

BiGmax Summer School 2021

Monday, 13 September 2021 - Friday, 17 September 2021

Virtual event

[https:](https://indico.mpi-magdeburg.mpg.de/e/bigmax_summerschool2021)

[//indico.mpi-magdeburg.mpg.de/e/bigmax_summerschool2021](https://indico.mpi-magdeburg.mpg.de/e/bigmax_summerschool2021)

Book of Abstracts



Program Overview

Monday, 13.9.: Machine Learning - Theory

[Zoom link](#), Meeting ID: 834 2962 3043, Passcode: 159390

14:00-14:15	Introduction	Christian Liebscher	
14:15-15:15	Lecture	Jörg Behler	Four Generations of Neural Network Potentials
15:15-16:15	Lecture	Abbas Ourmazd	Science in the Age of Machine Learning
16:15-17:00	Break	gather.town	
17:00-18:00	Tutorial	Milica Todorović	Active Learning with Bayesian Optimization
18:00-19:00	Tutorial	Luca Ghiringhelli	Compressed sensing meets symbolic regression

Tuesday, 14.9.: Machine Learning - Theory & Application

[Zoom link](#), Meeting ID: 839 8830 1195, Passcode: 323414

14:00-15:00	Tutorial	Ye Wei	Neural Networks
15:00-16:00	Lecture	Baptiste Gault	Machine-learning aided atom probe tomography
16:00-16:45	Break	gather.town	
16:45-17:45	Lecture	Sergei Kalinin	Machine Learning for Electron Microscopy
19:00-21:00	Posters	gather.town	

Wednesday, 15.9.: Learning from Complex Data

[Zoom link](#), Meeting ID: 835 1563 1096, Passcode: 220754

14:00-15:00	Tutorial	Niels Cautaerts	Tips, tricks, and tools for reproducible materials science
15:00-16:00	Lecture	Tess Smidt	Neural Networks with Euclidean Symmetry for Physical Sciences
16:00-16:45	Break	gather.town	
16:45-17:45	Tutorial	Tess Smidt	Eucleidian Neural Networks Hands On
17:45-18:45	Lecture	Ankit Agrawal	Artificial Intelligence and High-Performance Data Mining

Thursday, 15.9.: Multidimensional Data Analysis

[Zoom link](#), Meeting ID: 834 2603 9098, Passcode: 437058

14:00-15:00	Lecture	Gabriele Steidl	Variational methods in material sciences
15:00-16:00	Lecture	Francisco De La Pena	HyperSpy: theory and applications
16:00-16:45	Break	gather.town	
16:45-17:45	Tutorial	Francisco De La Pena	Tutorial: EELS analysis with HyperSpy

Friday, 17.9.: Open Data

[Zoom link](#), Meeting ID: 868 1635 2551, Passcode: 029857

14:00-15:00	Lecture	Nicole Jung	NFDI4Chem
15:00-16:00	Lecture	Christoph Koch	Structuring, analyzing, and harvesting big data in materials science electron microscopy
16:00-16:30	Break	gather.town	
16:30-17:30	Tutorial	Markus Scheidgen	Sharing, publishing, and managing computational materials science data with NOMAD

Contents

Lecture "Four Generations of Neural Network Potentials"	1
Lecture "Science in the Age of Machine Learning"	1
Tutorial "Active Learning with Bayesian Optimization"	2
Tutorial "Compressed sensing meets symbolic regression: learning interpretable models"	2
Tutorial "Neural Networks"	2
Lecture "Machine-learning aided atom probe tomography: status and (possible) directions"	3
Lecture "Machine Learning for Electron Microscopy: from Imaging to Atomic Fabrication"	3
Learning Dynamics of STEM by Enforcing Physical Consistency with Phase-Field Models	4
Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning	4
A materials informatics framework to discover patterns in atom probe tomography data.	5
Multidimensional Photoemission Spectroscopy: proposal & demonstration of data (infra)structure	5
Deep generative models for the design of Dual phase steel microstructures	6
Teaching solid mechanics to artificial intelligence	7
Tutorial "Tips, tricks, and tools for reproducible materials science"	7
Lecture "Neural Networks with Euclidean Symmetry for Physical Sciences"	8
Tutorial "Neural Networks with Euclidean Symmetry for Physical Sciences"	8
Lecture "Artificial Intelligence and High-Performance Data Mining for Accelerating Scientific Discovery"	9
Lecture "Variational methods in material sciences"	9
Lecture "Introduction to EELS analysis with HyperSpy"	10

