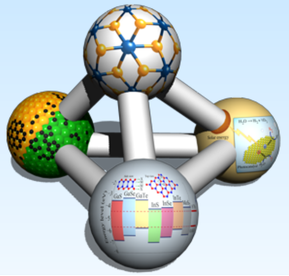
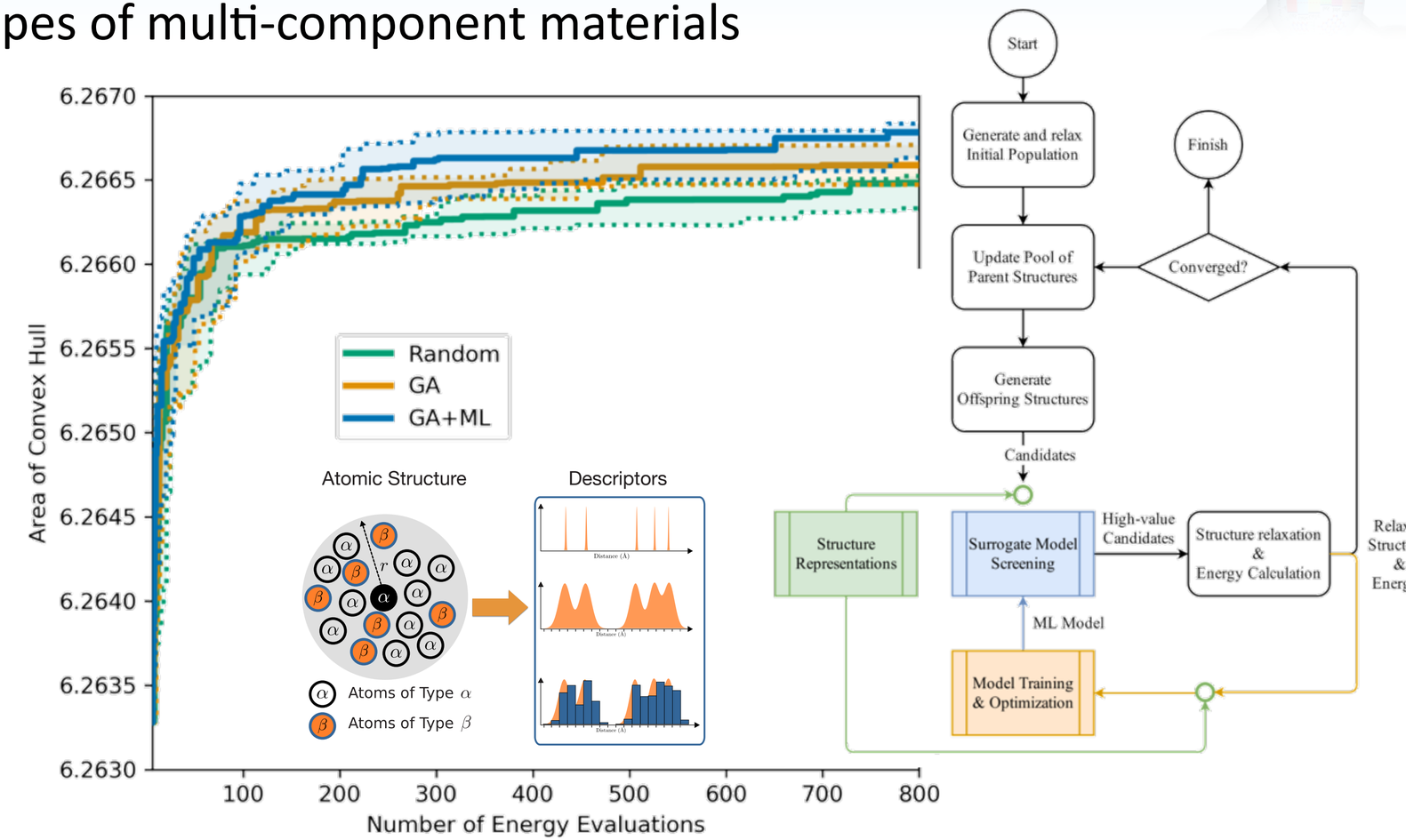
Machine-learning for the Exploration of Energy Landscapes

Stephen Xie, Shreyas Honrao, Anne Marie Tan, Halee Lester, and Richard G. Hennig
Materials Science and Engineering & Quantum Theory Project, University of Florida



- Machine learning of energy landscapes of multi-component materials
- Radial and angular distribution functions as descriptors
- On-the-fly machine learning of the relaxed energy landscape accelerates evolutionary structure searches
- Choice of stochastic screening criteria affects convergence behavior of GASP+ML



Powered by
MPInterfaces & materialsweb

rhennig@ufl.edu
http://hennig.mse.ufl.edu

Atomistic Simulations for Industrial Needs
August 5-7, 2020 • NIST • Gaithersburg, MD

Acknowledgment

- GASP genetic algorithm: **B. Revard**, W. Tipton, A. Yesupenko, **H. Lester**
- **Machine learning of energy landscapes: S. Xie**, **S. Honrao**, B. Antonio, **A. M. Tan, M. Rupp** (U. Konstanz)
- Machine learning of supeconductivity: P. Hirschfeld, J. Hamlin, G. Stewart
- Financial support by NSF, DOE, NIST
- Computational resources: HiPerGator@UF, NSF XSEDE, and Google Cloud Platform

NSF

NIST

Center for BRIGHT BEAMS
A National Science Foundation Science & Technology Center

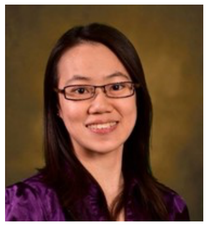
XSEDE

HiPerGator 2.0
The University of Florida Supercomputer

Office of Basic Energy Sciences
Serving the Present, Shaping the Future

2D Crystal Consortium
NSF Materials Innovation Platform

Google Cloud Platform



rhennig@ufl.edu
http://hennig.mse.ufl.edu

Machine Learning in Materials Science

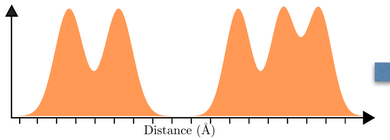
Supervised Learning

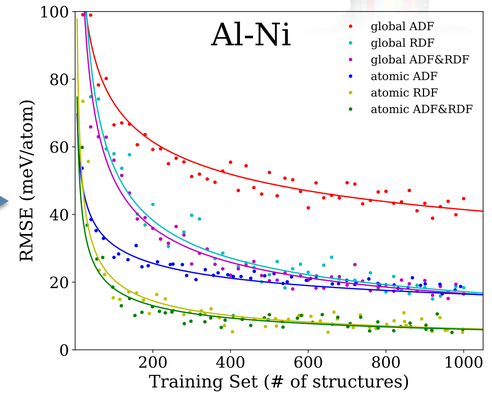
Big data

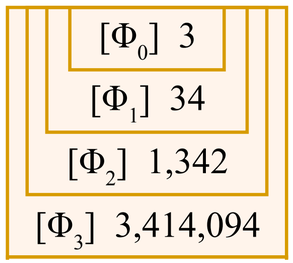
- High-throughput screening for properties
- Surrogate models for structure exploration

Small data

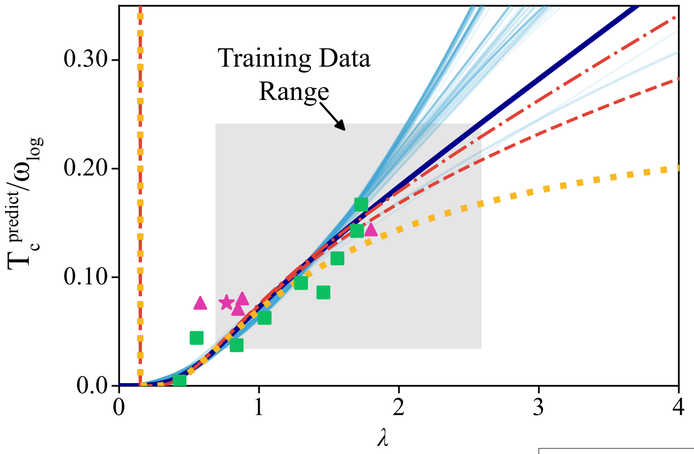
- Symbolic regression for analytic expressions





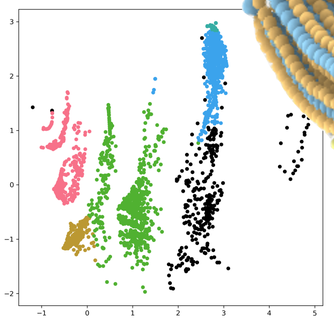


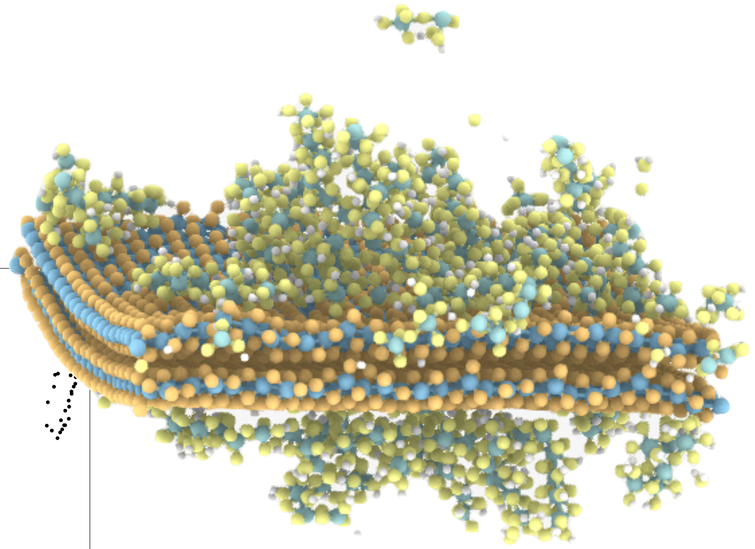
$$\omega_{\log}\ ,\ \mu^*\ ,\ \lambda\ ,\ \ldots$$
$$\omega_{\log}\times\lambda\ ,\ \sqrt{\mu^*}\ ,\ \lambda^3\ ,\ \ldots$$
$$\lambda^3\times(\omega_{\log}\times\lambda)\ ,\ \lambda^3+\sqrt{\mu^*}\ ,\ \ldots$$
$$\lambda^3\times(\omega_{\log}\times\lambda)\ /\ \left(\lambda^3+\sqrt{\mu^*}\right),\ \ldots$$



