ChemCatBio
Chemical Catalysis for Bioenergy

Artificial Intelligence for Catalysis

Rajeev Surendran Assary, *Argonne National Laboratory*

June 12, 2024

U.S. DEPARTMENT OF ENERGY | *Office of* ENERGY EFFICIENCY & RENEWABLE ENERGY
BIOENERGY TECHNOLOGIES OFFICE

Resources

- **Website: [chemcatbio.org]**
  - Tools and capabilities
  - Publications
  - Webinars
  - Interactive technology briefs
- **Tools:**
  - Catalyst Property Database
  - CatCost
- **Newsletter: The Accelerator**
  - Subscribe

Housekeeping

- Attendees will be in listen-only mode
- Audio connection options:
  - Computer audio
  - Dial in through your phone (best connection)
- Automated closed captions are available
- Use the Q&A panel to ask questions
- Technical difficulties? Contact Erik Ringle through the chat section, lower right of your screen
- Recording will be available at: https://www.chemcatbio.org/webinars.html

NOTICE: This webinar, including all audio and images of participants and presentation materials, may be recorded, saved, edited, distributed, used internally, posted on DOE's website, or otherwise made publicly available. If you continue to access this webinar and provide such audio or image content, you consent to such use by or on behalf of DOE and the Government for Government purposes and acknowledge that you will not inspect or approve, or be compensated for, such use.

Today's Speaker

Rajeev Surendran Assary
Argonne National Laboratory
assary@anl.gov

ChemCatBio
Chemical Catalysis for Bioenergy

CCPC
Consortium for Computational Physics and Chemistry
U.S. DEPARTMENT OF ENERGY

Argonne | INL | NETL | NREL | Oak Ridge National Laboratory | Pacific Northwest National Laboratory

A multi-scale problem ... A multi-lab solution

AI for Catalysis

Rajeev S. Assary, PhD

Materials Science Division, Argonne National Laboratory

ChemCatBio Webinar

assary@anl.gov | | @rsassary

Research Interests: Energy Storage Conversion

- Beyond-Li ion Energy Storage | Chemical Catalysis | Decarbonization

AI for Materials

- Molecules and materials for next generation e - storage

Autonomous Materials Discovery

- Accelerated identification and synthesis of optimal materials

Clean Energy and Sustainability

- Efficient carbon utilization, carbon-neutral energy systems, water

Source: Argonne National Laboratory

Imagine if ....

We could design, discover, demonstrate, and deploy *optimal catalysts* faster and faster with the help of AI

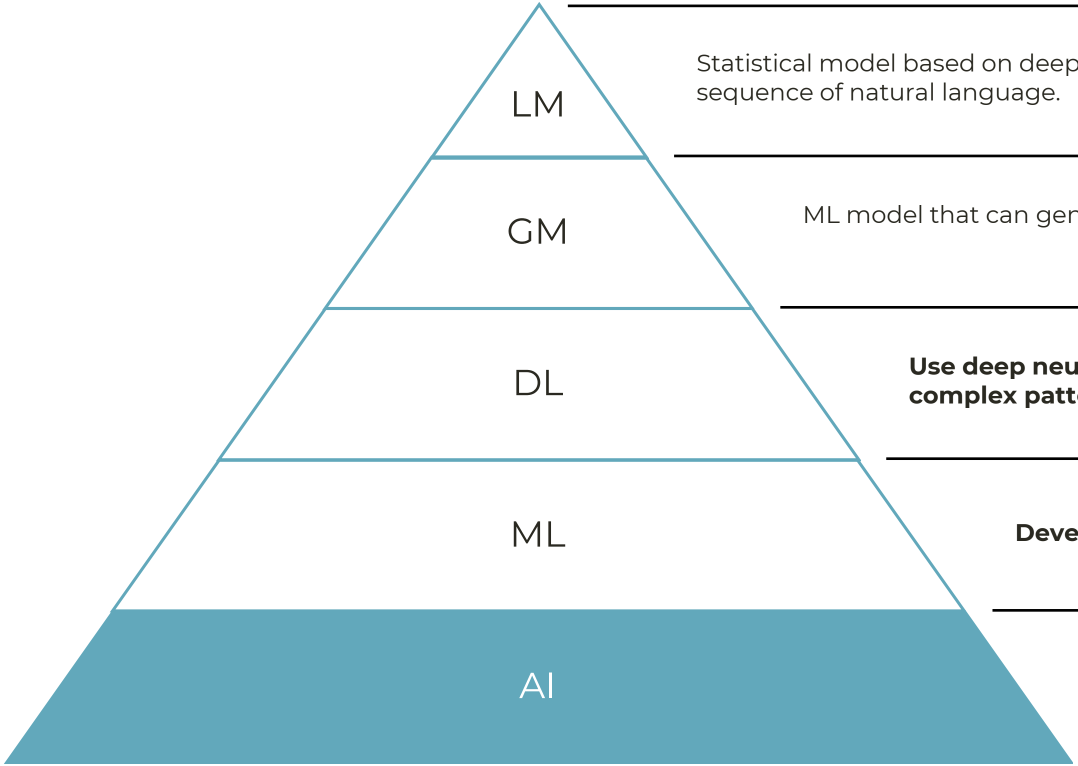
- Literature
- Experiments
- Analysis
- Modeling
- Data
- Any tools

Maximize

- Materials ($) || Operations ($) || Sustainability
- Low temp || Selectivity || Stability
- Green || Non-toxic || Recyclability
- Other objectives (*x*)

Optimal catalysts

AI and Tools



Statistical model based on deep learning of massive datasets used to predict words in a sequence of natural language.

ML model that can generate new data based on the patterns learned from input data

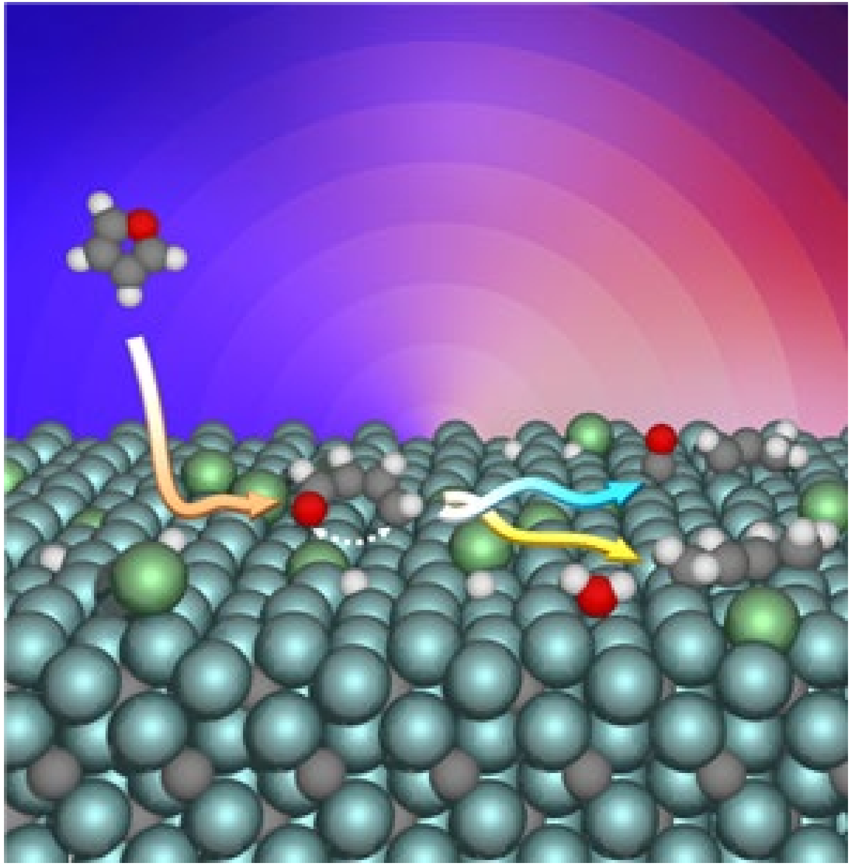
Use deep neural networks, as a mechanism for ML algorithms to learn complex patterns from data

