltonians was trained, the predicted Hamiltonians can be used to compute other electron density related properties, such as excitation energies.

Alexander Knoll introduced the high-dimensional neural network potentials (HDNNP) as another class of MLIPs. While already the previous generations of HDNNPs included local atomic charges, only the latest 4th generation also covers long range interactions such as dispersion, electrostatics, and non-local charge transfer by coupling the short-range atomic neural network with a second neural network predicting the long range electrostatic energy contribution based on an equilibration of atomic charges. Knoll demonstrated the functionality of the 4G-HDNNP on the wetting properties of a gold dimer on MgO and oxygen diffusion in PrCeO. However, including long-range interactions in MLIPs as well as extending the predictable properties to atomic charges and spins not only increases the computational cost but also the complexity of the models significantly at the cost of user-friendliness, robustness, and reproducibility. Hence, Knoll emphasized the importance of automatized workflows and analysis procedures introducing RuNNer2.0 and runnerase.

Prof. Reinhard Maurer picked up on the Prof. Ceriotti's topic of learning the electronic wave function via the Hamiltonian. He encouraged the community to reevaluate some decisions and conventions that were made a long time ago for first principles methods but might no longer be relevant for modern methods, such as Gaussian orbitals. He demonstrated the potential of electronic structure surrogate models with the NQCDynamics.jl toolbox for applications in the area of non-adiabatic processes, such as hot electron dynamics, surface reactions, electronic excitations, electron-phonon coupling, and electronic friction.

William Baldwin presented an overview of how long-range interactions are included in the MLIPs, categorizing the approaches in fixed point and charge equilibration schemes. Implementations of different complexity were compared, as implemented in the MACE framework. These surrogate models account for self-consistent Coulomb interactions and more faithfully describe a structure's response to an external electric field. Baldwin applied the method to examples from the charge

distribution of metal clusters in water to dipoles and screening potentials of water on metallic surfaces in an external electric field.

The initial discussion focused on the long range interactions. While important for specific materials and systems, the increased complexity and computational cost of ML models including long range interactions was critically discussed. A hierarchy of methods or embedding procedures for the more sophisticated long-range MLIPs was suggested. The discussion about the significance of long range interactions raised the question of how to evaluate, whether your model is accurate enough for your system of interest. Knoll suggested to validate ML models with experimental observables to verify, that the model can reproduce physics, while Prof. Ceriotti pointed out a heuristic approach utilizing the dynamical matrix. Long story short, if you can create a reliable model with sufficient accuracy for your purpose without long range interactions, the discussion suggested to neglect them in favour of more simple and effective, purely local methods. But there are applications, where this is not possible.

Next, the discussion turned to ML models for the Hamiltonian and electronic structure in contrast to learning the observables of interest directly. Models that learn properties such as the band gap directly do not learn inherent physical properties such as symmetry and lack transferability. Deriving observables from a machine learned Hamiltonian might suffer from a certain amount of error propagation, but learning the Hamiltonian is more stable, transferable, and gives you access to any property you can derive from the Hamiltonian, its eigenvalues or the electron density. On the other hand, the model might be too flexible, therefore including symmetry and optimized weights can stabilize the models. The benefit of direct learning of the Hamiltonian is that you can design the loss function with different eigenvalues or regularize the eigenvalues. But there are also some technical difficulties with the diagonalization procedure. Even the first foundation models for Hamiltonians are already available.

When the local additive ansatz of MLIPs was compared to the rotational symmetry constraint as another inductive bias, the consequences of dropping the local additive ansatz were discussed. When neglecting rotational symmetry, Pozdnyakov had shown in the previous session that ML models can learn inherent rotational equivariance. However, the local additive ansatz is a prerequisite for a model to be size extensive, i.e. to be trained on small structures but be transferable to large systems. Data augmentation schemes to teach the model size extensive characteristics might be possible, but certainly very costly.

Prof. Ceriotti raised the point, that there has not been much activity to connect atomistic machine learning with cheminformatics approaches, which go beyond the electronic structure and target higher level properties such as solubility or toxicity. This kindled a discussion on the gap between atomistic descriptors and complex, high-level, large scale experimental observables by including more physics in ML models and working with experimental data instead of first principles data.

