Materials Informatics: Accelerating Materials Research and Design with Artificial Intelligence

ICISE complex, Quy Nhon, Binh Dinh, Vietnam
August 23-25, 2024

ABSTRACT BOOKLET

https://icisequynhon.com/conferences/2024/materials-informatics

Contents

Acknowledgements

Emerging in the early 2010s and partly propelled by the White House's Materials Genome Initiative in 2011, materials informatics has rapidly developed, joining the main stream of materials research. In this emerging highly-interdisciplinary research and development field, Artificial Intelligence (AI)/Machine Learning (ML) techniques are developed to learn past materials data, creating different classes of AI/ML models, workflows, and strategies to complement and accelerate the traditional (physics- and chemistry-based) approaches in materials research and discovery. Synergistic efforts worldwide involving materials informatics, simulations, and physical experimentation have led to numerous discoveries in, for examples, battery materials, green energy materials, functional and sustainable polymers, alloys, and more.

This is the first workshop dedicated to materials informatics in Vietnam. We are delighted to welcome leading experts and pioneers, plenary speakers, distinguished guests, participants, and, especially, young scientists from Austria, Australia, China, France, Germany, India, Japan, Singapore, USA, and Vietnam. We are also lucky enough to have scientific contributions that cover a few critical topics of the field. They are (A) "*Fundamentals of Materials Informatics*" where the core ideas, techniques, methods, and success and inspiring stories are cultivated and described, (B) "*Basic Infrastructures of Materials Informatics*" where materials data, algorithms, software, recommender systems, and robotics and autonomy frameworks are developed, (C) "*Accelerated Methodologies for Materials Research*" where the traditional methods are significantly accelerated, and (D) "*Accelerated Materials Property Predictions & Design*" where AI/ML techniques help to explore the staggering chemical and parameter spaces, discovering new materials and formulations with desired performances. After the first two days devoted to nearly 30 lectures covering the topics, the last day of this workshop is reserved for (E) some practical hand-on tutorials, introducing some simple problems in the field and walking participants through the essential steps to solve them.

This workshop is organized by *Rencontres du Vietnam*, *International Centre for Interdisciplinary Science and Education* (ICISE), and *VinUni Center for Environmental Intelligence* (CEI) in Quy Nhon, Binh Dinh, Vietnam on August 23-25, 2024. We sincerely thank Prof. Jean Tran Thanh Van (Rencontres du Vietnam and ICISE), Dr. Tran Thanh Son (ICISE and Quy Nhon University), Prof. Phung Thi Viet Bac (CEI), Ms. Tran Thi Quynh Nhu (ICISE), and Mr. Nguyen Trong Nhan (ICISE) for enormous supports from the early days of the plan. To speakers, guests, participants, and organizing committee members, we thank you for your critical contributions, without which the workshop plan cannot be materialized. We wish you an enjoyable experience in Quy Nhon, and hope to welcome you next times.

Huan Tran, Georgia Institute of Technology, USA
Tuoc Vu, Hanoi University of Science & Technology, Vietnam
On behalf of the organizing committee

Organizers

International Centre for Interdisciplinary Science and Education

The International Centre for Interdisciplinary Science and Education (ICISE) is a unique science and education institution located in the coastal city of Quy Nhon, Vietnam. It welcomes national and international Conferences, specialized professional colloquia, as well as selected Thematic Schools, in particular for doctoral students and has full support and appreciation of the local and national authorities. The monument is housing a 350-seat auditorium, a conference room and more intimate seminar rooms along with offices for study and research. These spaces showcase spectacular views out over the sea and nearby mountains.

ICISE center - view from the west

Activity

Commissioned by the association "Rencontres du Vietnam" under the direction of its founders Prof. Jean Tran Thanh Van and Prof. Le Kim Ngoc, the purpose of ICISE is to bring together scientists from developed and emerging countries to host conferences as well as to nurture the association's long-held expertise in designing exceptional cultural and educational projects.

The Centre ICISE hosts between 10 and 12 high level international scientific conferences each year. Covering disciplines in the scope of fundamental or applied science mainly in the field of physics, the program will extend to cover a larger array of scientific disciplines as biology, medicine, social and human sciences. On average a conference gather 150 people during a six days period.

The location

Located a few kilometers away from the city centre of Quy Nhon, capital of Binh Dinh province, ICISE provides a peaceful environment, removed from the city traffic to facilitate more intimate dialogues and an exclusive focus on work exchanges. Its exceptional location on a wild 21.38 hectares (50 acres) site between mountains and sea, makes it an inspirational and energizing place.


ICISE centre between mountains and sea

The city of Quy Nhon is located on the east coast of Vietnam, just 45 minutes north by air from Ho-Chi-Minh City (Saigon) and 1 and 30 hour south from Hanoi. This economically growing city still preserved from mass tourism has retained its traditions including martial arts and reveals an authentic face of Vietnam.

History

In a unique adventure in the scientific world, the Association "Rencontres du Vietnam" under the direction of its founder Jean Tran Thanh Van, has developed a close network of international cooperation for scientific research around the entire world for over 30 years. It organizes high-level Conferences in Vietnam which attract research

scientists from developed and emerging nations allowing young Asian students to meet with top scientists all the way to Nobel Laureates.

Objectives

The success and reputation of previous Rencontres and conferences led to the creation of ICISE. The centre was born in the attempt to promote North-South cooperation and develop education through a privileged partnership.

The rapid economic development of emerging nations in the Asian-Pacific zone generally contrasts with their level of scientific and technological development. The increasing awareness of this gap has resulted in the high demand for collaborations, in fundamental research as much as in technological transfers and applications, which today opens up a large range of potential cooperations in Asia.

It is well known that this region of the world is characterized by the youth of its population, and by the constant progress achieved in university education. To give opportunities to students from all these countries to take part in scientific projects with a strong international dimension constitutes an issue and a challenge of first rate importance for the countries and institutions involved as much as for the individuals concerned.

To help the emerging nations to develop their potential fully, one must consider the scientific instruction of their youth. The actions taken at ICISE will serve to define the adaptation of new technologies to different economic, social, and environmental contexts, and this effort will open up new industrial and commercial opportunities.

VinUni Center for Environmental Intelligence

The Center for Environmental Intelligence (CEI) - VinUniversity represents a pioneering initiative at the intersection of advanced technology, and environmental science, and interdisciplinary research.

CEI's core strength lies in its diverse, world-class advisory board, comprising experts from leading institutions such as UC Berkeley, University of Notre Dame, and Carnegie Mellon University. This multidisciplinary team guides the center's research agenda, ensuring a comprehensive approach to environmental intelligence.

The center's research portfolio spans three key areas:

- Theme 1: Environmental Monitoring, Community Resilience
- Theme 2: Living Lab: Smart Cities, Energy Efficiency, and Green Technologies
- Theme 3: AI-powered Materials Discovery, Digital Materials Science

These pillars exemplify CEI's commitment to integrating cutting-edge technologies like AI, digital science, and advanced monitoring systems with environmental research. By fostering collaboration between technologists, social scientists, and environmental experts, CEI aims to generate innovative, holistic solutions to pressing environmental issues.

At the core of this commitment lies the establishment of new research facilities at VinUni, which will serve as a technology hub for future innovators and facilitate connections among experts from around the world to collectively solve sustainable development issues on a global scale. CEI harnesses advanced technology and interdisciplinary research to tackle urgent environmental challenges. Through the integration of technology, social sciences, digital materials, and sustainable energy, we strive to drive transformative solutions for a sustainable future.

Laurent El Ghaoui

SCIENTIFIC DIRECTOR

Organizing Committee

1. Huan Tran, Georgia Institute of Technology, USA
2. Tuoc Vu, Hanoi University of Science & Technology, Vietnam
3. Kien X. Ngo, Nano Life Science Institute, Kanazawa University, Japan
4. Tu C. Le, School of Engineering, RMIT University, Australia
5. Nam-Nhat Hoang, Vietnam National University, Hanoi, Vietnam
6. Thanh Tien Nguyen, Can Tho University, Vietnam
7. Vuong Minh Nguyen, Quy Nhon University, Vietnam
8. Viet Bac T. Phung, VinUniversity, Vietnam
9. Thanh Hai Phan, Quy Nhon University, Vietnam
10. Thanh-Son Tran, ICISE and Quy Nhon University, Quy Nhon, Vietnam
11. Thang Duy Dao, Microsystems Silicon Austria Labs GmbH, Villach, Austria
12. Son Tung Ha, A*STAR, Singapore
13. Hai Son Nguyen, Ecole Centrale de Lyon, France

International Advisory Committee

1. Prof. Tran Thanh Van, Rencontres du Vietnam & ICISE, Quy Nhon, Vietnam
2. Prof. Le Kim Ngoc, Rencontres du Vietnam & ICISE, Quy Nhon, Vietnam
3. Masato Yoshiya, Osaka University, Japan
4. An Bang Ngac, Vice Rector, Hanoi University of Science, Vietnam
5. Hieu-Chi Dam, Japan Advanced Institute of Science & Technology, Japan
6. Beatriz G Del Rio, Universidad de Valladolid, Spain
7. Christopher Künneth, University of Bayreuth, Germany
8. Toan T. Nguyen, Vietnam National University, Hanoi, Vietnam
9. Tomoyuki Yamamoto, Waseda University, Japan

General Information

The City of Quy Nhon

Quy Nhon, the seat (or capital) of Binh Dinh Province, is a coastal city located in central Vietnam, known for its scenic landscapes, historical sites, and a growing reputation as a tourist destination. As of 2022, its population was 481,110. Nestled between lush green mountains and the turquoise waters of the South China Sea, Quy Nhon boasts pristine beaches, picturesque landscapes, and a tranquil atmosphere that appeals to those seeking a more relaxed and off-the-beaten-path destination.

The city is known for its well-preserved Cham towers, ancient relics of the Champa Kingdom (an awesome article about the towers can be found by following this link), and its bustling fishing villages that give visitors a glimpse into traditional Vietnamese life. Quy Nhon also offers a variety of outdoor activities, from exploring the stunning Ky Co and Eo Gio beaches to hiking in the nearby hills.

With a growing reputation for its fresh seafood and friendly locals, Quy Nhon provides an authentic experience for travelers looking to explore Vietnam beyond the usual tourist spots. Whether you're interested in history, nature, or simply enjoying the serene coastal charm, Quy Nhon has something to offer for every type of travelers.

Quy Nhon in the map of Vietnam (from Wikipedia).

Currency

Vietnam's currency is the Vietnamese đồng, abbreviated as VND, and used for all transactions within the country. Bank notes are available in denominations ranging from 1,000 to 500,000 đồng. The current exchange rate is roughly 25,000 VND for 1 USD. Exchanging from USD and other common foreign currencies such as Japanese ¥ and European € to VND is quite convenience; you can do the exchange in the international airport when you enter Vietnam, or in many banks and agencies in the city.

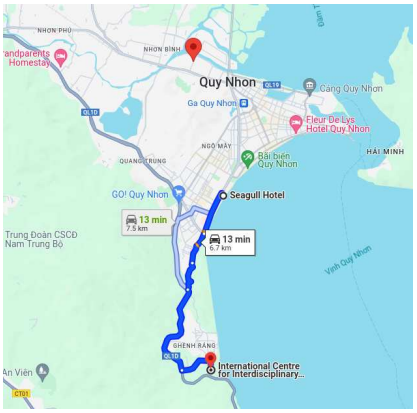
When traveling in Vietnam, it's advised to carry some cash, especially in smaller towns and rural areas where credit cards may not be widely accepted. ATMs are available in most cities and tourist areas, and they typically dispense larger denomination notes.