Unsupervised Learning

- Identification of new configurations in MD simulations





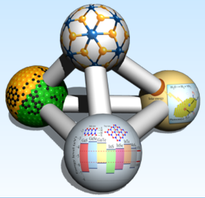
Powered by
MPInterfaces & materialsweb

Honrao, Xie, Hennig, J. Appl. Phys., in print (2020)
Honrao, Anthonio, Ramanathan, Gabriel, Hennig, Comp. Mater. Sci. 158, 414 (2019)
Xie, Stewart, Hamlin, Hirschfeld, Hennig, Phys. Rev. B 100, 174513 (2019)

Part I: Exploration of Materials Energy Landscapes by Evolutionary/Genetic Algorithms

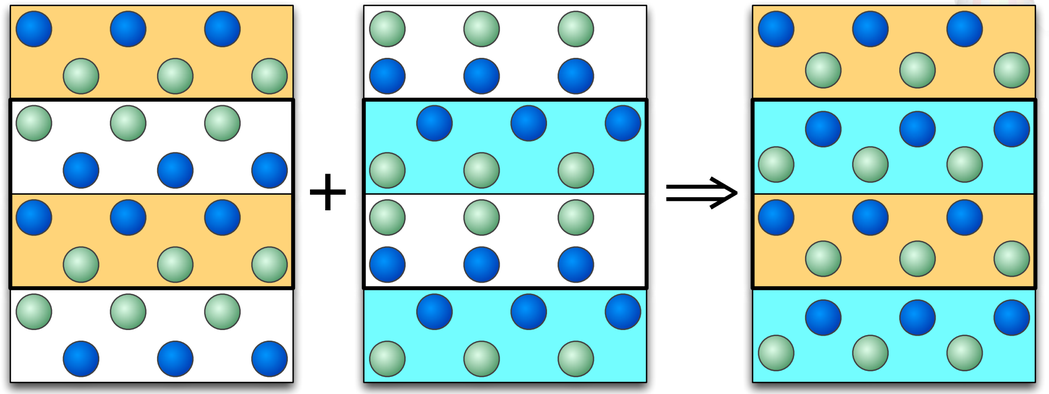
rhennig@ufl.edu
http://hennig.mse.ufl.edu

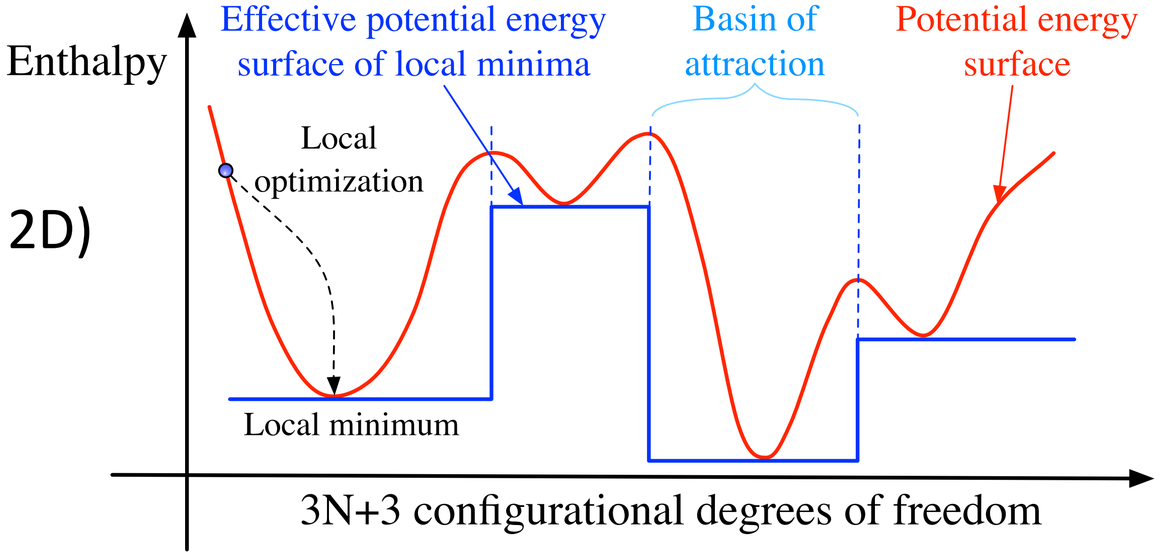
Genetic Algorithm Search for Crystalline Materials



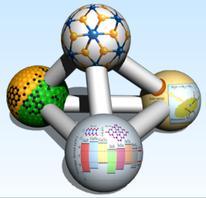
- Tipton & Hennig (2013): GASP
- GASP is a grand canonical evolutionary algorithm for global structure optimization
  - Generation with mutation & crossover
  - Local optimization through relaxation
  - Identify low-energy basins of attraction
- Revard et al. (2016): GASP-Python
  - Continuous generation of candidates
  - Support for low-dimensional materials (0D, 1D, 2D)
- Xie and Kolluru et al.: GASP 2.0
  - Support for ML and surface phases

Freely available on our group's Github page:
https://github.com/henniggroup/GASP-python

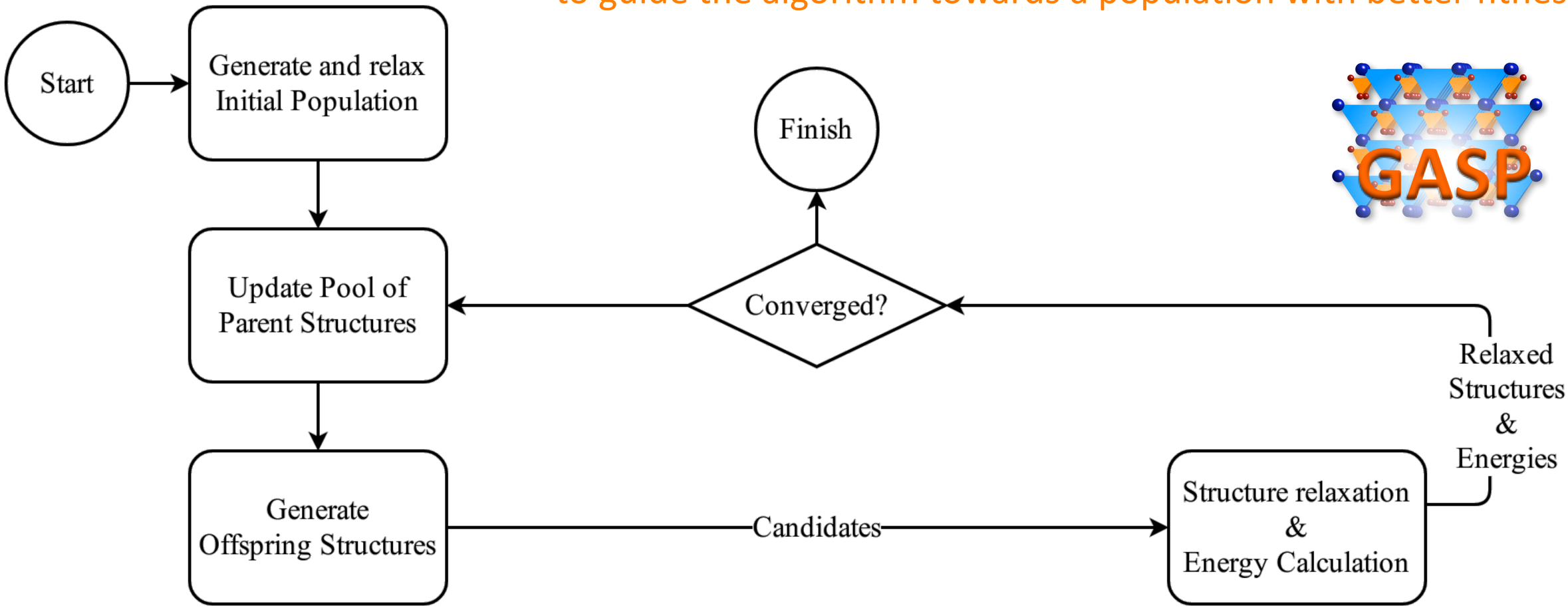




Genetic Algorithm for Structure Predictions (GASP)



Parent candidates with higher fitness are preferentially selected to guide the algorithm towards a population with better fitness.



