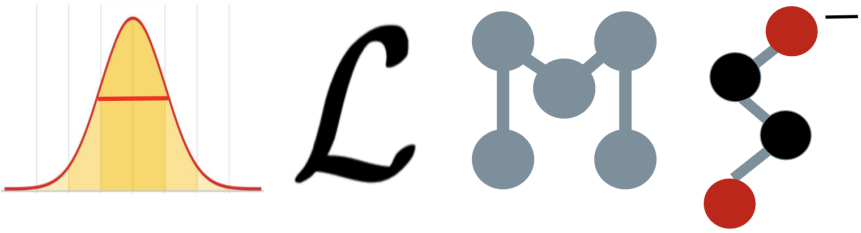
Actively Learning Materials Science

Aalto University, Espoo, Finland

27.02.-03.03.2023



Book of Abstracts

Contents

General

Organizers

Matthias Stosiek, Aalto University, Finland
Armi Tiihonen, Aalto University, Finland
Kevin Rossi, ETH Zurich, Switzerland
Milica Todorović, University of Turku, Finland
Patrick Rinke, Aalto University, Finland

Code of Conduct

We strive for making every attendee feel welcome and respected, regardless of gender identity, sexual orientation, disability, physical appearance, body size, race, nationality, religion, background, level of education, or socioeconomic status.

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.

Health and Safety Measures

When you are indoor and you are not speaking, eating, or drinking, you are suggested to wear a mask covering your mouth and nose, according to the latest World Health Organization. If you feel unwell, we kindly suggest you to not attend the conference in person and join the event online instead.

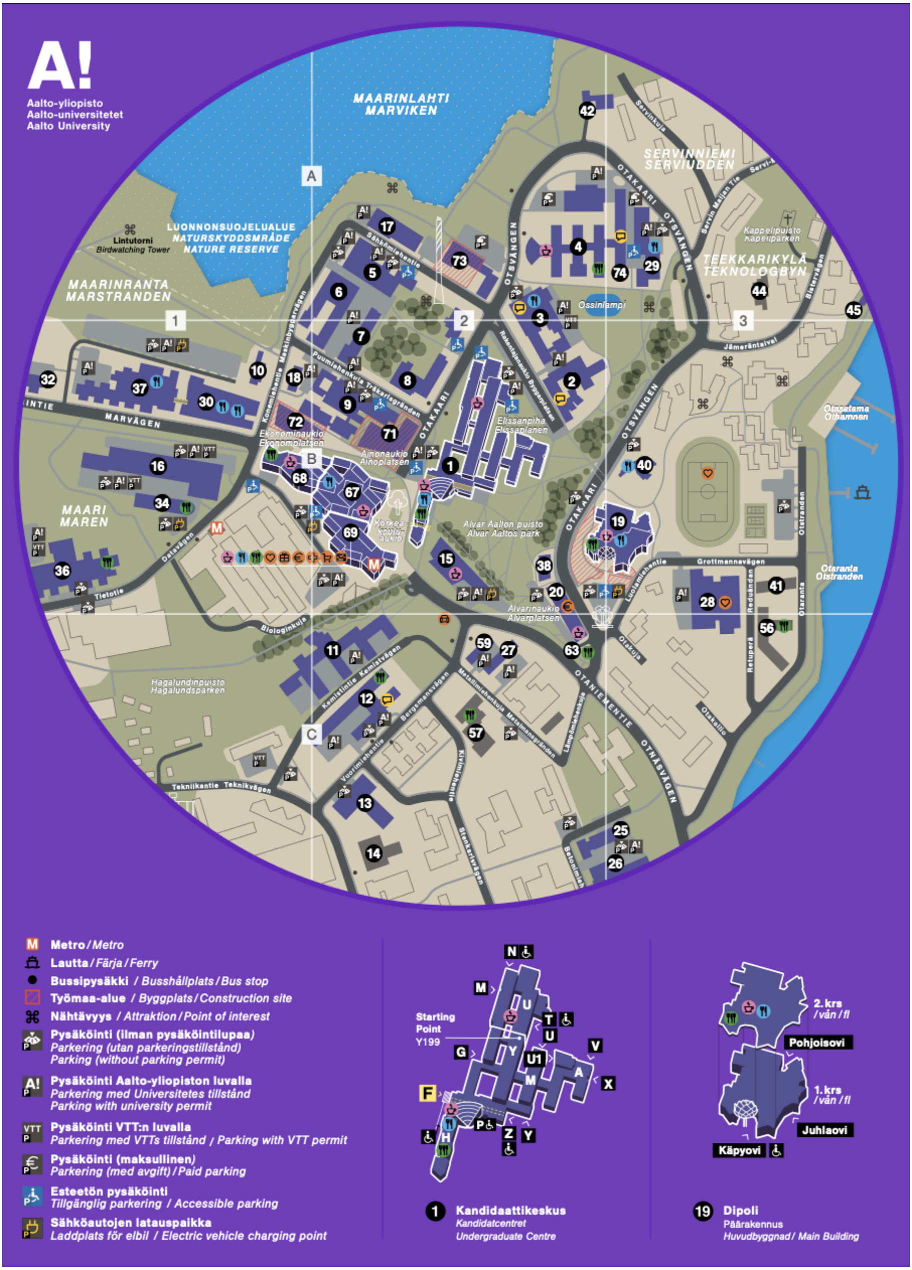
Every attendee should feel in a safe environment, make sure to agree and respect the right distance from others. If in doubt, a 1.5m distance is likely to ensure a lesser likelihood of COVID19 transmission.

Emergency numbers: 112 general (police, fire brigade, first aid).

Venues

Use next two pages images to familiarize with the Campus Map, to easily identify the location of Otakaari (Undergraduate Center building, labelles as building 1 in the map) and Kaleva Rooms (Dipoli building, labelled as building 19 in the map) at the Aalto University Campus, and other useful services.

Monday venue: Undergraduate Center, Otakaari - Room U1
Tuesday venue: Undergraduate Center, Otakaari - Room U4
Wednesday venue: Dipoli Building - Kaleva Room
Thursday venue: Dipoli Building - Kaleva Room
Friday venue: Dipoli Building - Kaleva Room


A!
Aalto-yliopisto
Aalto-universitetet
Aalto University
MAARINLAHTI
MARVIKEN
LUONNONSUOJELUALUE
NATURSKYDDSMRÅDE
NATURE RESERVE
Lintutorni
Birdwatching Tower
MAARINRANTA
MARSTRANDEN
SERVINNIEMI
SERVIUDDEN
TEEKKARIKYLÄ
TEKNOLOGBYN
Kappelipuisto
Kapellparken
Ossinlampi
MAARI
MAREN
Otsaatama
Otshamnen
Ekonominaukio
Ekonomplatsen
Ainonaukio
Ainoplatsen
Elissanpiha
Elissaplanen
Alvar Aallon puisto
Alvar Aaltos park
Alvarinaukio
Alvarplatsen
Hagalundinpuisto
Hagalundsparken
Otaranta
Otstranden
MARVÄGEN
OTAKAARI
OTSVÄNGEN
OTANIEMENTIE
OTNÄSVÄGEN
Jämeräntaival
Grottmannavägen
Metro / Metro
Lautta / Färja / Ferry
Bussipysäkki / Busshållplats / Bus stop
Työmaa-alue / Byggplats / Construction site
Nähtävyys / Attraktion / Point of interest
Pysäköinti (ilman pysäköintilupaa)
Parkering (utan parkeringstillstånd)
Parking (without parking permit)
Pysäköinti Aalto-yliopiston luvalla
Parkering med Universitetes tillstånd
Parking with university permit
Pysäköinti VTT:n luvalla
Parkering med VTTs tillstånd / Parking with VTT permit
Pysäköinti (maksullinen)
Parkering (med avgift) / Paid parking
Esteetön pysäköinti
Tillgänglig parkering / Accessible parking
Sähköautojen latauspaikka
Laddplats för elbil / Electric vehicle charging point
Starting Point
Y199
1 Kandidaattikeskus
Kandidatcentret
Undergraduate Centre
2. krs / vån / fl
Pohjoisovi
1. krs / vån / fl
Juhlaovi
Käpyovi
19 Dipoli
Päärakennus
Huvudbyggnad / Main Building

Otaniemi / Otnäs

A Blanc Ostoskeskus *Köpcentret / Shopping Centre* 20 B2
A Bloc Kauppakeskus / *Köpcentret Shopping Centre* 69 B2
A Grid Startup-hubi *Startup-hubb / Startup hub* 4 A2
Aalto Bioproduct Centre 13 C2
Aalto Inn 41 B3
Aalto Studios 29 A3
Aalto-yliopisto Junior *Aalto University Junior* 11 C2
Aalto-yliopistokiinteistöt Oy Aalto University Campus & Real Estate 2 B2
Arkkitehtuurityöpaja *Architecture Workshop* 27 C2
AVP Aalto Ventures Program 37 B1
AYY Aalto-yliopiston ylioppilaskunta / *AUS Aalto-universitetets studentkår / AYY Aalto University Student Union* 38 B2
Circular Raw Materials Hub 12 C2
Design Factory 25 C3
Dipoli Päärakennus *Huvudbyggnad / Main Building* 19 B3
Harald Herlin -oppimiskeskus *Harald Herlin-lärcentret Harald Herlin Learning Centre* 15 B2
Haukilahden lukio *Gymnasiet / High school* 14 C2
Ice Tank 16 B1
Kandidaattikeskus *Kandidatcentret Undergraduate Centre* 1 B2
Kauppakorkeakoulu *School of Business* 68 B1
Kemian tekniikan korkeakoulu *Högskolan för kemiteknik School of Chemical Engineering* 11 C2
Kirjaamo / *Registratur / Registry* 1 B2
Kirjasto / *Bibliotek / Library* 15 B2
Konetekniikan talo (K1) 8 B2
KY 10 B1
Maarintalo / *Marhuset Maari Building* 17 A2
Magneettitalo / *Magnethuset Magnetic Building* 74 A3
Marsio* 71 B2
Meritekniikka *Marinteknik / Marine Technology* 16 B1
Mediaspace 15 B2
Micronova 36 B1
Nanotalo 9 B2
OIH Open Innovation House 32 B1
Pysäköintitalo* / *Multi-storey car park / Parkeringshus* 73 A2
Rantasauna 45 A3
Saha KY's Festival Hall 18 B1
Servin Mökki 44 A3
Space 21 12 C2
Startup Sauna 26 C3
Terveysteknologian talo Health Technology House 3 A2
TF Teknologföreningen 40 B3
Tietotekniikan talo *T-huset / CS Building* 30 B1
Urban Mill 26 C3
Vesitalo / *Vattenbyggnaden Water Building* 34 B1
Väre Taiteiden ja suunnittelun korkeakoulu / *Högskolan för konst, design och arkitektur / School of Arts, Design and Architecture* 67 B2