Twin Cham towers in Quy Nhon. Photo from `https://toquoc.vn`.

Transportation

Quy Nhon is served by Vietnam Airlines, Bamboo Airways, and VietJet Air through Phu Cat Airport (IATA: UIH), with flights from/to Hanoi and Ho Chi Minh City. As there are more flights connecting Ho Chi Minh City to Quy nhon, it is more convenient to fly from foreign countries to this southern city than to Hanoi. From the airport, you can find multiple taxi options on arrival. The cost could range from about 250,000 VND (≃ 10 USD) for a compact sedan to 500,000 VND (≃ 20 USD) for a 7-seat SUV. Quy Nhon can also be reached by trains via Dieu Tri railway station, around 10 km (≃ 6 miles) to the west of Quy Nhon. Rental bikes and motorbikes are available for local explorations, and information can be obtained at hotel front desks.

During the workshop, shuttle buses are provided from Seagull hotel (downtown Quy Nhon) to ICISE. The trip takes about 10-15 minutes for about 7 km of distance. Please refer to the Agenda for the schedule.



Shuttle bus route between Seagull hotel and ICISE during the workshop. See Agenda for schedule.

Agenda

Day 1: August 23, 2024

Commute	
07.00 – 07.30	Shuttle bus from Seagull Hotel to ICISE
Registration & Opening	
07.30 – 08.30	Registration & poster hanging
08.30 – 09.00	Opening and photos
Plenary Section	**A: Fundamentals of Materials Informatics**
09.00 – 09.40	Plenary talk (David Winkler, PLN01)
09.40 – 10.20	Plenary talk (Claudia Draxl, PLN02)
10.20 – 10.50	Coffee break
10.50 – 11.20	Invited talk (Hiori Kino, INV01)
11.20 – 11.50	Invited talk (DAM Hieu-Chi, INV02)
Lunch	
11.50 – 13.00	Lunch at ICISE
Plenary Section	**B: Basic Infrastructures (Data, Algorithms, Software, Robotics, & Autonomy)**
13.00 – 13.30	Invited talk (Laurent El Ghaoui, INV03)
13.30 – 14.00	Invited talk (Ryo Yoshida, INV04)
14.00 – 14.20	Oral talk (Duc-Anh Dao, ORL04)
14.20 – 14.40	Oral talk (Minh-Quyet Ha, ORL02)
14.40 – 15.10	Coffee break
15.10 – 15.40	Invited talk (Kanta Ono, INV05)
15.40 – 16.10	Invited talk (Hiroyuki Hayashi, INV06)
16.10 – 16.30	Free time at ICISE
Commute	
16.30 – 17.00	Shuttle bus from ICISE to Seagull Hotel

Day 2: August 24, 2024

Commute		
08.30 - 09.00	Shuttle bus from Seagull Hotel to ICISE	
Parallel Sections	**C1: Accelerated Methodologies for Materials Research**	**D1: Accelerated Materials Property Predictions & Design**
09.00 – 09.30	Invited talk (Brandon Wood, INV07)	Invited talk (Hongbin Zhang, INV12)
09.30 – 10.00	Invited talk (N. T. Van-Oanh, INV08)	Invited talk (Tu Le, INV13)
10.00 – 10.30	Coffee break	
10.30 – 11.00	Invited talk (Sanjeev Nayak, INV09)	Invited talk (Thanh N. Truong, INV14)
11.00 – 11.30	Invited talk (Toan T. Nguyen, INV10)	Invited talk (Anh Pham, INV15)
11.30 – 12.00	Invited talk (Manh-Thuong Nguyen, INV11)	Invited talk (Sanjubala Sahoo, INV16)
Lunch		
12.00 – 13.30	Lunch at ICISE	
Parallel Sections	**C2: Accelerated Methodologies for Materials Research**	**D2: Accelerated Materials Property Predictions & Design**
13.30 – 14.00	Invited talk (Shusuke Kasamatsu, INV17)	Invited talk (Tien Quang Nguyen, INV20)
14.00 – 14.30	Invited talk (Dao Xuan Viet, INV18)	Invited talk (Van Doan Le, INV21)
14.30 – 15.00	Coffee break	
15.00 – 15.30	Invited talk (Kien Xuan Ngo, INV19)	Invited talk (Nguyen Thanh Tien, INV22)
15.30 – 15.50		Oral talk (Debarati Dey Roy, ORL05)
Poster Section		
15.50 – 16.30	Free time & move to Poster section	
16.30 – 17.30	Posters PST01 (Lam H. Nguyen), PST02 (Huu T. Do), PST03 (Huy Duy Nguyen), PST04 (Huu Doanh Nguyen), PST05 (Thai-Bach Le), PST06 (Tri M. Nguyen), PST07 (Erina Fujita), and PST08 (Aiko Takahashi)	
Gala Dinner		
17.30 – 18.00	Poster removal & move to dinner location at ICISE	
18.00 – 20.30	Gala dinner at ICISE	
Commute		
20.30 - 21.00	Shuttle bus from ICISE to Seagull Hotel	

Day 3: August 25, 2024

Commute		
07.00 - 07.30	Shuttle bus from Seagull Hotel to ICISE	
Parallel Sections	**E: Materials Informatics Tutorial**	**F: Local Excursion**
07.30 – 12.00	Handon tutorials, preregistered participants	Sightseeing, preregistered guesses
Closing		
12.00 – 12.30	Closing & travel awards	
Lunch		
12.30 – 13.30	Lunch at ICISE	
Commute		
13.30 - 14.00	Shuttle bus from ICISE to Seagull Hotel	

Technical Abstracts

Use of Materials Informatics to Discover New Materials: Theory and Successes

David A. Winkler
La Trobe Institute for Molecular Science, La Trobe University, Bundoora, Australia
Monash Institute of Pharmaceutical Science, Monash University, Parkville, Australia
School of Pharmacy, University of Nottingham, Nottingham, UK
Email: d.winkler@latrobe.edu.au

ID: PLN01
Plenary Talk
Aug. 23, 9.00 - 9.40

Keywords: Materials design, computational chemistry, machine learning, artificial intelligence, molecular modelling

Currently, we are experiencing amazing, paradigm shifting scientific developments. We understand that the size of small molecule and materials spaces is essentially, infinite, representing an inexhaustible supply of potential drugs and materials with useful properties. This has seen a rapid increase in automation and robotics, allowing synthesis of new molecules and materials and measurement of properties orders of magnitude faster. This has created massive databases of complex genetic, structural, chemical, property, and biological information.

Finding "islands of chemical utility" in a vast palette of possibilities and extracting meaning from massive databases has driven a rise in applications of AI and machine learning, and development of methods. There has been a parallel rise in applications to most aspects of modern life – medicine, finance, manufacturing, social media. Recently, we saw development of rapid, accurate quantum machine learning methods, generative methods to suggest new molecules or materials with improved properties, prediction of protein structures from sequence (AlphaFold), the beginning of general AI (ChatGPT), massive "make on demand" chemical libraries (ZINC-22), seminal work on autonomous chemical discovery, and increasing use of other AI methods (evolutionary algorithms) to discover molecules and materials with improved properties.

This presentation will discuss drivers for these developments, summarize contributions my teams have made to adapting and applying machine learning, and provide examples of applications to biomaterials and regenerative medicine, drug design, 2D and porous materials, nanomaterials, surface science, cancer diagnostics, corrosion control, and sustainable energy sources.

References

1. Li *et al.*, "Rational Atom Substitution to Obtain Efficient, Lead-Free Photocatalytic Perovskites Assisted by Machine Learning and DFT Calculations", Angew. Chem. Int. Ed. **62**, e202315002 (2023).
2. Surmiak *et al.*, "Machine Learning-Enhanced High-Throughput Fabrication and Optimization of Quasi-2D Ruddlesden-Popper Perovskite Solar Cells", Adv. Energy Mater. **13**, 2203859 (2023).
3. Mai *et al.*, "Machine Learning in the Development of Adsorbents for Clean Energy Applications", Adv. Sci. **9**, 2203899 (2022).
4. Mai *et al.*, "Machine Learning for Electrocatalyst and Photocatalyst Design and Discovery", Chem. Rev. **122**, 13478 (2022).
5. Muratov *et al.*, "A critical overview of computational approaches employed for COVID-19 drug discovery", Chem. Soc. Rev. **50**, 9121 (2021).
6. Vasilevich *et al.*, "Evolutionary design of optimal surface topographies for biomaterial", Sci. Rep. **10**, 22160 (2020).
7. Meftahi *et al.*, "Machine Learning Property Prediction for Organic Photovoltaic Devices", npj Comput. Mater. **6**, 166 (2020).
8. Muratov, *et al.*, "QSAR without Borders", Chem. Soc. Rev. **49**, 3525 (2020).

FAIR Data Towards New Insights into Materials

Claudia Draxl
Physics Department and CSMB, Humboldt-Universität zu Berlin, Berlin, Germany
Email: claudia.draxl@physik.hu-berlin.de

ID: PLN02
Plenary Talk
Aug. 23, 9.40 - 10.20

Data-centric approaches are becoming an integral part of our research in complementing established traditional paradigms. Given the variety of possible applications in materials science, the potential for gaining new insights through artificial-intelligence (AI) is enormous. The vast amounts of research data produced every day in the field represent a 21st-century gold mine. How can we turn these data into knowledge and value? A FAIR (Findable, Accessible, Interoperable, and Re-usable) data infrastructure plays a decisive role, because this gold mine is of little value if the data are not comprehensively characterized and made available. Only then, data can be readily shared and explored through statistical analysis and machine learning. Making data Findable and AI Ready (another interpretation of the acronym) will change the way how science is done today. With selected examples, I will show how Big Data from either computational or experimental work can be used to find trends and patterns, which would not be possible from single investigations. Thereby, I will also address the challenges when data potentially come from different sources and show how machine learning can be used for error quantification and data augmentation. I will ask how we can make use of existing data collections to find materials with desired properties and will show how ML can be combined with established methodology toward nonlinear modeling and classification.

Keywords: Artificial intelligence, data-centric science, data infrastructure, FAIR data, big data

Enhancing Materials Informatics through Ontologies and Functional Decomposition: Improving Data Traceability and Understanding

ID: INV01
Invited Talk
Aug 23, 10.50 - 11.20

Hiori Kino
National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki, Japan
Email: kino.hiori@nims.go.jp

Keywords: Materials Informatics, Ontologies, Functional Decomposition, Data Cleansing, NIMS Atomwork Advanced Database, Protégé, Data Traceability, Knowledge Hierarchy

Materials informatics is an interdisciplinary field, and it is difficult to understand as a whole because there are specialized areas even within materials scientists, as well as data scientists working on databases, scientists working on machine learning methods. Because of the difficulty of understanding disciplines, ontologies formally define concepts and their relationships in order to facilitate data sharing and reuse. Even so, it is difficult to create an exhaustive lexical hierarchy of concepts in materials science. This approach can be used in a simple way. In addition, other methods developed by ontology specialists can be used to improve the explanatory power of the ontology.

In materials science, data are not immediately ready for analysis upon collection. They require data cleansing, which is purpose-dependent. Efficiently cleansing large datasets necessitates identifying data nodes that fail to be transformed appropriately. The NIMS Atomwork Advanced Database is optimized for browser display, not data analysis. Using this database as an example, I describe the data structure with ontologies via Protégé and transform it into a graph database. This method improves data traceability, clarity, and understanding.