Finally, the issue of collaboration and insufficient exchange of codes and data within the community was debated. The idea to champion open source publication of both codes and datasets as well as minimum standards of documentation, curation, and continuous integration was supported. But on the other hand, the effort to accomplish this goal and to motivate scientists to invest their time was discussed controversially.

Session Magnetic Observables

The overview talk by Prof. Stefano Sanvito introduced the different properties characterising a magnetic material as well as the challenges inherent in predicting these magnetic properties with first principles methods. First, magnetic properties depend critically on extremely accurate energy calculations often requiring high-level methods such as coupled cluster (CCSD(T)) or spin-orbit coupling interactions on the level of spin-polarised or even non-collinear electronic structure calculations. Second, magnetic materials often necessitate the simulation of large systems and depend on long range interactions. Third, multiple magnetic phenomena pose open-boundary condition problems and complex, higher-order interactions such as spin-lattice interactions.

Prof. Sanvito presented the concept of non-self-consistent field (non-SCF) DFT first principles simulations. Based on the linear scaling, physically interpretable, transferable, rotationally invariant Jacobi-Legendre (JL) polynomial cluster expansion model, ML models can predict the ground state electron density, spin densities, and even non-collinear spin vector fields as initial guess providing a massive speed-up of DFT calculations. Due to the excellent extrapolation properties and analytical derivatives of JL, the ML models are transferable and also accurate for phonon and transition state predictions. With JL and similar spin force fields (FF), spin dynamics can be simulated.

Johannes Wasmer picked up on the cluster expansion and compared several ML models such as ACE, MACE, and SOAP for the example of learning the exchange interaction of magnetic impurities in a topological insulator Bi_2Te_3 in order to predict the critical temperature. He put particular emphasis on the importance of exchange, dissemination, and benchmarking of atomistic machine learning codes, workflows, tools, and datasets introducing the JuDFTteam (<https://github.com/JuDFTteam/best-of-atomistic-machine-learning>) and DAEMON network (cost-daemon.eu) activities.

Prof. Alessandro Lunghi moved on to magnetic single molecules and their spin relaxation based on spin-lattice decoherence. Linear scaling SNAP ML models are applied to learn not only the potential energy surface, but also phonons, components of the spin Hamiltonian, as well as the derivative of magnetic properties. Uncertainty-aware active learning strategies yield models that perform very well both in the structural space as well as in the chemical space, i.e. in the context of high-throughput screening of new material candidates.

Finally, Shuping Guo presented her work on predicting magnetic properties of double perovskites. Including onsite energies and d-state transfer integrals in the learning process, the double perovskites are first classified as anti-/ferromagnetic and the transition as well as the Weiss temperatures are predicted. Several of the compounds are identified as magnetically frustrated.

After discussing some specific questions concerning the individual talks and models, the discussion soon focused on the potential of ML models for the electron density. Based on the transferability of electron density predicting models, they can significantly speed up phase diagram investigations with a non-SCF DFT approach, even though the accuracy of the predicted electron density and energy might not be sufficient to determine the convex hull entirely without first principles. In principle, ML models predicting Hamiltonians and electron densities are equivalent. However, Hamiltonians depend on a basis, whereas the electron density though defined on a grid is more universal and is sufficient as input for several DFT codes to start from.

Besides, codes such as the Materials Learning Algorithms (MALA) can learn the local density of states, from which other properties can be derived without a final SCF cycle at the end. In other cases, learning the observable or property of interest directly, such as exchange coupling, can be easier or advantageous.

However, many magnetic properties depend on exceptionally accurate energy calculations necessitating high-level calculations. Furthermore, calculated properties, energies, and electron densities are naturally highly sensitive to physical approximations, such as exchange-correlation functional, which have to be chosen depending on the property of interest as well as the material system. Nevertheless, the first principles training and testing data for the ML models need to be robust and ideally generated automatically. But for example the automatized determination of the chemical active space (CAS) in CAS-SCF calculation is not always reliable and can taint the datasets. Considering the computational expense of high-level first principles calculations for magnetic properties, ML models are still located in the small data regime and need to be of limited complexity and not overly parametrized. Interesting future projects could work on ML datasets for active spaces or on ML on full configuration interaction (CI) to identify, which contributions actually matter.