Download Aalto Space for the latest campus map

KATUOSOITTEET
Gatuadresser / Street addresses

Betonimiehenkuja 3 26 C3
Betonimiehenkuja 5 25 C3
Ekonominaukio 1 68 B1
Kemistintie 1 11 C2
Konemiehentie 1 18 B1
Konemiehentie 2 30 B1
Konemiehentie 4 10 B1
Maarintie 6 32 B1
Maarintie 8 37 B1
Metallimiehenkuja 2 59 C2
Metallimiehenkuja 4 27 C2
Otakaari 1 1 B2
Otakaari 2 A* 71 B2
Otakaari 2 B* 72 B1
Otakaari 3 3 A2
Otakaari 4 8 B2
Otakaari 5 4 A2
Otakaari 7 29 A3
Otakaari 8* 73 A2
Otakaari 11 38 B2
Otakaari 12 42 A2
Otakaari 22 40 B3
Otakaari 24 19 B3
Otakaari 27 20 B2
Otaniementie 9 15 B2
Otaniementie 12 69 B2
Otaniementie 14 67 B2
Otaranta 2 56 C3
Otaranta 4 41 B3
Otaranta 6 28 B3
Puumiehenkuja 2 9 B2
Puumiehenkuja 3 7 B2
Puumiehenkuja 5 6 A2
Rakentajanaukio 4 2 B2
Sähkömiehentie 3 17 A2
Sähkömiehentie 4 5 A2
Tekniikantie 3 14 C2
Tietotie 1 A 16 B1
Tietotie 1 E 34 B1
Tietotie 3 36 B1
Vuorimiehentie 1 13 C2
Vuorimiehentie 2 12 C2

Palvelut / Tjänster / Services

* **Rakennustyömaa** / *Byggplats / Under construction*

MUUT PALVELUT
Andra tjänster / Other services

Aalto Shop 67 B2
Aaltopahvi 15 B2
A Bloc Damask 69 B2
Attitude Helsinki 69 B2
AYY-asuntotoimisto *AUS Bostadskontoret / AYY Housing Office* 38 B2
Kuntokeskus Liikku / *Fitness club Liikku / Fitnessklubb Liikku* 69 B2
Laude Sauna & Terrace 4 A2
Otahalli 28 B3
Radisson Blu Hotel 56 C3
Terveystalo Otaniemi 20 B2
Unigrafia Noutopiste *Avhämtningsplats / Pick-up point* 1 B2
Unisport 28 B3
YTHS / SHVS / FSHS 42 A2

RAVINTOLAT
Restauranger / Restaurants

Arvo 68 B1
Eat Poke HKI 69 B2
Factory Otaniemi 4 A2
Fafa's 69 B2
Fat Lizard 34 B1
Konnichiwa 69 B2
MAU-KAS 57 C2
Metso 19 B3
NOVA MAU-KAS 36 B1
Ranta 56 C3
Silinteri 1 B2
Tasteat 12 C2
Who The xxxx is Alice 69 B2
X-burger 63 C2

OPISKELIJARAVINTOLAT
Studentrest. / Student restaurants

Alvari 1 B2
Food & Co A Bloc 69 B2
Kipsari 67 B2
Kvarkki 3 A2
Reima 19 B3
Studio Kipsari 29 A3
Subway 30 B1
Tietotekniikantalo 30 B1
TUAS 37 B1
Täffä 40 B3

KAHVILAT
Kaféer / Cafés

Aino 1 B2
Cafeport Otaniemi 67 B2
Cafetoria Aalto 15 B2
Café Factory 4 A2
Café Reima 19 B3
Elissa 1 B2
Espresso House 69 B2
Kylteri 68 B1
Taproom d20 20 B2

Apteekki / *Apotek / Pharmacy* B2
Campus Membership coworking-hubi / *Coworking-hubb / Coworking hub* B2, A2, C2
Liikuntapalvelut *Idrottstjänster / Sports* B2, B3
Pakettiautomaatti / *Paketautomat / Parcel pickup point* B2
Pankkiautomaatti / *Bankautomat / ATM* B2
Postilaatikko / *Brevlåda / Letter Box* B2
Päivittäistavarakauppa / *Mataffär / Grocery store* B2
Taksiasema / *Taxistation / Taxi* B2

aalto.fi

Poster Session

The poster session will be held on Wednesday, 01 Mar 2023, 18.00-20.00h.
The posters will be displayed next to the conference area (Kaleva).
Suggested poster sizes are A0 portrait or A1 landscape.
During the poster session, a warm buffet and drinks will be served.

We acknowledge the generous support of the Advanced Science journal (Wiley and Sons) towards the award of best oral and best poster contribution.

Meals

Coffee breaks will be provided free of charge.

See map of campus for Aalto University Campus convenient lunch options.

Internet Access

EDUROAM is available at Aalto University.

If you do not possess an EDUROAM account, use the open Aalto University Network.

Tutorials Lecture and Internet Access

The hands-on session will be best followed with a laptop. Every participant is suggested to bring their own device to this end. 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.

Supported by

Programme

		Otakaari 1 (U4)	Otakaari 1 (U1)	KALEVA	KALEVA	KALEVA
Start	**End**	**Monday 27.2.**	**Tuesday 28.2.**	**Wednesday 1.3.**	**Thursday 2.3.**	**Friday 3.3.**
09:00	09:20					
09:20	09:40				Adam Foster	Sami Kaski
09:40	10:00				Xin Yang	Ulpu Remes
10:00	10:20		Armi Tiihonen	Yuri Lysogorskiy	Leonard Ng	Saso Dzeroski
10:20	10:40				Panel Discussion + Coffee	Panel Discussion + Coffee
10:40	11:00		Coffee Break	Coffee Break		
11:00	11:20					
11:20	11:40				A. N. Dorabati	Martin Pimon
11:40	12:00					
12:00	12:20		Armi Tiihonen	Yuri Lysogorskiy	Gareth Conduit	Julija Zadlav
12:20	12:40					
12:40	13:00					Remarks + Prizes
13:00	13:20			Lunch Break		
13:20	13:40	Registration + Remarks		Registration + Remarks		
13:40	14:00		Lunch Break		Lunch Break	
14:00	14:20					
14:20	14:40			Philippe Schwaller	Kim Jelfs	
14:40	15:00			Simona Capponi	Venkat Kapil	
15:00	15:20	Bjørk Hammer	Matthew Evans	Rasmus Kronberg	Esa Puukko	
15:20	15:40			Coffee Break	Coffee Break	
15:40	16:00	Coffee Break	Coffee Break		Claudio Zeni	
16:00	16:20			Tonio Buonassisi	M.-C. Chang	
16:20	16:40			Kristian Berland	Joakim Lofgren	
16:40	17:00				Juha Koivisto	
17:00	17:20	Bjørk Hammer	Matthew Evans	Maria Chan	Shijing Sun	
17:20	17:40					
17:40	18:00					
18:00	18:20					
18:20	18:40					
18:40	19:00					
19:00	19:20					
19:20	19:40			Poster Session + Dinner		
19:40	20:00					

Legend	
	Intro Lectures
	Tutorials
	Invited Talks
	Contributed Talks
	Industry Talks
	Food Breaks
	Other

Monday

13:00 - 13:45	Registration
13:45 - 14:00	Opening Remarks
14:00 - 15:20	**Bjørk Hammer** Intro Lecture: Machine learning acceleration of global structure optimization
15:20 - 16:00	Coffee Break
16:00 - 17:20	**Bjørk Hammer** Hands-on : Machine learning acceleration of global structure optimization

Machine learning acceleration of global structure optimization

Bjørk Hammer [1]

[1] *Aarhus University, Aarhus, Denmark*

Optimizing atomistic structure in a density functional theory (DFT) framework can be accelerated by on-the-fly construction of a machine learning potential. The ML-potential relies initially on very few data, but eventually attains a quality comparable to the full DFT calculations. The uncertainty of the ML-potential may further be used to steer the search into unknown territory, thereby constituting an active learning setting. In the presentation and tutorial, elements of an automated structural search method [1] relying on a local descriptor energy expression [2] will be covered. Also, an algorithm [3] that uses image recognition and reinforcement learning will be described.

[1] https://aip.scitation.org/doi/full/10.1063/5.0094165
[2] https://aip.scitation.org/doi/full/10.1063/5.0121748
[3] https://aip.scitation.org/doi/full/10.1063/1.5108871

Tuesday

9:00 - 10:20	**Armi Tiihonen** Intro Lecture: Gaussian processes and Bayesian optimization
10:20 - 11:00	Coffee Break
11:00 - 12:20	**Armi Tiihonen** Tutorial Lecture: Gaussian processes and Bayesian optimization
12:20 - 14:00	Lunch Break
14:00 - 15:20	**Matthew Evans** Intro Lecture: Open databases integration for materials design (OPTIMADE).
15:20 - 16:00	Coffee Break
16:00 - 17:20	**Matthew Evans** Hands-on Lecture: Open databases integration for materials design (OPTIMADE).

Gaussian processes and Bayesian optimization

Armi Tiihonen [1]

[1] *Aalto University, Helsinki, Finland*

Probabilistic machine learning methods provide means for estimating the reliability of machine learning predictions, which is valuable in model deployment. In this tutorial, two such methods, Gaussian processes for regression and Bayesian optimization for active learning, are introduced with examples from materials science. You will get to test these concepts during the hands-on coding part of the tutorial.