It is difficult to describe the world using the relation-based description method in which ontology excels. For this reason, ontology experts have proposed a functional decomposition tree, which describes the world in a purpose-oriented and recursive manner, using natural language to describe transformations as the main subject. This method can be used to hierarchically describe tacit knowledge that is frequently omitted in expert explanations and to improve the understanding of the people involved. Even if you do not actually use this technique, simply being aware of it can improve explanatory performance. For example, a lexical hierarchy of material science concepts can be defined by a functional decomposition tree. Another example is to describe a program in terms of a functional decomposition tree.

Development of Data-Driven Methods for Materials Discoveries: Representation, Learning, and Uncertainty Modeling and Management

ID: INV02
Invited Talk
Aug. 23, 11.20 - 11.50

DAM Hieu-Chi,[1,2] NGUYEN Duong-Nguyen,[1] HA Minh-Quyet,[1] VU Tien-Sinh,[1] NGUYEN Viet-Cuong,[3] KINO Hiori,[4] and MIYAKE Takashi[5]

[1] *Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan*

[2] *Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan*

[3] *HPC SYSTEMS Inc., 3-9-15 Kaigan, Minato, Tokyo 108-0022, Japan*

[4] *National Institute for Materials Science, 1-2-1 Sengen, Tsukuba 305-0044, Japan*

[5] *National Institute of Advanced Industrial Science and Technology, 1-1-1 Umezono, Tsukuba 305-8568, Japan*

Email: dam@jaist.ac.jp

Keywords: Machine Learning, Data Mining, Evidence Theory

In this presentation, we explain our efforts in developing data-driven methods for materials studies, utilizing machine learning (ML) and data mining to enhance materials research and design. Our focus is on methods that deepen the understanding of mechanisms governing material properties in both molecular and crystalline forms, derived from computational and experimental data.

Central to our research is the development of advanced ML and data mining models that enhance comprehension of structure-property relationships and predictive capabilities for materials properties. Our work starts with designing descriptors that incorporate structural information, electronic structures, and conventional physicochemical knowledge, facilitating property predictions through ML techniques [1]. Our recent use of interpretable deep learning with attention mechanism has advanced the understanding of structure-property relations in materials, contributing to more transparent and explainable models in materials science. These models not only predict material properties but also elucidate underlying factors, revealing key structure-property relationships [2].

We also focus on managing and quantifying uncertainty in inference and prediction to ensure robust outcomes. Using the Dempster-Shafer theory, we convert material data into evidence on similarities, appropriately modeling and managing uncertainty for data-driven inferences using unbalanced, contradictory, and biased materials data. We have developed evidence-based methods to evaluate material similarity and compositions, providing deeper insights and improving predictive accuracy. This approach has shown superior performance in identifying promising material compositions, validated through experimental studies [3].

The synergy of data-driven and conventional computational approaches in our studies accelerates materials discovery, emphasizing interpretable models and the explainability of results. This comprehensive approach demonstrates the transformative potential of data-driven methods in understanding and discovering novel materials.

References

1. T. L. Pham, H. Kino, K. Terakura, T. Miyake, K. Tsuda, I. Takigawa, and H. C. Dam, Sci. Technol. Adv. **18**, 756-765 (2017).
2. T.-S. Vu, M.-Q. Ha, D.-N. Nguyen, V.-C. Nguyen, Y. Abe, T. Tran, H. Tran, H. Kino, T. Miyake, K. Tsuda, and H.-C. Dam, npj Comput. Mater. **9**, 215 (2023).
3. M.-Q. Ha, D.-N. Nguyen, V.-C. Nguyen, T. Nagata, T. Chikyow, H. Kino, T. Miyake, T. Denœux, V.-N. Huynh, and H.-C. Dam, Nat. Comput. Sci. **1**, 470-478 (2021).

Finding Environmental-friendly Chemical Synthesis with AI and High-throughput Robotics

ID: INV03
Invited Talk
Aug. 23, 13.00 - 13.30

Vu Van Hao,[1,2] Dang Dinh Dang Khoa,[1] Le Duy Dung,[1,2] Nguyen Dang Tung,[1,2] and Laurent El Ghaoui[1,2,†]

[1] *Center for Environmental Intelligence, VinUniversity, Gia-Lam, Hanoi, Vietnam*

[2] *College of Engineering and Computer Science, VinUniversity, Gia-Lam, Hanoi, Vietnam*

Email: laurent.eg@vinuni.edu.vn

Keywords: LLM literature synthesis, MOFs, LLM experimental verification, Green, Robotic synthesis

Recent challenges with the environment have resulted in tremendous interest in Green Chemistry, which includes the design of chemical products and processes that reduce the use of environmentally harmful substances. Until now, finding new environmental chemical synthesis is largely a trial-and-error process, requiring trained expertise and a large amount of work. Here, we report a high-throughput process, combining AI techniques and robotic synthesis, allowing us to find a more environmentally friendly way to synthesize an existing material. The model materials in this study are to replace nitrate salts (NO_3-), which might be responsible for algae bloom if leaked into open water, by a chloride salt (Cl-), a natural abundant ion, in the synthesis of a metal-organic framework, Zn-HKUST-1. Our high-throughput process starts with using large language models (LLM)-based literature summary to create a database on the synthesis of Zn-HKUST-1 with NO_3-, so that optimized concentrations of Cl- can be suggested. Subsequently, these suggestions are tested with automatic robotic processes, increasing the speed and precision of the experiment, to find the optimal synthesis condition. Using this process, we successfully obtained MOFs (Metal Organic Frameworks) crystals from $ZnCl_2$ precursors. An AI-based learning process is developed, comparing the structural information of our synthesized materials with that of the database to confirm indeed the obtained MOFs are Zn-HKUST-1. This success proves that our process holds the promise to accelerate the discovery of new environmental-friendly materials in the near future.

Transferability and Scalability of Growing Computational Database in Sim2Real Materials Informatics

ID: INV04
Invited Talk
Aug. 23, 13.30 - 14.00

Ryo Yoshida
The Institute of Statistical Mathematics, Tokyo, Japan
Email: yoshidar@ism.ac.jp

Keywords: Polymer Database, Molecular Dynamics, Transfer Learning, Small Data.

In the past decade, machine learning has shown the potential to greatly accelerate the of discovery of new materials in various systems. However, the major obstacle in data-driven materials research, the lack of data resources, remains unresolved. High experimental costs and the cultural barrier of researchers to disclose their laboratory data contribute to this issue, making it difficult to solve in the short term. To address data scarcity, the development of open databases using computational methods such as first-principles calculations and molecular dynamics (MD) simulations is underway. However, for polymeric materials, no comprehensive database exists due to technical barriers in automating computer experiments. To generate data resources for machine learning, we developed RadonPy, an open-source software that fully automates polymer property calculations based on all-atom classical MD simulations [1]. Using RadonPy, we create the world's largest polymer property databases, spanning a broad chemical space of more than $10^5 - 10^7$ polymer species.

The methodology of transfer learning, particularly the simulation-to-real (Sim2Real) transfer learning, enables the integration of extensive simulation data with limited experimental data [2,3,4]. Transfer learning leverages data or pretrained models from a source domain to enhance tasks in a target domain. Here, we present a measure for quantitatively evaluating the transferability and scalability of growing computational materials database. Our work revealed a scaling law in Sim2Real transfer learning using the RadonPy database. By observing the scaling behavior of transferred predictors, we estimate their expected generalization performance achievable by further increasing simulation data, serving as an indicator of the database's potential value. Additionally, multidimensional scaling, considering both physical and computer experiments, provides a statistical estimate for the equivalent sample size of experimental and computational data. Furthermore, observing the scaling strength of individual materials, we can gain insights into material groups sharing physical mechanisms across different material systems.

References

1. Y. Hayashi, J. Shiomi, J. Morikawa, and R. Yoshida, "RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics", npj Comput. Mater. **8**, 222 (2022).
2. Wu *et al.*, "Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm", npj Comput. Mater. **5**, 66 (2019).
3. Yamada *et al.*, "Predicting materials properties with little data using shotgun transfer learning", ACS Central Science **5**, 1717 (2019).
4. Aoki *et al.*, "Multitask machine learning to predict polymer-solvent miscibility using Flory-Huggins interaction parameters", Macromolecules **56**, 5446 (2023).

Autonomous Materials Research utilizing Robots and AI

Kanta Ono
Department of Applied Physics, Osaka University, Osaka, Japan
Email: ono@ap.eng.osaka-u.ac.jp

ID: INV05
Invited Talk
Aug. 23, 15.10 - 15.40

Keywords: Robotics, Artificial Intelligence, Materials Informatics, Laboratory Automation, Autonomous Experimentation, Self-driving Laboratory

Recent advancements in robotics and artificial intelligence (AI) have revolutionized the field of materials research, offering unprecedented levels of automation and precision. Our latest research focused on two pioneering autonomous materials research systems: a force-controlled robotic mechanochemical synthesis system and an autonomous robotic experimentation system for powder X-ray diffraction.

The force-controlled robotic mechanochemical synthesis system utilizes a robotic powder grinding mechanism that applies a precisely controlled and constant mechanical force. This approach enhances reproducibility and allows for detailed analysis of reaction pathways. Our experiments with perovskite materials have demonstrated that this robotic method provides significantly higher reproducibility than traditional manual grinding and ball milling. We could dynamically control the reaction pathways by varying the grinding forces and speeds, leading to new insights and discoveries in chemical reaction mechanisms.

In parallel, our development of the autonomous robotic experimentation system for powder X-ray diffraction (ARE-XRD) has enabled real-time, in-situ monitoring of chemical reactions. This system automates the entire experimental process, from sample preparation to data analysis, allowing continuous, unattended operation. Integrating ARE-XRD with our robotic mechanochemical synthesis system has proven particularly powerful, providing real-time feedback and enabling precise control over experimental conditions.

These combined advancements highlight the potential of robotics and AI to revolutionize materials research. The ability to automate and precisely control experimental processes opens new avenues for discovering and understanding complex chemical reactions. In this talk, I will present the design, construction, and experimental results of our robotic mechanochemical synthesis system and the ARE-XRD system. We will discuss the broader implications of these technologies for materials science and outline future directions for integrating robotics and AI in scientific research.

Recommender Systems in Materials Science: Methods and Applications

ID: INV06
Invited Talk
Aug. 23, 15.40 - 16.10

Hiroyuki Hayashi,[1,†] Atsuto Seko,[1] and Isao Tanaka[1,2]
[1] *Department of Materials Science and Engineering, Kyoto University, Kyoto, Japan*
[2] *Nano Research Laboratory, Japan Fine Ceramics Center (JFCC), Nagoya, Japan*
Email: hayashi.hiroyuki.6s@kyoto-u.ac.jp

Keywords: Chemically relevant compositions, Machine learning, Recommender system, Parallel synthesis

A recommender system based on experimental databases aids in the efficient discovery of inorganic compounds. This presentation reviews studies on discovering unknown compounds using such systems. Two primary methods are highlighted: one using compositional descriptors derived from elemental features for machine learning and binary classification of chemical compositions registered in the inorganic crystal structure database (ICSD), and the other employing tensor decomposition techniques without descriptors. Predictive performance for unknown chemically relevant compositions (CRCs) was assessed by checking their presence in other databases.

Additionally, a synthesis-condition recommender system was developed using machine learning on a parallel experimental dataset collected through a polymerized complex method. This system's recommendation scores for unexperimented conditions were evaluated, leading to the discovery of two new pseudo-binary oxides.

Recommender systems are increasingly popular in fields like e-commerce and social networking, where they suggest products and content based on user data. This methodology is also applicable to materials discovery, where recommendation scores indicate the likelihood of finding relevant chemical compositions and synthesis conditions. The reviewed studies highlight the advantages of both tensor-based and descriptor-based approaches. Tensor-based methods are preferred when data is uniformly distributed, while descriptor-based approaches help avoid cold-start problems when few known CRCs are available.

