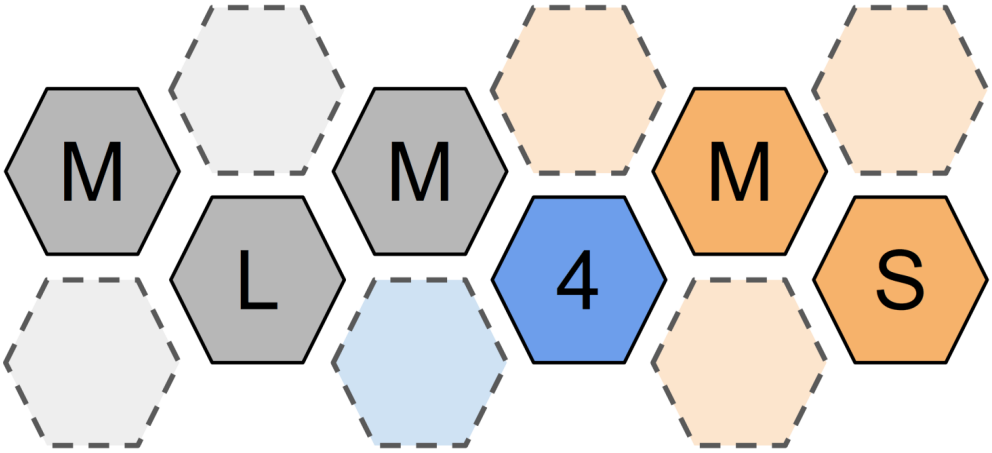


Machine Learning Modalities for Materials Science

13.05.2024 - 17.05.2024
Jožef Stefan Institute, Ljubljana Slovenia

Book of Abstracts

Contents

General

Organizers

Sašo Džeroski, Jožef Stefan Institute, Slovenia
Stefano de Gironcoli, Scuola Internazionale Superiore di Studi Avanzati, Italy
Patrick Rinke, Aalto University, Finland
Kevin Rossi, TU Delft, Netherlands
Sintija Stevanoska, Jožef Stefan Institute, Slovenia
Milica Todorović, University of Turku, Finland

The DAEMON COST Action CA22154 acts as a co-organizer of the event. COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation.

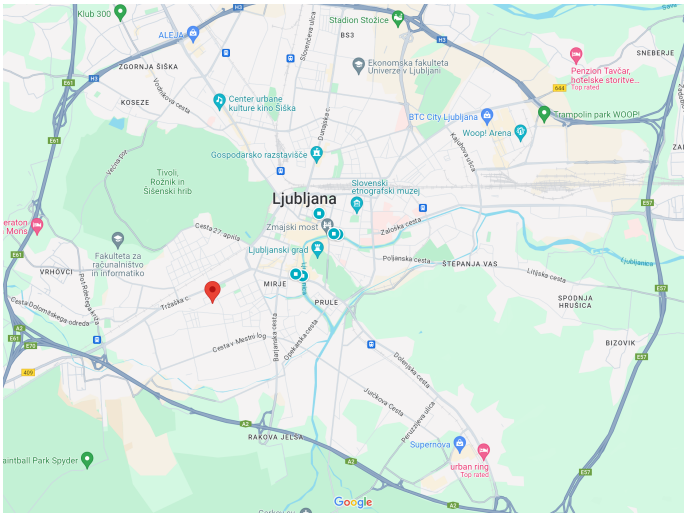
Funded by
the European Union

Contents

Venue

Main venue: Main Lecture Hall, Jožef Stefan Institute, Jamova cesta 39, Ljubljana, Slovenia (see also the map below).
The workshop location can be reached by:

- Walking from the center, which will take approximately 20 minutes
- By taking busses no. 1 (direction "Mestni Log", bus stop "Jadranska") or no. 6 (direction "Dolgi most", bus stop "Glince")



Poster sessions

There will be two poster sessions, held on Tuesday (14.05) and Wednesday (15.05), 17:40 - 20:00h. Both will be held in the Jožef Stefan Institute Main Building, Jamova cesta 39, Ljubljana, Slovenia.
The recommended poster size is A0 portrait.

Meals

Lunch will be provided free of charge on Tuesday, Wednesday, and Thursday.
Coffee breaks will also be provided to the participants.
Light finger food and beverages will be served during the poster sessions.
The social dinner will be held on Thursday evening at Union Brewery in Ljubljana, located on Celovska cesta 22, free of charge.

Internet Access

An EDUROAM account is required to access the internet through the ICTP network at the Main Venue. There is also a guest network available.

Software Requirements for Tutorials

A Google account is required to fully benefit from the Tutorial Lectures, as Google Colab will be mostly used as the tool to run calculations and examples. Python should also be installed on your machine.

Code of Conduct

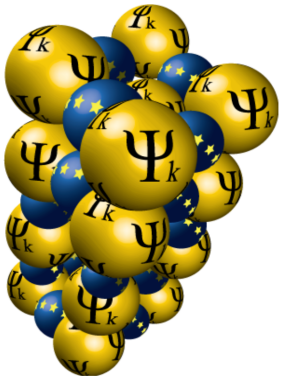
We strive for making every attendee feel welcome and respected. Do not hesitate to contact any of the organizers if you are the witness or the victim of any discrimination or harassment. Step up and speak out to stop any kind of inappropriate behaviour you witness. Bystander intervention creates safer communities and prevents harmful escalation.

Other

The Emergency number for Slovenia is 112.

Sponsors

We acknowledge the generous support of AI Journal, TAILOR, MAX, CECAM, PSI-K and SMASH.



Programme

	Monday (13.05)	Tuesday (14.05)	Wednesday (15.05)	Thursday (16.05)	Friday (17.05)
09:00		Kevin Jablonka 09:00 - 10:20	Christoph Koch 09:00 - 10:20	Andy Sode Anker 09:00 - 09:40	Teodoro Laino 09:00 - 09:40
09:20					
09:40				Javier Heras-Domingo 09:40 - 10:00	Lei Zhang 09:40 - 10:00
10:00				Nejc Hodnik 10:00 - 10:20	Matej Praprotnik 10:00 - 10:20
10:20		Coffee Break 10:20 - 11:00	Coffee Break 10:20 - 11:00	Panel Discussion + Coffee 10:20 - 11:20	Panel Discussion + Coffee 10:20 - 11:20
10:40					
11:00		Kevin Jablonka 11:00 - 12:20	Christoph Koch 11:00 - 12:20		
11:20				Vinko Sršan 11:20 - 11:40	Matilda Sipilä 11:20 - 11:40
11:40				Andrea Ruiz 11:40 - 12:00	Morgan Kerhouant 11:40 - 12:00
12:00	Registration 12:00 - 13:50			Helge Stein 12:00 - 12:40	Tian Xie 12:00 - 12:40
12:20			Closing: 10min		
12:40		Lunch 12:40 - 14:00	Lunch 12:40 - 14:00	Lunch 12:40 - 14:00	Closing: 10min
13:00					
13:20					
13:40	Opening: 10min				
14:00	Lilo Pozzo 14:00 - 15:20	Jörg Neugebauer 14:00 - 15:20	Opening: 20min	Dusan Strmcnik 14:00 - 14:20	
14:20			Ekin Dogus Cubuk 14:20 - 15:00	Saso Sturm 14:20 - 14:40	
14:40				Austin Zadoks 14:40 - 15:00	
15:00			Andreja Bencan 15:00 - 15:20	Franco Pellegrini 15:00 - 15:20	
15:20	Coffee Break 15:20 - 16:00	Coffee Break 15:20 - 16:00	Coffee Break 15:20 - 16:00	Coffee Break 15:20 - 16:00	
15:40					
16:00	Brenden Pelkie 16:00 - 17:20	Jörg Neugebauer 16:00 - 17:20	Martin Uhrin 16:00 - 16:20	Matthias Stosiek 16:00 - 16:20	
16:20			Christer Soderholm 16:20 - 16:40	Nataliya Lopanitsyna 16:20 - 16:40	
16:40			DAEMON 16:40 - 17:20	Emma King Smith 16:40 - 17:20	
17:00					

