Accelerating Materials Discovery using Machine Learning

Alexandru B. Georgescu[1], Peiwen Ren[1], Aubrey Toland[2], Yiqun Wang[1], Chris Karpovich[2], Wei Chen[1], Dan Apley[1], Ramin Bostanabad[3], Elsa Olivetti[2], James Rondinelli[1]

[1]Northwestern University, [2]MIT, [3]UC Irvine

My Website

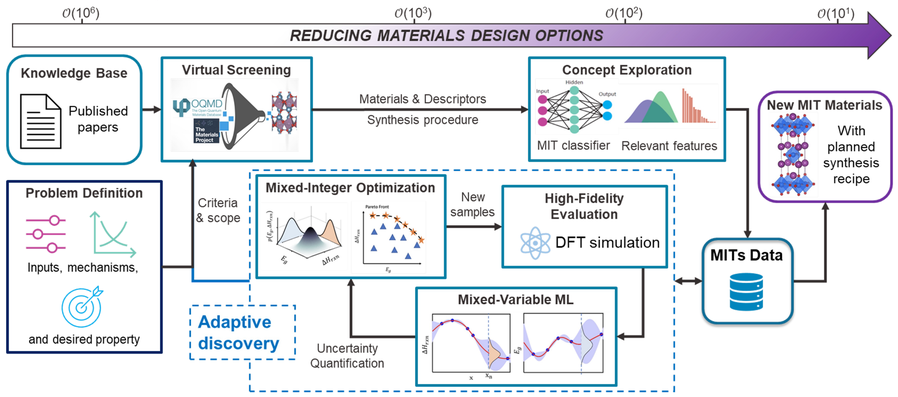
Clickable Poster

DE-AR0001209

Goals

Accelerate materials discovery and design using machine learning at every stage of the process:

- Build database of state of the art knowledge in the field
- Rapid screening of large numbers of new candidate materials
- Optimize materials families
- Provide synthesis recipes to accelerate experimental validation.
- Transform rare-event discoveries of materials to persistent discovery.



Metal-Insulator Transition (MIT) Materials

Switchable resistivity as a function of temperature/voltage

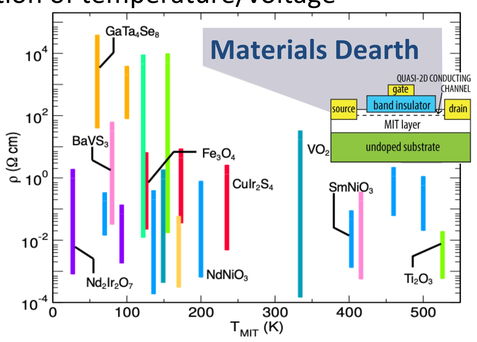
Key to new microelectronics.

Materials dearth, need :

- T_{MIT} > 300K
- High resistivity switch

Scarce & Heterogeneous :

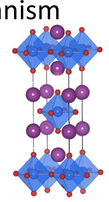
- Heterogeneous
- High fidelity calculations needed.



Density Functional Theory Calculations

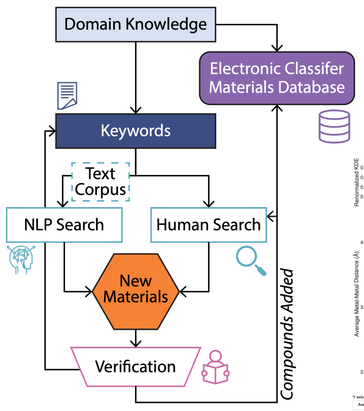
High throughput calculations not appropriate: complex mechanism

- High fidelity calculations: SCAN functional or DFT+U
- Identify MIT transition mechanism, as well as:
  - Switchability -> band gap in insulating state E_g
  - Chemical Stability: Enthalpy of formation ΔH
  - Synthesis reaction energies
- For Machine learning purposes, calculations have to be uniform
  - Constrained to a +U benchmarked for a materials class
  - SCAN exchange correlation potential

