MAX PLANCK RESEARCH NETWORK

BiG max on big-data-driven materials science

BigMax Workshop 2018 on Big-Data-Driven Materials Science

Kloster Irsee, Germany

April 10-13, 2018

Organizers
Jörg Neugebauer (Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf)
Peter Benner (Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg)

https://www.bigmax.mpg.de/10553/bddms2018

Contents

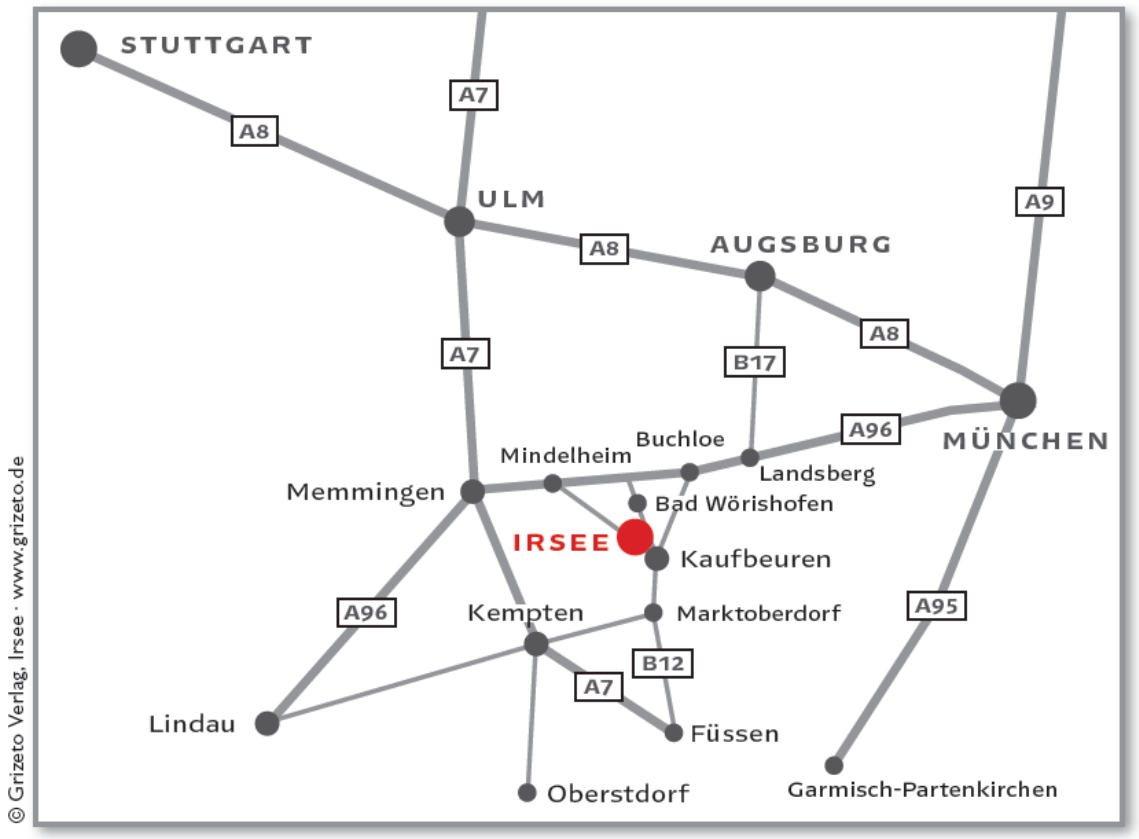
Materials science is entering an era where the growth of data from experiments and calculations is expanding beyond a level that is properly processable by established scientific methods. Dealing with this big data is not just a technical challenge, but much more: it offers great opportunities. Big-data analytics will revolutionize new material discovery and will make the successful search of structure-property relationships among multiple length and time scales possible.

Thus far, the search for new materials for new applications was limited to educated guesses mostly based on selective experiments. By tackling this big-data challenge with high-speed computing, extremely large, disparate databases and large-scale computations have to be dealt with. But recent advances in data mining will allow pattern recognition and pattern prediction in an unprecedented way. The outcome of the big-data-driven materials science approaches will then impact the way experiments and data analyses are done.