Monday

12:00 - 13:50	Registration
13:50 - 14:00	Opening Remarks
14:00 - 14:40	**Lilo Pozzo** Intro Lecture: Scientific advancements enabled with automated experimentation
14:40 - 15:20	**Brenden Pelkie** Intro Lecture: Open-source hardware solutions for automating experimental workflows
15:20 - 16:00	Coffe Break
16:00 - 17:20	**Lilo Pozzo & Brenden Pelkie** Tutorial Lecture: Autonomous experimentation platform demonstration showcasing the Jubilee lab automation platform.

Tutorial Lecture: Open-Science laboratory automation for AI-accelerated materials research and optimization

Lilo Pozzo[1], Brended Pelkie[1]

[1] *University of Washington, Seattle, United States of America*

Laboratory automation promises great benefits for materials research, in particular by enabling integration of machine learning (ML) and artificial intelligence (AI) experimental design strategies to guide the discovery of new materials. However, establishing new automation workflows for experiments can be challenging and requires researchers to develop solutions for many system-dependent requirements. Pursuing an Open-Science approach to automation affords researchers greater flexibility and creativity in meeting these requirements. This tutorial will present recent progress in laboratory automation in our group and provide guidance for researchers interested in working with automated experimentation by adopting Open-Science principles and community-driven research. We will share opportunities and challenges experienced by our group as we transitioned to embrace laboratory automation across projects. We have implemented automation workflows with many approaches, with a focus on practical, translatable, modular, scalable, and openly shared solutions. During this workshop, examples of implementations will be discussed, including the scientific advances that they have enabled. A practical introduction to some common automation practices will help new automation practitioners plan their first experiment. A demonstration of an Open-Hardware autonomous experimentation platform (Jubilee) will be given, showcasing the promise of flexible experimental automation integrated with machine learning decision making for materials optimization experiments. We will highlight opportunities for skillset development and community contributions toward the advancement of open laboratory automation.

Tuesday

9:00 - 10:20	**Kevin Jablonka** Intro Lecture: Large language models for materials science
10:20 - 11:00	Coffee Break
11:00 - 12:20	**Kevin Jablonka** Tutorial Lecture: Large language models for materials science
12:20 - 14:00	Lunch Break
14:00 - 15:20	**Jörg Neugebauer** Intro Lecture: Maximizing High-Throughput Discovery and Machine Learning Efficiency Through Computational Workflows
15:20 - 16:00	Coffee Break
16:00 - 17:20	**Jörg Neugebauer & Sarath Menon** Tutorial Lecture: Maximizing High-Throughput Discovery and Machine Learning Efficiency Through Computational Workflows
17:40 - 20:00	Poster session + Dinner

Large language models for materials science

Kevin Jablonka[1]

[1] *University of Jena, Jena, Germany*

In this session, we will discuss how large language models (e.g. decoder-only architectures like GPT) work and how we can apply them in materials science in chemistry. We will build a simple model from scratch in Python and discuss aspects (e.g. tokenization) that are specific to chemistry and materials science data. In a second part we will delve deeper in tool-augmented language models, in which we no longer rely on the reasoning of an LLM but give the LLM access to robust tools and knowledge bases.

References:

[1]: Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. Leveraging Large Language Models for Predictive Chemistry. Nat Mach Intell 2024, 1–9. https://doi.org/10.1038/s42256-023-00788-1

[2]: Jablonka, K. M.; Ai, Q.; Al-Feghali, A.; Badhwar, S.; D. Bocarsly, J.; M. Bran, A.; Bringuier, S.; Catherine Brinson, L.; Choudhary, K.; Circi, D.; Cox, S.; Jong, W. A. de; L. Evans, M.; Gastellu, N.; Genzling, J.; Victoria Gil, M.; K. Gupta, A.; Hong, Z.; Imran, A.; Kruschwitz, S.; Labarre, A.; Lála, J.; Liu, T.; Ma, S.; Majumdar, S.; W. Merz, G.; Moitessier, N.; Moubarak, E.; Mouriño, B.; Pelkie, B.; Pieler, M.; Caldas Ramos, M.; Ranković, B.; G. Rodriques, S.; N. Sanders, J.; Schwaller, P.; Schwarting, M.; Shi, J.; Smit, B.; E. Smith, B.; Herck, J. V.; Völker, C.; Ward, L.; Warren, S.; Weiser, B.; Zhang, S.; Zhang, X.; Ahmad Zia, G.; Scourtas, A.; J. Schmidt, K.; Foster, I.; D. White, A.; Blaiszik, B. 14 Examples of How LLMs Can Transform Materials Science and Chemistry: A Reflection on a Large Language Model Hackathon. Digital Discovery 2023, 2 (5), 1233–1250. https://doi.org/10.1039/D3DD00113J

[3]: Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; Amodei, D. Language Models Are Few-Shot Learners. arXiv July 22, 2020. http://arxiv.org/abs/2005.14165 (accessed 2022-06-22)

[4]: The annotated transformer: https://nlp.seas.harvard.edu/annotated-transformer/

[5]: White, A. D.; Hocky, G. M.; Gandhi, H. A.; Ansari, M.; Cox, S.; Wellawatte, G. P.; Sasmal, S.; Yang, Z.; Liu, K.; Singh, Y.; Ccoa, W. J. P. Assessment of Chemistry Knowledge in Large Language Models That Generate Code. Digital Discovery 2023, 2 (2), 368–376. https://doi.org/10.1039/D2DD00087C

[6]: White, A. D. The Future of Chemistry Is Language. Nat Rev Chem 2023, 1–2. https://doi.org/10.1038/s41570-023-00502-0

[7]: Hocky, G. M.; White, A. D. Natural Language Processing Models That Automate Programming Will Transform Chemistry Research and Teaching. Digital Discovery 2022, 1 (2), 79–83. https://doi.org/10.1039/D1DD00009H

[8]: LLM Powered autonomous agents: https://lilianweng.github.io/posts/2023-06-23-agent/

[9]: Yao, S.; Zhao, J.; Yu, D.; Du, N.; Shafran, I.; Narasimhan, K.; Cao, Y. ReAct: Synergizing Reasoning and Acting in Language Models. arXiv March 9, 2023. https://doi.org/10.48550/arX