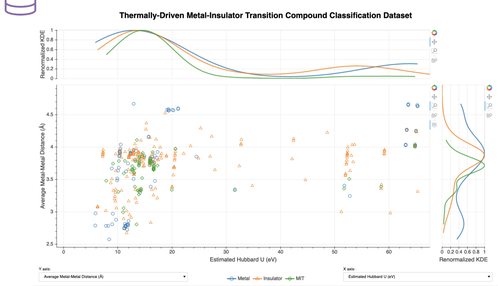


Virtual Screening: NLP-Assisted Database Creation

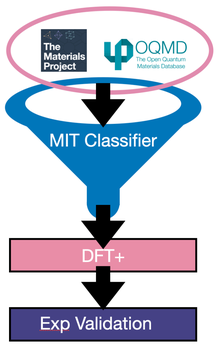
- Text corpus of over 4M articles → ~55,000 perovskite papers, ~10,000 MITs papers



- 68 MIT Compounds Identified
- 343 total compounds classified
- Database provided online



Concept Exploration: Rapid Filtration of Materials



Physically meaningful features

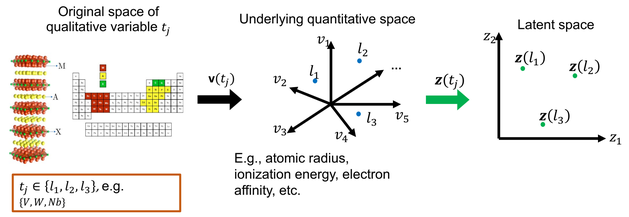
- Known: Hubbard U, Metal-Metal Distance
- New: Average Deviation of the Covalent Radius
- Easy to calculate, facilitating rapid screening
- Disseminated online via Binder: no installation
- Identified multiple new compounds for experimental validation

Concept Exploration: Synthesis Recipes

- Conditional variational autoencoder (CVAE), trained on ~15,000 known synthesis recipes
- Predicts synthesis (precursor + method) for target material
- Natural language processing (NLP): regex searching + rule-based approach, identified precursors and targets
- Extracted and classified synthesis methods (hydrothermal, sol-gel, etc.) and processing conditions

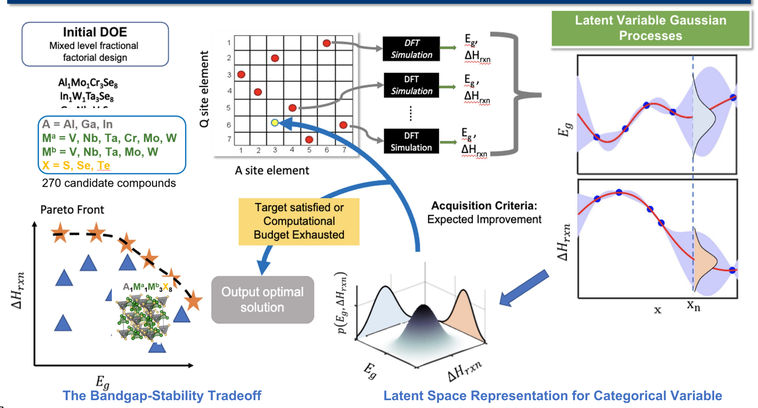
DFT-identified MIT Target	Precursor Predictions	Synthesis Operation Predictions
X–Y–N compound 1	['XCO3', 'Y2O3', 'N2O'] ['X(NO3)2', 'Y(NO3)2'] ['XCO3', 'Y(NO3)2'] ['XCl2', 'Y2O3', 'Z(NO3)3']	['prepare', 'dissolve', 'add', 'stir', 'dry'] ['dissolve', 'add', 'stir', 'transfer', 'fire', 'mix', 'seal', 'collect']