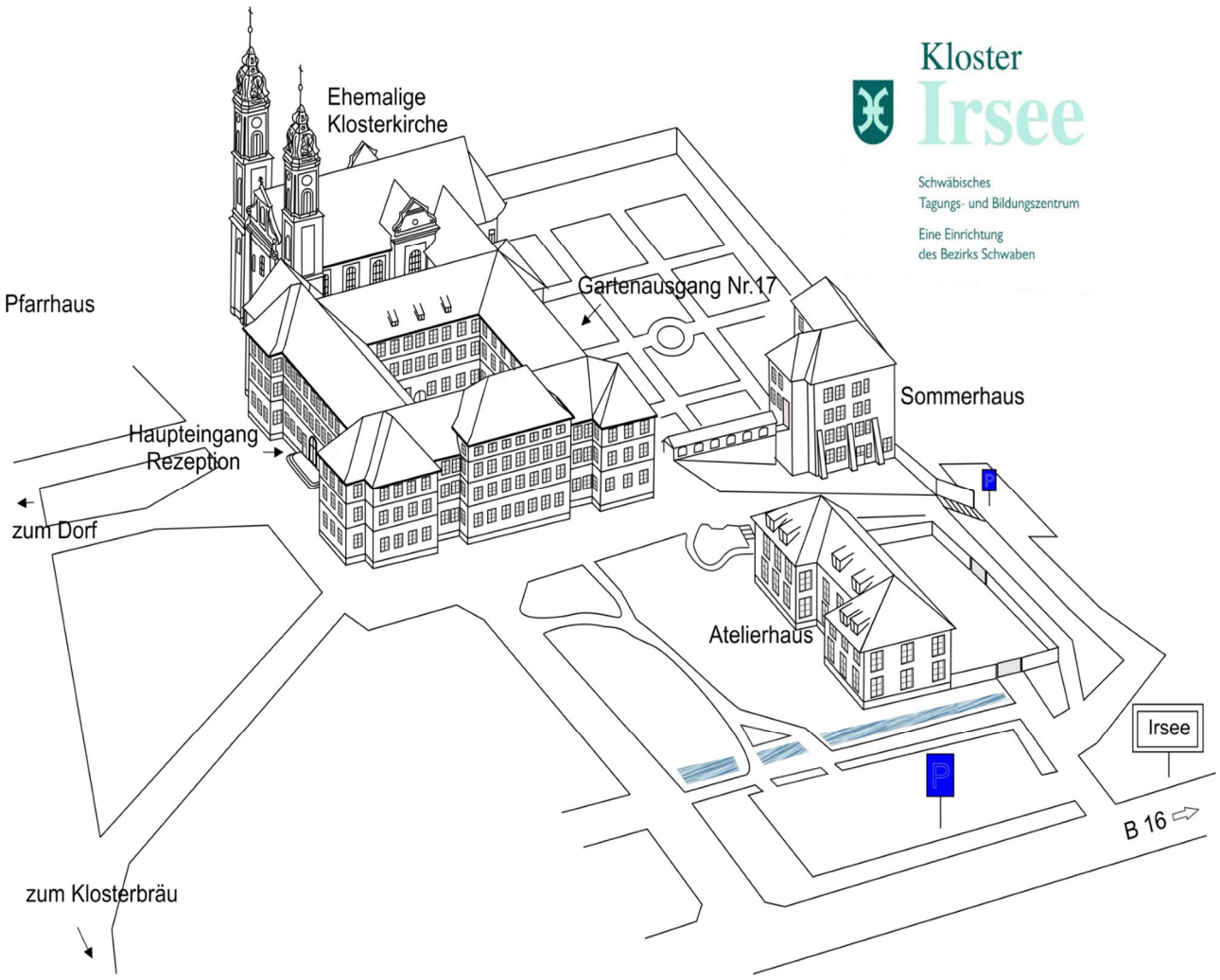
The involved Institutes of the Max Planck Research Network Big-Data-Driven Materials Science - BigMax cover a significant breadth in research areas, and they are convinced that the envisioned synergy will enable them and their MPIs to develop novel, domain-specific and property-specific methods to enter and shape the era of data-driven materials research. The goal of BigMax is to fully exploit these scientific potentials of materials science activities of the CPTS and to raise the consortium to world leadership in data-driven materials science.

At this workshop, members of BigMax will present relevant results that have been achieved during the first year to facilitate exchange and collaboration inside the network. Renowned experts of the field will give additional key note talks on recent advances in the field. Finally, there will also be potential new members of BigMax that will present their research and ideas to the network.

Kloster Irsee
Tagungs-, Bildungs- und Kulturzentrum des Bezirks Schwaben
Klosterring 4
D-87660 Irsee



There will be a shuttle service on Tuesday (arrival day) and Friday after lunch between the train station Kaufbeuren and Kloster Irsee.



Places

All talks and the poster parade will be held at the Festsaal (2nd floor).

The poster presentation will be held in the east wing of the 2nd floor.

Meals

- Lunch (buffet): at the Restaurant (ground level)
- Dinner (Tuesday and Thursday): at the monastery cellar
- Conference Dinner (Wednesday): at the restaurant (ground level)

Tuesday, April 10, 2018

14:00-15:00	Welcome snack and refreshment
15:00-15:30	Welcome
15:30-16:30	Stefano Curtarolo (Duke University): Data, disorder and materials (see p. 7)[1]
16:30-17:00	Coffee break
17:00-17:40	Claudia Draxl (HU Berlin): Making scientific data accessible: the materials encyclopedia (see p. 10)
17:40-18:20	Tristan Bereau (MPIP Mainz): Computational high-throughput screening of drug permeation through biological membranes (see p. 9)
18:30	Dinner
19:30	Guided tour through the monastery (German/English)