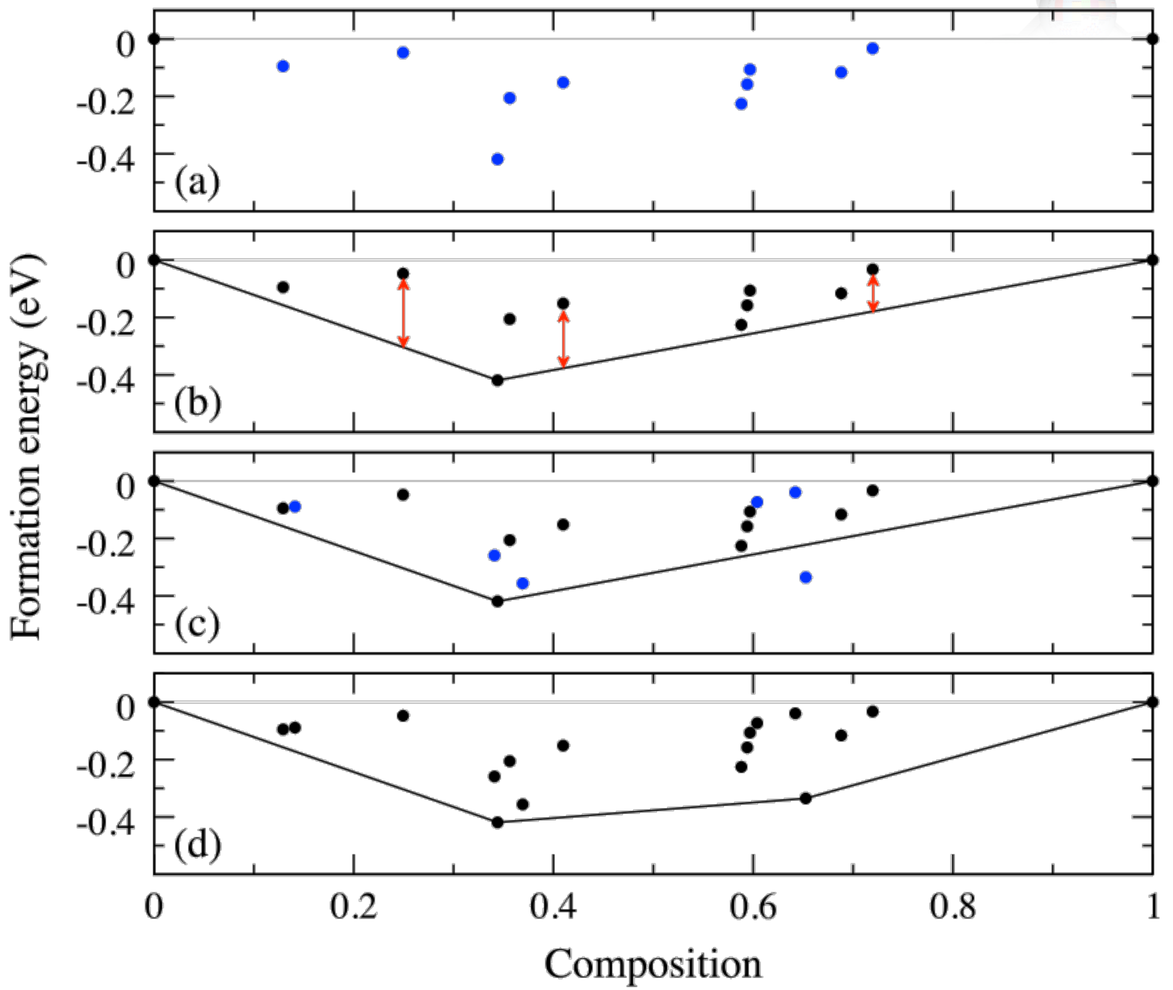
rhennig@ufl.edu
http://hennig.mse.ufl.edu

Fitness Criterion: Thermodynamic Stability

- **Formation Energy:** $E_{\mathrm{F}} = E_{\mathrm{tot}} - x_A E_A - x_B E_B$
- Distance from the convex hull ΔE_{H} is a measure of relative stability
- **Fitness:**

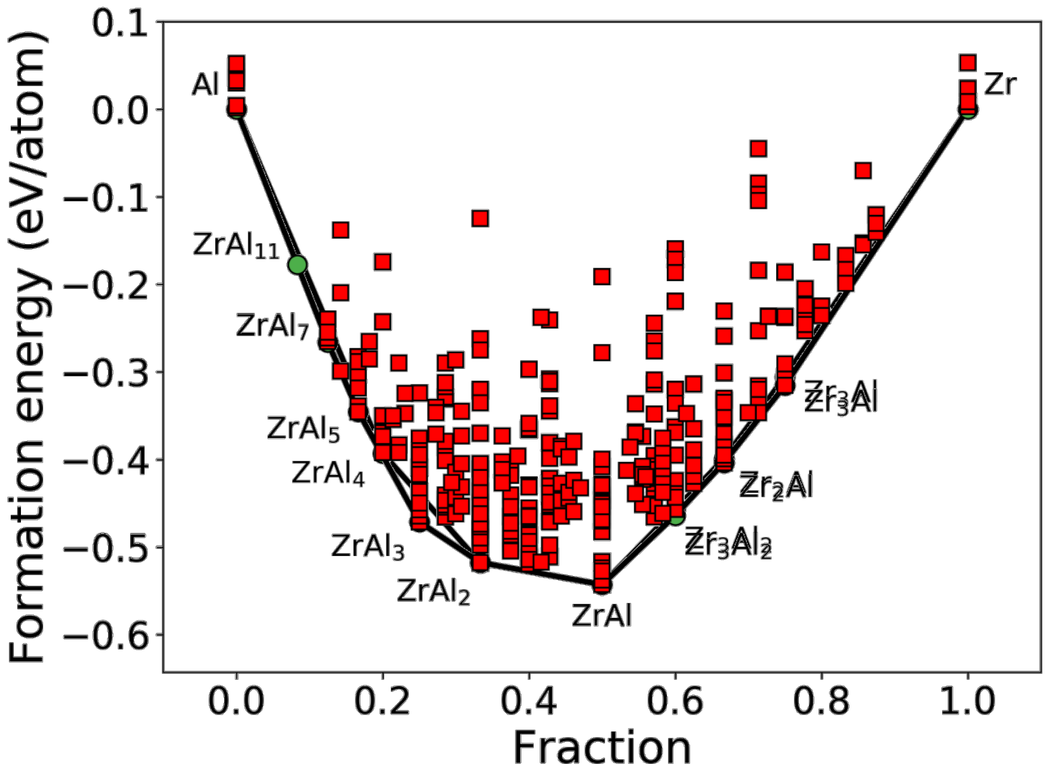
$$f = \frac{E_{\mathrm{H,max}} - E_{\mathrm{H}}}{E_{\mathrm{H,max}} - E_{\mathrm{H,min}}}$$

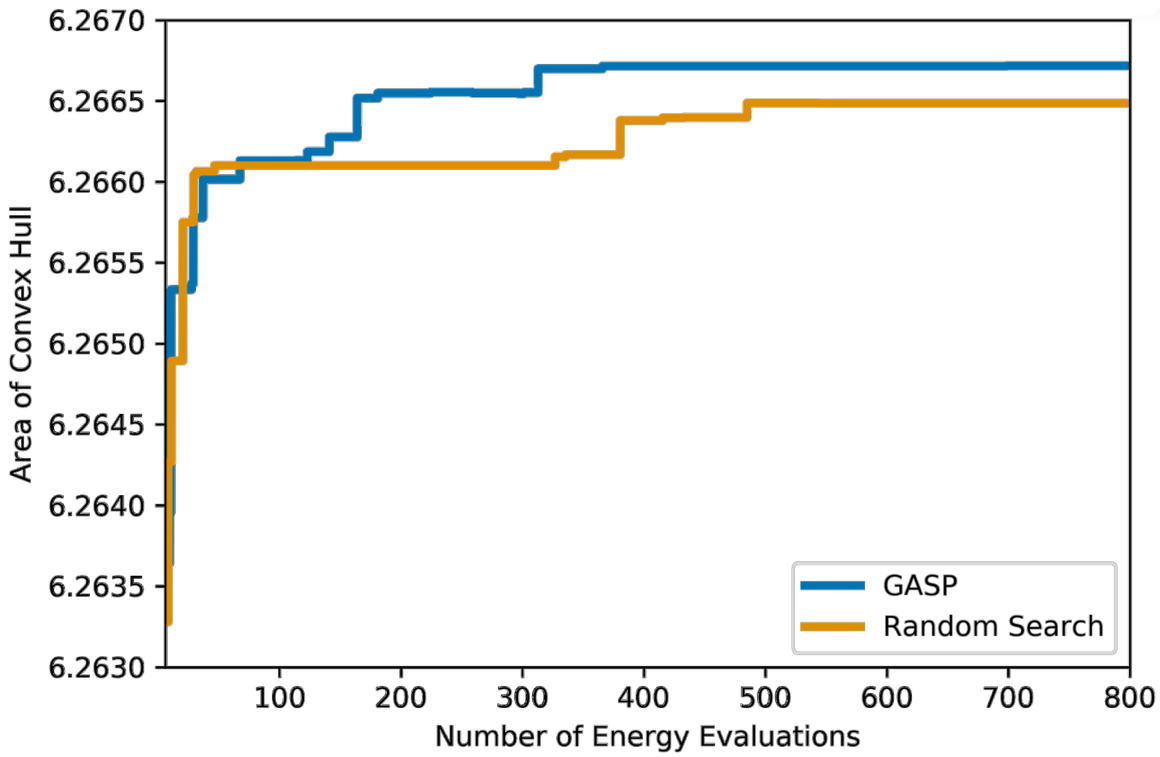
Parent candidate with higher fitness has higher probability of participating in mutation and crossover



rhennig@ufl.edu
http://hennig.mse.ufl.edu

Example: Zr-Al with EAM Potential

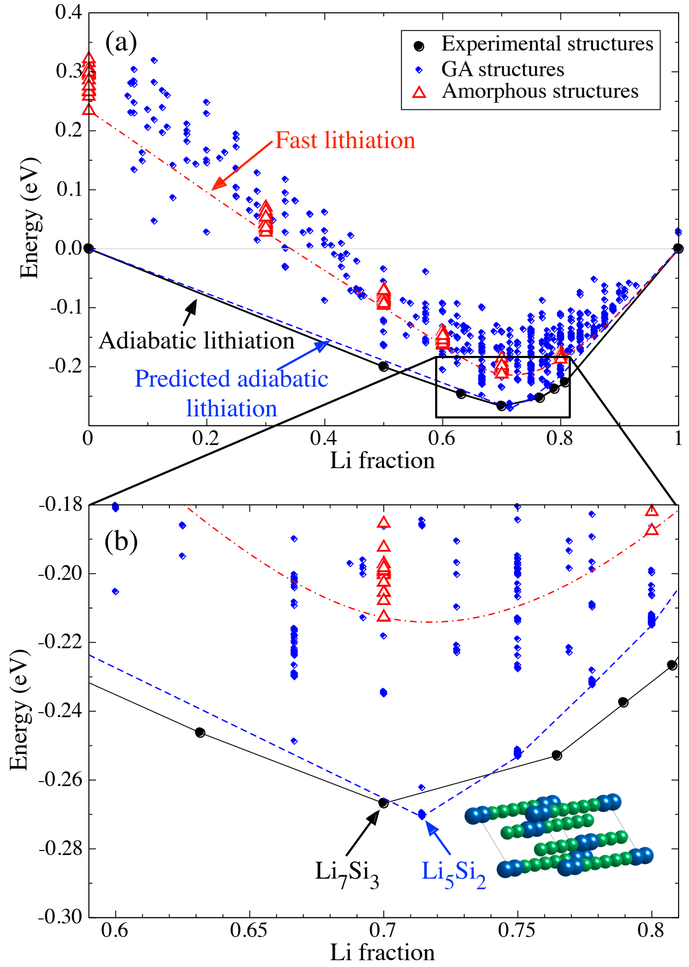




- Random search requires 2-3x more structure relaxations
- Genetic algorithm learns from previous structures