For synthesis-condition recommender systems, data acquisition speed is critical. Combining these systems with high-speed, automated synthesis robots can iteratively improve recommendation quality. Although still in its infancy, the application of recommender systems to various problems in materials science and technology holds significant potential.

Accelerating Simulations of Kinetic Processes in Energy Materials with Machine Learning and Data Science

ID: INV07
Invited Talk
Aug. 24, 9.00 - 9.30

Brandon C. Wood
Laboratory for Energy Applications for the Future (LEAF), Lawrence Livermore National Laboratory, Livermore, CA, USA
Email: brandonwood@llnl.gov

Keywords: Molecular dynamics

Many energy storage and conversion devices rely on complex chemical and transport processes that tend to occur at buried materials interfaces that are often difficult to probe experimentally. Computational simulations of these processes can help elucidate interface behavior, but the associated structural and compositional complexity is difficult to address with conventional approaches. Likewise difficult is addressing the wide ranges of length and time scales involved, which span from atomic to microscale and beyond. Recently, machine learning and data science techniques have emerged as important companions to direct simulations to bridge scales and address needed complexities in ways that are inaccessible to human intuition. I will provide a few key examples of how we are using machine learning and data science methods within our LLNL LEAF Center to accelerate and interpret multiscale simulations of kinetic processes in materials relevant for energy storage and production. First, I will illustrate how we integrate machine learning-trained molecular simulations, microstructure analysis, and feature extraction from mesoscale simulations to understand multiscale factors that govern ion transport in solid-state batteries and hydrogen storage materials. I will highlight how the data science approaches have elucidated previously unknown relationships between composition and microstructure in determining performance of these devices. Next, I will show how materials training data from molecular simulations are being implemented in surrogate models and integrated into larger-scale kinetic evolution models to directly simulate changes in the surface oxide composition of metals in electrochemical operating environments. These models reveal emergent phenomena at much longer time scales compared to molecular dynamics, with implications for understanding durability of components in operating energy conversion devices.

Machine Learning Interatomic Potentials for Metal Nanoparticles

R. Vangheluwe, M. Ritopecki, N. Bhatia, R. Tandiana, E. Brun, C. Sicard-Roselli, D. Domin, C. Clavaguéra, and N. T. Van-Oanh†
Institut de Chimie Physique (ICP), Université Paris-Saclay, CNRS, UMR8000, Orsay, France
Email: van-oanh.nguyen-thi@universite-paris-saclay.fr

ID: INV08
Invited Talk
Aug. 24, 9.30 - 10.00

Keywords: machine learning interatomic potentials, metal nanoparticles, bimetallic nanoparticles

Gold nanoparticles are of great interest in advanced radiotherapy where these nanoparticles are often used to enhance radiosensitivity. It has been demonstrated by experiments in our lab that interfacial water plays a central role in the production of OH radicals when irradiating solvated gold nanoparticles. Our group has experience simulating these systems using classical potentials [1,2,3]. However, these methods do not predict well the interaction energy between gold nanoparticles and water molecules. Due to the size of the systems and the timescale, direct first principle molecular dynamics simulations of these systems are prohibitively expensive; hence we employ in this work different machine learning strategies to tackle the problem:

- the delta-machine learning within the framework of Density Functional based Tight Binding (DFTB) model
- the Chebyshev Interaction Model for Efficient Simulation (ChIMES), a physics-informed machine learning interatomic potential (ML-IAP) [4,5]
- and the deep neural network ML-IAP of DeepMD model [6]

The main advantage of these machine learning potentials is that their cost is that of classical potentials, but their accuracy is comparable to quantum level calculations. In this work, we will present the characterization of the interaction energy and structural, dynamical and vibrational properties of the water surrounded gold nanoparticle using these machine learning methods [7]. A perspective of using these machine learning methods for studying the structure of bismuth-platinum nanoparticles is also presented [8].

References

1. R. Tandiana, E. Brun, C. Sicard-Roselli, D. Domin, N.-T. Van-Oanh and C. Clavaguera, J. Chem. Phys. **154**, 044706 (2021).
2. E. Brun, R. Tandiana, M. Gilles, Y. Cheref, N.-T. Van-Oanh, C. Clavaguera and C. Sicard-Roselli, J. Phys. B: At. Mol. Opt. Phys. **57**, 065002 (2024).
3. R. Tandiana et al., in preparation.
4. R. K. Lindsey, L.E. Fried, N. Goldman, J. Chem. Theory Comput. **13**, 6222 (2017).
5. R. K. Lindsey, L.E. Fried, N. Goldman, J. Chem. Theory Comput. **15**, 436 (2019).
6. H. Wang, L. Zhang, J. Han, and E. Weinan, Comput. Phys. Comm. **228**, 178-184 (2018).
7. M. Ritopecki et al., in preparation.
8. R. Vangheluwe et al., in preparation.

Quantum Materials Modeling and Electronic Structure Prediction for Varying Mechanical and Chemical Boundary Conditions

ID: INV09
Invited Talk
Aug. 24, 10.30 - 11.00

Devesh Kale,[1] Taradutt Pattnaik,[1] S. Pamir Alpay,[1,2] and Sanjeev K. Nayak[1]
[1] *Department of Materials Science and Engineering and Institute of Materials Science, University of Connecticut, Storrs, CT 06269, USA*
[2] *Department of Physics, University of Connecticut, Storrs, CT 06269, USA*
Email: sanjeev.nayak@uconn.edu

In quantum materials research, we have reached a stage where the exotic phenomena once predicted by quantum field theory are being observed in real materials. These developments bring significant expectations on computational materials science to reliably predict physical properties. Density functional theory (DFT) remains the most popular approach to modeling material systems in the computational materials domain. In this presentation, we will provide a comprehensive assessment of various DFT formalisms and their effectiveness in capturing the electronic structure of topological insulators (TIs), a class of quantum materials that has semiconducting properties in the bulk, but the surface develops a unique band structure. Our findings indicate that the generalized gradient approximation (GGA) and kinetic density functional (metaGGA) predict an outward relaxation of TI slabs, whereas the local density approximation (LDA) shows opposite behavior. Including van der Waals (vdW) interactions corrects the over-relaxations of the GGA, aligning the atomic positions closer to experimental findings. Based on a systematic computational study, we demonstrate that GGA with vdW treatment is an appropriate method for structural optimization. We show that mechanical boundary conditions (e.g., biaxial strain) and chemical boundary conditions (e.g., doping) can introduce TI properties in several layered tellurides, such as Bi_2Se_3, Bi_2Te_3, As_2Te_3, $ZrTe_5$, and SnTe. These two approaches are independent but can be combined to engineer TI states in high-Z and low-band gap compounds. Additionally, we will also briefly discuss our group's research efforts on optical qubits in 2D semiconductors with point defects.

Keywords: Quantum materials, Topological insulators, Telluride compounds, Qubits, 2D-semiconductrors

Comparing Receptor Binding Properties of SARS-CoV-2 and of SARS-CoV Virus by Using Unsupervised Machine Learning Models

ID: INV10
Invited Talk
Aug. 24, 11.00 - 11.30

Cao Phuong Cong, Hien T. T. Lai, Ly H. Nguyen, and Toan T. Nguyen†
Key Laboratory for Multiscale Simulation of Complex Systems, and Faculty of Physics, University of Science, Vietnam National University, 334 Nguyen Trai Street, Thanh Xuan, Hanoi 11400, Vietnam
Email: toannt@hus.edu.vn

This work continues our recent molecular dynamics investigation of the three systems of the human ACE2 receptor interacting with the viral RBDs of SARS-CoV virus and two variants of SARS-CoV-2 viruses. The simulations are extended and analyzed using unsupervised machine learning models to give complementary descriptions of hidden features of the viral binding mechanism. Specifically, the principal component analysis (PCA) and the variational autoencoder (VAE) models are employed, both are classified as dimensionality reduction approaches with different focuses. The results support the molecular dynamics results that the two variants of SARS-CoV-2bind stronger and more stable to the human ACE2 receptor than SARS-CoV virus does. Moreover, stronger bindings affect the structure of the human receptor, making it fluctuate more, a sensitive feature which is hard to detect using standard analyses. Unexpectedly, it is found that the VAE model can learn and arrange randomly shuffled protein structures obtained from molecular dynamics in time order in the latent space representation. This result potentially has promising ap-plication in computational biomolecules. One could use this VAE model to jump forward in time during a molecular dynamic simulation, and to enhance the sampling of protein configuration space.

Keywords: Coronaviruses, human ACE2, unsupervised machine learnings, enhanced sampling, molecular dynamics, variable autoencoder

Structural and Transport Properties of Actinide-Containing Molten Salts: Machine Learning-based Studies

ID: INV11
Invited Talk
Aug. 24, 11.30 - 12.00

Manh-Thuong Nguyen

Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99354, United States

Email: manhthuong.nguyen@pnnl.gov

Keywords: Machine learning interatomic potentials, molecular dynamics, molten salts, graph theory

Actinide-containing molten salts are a critical part of next generation nuclear reactors. Yet, studying their fundamental properties is highly challenging for both experiment and modeling. In this talk, I will present our recent efforts in understanding these materials at the atomistic level by using statistical mechanics and machine learning. First, I will discuss our development of pseudopotentials and basis sets for the actinide series [1,2] that allow for large scale ab initio molecular dynamics (AIMD) simulations. Next, I will talk about our AIMD studies and machine learning -based analyses of the coordination structure of single actinide ions in host molten salts [3]. Finally, I will present how we used machine learning interatomic potential molecular dynamics and graph theory to discover the breakdown of entropy-scaling laws of the diffusion of actinide ions and learn about their clustering [4,5].

References

1. J.-B. Lu, D.C. Cantu, C.-Q. Xu, M.-T. Nguyen, H.-S. Hu, V.-A. Glezakou, R. Rousseau, J. Li, J. Chem. Theory Comput. **17**, 3360 (2021).
2. M.-T. Nguyen, J. Zhang, D. C. Cantu, R. Rousseau, V.-A. Glezakou, "Tailored Computational Approaches to Interrogate Heavy Element Chemistry and Structure in Condensed Phase", in "Rare Earth Elements and Actinides: Progress in Computational Science Applications", ACS Publications, p. 219-245 (2021).
3. M.-T. Nguyen, B.A. Helfrecht, R. Rousseau, V.-A. Glezakou, J. Mol. Liq. **365**, 120115 (2022).
4. M.-T. Nguyen, R. Rousseau, P. D. Paviet, and V.-A. Glezakou, ACS Appl. Mater. Interfaces **13**, 53398-53408 (2021).
5. M.-T. Nguyen, V.-A. Glezakou, R. Rousseau, P. D. Paviet, Appl. Mater. Today **35**, 101951 (2023).

Inverse Design of Functional Materials

Hongbin Zhang
Theory of Magnetic Materials Group, TU Darmstadt, Otto-Berndt-Str. 3, 64287 Darmstadt, Germany
Email: hongbin.zhang@tu-darmstadt.de

ID: INV12
Invited Talk
Aug. 24, 9.00 - 9.30

Keywords:

Machine learning has been widely applied to obtain statistical understanding and rational design of advanced materials by mapping out the processing-(micro)structure-property–performance relationships. Focusing on the structure - property relationships, I am going to demonstrate the concept of inverse design and to showcase how it can be carried out in three different flavours, i.e., high-throughput (HTP) combinatorial computation, Bayesian optimization, and generative deep learning. For the HTP screening, it is essential to implement and execute automatized high-throughput workflows after identifying the essential benchmarking properties. I am going to show how we are using such workflows to screen for permanent magnets, magnetocaloric, photovoltaic, and thermoelectric materials, with a recent extension to magnetic high entropy alloys. Furthermore, in order to explore the vast design space more efficiently, we propose using Bayesian optimization as an efficient ranking machine with reliable recommendations based on balanced exploration and exploitation. Such an approach has been applied for various experimental characterization techniques such as x-ray absorption spectroscopy, high-resolution transmission electron microscopy, and inelastic neutron scattering. In particular, I am going to show demonstrate that a closed-loop adaptive design framework can be implemented and applied to optimize Invar alloys and magnetocaloric materials. Last but not least, I am going to give an overview on how generative deep learning can be applied to predict novel crystal structures and microstructures based on our recent implementation using the generative adversarial network and diffusion models.