Open databases integration for materials design (OPTIMADE)

Matthew Evans[1]

[1] *Universite de Louvain, Louvain, Belgium*

The availability of databases of crystal structures and associated measured or computed properties have enabled the proliferation of data-driven approaches to materials science. Increasingly, research in materials design and discovery requires automated or even autonomous access to such data, via application programming interfaces (APIs), be it for high-throughput screening, curating datasets for machine learning property prediction, or for benchmarking generative models. OPTIMADE is an API specification for exchanging materials data that has been designed and implemented by many leading crystal structure databases. By introducing a common API format for filtering, returning and structuring data, and by providing a decentralized registry of supported databases, OPTIMADE enables enhanced discoverability and data unification. By virtue of the participating databases, OPTIMADE provides uniform access to the most complete dataset of known and hypothetical inorganic crystal structures, albeit in a fractured and uncurated manner, which presents new challenges and opportunities. In this tutorial, the design of OPTIMADE will be introduced alongside the crystal structure databases that promoted its creation. Participants will be guided through open-ended exercises for accessing data via OPTIMADE, either with their own tools or those developed by the community, and suggestions will be provided as to how OPTIMADE could be incorporated into existing research workflows.

Wednesday

09:00 - 10:20	**Yury Lysogorskiy** Intro Lecture: Active learning for interatomic potentials development
10:20 - 11:00	Coffee Break
11:00 - 12:20	**Yury Lysogorskiy** Hands-on Lecture: Active learning for interatomic potentials development
12:20 - 13:40	Lunch + Registrations
13:40 - 14:00	Opening Remarks
14:00 - 14:40	**Philippe Schwaller** Bayesian optimisation for chemical reactions
14:40 - 15:00	**Simona Capponi** Batched Bayesian optimization for molecular structures
15:00 - 15:20	**Rasmus Kronberg** Machine learning and materials science on LUMI
15:20 - 15:40	Coffee Break
15:40 - 16:20	**Tonio Buonassisi** Inverse Design: Why Aren't We There Yet?
16:20 - 16:40	**Kristian Berland** Screening of low lattice thermal conductivity materials using active learning
16:40 - 17:20	**Maria Chan** Theory-informed AI for experimental data interpretation
18:00 - 20:00	Poster session + Dinner

Active learning for interatomic potentials development

Yury Lysogorskiy [1]

[1] *ICAMS, Ruhr University Bochum*

The development of any interatomic potential requires careful selection of training data. Much attention has been given to active learning in the development of interatomic potentials due to its ability to reduce the amount of data needed to accurately describe complex systems. An essential part of active learning is an Uncertainty Indication (UI) mechanism, which identifies regions at risk of inaccurate extrapolation. In this talk, we will consider various UI mechanisms for machine learning interatomic potentials and in more detail compare the performance of two approaches for UI of atomic cluster expansion models based on the D-Optimality criterion and ensemble learning. Additionally, we will discuss active exploration of new structures, which offers the potential for automated discovery of rare-event configurations.

Bayesian optimisation for chemical reactions

Philippe Schwaller [1]

[1] *Laboratory of Artificial Chemical Intelligence, Lausanne, Switzerland.*

Bayesian Optimisation is a powerful tool used in chemical reactions to optimise various parameters such as reaction conditions, catalysts, and reaction times to obtain the desired outcome. This optimization method uses probabilistic models and Bayesian inference to iteratively update its understanding of the system based on previous experimentation results, allowing for more efficient and effective use of limited resources. Bayesian optimisation has been shown to outperform traditional optimisation methods in terms of both efficiency and accuracy, making it a valuable tool in designing and optimising chemical reactions. I will give a brief overview of AI for chemical reactions, highlight key publications where Bayesian optimisation was applied to chemical reactions, discuss opportunities and challenges, and show preliminary results from the Schwaller Group at EPFL.

Batched Bayesian optimization for molecular structures

Simona Capponi [1]

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

Batch Bayesian optimization (BBO) has recently emerged as a successful tool for optimizing materials properties, because of its efficiency in optimizing expensive black-box functions, further accelerated by the possibility to evaluate these functions in parallel. However, while several frameworks have been proposed to solve problems defined in a continuous search space, the application of BBO to discrete spaces has been yet little investigated. To fill this gap we explore the applicability of LAW2ORDER, a state of the art BBO approach proposed by M.Welling and coworkers to solve optimization problems on permutations. The advantage of LAW2ORDER is that it is designed for large and discrete spaces, while it successfully balances information versus diversity in its batching policy, relying on a novel type of determinantal point processes. We test its performance when applied to the optimization of molecular properties. We report experimental results on a range of optimization tasks over a variety of organic molecular structures, which we encode with Coulomb matrices and with SOAP kernel descriptors. We also compare the performance of this method versus the more common BBO approach based on Thompson sampling, and we show that LAW2ORDER is always competitive with Thompson sampling across all the tasks.

Machine learning and materials science on LUMI

Rasmus Kronberg [1]

[1] *CSC – IT Center for Science Ltd, Helsinki, Finland*

LUMI [1] is a European pre-exascale supercomputer inaugurated in 2022 and hosted by CSC – IT Center for Science, Finland. Being the third fastest supercomputer in the world [2] with a hefty GPU-partition, LUMI enables unprecedented HPC workflows, especially within the domains of AI and high-performance data analytics. Materials science and computational chemistry applications ported to AMD GPUs will also greatly benefit from the performant MI250X accelerators, in addition to the auxiliary CPU-partition of more than 200k compute cores. LUMI provides thus a bleeding-edge computing environment for coupling AI with modern molecular dynamics and electronic structures codes.

In this contribution, I will present technical details about the LUMI supercomputer and highlight some pilot cases within AI and materials science that have already been run on LUMI [3,4]. The process of getting access to LUMI will be outlined as well, including how companies can utilize the computing resources of which 20% has been allocated to commercial users [5].

[1] https://www.lumi-supercomputer.eu/
[2] https://www.lumi-supercomputer.eu/lumi-again-among-the-fastest-and-greenest-supercomputers-in-the-world/
[3] https://www.lumi-supercomputer.eu/lumi-pilot-projects-selected/
[4] https://www.lumi-supercomputer.eu/lumis-second-pilot-phase-in-full-swing/
[5] https://www.csc.fi/solutions-for-business/

Inverse Design: Why Aren't We There Yet?

Tonio Buonassisi[1]

[1] *Massachusetts Institute of Technology , Cambridge, United States of America*

"Inverse design" describes the act of intentionally creating a material with a set of defined properties. It's accepted that only a small fraction of possible materials has been synthesized, and thus, an inverse-design tool can help navigate unexplored spaces more resource-efficiently than current approaches. Historically, inverse design was attempted using intuition and heuristics, later first-principles calculations including DFT, and more recently, machine-learning (ML) methods that leverage elements of the latter two as training data or constraints.In this seminar, I'll review the state of the art of ML-driven inverse design, and highlight outstanding gaps to be addressed. Highlights include novel materials representations that reflect underlying physics of elements and their diverse combinations, ML algorithms designed to capture and generalize patterns from known compounds, refinement approaches designed to eliminate "garbage-out" data, and early attempts to integrate inverse design into experimental workflows. I'll also share some of the first materials created using workflows that incorporate inverse design.

Screening of low lattice thermal conductivity materials using active learning

Kristian Berland [1]

[1] *Norwegian University of Life Sciences NMBU, Viken, Norway*

Identifying materials with low lattice thermal conductivity (LTC) is crucial for developing better thermoelectric materials, as many competing criteria must be fulfilled to obtain high efficiency. Screening with machine learning (ML) is attractive, but suffers from the scarcity of training data. Density functional theory (DFT) can provide additional training data but is limited by the high computational cost of LTC calculations. In a screening study of Half Heuslers, recently showed that identifying high value training data using principal component analysis [Comput. Mater. Sci. 202, 110938, 2022] can greatly accelerate ML training. To extend this screening to broader classes of cubic materials, we have adopted an active learning approach based on the uncertainty measures from Gaussian process regression (GPR), which was found superior to random sampling. The final model was trained on 268 LTCs obtained with TDEP, the largest training set generated so far, and used to predict the LTC of 1573 compounds. Several novel low-LTC compounds were identified.

Theory-informed AI for experimental data interpretation

Maria Chan [1]

[1] *Argonne National Laboratory, Lemont, United States of America*

In materials science, the combination of high throughput computational modeling and in situ/operando characterization experiments has given rise to significant challenges and opportunities. However, the integration of computational modeling with experimental measurements still often happens in a post-hoc and inefficient way. Data science techniques such as machine learning (ML), artificial intelligence (AI), and computer vision may be able to accelerate this integration. In this talk, we will discuss how we use AI/ML in conjunction with theory-based modeling to interpret experimental characterization data (such as x-ray scattering, spectroscopy, scanning probe microscopy, and electron microscopy). We will describe our FANTASTX (Fully Automated Nanoscale To Atomistic Structures from Theory and eXperiment) [1] and Ingrained [2] codes which allow such integration, as well as EXSCLAIM! [3] and Plot2Spectra [4] codes for extracting microscopy and spectroscopy data from scientific literature. The use of these codes to discover the structure of novel materials [5] as well as understand technologically important energy materials [6] will also be discussed.

[1] D. Unruh, V. S. C. Kolluru, A. Baskaran, Y. Chen, and Maria K. Y. Chan "Theory+AI/ML for Microscopy and Spectroscopy – Challenges and Opportunities", MRS Bulletin 2022.

[2] E. Schwenker, V. S. Chaitanya Kolluru, J. Guo, X. Hu, Q. Li, M. C. Hersam, V. P. Dravid, R. F. Klie, J. R. Guest, M. K. Y. Chan, "Ingrained: an automated framework for fusing atomic-scale image simulations into experiments," Small 18, 2102960 (2022).

[3] E. Schwenker, W Jiang, T Spreadbury, O Cossairt, MKY Chan, "EXSCLAIM!–An automated pipeline for the construction of labeled materials imaging datasets from literature" arXiv:2103.10631 under review.

