Data-centric Design of Combinatorial Materials Systems

Hengrui Zhang

PhD Candidate and Ryan Fellow

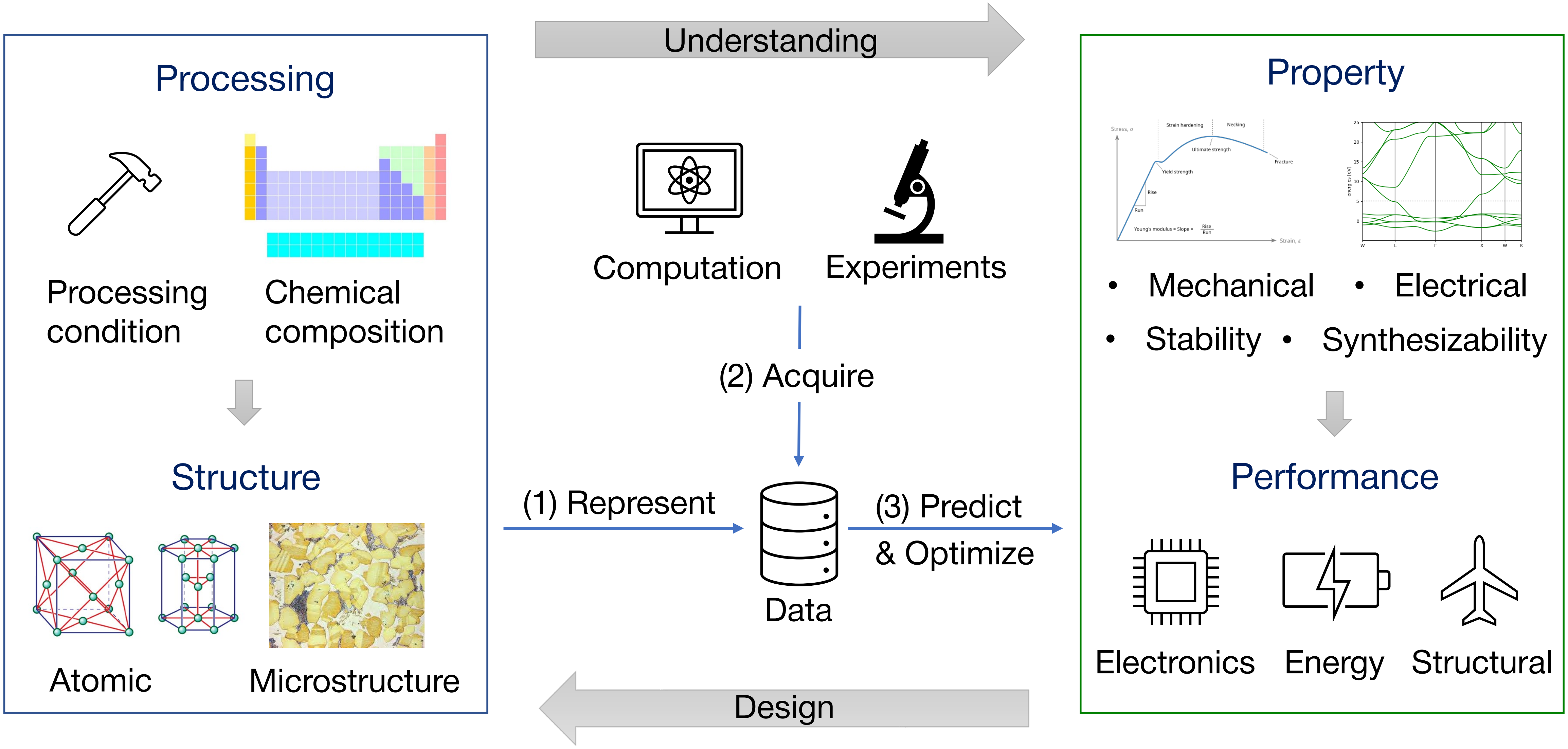
Northwestern University

Northwestern University

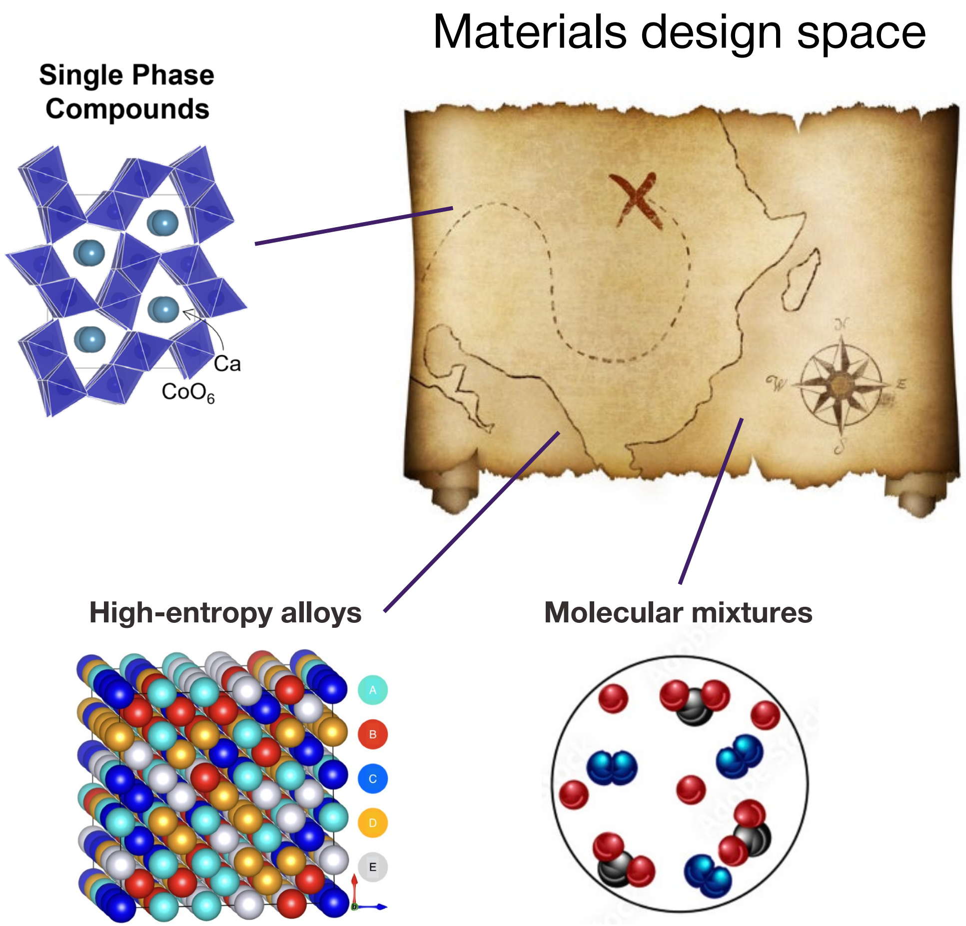
hrzhang@u.northwestern.edu