[10]: Karpas, E.; Abend, O.; Belinkov, Y.; Lenz, B.; Lieber, O.; Ratner, N.; Shoham, Y.; Bata, H.; Levine, Y.; Leyton-Brown, K.; Muhlgay, D.; Rozen, N.; Schwartz, E.; Shachaf, G.; Shalev-Shwartz, S.; Shashua, A.; Tenenholtz, M. MRKL Systems: A Modular, Neuro-Symbolic Architecture That Combines Large Language Models, External Knowledge Sources and Discrete Reasoning. arXiv May 1, 2022. http://arxiv.org/abs/2205.00445(accessed 2023-04-17)

[11]: Bran, A. M.; Cox, S.; White, A. D.; Schwaller, P. ChemCrow: Augmenting Large-Language Models with Chemistry Tools. arXiv April 12, 2023. http://arxiv.org/abs/2304.05376 (accessed 2023-05-22)

[12]: Boiko, D. A.; MacKnight, R.; Kline, B.; Gomes, G. Autonomous Chemical Research with Large Language Models. Nature 2023, 624 (7992), 570–578. https://doi.org/10.1038/s41586-023-06792-0

[13]: Bran, A. M.; Schwaller, P. Transformers and Large Language Models for Chemistry and Drug Discovery. arXiv October 9, 2023. https://doi.org/10.48550/arXiv.2310.06083

Maximizing High-Throughput Discovery and Machine Learning Efficiency Through Computational Workflows

Jörg Neugebauer[1], Sarath Menon[1]

[1] *Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf, Germany*

The advent of high-throughput computation and discovery combined with machine learning is revolutionizing the field of computational materials science. It enables the simulation of large systems and complex material properties with ab initio accuracy. However, the development of these data-driven activities is often a computationally complex and intensive task, requiring the combination and orchestration of multiple and often incompatible simulation codes. Automated, reliable, and robust computational workflows are required to design and execute the underlying complex simulation protocols. Using the pyiron framework (pyiron.org), the tutorial provides a hands-on introduction to all aspects of workflow design, testing, and execution, with a strong focus on materials science and atomistic simulations.

Wednesday

09:00 - 10:20	**Christoph Koch** Tutorial Lecture: Machine learning in electron microscopy and spectroscopy
10:20 - 11:00	Coffee Break
11:00 - 12:20	**Christoph Koch** Tutorial Lecture: Machine learning in electron microscopy and spectroscopy
12:20 - 14:00	Lunch + Registrations
14:00 - 14:20	Opening Remarks
14:20 - 15:00	**Andy Sode Anker** Machine learning for analysis of experimental scattering data in materials chemistry
15:00 - 15:20	**Andreja Benčan Golob** Addressing Challenges in 4D STEM Data of Ferroelectrics Using Machine Learning
15:20 - 16:00	Coffee Break
16:00 - 16:20	**Martin Uhrin** Symmetry-aware generative model for amorphous solids
16:20 - 16:40	**Christer Söderholm** Materials design of inorganic crystals with 3D transformers
16:40 - 17:20	**Kevin Rossi** DAEMON COST: a European network on Machine Learning for Materials Science
17:40 - 20:00	Poster session + Dinner

Machine learning in electron microscopy and spectroscopy

Christoph Koch[1]

[1] *Humboldt-Universität zu Berlin & Center for the Science of Materials, Berlin, Germany*

The application of Machine Learning (ML) has a long history in the analysis of experimental transmission electron microscopy (TEM) data, which includes images, diffraction patterns, spectra and multidimensional collections of them. One family of ML-methods with a long tradition is the collection of different multivariate statistical analysis (MSA) techniques, which include principal and independent component analysis (PCA & ICA), mainly for denoising or finding distinct chemical compositions or structure in experimental data sets. However, also deep learning techniques are being applied more and more, e.g. for automating the interpretation of experimental data or for producing more easily interpretable reduced representations of it. In this combination of lecture, demonstration and hands-on tutorial, an overview of a range of different ML-methods being applied in the field of TEM will be given, the potential that the novel materials discovery laboratory (NOMAD) and its online tools have for both computational and experimental materials research will be demonstrated, and a hands-on tutorial, where you will be given a chance to use these tools, will be given.

Machine learning for analysis of experimental scattering data in materials chemistry

Andy Sode Anker[1,2]

[1] *Department of Energy, Danish Technical University, Denmark*
[2] *Department of Chemistry, University of Oxford, England*

The rapid growth of materials chemistry data, driven by advancements in large-scale radiation facilities as well as laboratory instruments, has outpaced conventional data analysis and modelling methods, which can require enormous manual effort. To address this bottleneck, supervised machine learning (ML) models are frequently trained on large datasets of physics-based simulations with the aim of being applied for experimental scattering data analysis. I will demonstrate how we have used supervised ML to identify structural models from experimental scattering data.
However, ML models trained on simulated data often struggle to perform on experimental data. I will also introduce two primary challenges in the field: 1) handling data from structures not present in the training database and 2) accounting for experimental data that contains signals not included in the simulated data. Generative ML can be used to address both challenges by learning the underlying distribution of the data. I will discuss how we use generative ML to solve mono-metallic nanoparticles given pair distribution function data previously unseen by the model and how generative ML can be used to convert a simulated inelastic neutron scattering dataset into one that resembles an experiment, and vice versa.

Addressing Challenges in 4D STEM Data of Ferroelectrics Using Machine Learning

Andreja Benčan Golob[1]

[1] *Jožef Stefan Institute, Ljubljana, Slovenia*

The functionality of ferroelectric materials relies on the distribution of polarization, particularly the spontaneous dipoles formed by the separation of positive and negative charges within the unit cell. Structural imperfections, such as point defects at cationic and oxygen sites, significantly influence the polarization direction and magnitude. Quantifying these parameters is crucial for tailoring material properties. By employing scanning transmission electron microscopy (STEM) equipped with the latest 4D STEM pixelated detectors, we aim to visualize polarization fields. This technology enables detailed observations of local strain/electric fields and charge density distribution. However, the complexity and volume of data generated by 4D STEM necessitate robust computational methods for efficient processing and analysis. The current study on ferroelectrics is part of a larger Slovenian national project titled "4D STEM of Energy-related Materials down to Quantum Level." This project is dedicated to developing structural characterization through the integration of advanced electron microscopy techniques, particularly 4D STEM, with first-principles simulations and machine learning methodologies.

Symmetry-aware generative model for amorphous solids

Martin Uhrin[1], Anna Paulish

[1] *Multidisciplinary Institute in Artificial Intelligence, Université Grenoble Alpes, Grenoble, France*

Unlike their crystalline counterparts, amorphous solids cannot be described in a compact form such as a unit cell and basis due to their disordered nature. Nonetheless, there are statistical correlations in the spatial distributions of atoms that clearly distinguish them from a simple uniform distribution of uncorrelated points in space. These correlations define a manifold of reduced dimensionality compared to the full 3N degrees of freedom. What's more, just as a unit cell and basis allow us to build an entire crystal, a sufficiently complete description of the correlations can be sampled to build plausible amorphous structures of arbitrary size. These ideas are not new, and in fact form the basis of the well-known Reverse Monte Carlo (RMC) method, however, a typical shortfall is that only radial (two body) correlations are accounted for, leading to a highly incomplete description. In this work, we combine an invertible, rotationally invariant fingerprint of atomic environments, with a Variational Autoencoder neural network that captures the manifold of reduced dimensionality in the form of a latent space which can be easily sampled to generate new local environments. Using conditional generation, we can build entire unit cells by placing one environment at a time. As a demonstration of this approach, we show how it can be used to generate structures of amorphous silicon.