Tutorial "Introduction to EELS analysis with HyperSpy"	10
Lecture "NFDI4Chem"	10
Lecture "Structuring, analyzing, and harvesting big data in materials science electron microscopy"	11
Tutorial "Sharing, publishing, and managing computational materials science data with NOMAD"	11

Machine Learning - Theory**Lecture ”Four Generations of Neural Network Potentials”****Author:** Jörg Behler¹¹ *University of Göttingen*

A lot of progress has been made in recent years in the development of machine learning (ML) potentials for atomistic simulations [1]. Neural network potentials (NNPs), which have been introduced more than two decades ago [2], are an important class of ML potentials. While the first generation of NNPs has been restricted to small molecules with only a few degrees of freedom, the second generation extended the applicability of ML potentials to high-dimensional systems containing thousands of atoms by constructing the total energy as a sum of environment-dependent atomic energies [3]. Long-range electrostatic interactions can be included in third-generation NNPs employing environment-dependent charges [4], but only recently limitations of this locality approximation could be overcome by the introduction of fourth-generation ML potentials [5], which are able to describe non-local charge transfer using a global charge equilibration step. In this talk an overview about the evolution of NNPs will be given along with typical applications in large-scale atomistic simulations.

[1] J. Behler, *J. Chem. Phys.* 145 (2016) 170901.

[2] T. B. Blank, S. D. Brown, A. W. Calhoun, and D. J. Doren, *J. Chem. Phys.* 103 (1995) 4129.

[3] J. Behler and M. Parrinello, *Phys. Rev. Lett.* 98 (2007) 146401.

[4] N. Artrith, T. Morawietz, J. Behler, *Phys. Rev. B* 83 (2011) 153101.

[5] T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, *Nature Comm.* 12 (2021) 398.

Machine Learning - Theory**Lecture ”Science in the Age of Machine Learning”****Author:** Abbas Ourmazd¹¹ *University of Wisconsin Milwaukee***Corresponding Author:** ourmazd@uwm.edu

Traditionally the “best” observations are those with the largest signal from the most tightly controlled systems. In a wide range of phenomena – the dance of proteins in function, femtosecond breaking of molecular bonds, the gestation of fetuses – tight control is neither feasible, nor desirable. Modern machine-learning techniques extract far more information from sparse random sightings than usually obtained from set-piece experiments. I will describe on-going efforts to extract structural and dynamical information from noisy, random snapshots recorded with very poor, or non-existent timing information. Examples will include functional motions of molecular machines, and ultrafast dynamics of photo-induced reactions.

Machine Learning - Theory**Tutorial ”Active Learning with Bayesian Optimization”****Author:** Milica Todorović¹¹ *Turku University*

Gaussian process regression (GPR) is a kernel-based regression tool with intrinsic uncertainty estimation, which makes it well-suited to natural science datasets. In Bayesian optimization, GPR is coupled with acquisition functions for an active learning approach, where models are iteratively refined by addition of new data points with high information content. This tutorial will use the BOSS code to demonstrate the basic principles of Bayesian optimization and how it can be applied in N-dimensional atomistic structures search.

Machine Learning - Theory**Tutorial ”Compressed sensing meets symbolic regression: learning interpretable models”****Author:** Luca M. Ghiringhelli¹**Co-author:** Luigi Sbailò²¹ *Fritz Haber Institute of the Max Planck Society*² *NOMAD Lab at FHI*

In this tutorial, we introduce the AI technique of symbolic regression, combined with compressed sensing for the identification of compact, interpretable models. Specifically, we introduce the Sure-Independence Screening and Sparsifying Operator (SISSO), together with its recent variants.