https://hrzhang.me

Data-centric materials design



Structural complexity and representation



Combinatorial complexity at multiple scales

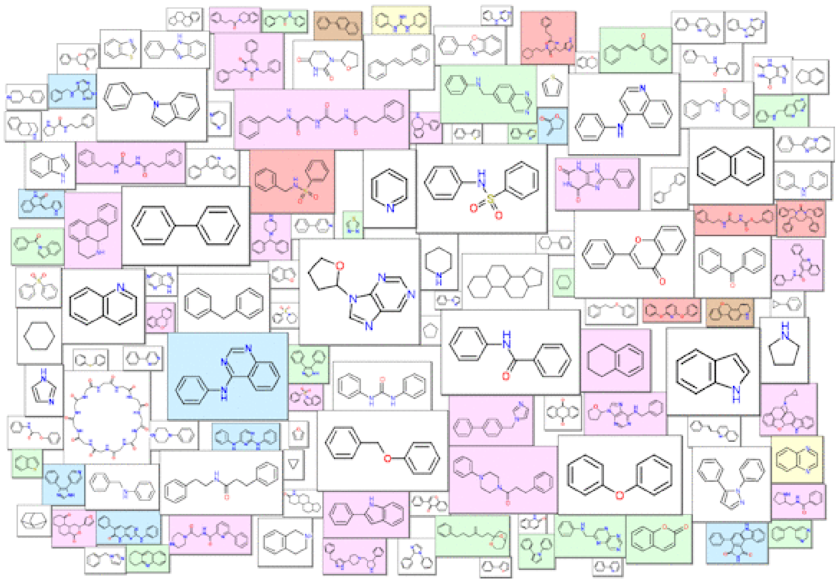
- Diverse constituents
- Arrangement of constituents
- Disordered