Materials design of inorganic crystals with 3D transformers

Christer Söderholm[1], Milica Todorović, Filip Ginter, Sampo Pyysalo

[1] *Aalto University, Finland*

Materials discovery and design through substitution of elements on different crystal lattice sites can be used to tune materials properties and offer solutions to many technological challenges. By using transformer networks, our idea is to discover new materials by predicting element substitutions similarly to masked word prediction in a sentence for natural language processing. The implementation uses a point cloud transformer network that receives 3D coordinates for atoms in unit cells, along with atomic numbers for the unmasked atoms. The model was trained in a self-supervised way on crystalline materials with masked atoms. Model performance evaluation included prediction-target element pairs visualised as a substitution matrix.
The substitution matrices showed clear trends that match established chemical rules for the substitution of elements, for example favouring substitution with elements in the same group. Accuracy for correctly predicting elements was 80.4% for the point cloud model, which also indicates the model has learnt chemical rules. Further tests included employing different materials representations; using Voronoi tessellation and coulomb matrix as descriptors improved the accuracy to 81.2% and 83.4% respectively. As a conclusion, transformer models utilizing 3D information about crystal structures can be used to generate probabilities for certain elements at given lattice sites, which enables the discovery of new materials through substitution with elements other than the most likely prediction.

DAEMON COST: a European network on Machine Learning for Materials Science

Kevin Rossi[1]

[1] *Technical University of Delft, Delft, Netherlands*

DAEMON COST [1,2] is a pan-European network consisting of 200+ members from 40+ EU countries, which focuses on capacity-building and research-coordination efforts, with the end goal of popularizing and democratizing emergent approaches, such as generative AI and automated labs, to accelerate materials discovery, design, and commercialization.
In this presentation I will discuss ongoing efforts and activities of the network towards this goal. Next, I will focus on the challenges and opportunities, which characterize materials acceleration in the specific context of a truly horizontal, inclusive, and pan-European network, together with the top-down and bottom-up level policies we aim to lobby for.

[1] https://www.cost.eu/actions/CA22154/

[2] https://cost-daemon.eu/

Thursday

09:00 - 09:40	**Ekin Dogus Cubuk** Scaling up computational materials discovery via deep learning
09:40 - 10:00	**Javier Heras-Domingo** Unlocking the Potential of EXAFS: Machine Learning Approaches for Spectroscopic Data
10:00 - 10:20	**Nejc Hodnik** Exploring Big Data for a Deeper Understanding of Electrocatalyst Behavior
10:20 - 11:20	Panel discussion and coffee
11:20 - 11:40	**Vinko Sršan** Deep learning-based drift correction in atomically resolved STEM images
11:40 - 12:00	**Andrea Ruiz** Quantitative description of metal center organization and interactions in single-atom catalysts
12:00 - 12:40	**Helge Stein** What comes after acceleration of research? What got us here?
12:40 - 14:00	Lunch
14:00 - 14:20	**Dušan Strmčnik** Machine Learning for Investigation of Nickel Surface Chemistry in Electrocatalytic Production of Hydrogen
14:20 - 14:40	**Sašo Šturm** Autonomous laboratory for sustainable research and discovery of new materials
14:40 - 15:00	**Austin Zadoks** Spectral Operator Representations
15:00 - 15:20	**Franco Pellegrini** LATTE: an atomic environment descriptor based on Cartesian tensor contractions
15:20 - 16:00	Coffee break
16:00 - 16:20	**Matthias Stosiek** Lignin Carbohydrate Complexes – Learning the Structure-Property Relation with Artificial Intelligence
16:20 - 16:40	**Lopanitsyna Nataliya** Revealing Chemical Pathways in Reaction Data through Noctis
16:40 - 17:20	**Emma King-Smith** Practical Machine Learning for Organic Small Molecule Modelling
19:00 - 20:40	Social Dinner

Scaling up computational materials discovery via deep learning

Ekin Dogus Cubuk[1]

[1] *Google DeepMind, London, United Kingdom*

Deep learning models are often evaluated on validation sets sampled from the same distribution as their training sets. In the natural sciences and engineering, however, models are evaluated for their ability to generalize beyond their "training set," either in applications of discovery or theoretical modeling. This dichotomy has caused confusion in deep learning, where methods like active learning and curriculum learning do not improve performance on independent and identically distributed (IID) academic datasets such as ImageNet, while being indispensable tools in real-life applications such as autonomous driving. With the increasing interest in using machine learning in the physical sciences, this dichotomy poses an obstacle to making meaningful progress. I will provide specific examples of this problem in the context of computational materials discovery, where graph neural networks that can predict the formation energy of inorganic crystals with unprecedented accuracy have been shown not to improve the efficiency of stable materials discovery at 0K. I will present our progress in addressing this challenge and discuss future work.

Unlocking the Potential of EXAFS: Machine Learning Approaches for Spec- troscopic Data

Javier Heras-Domingo[1], Andrea Ruiz-Ferrando, Stephan Pollitt, Sharon Mitchel, Olga Safonova, Javier Pérez-Ramírez and Núria López

[1] *Institute of Chemical Research of Catalonia (ICIQ), Tarragona, Spain*

Sustainability will only be achieved through the pursuit of precise control of materials, for instance utilizing minimal amounts of critical raw materials, which stands as the Holy Grail in heterogeneous catalysis. Heterogeneous catalysts featuring atomically dispersed active and stable metal atoms were coined in 2011 as single-atom catalysts (SACs). Since then, SACs exhibit remarkable and exciting performances in numerous industrially significant reactions. The host can be different materials, including oxides, and most commonly carbon-based ones. Due to the varied synthetic forms of carbon matrix SACs offer a very wide speciation (charge and coordination of the metal atom) imparting unique geometric and electronic properties needed in the catalysis field that requires to be identified via characterization tools like microscopy and X-ray absorption spectroscopy.

Generally, in heterogeneous catalysis (and more critically in SACs) there is no direct way to inverse engineer active sites (i.e., targeting a catalytic property and identifying the synthesis route) and thus material synthesis and testing is followed by strict and expensive characterization protocols aimed to provide hints for structures that can be then evaluated via Density Functional Theory (DFT) simulations to trace structure-activity relationships. The ultimate consequence is that the development of materials is limited by the characterization step, which the demand for synchrotron characterization techniques is skyrocketing, even if in many cases the data treatment is done manually. In addition, bench solutions like hiXAS (HP Spectroscopy) and easyXAFS are being developed and the cost of this equipment is getting reduced. Therefore, the amount of data and users is growing at a fast pace, highlighting the pressing need of leveraging the newest advances in data science and artificial intelligence (AI) to the field.

In this work, we have developed intelligent and automated EXAFS tools for the rational design of single-atom catalysts, which leverage artificial intelligence (AI) to enhance data extraction, pivoting to quantitative information. We leveraged automated workflows to generate high-quality theoretical data using state-of-the-art open-source software package Atomate. The collected data has been used to train a new Graph Neural Network model which is able to directly predict the EXAFS spectra of Platinum single atom catalysts (SACs) with outstanding performance.