[4] W. Jiang, K. Li, T. Spreadbury, E. Schwenker, O. Cossiart, M. K. Y. Chan, "Plot2Spectra: an Automatic Spectra Extraction Tool," Digital Discovery 1, 719-731 (2022).

[5] Q. Li, V. S. C. Kolluru, M. S. Rahn, E. Schwenker, S. Li, R. G. Hennig, P. Darancet, M. K. Y. Chan, M. C. Hersam, "Synthesis of borophane polymorphs through hydrogenation of borophene," Science, 371(6534), 1143-1148 (2021).

[6] X. Liu, G.-L. Xu, K. V. S. Kolluru, et al, "Origin and regulation of oxygen redox instability in high-voltage battery cathodes," Nature Energy 7, 808–817 (2022).

Thursday

09:00 - 09:40	**Adam Foster** Machine learning in scanning probe microscopy
09:40 - 10:00	**Xin Yang** Investigating oxygen reduction kinetics at gold-water interface using machine learning potentials
10:00 - 10:20	**Leonard Ng** High-throughput, single-experiment optimisation of roll-to-roll fabricated non-fullerene acceptor photovoltaics using machine learning
10:20 - 11:20	Panel discussion and coffee
11:20 - 11:40	**Amirhossein Naghdi Dorabati** Neural network interatomic potentials for nanoindentation MD simulations
11:40 - 12:20	**Gareth Conduit** The modern-day blacksmith
12:20 - 14:00	Lunch
14:00 - 14:40	**Kim Jelfs** Remembering the lab in computational molecular material discovery
14:40 - 15:00	**Venkat Kapil** The first-principles phase diagram of monolayer nanoconfined water
15:00 - 15:20	**Esa Puukko** Digital manufacturing in stainless steel industry
15:20 - 15:40	Coffee break
15:40 - 16:00	**Claudio Zeni** Exploring the robust extrapolation of high-dimensional machine learning potentials
16:00 - 16:20	**Ming-Chiang Chang** Integrated autonomous and user-guided active learning for targeted material synthesis
16:20 - 16:40	**Joakim Löfgren** Bayesian optimization for experimental materials design
16:40 - 17:00	**Juha Koivisto** Mapping fluid state properties of biofoams to solid state using Bayesian optimization
17:00 - 17:20	**Shijing Sun** Application of machine learning in EV battery R&D

Machine learning in scanning probe microscopy

Adam Foster [1]

[1] *Aalto University, Aalto, Finland*

Scanning Probe Microscopy (SPM) has been the engine of characterization in nanoscale systems in general, and the evolution of molecule-functionalized tips as a reliable tool for high-resolution imaging without material restrictions has been a breakthrough in studies of molecular systems [1]. In parallel, machine learning (ML) methods are increasingly being applied to data challenges in SPM. In particular, the success of deep learning in image recognition tasks has led to their application to the analysis of SPM images, especially in the context of surface feature characterisation and techniques for autonomously-driven SPM [2].

In this work, we explore the potential for using deep learning to aid in the interpretation of high resolution Atomic Force Microscopy (AFM) images. Building upon a deep learning infrastructure that matches a set of AFM images with a unique descriptor characterizing the molecular configuration [3], we further develop methods for electrostatic characterization [4] and direct atomic structure prediction with Graph Neural Networks [5]. In a challenging test of the approach, we combine ML structure predictions from AFM images of ice nanostructures with neural network potentials to try to predict complex, unknown geometries. Alongside this, we show how ML approaches can be used actively during SPM experiments to aid in both tip functionalization [6] and in the construction of nanostructures through atomic manipulation [7].

[1] N. Pavliček and L. Gross, Nat. Rev. Chem. 1 (2017) 5
[2] O.M. Gordon and P.J. Moriarty, Mach. Learn.: Sci. Technol. 1 (2020) 023001
[3] B. Alldritt, P. Hapala, N. Oinonen, F. Urtev, O. Krejci, F. F. Canova, J. Kannala, F. Schulz, P. Liljeroth, and A. S. Foster, Sci. Adv. 6 (2020) eaay6913
[4] Niko Oinonen, Chen Xu, Benjamin Alldritt, Filippo Federici Canova, Fedor Urtev, Ondřej Krejčí, Juho Kannala, Peter Liljeroth and Adam S. Foster, ACS Nano 16 (2022) 89
[5] Niko Oinonen, Lauri Kurki, Alexander Ilin and Adam S. Foster, MRS Bulletin 47 (2022)
[6] Benjamin Alldritt, Fedor Urtev, Niko Oinonen, Markus Aapro, Juho Kannala, Peter Liljeroth and Adam S. Foster, Comp. Phys. Comm. 273 (2022) 108258
[7] I-Ju Chen, Markus Aapro, Abraham Kipnis, Alexander Ilin, Peter Liljeroth and Adam S. Foster, Nat. Commun. 13 (2022) 7499

Investigating oxygen reduction kinetics at gold-water interface using machine learning potentials

Xin Yang [1]

[1] *Technical University of Denmark*

Equivariant graph neural network (GNN) based accurate surrogate potentials can accelerate the speed of performing ab-initio molecular dynamics (AIMD) after learning on representative structures in a data efficient manner. In this study, we combined uncertainty- aware GNN potentials and enhanced sampling to investigate the reactive process of the oxygen reduction reaction (ORR) at Au(100)-water interface. By using a well-established active learning framework based on CUR matrix decomposition, we can evenly sample equilibrium structures from MD simulations and non-equilibrium reaction intermediates that are rarely visited during the reaction. The trained GNNs have shown exceptional performance in terms of force prediction accuracy, the ability to reproduce structural properties, and the low uncertainties when performing MD and metadynamics simulations. Furthermore, the collective variables employed in this work enables an automatic search of reaction pathways and provide a detailed understanding towards the ORR reaction mechanism on Au(100). Our simulations identify an associative reaction mechanism where adsorbed O2 reacts with water to form hydroxyls through an *OOH transition state. The reaction proceeds without formation of *O with a low reaction barrier of 0.3 eV. The low barrier agrees with the fast reaction kinetics observed experimentally. The methodology employed in this study can pave the way for modeling complex chemical reactions at electrochemical interfaces with an explicit solvent at ambient conditions.

High-throughput, single-experiment optimisation of roll-to-roll fabricated non-fullerene acceptor photovoltaics using machine learning

Leonard Ng [1]

[1] *Nanyang Technological University, Singapore, Singapore*

Roll-to-roll (R2R) printed photovoltaics (PVs) can reduce solar energy costs with their flexible form factors and their manufacturing footprint. Here we demonstrate a custom-built 'MicroFactory', a high-throughput fabrication, characterisation and data collection research platform that leverages on machine and deep learning (ML/DL) techniques for R2R printed PVs. Consisting of a R2R fabricator, characteriser, and a cloud database, we use the MicroFactory to fabricate, characterise and record manufacturing parameters of 11,800 non-fullerene acceptor (NFA) organic photovoltaic devices (OPV) devices in a single experiment. This rich dataset allows us to run 17 ML/DL models. Using the best performing model, chosen using Euclidean distance, we make predictions on 1,000,000 simulated devices to extract four donor:acceptor (D:A) formulas. We use these formulas to fabricate 2,400 enhanced devices which produced a peak power conversion efficiency (PCE) of 9.35%, a 87% improvement from a previous record efficiency of 5% for printed R2R PVs.

Neural network interatomic potentials for nanoindentation MD simulations

Amirohossein Naghdi Dorabati [1]

[1] *Nomaten CoE, Otwock, Poland*

Numerical investigations of nano-mechanical testing for metals and alloys require accurate interatomic potentials that may predict configurational energies and interatomic forces, consistent with ab initio calculations. In this work, we investigate crystalline molybdenum (Mo), a viable candidate for extreme environments, functioning at elevated temperatures. Mechanical properties of Mo, such as nanoindentation hardness, display non-trivial temperature dependence that requires further validation and deeper understanding, beyond classical force fields methods. In this work, we create a Neural-network interatomic potential (NNIP) for nanoindentation of pure crystalline Mo to investigate mechanisms of dislocation nucleation and evolution at multiple temperatures, up to 1000K. Compared to common machine learned potentials (MLPs) in the literature and by employing a similarity measure between an indented sample and MLP datasets, we found very relevant configurations to add to common MLP datasets that were missed before. Elastic constants, dislocation densities, strain maps and slip traces as a function of indentation depth of the system are compared with embedded atom method (EAM) potentials and the advantages and limitations of NNIPs over traditional potentials are reported.

The modern day blacksmith

Gareth Conduit [1]

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

We present a machine learning methodology, Alchemite, that uses active learning to request and exploit multiple sources of materials information spanning experimental data to computer simulations. Alchemite judiciously requests only those computations and experimental properties required, resulting in a sparse data set, and then exploits property-property correlations to impute whilst training the final model.

We illustrate the approach with two case studies: Firstly, we design and experimentally verify a nickel-base superalloy for direct laser deposition. Starting from a training set comprising just eight results, the machine learning tool juxtaposes complementary material properties to circumvent missing data [1]. Secondly, we use Alchemite to guide the commissioning of a new additive manufacturing facility. Alchemite proposes which tests should be performed, in order to get the facility up and running as quickly as possible. Finally, we discuss further case studies that highlight the breadth of applications of the generic approach.

[1] Materials Design 168, 107644 (2019)

Remembering the lab in computational molecular material discovery

Kim Jelfs [1]

[1] *Imperial College London, London, United Kingdom*

We have been developing computational software towards assisting in the discovery of molecular materials with targeted structures and properties. Whilst initially we have focused upon porous molecular materials, we will also address the ways in which our approach is generalisable to other molecular materials and their applications, including as organic semiconductors or for photocatalysis. Our evolutionary algorithm automates the assembly of hypothetical molecules from a library of precursors. Our approach has already suggested promising targets that have been synthetically realised. We have also examined the application of both supervised machine learning and explainable graph neural networks for the rapid prediction of porous molecules' properties. Finally, we have trained a model (the Materials Precursor Score, MPScore) to guide our predictions to select materials that have a high chance of being synthesisable in the laboratory.