Developing algorithms that can learn from data

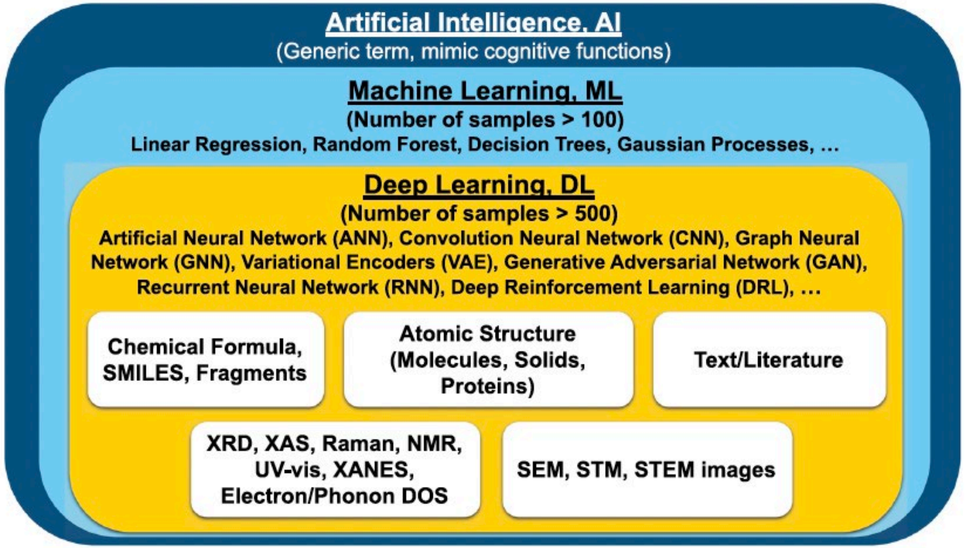
Creating intelligent agents that can reason, learn, and act autonomously (AIM)

How Can Computations & AI Help Catalysis ?

- **AI-Guided Design**
  - Informatics: Create and manage properties data
  - Optimization: Optimal materials and conditions
- **Creating New Materials**
  - Automated chemical synthesis
  - Automated materials characterization
- **Catalysis and Democracy**
  - Public-accessible information
  - Catalyst knowledge accessible for everyone
  - Foundational models (GPTs)



Overview of AI in Materials Research

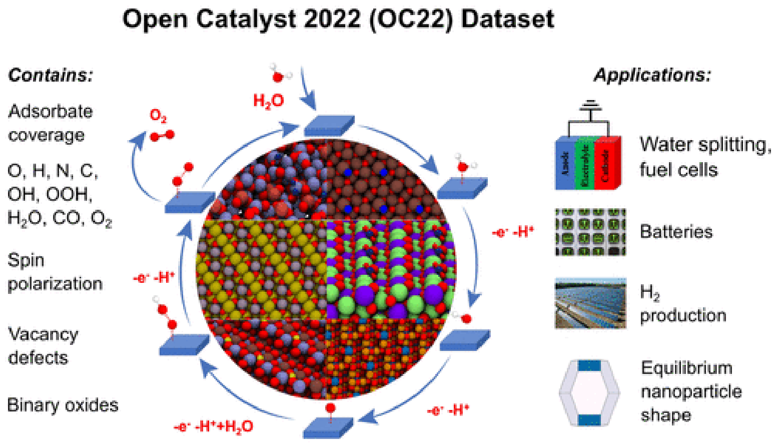


Given *x* and *y*, find a solution *f*(*x*) that predicts *y*

$y = \sigma(W X + b)$ (Fits any function)

σ : non-linear activation

W: trainable weights



- 1.3 million DFT calculations
- Structure and BE of intermediates
- HER, OER, Red (CO2, N2)
- No kinetics

Materials Data Ecosystem

- Catalyst data ecosystem services and tools are being built and populated (MDF, Datahub)

DataHub
ChemCatBio

The Materials Project

AFLOW
Automatic - FLOW for Materials Discovery

OpenAIRE

funcX

MaterialsCommons

4CeeD

matminer

NOMAD

MaterialsIO

Citrination

Materials Resource Registry
Part of the Materials Genome Initiative

MATERIALS DATA FACILITY

MATERIALSCLOUD

OQMD
The Open Quantum Materials Database

Connecting the Research Universe

NanoMine

DLHub
Data and Learning Hub for Science
https://www.dlhub.org

EXABYTE.IO

MaRDA

MLExchange

pymatgen

MAST
MAterials Simulation Toolkit

MATERIALS INNOVATION NETWORK

Many others…

AI for Catalysis: Data-Driven "Cat" Exploration

Nudged Elastic Band Calculations Accelerated with Gaussian Process Regression Based on Inverse Interatomic Distances

Olli-Pekka Koistinen, Vilhjálmur Ásgeirsson, Aki Vehtari, and Hannes Jónsson*

Cite this: J. Chem. Theory Comput. 2019, 15, 12, 6738–6751
Publication Date: October 22, 2019
https://doi.org/10.1021/acs.jctc.9b00692

npj | computational materials
www.nature.com/npjcompumats

REVIEW ARTICLE OPEN

Machine-learning atomic simulation for heterogeneous catalysis

Dongxiao Chen[1], Cheng Shang[1,2] and Zhi-Pan Liu[1,2,3]

Heterogeneous catalysis is at the heart of chemistry. New theoretical methods based on machine learning (ML) techniques that emerged in recent years provide a new avenue to disclose the structures and reaction in complex catalytic systems. Here we review briefly the history of atomic simulations in catalysis and then focus on the recent trend shifting toward ML potential calculations. The advanced methods developed by our group are outlined to illustrate how complex structures and reaction networks can be resolved using the ML potential in combination with efficient global optimization methods. The future of atomic simulation in catalysis is outlooked.

npj Computational Materials (2023)9:2 ; https://doi.org/10.1038/s41524-022-00959-5

Review Article | Published: 23 February 2023

Bridging the complexity gap in computational heterogeneous catalysis with machine learning

Tianyou Mou, Hemanth Somarajan Pillai, Siwen Wang, Mingyu Wan, Xue Han, Neil M. Schweitzer, Fanglin Che & Hongliang Xin

Nature Catalysis **6**, 122–136 (2023) | Cite this article

Perspective

Machine Learning for Heterogeneous Catalyst Design and Discovery

Bryan R. Goldsmith, Jacques Esterhuizen, and Jin-Xun Liu
Dept. of Chemical Engineering, University of Michigan, Ann Arbor, MI 48109-2136

Christopher J. Bartel
Dept. of Chemical and Biological Engineering, University of Colorado Boulder, Boulder, CO 80309

Christopher Sutton
Fritz-Haber-Institut der Max-Planck-Gesellschaft, Theory Dept., Faradayweg 4-6, Berlin D-14195, Germany

DOI 10.1002/aic.16198
Published online May 25, 2018 in Wiley Online Library (wileyonlinelibrary.com)