The methodology starts from a set of candidate features, provided by the user, and it builds a tree of possible mathematical expression, involving linear and nonlinear operators, up to a given complexity. A compressed sensing solver finds, among billions or trillions of candidate expressions those that better explain the training data.

We will show demonstrative applications to materials science, including prediction of perovskite-materials stability and topological-insulators identification.

Machine Learning - Theory & Application**Tutorial ”Neural Networks”****Author:** Ye Wei¹¹ *Max-Planck-Institut für Eisenforschung GmbH*

In this tutorial, we discuss a neural network application in materials discovery. More specifically, we will showcase how to accelerate functional high entropy alloy discovery using neural network based generative model and ensemble model for the regression task. Therefore, this tutorial consists of two parts. Firstly, we discuss in detail how to construct an alloy generation scheme based on generative neural network model, attribute classifier, stochastic sampling, and density estimation model. Secondly, we demonstrate a systematic approach to combine Bayesian optimization, neural network and gradient boosting decision tree to achieve a two-stage Ensemble Regression Model (TERM). The first stage concerns composition-based regression models, aiming at fast and large-scale composition inference. The top results from

the first-stage model enter a more refined model, where density functional calculations and thermodynamic calculations are included as part of the model input. Finally, the TERM outputs were evaluated based on a rank-based policy.

Machine Learning - Theory & Application

Lecture "Machine-learning aided atom probe tomography: status and (possible) directions"

Authors: Alaukik Saxena¹; Baptiste Gault¹; Christoph Freysoldt¹; Ye Wei¹; Yue Li¹

¹ *Max-Planck-Institut für Eisenforschung GmbH*

Corresponding Author: b.gault@mpie.de

Atom probe tomography (APT) is a materials analysis technique that provides sub-nanometer resolution compositional mapping. The data is in the form of a point cloud containing often millions of atoms, and to each of these points is associated an elemental nature. By interrogating the point cloud, the local composition of a material or a phase of a specific microstructural feature can be reported. APT is often referred to as "data-intensive" technique, and has long made use of many clustering-type techniques (DBSCAN, NN etc.) to facilitate data extraction, which are all now often classified as belonging to machine-learning.

In this presentation, I will review some of the recent developments from MPIE in the application of machine-learning techniques to atom probe analysis workflows – i.e. beyond just extraction of data from the point cloud – targeting faster and more efficient, reliable and reproducible data analysis.

Machine Learning - Theory & Application

Lecture "Machine Learning for Electron Microscopy: from Imaging to Atomic Fabrication"

Author: Sergei Kalinin¹

¹ *Oak Ridge National Laboratory*

I will discuss recent progress in automated experiment in electron microscopy, ranging from feature to physics discovery via active learning. The applications of classical deep learning methods in streaming image analysis are strongly affected by the out of distribution drift effects, and the approaches to minimize though are discussed. We further present invariant variational autoencoders as a method to disentangle affine distortions and rotational degrees of freedom from other latent variables in imaging and spectral data and decode physical mechanisms. Extension of encoder approach towards establishing structure-property relationships will be illustrated on the example of plasmonic structures. Finally, I illustrate transition from post-experiment data analysis to active learning process. Here, the strategies based on simple Gaussian Processes often tend to produce sub-optimal results due to the lack of prior knowledge and very simplified (via learned kernel function) representation of the system. Comparatively, deep kernel learning (DKL) methods allow to realize both the exploration of complex systems towards the discovery of structure-property relationship, and enable automated experiment workflows targeting physics (rather than simple spatial feature) discovery. The latter is illustrated via experimental discovery of the edge plasmons in STEM/EELS in MnPS₃, a lesser-known 2D material.

This research is supported by the by the U.S. Department of Energy, Basic Energy Sciences, Materials Sciences and Engineering Division and the Center for Nanophase Materials Sciences, which is sponsored at Oak Ridge National Laboratory by the Scientific User Facilities Division, BES DOE.