AI/ML models need readable data

How do we **represent** the materials to capture critical information?

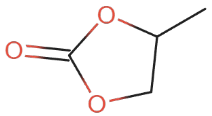
R. Singh et al. *Nat. Comput. Sci.* (2021)

Molecular mixture electrolytes

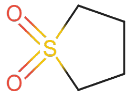


~160 molecules, ~40 salts

- $C(200,4) > 10^7$ compositions
- Infinite if considering fractions

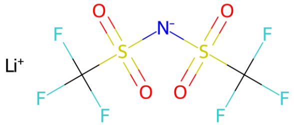


Solvent 1
w_1

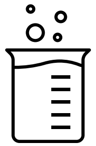


Solvent 2
w_2

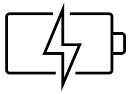
…



Li^+ salt
w_s



Molecular mixture

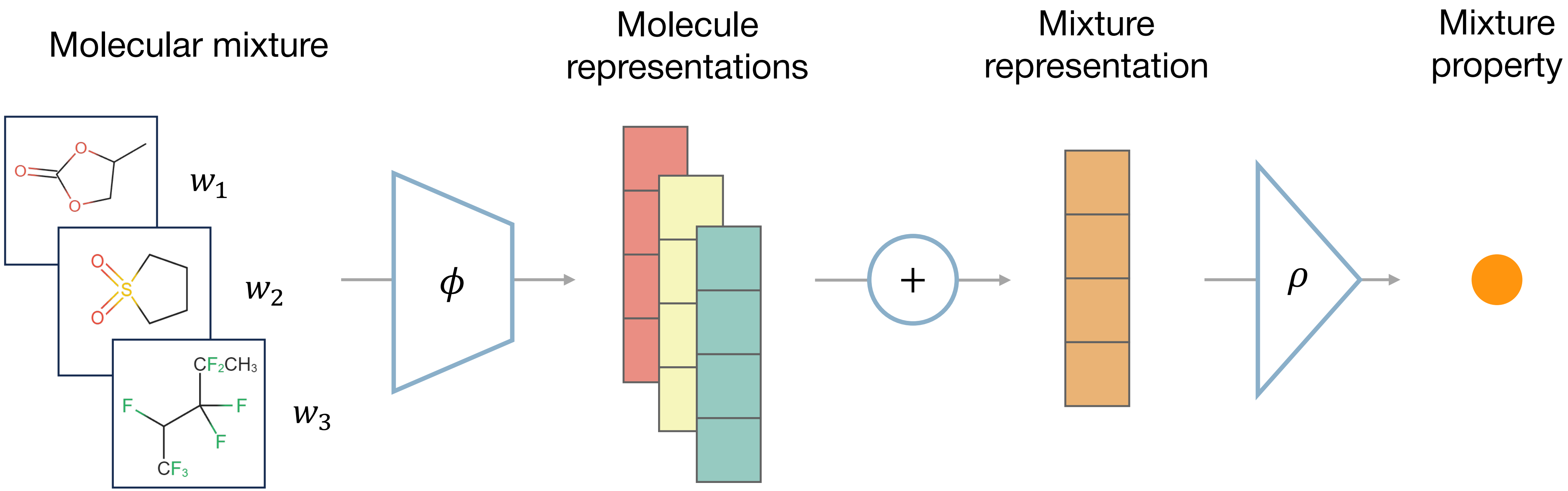


Battery electrolyte

- Molecules have diverse chemistry and geometries
- Mixture = constituents + molecular interactions