Machine Learning for Materials, a Journey from Artificial Intelligence to Intelligent Materials

Tu Le
School of Engineering, RMIT University, Melbourne, VIC 3000, Australia
Email: tu.le@rmit.edu.au

ID: INV13
Invited Talk
Aug. 24, 9.30 - 10.00

The design and synthesis of materials with useful, novel properties is one of the most active areas of contemporary science, generating a veritable explosion of scientific activity in areas such as biomaterials, cell and tissue engineering, organic photovoltaics and light-emitting materials, and nanomaterials for a myriad of medical and non-medical applications. This new era of materials design and discovery covers many disciplines from chemistry and biology to physics and engineering. Conventionally, it takes at least 20 years to move a material from initial discovery to the marketplace. To accelerate the pace of novel materials discovery, computational methods such as artificial intelligent machine learning techniques can be used to construct predictive materials property models and allow rapid scanning of large chemical datasets to systematically identify attractive candidates for specific applications. This presentation will showcase recent studies on data-driven design of functional materials for a broad spectrum of applications such as drug delivery, antifouling materials, and CO_2 capturing materials.

Keywords:

Applications of Machine Learning in Predicting Electronic Properties in Organic Semiconductor Materials

ID: INV14
Invited Talk
Aug. 24, 10.30 - 11.00

Lam H. Nguyen,[1] Tuan H. Nguyen,[2] Khang M. Le,[1] and Thanh N. Truong[3,†]

[1] *Faculty of Chemistry, VNUHCM-University of Science, 227 Nguyen Van Cu Street, Ho Chi Minh City 700000, Vietnam*

[2] *Department of Chemistry, Emory University, Atlanta, GA 30322, United States*

[3] *Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States*

Email: Thanh.Truong@utah.edu

Predicting electronic properties like the band gaps, ionization potentials, and electron affinities is crucial for designing organic semiconductor materials. Traditionally, one common and efficient way to support this design process is by building quantitative structure-property relationship (QSPR) models. We have developed a new quantum mechanical based QSPR models called Degree of Partial π Overlap (DPO) for predicting these properties in polycyclic aromatic hydrocarbons (PAHs) and their derivatives, which are known to serve as core platform in organic semiconductor material synthesis. These DPO models are shown to achieve accurate predictions within 0.1 eV. However, a significant challenge in QSPR is to derive accurate model descriptors, which is time-consuming and resource intensive. To overcome these obstacles, we utilize machine learning (ML) techniques. The ML models in our studies optimize parameters faster, reduce data requirements, and improve generalization using simpler inputs like SMILES strings of molecular structures instead of detailed descriptors. This transition from manual optimization QSPR methods to practical ML based techniques while maintaining accuracy within an acceptable range, would advance the future of material design.

Keywords: Organic Semiconductor Materials, Machine Learning, Degree of π-Orbital Overlap, Electronic Properties, Quantitative Structure-Property Relationship (QSPR)

References

1. L. H. Nguyen and T. N. Truong, ACS Omega **3**, 8913-8922 (2018).
2. L. H. Nguyen, T. H. Nguyen, and T. N. Truong, ACS Omega **4**, 7516-7523 (2019).
3. T. H. Nguyen, L. H. Nguyen, and T. N. Truong, ACS Omega **7**, 22879-22888 (2022).
4. T. H. Nguyen, K. M. Le, L. H. Nguyen, and T. N. Truong, ACS Omega **8**, 464-472 (2023).
5. T. H. Nguyen, K. M. Le, L. H. Nguyen, and T. N. Truong, ACS Omega **8**, 38441-38451 (2023).

Elucidating Structural Heterogeneity of Materials with Spectroscopy, Machine Learning, and Atomistic Simulations

ID: INV15
Invited Talk
Aug. 24, 11.00 - 11.30

Anh Pham

Quantum Simulations Group & Computational Chemistry and Materials Science Summer Institute, Lawrence Livermore National Laboratory, Livermore, CA 94550, United States

Email: pham16@llnl.gov

Keywords:

Precise determination of atomic structural information in functional materials holds transformative potential and broad implications for emerging technologies. Spectroscopic techniques, such as X-ray absorption near-edge structure (XANES), have been widely used for material characterization; however, extracting chemical information from experimental probes remains a significant challenge, particularly for disordered materials. We present an integrated approach that combines atomic simulations, data-driven techniques and experimental measurements to investigate chemical speciation of amorphous systems, including carbon and carbon nitrides as examples. We discuss the development of machine learning potentials that can efficiently explore the vast configuration space of the systems. By employing statistical methods, this structural database enables the elucidation of the most representative local structures and how they evolve with chemical compositions and density. Density functional theory simulations are used to establish a correlation between the local structure and spectroscopic signatures, which then serves as the basis for interpreting and extracting chemical content from experimental data. Although our framework is specifically demonstrated to XANES, the approach described herein is readily adaptable as applied to other experimental characterization probes and materials classes.

Quantum theoretical modeling of smart catalyst design for clean energy and sustainability

ID: INV16
Invited Talk
Aug. 24, 11.30 - 12.00

Sanjubala Sahoo

Department of Materials Science & Engineering, Institute of Materials Science, University of Connecticut, Storrs, CT, United States

Email: sanjubala.sahoo@uconn.edu

Keywords:

The reliance on fossil fuels has caused major threats to the global climate and environment. As a result, the need for clean energy and environmental sustainability has become a pressing issue. In this regard, hydrogen economy is a transformative vision for a sustainable energy future. Novel catalyst development is important to promote the hydrogen economy for reducing carbon footprint. Such research has been at the forefront and is a key mission for governments worldwide. One of the major roadblocks to implementing the hydrogen economy in terms of technology is the sparse knowledge of physical processes involving the production of hydrogen and its utilization. Hence, there is an active need to explore and understand the atomic and molecular processes occurring at the surfaces/interfaces with hydrogen as a component.

In this presentation, an overview of some of our efforts on critical aspects of catalyst design for hydrogen utilization will be discussed. The research is aimed for energy storage and conversion devices. One such example is the state-of-the-art anion exchange membrane fuel cells (AEMFCs) which offer significant potential for reducing carbon emissions and enhancing energy efficiency [1,2]. The investigations demonstrate a combination of theory and experimental results toward the best catalyst development for hydrogen oxidation reaction (HOR) [1,2]. A set of key descriptors that primarily govern the HOR activity in bi-functional rare-earth transition metal oxides are figured out. The results demonstrate a relatively high HOR performance for the bi-functional system than that of the individual components indicating excellent agreement between the first-principles theory and experimental observations.

References

1. S. Sahoo, D. R. Dekel, R. Maric, and S. P. Alpay, "Atomistic insights into the hydrogen oxidation reaction of palladium-ceria bifunctional catalysts for anion-exchange membrane fuel cells", ACS Catal. **11**, 2561 (2021).
2. M. V. Pagliaro, C. Wen, B. Sa, B. Liu, M. Bellini, F. Bartoli, S. Sahoo, R. K. Singh, S. P. Alpay, H. A. Miller, and D. R. Dekel, "Improving Alkaline Hydrogen Oxidation Activity of Palladium through Interactions with Transition-Metal Oxides", ACS Catal. **12**, 10894 (2022).

On-Lattice Machine Learning Model for Lattice Monte Carlo Simulations of Complex Oxides with First-Principles Accuracy

ID: INV17
Invited Talk
Aug. 24, 13.30 - 14.00

Shusuke Kasamatsu
Faculty of Science, Yamagata University, Yamagata-shi, Yamagata, Japan
Email: kasamatsu@sci.kj.yamagata-u.ac.jp

Keywords: First-principles Calculation, Monte Carlo Calculation, Machine Learning

First-principles calculation based on the laws of quantum physics has established itself as a powerful tool for predicting materials properties from atomistic structures. However, its high computational cost means that it is rather limited in the ability to cover the combinatorially exploding compositional, structural, and configurational space required for materials design. This has started to change in recent years with the development of machine-learning surrogate models that closely mimic first-principles calculations at a small fraction of the computational cost. Especially, machine learning potentials that reproduce interatomic forces and total energies from given atomic coordinates has been instrumental in enabling large-scale long-time molecular dynamics (MD) simulations with first principles accuracy [1]. Here, we discuss the application of such potential models to lattice Monte Carlo (MC) simulation, which, unlike MD, enables fast equilibration and sampling of atomistic configurations at finite temperature even for slowly relaxing systems. This is especially advantageous for computational prediction of the configurations and physical properties of many-component crystalline systems such as metal alloys and complex oxides. To maximize the efficiency for lattice MC simulation, we propose to train machine learning potentials to predict the total relaxed energy from ideal on-lattice coordinates. This on-lattice model is trained in an iterative manner: an initial (possibly inaccurate) model is used to generate "thermodynamically balanced" MC samples, which are added to the training set if the error is larger than a given threshold. The model training and data set generation using the model is performed iteratively until sufficient accuracy is reached. A python framework for facilitating this process has been implemented in our open-source abICS code [2,3]. In the talk, I will discuss the application of this approach to ion conducting oxides and their interfaces [3,4].

References

1. J. Behler and G. Csányi, Eur. Phys. J. B **94**, 142 (2021).
2. S. Kasamatsu *et al.*, J. Chem. Phys. **157**, 104114 (2022).
3. S. Kasamatsu *et al.*, Sci. Technol. Adv. Mater. Meth. **3**, 2284128 (2023).
4. K. Hoshino, S. Kasamatsu *et al.*, Chem. Mater. **35**, 2289 (2023).

Studying Additional Third-Order Transitions in the Two-Dimensional Ising Model via Machine Learning

ID: INV18
Invited Talk
Aug. 24, 14.00 - 14.30

Pham Long Nhat,[1] Nguyen Van Viet Hoang,[1] Nguyen Duc Dung,[1] Vu Hai,[2] Hoang Van Nam,[2] Duong Xuan Nui,[1] and Dao Xuan Viet[1,†]

[1] *School of Material and Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam*

[2] *School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam*

Email: viet.daoxuan@hust.edu.vn

We study the phase transitions of the two-dimensional (2D) Ising model using machine learning techniques. The 2D Ising model, a cornerstone of statistical mechanics, is known for its well-documented second-order ferromagnetic-paramagnetic phase transition. Recent studies have suggested the presence of additional third-order ferromagnetic-ferromagnetic transitions within the ferromagnetic phase and another third-order transition within the paramagnetic phase. To further investigate these complex phenomena, we apply machine learning methods to analyze the phase transitions in the 2D Ising model. Our approach facilitates the identification and characterization of these transitions, providing deeper insights into the critical behavior of the 2D Ising model.