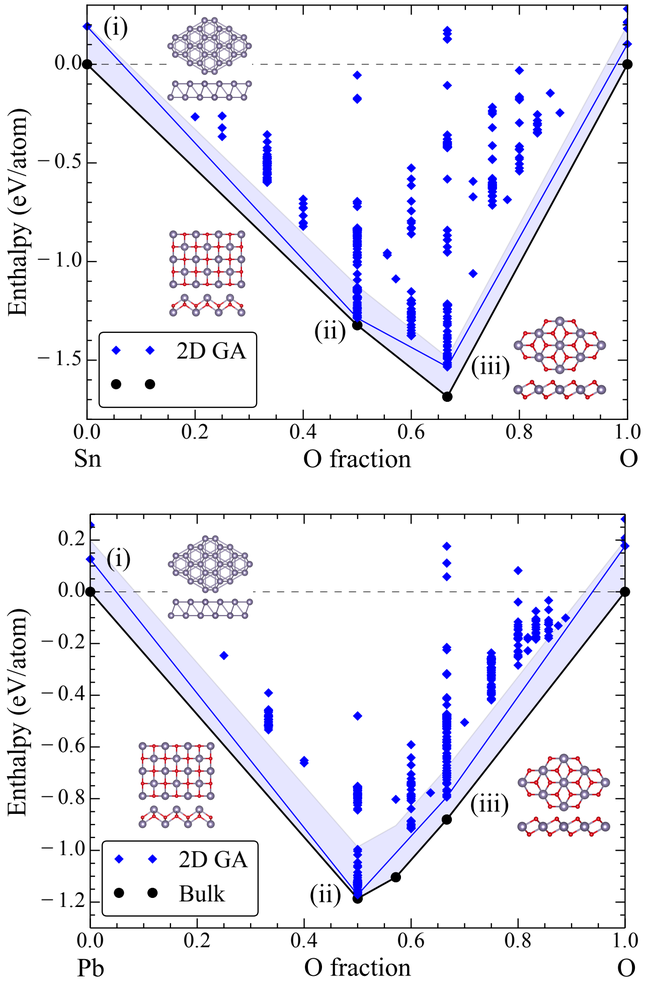
rhennig@ufl.edu
http://hennig.mse.ufl.edu

Li-Si for Batteries



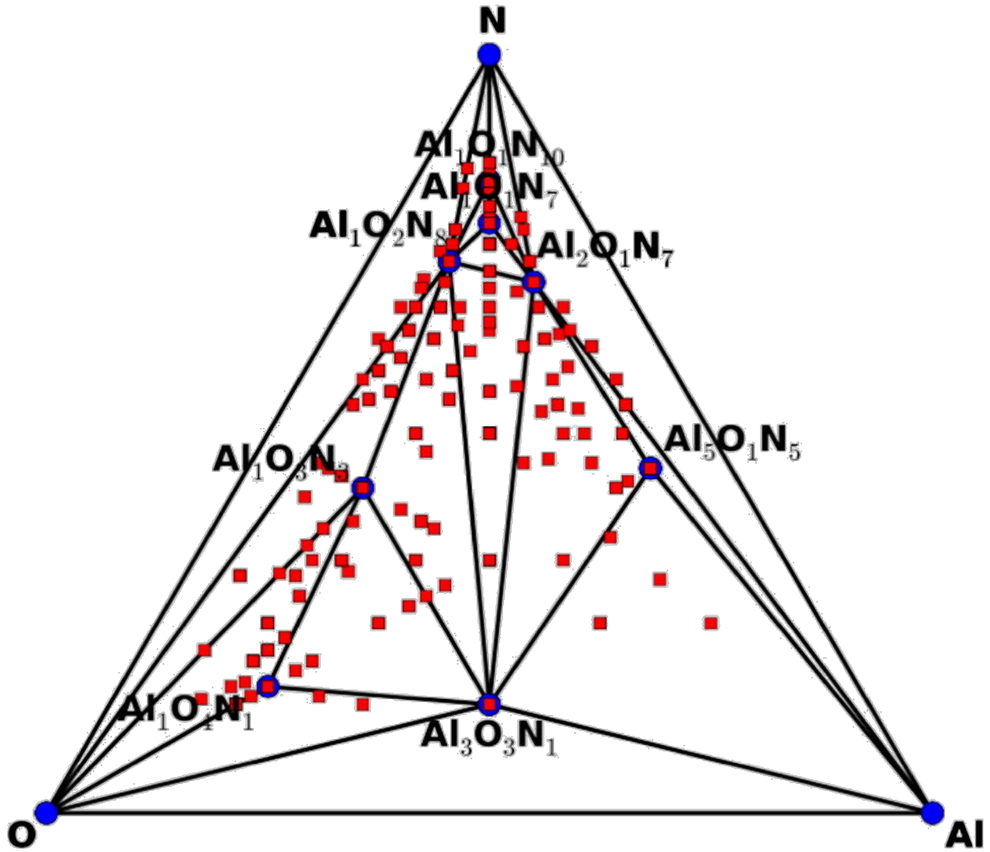
Tipton, Bealing, Mathew, Hennig,
Phys. Rev. B 87, 184114 (2013)

2D Materials for Dielectrics



Singh, Revard, Ramanathan, Ashton, Tavazza, Hennig
Phys. Rev. B 95, 155426 (2017).

Multinary Phase Diagrams (COMB)



Choudhary, Liang, Mathew, Revard, Chernatynskiy,
Phillpot, Hennig, Sinnott, Comp. Mater. Sci. 113, 80 (2016).

Can we use surrogate models to discover new phases with fewer energy evaluations?

Powered by
MPInterfaces & materialsweb

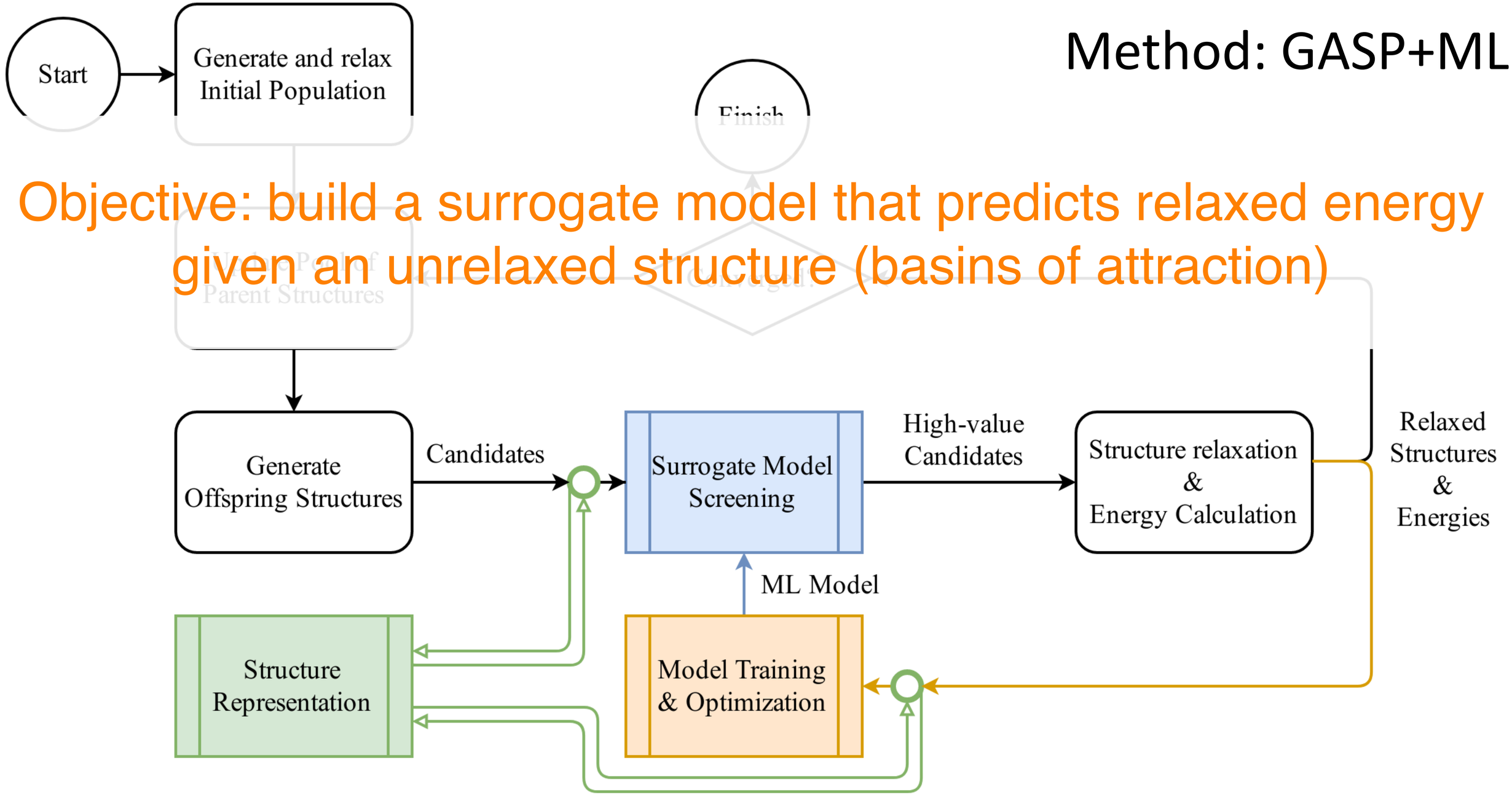
rhennig@ufl.edu
http://hennig.mse.ufl.edu

Part II: Machine Learning of Energy Landscapes

rhennig@ufl.edu
http://hennig.mse.ufl.edu

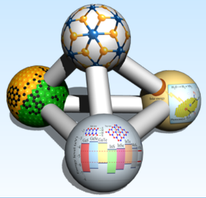
Method: GASP+ML

Objective: build a surrogate model that predicts relaxed energy given an unrelaxed structure (basins of attraction)



rhennig@ufl.edu
http://hennig.mse.ufl.edu

Structure Representation for Machine Learning



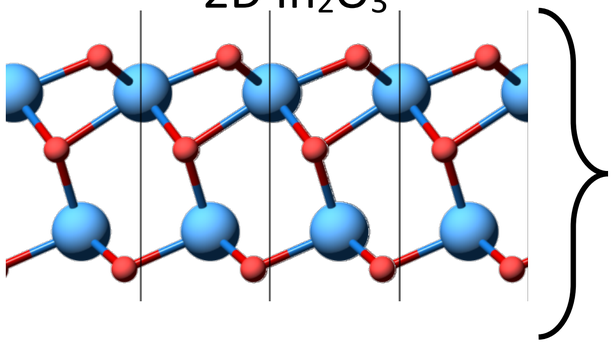
- Representation or descriptor:
  Numeral encodings of atomic structure with constant-length vectors
- Requirements for accurate ML:
  - Invariance
  - Uniqueness
  - Smoothness