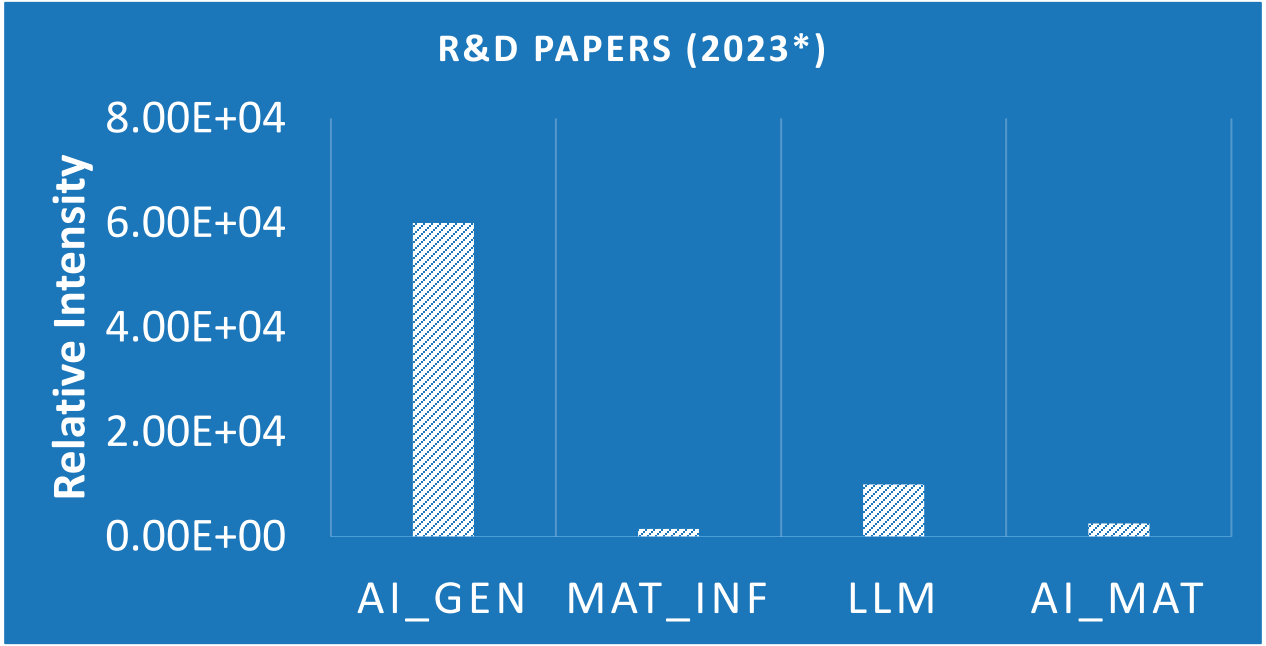
Perspective | Published: 26 January 2023

Exploring catalytic reaction networks with machine learning

Johannes T. Margraf, Hyunwook Jung, Christoph Scheurer & Karsten Reuter

Nature Catalysis **6**, 112–121 (2023) | Cite this article

Emerging AI for Materials Science and Chemistry



Tools Available for Catalyst Design

1

- **Physics**-based simulations
- DFT/MD calculations

- **Expensive**

2

- **Statistical Models**
- Analytical models
- Monte Carlo simulations

3

- **ML/DL**

- **Data-Driven Catalysis Science**
- Inverse design of materials
- Deep geometric learning
- Imaging and characterization

4

- **Bayesian Approaches**
- On the fly optimization
- Design of the experiments

- **Self-supervised**
- **Autonomous**

5

- **Natural language**
- **Data from everywhere!**

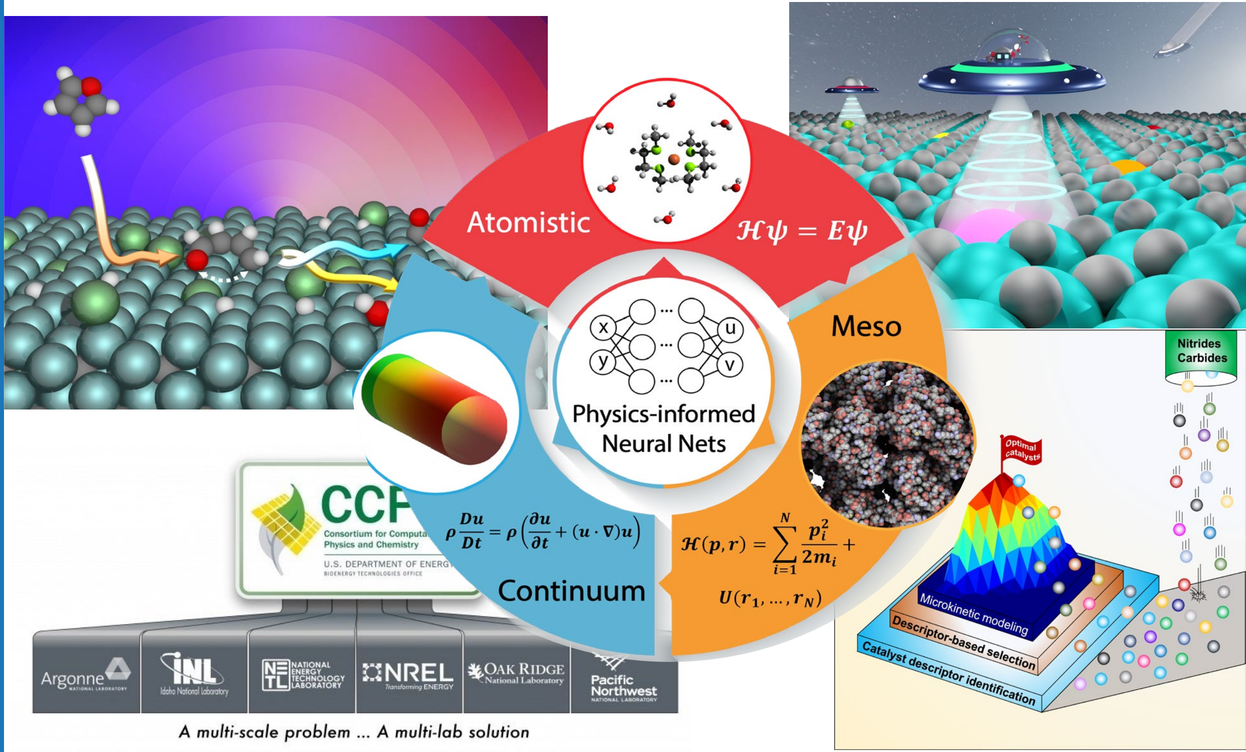
- **LLM for Everything**
- Generative modeling
- GPT4 or better…

6

- **Beyond GPT 4**
- Multi-modality
- Autonomous agents

- **Need general and reliable data**

Multiscale Modeling With the Help of AI



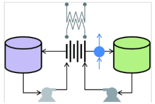
Achieving bigger time/length scales

Design of Molecules and Catalysts

Discover new Molecules

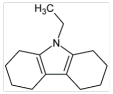
Energy Storage

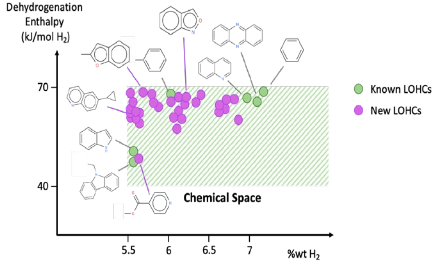
Grid storage based on organic materials

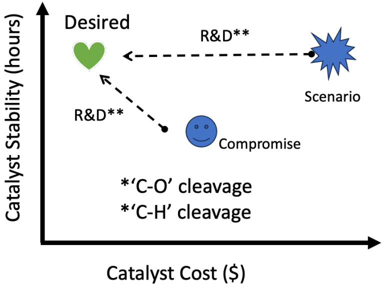


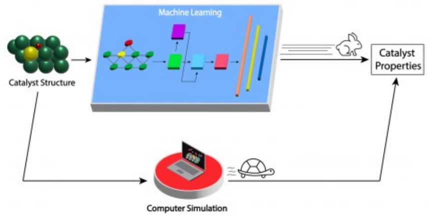
Hydrogen Storage

Liquid Organic Hydrogen Carriers









Liquid Organic Hydrogen Carriers

Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

Check for updates

Cite this: DOI: 10.1039/d3dd00123g

Uncovering novel liquid organic hydrogen carriers: a systematic exploration of chemical compound space using cheminformatics and quantum chemical methods†

Hassan Harb, [a] Sarah N. Elliott, [b] Logan Ward, [c] Ian T. Foster, [c] Stephen J. Klippenstein, [b] Larry A. Curtiss [a] and Rajeev Surendran Assary [*a]

Catalyst: Property Prediction

Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Digital Discovery*, 2023, 2, 59