Exploring Big Data for a Deeper Understanding of Electrocatalyst Behavior

Nejc Hodnik [1]

[1] *National Institute of Chemistry, Ljubljana, Slovenia*

Electrocatalysis is experiencing a growing trend each year, presenting a promising shift for the chemical industry to rely solely on electrical energy sourced from sustainable options like solar and wind power. Through electrochemical energy conversion, we can efficiently store this green energy in chemical bonds. However, the study of electrocatalysts is challenging, requiring numerous trial-and-error tests to understand structure-activity and -stability behaviors. Electrochemical methods are ideal for automation, facilitating a transition to high-throughput techniques. Integration with automated characterization techniques such as electron microscopy, X-ray diffraction, and X-ray photoelectron spectroscopy results in a substantial generation of data.

The systematic acquisition, storage, processing, and analysis of this data becomes pivotal. I firmly believe that a wealth of untapped insights into electrocatalytic behavior await discovery through the application of machine learning approaches. In my talk, I will showcase a few examples illustrating how these approaches can be exploited. Such strategies have the potential to revolutionize our comprehension, exploration, and enhancement of electrocatalysis, thereby paving the way for new horizons in sustainable energy development.

Deep learning-based drift correction in atomically resolved STEM images

Vinko Sršan[1], Marjan Stoimchev, Dragi Kocev, Sašo Šturm, Sašo Džeroski

[1] *Jožef Stefan Institute, Ljubljana, Slovenia*

The scanning transmission electron microscope (STEM), powered by spherical aberration corrector, poses a powerful material characterization tool at atomic resolution, which is nowadays only limited by the microscope's stability and the sample itself. One of such disturbances is specimen and/or stage drift, leading to distorted images. In this case, the physical information obtained about the material is fundamentally flawed, since the lattice parameters and angles may not match the fundamental truth of the underlying crystal structure. To date, many classical warping algorithms have been developed to correct STEM images for drift, but their effectiveness relies on users' ability to adequately assign warping parameters; therefore, modern developments have shifted towards the application of machine learning-based algorithms.
In this work, we propose a deep learning (DL) approach for drift correction based on convolutional autoencoders. The model was trained on 45 000 image pairs consisting of drift-free and drifted images, respectively, which were generated through multislice STEM simulations and included several image distortion types (such as linear and nonlinear drift, scanning noise, scan skips etc.). The loss function used to train the neural network was the standard mean square error (MSE) between the corrected and drift-free images, which proved to be successful in correcting drifted images.
The performance of the model was assessed on separate set corresponding to different kinds of drifts, validating its effectiveness and robustness against different signal-to-noise ratio values. More specifically, we obtained 0.3% error in mean pixel intensity correlation between undrifted and corrected image, and 45.6 peak signal-to-noise ratio (PSNR) values for performance.
Our findings suggest that the proposed approach would yield fast and accurate drift correction of STEM images, resulting in a significant reduction of time needed for data analysis while providing enhanced image quality.

Quantitative description of metal center organization and interactions in single-atom catalysts

Andrea Ruiz[1], Kevin Rossi, Dario Faust Akl, Victor Gimenez Abalos, Javier Heras-Domingo, Romain Graux, Dario Garcia-Gasulla, Núria López, Javier Pérez-Ramírez, Sharon Mitchell

[1] *Institute of Chemical Research of Catalonia (ICIQ), Tarragona, Spain*

Ultra-high-density single-atom catalysts (UHD-SACs) unlock potential cooperative effects between adjacent metal centers. Achieving precise metal center proximity relies on intricate synthesis techniques, which are influenced by factors such as support sites, metal adsorption, and synthesis conditions. Characterizing metal centers organization mainly depends on aberration-corrected scanning transmission electron microscopy (AC-STEM), but current methods mostly visually compare metal center dispersion, offering limited quantitative information about atom density and proximity. Machine learning provides a solution to standardize image analysis and overcomes manual detection processes. This work is based on the continuation of a previous deep-learning method for the automated atomic detection in image analysis based on a representative case study (low-density Pt single atoms based on functionalized carbon). With the extension of the pipeline to UHD SACs where multimers and different chemical species can be detected, we quantitatively characterize metal center identity, spatial organization and interactions in both mono- and multimetallic UHD-SACs across various surface atom densities. This analysis establishes criteria for metal center isolation and characterizer neighbor distance distributions relative to metal surface density, thus providing a basis for establishing quantitative structure-performance relationships.
This work represents a transformative shift in characterizing UHD-SACs, moving from qualitative to systematic, quantitative methods employing microscopy and machine learning. The development of dedicated machine learning algorithms for precise extraction of atomistic details promises not only to enhance our fundamental understanding of catalytic materials but also to enable the engineering of structure-property relationships.

What comes after acceleration of research? What got us here?

Helge Stein[1]

[1] *Technical University of Munich, Munich, Germany*

Automation and artificial intelligence have been rapidly evolving, allowing experimental chemists to accelerate discovery and scale-up processes. However, we must reflect on the journey that led us here and ponder what lies ahead. This talk will review the (r)evolutionary history of high-throughput experimentation (HTE), inverse design, and recent developments in materials acceleration platforms (MAPs). We'll explore the general concepts, robots, and machine learning algorithms that enabled us to test and manage experiments on millions of catalysts and build and test hundreds of batteries daily. However, speed alone is meaningless if research tasks are not integrated. We must go beyond intra-lab integration and accelerate science across labs and countries. This talk will highlight the fundamental challenges in accelerating research and what we must do to expedite the proliferation of insights and tools from the lab bench to factories and back. Throughout the journey, we'll examine the milestones that paved the way for the acceleration of research. From the early adoption of automation to the integration of machine learning, we'll explore the key developments that have shaped the landscape of experimental chemistry. Finally, the talk will conclude with demonstrations of accelerated developments from our lab that we successfully scaled to industry, showcasing the potential of these advancements in real-world applications. Join me as we reflect on the past, celebrate the present, and envision the future of accelerated research in chemistry and beyond.

Machine Learning for Investigation of Nickel Surface Chemistry in Electrocatalytic Production of Hydrogen

Boštjan Gec, Dževad K. Kozlica, Matjaž Finšgar, Ljupčo Todorovski Sašo Džeroski, Dušan Strmčnik[1]

[1] *National Institute of Chemistry, Ljubljana, Slovenia*

The increasing energy demand is becoming one of the major global challenges of modern time, causing both economic and existential crises. One of the alternative and clean ways of producing, capturing, and utilizing energy is converting sustainable energy into hydrogen gas, which is an ideal candidate because of its high energy density and zero emission of pollutants during (electro)chemical conversion. Electrochemical production of hydrogen through alkaline water and chlor-alkali electrolysis are two of the more promising approaches that fit into the clean and sustainable energy cycle. Currently, one of the most used metals in industrial-scale electrolytic hydrogen production is nickel.

While Ni has historically been one of the go-to materials for hydrogen evolution production (HER), the lack of fundamental understanding of the interfacial processes on the atomic/molecular level has hindered major progress. Ni has a very rich and complex surface chemistry, making the analysis and comparison of individual samples challenging and the fundamental insights extremely elusive.

We will present the utilization of hierarchical agglomerative clustering methods and symbolic regression methods to extract meaningful insights from large datasets generated through advanced analytical techniques such as TOF-SIMS, XPS and electrochemical methods on Ni-based materials and identify the relationships between surface characteristics and material (electro)chemical properties.