Ertl & Rohde, *J. Cheminformatics* 4, 12 (2012)

Modeling mixture as graph set



Graph neural network (GNN)

- Capture chemistry and geometry

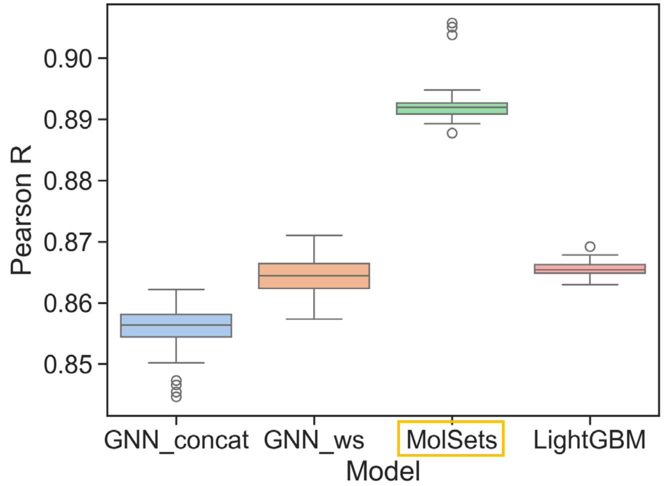
Attention mechanism (as in GPT)

- Learns relative importances
- Permutation invariant

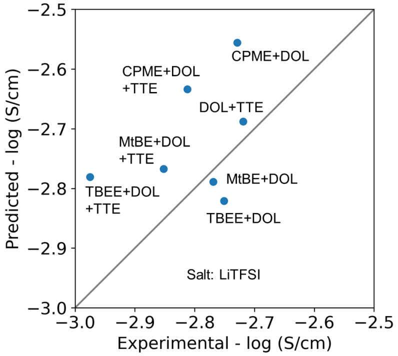
H. Zhang et al. *PRX Energy* (2024)

Benefits and applications

Superior accuracy in benchmark tests



Exp. verified (Tianxing Lai @ UT Austin)



Virtually screened >10,000 new candidate mixtures

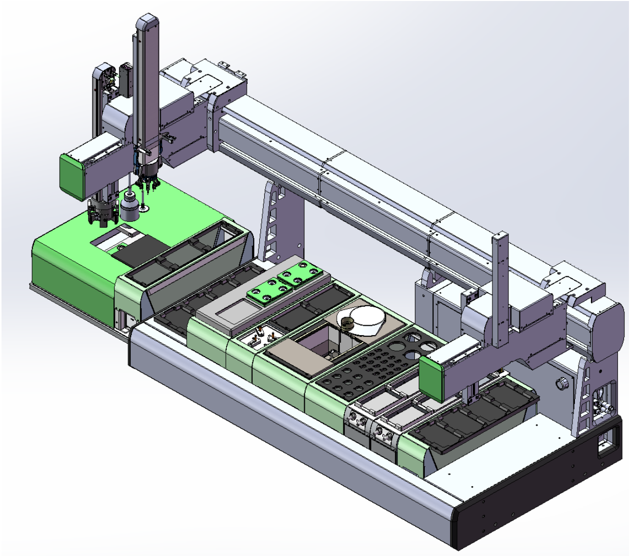
Predicted room temperature electrical conductivity of molecular mixtures

Zhang, Hengrui [1]; Lai, Tianxing [2]; Chen, Jie [1]; Manthiram, Arumugam [2]; Rondinelli, James [1]; Chen, Wei [1]