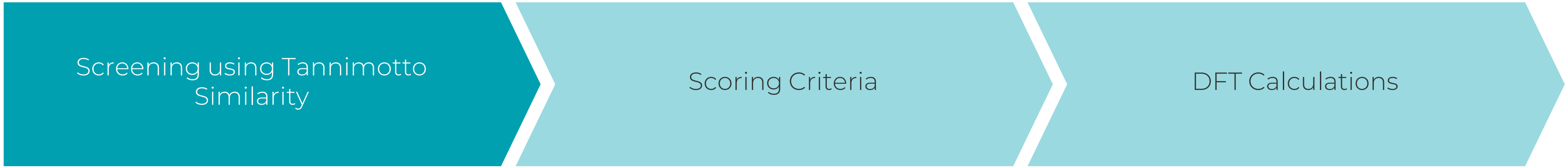
Accelerating the evaluation of crucial descriptors for catalyst screening *via* message passing neural network†

Hieu A. Doan, [*a] Chenyang Li, [a] Logan Ward, [b] Mingxia Zhou, [ac] Larry A. Curtiss [a] and Rajeev S. Assary [*a]

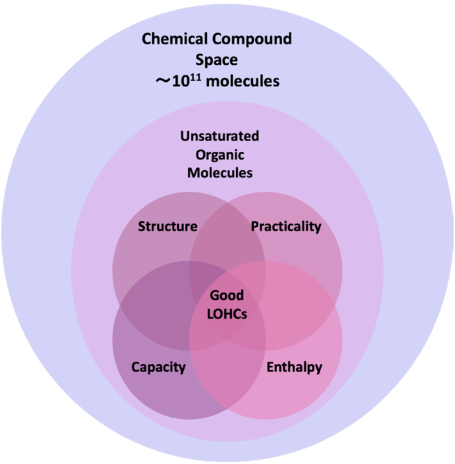
New LOHC From 166 Billion Molecules

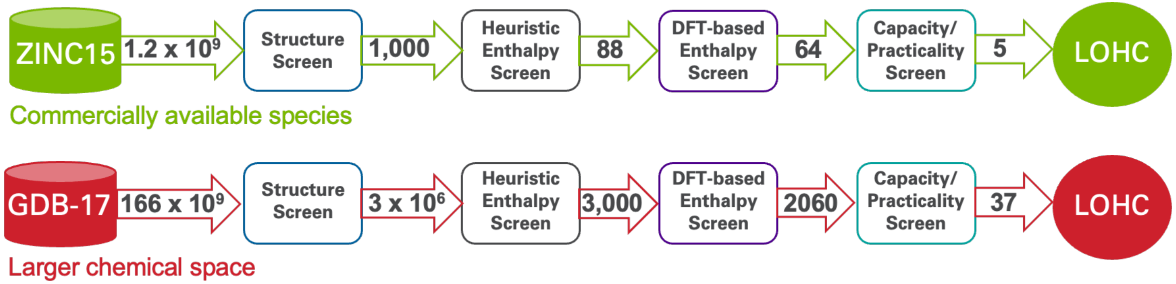
166 billion to 1 milliion

3000 de/hydrogenation calculations



1 million to 3000

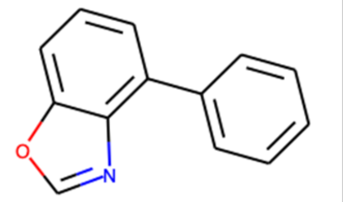




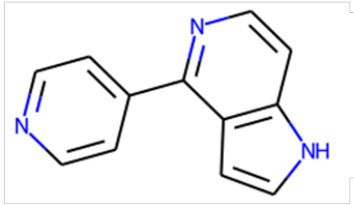
https://github.com/HydrogenStorage/screening-large-databases

https://github.com/HydrogenStorage/LOHC

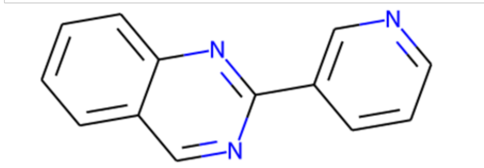
Molecular Screening to Experiments



ΔH / 10 = 6.50
E = 6.45



ΔH / 10 = 5.81
E = 5.83

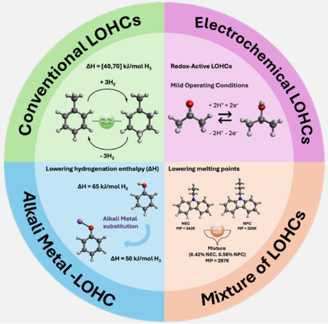


ΔH / 10 = 6.10
E = 6.10

- Catalyst design and decarbonization center
- In-house developed catalysts
- Performance evaluation
- Catalyst degradation studies

https://github.com/HydrogenStorage/screening-large-databases

https://github.com/HydrogenStorage/LOHC



Molecular Screening From Large Chemical Space

Argonne NATIONAL LABORATORY

RESEARCH WORK WITH US COMMUNITY ABOUT US

PRESS RELEASE | ARGONNE NATIONAL LABORATORY

AI helps whittle down candidates for hydrogen carriers in liquid form from billions to about 40

Energy could be stored in hydrogen carrier molecules to be used as fuel

BY JOSEPH E. HARMON | JANUARY 10, 2024

In an AI-based exploration of 160 billion organic molecules, Argonne scientists identified about 40 liquid hydrogen carriers that could one day fuel cars, trucks, buses, trains and ships and generate energy for consumers.

In a computational study leveraging artificial intelligence (AI), scientists at the U.S. Department of Energy's (DOE) Argonne National Laboratory assessed 160 billion molecules, a number exceeding the people born in the entire span of human history. Their goal was to screen the molecules for suitability as liquid carriers of hydrogen.

Liquid hydrogen carriers could one day fuel cars, trucks, buses, trains,

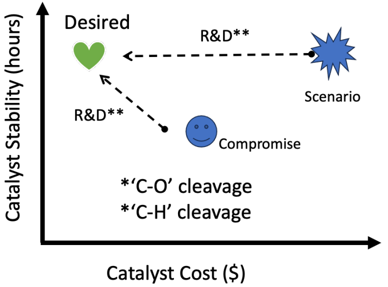
News
Media Contacts
Experts Guide
Press Releases
Feature Stories
In the News
Media Tips
Subscribe to Argonne

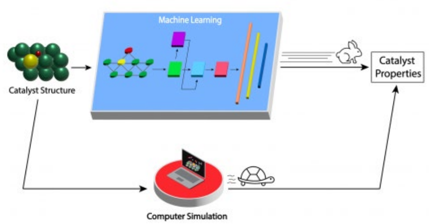
SHARE

Liquid Organic Hydrogen Carriers

Design of Molecules and Catalysts

- Identify active and inexpensive deoxygenation catalyst
- Utilize data-driven models
- Develop AI capability at CCPC





Catalyst: Property prediction

Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

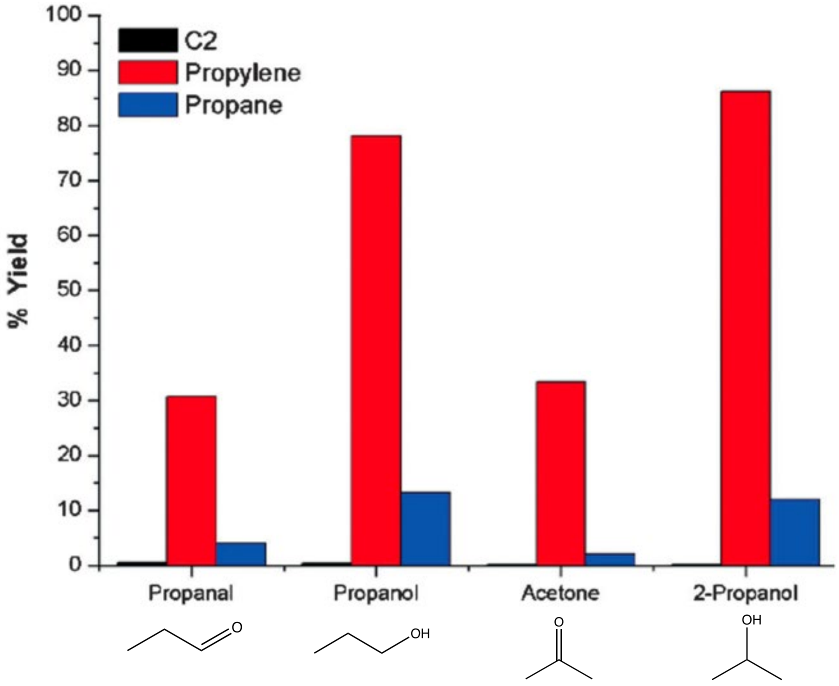
Check for updates

Cite this: *Digital Discovery*, 2023, **2**, 59

Accelerating the evaluation of crucial descriptors for catalyst screening *via* message passing neural network†