Wednesday, April 11, 2018

9:00-10:00	Gábor Csányi (University of Cambridge): On creating databases for machine learned interatomic potentials (see p. 7)
10:00-10:30	Coffee break
10:30-11:10	Paolino De Falco (MPIKGF Potsdam), Pawan Goyal (MPIDCTS Magdeburg): 3D SAXS tomography for bio-materials (see p. 10)
11:10-11:50	Shyam Katnagallu (MPIE Düsseldorf): Applied machine learning in field ion microscopy aided by field ion image simulation (see p. 11)
12:00-14:00	Lunch
14:00-14:30	Andrew Breen (MPIE Düsseldorf): Automatic crystallographic analysis of atom probe data utilising machine/deep learning algorithms (see p. 15)
14:30-15:00	Ralph Ernstorfer, Patrick Xian (FHI Berlin): Multidimensional photoelectron spectroscopy (see p. 15)
15:00-15:30	Christian Liebscher (MPIE Düsseldorf): Big datasets in time-resolved and multi-dimensional STEM (see p. 16)
15:30-16:00	Coffee break
16:00-16:30	Christoph Freysoldt (MPIE Düsseldorf): Automatic feature extraction from multi-dimensional STEM data (see p. 16)
16:30-17:00	Isabelle Mouton, Leigh Stephenson (MPIE Düsseldorf): Machine-learning approaches to reveal the position and identity of each atom in a material (see p. 17)
17:00-17:30	Mariana Rossi (FHI Berlin): Towards an accurate, high-throughput framework for the prediction of anharmonic free energies in molecular crystals (see p. 18)
17:30-19:00	Meeting Executive committee
19:00	Conference dinner

[1] Pages where to find abstracts are given in brackets.

Thursday, April 12, 2018

9:00-10:00	Vidvuds Ozolins (Yale University): Compressive sensing in computational materials science (see p. 7)
10:00-10:30	Coffee break
10:30-11:10	Matthias Rupp (FHI Berlin): Kernel-based machine learning for materials (see p. 13)
11:10-11:40	Anna Lehner (Fraunhofer IWM, Freiburg): Material properties in material data: A machine-learning approach for finding new hard-magnetic phases (see p. 12)
12:00-14:00	Lunch
13:00-14:30	PI/deputy meeting
14:30-15:10	Jilles Vreeken (Saarland University, MPIINF Saarbrücken): Interpretable local modelling for data-driven science (see p. 14)
15:10-15:50	Luca Ghiringhelli (FHI Berlin): Discovering interpretable descriptors through compressed sensing (see p. 11)
15:50- 18:30	Poster parade, poster session (and coffee)
18:30	Dinner
	possibly additional PI meeting

Friday, April 13, 2018

9:00-9:40	Angelo Ziletti (FHI Berlin): Automatic classification of pristine and defective crystal structures from noisy data using deep learning (see p. 14)
9:40-10:20	Michael Ashton (MPIE Düsseldorf): First-principles calculations of field evaporation in atom probe tomography (see p. 9)
10:20-10:40	Coffee break
10:40-11:20	Zirong Peng (MPIE Düsseldorf): Towards automated information extraction from atom probe tomography data (see p. 13)
11:20-12:00	Alex Eisfeld, Ulf Saalmann, Jan-M. Rost (MPIPKS Dresden): Machine learning in AMO physics (see p. 11)
12:00-12:10	Conclusion Remarks
12:10-14:00	Lunch

On creating databases for machine learned interatomic potentials

Gábor Csányi
University of Cambridge, UK

The last few years have seen fervent activity in applying machine learning techniques to create interatomic potentials and force fields, with lots of approaches being tried: kernel methods, neural networks, scattering transforms, symmetrised polynomials, etc. Much less attention has been devoted to thinking about the databases that these approximants are fit to. My talk will attempt to begin to address this, using the case study of a very extensive database of periodic structures of silicon, and a correspondingly extensive suite of benchmark tests that a materials modeller would use to judge the quality of a potential. I will argue that alternatives to Boltzmann sampling to generate the database will be important in the future.

Data, disorder and materials

Stefano Curtarolo
Duke University, Durham, USA

Critical understanding of large amount of data leads to new ideas for discovering materials: tangible examples reveal the future revolution.

Compressive sensing in computational materials science

Vidvuds Ozolins
Yale University, New Haven, USA

Compressive sensing allows to reconstruct the data (e.g., image) from an incomplete set of quasi-random measurements in a mathematically rigorous and efficient way. It is an important tool in modern information theory where it provides automatic recognition of sparsity. In this talk, we will highlight how compressive sensing is used to build highly accurate and systematically improvable physics models based on first-principles quantum mechanical calculations. This approach allows to extend the reach of DFT to very large system sizes (millions of atoms) and nanosecond time scales, enabling direct simulation of the thermodynamic and thermal transport properties of complex materials. Examples from the field of metallic alloys and thermoelectric materials will be presented to illustrate the advantages of this approach.

We also discuss a recently developed method for calculating Wannier functions (WFs) of periodic solids directly from a modified variational principle for the energy. Spatial localization is achieved by adding a weighted regularization term to the energy functional, chosen as the L1 norm of the Wannier function. This approach results in ”compressed” WFs with compact support (i.e., WFs are nonzero in a finite spatial region), and one parameter automatically controls the trade-off between the accuracy of the total energy and the localization of the WFs. In periodic solids, the compressed WFs can be chosen as irreducible

representations of the crystal symmetry group, allowing for an efficient construction of symmetry-preserving tight binding models. Efficient numerical algorithms for the variational minimization problem are demonstrated. Applications to periodic crystals and topological insulators are discussed to illustrate the advantages of this approach, and implementation within the density-functional theory is demonstrated.

First-principles calculations of field evaporation in atom probe tomography

Michael Ashton, Arpit Mishra, Christoph Freysoldt, and Jörg Neugebauer
Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf

Atom probe tomography (APT) has developed into a key technique for studying 3D element distribution in complex microstructures at near-atomic resolution. At the heart of the technique is the evaporation of single atoms from a very sharp tip exposed to huge electric fields (10^{11} V/m). However, commonly used geometric reconstruction algorithms fail to take into account any details of the evaporation mechanism, such as differences in evaporation between chemical species or from different sites. To shed light on the factors influencing electric field evaporation, we study desorption from various sites (ad-atom, steps, kinks) on prototypical metal surfaces by means of density-functional theory calculations. From these calculations, we obtain electric-field-dependent evaporation barriers, which will be used in subsequent APT simulations. We also find evidence for the roll-over effect at steps that has been proposed to explain orientational bias in evaporation trajectories.

