Hybrid Modeling of Mixed Perovskite Oxides with Gaussian Process Regression: Overcoming Data Scarcity

Franz M. Rohrhofer[1,*], Jean-François Paul[2], Bernhard C. Geiger[1], Elise Berrier[2]

1) Know Center Research GmbH, Sandgasse 34, 8010 Graz, Austria
2) Université Lille, CNRS UMR8181, Unité de Catalyse et de Chimie du Solide, UCCS, F-59655 Villeneuve d'Ascq, France
*) Corresponding author: frohrhofer@know-center.at

The KNOWSKITE-X Project

Development of **science-based electrode materials** for reversible **chemical-to-power cells** (fuel cell, electrolytic cell), enabling efficient energy conversion and storage.

- **Key Benefits:**
  - Supports renewable energy integration by storing excess power as carbon-free fuel.
  - Focuses on **Mixed Perovskite Oxides** with reduced critical content while maintaining high performance and economic viability.
- **Objective:** uncovering correlations between composition, structure, activity, and performance in Perovskite-based electrode materials using **multi-scale modeling**, **advanced characterization**, and **machine learning**.

Problem Formulation

Modeling Mixed Perovskite Oxides is challenging due to

- **Data scarcity** in high-fidelity simulations and experimental measurements.
- **Compositional complexity** due to the vast number of elements considered in the composition[1].
- **Computational expenses** as first-principles methods, like Density Functional Theory (DFT), are accurate but costly.

Hybrid modeling with Gaussian Process Regression can help in modeling Mixed Perovskite Oxides by:

- Learning structure-property relationships **from limited data**.
  - Predicting key material properties, like DFT energies, **accurately & efficiently**.
    - Reducing reliance on expensive simulations or experiments.

Gaussian Process Regression

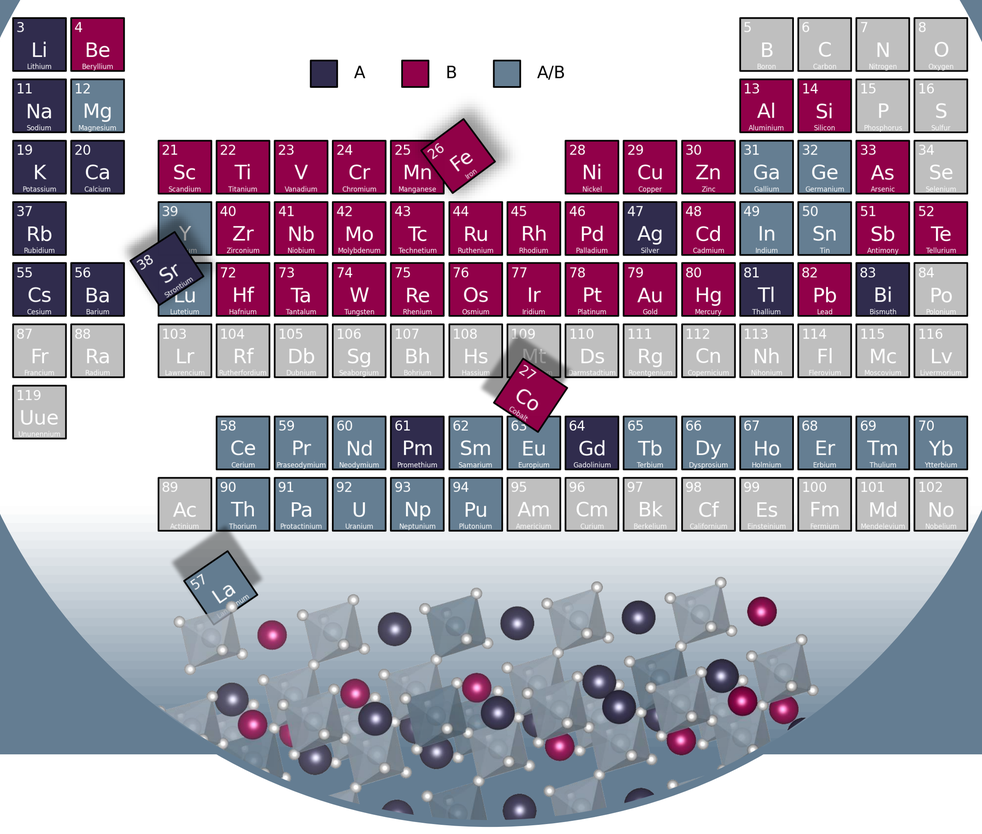
- **Models target property** E as non-linear, continuous function, $E = f(x)$.
- **Fits data examples** $\{(x^i, E^i)\}_{i=1}^N$ to learn underlying function f.
- **Measures similarity** through kernel $k(x, x')$ with $k \to 1$ as $\|x - x'\| \to 0$.
- Relies on **structure representation** x.