Poster Session: Machine Learning -Theory & Applications / Open Data

Learning Dynamics of STEM by Enforcing Physical Consistency with Phase-Field Models

Authors: Christian Liebscher¹; Christoph Freysoldt¹; Jaber Mianroodi¹; Lekshmi Sreekala²; Ning Wang¹; Pawan Goyal³; Peter Benner³; Sandeep Reddy Bukka³

¹ *Max-Planck-Institut für Eisenforschung GmbH*

² *Max-Planck-Institut für Eisenforschung GmbH*

³ *Max Planck Institute for Dynamics of Complex Technical Systems*

Corresponding Authors: l.sreekala@mpie.de, goyalp@mpi-magdeburg.mpg.de

In this poster, we present our research goals of a recently BiGmax funded project towards learning dynamics of scanning transmission electron microscopy (STEM) by incorporating physical consistency with phase-field models. The primary idea of this project is to develop machine learning (ML)-based modeling of an interpretable coarse-grained dynamic model utilizing in situ STEM video sequences fulfilling a suitable dynamical phase-field equation. The modeling approach aims to discover governing equations by utilizing the video sequence data and prior physics knowledge that is directly compatible with analytic theories or subsequent ML-based analysis.

Poster Session: Machine Learning -Theory & Applications / Open Data

Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning

Author: Andreas Leitherer¹

Co-authors: Angelo Ziletti¹; Luca M. Ghiringhelli¹

¹ *Fritz Haber Institute of the Max Planck Society*

Corresponding Author: leitherer@fhi-berlin.mpg.de

Due to their ability to recognize complex patterns, neural networks can drive a paradigm shift in the analysis of materials science data. Here, we introduce ARISE, a crystal-structure identification method based on Bayesian deep learning. As a major step forward, ARISE is robust to structural noise and can treat more than 100 crystal structures, a number that can be extended on demand. While being trained on ideal structures only, ARISE correctly characterizes strongly perturbed single- and polycrystalline systems, from both synthetic and experimental resources. The probabilistic nature of the Bayesian-deep-learning model allows to obtain principled uncertainty estimates, which are found to be correlated with crystalline order of metallic nanoparticles in electron-tomography experiments. Applying unsupervised learning to the internal neural-network representations reveals grain boundaries and (unapparent) structural regions sharing easily interpretable geometrical properties. This work enables the hitherto hindered analysis of noisy atomic structural data from computations or experiments.

Poster Session: Machine Learning -Theory & Applications / Open Data**A materials informatics framework to discover patterns in atom probe tomography data.****Author:** Alaukik Saxena¹**Co-authors:** Nikita Polin ²; Dierk Raabe ²; Benjamin Berkels ³; Baptiste Gault ²; Christoph Freysoldt ²¹ *Max-Planck-Institut für Eisenforschung GmbH (Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE))*² *Max-Planck-Institut für Eisenforschung GmbH*³ *RWTH Aachen University***Corresponding Author:** saxenaalaukik93@gmail.com

To quantify chemical segregation at multiple length scales in APT in a semi-automatic way, we propose a multi-stage strategy. First, we collect composition statistics from APT datasets for 2x2x2 nm voxels. These voxel compositions are then clustered in compositional space using Gaussian mixture models to automatically identify key phases. Next, based on this compositional classification we employ DBSCAN in physical space at voxel resolution to detect individual precipitates at a small fraction of the effort needed for single-atom-based algorithms. This framework was used to identify and disentangle plate-like Zr-rich and topologically complex Cu-rich precipitates in two APT datasets from Fe-doped Sm-Co magnets, each containing approximately 500 million ions. Upon segmentation of each precipitate using DBSCAN a new approach based on principle component analysis (PCA) is applied to 2D slices of the complex Cu-rich precipitates to further decompose them into approximate planar regions along with their junctions. For each precipitate, the actual distribution of atomic fractions is compared to the expected distribution of a random alloy. This step helps to quantitatively assess clustering within a given precipitate for each atomic species. Finally, for each quasi-planar precipitate, a triangular grid is superimposed to investigate in-plane compositional fluctuations, thickness, and 1D composition profiles.