Computational high-throughput screening of drug permeation through biological membranes

Roberto Menichetti, Kiran H. Kanekal, Kurt Kremer, and Tristan Bereau
Max Planck Institute for Polymer Research, Mainz

The partitioning of small molecules in cell membranes---a key parameter for pharmaceutical applications---typically relies on experimentally-available bulk partitioning coefficients. Computer simulations provide a structural resolution of the insertion thermodynamics via the potential of mean force, but require significant sampling at the atomistic level. Here, we introduce high-throughput coarse-grained molecular dynamics simulations to screen thermodynamic properties. This application of physics-based models in a large-scale study of small molecules establishes linear relationships between partitioning coefficients and key features of the potential of mean force. This allows us to predict the structure of the insertion from bulk experimental measurements for more than 400,000 compounds. The potential of mean force hereby becomes an easily accessible quantity---already recognized for its high predictability of certain properties, e.g., passive permeation. Further, we demonstrate how coarse graining helps reduce the size of chemical space, enabling a hierarchical approach to screening small molecules.

3D SAXS tomography for bio-materials

Paolino De Falco [a], Pawan Goyal [b], Martin Stoll [c], Wolfgang Wagermaier [a],
Richard Weinkamer [a], Peter Benner [b], and Peter Fratzl [a]
[a] Max Planck Institute of Colloids and Interfaces, Potsdam
[b] Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg
[c] Technische Universität Chemnitz, Chemnitz

The structural complexity of biological materials requires new methodology of material characterization in three and even four dimensions including time. We explore new 3D imaging methods based on X-ray scattering using synchrotron sources that can provide important information on the nanostructure of materials. In a tomography experiment at a synchrotron source, data collection requires several hours. Consequently, the size of relevant data produced by this new approach tremendously grows resulting also in an increasing need of computational power and time for 3D reconstruction of data. The focus of our research is the development of a new methodology for fast characterizations of the 3D nanostructure of bone. On the nanoscopic length scale, bone is a composite of a fibrous collagen matrix in which inorganic calcium phosphate particles are incorporated. The mineral particles decisively contribute to the high mechanical stiffness and strength of bone material. We aim to elucidate the 3D distribution of mineral particle sizes within a certain bone volume, which is a relevant parameter to characterize the influence of bone diseases on the bone's mechanical properties. We present first results from our approach based on SAXS (small angle X-ray scattering) tomography experiments at the Berlin synchrotron BESSY and initial ideas on mathematical algorithms for 3D image reconstruction.

Making scientific data accessible: the materials encyclopedia

Claudia Draxl
Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin
and Fritz Haber Institute of the Max Planck Society, Berlin

Knowledge and understanding of materials is based on their characterization by a variety of measured and/or computed properties. This includes structural features, mechanical and thermal behavior, electronic and magnetic properties, the response to light, and more. For the computational side, the NOMAD Laboratory (Novel Materials Discovery) Centre of Excellence (https://nomad-coe.eu) has created a data infrastructure to collect, share, and keep data available for at least ten years. The variety of data uploaded to the NOMAD Repository allows for in-depth explorations and discoveries. The NOMAD Encyclopedia is a graphical user interface that serves this purpose. I will review how it makes millions of calculations accessible, demonstrate its capabilities, and discuss our plans to extend the data archive by experimental data.

Machine learning in AMO physics

Alex Eisfeld, Ulf Saalmann, and Jan-M. Rost
Max Planck Institute for the Physics of Complex Systems, Dresden

We have explored machine learning concepts for three typical problems in atomic, molecular and atomic physics problems.
Firstly, we have used Gaussian processes in the context of excitonic Rydberg problems, secondly we have trained a convolutional network to predict the response of an atomic system to variable shapes of a strong short laser pulse, and thirdly, we want to filter out correlated signals in very noisy spectra. In the long run, our goal is to better understand the resulting structure of trained network with respect to the original physics problem.

Discovering interpretable descriptors through compressed sensing

Luca Ghiringhelli
Fritz Haber Institute of the Max Planck Society, Berlin

The number of possible materials is practically infinite, while only few hundred thousands of (inorganic) materials are known to exist and for few of them even basic properties are systematically known. In order to speed up the identification and design of new and novel optimal materials for a desired property or process, strategies for quick and well-guided exploration of the materials space are highly needed. A desirable strategy would be to start from a large body of experimental or theoretical data, and by means of "(big-)data-analytics" methods, to identify yet unseen patterns or structures in the data. This leads to the identification of maps (or charts) of materials where different regions correspond to materials with different properties. The main challenge on building such maps is to find the appropriate descriptive parameters (called descriptors) that define these regions of interest.
Here, I will present methods for the machine-aided identification of descriptors and materials maps applied to the metal/insulator classification, the prediction of novel 2D topological insulators, and the construction of a tolerance factor for the stability of perovskites.