Hieu A. Doan, *[a] Chenyang Li, [a] Logan Ward, [b] Mingxia Zhou,[ac] Larry A. Curtiss[a] and Rajeev S. Assary *[a]

Example Problem: Catalytic Fast Pyrolysis With Mo_2C



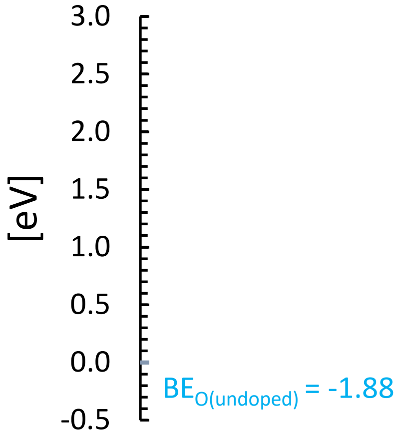
- ❑ High activity and selectivity toward C-O/C=O bond cleavage

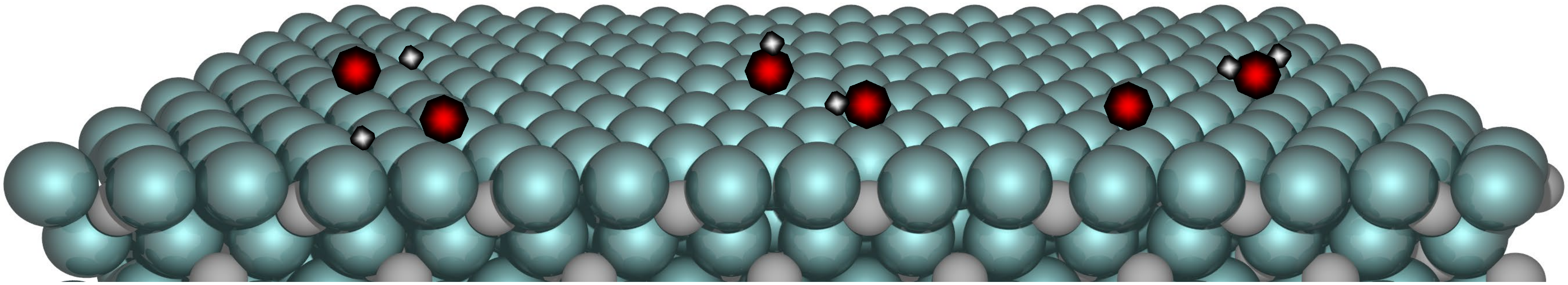
- ❑ Susceptible to deactivation by surface oxidation and coke formation

→ Excess H_2 co-feed required

Deoxygenation: Ren, H, Vlachos, D., Chen, J. G., et al., *ChemSusChem* 2013, 6, 798-801
CO_2 reduction: Khoshooel, M. A, Snurr, R, Furha, O., et al, *Science*, 2024, 384, 540-546