Keywords: Ising model, Machine learning, Third-order phase transition

An Innovative Approach for Watching Dynamic Conformational Changes in Proteins at Ultrafast and Atomistic Levels

ID: INV19
Invited Talk
Aug. 24, 15.00 - 15.30

Kien Xuan Ngo
Nano Life Science Institute (WPI-NanoLSI), Kanazawa University, Kanazawa, Ishikawa 920-1192, Japan
Email: ngoxuankien@staff.kanazawa-u.ac.jp

Keywords: High speed atomic force microscopy, SAI-AFM, alpha actinin, actin

Addressing the temporal and spatial resolution discrepancies between experimental high-speed atomic force microscopy (HS-AFM) data and simulation data necessitates a comprehensive strategy. This strategy employs advanced computational methods such as novel molecular dynamics (MD) and coarse-grained (CG) models, as well as artificial intelligence (AI) techniques. The primary challenge in studying "Protein Dynamics" in real time is bridging the time scale gap: HS-AFM captures dynamics in the millisecond to second range, while simulations typically operate from femtoseconds to microseconds. Additionally, simulations face spatial modeling constraints, limiting the study of large, complex biological systems. HS-AFM, however, allows for the construction of detailed protein models, enabling high temporal imaging of their structures and dynamics during functional activity.

We are developing an innovative method called Simulation Artificial Intelligence-Atomic Force Microscopy (SAI-AFM) to visualize dynamic structural changes in proteins at ultrafast and atomistic levels. This novel workflow integrates multiscale computational models (MDs and CGs) with AI-driven analysis and HS-AFM data across multiple timescales and spaces. The advanced SAI-AFM technology allows for concurrent observation of protein structures, dynamics, and functions at the ultrafast and atomistic levels, surpassing individual methods.

Our goals include:

1. Creating a library of experimental PDBs, including saiPDBs, that accurately reflect the atomic structures of protein dynamics in real time, bridging the time and space gaps between individual methods.
2. Applying SAI-AFM to uncover dynamic conformational changes in alpha actinin at super-high spatiotemporal resolutions.

Accelerated Discovery of Cathode Materials for Li-ion Batteries using General-Purpose Neural Network Potentials

ID: INV20
Invited Talk
Aug. 24, 13.30 - 14.00

Tien Quang NGUYEN and Michihisa KOYAMA
Research Initiative for Supra-Materials, Shinshu University (〒380-8553 Nagano, Nagano-shi, Wakasato 4-17-1), Japan
Email: quang@shinshu-u.ac.jp

Keywords: Li-ion Battery, Neural Network Potential, Multi-Canonical Sampling, Global Optimization

With the rapid development of mobile technology and electric vehicles, research and improvement of Li-ion batteries (LIBs) have recently garnered significant attention from engineers and scientists. A modern Li-ion battery consists of two electrodes separated by a non-aqueous liquid electrolyte. Typically, the anode is made of graphite, while the cathode is either a layered oxide or spinel. During charging and discharging, Li ions move between the electrodes. Despite their high energy density, great power, and long-life cycle, Li-ion batteries still face operational challenges such as electrode degradation. The cathode, in particular, encounters significant issues due to the insertion and removal of Li ions, leading to phase transformations, surface reconstruction, and electrolyte decomposition in contact with the cathode. Consequently, discovering new cathode materials is essential. Traditionally, materials discovery often consumes a vast amount of time and resources in experiments and computational modelling/simulation because of the limitations of experimental conditions or theoretical foundations. Thus, it is crucial to approach this task in a new way. In recent years, the availability of large experimental and simulation datasets combined with advanced algorithms and computational resources has heightened interest in data-driven and machine learning methods for materials discovery, achieving notable improvements in both time efficiency and prediction accuracy. One promising approach is the use of neural network potentials (NNPs), which model atomic interactions by mimicking neural networks in the human brain. These potentials retain the accuracy of quantum mechanical calculations while operating at the speed of force field calculations.

In this talk, we will present our current development of a high-throughput screening scheme for exploring multi-component cathode materials in Li-ion batteries using general-purpose NNPs [1], in combination with global optimization techniques. From this scheme, several candidate materials will be revealed. Additionally, we will discuss the effects of mixing entropy in multi-component cathode materials by employing the general-purpose NNP alongside the multi-canonical sampling-based Wang-Landau method [2-4] to study the thermodynamic properties and cation ordering within the transition metal layers of $LiNi_xCo_yMn_{1-x-y}O_2$ (NCM) layered oxides. With the acceleration provided by the general NNP, configurational spaces of several NCM compounds are explored, and configurational density of states are calculated. For each compound, different transition temperatures are identified from the calculation of configurational specific heats, depending on the compositions of metal components. The approaches introduced here not only accelerate the discovery of high-performance cathode materials but also provide insights into the thermodynamic behavior and stability influenced by cation mixing.

References

1. S. Takamoto *et al.*, Nat. Commun. **13**, 2991 (2022).
2. B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249-253 (1991).
3. F. Wang and D. Landau, Phys. Rev. Lett. **86,** 2050-2053 (2001).
4. T. Q. Nguyen, Y. Nanba, and M. Koyama, J. Comput. Chem. Jpn. **21**, 111-117 (2022).

Inverse Design in Photonics Using Finite Different Time Domain and Gradient-based Methods

ID: INV21
Invited Talk
Aug. 24, 14.00 - 14.30

Van Doan Le,† Rémi Colom, and Samira Khadir
Université Côte d'Azur, CRHEA, CNRS, 06560 Valbonne, France
Email: van-doan.le@crhea.cnrs.fr

Inverse design in photonics is a technique that leverages computational algorithms to create photonic components by specifying desired optical performance and subsequently determining the physical structure that achieves it. This process utilizes the adjoint method to compute the gradient of the objective function with respect to design parameters. In this tutorial, we demonstrate the efficient design of a photonic component, beginning with the solution of Maxwell's equations using the finite-difference time-domain (FDTD) method. Computational algorithms are employed to iteratively modify the design and evaluate its optical behavior against defined objectives. The optimization process continues until it converges to a design that meets the specified performance metrics. Additionally, the necessity of incorporating fabrication constraints into the inverse design optimization is discussed.

Keywords: Inverse design, Photonics, FDTD, Gradient-based method, Adjoint method

On-the-fly Machine Learning Potential Accelerates Accurate Lattice Thermal Conductivity Prediction of the Pentagonal Structured Monolayer Materials

ID: INV22
Invited Talk
Aug. 24, 15.00 - 15.30

Nguyen Thanh Tien,[1,†] Vo Khuong Dien,[2] and Pham Thi Bich Thao[1]

[1] *College of Natural Sciences, Can Tho University, 3-2 Road, Can Tho City 94000, Vietnam*

[2] *Department of Electrophysics, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan*

Email: nttien@ctu.edu.vn

In this study, we propose an innovative approach that combines the phonon Boltzmann transport equation (PBTE) with on-the-fly machine learning potential (FMLP) to accelerate the accurate prediction of lattice thermal conductivity of pentagon structured monolayer materials with composite two types of atoms. We overcome the traditional trade-off between accuracy and speed by integrating the precision of density functional theory (DFT) with the computational efficiency of empirical potentials. This way to hugely speed up these calculations is by using force fields and parametrizations of the potential energy. Our method is applied to predict the thermal conductivity of various penta-structured monolayers, revealing insights into their thermal transport properties. These findings have implications for designing advanced materials for thermoelectric applications and thermal management.

Keywords: Lattice thermal conductivity, monolayer material, pentagonal structures, machine learning

An Interpretable Data-Driven Approach for Decision-Making in Materials Discovery Amidst Uncertainties

ID: ORL02
Oral Talk
Aug. 23, 14.20 - 14.40

Minh-Quyet Ha and Hieu-Chi Dam
Japan Advanced Institute of Science and Technology, Nomi, Japan
Email: dam@jaist.ac.jp

Advanced materials exhibit exceptional properties and complex structures due to their multiple constituent elements, offering a wide range of potential candidates for new material development. However, the experimental validation and simulation of these candidates demand significant resources. Therefore, material scientists must make informed decisions about candidate selection while managing resource constraints. The primary challenge lies in evaluating the rationality of each decision, balancing two key objectives: exploitation and exploration. Exploitation focuses on maximizing immediate benefits by selecting candidates with high-performance potential based on existing knowledge. Exploration, on the other hand, seeks to expand understanding by choosing lesser-known or underexplored candidates. To quantitatively assess the rationality of these decisions, it is essential to model the uncertainties associated with their outcomes. Two main types of uncertainties are considered: aleatoric uncertainty, which arises from inherent randomness and unpredictable individual events, and epistemic uncertainty, which results from a lack of knowledge or data and can be reduced with additional information. Exploitation typically favors candidates with a higher probability of success but must address aleatoric uncertainty. Exploration broadens the understanding of the materials space, contending with aleatoric and epistemic uncertainties. This research introduces a data-driven approach that models the uncertainties of decision outcomes, supporting scientists in making rational choices. The framework employs quantitative methods to assess material similarity, providing a basis for analogy-based, intuitive, and interpretable reasoning. This approach is further supported by a quantitative model for uncertainty grounded in the Dempster-Shafer theory, a generalization of Bayesian theory, enabling the quantitative measurement of both aleatoric and epistemic uncertainties.

Keywords: Material informatics, Decision-making, Uncertainty modelling, Dempster-Shafer theory

Material Dynamics Analysis with Deep Generative Model

Duc-Anh Dao,[1] Tien-Sinh Vu,[1] Yoshifumi Oshima,[1] Masahiko Tomitori,[1] Yukio Takahashi,[2] Hieu-Chi Dam,[1,2,3†]

[1] *Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan*

[2] *International Center for Synchrotron Radiation Innovation Smart (SRIS), Tohoku University, 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan*

[3] *ESICMM, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan*

Email: dam@jaist.ac.jp

ID: ORL04
Oral Talk
Aug. 23, 14.00 - 14.20

Studying material dynamics is essential for predicting degradation, failure, and specific behaviors under various conditions, enabling efficient material design, enhanced performance, and driving innovation in materials science. Microscopic techniques have advanced significantly to document the dynamic evolution of materials during experimental progression. However, experimental observations from microscopy images are inherently bias to drive subsequent analysis to observable behaviors and excluding those in unobserved scenarios. To address this issue, we designed an analysis framework integrating a deep generative model to synthesize hypothetical material states that could emerge under realistic experimental conditions. This approach enables Monte Carlo simulations of material state variations driven by dynamic progression, reflecting more generalized and predictive dynamic behaviors. Applied to various microscopy image datasets, our framework effectively uncovers hidden physical phenomena in materials science experiments. Furthermore, this research supports a shift towards data-driven methodologies, advocating the adoption of deep learning technologies to foster innovative research practices in materials science.

Keywords: material dynamics, data analysis, microscopy images, deep generative model

Study of the IV characteristics for quantum-confined bio-molecular nanostructure: DFT and Machine Learning based combined framework

ID: ORL05
Oral Talk
Aug. 24, 15.30 - 15.50

Debarati Dey Roy,[1,†] Pradipta Roy,[2,†] and Nguyen Thanh Tien[3]

[1] *Department of Electronics & Communication Engineering, B. P. Poddar Institute of Management & Technology. 137, V. I. P Road, Kolkata-700052, West Bengal, India*

[2] *Dept. of Computer Application, Dr. B. C. Roy Academy of Professional Courses. Fuljhore, Jemua Road, Durgapur-713206. West Bengal, India*

[3] *College of Natural Sciences, Can Tho University, Campus II, Can Tho University, 3/2 street, Ninh Kieu District, Can Tho City, Vietnam*

Email: debaratidey24@gmail.com, pradiptoroy@gmail.com

Keywords: Machine Learning, DFT, Artificial Intelligence, I-V Characteristics, Conductivity

Machine Learning (ML) application is a boon for electronic structure prediction. The combined framework of Density Functional Theory (DFT) and ML algorithms makes the prediction work easier for Nanotechnologists. The algorithmic approach of ML towards the electronic structure theory permits the prediction of electronic characterization for molecular models. The time-consuming task of electronic structural prediction for quantum confinement into nano-dimension systems became easier with the introduction of Artificial Intelligence (AI). This article presents the best-fitted prediction model for the quantum-confined Adenine molecular nanoribbon using DFT and ML-based combined framework. Our framework introduces an algorithm to predict electronic characteristics, such as the current-voltage response graph and the graphical presentation of the conductivity with high accuracy. The regression learner will predict the graphical representation of various electronic characteristics of the quantum-confined bio-inspired nanostructure. A learnable real-space Hamiltonian procedure and k-point sampling are introduced to extract the information and feed it into the machine to train it to create a future prediction model for this quantum-confined Adenine nanoribbon. The best-fitted regression learner provides approximately 85% accuracy for this quantum-confined Adenine nanoribbon structure. This framework provides a faster and more accurate prediction model for electronically characterizing the quantum-confined bio-inspired nanoscale structure [1-4].