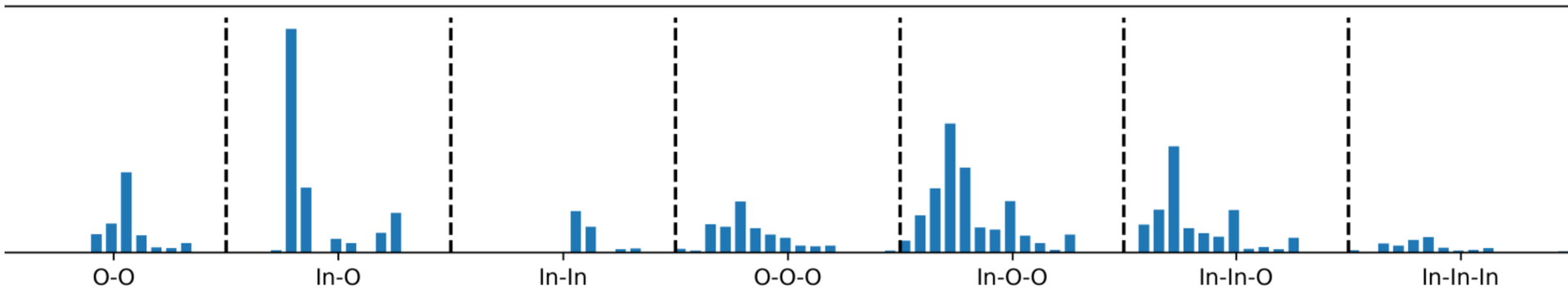
Many Body Tensor Representation

- Based on histograms of distances and angles in structure with smearing
- Sum, arrange by element combinations, concatenate

Goal: Use regression to relate constant-length vector representations to relaxed formation energy

2D In_2O_3

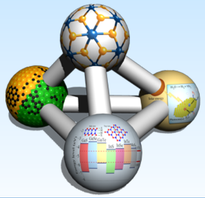




H. Huo, M. Rupp 2017 (arXiv: 1704.06439)

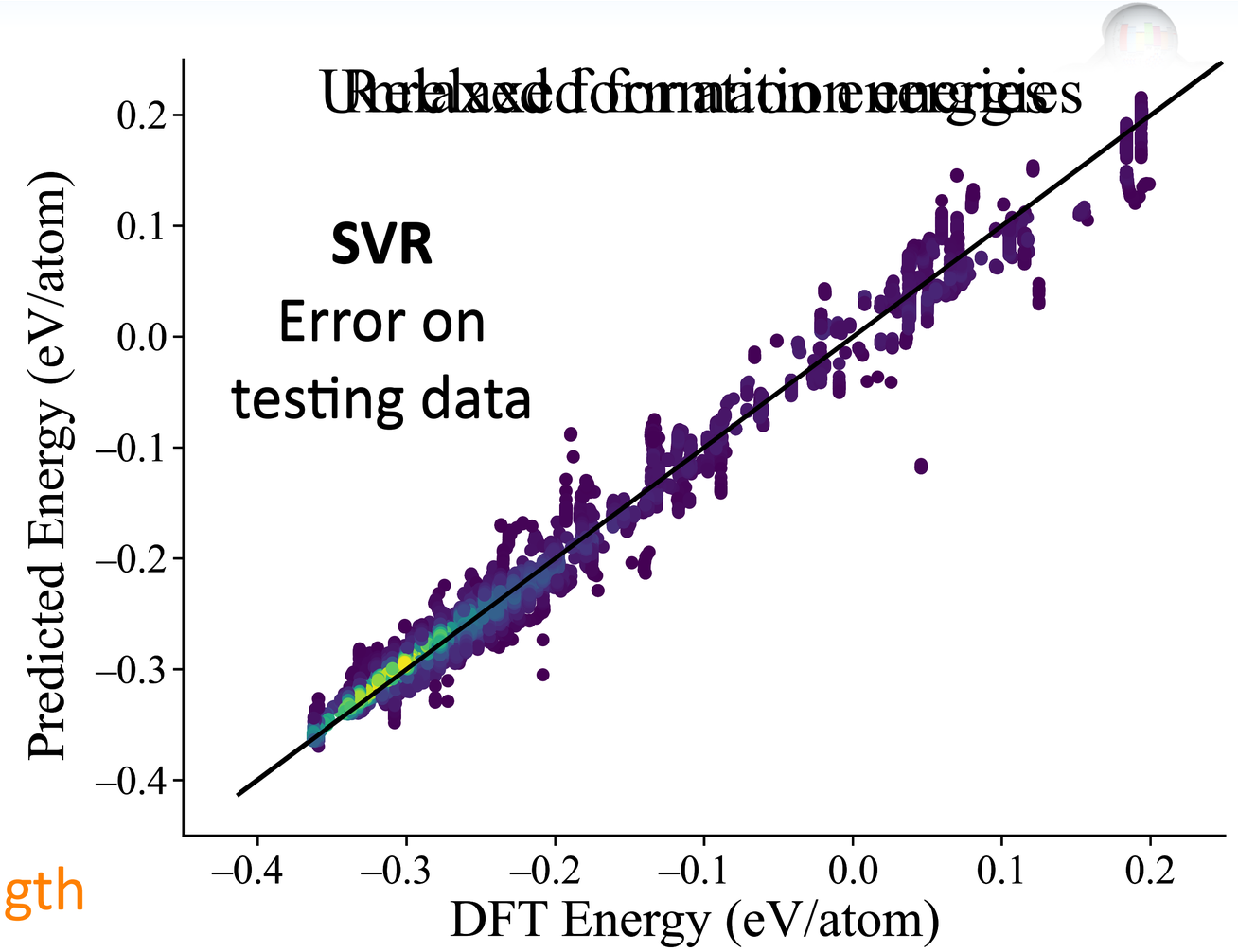
Powered by
MPInterfaces & materialsweb

Learning Energies from Structure Representations



Honrao et al. (2019)

- **Support Vector Machine** and **Kernel Ridge Regression** benchmarks
- Successfully learned the **relaxed formation energies from unrelaxed structures** using distance and angle based representation
- Bulk Li-Ge dataset
- Gaussian Kernel
- Length scale and Regularization parameter optimized by cross-validated random search

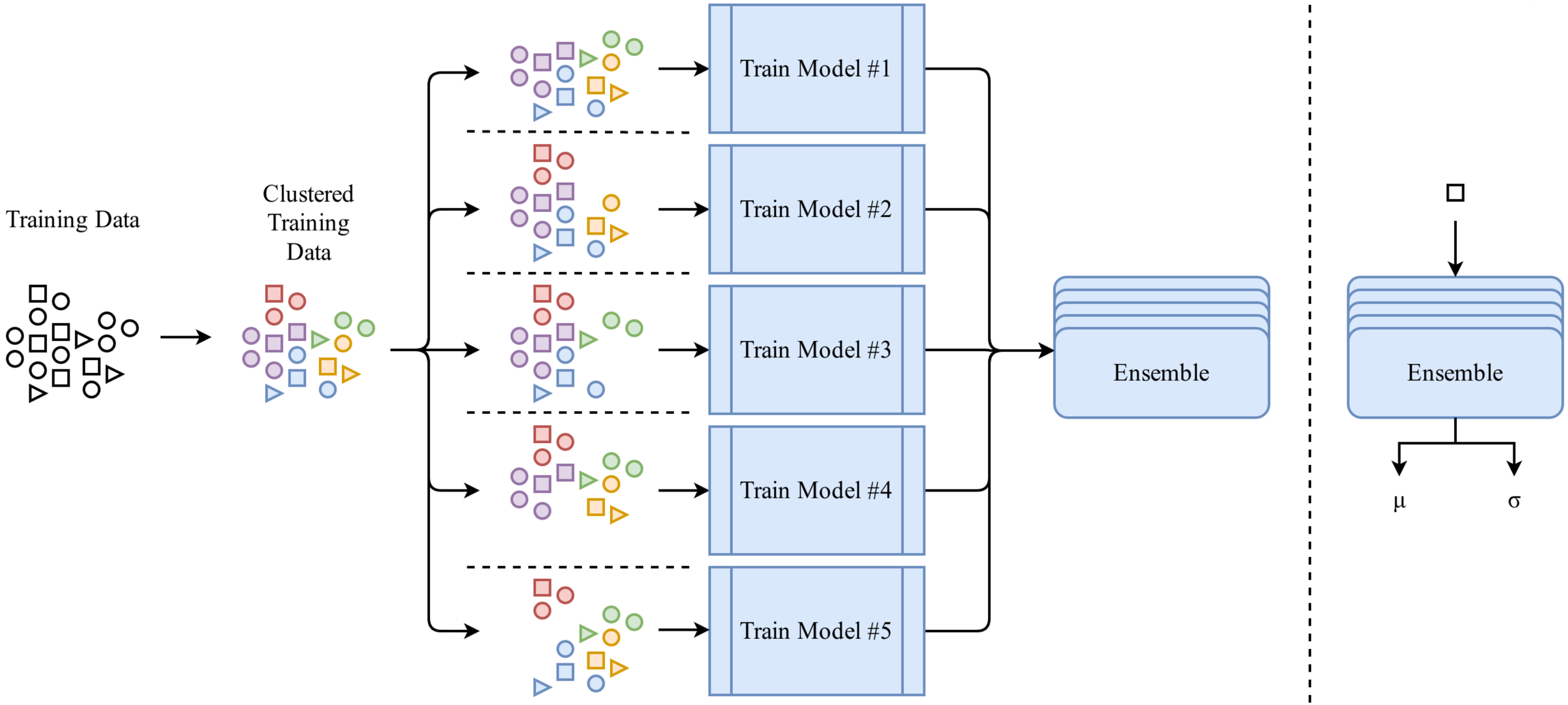


Goal: use regression to relate constant-length vector representations to relaxed formation energy