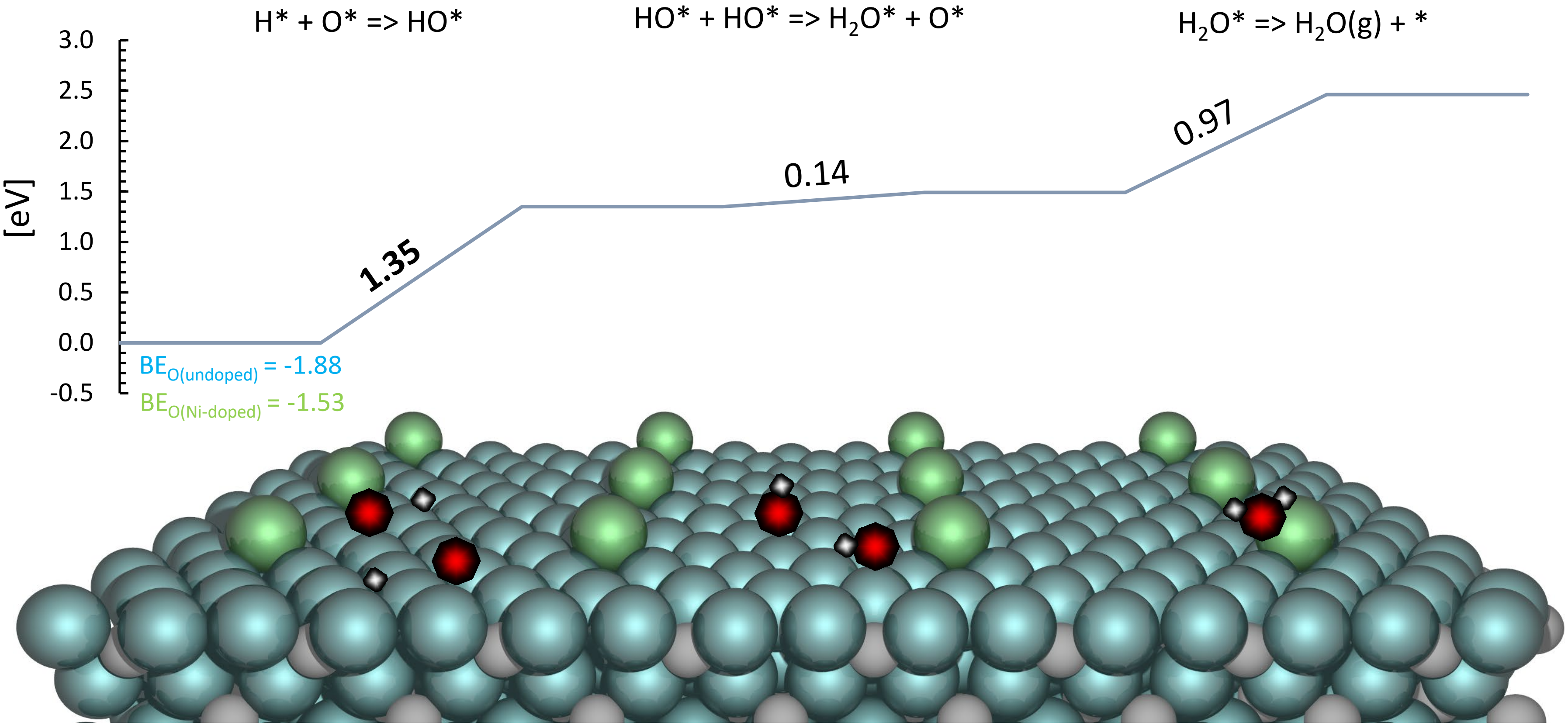
O* Removal via H_2O Formation





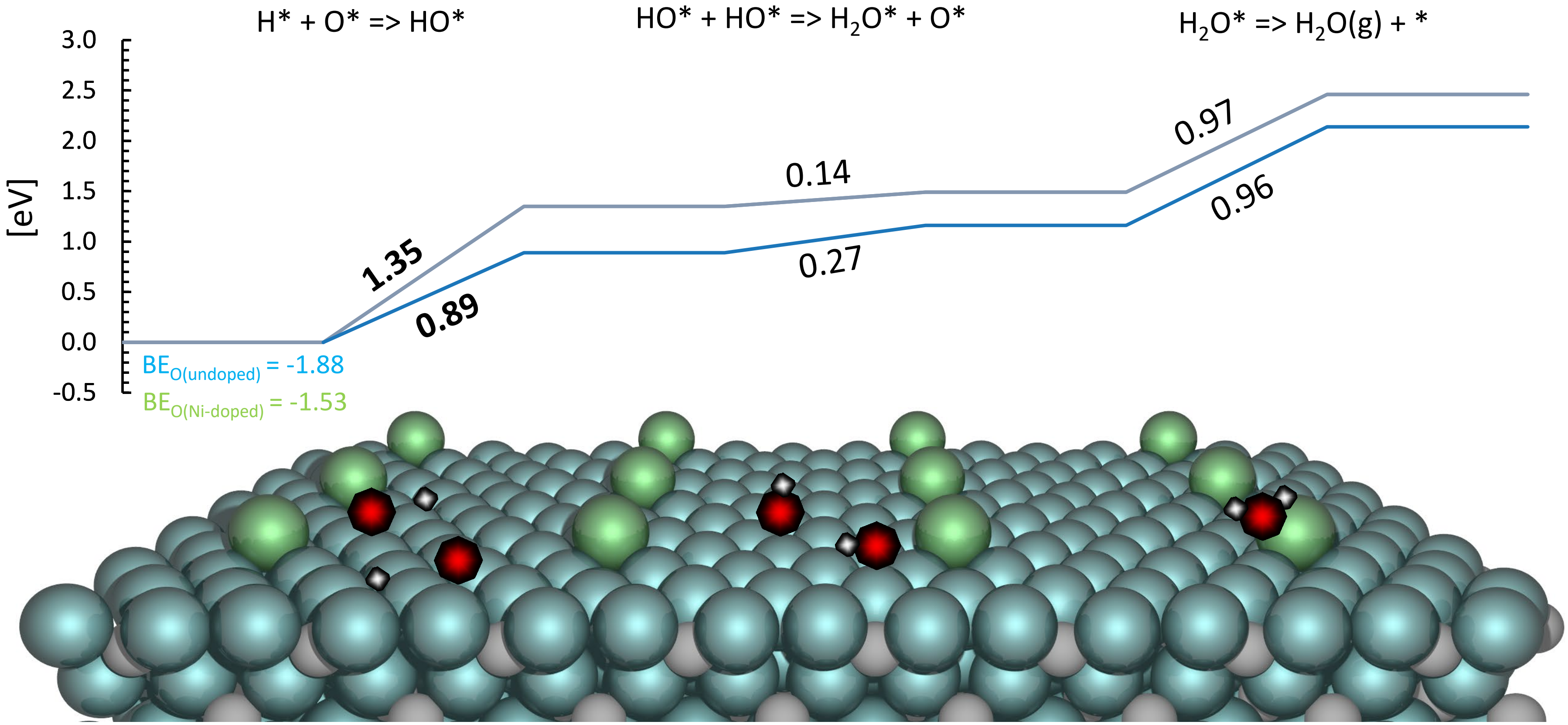
Zhou, Assary, et al., *J. Phys. Chem. C* 2018, 122, 1595-1603

O* Removal via H_2O Formation



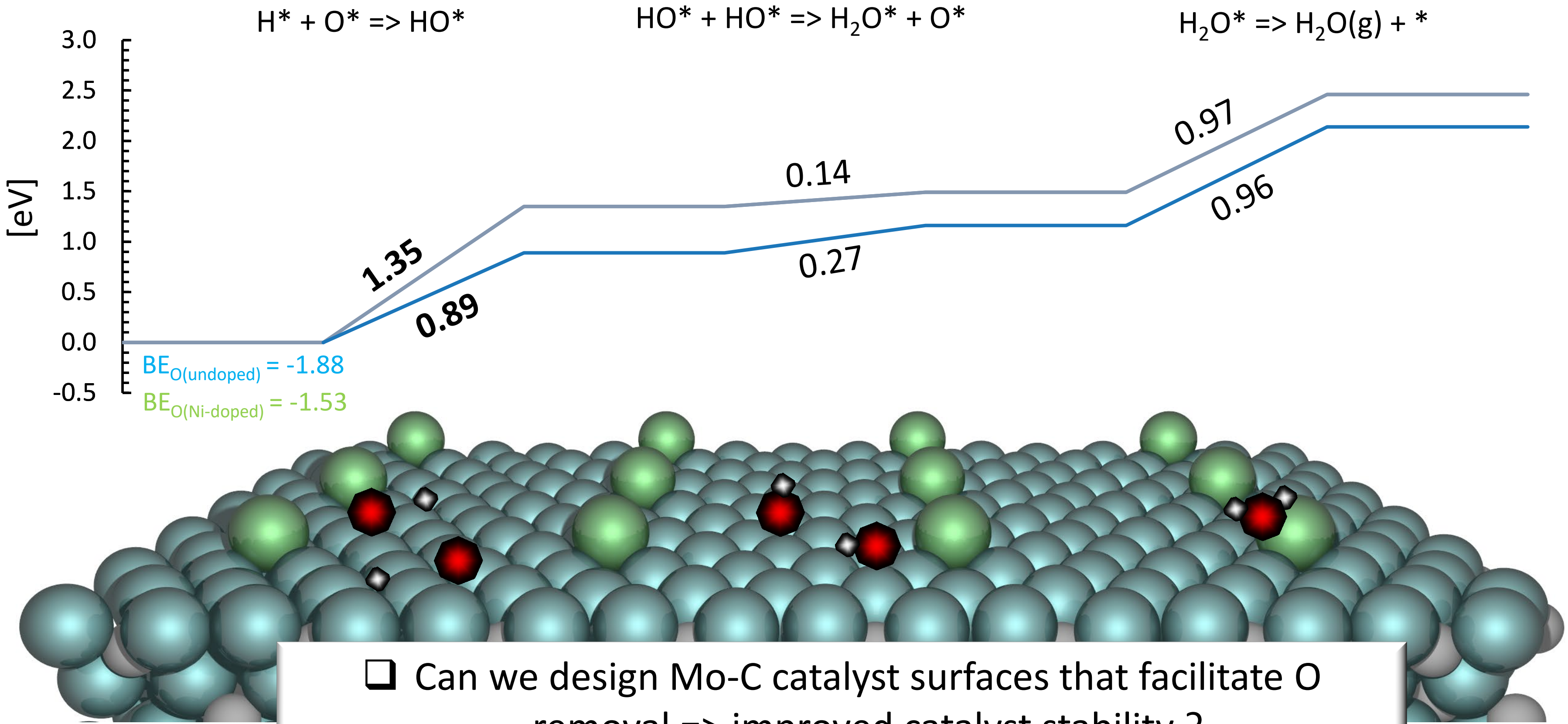
Zhou, Assary, et al., *J. Phys. Chem. C* 2018, 122, 1595-1603

O* Removal via H_2O Formation



Zhou, Assary, et al., *J. Phys. Chem. C* 2018, 122, 1595-1603

O* Removal via H_2O Formation



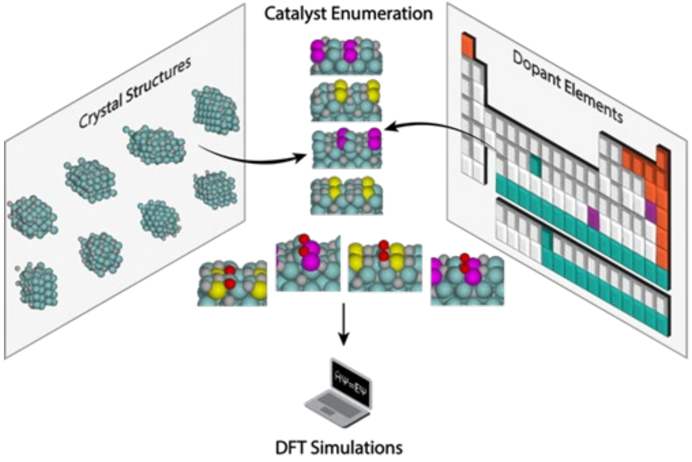
- ❑ Can we design Mo-C catalyst surfaces that facilitate O removal => improved catalyst stability ?