Applied machine learning in field ion microscopy aided by field ion image simulation

Shyam Katnagallu, Ali Nematollahi, Blazej Grabowski, Jörg Neugebauer, Dierk Raabe, and Baptiste Gault
Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf

Field ion microscopy (FIM) allows to image individual surface atoms by exploiting the ionization of an imaging gas by an intense electric field in the vicinity of a needle-shaped specimen. Widespread use of atomic resolution imaging by FIM has been hampered by a lack of efficient image processing/data extraction tools. We employ advanced data extraction tools for automated detection of atoms and lattice defects for materials characterization. We also show the use of machine learning (ML) approaches for data

extraction and develop simple FIM simulations to aid ML by providing a perfectly labelled dataset on which the machine can be trained. The evaporation sequence in these simulations is based on nearest neighbor criterion used in ref [1]. After a brief overview on a new method developed for FIM data extraction we will describe the use of these FIM simulations for a quantitative analysis of distortions in experimental FIM images. These ideas are discussed towards improving the spatial precision of FIM data reconstructions.

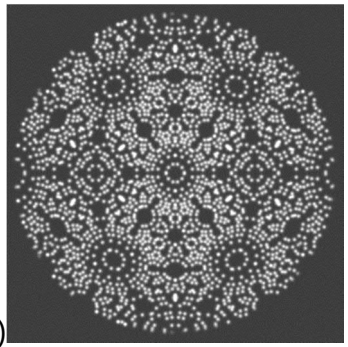
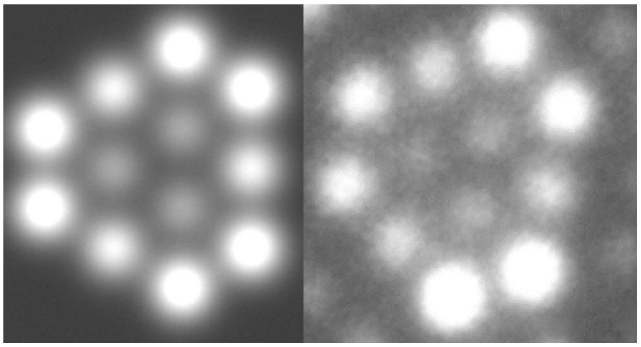
(a)  (b) 

Figure 1: (a) FIM simulation for a tip with 50 nm radius. (b) Simulation (left) of an outermost (222) bcc plane and the corresponding experimental (right) tungsten (222) plane's FIM image.

References

[1] A. J. W. Moore and J. A. Spink , "Field evaporation from tungsten and the bonding of surface atoms", Surf Sci, vol. 12 (1968), pp. 479–496.

[2] Drs. Michal Dagan, Paul Bagot and Prof. Micheal Moody are greatfully acknowledged for providing FIM data on tungsten.

Material Properties in Material Data: A machine-learning approach for finding new hard-magnetic phases

J.J. Möller, A. Lehner, D.F. Urban, C. Elsässer, and P. Gumbsch
Fraunhofer IWM, Freiburg

Data-analytics and machine-learning (ML) approaches are increasingly becoming useful as tools for the discovery and development of materials. At Fraunhofer IWM we are committed to the global vision of driving scientific and industrial exploitation of digital data information for material and processing technologies. Our efforts address the generation of (big-)data infrastructures for digital-twin representations of material systems, digital monitoring protocols for materials in products along process chains and life cycles, and the design of new materials with desired properties.

As a specific example, we present our use of kernel-based ML methods to predict optimal chemical compositions of intermetallic hard-magnetic phases for novel permanent magnets, which are key components for green-energy exploitation technologies. The hard-magnetic-property data set used for training and testing the ML models were obtained by combinatorial high-throughput screening (HTS) calculations using methods of density-functional theory. For encoding the information about structural, chemical, and magnetic properties in the HTS data set in a suitable digital form for ML tools, we use several, known

and novel material descriptors and we explore the predictive power of ML models built upon these and trained with HTS data. Such ML models with reasonably chosen descriptors and sensibly fitted parameters are indeed capable to predict promising structure-composition-property relationships for multi-component substitutes of state-of-the-art magnetic materials like $Nd_2Fe_{14}B$ – with similar intrinsic hard-magnetic properties but no or less amounts of critical rare-earth elements.

Towards automated information extraction from atom probe tomography data

Zirong Peng [a], Yifeng Lu [b], Leigh T. Stephenson [a], Shyam S. Katnagallu [a], Isabelle Mouton [a], Francois Vurpillot [c], Dierk Raabe [a], and Baptiste Gault [a]
[a] Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf
[b] Database Systems and Data Mining Group, Ludwig-Maximilians-Universität München
[c] Normandie Université, UNIROUEN, INSA Rouen, CNRS, Rouen, France

Atom probe data take the form of complex three-dimensional point cloud from which compositional and sometimes structural information must be extracted. This is often done in a rather simplistic way using regions-of-interest with pre-defined shapes that might not be suitable for complex microstructural features. We introduce here an efficient, automated computational approach for extracting features within atom probe tomography datasets, enabling to quantitatively map their thickness, composition, as well as the Gibbsian interfacial excess of each solute. We also introduce and discuss some possible pathways to enhance the accuracy of the compositional analysis by looking for specific patterns in the order in which ions are detected and retrieving the identity of ions that were previously discarded as background.