Poster Session: Machine Learning -Theory & Applications / Open Data**Multidimensional Photoemission Spectroscopy: proposal & demonstration of data (infra)structure****Authors:** Tommaso Pincelli¹; Steinn Ymir Agustsson²; Rui Patrick Xian¹; Dmytro Kutnyakhov³; Yves Acremann⁴; Michael Hartelt⁵; Maciej Dendzik⁶; Shuo Dong¹; Samuel Beaulieu⁷; Julian Maklar¹; Markus Scheidgen⁸; Laurenz Rettig¹; Ralph Ernstorfer¹¹ *Fritz Haber Institute of the Max Planck Society*² *JGU Mainz*³ *DESY Photon Science, 22607, Hamburg, Germany*⁴ *Laboratory for Solid State Physics, ETH Zurich, 8093, Zurich, Switzerland*⁵ *Technical University Kaiserslautern*⁶ *KTH Royal Institute of Technology, Stockholm*⁷ *CELIA, CNRS-Université de Bordeaux*⁸ *Humboldt-Universität zu Berlin***Corresponding Authors:** pincelli@fhi-berlin.mpg.de, steinny@gmail.com

The complexity of photoemission data is rapidly increasing, as new technological breakthroughs have enabled multidimensional parallel acquisition of

multiple observables. Most of the community is currently using heterogeneous data formats and workflows.

We propose a new data format based on NeXus, a hierarchically organized hdf5 structure. The aim is to immediately enable preprocessed data and metadata shareability according to FAIR principles, employing the existing storage and archiving infrastructure such as Zenodo, OpenAIRE and Nomad/FAIRmat. Ultimately, the multidimensional photoemission spectroscopy (MPES) format is designed to allow high-performance automated access, providing experimental databases for high-throughput material search.

The MPES format is based on the creation of a standardized set of classes that univocally identify experimental observables and metadata. Such dictionary is complemented by “application definitions”, i.e. ontologies that constrain the existence of specific elements in a file.

Our approach involves reaching out to the community using a [website](#) with a wiki structure. By this medium, we wish to favour acceptance, to avoid conflict with different low-level preprocessing workflows, and to create continuously updated documentation.

As a demonstrator of the potential of our approach, we present the workflow we developed for our data pipeline, originating from time-resolved angularly-resolved photoemission spectroscopy.

Poster Session: Machine Learning -Theory & Applications / Open Data

Deep generative models for the design of Dual phase steel microstructures

Author: Navyanth Kusampudi¹

¹ *Max-Planck-Institut für Eisenforschung GmbH*

Corresponding Author: n.kusampudi@mpie.de

Dual Phase (DP) steels are an important family of steel grades used widely in the automotive industry because of their beneficial properties such as high ultimate tensile strength (GPa range), low initial yield stress, and high early-stage strain hardening. The DP steel microstructure consists of soft ferritic grains, which are mainly responsible for ductility, and hard martensitic zones, which give these steels their strength. However, due to the huge parameter space, establishing the microstructure-property relationship in DP steels is a combinatorial challenge. To tackle this challenge we use data from crystal plasticity simulations and train machine learning models to automatically extract the patterns from DP steel microstructures without any need of hand-designed features. In particular, I will present my work on using Variational Auto Encoders (VAEs, a generative machine learning model) to learn the low dimensional latent representations and to generate synthetic EBSD data for DP steels. Also, I will present, that the learned latent space can serve as a design space for microstructure design.

Poster Session: Machine Learning -Theory & Applications / Open Data**Teaching solid mechanics to artificial intelligence****Authors:** Jaber Mianroodi¹; Nima H. Siboni¹; Dierk Raabe¹¹ *Max-Planck-Institut für Eisenforschung GmbH***Corresponding Author:** j.mianroodi@mpie.de

I will present our latest progress in using machine learning for solving non-linear solid mechanics. The presentation will be based on our published work here: <https://www.nature.com/articles/s41524-021-00571-z>

Learning from Complex Data**Tutorial ”Tips, tricks, and tools for reproducible materials science”****Author:** Niels Cauttaerts¹¹ *Max-Planck-Institut für Eisenforschung GmbH***Corresponding Author:** n.cauttaerts@mpie.de

The ability to replicate results is a key characteristic of quality science, and is growing ever more important in light of the replication crisis [1, 2].

A study can rarely be repeated using only the minimalistic descriptions provided in the “Materials and methods” section in a paper.