Powered by
MPInterfaces & materialsweb

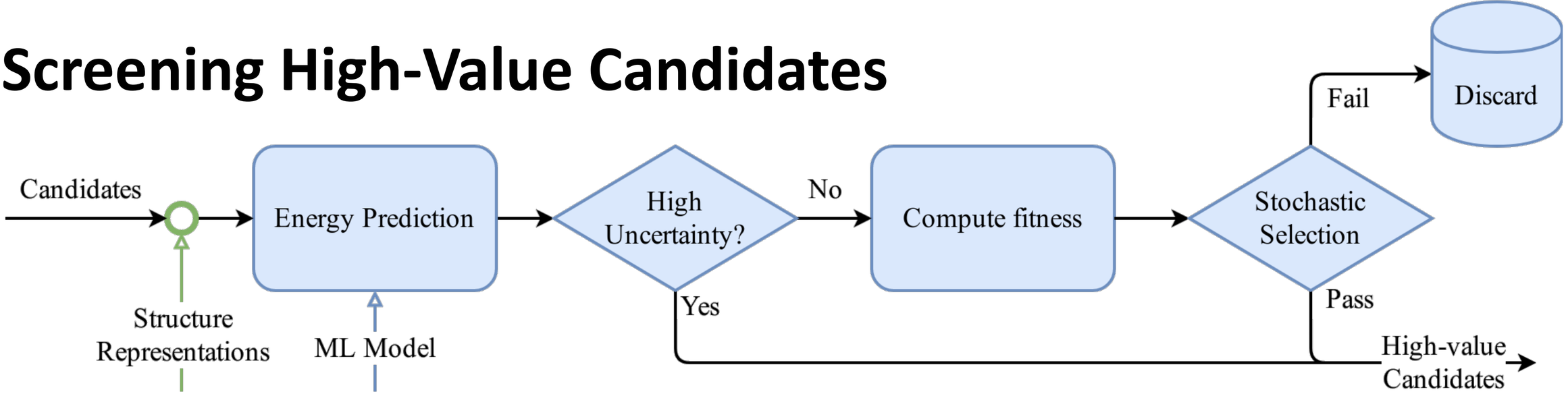
Honrao, Anthonio, Ramanathan, Gabriel, Hennig, Comp. Mater. Sci. 158, 414 (2019)

Ensemble Learning for Estimating Predictive Uncertainty



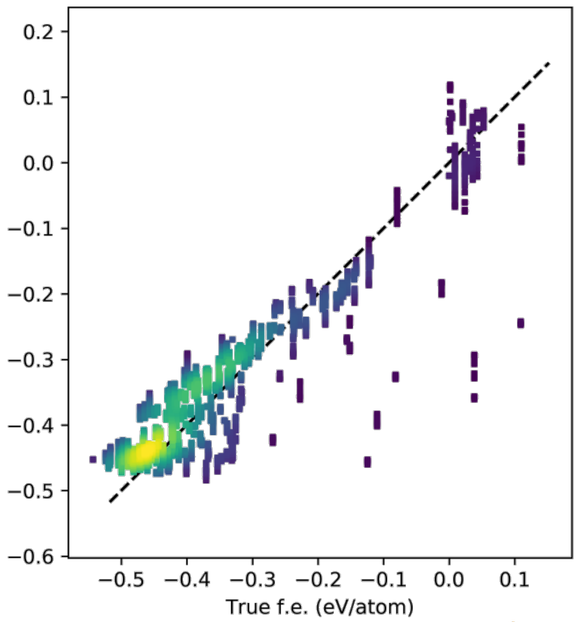
rhennig@ufl.edu
http://hennig.mse.ufl.edu

Screening High-Value Candidates

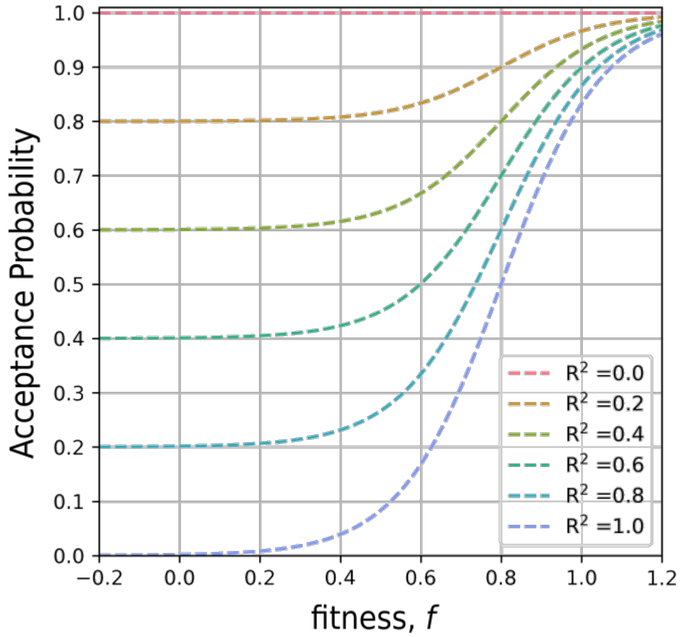


Acceptance probability depends on the predicted fitness and the expected performance of the model