Kernel-based machine learning for materials

Matthias Rupp
Fritz Haber Institute of the Max Planck Society, Berlin

Computational discovery and optimization of novel materials requires accurate treatment on the atomic scale. While numerical approximations to the electronic structure problem enable this in principle, their applicability is severely limited by their high computational cost. In high-throughput settings, machine learning can reduce these costs significantly by interpolating between reference calculations. I will outline kernel-based machine learning approaches for accurate interpolation of electronic structure calculations. For these approaches, a numerical representation of atomistic systems that supports interpolation is crucial. Using our recently introduced many-body tensor representation, I will present empirical evidence for predictions of ab initio formation enthalpies with errors in the single digit eV/atom range on benchmark datasets of crystal structures, determination of phase diagrams of Pt-group / transition-metal binary systems, as well as initial results for energy and band gap predictions of Al-Ga-In sesquioxides.

Interpretable local modelling for data-driven science

Jilles Vreeken
Saarland University and Max Planck Institute for Informatics, Saarbrücken

Machine learning methods for global model induction, like support vector machines or artificial neural networks, are nowadays applied in a wide range of data-driven applications. Therefore, they appear like a natural tool also for scientific data analysis. However, although their models can reach astounding accuracies, they tend to offer surprisingly little insight into the underlying domain.

Local modeling methods like subgroup discovery address this concern by being potentially agnostic about parts of the input space in order to focus on specific effects that can be modeled in simple terms with high precision---in particular those that are unusual or outstanding given the global picture. Additionally, they achieve interpretability by using discrete symbols that correspond to meaningful notions of the discovery domain.

In this talk, I show how subgroup discovery was successfully used to discover novel insights from ab initio materials computations as well as to diagnose ML models for high throughput screening. Conversely, I show how the requirements of scientific discovery motivated novel extensions to the subgroup discovery framework and stimulated corresponding algorithmic innovations.

Automatic classification of pristine and defective crystal structures from noisy data using deep learning

Angelo Ziletti
Fritz Haber Institute of the Max Planck Society, Berlin

Big data is emerging as a new paradigm in materials science. A vast amount of three-dimensional structural data is provided by both computational repositories (e.g. http://nomad-coe.eu) and experiments (e.g. atom probe tomography). Computational methods that automatically and efficiently detect long-range order are of paramount importance for materials characterization and analytics. Current methods are either not stable with respect to defects, or base their representation on local atomic neighborhoods, which in turn makes it difficult to detect “average” long-range order.

In the proposed approach, for a given crystal structure we first calculate its diffraction pattern, expand it on spherical harmonics, and then use a neural-network model to obtain a compact, low-dimensional representation. We apply this workflow to a subset of materials from the Novel Materials Discovery (NOMAD) Archive, and show that our deep-learning-based approach compactly encodes structural information, is robust to defects (e.g. point defects, and/or strain), and allows to build easily interpretable structural-similarity maps.

This work received funding from the NOMAD Laboratory, a European Center of Excellence.

Automatic crystallographic analysis of atom probe data utilising machine/deep learning algorithms

Andrew Breen and Baptist Gault
Max-Planck-Institut für Eisenforschung, Düsseldorf

Atom probe tomography (APT) is a uniquely powerful micro-analytical technique capable of reconstructing the position and chemical identity of millions of atoms from a material specimen in 3D. The technique works by field evaporating ions from a sharp-needle shaped specimen towards a single-ion, position-sensitive time-resolved detector. When polycrystalline materials are analysed, crystallographic information can be retrieved from variations of the point density that form a pattern of poles and zone lines often observed on the detector. This is caused by atomic terracing of the tip surface and an associated change in local electric field which causes slight trajectory aberrations. Such information is particularly useful for the purposes of calibrating the resulting reconstruction and measuring relative crystal orientation – it is possible to completely characterise the crystallographic nature of individual grain boundaries (5 degrees of freedom) and compare this directly to atomic segregation of different elements to the interface. Such information has a profound influence on our understanding of the mechanisms governing mechanical and functional properties of materials.

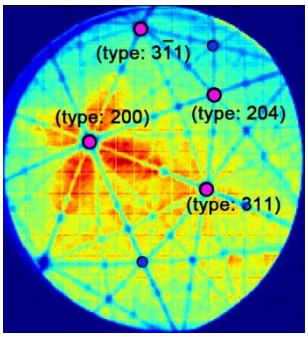


Figure 1: detector map from a pure-Al analysis showing a pattern of poles and zone lines. The position and identity of the poles was obtained from a preliminary implementation of a ML-based protocol.

However, currently the interpretation of this crystallographic information is done manually by the scientist and is very error prone – as a result is largely under-utilised. Here, we propose to fully deploy machine learning algorithms and techniques to automatically extract crystallographic patterns observed on atom probe detector maps. Preliminary results using neural networks have shown promising initial results, but the challenges and opportunities for materials physics to correlate atomic scale compositional and structural information are enormous, thereby justifying the request for a full-time position for a PhD student for three-year period.