It is therefore important to properly document the entire knowledge generation pipeline in such a way that it could be repeated by anyone with minimal effort.

In this tutorial we will look at the tools and techniques by which we can improve the reproducibility of an experimental study through a single comprehensive example from electron microscopy.

The session will cover:

- [eLabFTW](#), an electronic lab notebook application with a python API, which can be used to document samples and experiments.
- Various strategies of working with experimental metadata, including the HDF5 file format and JSON.
- Strategies and best practices for reproducible data analysis pipelines in jupyter notebooks, using git, conda-forge, mybinder and docker.

[1] Ioannidis JPA, Why most published research findings are false. *PLoS Med* 2(8): e124 (2005).

[2] Schooler, J., Metascience could rescue the ‘replication crisis’. *Nature* 515, 9 (2014).

Learning from Complex Data

Lecture "Neural Networks with Euclidean Symmetry for Physical Sciences"

Authors: Tess Smidt¹; Mario Geiger²

¹ *Berkeley Lab*

² *EPFL*

Atomic systems (molecules, crystals, proteins, etc.) are naturally represented by a set of coordinates in 3D space labeled by atom type. This is a challenging representation to use for machine learning because the coordinates are sensitive to 3D rotations and translations and there is no canonical orientation or position for these systems. One motivation for incorporating symmetry into machine learning models on 3D data is to eliminate the need for data augmentation — the 500-fold increase in brute-force training necessary for a model to learn 3D patterns in arbitrary orientations.

Most symmetry-aware machine learning models in the physical sciences avoid augmentation through invariance, throwing away coordinate systems altogether. But this comes at a price; many of the rich consequences of Euclidean symmetry are lost: geometric tensors, point and space groups, degeneracy, atomic orbitals, etc.

We present a general neural network architecture that faithfully treats the equivariance of physical systems, naturally handles 3D geometry, and operates on the scalar, vector, and tensor fields that characterize them. We describe how the networks achieves equivariance, demonstrate the capabilities of our network using simple tasks, and provide coding examples to build these models using e3nn: a modular framework for Euclidean Neural Networks (<https://e3nn.org>).

Learning from Complex Data

Tutorial "Neural Networks with Euclidean Symmetry for Physical Sciences"

Authors: Mario Geiger¹; Tess Smidt²

¹ *EPFL*

² *Berkeley Lab*

Atomic systems (molecules, crystals, proteins, etc.) are naturally represented by a set of coordinates in 3D space labeled by atom type. This is a challenging representation to use for machine learning because the coordinates are sensitive to 3D rotations and translations and there is no canonical orientation or position for these systems. One motivation for incorporating symmetry into machine learning models on 3D data is to eliminate the need for data augmentation — the 500-fold increase in brute-force training necessary for a model to learn 3D patterns in arbitrary orientations.

Most symmetry-aware machine learning models in the physical sciences avoid augmentation through invariance, throwing away coordinate systems altogether. But this comes at a price; many of the rich consequences of Euclidean symmetry are lost: geometric tensors, point and space groups, degeneracy, atomic orbitals, etc.

We present a general neural network architecture that faithfully treats the equivariance of physical systems, naturally handles 3D geometry, and operates on the scalar, vector, and tensor fields that characterize them. We describe how the networks achieves equivariance, demonstrate the capabilities of our network using simple tasks, and provide coding examples

to build these models using e3nn: a modular framework for Euclidean Neural Networks (<https://e3nn.org>).

Learning from Complex Data

Lecture "Artificial Intelligence and High-Performance Data Mining for Accelerating Scientific Discovery"

Author: Ankit Agrawal¹

¹ *Northwestern University*

The increasing availability of data from the first three paradigms of science (experiments, theory, and simulations), along with advances in artificial intelligence and machine learning (AI/ML) techniques has offered unprecedented opportunities for data-driven science and discovery, which is the fourth paradigm of science. Within the arena of AI/ML, deep learning (DL) has emerged as a game-changing technique in recent years with its ability to effectively work on raw big data, bypassing the (otherwise crucial) manual feature engineering step traditionally required for building accurate ML models, thus enabling numerous real-world applications, such as autonomous driving. In this talk, I will present our ongoing research in AI and high performance data mining, along with illustrative real-world scientific applications. In particular, we will discuss approaches to gainfully apply DL on big data (by accelerating DL and enabling deeper learning) as well as on small data (deep transfer learning) in the context of materials science. I will also demonstrate some of the software tools developed in our group.