Autonomous laboratory for sustainable research and discovery of new materials

Sašo Šturm[1], Matej Komelj, Barbara Ljubec Božiček, Belisa Alcantara Marinho

[1] *Jožef Stefan Institute, Ljubljana, Slovenia*

We introduce a pioneering project that aims to reshape materials discovery by establishing an autonomous lab dedicated to sustainable research and the development of new materials. The new way of performing research will be based on an inverse materials design approach, where the desired end material property drives the entire discovery process. For instance, our focus on high-entropy oxides, a novel material type with promising catalytic properties and a theoretical potential for over a million different catalytic sites, exemplifies our approach. Nevertheless, turning theoretical possibilities into actual materials efficiently is challenging with traditional experimentation methods. This is even more intensified in studies with high entropy oxides due to their unusual composition containing five or more principal metal cations and oxygen ions in a single-phase crystal structure, which gives thousands of possibilities for their final composition and structure.

Our research plan is to merge state-of-the-art robotic and automation synthesis and characterisation tools, generating extensive data. This data will be enhanced by machine learning (ML) and artificial intelligence (AI) supported by ab-initio and quantum-computing-based simulations, and fed into a continuous process for materials optimisation, along with insights from fundamental modelling. This truly autonomous approach to materials discovery represents a significant shift in materials discovery, with machines actively augmenting human researchers' abilities.

Spectral Operator Representations

Austin Zadoks[1], Antimo Marrazzo, Nicola Marzari

[1] *EPFL, Lausanne, Switzerland*

Machine learning in atomistic materials science has grown to become a powerful tool, with most approaches focusing on atomic arrangements, typically decomposed into local atomic environments. This approach, while well-suited for machine-learned potentials, is conceptually at odds with learning complex intrinsic properties, often driven by the spectral properties (e.g., band gaps or mobilities) of materials, which cannot be readily atomically partitioned. For such applications, methods which represent electronic rather than atomic structure are promising. We discuss a general framework for these spectral operator representations (SOREPs) as electronic structure descriptors which take advantage of the natural symmetries and inherent interpretability of physical models. Using this framework, we formulate simple SOREPs and apply them to distinguishing carbon nanotube and barium titanate polymorphs and to the discovery of novel transparent conducting materials (TCMs) in the Materials Cloud 3D database (MC3D) using a random forest classifier. By training only on 1% (N=222) of materials in the MC3D, our model is able to correctly label 85% of entries in the database which meet common screening criteria for promising TCMs.

LATTE: an atomic environment descriptor based on Cartesian tensor contractions

Franco Pellegrini[1], Stefano de Gironcoli, Emine Kucukbenli

[1] *SISSA, Trieste, Italy*

We present the Local Atomic Tensors Trainable Expansion (LATTE), a new descriptor for local atomic environments to be used in combination with machine learning models for the construction of interatomic potentials. LATTE allows for the efficient construction of a variable number of many-body terms with learnable parameters, resulting in a descriptor that is efficient, expressive, and can be scaled to suit different accuracy and computational cost requirements. We compare this new descriptor to existing approaches on several systems, showing it to be competitive with very fast potentials at one end of the spectrum, and extensible to an accuracy close to the state of the art.

Lignin Carbohydrate Complexes – Learning the Structure-Property Relation with Artificial Intelligence

Matthias Stosiek[1], Joakim Löfgren, Daryna Diment, Davide Rigo, Marie Alopaeus, Chunlin Xu, Michael Hummel, Mikhail Balakshin, Patrick Rinke

[1] *Aalto University, Aalto, Finland*

The potential of lignin as an abundant, underutilized biopolymer is increasingly being realized. A key challenge for the targeted production of lignins remains the poorly understood relation between lignin properties and its complex structure. Novel artificial intelligence (AI) methods could reveal such structure-function relationships but remain elusive in biomaterials research. Here, we present our AI study of the structure-function relationships in lignin carbohydrate complexes (LCC).
LCCs are extracted from birch wood combining the Aqua Solv Omni (AqSO) biorefinery process [1] with AI-guided data acquisition [2] by varying reactor temperature (T), reaction severity (P-factor) and liquid to solid (L/S) ratio. Each LCC sample is characterized with 2D nuclear magnetic resonance (NMR) spectroscopy. The 62 NMR spectra are complemented with measurements of key lignin properties. As first examples, we focus on the antioxidant activity, glass transition temperature, molecular weight, surface tension and degradation metrics.
The analysis of 2D NMR spectra with AI is made feasibly through dimensional reduction of the 1024x1024 pixels of each NMR spectrum. We divide the spectra into rectangular sectors, which are integrated to a single number and choose 70 of these integrated sectors (IS) with the highest variance across samples.
To establish structure-function relationships, we first correlate these ISs of the NMR spectra with the corresponding property measurements. The employed random forest regression (RFR) model shows good predictive capabilities. Subsequently, we use a RFR feature importance analysis to identify structural features that correlate with each property, see Figure 1, and provide a chemical interpretation of our findings.

[1] D. Tarasov, P. Schlee, A. Pranovich, A. Moreno, L. Wang, D. Rigo, M. H Sipponen, C. Xu, M. Balakshin, Green Chem. 24, 6639 (2022)

[2] J. Löfgren, D. Tasarov, T. Koitto, P. Rinke, M. Balakshin, M. Todorović, ACS Sustainable Chemistry & Engineering 10, 29 (2022)

Revealing Chemical Pathways in Reaction Data through Noctis

Lopanitsyna Nataliya[1]

[1] *Syngenta, Basel, Switzerland*

The Design-Make-Test-Analyze (DMTA) cycle is critical to the chemical industry, and its optimization is essential to handle challenges and ensure future readiness. Digital transformation of the cycle's steps can leverage both legacy and newly generated data to speed up new product development by reducing experimental workload. In particular, the design phase is strongly reliant on human expertise and struggle with analyzing large data bases. As a result, the development of cheminformatics tools becomes essential, enabling chemists to efficiently record, share, and analyze data while preventing oversights and eliminating duplication of work.
In this talk, I will outline the current challenges in synthesis design practices, highlighting the importance of advancing FAIR data standards and data quality in general, as well as the crucial roles that AI plays in small molecule research. I will also introduce Noctis, a Python toolkit for cheminformatics, designed to exploit chemical relations between reaction datapoints to construct knowledge graphs. Noctis offers functionality to build the reaction network combining data from different sources and an advanced suite of tools to interact with it. With the help of Noctis, one can mine synthetic routes from experimental data and combine them with reactivity model predictions to support data-driven synthesis design. This lays the foundation for our bigger ambition to use the knowledge graph as a robust base for developing a language model-driven recommendation system.

Practical Machine Learning for Organic Small Molecule Modelling

Emma King-Smith [1]

[1] *University of Cambridge, Cambridge, United Kingdom*

Synthetic chemistry has many open challenges: how reaction yields change as reactants and conditions change, how molecules interact with the human body, or the full underlying mechanisms of some workhorse reactions. Machine learning (ML) has seen enormous strides in modelling the world's "black boxes": from image processing and recognition that rival human ability, consistently beating human players in a variety of games, to the amusing ruminations of the latest large language models. Due to the low standardization of data, few large chemistry-focused datasets, and the mere fact that molecules are difficult systems to model, ML has historically struggled to make headway in the chemical sciences. Recent developments in ML models and increased access to open-source chemistry datasets have opened the door to practical ML models, including DFT and molecular property predictions and biological activity predictions. Herein, we present case studies utilizing recent and classic ML methods to further our predictive ability and understanding of synthetic chemistry.