Multidimensional photoelectron spectroscopy

R. Patrick Xian, Michele Puppin, Laurenz Rettig, and Ralph Ernstorfer
Fritz Haber Institute of the Max Planck Society, Berlin

Multidimensional photoemission spectroscopy (MPES) [1] is an emerging data-rich field that is capable of rapid mapping of the full band structure of materials by simultaneous measurements of the momentum components (k_x, k_y) of photoemitted electrons parallel to the specimen surface as a function of emission energy (E). In addition, information on the spin polarization (S) of electronic bands is obtainable with spinresolving detectors [1], while

the time (t)-dependence of nonequilibrium electronic properties can be studied in time-resolved measurements, when an ultrafast laser system is used to trigger electronic dynamics, thereby gaining access to the excited state band structure[2]. The measurements generate multidimensional datasets with dispersive structures in the (k_x, k_y, E, S) coordinates and location dependent changes along the time axis. Relating such datasets to physical quantities like the electron selfenergy requires detailed comparison with theoretical calculations in multiple parametric dimensions, but, currently, an efficient method to extract the global band structure parameters (including the position, intensity, linewidth, etc.) is lacking. We propose a data processing pipeline incorporating optimization methods and computer vision to achieve these goals. Building on this, we call for the creation of a data format for representing and storing experimental band structure mapping data to aid in materials design.

[1] Schönhense et al. New J. Phys., accepted (2017); Chernov et al. Ultramicroscopy 159, 453 (2015); Medjanik et al. Nat. Mater. 16, 615 (2017); Kutnyakhov et al. Sci. Rep. 6, 29394 (2016).

[2] Schmitt et al. Science 321, 1649 (2008); Bertoni et al. Phys. Rev. Lett. 117, 277201 (2016); Puppin et al, in preparation.

Automatic feature extraction from multi-dimensional STEM data

Christoph Freysoldt
Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf

Scanning transmission electron microscopy (STEM) provides valuable insights into the microstructure of complex structural materials such as modern steels. Recent advances in instrumentation do not only allow to image materials with atomic spatial resolution quickly enough to follow their dynamic evolution, but also to simultaneously record spectral data such as energy loss at each pixel. The impressive size (100-1000 GB per experiment) of the resulting four-dimensional data sets (4D = x, y, energy loss, time) call for an automatized extraction of the relevant features (grain boundaries, stacking faults, dislocations, and diffusion events). I will demonstrate how method development guided by physical understanding can lead to very efficient algorithms. Along these lines, I propose a strategy to approach big data analytics of 4D STEM data sets.

Big datasets in time-resolved and multidimensional STEM

Christian Liebscher
Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf

The emerging field of Scanning Transmission Electron Microscopy (STEM) offers versatile spectroscopic techniques to study materials at the atomic scale in terms of their structure, chemistry and bonding. When the electron beam is scanned over the sample, multiple spectra can be acquired at each probe position. The resulting multidimensional datasets (4D, 5D) are enormously rich in information, but drastically expand in data size (up to TB) and

complexity. The handling and analysis of these gargantuan datasets is one of the major challenges in the field of electron microscopy for the next century.
In the current presentation, we are illustrating the current state of the art in data generation in STEM, as well as recent imaging and spectroscopic techniques. A novel 4D-STEM method is presented, were in each electron probe position a full electron beam diffraction pattern is acquired. We also highlight approaches for employing STEM to investigate dynamic events in a material at the atomic scale under the stimulus of the electron beam. A first preliminary extension is presented, where time-resolved imaging is coupled with energy-dispersive X-ray spectroscopy. In the remainder of the presentation, the novel concepts for the acquisition of time-resolved and multidimensional STEM datasets will be introduced and how they will be coupled to in-situ experiments for observing the dynamic evolution of microstructure.

Machine-learning approaches to reveal the position and identity of each atom in a material

Shyam Katnagallu, Isabelle Mouton, Leigh Stephenson, Baptiste Gault, and Dierk Raabe
Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf

Field ion microscopy (FIM) achieved the first atomic-scale surface imaging in 1955. From FIM developed the first atom probe tomography (APT) instruments and, with modern machines, researchers can easily reconstruct relatively large volumes with sub-nanometre resolution, chemically identifying each atom. Both techniques are identical in that they employ a high electric field applied to a needle-like specimen ($r < 100$ nm), the differences being that FIM uses an imaging gas which is ionised and projected from nearby the specimen surface, while APT operates in an ultrahigh vacuum with the field sufficiently high to ionise and "evaporate" the surface atoms of the material itself. Reconstructing 3D volumes from the impacts of these field evaporated atoms is tricky and involves assumptions involving surface geometry and ion flight paths. A brief overview of how these assumptions limit the technique is given.
We have operated the Cameca LEAP 5000XS at MPIE Düsseldorf simultaneously in both FIM- and APT-modes. It was not designed for this, and doing so raises many technical issues, but preliminary data promises the ability to combine the position-resolving power of FIM with the time-of-flight mass spectrometry of APT. This potentially could have a massive impact in the material science community which will only be felt once we can handle and interpret the massive datasets correctly; simulation, machine learning techniques and basic data analysis will all need to be developed for this application.
We demonstrate the data challenges presented for our early experiments upon pure Al, pure W and simple alloys. We discuss these early results, both in the context of materials characterisation and how we can improve the data through experiment modification.