References

1. D. D. Roy and D. De, "Predicting model of I–V characteristics of quantum-confined GaAs nanotube: a machine learning and DFT-based combined framework", J. Comput. Electron. **22**, 999-1009 (2023).
2. D. D. Roy, P. Roy, and D. De, "Machine learning and DFT-based combined framework for predicting transmission spectra of quantum-confined bio-molecular nanotube". J. Mol. Model. **29**, 338 (2023).
3. D. Dey, P. Roy, and D. De, "Electronic characterisation of atomistic modelling based electrically doped nano bio p-i-n FET", IET Comput. Digit. Tech. **10**, 273-285 (2016).
4. O. T. Hofmann, E. Zojer, L. Hörmann, A. Jeindl, and R. J. Maurer, "First-principles calculations of hybrid inorganic–organic interfaces: from state-of-the-art to best practice", Phys. Chem. Chem. Phys. **23**, 8132-8180 (2021).

Quantitative Structure–Property Relationship in Predicting Electronic Properties of Polycyclic Aromatic Hydrocarbons and Derivatives from Degree of π-Orbital Overlap to High-Throughput Screening

ID: PST01
Poster
Aug. 24, 16.00 - 17.30

Lam H. Nguyen,[1] Tuan H. Nguyen,[2] Khang M. Le,[1] and Thanh N.Truong[3,†]

[1] *Faculty of Chemistry, VNUHCM-University of Science, 227 Nguyen Van Cu Street, Ho Chi Minh City 700000, Vietnam*

[2] *Department of Chemistry, Emory University, Atlanta, GA 30322, United States*

[3] *Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States*

Email: Thanh.Truong@utah.edu

In our quest to revolutionize the prediction of electronic properties in organic semiconductors, we have advanced the ability to predict key properties such as the Highest Occupied Molecular Orbital-Lowest Unoccupied Molecular Orbital (HOMO-LUMO) gap, ionization potentials, and electron affinities for polycyclic aromatic hydrocarbons (PAHs) and their derivatives. Initially, we manually developed QSPR models using the "degree of π-orbital overlap" (DPO) descriptor, achieving accurate predictions within 0.10 eV using Density Functional Theory. However, defining descriptor notation was time-consuming and resource-intensive. To enhance efficiency and precision, we transitioned to machine learning (ML) techniques. ML optimized the DPO parameters, reducing data requirements while maintaining high accuracy. Using the Weisfeiler-Lehman graph kernel method and Gaussian process regressor, our ML-enhanced models predicted electronic properties across 2131 aromatic molecules with remarkable precision, achieving root-mean-square deviations as low as 0.15 eV. This blend of DPO descriptors and ML techniques confirms the robustness of the DPO model and sets the stage for future advancements in organic semiconductors. Our journey from manual methods to sophisticated ML techniques highlights the dynamic evolution of our approach and its transformative impact on predicting electronic properties of PAH-based organic materials.

Keywords: Organic Semiconductors, Machine Learning, Degree of π-Orbital Overlap, Electronic Properties, Quantitative Structure–Property Relationship (QSPR)

References

1. L. H. Nguyen and T. N. Truong, ACS Omega **3**, 8913-8922 (2018).
2. L. H. Nguyen, T. H. Nguyen, and T. N. Truong, ACS Omega **4**, 7516-7523 (2019).
3. T. H. Nguyen, L. H. Nguyen, and T. N. Truong, ACS Omega **7**, 22879-22888 (2022).
4. T. H. Nguyen, K. M. Le, L. H. Nguyen, and T. N. Truong, ACS Omega **8**, 464-472 (2023).
5. T. H. Nguyen, K. M. Le, L. H. Nguyen, and T. N. Truong, ACS Omega **8**, 38441-38451 (2023).

Unravelling polaron and bipolaron in (Li, Na)-doped V2O5 materials: DFT+U computational method

ID: PST02
Poster
Aug. 24, 16.00 - 17.30

Huu T. Do
Department of Chemical Engineering, Chicago, Illinois, USA
Email: huutdo09@gmail.com

Pristine and (Li, Na)-doped (α, β)-V2O5 polymorphs emerging as quintessential exemplars in manifold of practical applications, especially for new generations of (Li, Na)-battery cathode materials, as well as for probing exotic fundamental electronic properties. α-V2O5 characterize as a d^0 charge-transfer insulator with strong O-p–V-d hybridization together with a large band gap, while β-$Na_{0.33}V_2O_5$ exhibits metal-insulator transition accompanied by a charge density wave (CDW) gap. With the highest oxidation state d^{+5} and layer structure, V_2O_5 has large potential to intercalate mobile alkaline (Li, Na) and alkaline earth elements which donate electrons to the framework. Consequently, the extra electrons routinely induce a polaronic mechanism in which they couple with available lattice distortions. It is a perennial issue that lonely density functional theory (DFT) faces challenging to approach bandgap and strongly correlated properties. In this talk, we perform the rigorous Hubbard U correction (DFT+U) is to characterize exactly the band gap of d-state charge-transfer V_2O_5 insulator as well as the CDW gap of β-$Na_{0.33}V_2O_5$. Remarkably, our calculations showcase capacity to unravel the presence of the free polaron in Li-doped α-V_2O_5 as well as the coincident quantum criticality of bipolaron-to-polaron, and energetic favorable antiferromagnetic-to-ferromagnetic transitions in β-phase.

Keywords: V_2O_5, β-$Na_{0.33}V_2O_5$, charge-transfer insulator, polaron, bipolaron-to-polaron transition, ab initio method

Prediction of carbon clusters using machine-learning potential

Huy Duy Nguyen
University of Science, Vietnam National University, Hanoi, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam
Email: huydn@hus.edu.vn

ID: PST03
Poster
Aug. 24, 16.00 - 17.30

In our investigation of twelve-atom carbon clusters, we employ the recently developed machine-learning potential GAP-20 to predict the low-energy isomers. The GAP-20 model exhibits good agreement with density-functional theory calculations for most isomers, accurately capturing geometric structures and average C-C bond lengths. However, the GAP-20 significantly underestimates the energies associated with cage-like structures, leading to an incorrect ground state prediction. The Jahn-Teller distortion associated with the monocyclic ring is also absent. This intriguing behavior prompts further exploration. Comparing cohesive energies, the GAP-20 performs well for monocyclic rings. In addition, it also reveals two novel low-energy isomers composed of multicyclic rings. These previously undiscovered structures challenge our existing understanding and open up exciting avenues for future research. The GAP-20 provides valuable insights into the intricate landscape of carbon clusters by shedding light on these multicyclic configurations.

Keywords: Carbon clusters, Low-energy isomers, Machine-learning potential

Artificial Intelligence/Machine Learning Advances in Screening Lithium-ion Battery Material Guide to Electrodes and Electrolytes

ID: PST04
Poster
Aug. 24, 16.00 - 17.30

Huu Doanh Nguyen,[1,2] Lam Tien Pham,[3] and Viet Bac T. Phung[1,2,†]

[1] *Digital Materials Science Lab, Center for Environmental Intelligence, VinUniversity, Hanoi, Vietnam*

[2] *College of Engineering & Computer Science, VinUniversity, Hanoi, Vietnam*

[3] *Phenikaa Institute for Advanced Study (PIAS) and Faculty of Computer Science, Phenikaa University, Ha Noi, Vietnam*

Email: bac.ptv@vinuni.edu.vn

This paper aims to demonstrate the advancements in utilizing artificial intelligence/machine learning (AI/ML) models to screen materials for lithium-ion batteries (LIBs). These advancements primarily focus on active electrodes and solid electrolytes, which are crucial for the next generation of LIBs. Consequently, extensive research and numerous papers have been published, showcasing various ML models used in LIB materials design and synthesis. The mathematical descriptors, which are key elements for ML models, vary depending on the target, including Fingerprints, SMILES, Potentials, Weighted Graph, Coulomb Matrix, Bag of Bonds/Fragments, and 3D Geometry [1]. Despite this, complete atomic representations of broad chemical spaces are still being researched, and developing a general model for all circumstances remains challenging. Therefore, a summary and guide of which model is suitable for each case of utilizing AI/ML methods is essential, and important in addressing the challenges of AI/ML accelerators in battery design. By integrating computational models with experimental data, AI/ML methods streamline the discovery process, predict material behavior, and facilitate the development of next-generation lithium-ion batteries. The potential impact of these technologies on sustainable energy storage solutions is significant, with promising breakthroughs in energy density, charging rates, and overall battery lifecycle management [2].

Keywords: Artificial Intelligence (AI), Machine Learning (ML), Lithium-ion batteries (LIBs), Electrode materials, Solid electrolyte, ML descriptors

References

1. B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: Generative models for matter engineering", Science **361**, 360 (2018).
2. T. Lombardo, M. Duquesnoy, H. El-Bouysidy, F. Årén, A. Gallo-Bueno, P. B. Jørgensen, A. Bhowmik, A. Demortière, E. Ayerbe, F. Alcaide, and M. Reynaud, "Artificial intelligence applied to battery research: hype or reality?", Chem. Rev. **122**, 10899 (2021).

Evaluating Contrastive Learning methods performance on Many-body UMLIPs architecture

ID: PST05
Poster
Aug. 24, 16.00 - 17.30

Thai-Bach Le,[1] Huu T. Do,[2] and Truong Son Hy[3]
[1] *Fulbright University Vietnam, Ho Chi Minh City, Vietnam*
[2] *University of Illinois, Chicago, Illinois, USA*
[3] *Indiana State University, Indiana, USA*
Email: TruongSon.Hy@indstate.edu

Pre-training techniques leverage unlabeled data to address the challenges of scarce and resource-intensive labeled datasets, while enhancing out-of-distribution prediction performance. Recent advancements in Universal Machine Learning Interatomic Potential (MLIP) architectures which incorporated many-body interaction techniques, have achieved a balance between ab initio accuracy and the computational efficiency of force field methods. Despite extensive research on pre-training techniques, most studies have focused on simpler architectures like GemNet and GIN on organic domain, leaving a gap in evaluating MLIP architectures and their application to inorganic materials. In this study, we assess node-level pre-training techniques applied to the M3GNet architecture. We utilized the OQMD Dataset with The Materials Project for unsupervised learning, and employ The Materials Project with Formation Energy for the downstream task. Our findings are consistent with prior works, demonstrating the limitations of naive contrastive learning methods as disrupt causation in dataset molecular structures. Additionally, our results indicate promising potential for generative self-supervised training in future MLIPs models toward transfer learning into specific domains.