Following Wasmer's introduction to best-of-atomistic-machine-learning and cost-daemon, the role of well curated repositories collecting, benchmarking, and disseminating codes and datasets in the area of atomistic ML was addressed again. Despite their undisputed importance, their cost in invested time and manpower was criticized. Funding rarely stretches to code or repository maintenance and very often code or data repositories and their maintainers and depositors are not given any credit in subsequent publications, especially when code or data were not published as journal contribution before. Furthermore, increased interaction with the users was recommended, where users can comment on experience, quality, usability of data and codes.

Based on the talks by Wasmer and Guo, the discussion moved to the Heisenberg model, which does not consider longitudinal spin relaxation and neglects electron-hole excitations. However, it is an easy approximation, which provides good results for many problems as well as decent thermodynamics.

The plenary discussion was closed with the outlook that there is still a long way to go to replace the full time-dependent (TD) DFT simulations of magnetic systems with coupled spin-lattice MLIPs MD.

Session Spectroscopic Observables I

The first session on spectroscopic observables was opened by Prof. Patrick Rinke, who in his overview talk introduced general concepts, successes, and challenges of ML applications in spectroscopy. ML can be applied to facilitate the two main goals and workflows at the interface between theoretical and experimental spectroscopy: firstly, spectra prediction, which is common in computational materials modelling, and secondly, property inference from spectra, which is more typical for experimental approaches. The latter predicts material structures and properties from spectral input or classifies the input according to different categories. The former can be facilitated by ML in different ways. The most common are: directly learning structure-spectrum correlations, accelerating the spectroscopy method itself or learning the Hamiltonian, from which spectral properties can be derived. One example of accelerated methodology is infrared (IR) spectroscopy, for which MLIPs can be trained and then used to speed up MD calculations. Paired with an ML model for the dipole, the IR spectrum can then be extracted from the auto-correlation function over the MD trajectory.

Prof. Rinke illustrated the different concepts with examples from a broad range of spectroscopy methods, such as molecular excitation spectra, core electron spectroscopy, UV/Vis spectroscopy, and structure/property inference from 2D nuclear magnetic resonance (NMR) spectroscopy. He emphasized the sparsity and lack of diversity in available spectroscopic datasets as well as the computational cost of establishing accurate first principles data sets, e.g. on the level of G_0W_0 theory, as prominent problems.

Dr. Tigany Zarrouk presented another example of ML facilitated structure inference from spectral data - X-ray photoelectron spectroscopy (XPS) on oxygenated amorphous carbon. He combined a ML-based XPS model trained from first principles GW data with a MLIP for the C-O system. With a grand-canonical Monte Carlo within a modified Hamiltonian formalism, the XPS spectra are deconvoluted into motif contributions bridging the gap between experiment and theory and providing access to realistic structural models by an experiment-driven materials modelling approach.

Prof. Rose Cersonsky moved on to the rather new research field of optical properties of photonic crystals, which requires the efficient computation of photonic band structures for three-dimensional patterns of materials with different permittivity. ML models show a big potential on providing surrogate models for photonic structures to derive new insights into the structure-property landscape. However, the scale-covariance of the governing equations poses a challenge for ML models.

Finally, the application of ML models in the area of X-ray spectroscopy was presented by Clelia Middleton. She illustrated how the ML model XANESNET, which predicts K- and L-edges of various transition metals based on the atomistic structure, was extended to light, non-metallic elements by the introduction of an additional descriptor for the electronic properties. The descriptor is based on the partial density of states (p-DOS) and enables the prediction of sulphur K-edges, which Middleton demonstrated for the analysis of the products of the photosimulated ring-opening of thiophenone with X-ray absorption near edge structure (XANES) spectroscopy.

Based on the conceptual strategies of how to integrate ML in theoretical spectroscopy introduced by Prof. Rinke, the discussion started on which was the correct or optimal strategy. The best possible strategy depends on the availability of

data or the cost of generating the required data for training the model and the trade off of the cost of data generation versus the gain in efficiency by the ML model.