Author affiliations

Published Mar 06, 2024 on Dryad. https://doi.org/10.5061/dryad.pnvx0k6w1

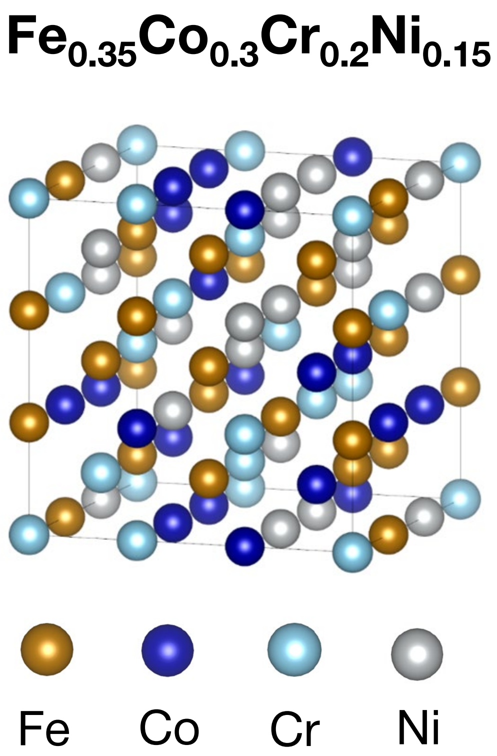
Subject keywords



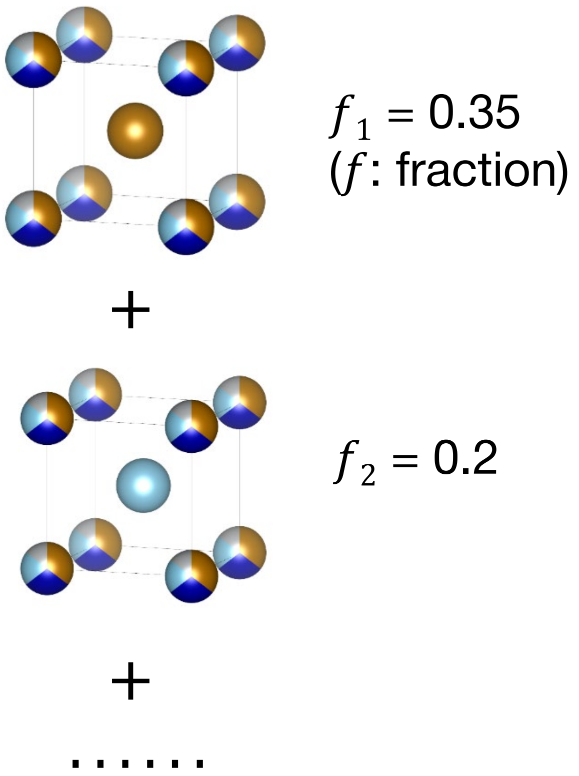
High-throughput, autonomous experiments (Jeffrey Lopez @ NU)

Extend to high-entropy alloys

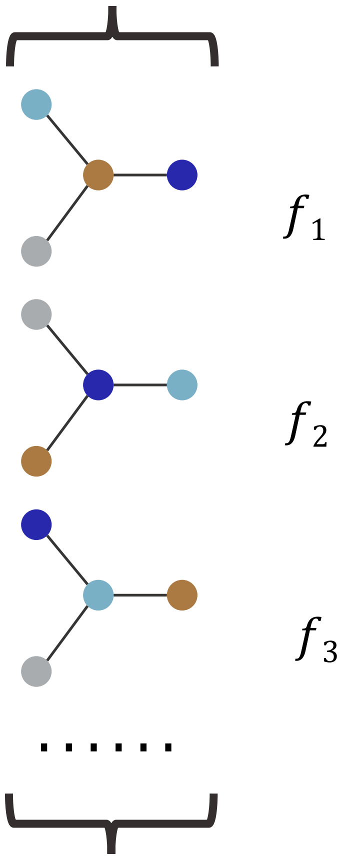
(Mentoring Ms. Ruishu Huang @ UW Madison)



HEA structure



Collection of local environments (LEs)