Towards an accurate, high-throughput framework for the prediction of anharmonic free energies in molecular crystals

Mariana Rossi
Fritz Haber Institute of the Max Planck Society, Berlin

Molecular crystals present multiple competing polymorphs having similar binding energies. Since these polymorphs exhibit different physicochemical properties, practical aspects related to their manufacturing, storage and usage can be largely affected by the delicate energetic balance between different forms. In our previous work [1] we have established that mainly anharmonic contributions, but also nuclear quantum effects, are necessary to compute accurate free energy differences in molecular crystals, making these contributions necessary to obtain predictive results. However, as it is rather computationally expensive to obtain accurate free energies for these crystals, in this project we intend to exploit different methods with which anharmonic effects can be included in the calculation of free energies, in order to obtain a framework that can be streamlined and used for high-throughput predictions. In particular we wish to employ the method of (ab initio) self-consistent phonons including lattice expansions in our simulations. Being successful at this step, and having a good database of accurate anharmonic free energies, machine learning methods could be used to estimate these contributions.
Finally in order to characterize vibrational fingerprints of these crystals we will employ kernel-ridge regression methods in order to learn polarizability tensors calculated from density-functional perturbation theory and obtain fully anharmonic vibrational Raman spectra at very low computational cost.

[1] Rossi, M.; Gasparotto, P.; Ceriotti, M. Phys. Rev. Lett. 117, 115702 (2016).

List of Posters

Presenter	Affiliation	Title
Christopher Bentley	MPI for the Physics of Complex Systems, Dresden	Investigating complex spectra and optimizing atomic setups
Arghya Dutta	MPI for Polymer Research, Mainz	High-throughput coarse-grained simulations of drug-membrane permeability
Benedikt Hoock	Humboldt-Universität zu Berlin	Compressed-sensing-based feature selection strategies in materials science: Defining the "best model"
Janis Kalofolias	Saarland University, Saarbrücken	Mining representative subgroups
Shyam Katnagallu	MPI für Eisenforschung GmbH, Düsseldorf / Poster together with Max Planck Computing and Data facility, Garching	Automated calibration of atom probe reconstructions using GPU based fourier analysis in a python workbench
Martin Kuban	Humboldt-Universität zu Berlin	Electronic structure fingerprints as similarity measures
Panagiotis Mandros	MPI for Informatics, Saarbrücken	Discovering reliable approximate functional dependencies
Alexander Marx	MPI for Informatics, Saarbrücken	Towards inferring causes for material properties
Aliaksei Mazheika	Fritz Haber Institute of the Max Planck Society, Berlin	Compressed-sensing analysis of CO_2 activation on oxide catalysts
Lukas Morand	Fraunhofer Institute for Mechanics of Materials IWM, Freiburg	Application of machine learning to forming processes: Microstructure-property relations and process optimization
Benjamin Regler	Fritz Haber Institute of the Max Planck Society, Berlin	Identifying actuators in materials science with information theory
Antonio Sanna	MPI of Microstructure Physics	Superconductivity, doped insulators and descriptors
Maria Troppenz	Humboldt-Universität zu Berlin	Cluster expansions with CELL: applications to simple and complex
Ye Wei	MPI für Eisenforschung GmbH, Düsseldorf	Machine learning crystal orientations from atom probe tomography datasets
Siyuan Zhang	MPI für Eisenforschung GmbH, Düsseldorf	Supervision on multi-dimensional data from electron microscopy

Name	Institute
Ahmetcik, Emre	FHI
Ashton, Michael	MPIE
Baghery, Mehrdad	MPIPKS
Bauer, Stefan	MPIIS
Benner, Peter	MPIDCTS
Bentley, Christopher	MPIPKS
Bereau, Tristan	MPIP
Breen, Andrew	MPIE
Carbogno, Christian	FHI
Csányi, Gábor	Univ. of Cambridge, UK
Curtarolo, Stefano	Duke Univ., USA
De Falco, Paolino	MPIKGF
Di Bernardo, Guiseppe	MPCDF
Draxl, Claudia	HU Berlin
Dutta, Arghya	MPIP
Eisfeld, Alexander	MPIPKS
Ernstorfer, Ralph	FHI
Fratzl, Peter	MPIKGF
Freysoldt, Christoph	MPIE
Gault, Baptiste	MPIE
Ghiringhelli, Luca	FHI
Goyal, Pawan	MPIDCTS
Heinzel, Stefan	MPCDF
Hoock, Benedikt	HU Berlin
Janßen, Jan	MPIE
Kalofolias, Janis	MPIINF
Katnagallu, Shyam S.	MPIE
Kour, Kirandeep	MPIDCTS
Krauter, Hanna	FHI
Kremer, Kurt	MPIP
Kuban, Martin	HU Berlin
Kumar Giri, Sajal	MPIPKS
Lehner, Anna	KIT;IWM
Liebscher, Christian	MPIE
Mandros, Panagiotis	MPIINF
Marx, Alexander	MPIINF
Mazheika, Aliaksei	FHI
Morand, Lukas	KIT;IWM
Mouton, Isabelle	MPIE
Neugebauer, Jörg	MPIE
Ozolins, Vidvuds	Yale Univ., USA
Peng, Zirong	MPIE
Rampp, Markus	MPCDF
Regler, Benjamin	FHI
Rigamonti, Santiago	HU Berlin

Rossi, Mariana	FHI
Rost, Jan-Michael	MPIPKS
Rupp, Matthias	FHI
Saalmann, Ulf	MPIPKS
Sanna, Antonio	MPIMSP
Scheffler, Matthias	FHI
Scheidgen, Markus	HU Berlin
Stephenson, Leigh	MPIE
Troppenz, Maria	HU Berlin
Vreeken, Jilles	MPIINF
Wei, Ye	MPIE
Xian, Patrick	FHI
Zhang, Siyuan	MPIE
Ziletti, Angelo	FHI