Cost-Efficient Estimation of Crucial Descriptors

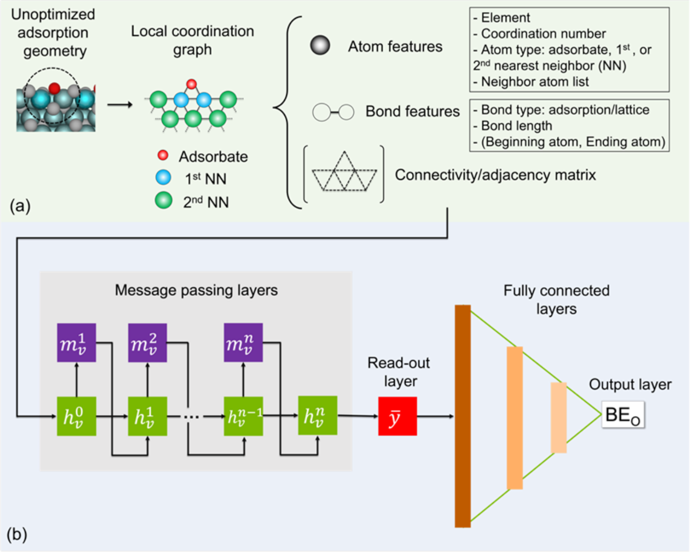
- **Develop data of <u>oxygen binding energy</u>**
  - Structure (graph) fits against binding energy
  - Develop a deep learning model
- **In future, provide a graph (approximate str)**
  - Model predicts oxygen binding energy (*ms* vs 4 hours)
  - Uses: Wulf construction, catalyst–regional reactivity

Data Generation for Catalyst Screening

- *In silico* Data: Development of 20K VASP Deoxygenation
- Develop a Graph Neural Network for Property Prediction



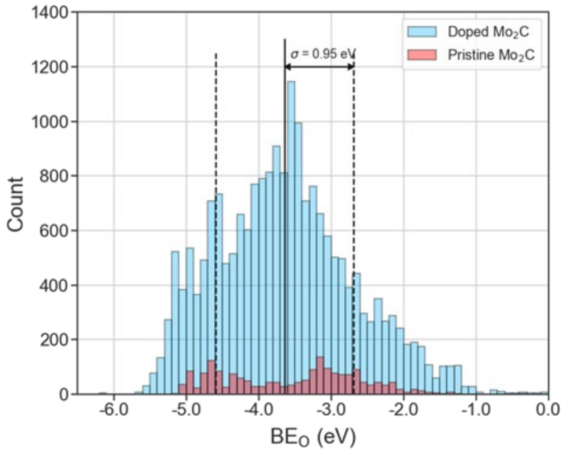
High-throughput structure enumeration and data generation (20K) for oxygen adsorption on pristine and doped Mo_2C catalyst surfaces



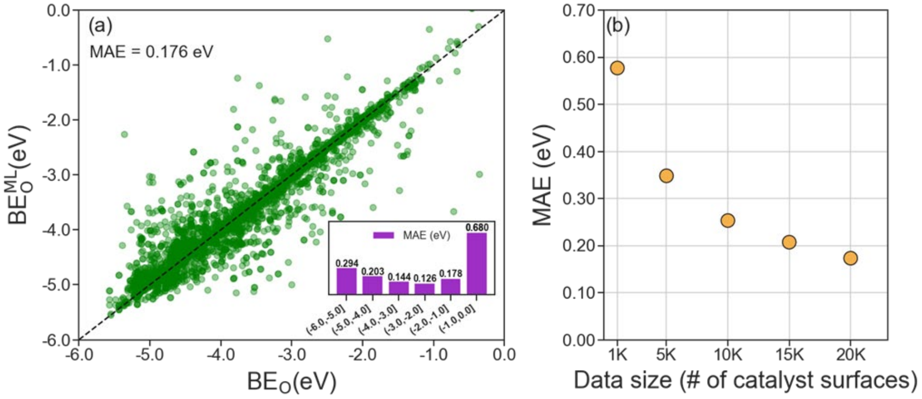
Message passing neural network using local coordination graph representation (LCG-MPNN) for predicting oxygen binding energies (BEO) on Mo_2C catalyst surfaces.

Comparison of Computed vs Learnt Data

- 20 K adsorption energy calculations → BE –model results comparable with binary and tertiary alloys (MAE 0.1 to 1.0 eV, a year ago!)



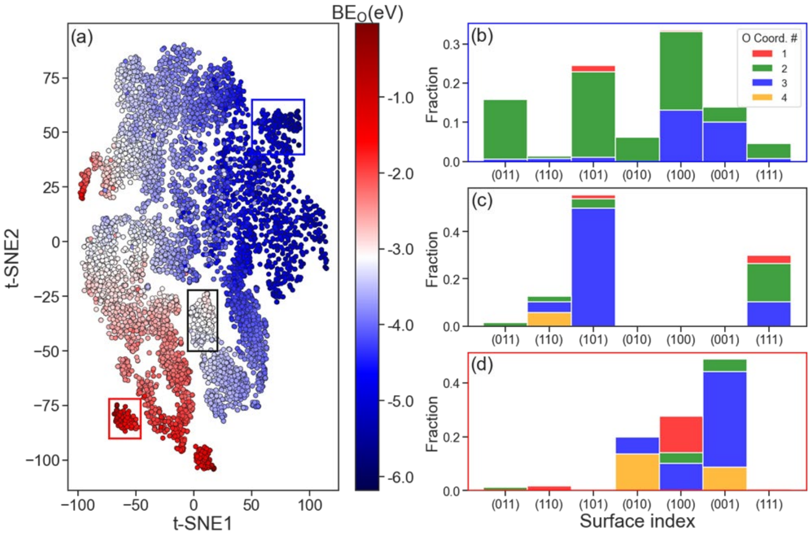
Oxygen binding energy distribution on pristine and doped Mo_2C facets



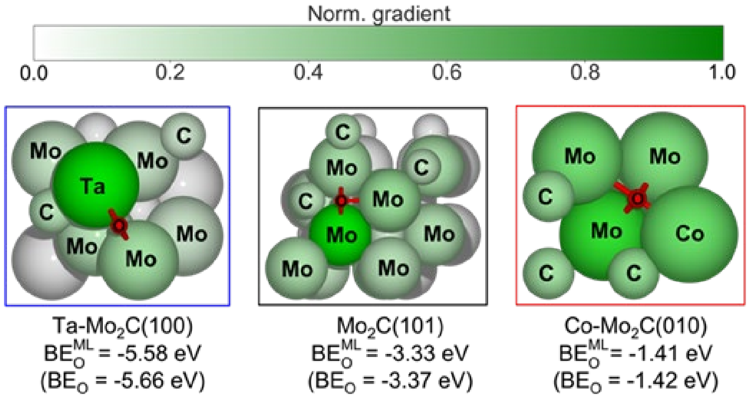
Parity plot of oxygen binding energies predicted by LCG-MPNN (BE_ML) and computed by DFT (BE_DFT) on the test set

https://github.com/MolecularMaterials/nfp

Analysis of Descriptor Trends and Feature Importance



2D t-distributed stochastic neighbor embedding (t-SNE) plot of graph-level features from the readout layer.



Graphical illustration of atom contribution to the prediction of oxygen binding energy.

Toward Accelerated Catalyst Sesign

Deoxygenation Catalysts

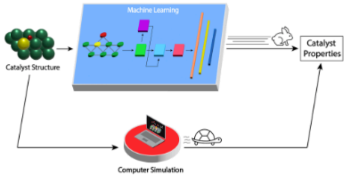
PRESS RELEASE | ARGONNE NATIONAL LABORATORY

Machine learning model speeds up assessing catalysts for decarbonization technology from months to milliseconds

BY JOSEPH E. HARMON | FEBRUARY 28, 2023

Scientists create computational model for identifying low-cost catalysts that convert biomass into fuels and useful chemicals with low carbon footprint.