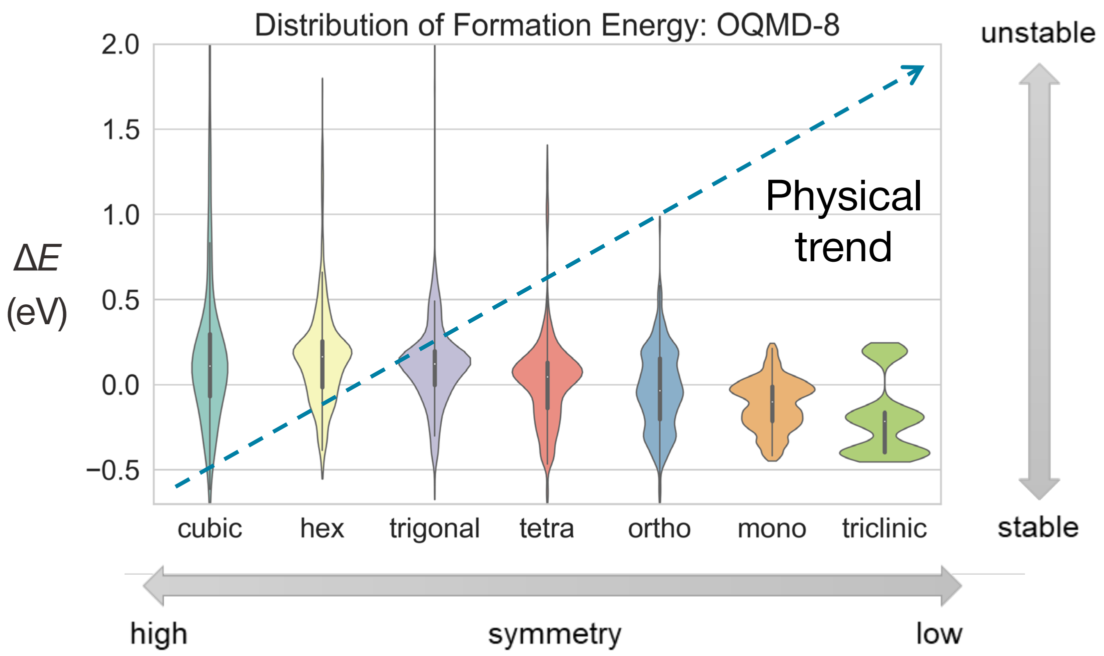


LE graph set

H. Zhang, R. Huang et al. *Mach. Learn: Sci. Tech.* (2025)

Examine data behind ML models…

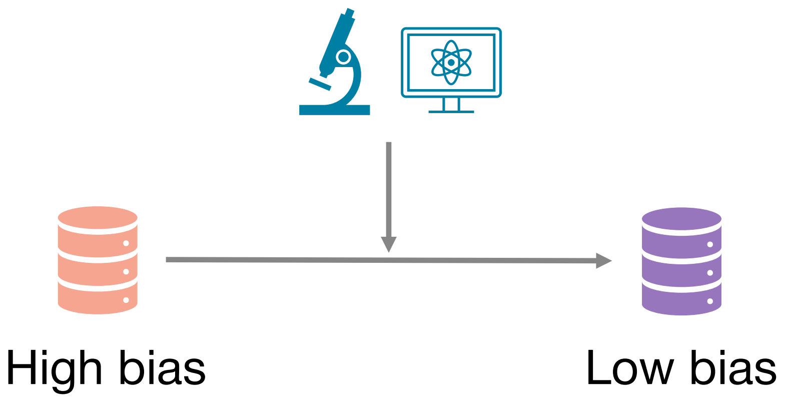
Data **bias** in materials data platforms, e.g., OQMD:



Source: unbalanced focus and knowledge
Biased data → problematic ML models.

Entropy-targeted active learning (ET-AL)

- A metric to quantify bias
- A framework to mitigate bias
    - Guide new data **acquisition**
- Construct high-quality database



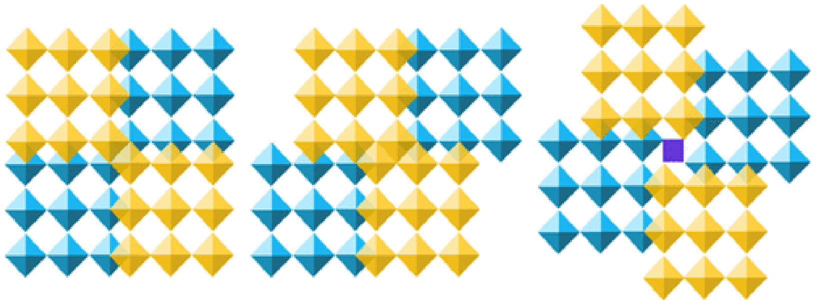
H. Zhang et al. *Applied Physics Reviews* (2023)