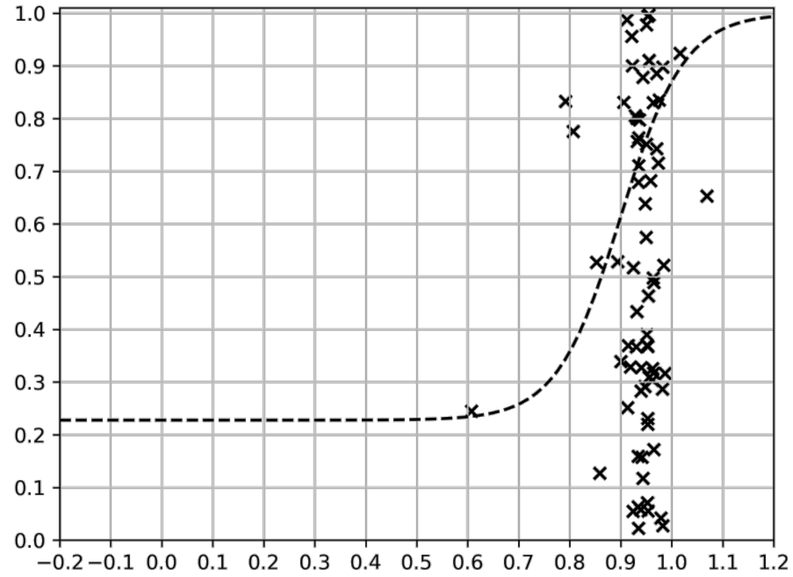
1) Ensemble R^2 from cross validation



2) Select Acceptance Probability Curve



3) Accept/Reject new candidates stochastically

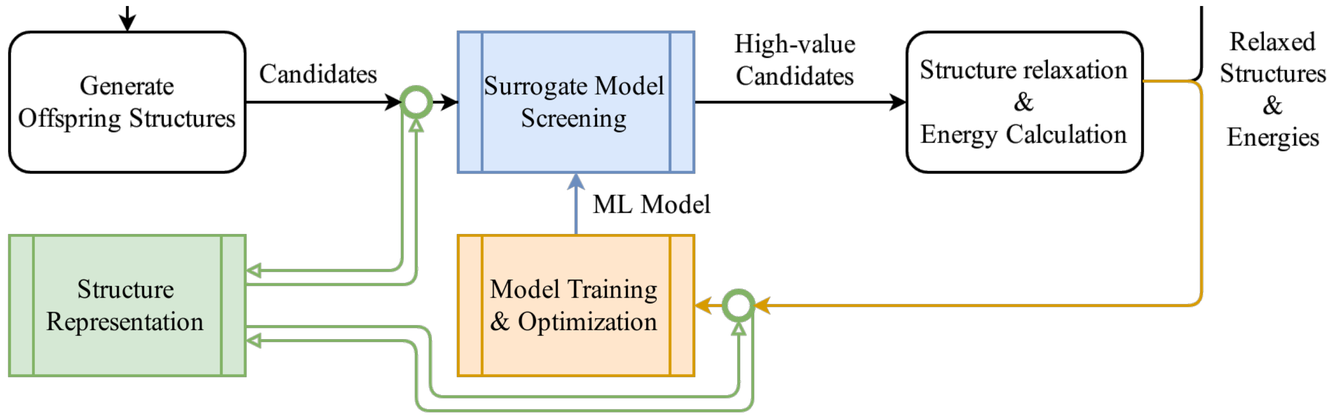


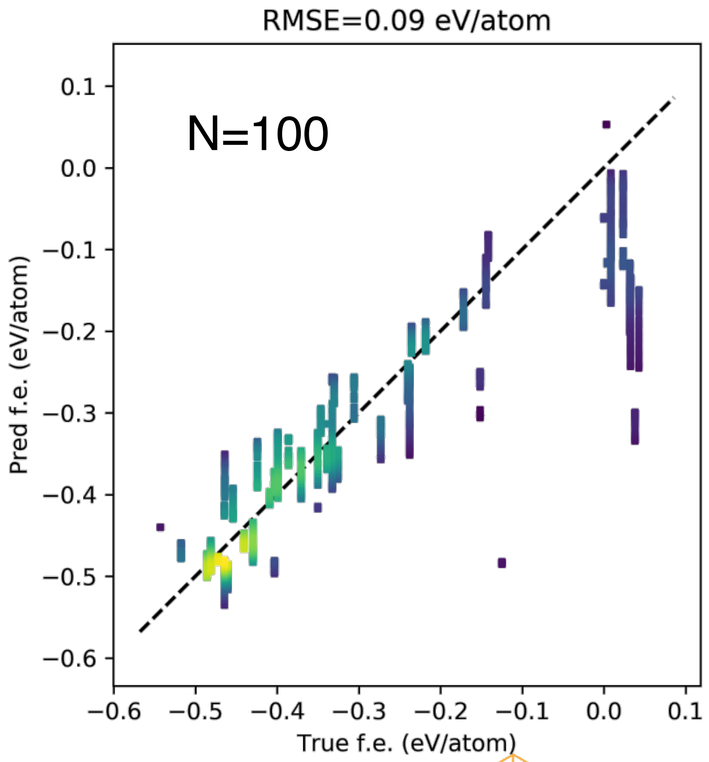
Powered by
MPInterfaces & materialsweb

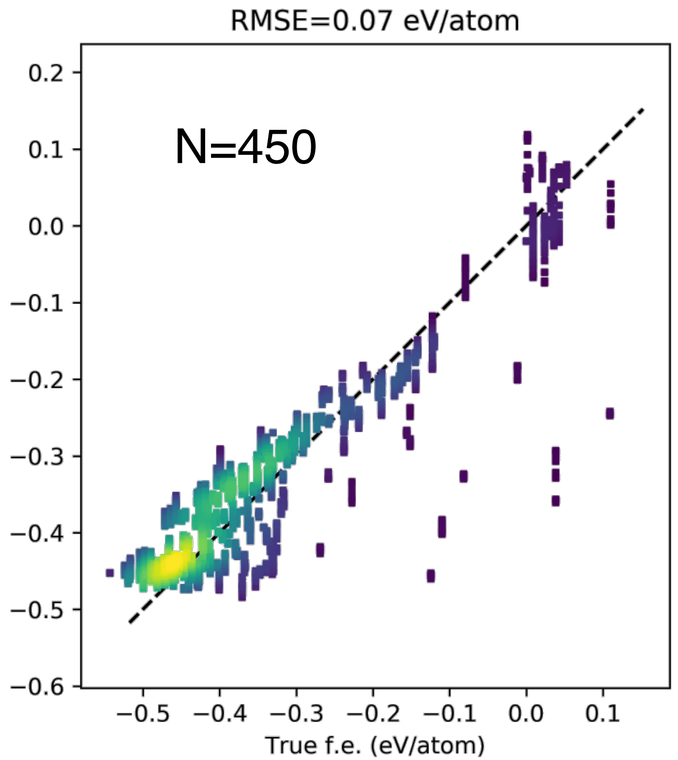
rhennig@ufl.edu
http://hennig.mse.ufl.edu

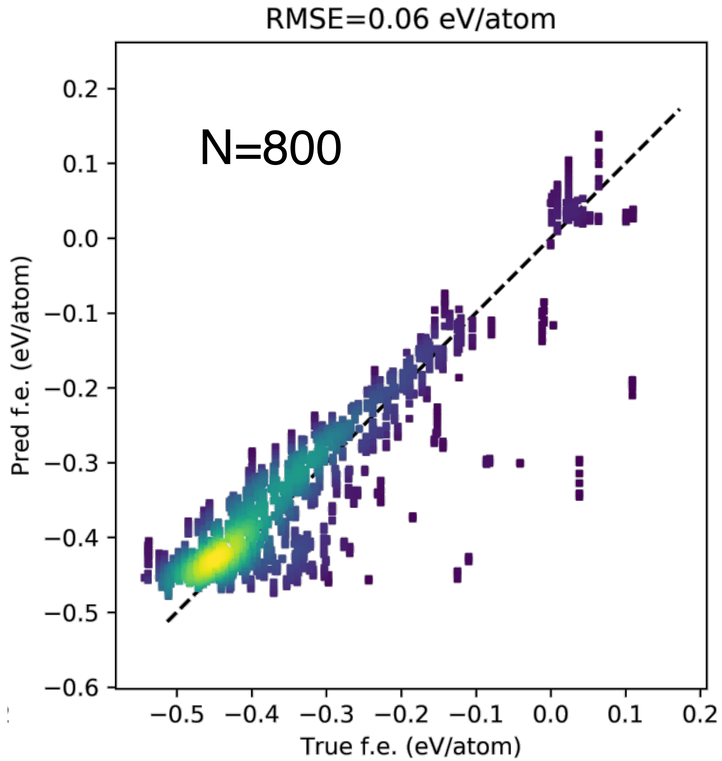
On-the-Fly Machine Learning

- Retrain the ensemble periodically as new relaxations are completed
- Recompute ensemble R^2 and select acceptance probability curve





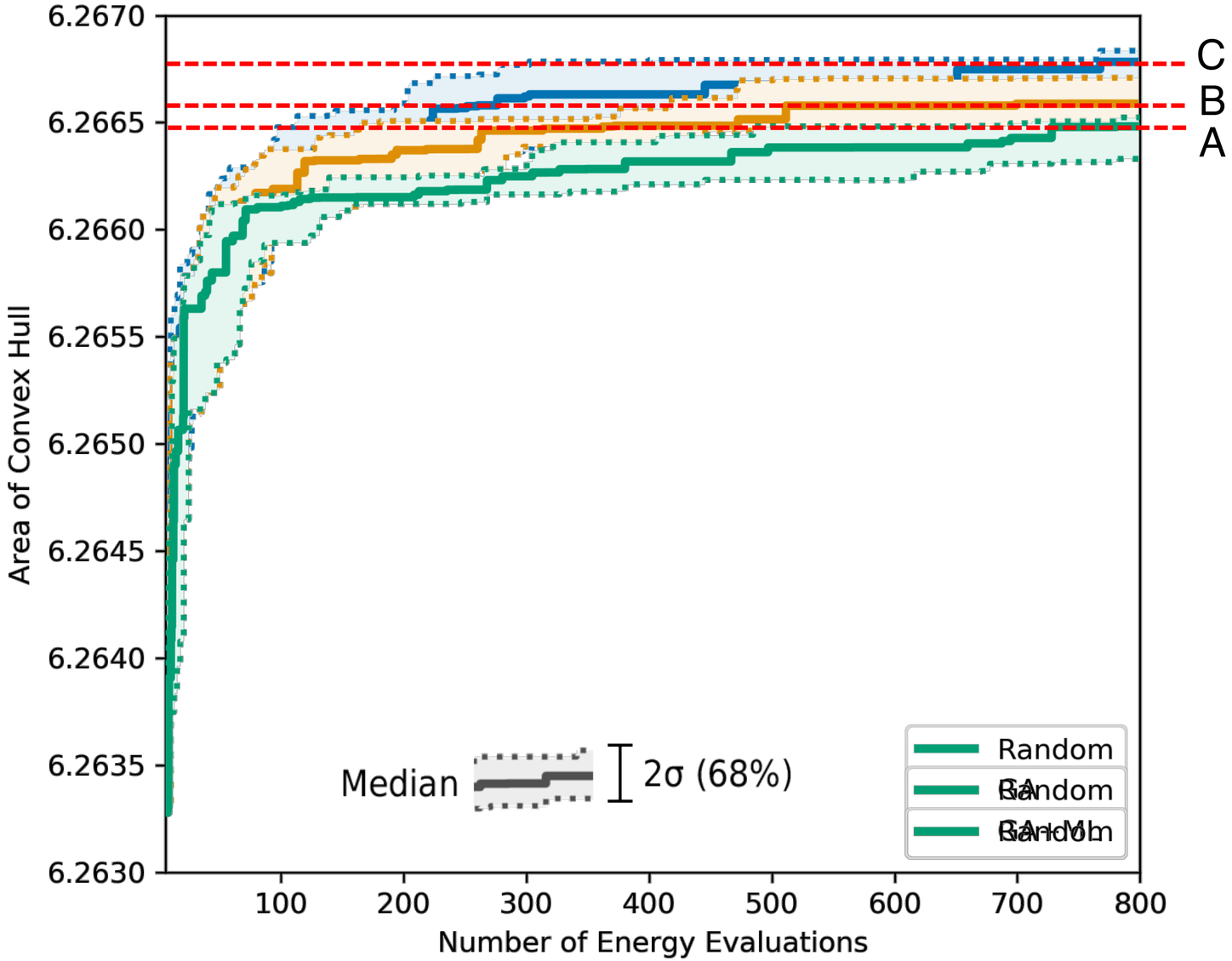




Powered by
MPInterfaces & materialsweb

rhennig@ufl.edu
http://hennig.mse.ufl.edu

Application #1: Zr-Al (EAM)