The advantages and disadvantages of learning a spectrum as a whole versus learning of the state eigenvalues were debated. Learning eigenvalues over a chemical space is challenging since the number of relevant states varies for each system. Additionally, the intensity and line broadening of peaks need to be learned, simulated or ultimately fitted. This raised the question of how to compare spectra and quantify agreement. Systematic shifts, different line broadening or different intensity ratio decreases the direct overlap of two spectra, although the agreement from a spectroscopist's point of view might still be quite good. Different loss functions and similarity coefficients are available. The approach to identify candidate structures with best similarity, provide them with a ranking, and ceding the final judgement to the user was suggested, but this approach can only find already known structures and does not allow for the identification of new compounds.

In the context of comparing spectra, the issue of reproducibility, accuracy, and reliability of experimental spectra and their sensitivity to different spectrometers, sample preparation, and references were controversially discussed. But also the accuracy of theoretical spectra was under scrutiny. In particular, the finite cluster approach with H-saturation by Prof. Rinke was addressed. The approach, which is necessary to saturate dangling bonds and ensure closed-shell configurations, in general does not result in additional artifacts in the spectroscopic calculations but can influence the convergence.

The increased complexity of recently studied materials and the ensuing challenges to spectroscopic calculations as well as ML predictions were debated. The more local the spectroscopic observable, e.g. core electron energies, the more reliably the property can be expressed in the local descriptor expansion and computed in finite model systems. Less local properties might be more challenging. Another factor is the scaling property of the ML models with number of species and system size.

Since most ML models presented in this and previous sessions were based on first principles training data, the relevance of experimental data was disputed. Naturally, all the ML models for spectroscopic observables are ultimately motivated by bridging theoretical models of material systems with experimental observations for validating theoretical models as well as facilitating the interpretation of experimental spectra (inference). The question how to make experimental data more relevant for ML model training was addressed and resulted in a detailed discussion about FAIR data management and the significance of documented metadata, in particular of prior technical or expert knowledge. Moreover, intensified communication to understand the scientific questions, technical details, and experimental problems better was emphasized. Even a basic definition such as, what is a material, might be fundamentally different between experiment and modelling and the transfer between an experimental sample and the atomistic structure is often not straightforward. Finally, determining the reproducibility of experimental data by repeating the measurement on the same sample as well as an identically produced sample was suggested, as it would increase confidence in experimental data, provide an uncertainty measure, and help identify outliers and false data.

Session Spectroscopic Observables II

The second overview talk about spectroscopic observables was given by Prof. Rebecca Nicholls. She introduced core-loss spectroscopy by electron or X-ray beams, which can provide information about bonding and atomistic structure of materials. Depending on the applied energy, phonons, plasmons or single electron excitations are addressed and the depth of excitation in the material varies making it a surface sensitive method. Electron energy loss spectra (EELS) can be simulated from first principles by computing the double differential cross section and comparison with calculated p-DOS allows for conclusions about the bonding situation. Prof. Nicholls presented the significance of theoretical simulations for experimental analysis with examples for understanding superconductivity in REBCO materials, the surface structure of LiNiO_2 battery materials, and the Zr suboxide in the ZrO_2/Zr interface. While theoretical simulations can facilitate the interpretation of experiments, Prof. Nicholls stresses the importance of keeping the scientific question in focus. Considering the computational cost of EELS simulations and the materials' complexity in many applications, ML has a large potential of streamlining core-loss spectroscopic computations and boosting the understanding and design of materials.

Prof. Josef Granwehr presented electron paramagnetic resonance (EPR) spectroscopy as a method for spectro-electrochemical in operando spectroscopy of organic radical batteries. The experimentally measurable g-value proves to be sensitive to the state-of-charge (SoC). In order to elucidate the correlation between SoC and structural and redox information, MD simulations combined with DFT-trained ML models predicting g-values from atomistic structures demonstrate, that the g-values is not only sensitive to the global radical concentration but also to local, heterogeneous concentrations during cycling, which also impacts the line shape. The accuracy of the ML model for predicting g-values is comparable if not superior to experimental uncertainties and transferable to different radical concentrations. Prof. Granwehr also introduced the neighbourhood analysis in order to unravel the impact of local structural features on the observable hyperfine coupling.