Summary of contributions

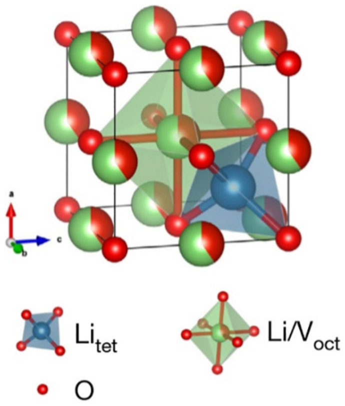
- ❖ Generic **representation** and ML model that address structural complexity
- ❖ Data **acquisition** for design in a broad composition–structure space

Combinatorial complexity formed by diverse constituents × configurations

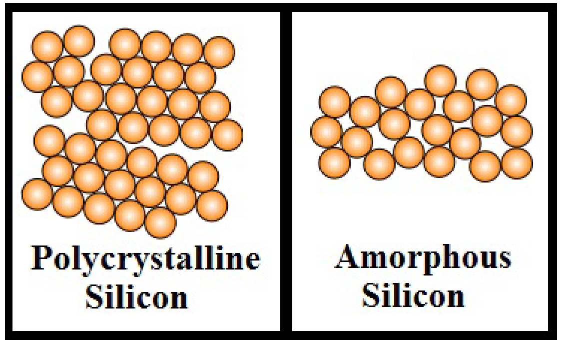
Disordered materials in energy and electronic technologies



Digital superlattices



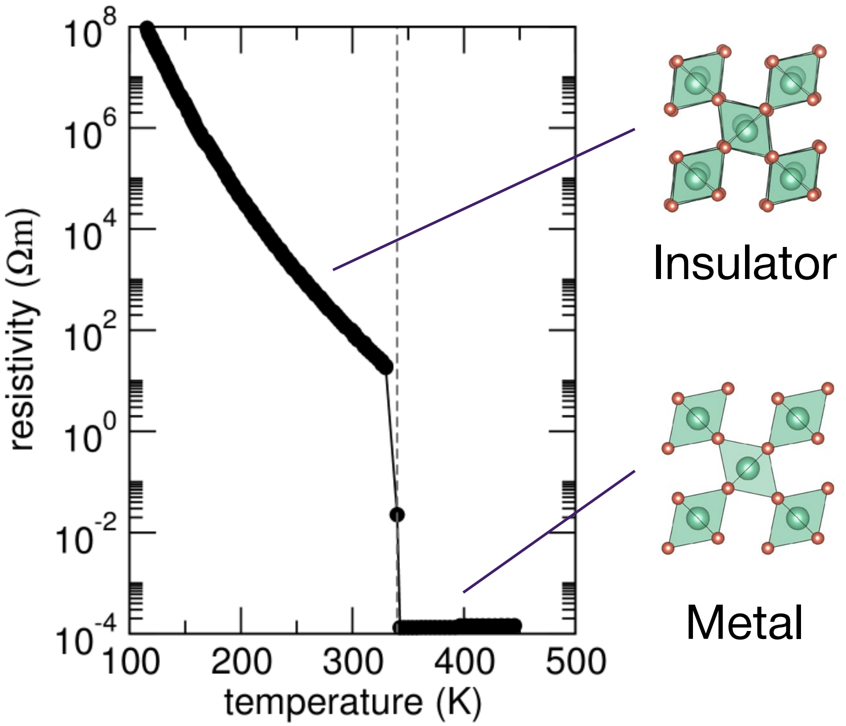
Disordered rock salt (DRX) in Li batteries