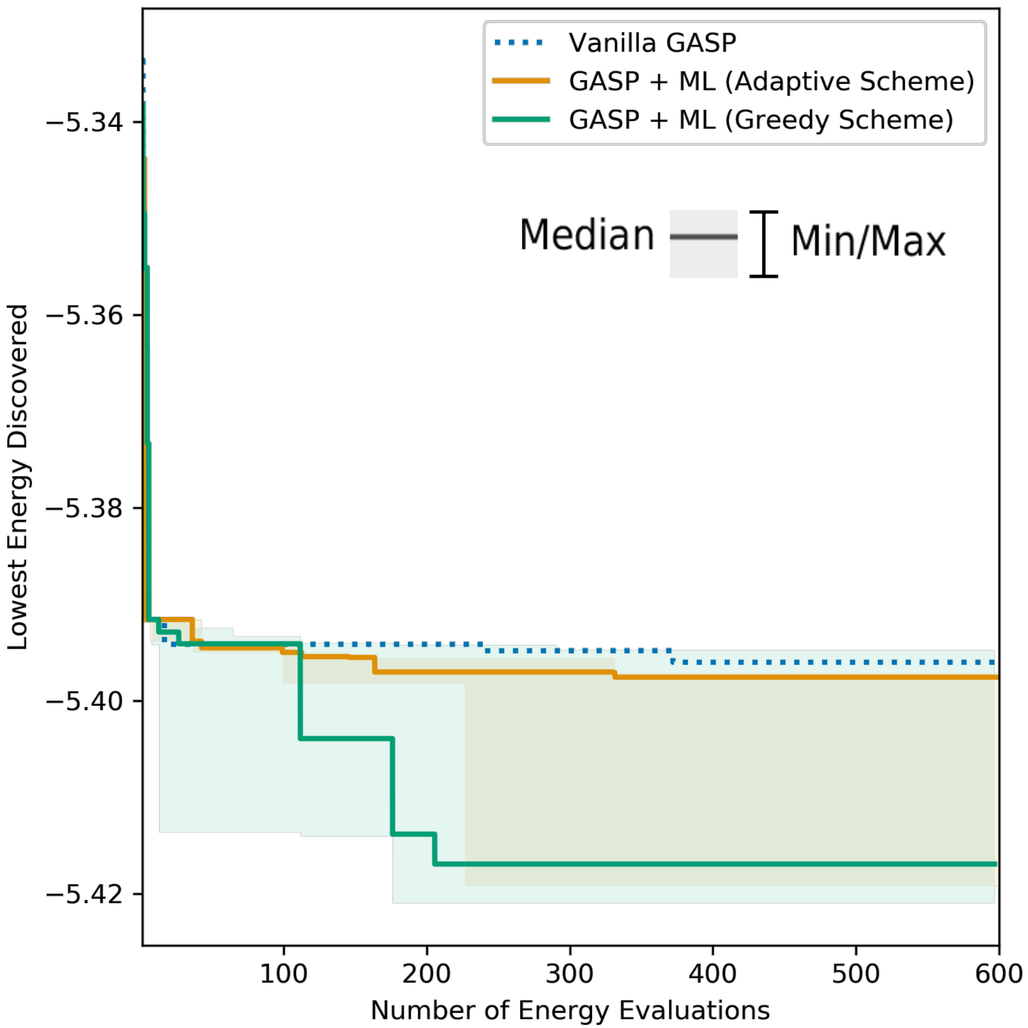


	Random Search	GASP	GASP + ML
A	781	368	217
B	-	743	261
C	-	-	767

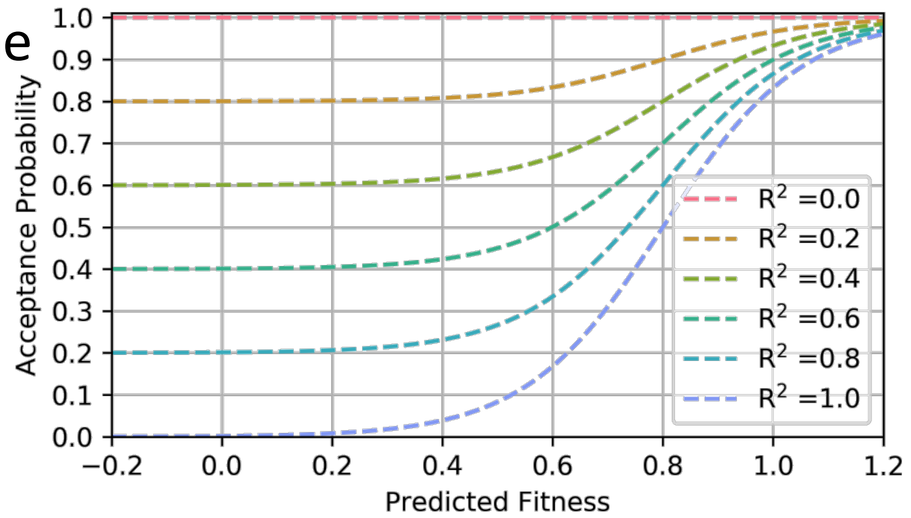
Powered by
MPInterfaces & materialsweb

rhennig@ufl.edu
http://hennig.mse.ufl.edu

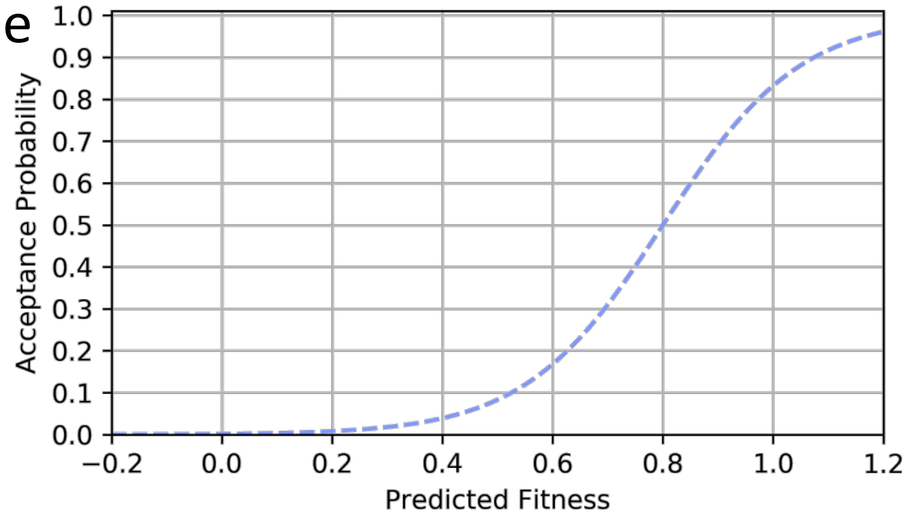
Application #2: In_2O_3 Monolayers (DFT)



Adaptive Scheme



Greedy Scheme



rhennig@ufl.edu
http://hennig.mse.ufl.edu

Future Work: Improving Model Performance

Other representations	Symmetry Functions SOAP Local MBTR Schnet	Behler 2015 Bartok 2014 Rinke 2019 Schutte 2018
Other Models	Support Vector Machines Artificial Neural Networks	Honrao 2018 Roitberg 2018
Learning with Forces	sGDML Truncated Taylor Series	Tkatchenko 2019 Artrith 2020
Uniqueness of descriptions	Completeness	Csanyi 2020

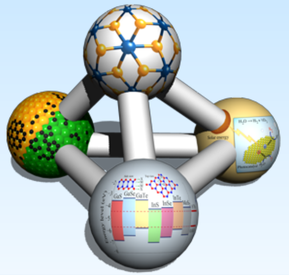
Important: Full on-the-fly ML cycle needs to be faster than DFT

- Representation
- Optimization of model hyperparameters
- Ensemble training
- Prediction

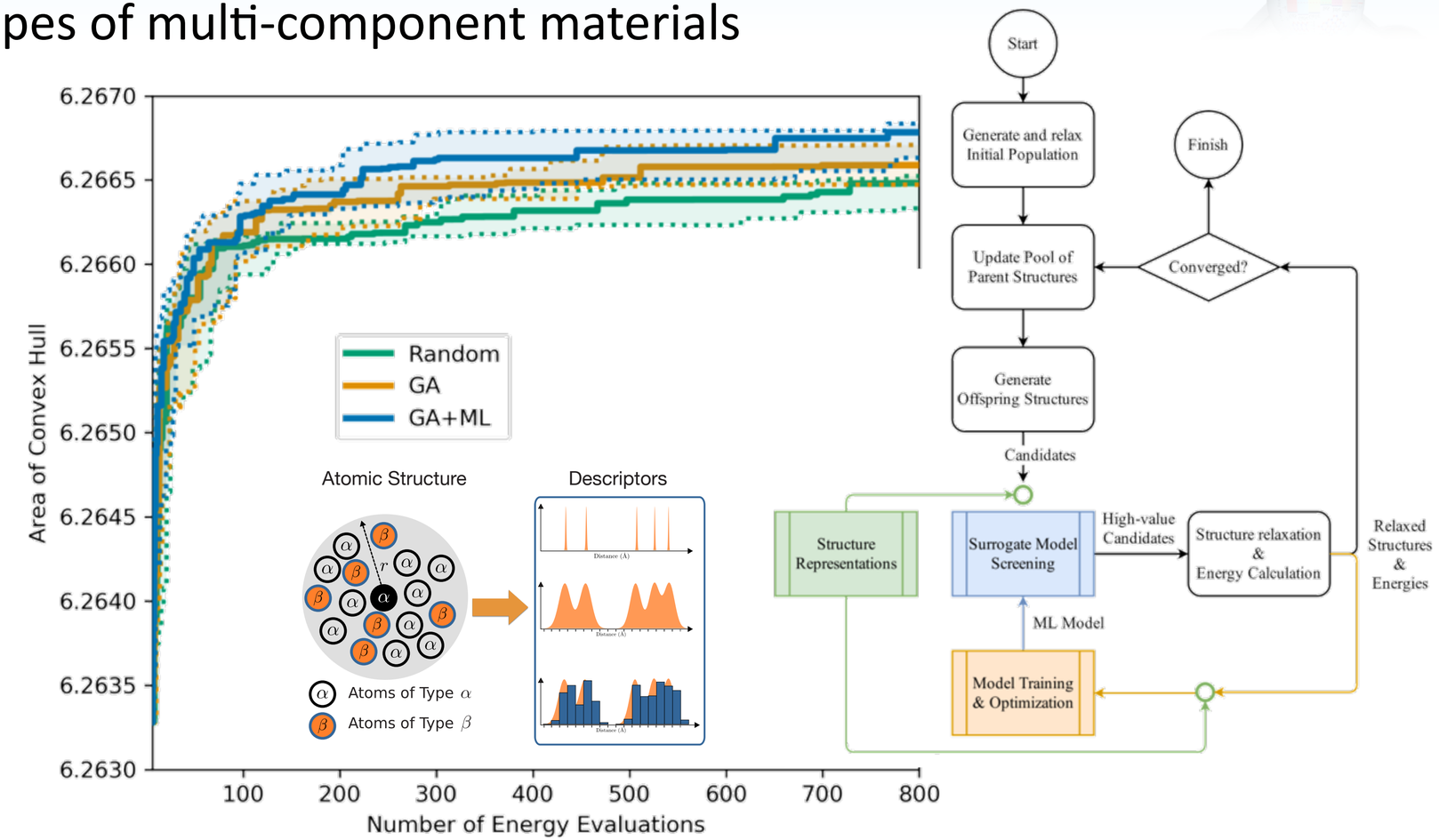
rhennig@ufl.edu
http://hennig.mse.ufl.edu

Machine-learning for the Exploration of Energy Landscapes

Stephen Xie, Shreyas Honrao, Anne Marie Tan, Halee Lester, and Richard G. Hennig
Materials Science and Engineering & Quantum Theory Project, University of Florida



- Machine learning of energy landscapes of multi-component materials
- Radial and angular distribution functions as descriptors
- On-the-fly machine learning of the relaxed energy landscape accelerates evolutionary structure searches
- Choice of stochastic screening criteria affects convergence behavior of GASP+ML



Honrao, Xie, Hennig, J. Appl. Phys., in print (2020)
Honrao, Anthonio, Ramanathan, Gabriel, Hennig, Comp. Mater. Sci. 158, 414 (2019)
https://github.com/henniggroup/gasp-python

Powered by
MPInterfaces & materialsweb

Atomistic Simulations for Industrial Needs
August 5-7, 2020 • NIST • Gaithersburg, MD