Prof. Claudia Draxl discussed the challenge of defining and quantifying similarity or discrepancies in spectra. Both experimental and theoretical spectra generation are subject to errors and inaccuracies due to deviating experimental conditions or different computational methods. By introducing spectral fingerprinting and different similarity coefficients such as the Tanimoto similarity, the comparison of spectra becomes more reproducible and robust. Un-/supervised learning of fingerprints allows to find patterns and trends rather than absolute differences. Prof. Draxl illustrated the spectral fingerprinting on screening for similar DOS of perovskites, on optical spectra of layered boron nitride, and different spectra of Ag. She referenced the C2DB database of 2D materials and presented the NOMAD infrastructure for FAIR dataset management.

Prof. Stefan Sandfeld's talk extended the session to ML applications beyond first principles atomistic applications. He highlighted the general challenges of scientific ML in contrast to computer science, such as sparse or noisy data, the necessity to generate accurate, transferable models that show an efficiency gain with respect to classical models, and the trend towards highly specialized models which are difficult to generalize to other scientific domains. Prof. Sandfeld presented examples from the area of image processing of electron microscopy, where U-net architectures are successfully applied for feature detection in time-series data and synthetic data generation replace time-consuming and less reproducible labelling procedures. For inverse problems and inverse dimensionality reduction, he illustrated a GAN model

that produces Ising systems for specific temperatures and pointed out the potential of latent space design.

The panel discussion started with the sensitivity of spectroscopy in general and the DOS and EELS in particular to temperature, defects, and surfaces and how well theoretical methods can replicate those effects. Of course, theoretical methods are based on certain approximation such as periodic boundary conditions, nevertheless they are in principle able to model complex systems as long as the atomistic model of the system is as close to the real sample in experiment as possible. The importance of sharing real data not only idealized, perfect results on both the experimental as well as theoretical side was emphasized. The interesting things happen, when results are unexpected and not ideal. Furthermore, it is also important to learn from failures. Therefore, recording of as much metadata as possible is necessary to identify the cause of failures or deviations.

The discussion then focused on the interface between experiment and theory and the process of sharing data. In the crystallography community, frameworks and platforms for sharing and documenting data already exist and work quite well. However, data once published are rarely updated with further knowledge or commented by users. Materials, which are not perfect but feature impurities or dopants are rarely published or more specific metadata is missing. There are infrastructures such as NOMAD for data and metadata management also including additional information such as synthesis procedures. Since scientists from many disciplines such as chemistry, computational sciences, and physics come together in material science, also the documentation of essential knowledge about the material, experimental procedure or simulation was deemed necessary, that might not be obvious to people from a different field of research.

Finally, the lack of motivation to invest time and resources in diligent and comprehensive data management was critically discussed. In other areas such as the food industry, pressure applied by official agencies has resulted in documentation of data and traceability of supply chains. The pressure in science was deemed insufficient also since consistent quality standards for data management are not defined. Moreover, methods to unify data from different sources need to be developed and established. Documentation of datasets is key as well as generally accepted reference data, reference codes, and reference samples or examples.

Session Electronic Structure & Long Range Interactions III

Luca Leoni presented his work on utilizing MLIPs for the simulation of polaron hopping dynamics. The simulation of polaron mobility with first principles methods is difficult due to the charge and excess magnetic moment of the polaron, requires large supercells, and is computation-intensive. Leoni demonstrated the gain in efficiency by replacing first principles methods with MLIPs based on equivariant graph neural networks and trained on ab initio MD (AIMD), where the local descriptor is extended by the polaron degree of freedom. Applications on small polaron mobility in MgO, TiO₂, and F-doped TiO₂ illustrate the accelerated performance as well as good accuracy.