The first-principles phase diagram of monolayer nanoconfined water

Venkat Kapil [1]

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

Water in nanoscale cavities is ubiquitous and of central importance to everyday phenomena in geology and biology. However, the properties of nanoscale water can be substantially different from those of bulk water, as shown, for example, by the anomalously low dielectric constant of water in nanochannels, near frictionless water flow, or the possible existence of a square ice phase, and highly-relevant for technological applications in nanofluidics, electrolyte materials, and water desalination. Here we alleviate conventional challenges in first-principles modeling by combining quantum Monte Carlo and an active learning scheme to develop a machine-learning potential across the phase diagram of nanoscale water. We perform advanced quantum and classical statistical simulations to estimate the first published first-principles phase diagram of monolayer water in a graphene-like channel. In addition to multiple molecular phases with melting temperatures varying non-monotonically by more than 400 kelvins with pressure, we predict a hexatic phase, which is an intermediate between a solid and a liquid, and a superionic phase with a high electrical conductivity exceeding that of battery materials. This suggests that nanoconfinement could be a promising route toward superionic behavior under easily accessible conditions.

[1] Kapil et. al "The first-principles phase diagram of monolayer nanoconfined water", Nature, 2022

Digital manufacturing in stainless steel industry

Esa Puukko [1]

[1] *Outokompu Oy, Helsinki, Finland*

This presentation demonstrates how one steel plant has improved its capability to use data more efficient and thereby gain business benefits. User cases from different needs and aspects will be highlighted. Challenges to execute digitalization roadmap will be openly shared for fruitful discussion.

Exploring the robust extrapolation of high-dimensional machine learning potentials

Claudio Zeni[1],

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

We show that, contrary to popular assumptions, predictions from machine learning potentials built upon high dimensional atom-density representations almost exclusively occur in regions of the representation space which lie outside the convex hull defined by the training set points. We then propose a perspective to rationalize the domain of robust extrapolation and accurate prediction of atomistic machine learning potentials in terms of the probability density induced by training points in the representation space.

Integrated autonomous and user-guided active learning for targeted materials synthesis

Ming-Chiang Chang [1]

[1] *Cornell University, Ithaca, United States of America*

Recently, incorporating active learning algorithms into material experiment workflows has been shown to be an effective way to accelerate the exploration of complex material design space. In previous studies, we developed the scientific autonomous reasoning agent (SARA) which combines automated non-equilibrium laser-spike annealing process, optical material characterization and Gaussian process with specially designed kernel into a close-loop autonomous cycle to efficiently construct processing phase diagrams using the gradient in optical signals. However, like many active learning methods, the input characterization data was not preprocessed to extract physically meaningful values for the AI agent to do physical reasoning. In this study, combining SARA's infrastructure and a new rapid and probabilistic XRD phase labeling algorithm, we demonstrate that by switching between exploration and exploitation mode, the autonomous AI agent can not only explore the phase space efficiently but also utilize the dense XRD information to search for the condition that can synthesize highly phase-purity structures. In the future, with such physical input, various acquisition functions can be design for the AI agent to pursuit different design goal and opens the opportunity for on-the-fly modeling of structure-to-property relationships and subsequently doing efficient multi-property optimization.

Bayesian optimization for experimental materials design

Joakim Löfgren [1]

[1] *Aalto University, Aalto, Finland*

A key challenge in materials science is to optimize the outcome of an experimental process (e.g., material properties, yield) in terms of input variables (e.g., precursor composition, environmental conditions). To solve such problems, Bayesian optimization (BO) has emerged as a powerful approach that guides the experimental data collection in an adaptive manner such that optimal outcomes and input variables can be identified from as few experiments as possible.

In this talk I highlight our recent efforts in promoting more widespread adoption of BO in experimental communities by increasing the accessibility in terms of software, as well as provide solutions for complex tasks such as batch acquisitions and multiple objectives. We apply these tools to a problem in sustainable biochemical engineering where lignin, an important biopolymer, is extracted from birch wood using a hydrothermal treatment method. I show how BO can be used to quickly identify optimal processing conditions for obtaining lignin that satisfies multiple optimization objectives: high yield and specific chemical structure. This approach allows the extracted lignin to be tailored for specific applications. I furthermore discuss how to make informative batch acquisitions to increase the efficiency of the data collection.

Mapping fluid state properties of biofoams to solid state using Bayesian optimization

Juha Koivisto [1]

[1] *Aalto University, Aalto, Finland*

Understanding of the viscoelastic behavior of a polymer is a prerequisite for its thermomechanical processing beyond laboratory scale to produce solid foams. Utilizing rheological and mechanical characterization is a powerful tool to comprehend the complex nature of macromolecular materials. Nevertheless, it consumes time as multiple experiments under several conditions are needed to visualize the non-linear behavior of materials under varying conditions. The work hereunder examines the fluid state rheology of cellulosic aqueous suspensions prepared using cellulose fibers as the dispersed phase (Refcell and Storacell) and methylcellulose (MC) as the polymeric matrix. Interfacial phenomena between MC and cellulose fibers arise in particle laden systems with supramolecular structures formed by non-covalent interactions dependent on the concentrations of raw materials. Changing the ingredients of the suspension changes both the rheological properties of the fluid as well as the mechanical properties of the dried foam. Here, I will show that the number of experiments can be reduced and still obtain a reliable mapping from ingredients to fluid state properties to solid state properties utilizing Bayesian optimization with Gaussian processes.

Application of machine learning in EV battery R&D

Shijing Sun [1]

[1] *Toyota Research, United States of America*

Recent advances in battery technology have attracted enormous attention from both academia and industry. This talk will focus on how machine learning techniques can be used to help researchers better understand battery degradation and to accelerate aging tests. The BEEP platform developed at Toyota Research Institute automates battery cycling experiments and subsequent data featurization. The system can make early outcome predictions to reduce testing time and conduct closed-loop experiments via Bayesian optimization to minimize the total number of tests. More recently, machine learning-based methods for state of health evaluation, prediction, and optimization have been researched leveraging the BEEP platform, enabling rapid iterations of battery development.

Friday

09:00 - 09:40	**Samuel Kaski** Collaborative AI for assisting virtual laboratories
09:40 - 10:00	**Ulpu Remes** Multi-fidelity Bayesian optimization structure search
10:00 - 10:20	**Sašo Džeroski** Semi-supervised and active multi-target regression for material design
10:20 - 11:20	Panel discussion and coffee
11:20 - 11:40	**Martin Pimon** Electrons go nuclear: a unique interaction with Thorium-229
11:40 - 12:20	**Julija Zavadlav** Deep molecular modeling: the impact of training objective
12:20 - 12:40	Closing remarks and prize announcement

Collaborative AI for assisting virtual laboratories

Samuel Kaski [1]

[1] *Aalto University, Aalto, Finland*

I will discuss two ideas: (1) virtual laboratories for science and RD, aiming to introduce an interface between algorithms and domain research that enables AI-driven scale advantages, and (2) AI-based 'sidekick' assistants. The purpose of the assistants is to help other agents reach their goals, even when they are not yet able to specify the goal explicitly, or it is evolving. Such assistants can help with prior knowledge elicitation, at the simplest, and zero-shot assistance as the worst case. Ultimately they should be helpful for human domain experts in running experiments and solving research problems in virtual laboratories. I invite researchers to join the virtual laboratory movement: domain scientists by hosting a virtual laboratory in their field, methods researchers by contributing new methods to virtual laboratories, and human-in-the-loop ML researchers by developing the assistants.

Multi-fidelity Bayesian optimization structure search

Ulpu Remes [1]

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

Bayesian optimization (BO) is an active learning method for optimising unknown functions that are expensive to evaluate. BO employs a statistical model to make smart decisions on sampling the search space and can combine multiple information sources to accelerate optimization. For example, multi-fidelity BO (MFBO) methods combine expensive target function evaluations with cheaper and less accurate support function evaluations. MFBO methods considered in this work include i) target function optimization initialised with support function data collected in a separate optimization run and ii) multi-task optimization that samples both the support and the target function.

We evaluate MFBO in a structure search task where the information sources are total energy calculation methods used in computational materials science. Total energy calculation can be performed at different chemical accuracy levels, whereby increased accuracy usually implies increased computation time. Here we use molecular mechanics and density-functional theory. We use an intrinsic coregionalization model to capture correlation between energy calculations, and our results demonstrate how high correlation between energy calculations can be exploited to dramatically reduce BO computational costs in a molecular conformer search task.

Semi-supervised and active multi-target regression for material design

Sašo Džeroski [1]

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

Machine learning (ML) methods for multi-target regression (MTR), i.e., MTR trees and tree ensembles, have been recently used in material design. They have learned models that simultaneously predict multiple material properties from material composition or from processing parameters. They have been used in fully supervised mode, where all the target values of all the training examples are known, i.e., all examples are labeled. In practice, unlabeled data are often available and can be beneficial to the learning process: ML can use the data unlabeled (semi-supervised learning) or can ask the user to label some of the examples, which it expects will most improve the learned model (active learning). Recently, semi-supervised variants have been developed of tree-ensemble approaches for MTR, based on the self-training approach and on the approach of predictive clustering. We plan to apply these two semi-supervised MTR approaches to versions of the datasets already addressed, extended with unlabeled data. We also plan to adapt the self-training approach towards active learning. Namely, it includes approaches to estimating the reliability of its predictions: These estimates can be used to focus on the least reliable predictions that should be labeled in active learning.

Friday

Electrons go nuclear: a unique interaction with thorium-229

Martin Pimon [1]

[1] *Technical University of Vienna, Vienna, Austria*

The thorium-229 isotope features an exceptionally low lying nuclear level in the range of typical electronic energies. In wide band gap materials, a peculiar electron-nucleus interaction can be explored, where a correct physical description of such a system can only be achieved by combining central approaches of quantum chemistry, condensed matter physics and nuclear physics. In order to estimate energy transfer rates between thorium nuclei and the electronic structure, an accurate study of excited state properties using many-body approaches is required. Finding optimized materials, which tailor this electron-nucleus coupling, would enhance the interaction rate by orders of magnitude compared to current experimental designs. However, the substantial computational demands and the large composition space prohibit a conventional high-throughput approach. Actively learning the coupling strength based on key features of the electronic environment has enormous potential to accelerate the search of such optimized materials. An immediate application is the improvement of the measurement accuracy in a novel solid-state "nuclear clock".