Prediction
(on new data x^*)

$$E(x^*) = \sum_{i=1}^{N} \lambda^i \, k(x^i, x^*)$$

Mixed Perovskite Oxides

$$A_\alpha A'_{1-\alpha} B_\beta B'_{1-\beta} O_{3\pm\delta}$$



Structure Representation

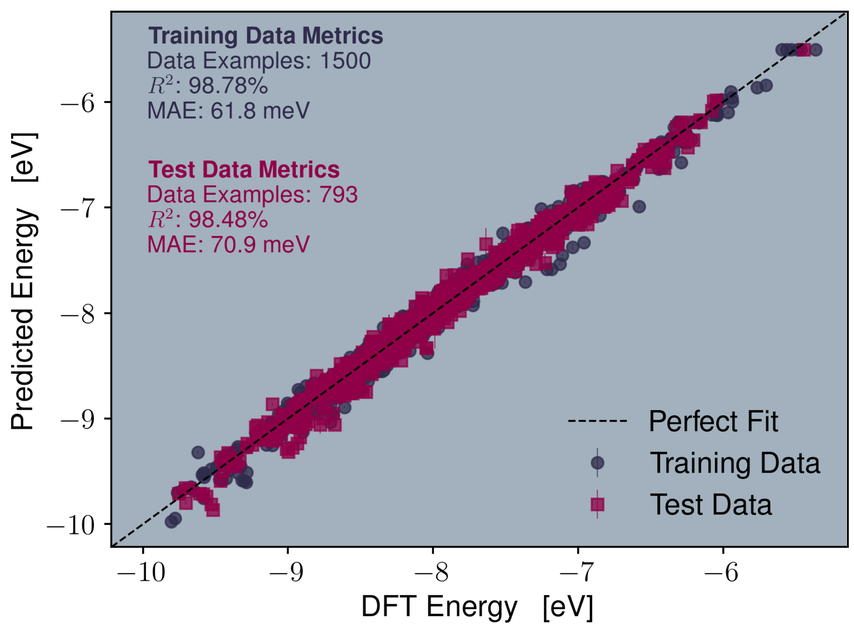
- **Uses descriptors/features** to define space in which similarity is measured.
- **Compositional features** by group & period of elements A,A',B, and B'.
- **Doping features** by A- and B-site doping levels α and β.

Composite kernel
(to measure similarity)

$$k = \underbrace{\underbrace{k_\alpha}_{\text{doping}} \cdot \underbrace{[k_A + k_{A'}]}_{\text{composition}}}_{A-\text{site}} + \underbrace{\underbrace{k_\beta}_{\text{doping}} \cdot \underbrace{[k_B + k_{B'}]}_{\text{composition}}}_{B-\text{site}}$$

Hybrid Modeling of DFT Energies

… under Data Richness



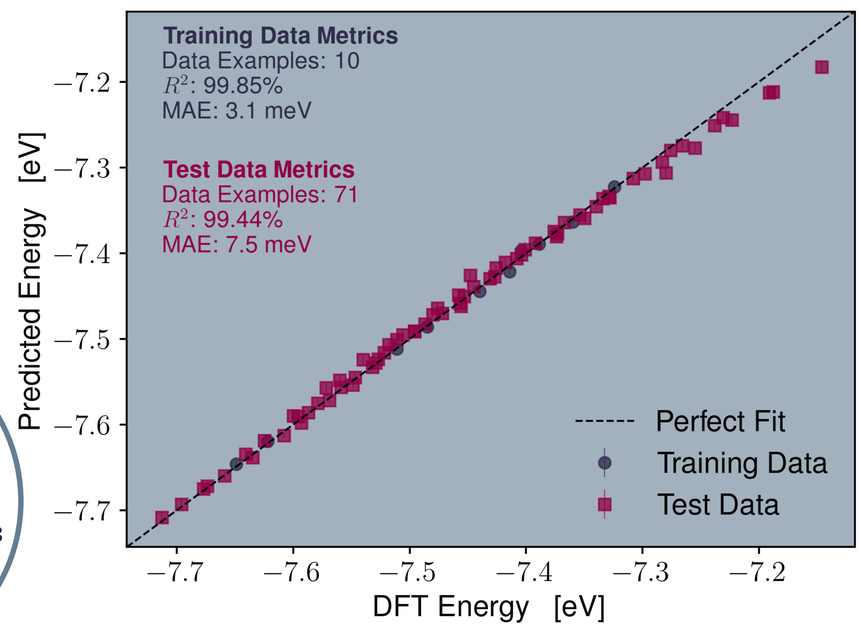
Data mined from the **Materials Project**[2]

$$A_\alpha A'_{1-\alpha} B_\beta B'_{1-\beta} O_3$$

Generated Data

$$La_\alpha Sr_{1-\alpha} Co_\beta Fe_{1-\beta} O_3$$

… under Data Scarcity



KNOWSKITE-X
Knowledge-driven fine-tuning of perovskite-based electrode materials for reversible Chemicals-to-Power devices
knowskite-x.eu
Funded by the EU

References

[1] Tao, Qiuling, et al. "Machine learning for perovskite materials design and discovery." Npj computational materials 7.1 (2021): 23.
[2] Jain, Anubhav, et al. "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation." APL materials 1.1 (2013).