Friday

09:00 - 09:40	**Teodoro Laino** Fueling the Digital Chemistry Revolution with Language and Multimodal Foundation Models
09:40 - 10:00	**Lei Zhang** A Comparative Study of Machine Learning Models and Vector Analysis Techniques for Improved Prediction of Quaternary Material Systems Based on Word Embeddings
10:00 - 10:20	**Matej Praprotnik** Developing an Implicit Solvation Machine Learning Model for Molecular Simulations of Ionic Media
10:20 - 11:20	Panel discussion and coffee
11:20 - 11:40	**Matilda Sipilä** Application of the Question Answering method to extract information from materials science literature
11:40 - 12:00	**Morgan Kerhouant** Enterprise deployment, scaling and democratisation of R&D models
12:00 - 12:40	**Tian Xie** MatterGen: a generative model for inorganic materials design
12:40 - 13:00	Closing remarks

Fueling the Digital Chemistry Revolution with Language and Multimodal Foundation Models

Teodoro Laino [1]

[1] *IBM Zurich, Zurich, Switzerland*

In the last years, natural language processing models have emerged as one of the most effective, scalable approaches for capturing human knowledge and modelling chemical processes in organic chemistry. Its use in machine learning tasks demonstrated high quality and ease of use in problems such as predicting chemical reactions [1-2], retrosynthetic routes [3], digitizing chemical literature [4], predicting detailed experimental procedures [5], designing new fingerprints [6] and yield predictions [7]. In this talk, I'll talk about the impact of language models in chemistry by highlighting the critical role of NLP architectures in implementing the first cloud-based AI-driven autonomous laboratory [8] and in the potential of multimodal foundation models in addressing the data capture problems in experiments performed by human scientists.

References:

[1]: IBM Research Europe, Chem. Sci., 2018, 9, 6091-6098

[2]: IBM Research Europe, ACS Cent. Sci. 2019, 5, 9, 1572-1583

[3]: IBM Research Europe, Chem. Sci., 2020, 11, 3316-3325

[4]: IBM Research Europe, Nat. Comm., 2020, 11, 3601

[5]: IBM Research Europe, Nat. Comm., 2021, 12, 2573

[6]: IBM Research Europe, Nat. Mach. Intel., 2021, 3, 144–152

[7]: IBM Research Europe, Mach. Learn.: Sci. Technol., 2021, 2, 015016

[8]: https://rxn.res.ibm.com

A Comparative Study of Machine Learning Models and Vector Analysis Techniques for Improved Prediction of Quaternary Material Systems Based on Word Embeddings

Lei Zhang[1], Lars Banko, Wolfgang Schuhmann, Alfred Ludwig, Markus Stricker

[1] *Ruhr-University Bochum, Bochum, Germany*

The prediction and extrapolation of material properties of yet-unknown materials is a significant challenge in materials discovery. We present three models fitted with two ternary materials property datasets to predict performance within their shared quaternary system. The first uses standard Gaussian Process Regression (GPR) based on elemental compositions. The second is also a GPR, however, not based on elemental composition but a representation of 'composition' in form of word embedding-derived materials vectors. The third model advances the methodology by employing the 'standard vector method', which synthesizes weighted word embedding-based vector representations of material properties to establish a benchmark vector, facilitating the prediction of material performance in the quaternary system by assessing similarity to this idealized reference.
With this we demonstrate the effectiveness to integrate meaningful word embedding-based representations of materials when materials data is scarce. Such robust predictions of higher-dimensional compositions spaces allows effective screening of the materials space based on available lower-dimensional composition-property spaces.

Developing an Implicit Solvation Machine Learning Model for Molecular Simulations of Ionic Media

Matej Praprotnik [1]

[1] *National Institute of Chemistry, Ljubljana, Slovenia*

Molecular simulations of biophysical systems require accurate modeling of their native environment, i.e., aqueous ionic solution, as it critically impacts the structure and function of biomolecules.
At the same time, the models should be computationally efficient to enable simulations of large spatiotemporal scales. In this talk, I will present a deep implicit solvation model for sodium chloride solutions that satisfies both requirements.
Owing to the use of a neural network potential, the model can capture the many-body potential of mean force, while the implicit water treatment renders the model inexpensive.
I will demonstrate the aplicability of our approach for pure ionic solutions and a solvated DNA molecule. In both cases, the structural properties are in good agreement with all-atom molecular simulations, showcasing a general methodology for the efficient and accurate modeling of ionic media.

Application of the Question Answering method to extract information from materials science literature

Matilda Sipilä[1], Farrokh Mehryary, Sampo Pyysalo, Filip Ginter, Milica Todorović

[1] *University of Turku, Turku, Finland*

Scientific text is a promising source of data in materials science, and there is ongoing research on how to utilise textual data in materials discovery. In addition to the more established approaches like named entity recognition or dictionary-based methods, new machine learning tools such as question answering (QA) are becoming available. The advantages of this method are that it is easy to scale and that it does not require manual text labeling or annotating work, but there may be some loss in precision compared to other methods.
We tested the performance of the QA method on the well-known task of information extraction. We extracted bandgap values of halide perovskite materials from scientific literature. Large language models (BERT models) were tuned towards a specific QA task and then used to select the correct answer for the question about materials properties. In comparison to more established methods, the QA method performed well, and we were able to extract correct information from text. This information can be used to map the space of materials properties and find promising new materials solutions. The potential in QA method lies in versatility, accessibility and scalability, since it is easy to use even for researchers with no previous knowledge of language technology and can be easily scaled to extract different materials and properties.

Enterprise deployment, scaling and democratisation of R&D models

Morgan Kerhouant[1]

[1] *Quaisir, London, United Kingdom*

We highlight the challenges in getting a modelling workflow running on a user's laptop to deploying it at scale and as a service on the cloud. We talk about the democratisation aspects and making these tooling accessible to all internal stakeholders within an organisation. We talk about the issues from the perspective of multiple stakeholders, connectivity and collaboration pains and the value of streamlining the process at a platform level to accelerate R&D cycles and decision making. To put things into perspective we use an example of interfacing grammatical evolution with a computational chemistry simulator for materials discovery.

MatterGen: a generative model for inorganic materials design

Tian Xie[1]

[1] *Microsoft Research Cambridge, Cambridge, United Kingdom*

The design of functional materials with desired properties is essential in driving technological advances in areas like energy storage, catalysis, and carbon capture. Traditionally, materials design is achieved by screening a large database of known materials and filtering down candidates based on the application. Generative models provide a new paradigm for materials design by directly generating entirely novel materials given desired property constraints. In this talk, we present MatterGen, a generative model that generates stable, diverse inorganic materials across the periodic table and can further be fine-tuned to steer the generation towards a broad range of property constraints. To enable this, we introduce a new diffusion-based generative process that produces crystalline structures by gradually refining atom types, coordinates, and the periodic lattice. We further introduce adapter modules to enable fine-tuning towards any given property constraints with a labeled dataset. Compared to prior generative models, structures produced by MatterGen are more than twice as likely to be novel and stable, and more than 15 times closer to the local energy minimum. After fine-tuning, MatterGen successfully generates stable, novel materials with desired chemistry, symmetry, as well as mechanical, electronic and magnetic properties. Finally, we demonstrate multi-property materials design capabilities by proposing structures that have both high magnetic density and a chemical composition with low supply-chain risk. We believe that the quality of generated materials and the breadth of MatterGen's capabilities represent a major advancement towards creating a universal generative model for materials design.