Next time you drive past farms or prairies and ponds on a rural road, look around. They are a rich source of biomass. That includes corn, soybeans, sugar cane, switchgrass, algae and other plant matter. These carbon-rich materials can be converted to liquid fuels and chemicals with many possible applications. There is enough biomass in the United States, for example, to produce renewable jet fuel for all air travel.



The newly developed machine learning model greatly speeds up assessing the properties of molybdenum carbide catalysts for biomass conversion to useful products (top path) compared with current computer simulation methods (bottom path). (Image by Argonne National Laboratory.)

A major stumbling block at present is lack of effective, low-cost catalysts needed to transform biomass into biofuel or other useful products. Researchers at the U.S. Department of Energy's (DOE) Argonne National Laboratory report developing an artificial intelligence-based model to speed up the process for engineering a low-cost catalyst based on molybdenum carbide.

> *"…with our deep learning model we can now do accurate and inexpensive calculations for tens of thousands of structures in milliseconds. It is materials screening on steroids."*
> *— Hieu Doan, Argonne assistant scientist*

"Biomass is an organic material, meaning it is full of carbon," said Rajeev Assary, group leader in Argonne's Materials Science Division (MSD). "The ultimate goal is to cheaply transform that

News

- Media Contacts
- Experts Guide
- Press Releases
- Feature Stories
- In the News
- Social Media
- Media Tips
- Subscribe to Argonne

SHARE

CONTACT US

For More Information or Media Related Inquiries

+1-630-252-5580 | media@anl.gov

CONTACT US

To discuss a potential research partnership with Argonne or to inquire about licensing Argonne technology

+1-800-627-2596

Integrating LLMs in Catalyst Discovery Sciences

- **"LLM can do jaw-dropping things. But nobody knows exactly why"**
  - MIT Technology Review, March 2024

- **Augment Existing LLMs**
  - Chain of thought prompting and data extraction
- **Combine and Enhance Models**
  - Integrate the data with relevant LLM models
  - Introduce additional regression layers
- **Fine Tuning to Specific Problems**
  - Develop targeted data for specific needs
  - Develop RAG systems for catalysts

- **ReactionT5**
- **T5Chem**
- **ChemBench**
- **ChemCrow**
- **ChemReasoner**
- **ChemBERTa-2**
- **Chem-LLM**
- **RetroBioCat**
- **ChemGPT**
- **MG-BERT**
- **DrugGPT**
- **SciBERT**
- **MatBERT**
- **MatSciBERT**
- **MaterialsBERT**
- **MatSCQA**
- **CataLM**

Large language models and automated agents for chemistry and materials

“Soon…Hey Siri, can you recommend me….?

- **“Hey Siri, can you recommend me the *best naturally abundant, green, heterogeneous catalyst* for deoxygenation reaction of 2-butanol? I am willing to spend $1 per Kg of the catalyst.”**
  - Certainly!
    - (Start with positive assurance – management 101)
  - Catalysts for 2-butanol – if not available
    - finding alternatives for you!
  - Dollar information is not available.
    - Processed data:
      - Here are the tools for catalyst cost prediction (use Cat Cost)…
  - Sorry, best catalysts are all precious metal catalyst (Truth!)
    - However, there Fe-Pt-U alloy can do something (Hallucinations or Truth!)

Summary

- **AI combined with atomistic modeling can provide valuable insights**
  - ➢ Molecular Discovery from Large Chemical Space
  - ➢ Data-driven Materials Property Prediction
- **Near-term: AI will advance catalyst design via:**
  - Adsorption energy and microkinetic modeling
  - Experimental validation is crucial
- **Red hot: There is vast opportunity for us (catalysis community), to utilize growing strengths of AI (LLM, DL) and HPC/HTE**
  - AI tools of synthesis of catalysts
  - AI for catalyst characterization

Great Catalysts = Great Quality of Life

A multi-scale problem ... A multi-lab solution

This work was performed in collaboration with the Chemical Catalysis for Bioenergy Consortium (ChemCatBio, CCB), a member of the Energy Materials Network (EMN)

Thank you

chemcatbio.org

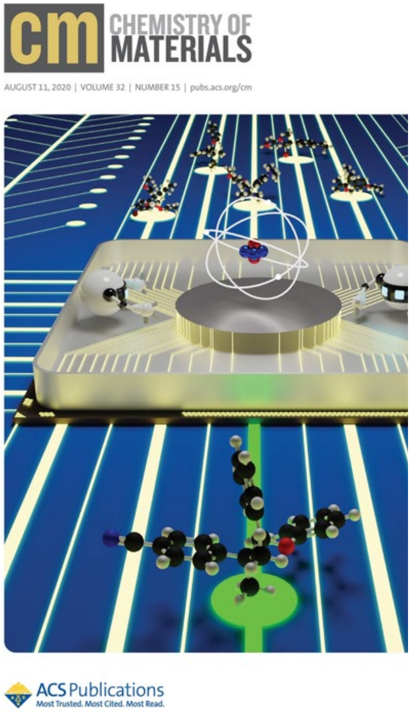
Q&A

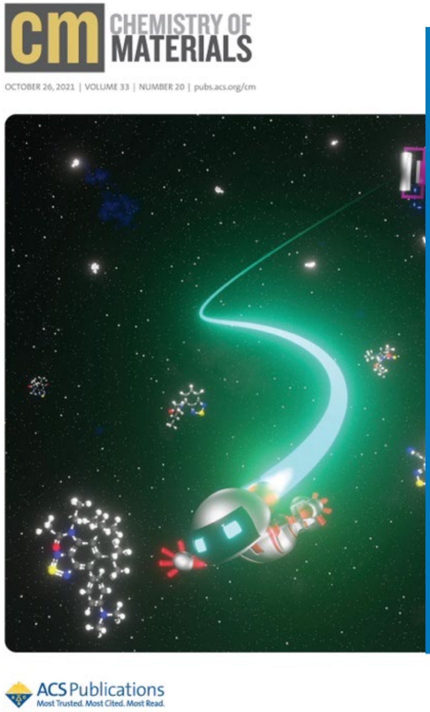
Rajeev Surendran Assary
Argonne National Laboratory
assary@anl.gov

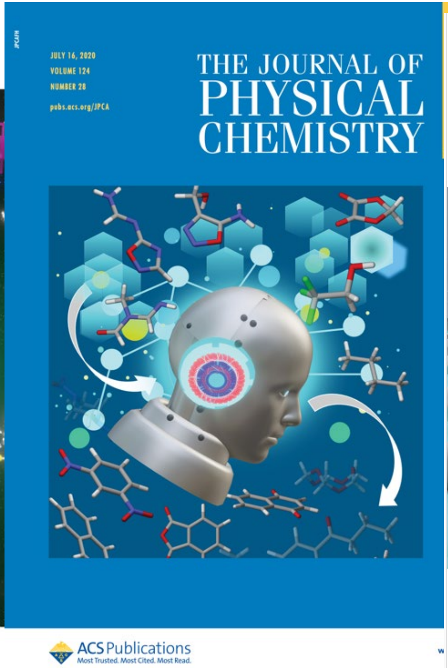
AI help in predicting optimal properties

New Molecules

New Catalysts







Acknowledgements

- Larry Curtiss
- Jim Parks
- Hieu Doan
- Mingxia Zhou
- Hassan Harb
- Logan Ward
- Josh Schaidle
- Sarah Elliott
- Stefan Klippenstein
- Ian Foster
- Jeremy Leong
- ALCC Computing (THETA)
- LCRC (ANL)

The CCPC constructs and utilizes computational models to discover new materials for bioenergy applications and assist in the cost-effective scale-up of bioenergy technologies to commercially relevant scales.