Deep molecular modeling: the impact of training objective

Julija Zavadlav [1]

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

Molecular modeling has become a cornerstone of many disciplines, including material science. However, the quality of predictions critically depends on the employed model that defines particle interactions. A class of models with tremendous success in recent years are neural network (NN) potentials due to their flexibility and capacity to learn many-body interactions. However, their accuracy critically depends on the training data and objective. Traditionally, these models are trained bottom-up via force matching. We recently showed that training based on relative entropy results in a more accurate potential energy surface and requires less data. Nevertheless, bottom-up data is frequently unavailable or insufficiently accurate. Top-down approaches are advantageous as the NN potential is learned directly from experimental data. We developed the Differentiable Trajectory Reweighting (DiffTRe) method, bypassing differentiation through the MD simulation. Leveraging thermodynamic perturbation theory, we avoid exploding gradients and achieve around two orders of magnitude speed-up in gradient computation for top-down learning. The effectiveness of DiffTRe is showcased on diverse molecular systems and modeling scales.

Posters

Mahmoud Attia, French Atomic Energy and Alternative Energies Commission -

Designing solid-state batteries (SSBs) requires the design of highly efficient solid electrolytes (SSEs) that exhibit high ionic conduction properties. Though they have numerous drawbacks, liquid organic-based electrolytes are the ones most commonly used in commercial batteries. They are not only dangerous and less efficient but also degrade faster which decreases their lifetime for long-term applications. Hence the necessity for a safer and longer-lived alternative for batteries is a must. To this end, SSEs are now at the forefront of all materials, as they combine factors of safety (non-flammable), efficiency, lower cost, and ease of fabrication. Among them, doped – LLZO (Li7La3Zr2O12) is considered of the most promising ceramics serving this purpose. Herein, we aim to model Li diffusion mechanisms and their respective spectroscopic properties (NMR and EIS) using Multiscale simulation approaches by harnessing both (i) short-time and length scales (from DFT and AIMD3) and (ii) long-time and length scales (from classical MD and KMC4) to define a multi-scale methodology for understanding the impact of doping have on Li mobility in the material in question by linking the simulations data to the microscopic diffusion measurements (NMR and EIS). Our aim is to develop a Kinetic Monte Carlo model, parameterized with inputs from MD simulations (both classical and ab-initio) capable of predicting NMR properties such as 7Li NMR lineshapes and nuclear relaxation times. Density-based clustering approaches are employed to analyze Lithium ion trajectories. Additionally, standard DFT-GIPAW calculations are performed for predictions of MAS-NMR5 spectra of 27Al in Al–doped LLZO (LLAZO). By adopting the same methodology, Na diffusion in NASICON6 and Sc+3 – doped NASICON has been investigated and analyzed.

Nitik Bhatia, Aalto University - Accelerating the accurate description of functionalized gold nanoparticles with delta machine learning

The complexity of gold nanoparticles (GNPs) and their interaction with solvents and biological systems makes the application of accurate density functional theory (DFT) or post-Hartree-Fock methods intractable. Instead, cheaper but less accurate semi-empirical methods like the density functional tight-binding model (DFTB) are often used. In this work, we have used Δ-machine learning (Δ-ML) to bridge between DFTB and DFT to make DFTB more accurate without losing its computational advantage. Our ΔML method employs a Behler-Parrinello neural network and atom-centered symmetric functions (ACSFs) as descriptors to describe local atomic environments [1]. The neural network is trained on Au79 GNP with one adsorbed water molecule (3000 geometries) to learn and predict the energetics of their interaction. The (Δ-ML)-based DFTB method predicts the interaction energy 4-5 times faster than DFT with a MAE just under 0.8 kcal/mol. Furthermore, the model is transferable to other cluster sizes (Au39 and Au55) with an accuracy of 1.0 kcal/mol. However, a systematic correction term has to be applied.

Behler, J. (2011). The Journal of Chemical Physics, 134(7), 074106.

Jari Järvi, Aalto University - Identifying molecular adsorbates on electronically decoupled graphene with Bayesian inference

Identifying stable configurations of molecular adsorbates requires thorough exploration of the adsorption energy surface (AES) with ab initio energy sampling. This is prohibitively expensive for density-functional theory (DFT) using conventional AES exploration methods. Bayesian Optimization Structure Search (BOSS) [1] is a new active learning method that combines DFT with Bayesian inference for accurate global structure search. BOSS applies strategic sampling to compute the complete AES with a small number of costly DFT calculations and thus facilitates the identification of stable adsorbate structures.
We apply BOSS to study the adsorption of two prototypical charge transfer molecules, F4TCNQ and TTF, on electronically decoupled graphene, Gr/O/Ir(111) [2]. We accelerate DFT energy sampling by approximating the complex substrate with charged freestanding Gr. We identified the most stable single-molecule configurations, which are candidate structures for thin-film charge transfer complexes. We extended the BOSS search to molecule pair configurations and analyzed their intermolecular interactions to gain insight on film morphologies. This study demonstrates the power of novel active learning methods in detecting molecular adsorbates.
[1] M. Todorović et al., npj Comput. Mater. 5, 35 (2019) ; [2] J. Järvi et al., Phys. Rev. B 105, 195304 (2022).

Nima Emami, University of Turku - Machine Learning Optimzation of Materials Flow in Battery Recycling

Significant activities are dedicated to researching new battery recycling processes, including efforts to implement emerging technologies claiming a broader recovery of materials at an industrial scale. With the recent rise in capabilities of AI and Machine learning models, the new recycling processes can be further optimized to achieve better recovery of materials. Machine learning models can be implemented in the design phase of processes to optimize each step in the process and the whole process in general.

Eduard Hahn, University of Hamburg - Predicting Exchange Spin Coupling within Dinuclear Copper Complexes using Machine Learning

Property tailored machine learning models based on Gaussian process regression have proven capable of predicting the strength of the exchange spin coupling within subsets of dinuclear copper complexes while requiring only a fraction of the computational cost at accuracies close to the reference calculations obtained by utilizing density functional theory (DFT) [1]. For the transformation of the molecular structure into machine-readable input, a descriptor based on chemical intuition is utilized, containing features known to correlate with the coupling constant, like specific bond lengths, the metal-bridge angles, and the specification of outer ligands [2]. This problem-tailored low-dimensional descriptor already outperforms several sophisticated off-the-shelf high-dimensional descriptors while greatly reducing the model's training time. Furthermore, the descriptor enables the straightforward implementation of a multi-layer machine learning system that deals with a classification/clustering problem (sorting structures by their bridge type) on the one hand, and the actual regression task (local structure-based training) on the other hand. In addition, the local machine learning models can be trained by exploiting structure-dependent descriptor modifications (bridge-type optimized feature combination/selection) allowing structures with significantly different contributions of through-bond coupling and through-space coupling to be treated differently, further improving the predictive performance.
[1] P. Bahlke, N. Mogos, J. Proppe, C. Herrmann, J. Phys. Chem. A 2020, 124, 42,

8708–8723; [2] P. J. Hay, et al. , J. Am. Chem. Soc. 1975, 97, 4884–4889.

Kameyab Raza Abidi, University of Jyväskylä - Optimizing DFT to study two-dimensional metals

The 2D metals are identified as non-layered structures in contrast to the covalent 2D materials like graphene. 2D metals are beginning to gain traction at a time when an assortment of density functional theory (DFT) based methods can be implemented to determine the properties and assist in design. To determine what level of DFT attributes are needed to study 2D metals, we systematically scanned the various DFT methods. A wide range of exchange-correlation (xc) functionals were examined, ranging from LDA to GGA to hybrid. To capture the effect of dimensionality one-dimensional (1D) chain, 2D (honeycomb, square, and hexagonal), and 3D systems are taken into consideration. The performance of xc functionals was compared by translating energetics into coordination number (C) and assessing how they approach systems in different dimensions. This work aids in the methodological developments for theoretical modelling of 2D metals which will facilitate the interaction between theory and experiment. The results have the potential to provide insight into choosing which DFT methods would be the most appropriate for studying 2D metals. For data-driven discoveries of 2D metals, the time-scale comparison of each method revealed the optimal DFT method for generating enormous DFT data for incorporating machine learning algorithms.

Jesper Byggmästar, University of Helsinki - Simple machine-learned potentials for high-entropy alloys

Development of machine-learned interatomic potentials for materials is typically best done with a delicate balance between human expertise and active learning. I will discuss the development of accurate machine-learned potentials for refractory high-entropy alloys, i.e. alloys consisting of many refractory metals (e.g. W, Mo, Ta) which leads to many unique or enhanced material properties. The potential uses Gaussian process regression with simple low-dimensional descriptors, resulting in computationally fast potentials that enables molecular dynamics simulations of millions of atoms for tens of nanoseconds. Our work also demonstrates that for many-element materials, using simple descriptors is enough to reach competitive accuracy, and even outperforms standard high-dimensional descriptors. The machine-learned potential allows us to investigate novel properties of defects, short-range ordering, and segregation in the MoNbTaVW high-entropy alloy.

Akshay Krishna Ammothum Kandy, Centre for Materials Elaboration and Structural Studies CNRS - Phi-LassoLars machine learning potentials for materials modelling

Machine-learning interaction potentials (MLIP) have been recently proposed to bridge the gap between quantum accurate calculations and fast empirical modelling. The main principle consists in using a large set of quantum-accurate calculations to adjust the parameters of a universal mathematical formulation that should represent the interaction potential. Lots of different approaches have been proposed including Artificial Neural Networks, Gaussian approximation potentials, Linearealized potentials, Spectral Neighbor Analysis Potentials, Symmetric Gradient Domain Machine learning and Moment Tensor Potentials. However, the current MLIPs are based on local environment descriptors and often lack physically relevant long-ranged electrostatic interactions. In this contribution, we will present an alternative machine-learning method that we are currently developing called the Electrostatic Physical LassoLars Interaction Potential (EPLIP). It combines a physically motivated mathematical formulation for the potential along with long-ranged Coulomb descriptors with a

scaled point charge approach. We use Zinc oxide as an example to illustrate the advantages of the new method.