Posters

Posters - Tuesday Session

Boštjan Gec
Using equation discovery to analyze ToF-SIMS data in materials science for energy production
Eleftherios Christofi
Physics-Informed Deep Learning Approach for Reintroducing Atomic Detail in Coarse-Grained Configurations of Multiple Poly(lactic acid) Stereoisomers
Sintija Stevanoska
A closer look at sustainability criteria for high entropy alloys
Anastasija Manojlovska
Zero-shot image segmentation for analysis of nanomagnetic particles
Sevra Çiçekli
Specification of a general method of Sustainable Materials by Design: application to batteries
Abdurrahman Adhyatma
Data-driven multimodal XPS study of molecular adsorbates on atmospheric aerosol nanoparticles
Davide Bidoggia
Dynamics of structural phase transitions in monolayer TMDs with a neural-network interatomic potential
Javier Heras-Domingo
Unlocking the Potential of EXAFS: Machine Learning Approaches for Spectroscopic Data
Prajwal Dattatray Pisal
Leveraging an Adsorption Energy-based Descriptor for Discovery of CO2 Hydrogenation Catalysts
Sebastian Utz
First Principles Study of the Interfacial Stability between Alkali Metal Anodes and their Solid Electrolytes
Bernhard Kretz
Machine-learning DFT-based band gaps of carbon nanostructures
Abdoul FataohKabore
Artificial Intelligence for Predicting New High Entropy Materials
Francesca Menescardi
Molecular dynamics strategies to determine the melting curve of CaO employing NN potentials
Arunima Singh
Machine-learning based optimization of hafnium-zirconium oxide compositions for stable ferroelectric applications
Konstantinos Konstantinou
Tuning oxidized crystalline semiconductor surfaces with first principles modelling and Bayesian optimization
Bruno Mladineo
A Machine Learning Approach to Study of Thermosalient Molecular Crystals
Karol Frydrych
Long-short term memory networks for parameter optimization: application to cyclic deformation of electrodeposited copper
Yuhao Zhang
Machine Learning Optimization of Thermally Activated Nylon Actuator Coils
Miha Hotko, Aleš Marsel
From Data to Discovery: Designing a Relational Electrocatalyst Database for Machine Learning-Driven Pattern Recognition
Jaka Olenik
ML assisted SERS Recognition of Trace Explosives
Enrico Pedretti
Xsorb: a software to automate the study of molecular adsorption on crystal surfaces
Marcin Maździarz
Suitability of the available interatomic potentials for the modeling of 2D materials
Nikolaj Rønne
Generative diffusion model for surface structure discovery
Ožbej Vodeb
Use of artificial intelligence in extracting kinetic parameters of an electrochemical reaction

Luca Leoni
Global sampling of Feynman diagrams trhough Normalizing FLows
Ivor Lončarić
Thermomechanical Properties of Molecular Crystals from Machine Learning Potentials
Pol Sanz
Neural network potential training framework for catalyst characterization
Dusabimana Muhire Fabrice
Unleashing the PtNP through Machine Learning-Driven Exploration of Electronic Properties
Kamila Kazmierczak
Quantitative Structure – Activity/Properties Relationships (QSA/PR) in Polymerization from Industrial Side
Bojana Paskaš Mamula
Exploring Metal Hydride Stability: Development and Application of the MetalHydrideEnth Machine Learning Model
Antoni Wadowski
Machine learning for metallic glasses
Sepideh Baghaee Ravari
Leveraging Knowledge Graphs for Extraction and Linkage of Information from Unstructured Data
Han Le
Design of Experiments: A Quantitative Comparison of Bayesian Optimization with Response Surface Methodology
Mahboubeh Shabani
Can a Neural Network learn the laws of friction?

Posters - Wednesday Session

Elif Muslu
Optimizing Silicon Thin-Film Anodes for Lithium-Ion Batteries: A Machine Learning Approach to Improving Discharge Capacity
Marjan Stoimchev
Deep learning based drift correction in atomically resolved STEM images
Anton Kokalj
Estimating the energy of Bi atoms' configurations with machine learning
José Oliveira, Patrícia Ramos
Enhancing Time Series Forecasting in Materials Science through Large Pre-Trained Models and Transfer Learning
Reza Gholizadeh, Blaž Likozar, Matej Huš
Synergistic Contribution of Machine Learning and Multiscale Modeling in the Field of Electrochemical Reduction of CO2 to Value-Added Chemicals
Amirhossein Naghdi Dorabati
Unbiased Atomistic Predictions of Crystal Dislocation Dynamics using Bayesian Force Fields
Alejandro Cañete Arché
Automated Workflow for Surface Coverage Analysis under Electrochemical Conditions
Ervin Rems
On the thermodynamic stability of the MXene family
Laura Mismetti
Automated ab initio-accurate atomistic simulations of dissociated dislocations
Minggao Feng
MCRT: A pre-trained foundation model for molecular crystals
Ziga Casar
Resolving the Structure of Nanocrystalline Calcium Silicate Hydrate: Marriage of experiments and atomistic modeling
Christian Camacho
Multi-criteria foamed glass design with multi-target regression and multi-objective optimization
Mohammed Zniber
Data digitalization strategies for AI-driven experimentation
Lucas Bandeira
Machine Learning Interatomic Potentials for Atmospheric Chemistry
Tatu Linnala
Performance comparison of heteroscedastic and homoscedastic noise models in Bayesian optimization
Layrton José Souza da Silva

Application of Neural Network-driven image processing for semantic segmentation of in-situ Liquid-Cell TEM images

Parisa Piran

Transformer Neural Networks for predicting the functional properties of materials

Benjamin Udofia

Identification of dislocation structures in experimental Laue microdiffraction patterns

Houlong Zhuang

Discovery of Novel High-Entropy Materials via Quantum Machine Learning

Tomasz Galica

Visual representation of materials and their properties

İbrahim Aydemir

Band gap estimation of ZnO through common technology and AI models

Bakhtiyar Mammadli

Detecting Plasticity in Material Mechanical Deformations through Digital Image Correlation and Unsupervised Machine Learning

Jan Vidergar

Algorithms for the segmentation of platinum alloy nanoparticles in electron micrographs

Aleksandar Živković

Probing Oxide Interfaces: Revealing Hidden Phenomena via Computational Modelling

Katarina Batalović

High-throughput search for novel Van der Waals (vdW) heterostructures

Jakob Baumsteiger

Exploring Noncollinear Magnetic Energy Landscapes with Bayesian Optimization

Elisa Damiani

High throughput first-principle prediction of tribological properties of solid-solid interfaces

Mina Taleblou

Handling many atomic species with LATTE

Ana Rebeka Kamšek

Extracting structural features of Pt-alloy nanoparticles using identical-location 4D-STEM and unsupervised algorithms

Federico Orlando

AMaRaNTA: An AiiDA-based Workflow To Automate Calculations Of Exchange Parameters In 2D Magnets

Naveed Ashraf

Reverse the effect of CO2 on climate change and turns into green energy

Muhammad Awais

Forging a sustainable future: A computational approach to CO2 reduction