Multidimensional Data Analysis

Lecture "Variational methods in material sciences"

Author: Gabriele Steidl¹

¹ *TU Berlin*

Corresponding Author: steidl@math.tu-berlin.de

Variational methods are powerful tools in image processing. Basically we are searching for a suitable mathematical model (function) consisting of a data term and a prior which minimizer provides a solution of the task at hand and can be computed in an efficient, reliable way. Typically this leads to non-smooth, high-dimensional optimization problems. This talk deals with recent results obtained by applying variational methods for different tasks in material sciences as

- crack detection using optical flow models in image sequences,
- determination of deformation fields in electron backscatter diffraction image sequences,
- superresolution of material images by learned patch-based priors,

and

- denoising of FIB images with directional total variation priors.

Multidimensional Data Analysis**Lecture "Introduction to EELS analysis with HyperSpy"****Author:** Francisco De La Peña¹¹ *Université de Lille*

HyperSpy (<https://hyperspy.org/>) is an open-source Python package for the analysis of multi-dimensional datasets. In its fourteen years of existence, its community of developers has taken it from being a simple collection of scripts for electron energy-loss data analysis to become the core of a multidisciplinary software ecosystem. In this talk, I will its evolution, ecosystem, main features and structure—all illustrated with a bit of theory and applications. I will also briefly describe how to contribute to HyperSpy and how to create your own HyperSpy extension.

Multidimensional Data Analysis**Tutorial "Introduction to EELS analysis with HyperSpy"****Author:** Francisco De La Peña¹¹ *Université de Lille*

HyperSpy (<https://hyperspy.org/>) is an open-source Python package for the analysis of multi-dimensional datasets. In its fourteen years of existence, its community of developers has taken it from being a simple collection of scripts for electron energy-loss data analysis to become the core of a multidisciplinary software ecosystem. In this talk, I will its evolution, ecosystem, main features and structure—all illustrated with a bit of theory and applications. I will also briefly describe how to contribute to HyperSpy and how to create your own HyperSpy extension.

Open Data**Lecture "NFDI4Chem"****Author:** Nicole Jung¹¹ *Karlsruhe Institute of Technology*

Open Data

Lecture ”Structuring, analyzing, and harvesting big data in materials science electron microscopy”

Authors: Christoph Koch¹; Markus Kühbach²; Sherjeel Shabih¹

¹ *Humboldt-Universität zu Berlin*

² *Max-Planck-Institut für Eisenforschung GmbH*

Corresponding Author: christoph.koch@hu-berlin.de

Every day, experimental materials science data is being collected in thousands of laboratories around the world. However, the diversity of instruments, vendor software packages and (proprietary) data formats, lab cultures, and the focus mostly on new discoveries causes most of this data to end up in a black hole in terms of accessibility to the scientific community (including in many cases the lab in which the data was acquired). Using the example of transmission electron microscopy we will present our attempts to encourage scientists to annotate and contribute more of their data to the scientific community and benefiting themselves in the process. We are striving to achieve this by being able to offer relevant online data processing and analyzing capabilities and tools to annotate data in more and more automated ways. As more and more data is being accumulated, big data techniques can be applied to benefit from the added value that sharing of experimental data sets produces.

Open Data

Tutorial ”Sharing, publishing, and managing computational materials science data with NOMAD”

Author: Markus Scheidgen¹

¹ *Humboldt Universität zu Berlin / Fritz Haber Institut der Max Planck Gesellschaft*

Corresponding Author: markus.scheidgen@physik.hu-berlin.de

The main focus of this tutorial is the FAIR sharing of materials science data and how to do it with NOMAD. We will be covering the publication of new data and the exploration and download from NOMAD’s existing data; both through our browser-based interface and APIs.