Jyothika Pillay, Friedrich Schiller Universität Jena - Understanding the Interaction of Organic Molecules and Metal Ions by Molecular Machine Learning

Transition metal or metal ion interactions with organic molecules or inorganic ions lead to the formation of transition metal complexes. These complexes have great importance and are used in a variety of fields, from medicine to energy conversion. The large number of metal ions as well as organic or inorganic ligands opens up a vast field of potential metal-ligand complexes, each with their own distinct properties. This project explores the application of supervised machine learning (ML) algorithms for investigating the formation of metal-ligand complexes and their properties. Experimentally, nuclear magnetic resonance (NMR) chemical shifts, being a direct consequence of the local chemical environment of the atoms, are especially suitable to establish such relationships. In the first stage, we use ML to predict the molecular properties of several complexes including NMR shifts using the chemical structure of the ligand and the sort of metal ion. To this end, we rely on structures obtained from the transition-metal quantum mechanical (tmQM) dataset and complement it with results from GIAO calculations. Eventually, our results might allow to identify correlations in the NMR data and to facilitate cross-correlating theoretical and experimentally obtained data.

Alfonso Gallo, CIC Energigune - Is my crystal too distorted? Unsupervised ML to classify Li-argyrodite solid-state electrolytes wrt geometrical distortion

High-throughput approaches in computational materials discovery often yield a combinatorial explosion that makes the exhaustive rendering of complete structural and chemical spaces impractical. A common bottleneck when screening new compounds with archetypal crystal structures is the lack of fast and reliable decision-making schemes to quantitatively classify the computed candidates as inliers or outliers (too distorted structures). Machine learning-aided workflows can solve this problem and make geometrical optimization procedures more efficient. However, for this to occur, there is still a lack of appropriate combinations of suitable geometrical descriptors and accurate unsupervised models which are capable of accurately differentiating between systems with subtle structural changes. Here, considering as a case study the compositional screening of cubic Li-argyrodites solid electrolytes, we tackle this problem head on. We find that Steinhardt order parameters are very accurate descriptors of the cubic argyrodite structure to train a range of common unsupervised outlier detection models. And, most importantly, the approach enables us to automatically classify crystal structures with uncertainty control. The resulting models can then be used to screen computed structures with respect to an user-defined error threshold and discard too distorted structures during geometrical optimization procedures. Implemented as a decision node in computer-aided materials discovery workflows, this approach can be employed to perform autonomous high-throughput screening methods and make the use of computational and data storage resources more efficient.

Dario Massa, NCBJ, NOMATEN CoE - Alloy informatics using charge profile for hydrogen energy applications

We propose 'alloy informatics' as a prototype predictive approach of defect related properties in bulk crystals, based on charge profiles derived from first-principle calculations. We demonstrate this framework in the context of hydrogen defects in FCC crystals. Ab-initio analysis of differential electron charge density profiles in pure FCC metallic bulk crystals with and without hydrogen are performed. Radial density Distribution Functions (RDF) of defect-induced differential charge density perturbations effectively capture defect-crystal interaction properties. The correlation analysis of atomic-structural-density datasets characterizes the properties of hydrogen differential charge densities and RDF profiles, as well as the bulk metallic crystal. We demonstrate through Nudge-Elastic-Band calculations, that the charge spatial extensions are intimately correlated with hydrogen migration energy barriers. Finally, combined unsupervised machine learning techniques unveil the emergence of behavioral classes, and the role of each original feature in driving their memberships and distributions.

Felix Arendt, Friedrich-Schiller University Jena - Construction of Machine Learning descriptors for modeling glass properties

There are three sources of information for modelling glasses: composition, structure, and processing parameters. Glass composition is the most readily available information, so it is desirable to extract as much performance from it as possible. An approach to this is ab initio descriptors. Here ab initio derived descriptors refer to mean values of the ab initio calculated properties of the glass components, weighted according to their respective mole fraction. Ab initio descriptors can reproduce the performance of their compositional counterparts for a large, comprehensive data set, and in some cases even improve it for a smaller data set covering a variety of oxides, chalcogenides, and metallic glasses.
Subsequently, the glass structure is used as an additional source of information to improve the predictive performance of the ML models. High-throughput molecular dynamics simulations of glasses are conducted. By treating the relaxed cells as 3D images, we intend to apply image recognition methods such as beta-variational autoencoders to extract the structural information. Thus, for this combined approach of novel descriptors and a robotic melting facility, active learning could be a valuable tool for efficiently allocating robotic and simulation resources for accelerated glass design.

Lincan Fang, Aalto University - Exploring the Conformers of an Organic Molecule on a Metal Cluster with Bayesian Optimization

Finding low-energy conformers of organic molecules is a complex problem due to their flexibility and the high dimensionality of the search space [1]. When such molecules decorate nanoclusters, the search complexity is exacerbated by the presence of the cluster and other surrounding molecules. To address this challenge, we modified our previously developed active learning molecular conformer search method based on Bayesian optimization [1,2] and density functional theory. Especially, we have developed and tested strategies to avoid the steric clash between the molecule and the cluster, which enables us to sample the entire, available phase space. In this work, we chose a cysteine molecule on a well-studied gold-thiolate cluster as a model system to test and demonstrate our method and investigate how the gold-thiolate cluster affects the structure of the cysteine ligand. We obtained the low-energy conformers of cysteine with 1000 single point energy calculations and tens of structure optimizations. We found that the adsorbed cysteine conformers inherit the hydrogen bond types from isolated conformers. However, the conformer energy ranking is reordered and the energy spacing of conformers is not the same on the cluster as in the gas phase.

[1] L. Fang, E. Makkonen, M. Todorović, P. Rinke, X. Chen, JCTC 17, 1955 (2021); [2] M. Todorović, M. U. Gutmann, J. Corander, P. Rinke, npj Comput. Mater. 5, 35 (2019)."

Soheil Ershadrad, Uppsala University - Prediction of Structural Phases in 2D FenGeTe2 Magnetic Compounds via Evolutionary Search Algorithm

A quest is underway to predict and synthesize stable 2D magnetic structures that are functional above room temperature. The recent synthesis of FenGeTe2 (n = 3, 4 and 5) vdW metallic compounds, with their room temperature transition temperature, has brought the scientific community one step closer to incorporating 2D magnets into electronic devices. It has been suggested in both experimental and theoretical studies that the unanswered complications in the magnetic properties of FenGeTe2 compounds are related to defects and disorders in their crystal structure. By using density functional theory calculations in conjunction with evolutionary search algorithms, we investigated the magnetic properties of the various configurational phases of FenGeTe2 systems. It has been demonstrated in our studies that a swapping of Fe atoms occurs in the outermost layer of 2D Fe5GeTe2 which is capable of explaining the trimerization of this compound. Our study proposes the use of an evolutionary search algorithm as a method of high-throughput prediction of metastable phases of magnetic materials with complicated crystal structures, such as the FenGeTe2 family.

Prajwal Dattatray Pisal, Aalto University - A data-driven approach toward designing efficient catalysts for CO2 to methanol conversion

To facilitate the efficient conversion of CO2 to methanol, new catalysts are required. Despite advances in high-throughput experimental and computational materials screening, catalyst discovery is challenging because the available materials space is vast and catalyst testing is slow. In this work, we have begun to compile a dataset of experimentally reported catalysts and their conversion efficiencies starting from the data set by Bahri et al. [1]. We build a random forest model that uses the Magpie descriptor [2] as input for each material. We trained the model on 124 catalytic materials. Methanol yield predictions on a test set of 24 catalysts reveal an accuracy of 0.007. Principal component analysis (PCA) gives insight into how the attributes are correlated and aids in dimensionality reduction. Future work will embed our random forest predictor into an active learning materials discovery workflow.
[1] S. Bahri et al., Journal of Cleaner Production 339 (2022) 130653 ; [2] L. Ward, et al., npj Computational Materials 2 (2016) 1–7.

Kunal Ghosh, Aalto University - Assessing the potential of active machine learning for curating molecular datasets

The performance of machine learning models depends on the size and quality of materials data, but dataset curation is often overlooked. We assess active learning for iterative curation of materials datasets, to produce accurate models with fewer data points.

Our active learning (AL) procedure is based on Gaussian process regression (GPR), a machine learning model which accepts the many-body tensor representation (MBTR) as input and predicts the highest occupied molecular energies (HOMO). We explore seven different AL acquisition strategies combining GPR prediction uncertainty with clustering based on structural diversity, and compare them to random sampling. Benchmarking the results on QM9, AA, and OE62 molecular datasets.

Smart sampling of molecules with active learning can improve model quality and prevent overfitting. Active learning can enable efficient searches of molecules by compiling a training set of molecules with the specified target property. An ML model trained on this dataset can subsequently be used to search for similar molecules. In a task designed to search for molecules with property values in a pre-determined range. Our proposed strategy was able to find the same number of molecules as random, with 10,400 fewer training labels.

Ashna Jose, University Grenoble Alpes - Regression tree-based active learning: An application to Metal Organic Frameworks

Predicting material properties using machine learning algorithms has played a crucial role in scanning large databases recently, but they often require large labeled training sets, which is not always feasible. To overcome this, we propose a regression tree-based active learning algorithm, and use it with feature space based methods to limit the size of the training set while maximizing the prediction quality. Through experiments on numerous datasets commonly used as benchmarks in machine learning, we demonstrate that our method is effective in learning a regression model from a very limited labeled dataset, even with many features, and has an appreciably low variance. We also demonstrate the efficiency of our method on a dataset of Metal Organic Frameworks (MOFs) that can employ an electronic transition to efficiently desorb gas, the thermodynamic quantities relevant to which are very time consuming to compute using electronic structure methods.