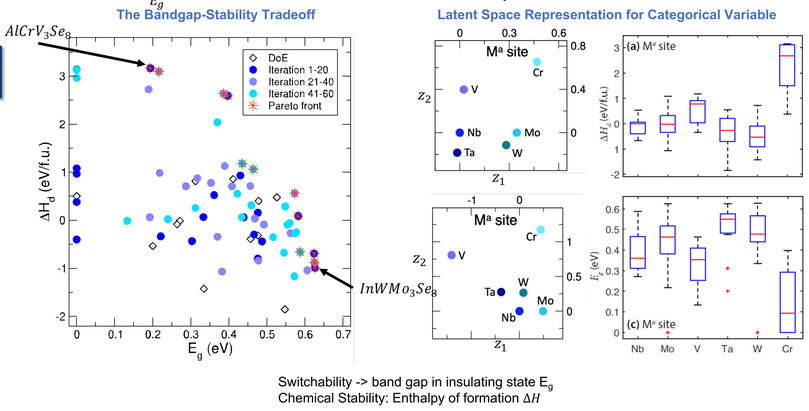
Mixed-Variable Latent Variable GP (LVGP)



- Advanced novel latent variable Gaussian process (LVGP) modeling approach for efficient mixed-variable machine learning (ML)
- High efficiency through physics-based dimension reduction
- Multi-response LVGP integrated with Bayesian Optimization for multicriteria mixed-integer optimization
- Bayesian LVGP improves performance and robustness

Adaptive Discovery





Switchability -> band gap in insulating state E_g

Chemical Stability: Enthalpy of formation ΔH

- 25% of spinel design space -> Pareto Frontier -> 12 compounds
- New Ruddlesden-Popper perovskite oxides (under preparation):
  - Insulating state stabilized by magnetism + Jahn Teller
  - Multiple never previously studied experimentally
  - Currently synthesized: Kumah Group at NCSU

Open Access Tools:

1) MIT and Stoichiometrically Related Materials Database: https://mtd.mccormick.northwestern.edu/mit-classification-dataset/
2) Classifier for rapid filtration of materials: https://tinyurl.com/mit-classifiers
3) LVGP Open access Code (Downloaded 15000 times): https://cran.r-project.org/web/packages/LVGP/index.html

Selected Publications:

1) A. B. Georgescu, P. Ren, A Toland, S. Zhang, K. Miller, D. Apley, E. Olivetti, N. Wagner, E. Olivetti, 'Database, Features and Machine Learning Model to Identify Thermally Driven Metal-Insulator Transition Compounds', Chem. Matter, DOI: 10.1021/acs.chemmater.1c00905, (2021)
2) A.B. Georgescu, A.J. Millis, "Quantifying the role of the lattice in metal–insulator phase transitions", In Press at Communications Physics, arXiv: arxiv.org/abs/2105.02271

Selected Publications (continued):

3) Alexandru B. Georgescu, Andrew J. Millis, James M. Rondinelli, 'Trigonal Symmetry Breaking and its Electronic Effeccts in Two-Dimensional Dihalides and Trihalides', arxiv.org/abs/2110.04665
4) Wang, Y., Iyer, A., Chen, W., and Rondinelli, J., "Featureless adaptive optimization accelerates functional electronic materials design", App. Phys. Rev., DOI: 10.1063/10.0002459, **cover article**, 2020
5) Zhang, Y., Apley, D., and Chen, W., "Bayesian Optimization for Materials Design with Mixed Quantitative and Qualitative Variables", *Scientific Report*, 10, Article number: 4924, 2020.
6) C. Karpovich, E. Pan, Z. Jensen, E. A. Olivetti. Interpretable Machine Learning Enabled Inorganic Reaction Classification and Synthesis Condition Prediction. *In preparation*
7) Y. Wang, XJ Zhang, F. Xia, E.A. Olivetti, S.D. Wilson, R Seshadri, J.M. Rondinelli, 'Learning the crystal structure genome for property classification', Physical Review Research, 4, 023029, 2022

Technology To Market:

- **Provisional patent** for the entire computational design framework
- Invention disclosure and "**Click License**" of LVGP code being adopted by 3DS into commercial software (Isight & the Simulia Execution Engine)
- **Benchmark** of LVGP with Citrine's random forest-based techniques
- **Industry collaboration** with QuesTek to integrate LVGP+BO into alloy design workflow
- Multiple materials undergoing experimental validation