The ensuing discussion addressed the description of the polaron degree of freedom in the feature vector of the local descriptor. An explicit encoding of the polaron localization is necessary, since the geometric distortions due to the polaron are much smaller than thermal displacements. Hence, the polaron position is not indirectly encoded in and cannot be inferred from the atomic structure. In the current descriptor, the polaron position is encoded as a kind of charge state encoding with a 0/1 integer added to the local atomic descriptor for whether the polaron is localized on this atom or not. The integer makes the energy landscape discontinuous with an energy discontinuity on the order of meV/structure. Defining the polaron position with a floating point number improves continuity but also allows for the description of polaron delocalization. However, quantifying the amount of polaron localization from DFT calculations is not straightforward and unambiguous. Alternatively, a ML model predicting the charge density could be applied, but would lose efficiency for long timescale MD simulations.

In order to improve the energy barrier predictions in the TiO₂ example, additional sampling around the energy barrier had been applied. Sampling of the PES purely by AIMD naturally puts a strong emphasis on the minimum energy areas. Especially when upscaling the method to larger structures, pure AIMD training might not be the method of choice, since you need at least one jump of the type you are looking for in the training data. Replacement or extension of the AIMD training data with nudged elastic band (NEB) might be a solution for more homogeneous, effective sampling and training of the model.

The question about the suitability of the ML model for large polarons was answered negatively by Leoni. Even if the model may be able to treat such cases it would be impossible to create a database needed for training due to the need of supercells that, to this day, not allow for long enough MD runs or even NEB computations to be performed in a suitable time. Alternatively, large polarons could be described via a ML model for the Hamiltonian.

Bartosz Brzozowski introduced the MALA package, a graph neural network based ML framework for predicting electronic structures. By combining SNAP descriptors with multi-layer perceptrons mapping the bispectrum components to local DOS (LDOS), a linear scaling, rotationally invariant ML model for the electronic structure is generated. First applications to the liquid-solid phase diagram of molecular hydrogen at various pressures and temperatures show good results at chemical accuracy and illustrate that the attention neural network model selectively focuses on relevant pieces of information. The extension to SE(3)-equivariant models by tensor field network convolution can directly predict higher rank observables as well as LDOS.

The subsequent discussion initially focused on placing the MALA model in the bigger picture of ML models for electronic structure. There are other graph neural network models around. Whether there is already a model available that combines SE(3) equivariance and attention graph neural networks and how MALA compares performance-wise was left open. Brzozka introduced the number of electrons as a metric for the error of the model, derived by integrating over the electron DOS to get the LDOS, which not always reproduces the correct number of electrons but was deemed manageable. The temperature degree of freedom of the LDOS was mentioned, which raises or lowers the electron energy levels independent of the ionic temperature and can be included in the model.

The extension of the MALA package to multi-species systems was addressed. It will change the scaling properties but enable the application to the future periodic systems of interest such as proteins and superconductors.

Session Databases & Reaction Networks

The final session was opened by Prof. Johannes Margraf presenting the topic of ML in chemical reaction space. He introduced the exploration of chemical reaction networks of organic molecules as a crucial step within the multi-scale approach towards microkinetics of simulating gas phase reactions and heterogeneous catalysis. By thoroughly sampling of the PES of the reactants including global optimization of the conformational space as well as rare-event sampling for the kinetics, the rather complex reaction networks, that grow exponentially with number of atoms, can be simplified and pruned systematically. However, in order to study reactivity, radicals and open-shell molecules need to be considered, too. ML models allow the efficient exploration of larger reaction networks. Prof. Margraf demonstrated on various examples, such as combustion and heterogeneous catalysis on Rh surfaces, that ML models trained on atomic energies are superior in predicting reaction energies and that atomic energies are superior weights for reaction paths than total energies. Furthermore, he compared different free energy approximation and demonstrated their impact on kinetic rate constants.

Dr. Jonathan Schmidt presented a prototype-based high-throughput material screening for perovskites as well as mixed perovskites with graph attention neural networks searching for thermodynamically stable and metastable materials in convex hull simulations. Symbolic regression models identified DOS, Fermi level and the element in the Z-position as decisive for thermodynamic stability and superconductivity properties. In the course of those and similar projects, the Alexandria database was developed, which comprises 4.5 Mio crystal structures as well as convex hull data for over 115,000 structures on the level of PBE, SCAN, and PBEsol. More recently, the database was extended by 2D and 1D materials by transfer learning. In combination with other datasets such as ICSD or materials.colabfit.org, it was used to train and benchmark universal MLIPs and generative models such as MatterGen and MatterSim.