Keywords: Pretraining techniques, Universal Machine Learning Interatomic Potential (MLIP), Many-Body Interaction, Contrastive Learning, Generative Self-Supervised Training, GNN, M3GNet Architecture, OQMD Dataset, The Materials Project

Applications of generative molecular AI in drug design

Tri M. Nguyen[1] and Thanh N. Truong[2]

[1] *Faculty of Chemical Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam*

[2] *Department of Chemistry, University of Utah, Salt Lake City, Utah, United States*

Email: Thanh.Truong@utah.edu

ID: PST06
Poster
Aug. 24, 16.00 - 17.30

Drug development is a traditionally long and very expensive task consisting of many steps including lead discovery and lead optimization. For this reason, a popular approach due to its cost-effectiveness currently is drug repurposing, in which one attempts to modify or even use an existing drug for new disease targets. Recent advancements in artificial intelligence (AI) have shown great potential in streamlining this process. Many applications of machine learning methodology in this approach necessitate an AI model capable of modifying the molecular structure of an initial drug according to relevant metrics to meet specific therapeutic requirements. In this study, we assess the performance of these AI models for molecular generation in drug design, focusing on their effectiveness in generating druglike and plausible molecules. We place particular emphasis on how the representation of molecules as input affects the training and outcomes of the models.

Keywords: generative AI, drug design, molecular representation

Development of Quasicrystals Datasets and Applications to Machine Learning

ID: PST07
Poster
Aug. 24, 16.00 - 17.30

Erina Fujita,[1†] Chang Liu,[1] Yukari Katsura,[2,5,6] Kaoru Kimura,[1] Asuka Ishikawa,[3] Ryuji Tamura,[3] Tomoya Mato,[2] Koichi Ktahara,[4] Keiichi Edagawa,[7] Ryo Yoshida[1]

[1] *The Institute of Statistical Mathematics (ISM), Tachikawa, Tokyo, Japan*
[2] *National Institute for Materials Science (NIMS), Tsukuba, Ibaraki, Japan*
[3] *Tokyo University of Science, Katsushika- ku, Tokyo, Japan*
[4] *National Defense Academy, Yokosuka, Kanagawa, Japan*
[5] *Tsukuba University, Tsukuba, Ibaraki, Japan*
[6] *RIKEN, Chuo-ku, Japan*
[7] *The University of Tokyo, Meguro-ku, Tokyo, Japan*
Email: `fujita-e@ims.ac.jp`

Keywords: Materials Informatics, Open Data, Experimental Data, Quasicrystal, Physical Properties

Quasicrystals are solid-state materials with unique symmetry not allowed in ordinal crystals and they were first discovered in 1984. In recent years, machine learning has been employed to explore quasicrystals with unique properties inherent to quasiperiodic systems. However, the lack of open data on quasicrystal composition, structure, and physical properties has hindered the widespread use of machine learning in quasicrystal research. As the quasiperiodic materials exhibit distinct characteristics different from conventional periodic systems, experimentally observed data is considered to play a major role of material informatics in this field.

Comprehensive literature review and manual data extraction were conducted to develop open datasets consisting of compositions, structure types, phase diagrams, sample preparation processes and temperature-dependent thermal, electrical and magnetic properties for a wide range of stable and metastable quasicrystals and approximants [1].

Those datasets were applied to machine learning. Specifically, (1) prediction and validation of quasicrystal forming compositional regions by machine learning [2, 3], (2) prediction of the existence of quasicrystals from experimental XRD pattern images [4], and (3) exploration of thermal diode materials using experimental temperature-dependent properties. Further research utilizing those datasets are expected to accelerate the understanding of QC characteristics.

References

1. HYPOD: `https://doi.org/10.6084/m9.figshare.25650705.v2`
2. C. Liu, E. Fujita, Y. Katsura, Y. Inada, A. Ishikawa, R. Tamura, K. Kimura, and R. Yoshida, "Machine learning to predict quasicrystals from chemical compositions", Adv. Mater. **33**, 2102507 (2021).
3. C. Liu, K. Kitahara, A. Ishikawa, T. Hiroto, A. Singh, E. Fujita, Y. Katsura, Y. Inada, R. Tamura, K. Kimura, and R. Yoshida, "Quasicrystals predicted and discovered by machine learning", Phys. Rev. Mater. **7**, 093805 (2023).
4. H. Uryu, T. Yamada, K. Kitahara, A. Singh, Y. Iwasaki, K. Kimura, K. Hiroki, N. Miyao, A. Ishikawa, R. Tamura, and S. Ohhashi, "Deep learning enables rapid identification of a new quasicrystal from multiphase powder diffraction patterns", Adv. Sci. **11**, 2304546 (2024).

Industry-Academia Consortium for Co-creating Polymer Property Database

Aiko Takahashi, Yoshihiro Hayashi, and Ryo Yoshida
The Institute of Statistical Mathematics, Tokyo, Japan
Email: a-taka@ism.ac.jp

ID: PST08
Poster
Aug. 24, 16.00 - 17.30

Keywords: Polymers, industry-academia consortium, supercomputer, database, molecular dynamics, materials informatics.

Background and Objective. An interdisciplinary field of materials research called "Materials Informatics (MI)" is gaining significant attention. However, the development of polymeric material database to be utilized for data-driven research has been significantly delayed due to several reasons such as (1) the complexity of the materials space, (2) the high cost of data acquisition, and (3) the lack of incentive for researchers to disclose their laboratory data.

To overcome the hurdle of limited data resource for polymeric materials, we have developed a fully automated polymer property calculation system RadonPy [1,2] and a large computational database. The database is used as a dataset of machine learning, specifically Simulation-to-Real (Sim2Real) transfer learning: a source model pretrained with the RadonPy database is fine-tuned to obtained a highly generalizable predictor for real-world systems using limited experimental data.

Currently, an industry-academia consortium has been formed to develop one of the world's largest open databases for polymer properties using RadonPy. The goal is to create a comprehensive database encompassing more than 100,000 molecular skeletons with a wide variety of thermophysical, mechanical, electrical and optical properties for various polymer systems.

Consortium Overview. The consortium currently consists of 241 members from 1 national research institute, 9 universities, and 36 companies, collaboratively producing and sharing data across organizational boundaries. Since 2021, our project has been supported by the "Fugaku" supercomputer program, providing massive amounts of computational resources. This enabled the calculation of various properties for approximately 110,000 amorphous polymers using RadonPy. Our database successfully revealed a polymer property world map. The monthly meetings and Slack communication facilitate collaboration and knowledge exchange, fostering the development of members who may not have expertise in MD simulation or high-performance computing.

Outlook. The consortium aims to release an open database of polymer properties encompassing over 100,000 molecular skeletons. Starting this fiscal year, it seeks to involve automated synthesis and experimental research groups to establish a foundation for data-driven polymer materials research.

References

1. Yoshihiro Hayashi, Junichiro Shiomi, Junko Morikawa, Ryo Yoshida, "RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics", npj Comput. Mater. **8**, 222 (2022).
2. Source code: https://github.com/RadonPy/RadonPy

Participants

We will progressively update this list. If you note any missing names and/or and incorrect information, please let us know at huantd@gmail.com (Huan Tran). We appreciate your helps.

No.	First Name	MI	Last Name	Affiliation	Country
1	Nam Son		Bui	VNU University of Natural Science	Vietnam
2	Hieu-Chi		Dam	Japan Advanced Institute of Science and Technology & Tohoku University	Japan
3	Phuc Dam		Dang	Can Tho University	Vietnam
4	Duc Anh		Dao	Japan Advanced Institute of Science and Technology	Japan
5	Viet	X.	Dao	Hanoi University of Science and Technology	Vietnam
6	Debarati		Dey Roy	B. P. Poddar Institute of Management and Technology	India
7	Ngoc Anh		Doan	VNU University of Natural Science	Vietnam
8	Claudia		Draxl	Humboldt-Universität zu Berlin	Germany
9	Laurent		El Ghaoui	VinUniversity	Vietnam
10	Erina		Fujita	The Institute of Statistical Mathematics	Japan
11	Cap		Ha	ON Vietnam LLC	Vietnam
12	Hoang		Ha	Can Tho University	Vietnam
13	Minh Quyet		Ha	Japan Advanced Institute of Science and Technology	Japan
14	Hiroyuki		Hayashi	Kyoto University	Japan
15	Nam Nhat		Hoang	Vietnam National University	Vietnam
16	Huy		Huynh Anh	Can Tho University	Vietnam
17	Shusuke		Kasamatsu	Yamagata University	Japan
18	Hiori		Kino	National Institute for Materials Science	Japan
19	Nhat Thanh		Le	Can Tho University	Vietnam
20	Thai-Bach		Le	Fulbright University Vietnam	Vietnam
21	Tu		Le	RMIT University	Australia
22	Van Doan		Le	Université Côte d'Azur	France
23	Sanjeev		Nayak	University of Connecticut	USA
24	Xuan Kien		Ngo	Kanazawa University	Japan
25	Cong Dat Vinh		Nguyen	Can Tho University	Vietnam
26	Duc Quan		Nguyen	VNU University of Natural Science	Vietnam
27	Duy Huy		Nguyen	Vietnam National University	Vietnam
28	Hai Dang		Nguyen	Nam Can Tho University	Vietnam
29	Lam	H.	Nguyen	VNUHCM University of Science	Vietnam
30	Manh-Thuong		Nguyen	Pacific Northwest National Laboratory	USA
31	Minh Tri		Nguyen	VNUHCM University of Science	Vietnam
32	Phuc		Nguyen	Can Tho University	Vietnam
33	Toan	T.	Nguyen	Vietnam National University	Vietnam
34	Thanh Tien		Nguyen	Can Tho University	Vietnam
35	Thi Ngoc Mai		Nguyen	VNU University of Natural Science	Vietnam
36	Tri	M.	Nguyen	Ho Chi Minh City University of Technology	Vietnam
37	Son		Nguyen	VNU University of Natural Science	Vietnam
38	My Duyen		Nguyen	VNU University of Natural Science	Vietnam
39	Tien Quang		Nguyen	Shinshu University	Japan
40	Viet Anh		Nguyen	Chinese University of Hong Kong	China
41	Viet Cuong		Nguyen	Intelligent Integration Co., Ltd.	Vietnam
42	Loc		Nguyen Duc	University of Natural Science	Vietnam
43	Doanh		Nguyen Huu	VinUniversity	Vietnam
44	Lan Anh		Nguyen Thi	VNU University of Natural Science	Vietnam

45	Van-Oanh		Nguyen-Thi	Université Paris-Saclay	France
46	Anh		Pham	Lawrence Livermore National Laboratory	USA
47	Phong		Pham	Hanoi University of Science and Technology	Vietnam
48	Binh		Phan	VNU University of Science	Vietnam
49	T Viet Bac		Phung	VinUniversity	Vietnam
50	Kanta		Ono	Osaka University	Japan
51	Khuong	P.	Ong	Institute of High Performance Computing	Singapore
52	Dibya Prakash		Rai	Mizoram University	India
53	Pradipta		Roy	Dr. B. C. Roy Academy of Professional Courses	India
54	Aiko		Takahashi	The Institute of Statistical Mathematics	Japan
55	Huan		Tran	Georgia Institute of Technology	USA
56	Van Quang		Tran	VNU University of Engineering and Technology	Vietnam
57	Thanh	N.	Truong	University of Utah	USA
58	Ngoc Linh		Vu	Vietnam National University	Vietnam
59	Ngoc-Tuoc		Vu	Hanoi University of Science and Technology	Vietnam
60	Sanjubala		Sahoo	University of Connecticut	USA
61	David		Winkler	La Trobe institute for Molecular Science	Australia
62	Brandon		Wood	Lawrence Livermore National Laboratory	USA
63	Ryo		Yoshida	The Institute of Statistical Mathematics	Japan
64	Hongbin		Zhang	Technische Universität Darmstadt	Germany