Amorphous semiconductor in electronic devices

M. Saber et al. *Inorganic Chemistry* (2023)

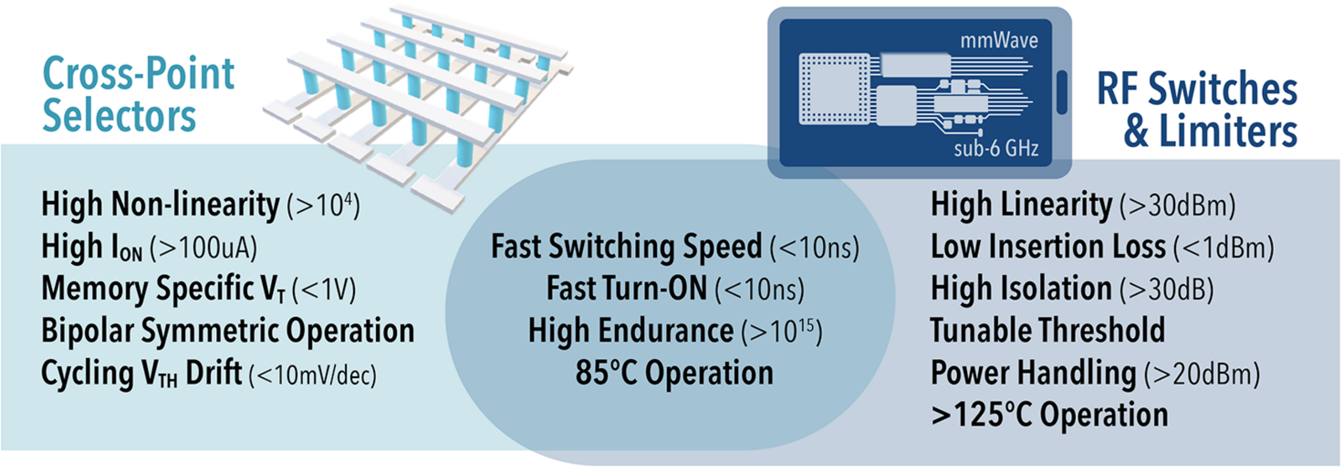
Designing combinatorial materials

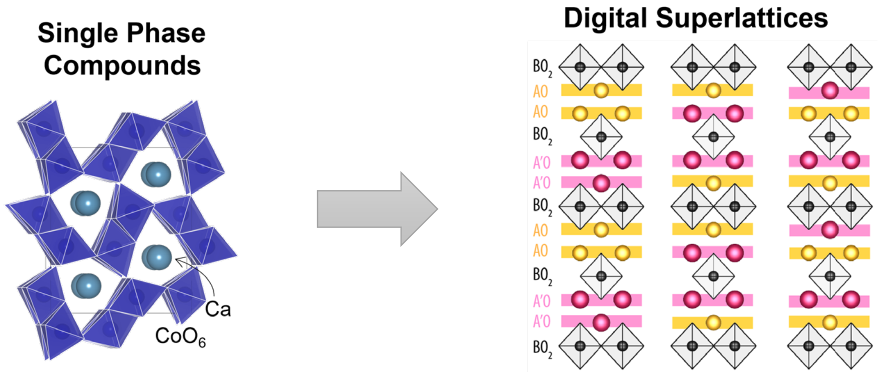


Metal–insulating transition (MIT) materials

- Tradeoff btw. resistivity change $\Delta\rho/\rho$ and temperature T_{MIT}

MIT materials are useful in microelectronic devices:





H. Zhang et al. *Accounts of Materials Research* (2025)

Acknowledgments

Dr. Wei Chen's
IDEAL Lab

Dr. James Rondinelli's
MTD Group

Ted Sargent
@ NU

Daniel Apley
@ NU

Vinayak Dravid
@ NU

Elsa Olivetti
@ MIT

A. Manthiram
@ UT Austin

Divine Kumah
@ Duke

Jie Chen
@ VT

Wei Chen
@ TAMU

A. Georgescu
@ IUB

NSF

ARPAE

CHIMaD

Northwestern
INTERNATIONAL INSTITUTE FOR NANOTECHNOLOGY