Prof. Olexandr Isayev introduced the latest developments in the AIMNet family of chemically inspired deep neural network potentials. The universal, transferable MLIPs for the most common elements in organic molecules are trained on highly accurate (e.g. def2-tzvpp, CCSD(T)) first principles calculations. The recent advances of AIMNet, the fully reactive, linear scaling AIMNet2, also considers open-shell as well as charged molecules of different multiplicity and adds physics-based long-range terms to the ML-parametrized, additive short-range energy. Prof. Isayev demonstrated the potential and transferability of various AIMNet models for chemical reaction network exploration, transition state optimization, reaction thermochemistry, and even identification of key reaction mechanisms in complex reactions such as Diels-Alder, click chemistry, and Suzuki cross-coupling reactions. By combination with nano-reactor MD, the MLIPs are extended to extreme dynamics at high temperature.

Finally, Dr. Pierre-Paul De Breuck addressed the scarcity, limited transferability, and comparability of datasets. In all the areas of material science, high-fidelity data is much scarcer than low-fidelity data both in theory and experiment. Furthermore, first principles data is usually not directly comparable with experimental data. Dr. De Breuck presented MODNet, a feed-forward neural network based on material-optimal, physically meaningful features and joint learning, which balances small datasets with multi-fidelity datasets of varying accuracy and simultaneously provides

a confidence measure. The method is benchmarked for electronic band gap predictions and refractive indices of optical materials as well as TRIP-TWIP Ti alloys.

The panel discussion picked up on the last talk by Dr. De Breuck and enlarged upon transfer learning versus delta learning and the onion model approach with asymptotic error. The denoising approach with multi-fidelity data was debated, which provided good results but is difficult to analyse and to reproduce for different data. Also, whether the deviation of different DFT data and in particular with respect to experimental data could be defined as noise despite lacking a typical noise distribution was discussed. Experimental data is also particularly difficult to assign with individual fidelity labels. If consistent metadata is missing, it requires the analysis of the original publications. The data requirements for transfer learning of multi-fidelity data depend very much on the distribution of the data in the fidelity space. The larger the difference between the high and low fidelity data, the more data in general and the more data in particular in the high fidelity class are required for reliable accuracy.

The transfer learning was also addressed in the context of Dr. Schmidt's study of thermodynamic stability to transfer convex hulls to convex hulls of increased fidelity. For magnetic materials, the convex hull can also serve as starting point for iterating through different magnetic configurations in supercells.

Prof. Isayev's state-of-the-art AIMNet models including charge and spin attracted further interest. In contrast to previous AIMNet generations, the second generation is trained on hybrid DFT, since CCSD(T) is unsuitable for open-shell datasets. A single model can predict different spin states, open-shell systems, and biradicals as well as differently charged and neutral components simultaneously. The larger error on the transition state prediction is not only due to the higher degree of flexibility but also due to the large radical structure, which requires extrapolating of the AIMNet model, which is trained on small organic structures with less than 50 atoms.

The impact of solvent on Prof. Isayev's examples on click chemistry and Suzuki coupling were debated. The AIMNet models are exclusively trained on gas phase molecules and also the predictions neglect solvents. Solvents might be included implicitly in the neural network, which also allows for description of the response to charges.

Similarly, the approximation of a static catalyst surface in Prof. Margraf's modelling of adsorbates and catalytic reaction networks was addressed. An explicit treatment of the surface activity might be feasible for small reaction networks, but is still beyond the computation limits for complex catalytic reactions.

The ML models presented in this and previous sessions focused on predicting the target properties and thermodynamic stability. More advanced approaches are suggested to additionally include kinetic information about the synthesizability in the prediction score as well as knowledge about which elements the experimental community usually works with.

Finally, the discussion moved again to the foundation models. For several of the presented applications, the MACE-MP-0 or similar foundation models do not work out of the box. Either the accuracy is not sufficient or the system is not stable. Nevertheless, the foundation models prove to be good starting points for fine-tuning with active learning and adaption of the models for the specific application.

We thank all the participants, online and in person, for their contributions to the panel discussions and the speakers for their commitment and the lively exchange.