Kameyab Raza Abidi, University of Jyväskylä - Optimizing DFT to study two-dimensional metals

The 2D metals are identified as non-layered structures in contrast to the covalent 2D materials like graphene. 2D metals are beginning to gain traction at a time when an assortment of density functional theory (DFT) based methods can be implemented to determine the properties and assist in design. To determine what level of DFT attributes are needed to study 2D metals, we systematically scanned the various DFT methods. A wide range of exchange-correlation (xc) functionals were examined, ranging from LDA to GGA to hybrid. To capture the effect of dimensionality one-dimensional (1D) chain, 2D (honeycomb, square, and hexagonal), and 3D systems are taken into consideration. The performance of xc functionals was compared by translating energetics into coordination number (C) and assessing how they approach systems in different dimensions. This work aids in the methodological developments for theoretical modelling of 2D metals which will facilitate the interaction between theory and experiment. The results have the potential to provide insight into choosing which DFT methods would be the most appropriate for studying 2D metals. For data-driven discoveries of 2D metals, the time-scale comparison of each method revealed the optimal DFT method for generating enormous DFT data for incorporating machine learning algorithms.

Matilda Sipilä, University of Turku - Applying natural language processing to materials science

Scientific text is still an underutilized source of data in materials science. The aim of my research is to discover previously unknown relationships between materials, their properties and applications using natural language processing, NLP, applied to the materials science scientific text. This can be done via training large deep learning models called transformers, the state-of-the-art in NLP, with materials science text and towards some certain task. These tasks can vary from extracting plain material names from text to extracting multiple linked words, like material, property, value and unit. The emphasis of the research is in the perovskite materials, which are promising candidates for new generation solar cells. The ultimate goal of my research is to find new materials with superior functional properties.

Abdoulatif Cisse, University of Liverpool - Chemist-in-the-loop: Crowdsourcing Chemical Reasoning

The benefits of human-in-the-loop in machine learning are established. However, only some studies have considered a collaboration where the human contribution is crowdsourced and governs the data and model. We propose a new approach to solving chemistry optimisation problems by a robot utilising crowdsourced experience and reasoning given in real-time by a crowd of expert chemists. The computer uses Bayesian optimisation to solve a chemistry problem. Using a custom-developed interactive website, chemists in different locations can get insights into that opaque optimiser and guide it through crowdsourced constraints and conjectures. This translates into a more adventurous exploration of the problem search space and faster convergence. In a designed football free-kick experiment serving as a proof of concept, the computer is trying to find the set of velocity and angular features minimising the distance between a ball's path and a point target. The experts were able to visualise the search space and the evolution of both the features and model behaviour. As a result, the users added their derived constraints online into the model. This led to the computer reaching its best performance in fewer iterations.

Martin Petersen, Technical University of Denmark - Quest for outperforming cathode materials for Na-ion batteries

Na-ion batteries (NIBs) are considered sustainable, and lower-cost alternatives to Li-ion batteries (LIB), but suitable cathode materials for NIB are yet to be discovered. Because of the similarities between Na and Li, the olivine cathode materials NaMPO4 (M=Fe, Mn, Ni, and Co) have a particular interest, which also reasons from its possible high energy capacity and the transition metal ions (M) high abundance. Continuing previous studies on these materials, ab initio molecular dynamic (AIMD) simulations with density functional theory (DFT) have been performed for a few picoseconds and used to train a machine learning force field (ML-FF) for the four transition metal olivine systems. With the ML-FF, AIMD can be performed for several ns, and using active learning (AL), in particular the BMDAL method, different configurations can be selected for efficiently retraining the ML-FF. Continuing the workflow of training, predicting, selecting, and labeling makes it possible to achieve the high prediction accuracy needed to capture the kinetics of NaMPO4 using AIMD. Efficiently upgrading the ML-FF, to account for olivines with transition metal mixing, can be done by letting the AL method select the structures to be labeled for better retraining the ML-FF.

Noël Jakse, SIMAP, Université Grenoble-Alpes - Regression tree-based active learning: An application to Metal Organic Frameworks

Predicting material properties using machine learning algorithms has played a crucial role in scanning large databases recently, but they often require large labeled training sets, which is not always feasible. To overcome this, we propose a regression tree-based active learning algorithm, and use it with feature space based methods to limit the size of the training set while maximizing the prediction quality. Through experiments on numerous datasets commonly used as benchmarks in machine learning, we demonstrate that our method is effective in learning a regression model from a very limited labeled dataset, even with many features, and has an appreciably low variance. We also demonstrate the efficiency of our method on a dataset of Metal Organic Frameworks (MOFs) that can employ an electronic transition to efficiently desorb gas, the thermodynamic quantities relevant to which are very time consuming to compute using electronic structure methods.

Mohammed Guerboub, University of Strasbourg-IPCMS - Machine learning potential and first-principles molecular dynamics modelling of amorphous Ge2Sb2Te5 systems: insights into their structural and thermal properties

Phase Change Memory (PCM) is a promising non-volatile memory (NVM), they have many applications starting from memories to neuromorphic computing and storage applications, and it exhibits a high contrast between amorphous and crystalline phases[1]. This type of application leans on the difference's behaviour of amorphous and crystalline phases, especially in terms of resistivity contrast, which is exploited in PCM devices. The switching between two phases can be induced by electrical, thermal activation, or by optical pulse. The Ge-Sb-Te ternary system represents one of the most technologically relevant family of materials for PCM. In the present work, I will first assess the quantitative degree of agreement with experimental findings in terms of the structural properties of amorphous Ge2Sb2Te5 (GST225). This work is based on the exploitation of the predictive power of first-principles molecular dynamics (FPMD)[2] and a machine learning interatomic potential (MLIP) based on Gaussian Approximation Potentials (GAP) and grounded on Gaussian Process Regression (GPR). [3] Secondly, both FPMD and MLIP will be combined with the approach-to-equilibrium MD (AEMD) [4] in order to assess the thermal properties of amorphous GST225.

[1] Zhang et al. Nature Rev. Mater. 4, 150 2019; [2] Car and Parrinello Phys. Rev. Lett. 55 2471 1985; [3] Albert Bartok-Partay et al. Quantum Chemistry, 115(16), pp. 1051-1057 2015 ; [4] E. Lampin et al, J. Appl. Phys., 2013, 114, 033525.

Sintija Stevanoska, Jozef Stefan Institute - Semi-supervised and active multi-target regression for material design

Machine learning (ML) methods for multi-target regression (MTR), i.e., MTR trees and tree ensembles, have been recently used in material design. They have learned models that simultaneously predict multiple material properties from material composition or from processing parameters. They have been used in fully supervised mode, where all the target values of all the training examples are known, i.e., all examples are labeled. In practice, unlabeled data are often available and can be beneficial to the learning process: ML can use the data unlabeled (semi-supervised learning) or can ask the user to label some of the examples, which it expects will most improve the learned model (active learning). Recently, semi-supervised variants have been developed of tree-ensemble approaches for MTR, based on the self-training approach and on the approach of predictive clustering. We plan to apply these two semi-supervised MTR approaches to versions of the datasets already addressed, extended with unlabeled data. We also plan to adapt the self-training approach towards active learning. Namely, it includes approaches to estimating the reliability of its predictions: These estimates can be used to focus on the least reliable predictions that should be labeled in active learning.

Aslak Fellman, University of Helsinki - Machine Learning interatomic potentials for FCC high-entropy alloys

Machine learning (ML) approaches are increasingly becoming common place in the development of interatomic potentials. ML potentials have shown excellent accuracy compared to traditional analytical potentials. I will present my current work in developing a ML potential for FCC high-entropy alloy AlCrCuFeNi. High-entropy alloys have shown exciting material properties and are currently a very active field of research within materials research. The creation of new interaction models is crucial for the accurate simulation of these complex alloys. The potentials in this work were developed using the Gaussian approximation potential (GAP) method. By using low-dimensional descriptors we are able to create computationally fast potentials that allow for the simulations of large-scale systems. I will discuss the process and considerations when it comes to creation of high-entropy alloy potentials using machine learning.

Han Le, Aalto University - Design of Experiments: Quantitative Comparison of Bayesian Optimization with Response Surface Methodology

Optimizing measured quantities, e.g., the yield of a chemical process, can be costly and time-consuming. Design of Experiment (DoE) methods formalize this process of finding optimal experimental conditions. Response surface methods (RSM), in which a model of the measured quantity is fitted to the experimental conditions, are a common DoE approach. However, RSM has several disadvantages, e.g., (i) the formulated model typically only considers second-order polynomials, neglecting higher-order effects; (ii) choices for experimental conditions are not updated based on knowledge obtained from completed experiments.

Bayesian optimization (BO) has emerged as an alternative, adaptive DoE method. In this work, we compare the Bayesian Optimization Structure Search (BOSS) code, which has been used for DoE [1], to the central composite design in RSM. To do this, we generate a large set of multivariate polynomial functions with (i) dimensions from 2 to 4; (ii) varying standard deviation of simulated Gaussian noise; (iii) varying weight of third-order terms. For all investigated dimensions, we identify a broad regime, in which BOSS performs better than RSM as the weight of the third-order terms and the noise increase.

[1] J. Löfgren, D. Tarasov, T. Koitto, P. Rinke, M. Balakshin, and M. Todorovic, "Machine learning optimization of lignin properties in green biorefineries," ACS Sustainable Chemistry Engineering, vol. 10, no. 29, pp. 9469–9479, 2022.

Rasmus Kronberg, CSC – IT Center for Science Ltd. - CSC Services for Computational Research

CSC – IT Center for Science Ltd. provides world-class services for computational research, both for academia and enterprises. This includes high-quality science support, data storage solutions and high-performance computing (HPC) resources such as the national supercomputers Puhti and Mahti. Moreover, CSC also hosts the European pre-exascale supercomputer LUMI, which was inaugurated in 2022 and has now reached general availability1. LUMI is currently the third fastest supercomputer in the world with a hefty GPU-partition that enables unprecedented HPC workflows, especially within AI and high-performance data analytics. Materials science and computational chemistry applications ported to AMD GPUs will also greatly benefit from the performant MI250X accelerators, in addition to the auxiliary CPU-partition of more than 200 000 compute cores. This contribution presents the service portfolio of CSC and outlines which types of use cases are best suited for each resource. The poster presentation will also serve as an opportunity for you to discuss your own research and development workflow and learn how CSC could support you in deploying the available HPC resources as efficiently as possible.

[1] https://www.lumi-supercomputer.eu/(Accessed 13.2.2023).