Carl von Ossietzky
Universität
Oldenburg


# Machine Learning-Assisted Design of Cesium-Based Photocathodes from First Principles

Holger-Dietrich Saßnick[1] and Caterina Cocchi[1,2]

[1]Carl von Ossietzky Universität Oldenburg, Physics Department, 26129 Oldenburg
[2]Humboldt-Universität zu Berlin, Physics Department and IRIS Adlershof, 12489 Berlin

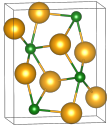
18th September, 2024

Carl von Ossietzky Universität Oldenburg

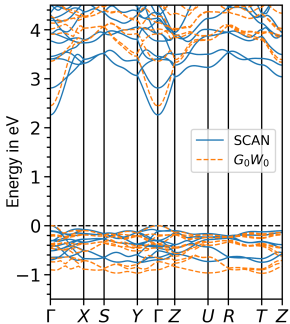
*Ab initio* methods like density functional theory (DFT) allow the calculation of material's properties without empirical parameters.

*Ab initio* methods like density functional theory (DFT) allow the calculation of material's properties without empirical parameters.

## Electronic Properties



$Cs_2Te$
*Pnma* [62]

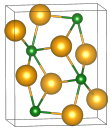



Cocchi and Saßnick, *Micromachines* **2021**, 12(9), 1002;
doi: 10.3390/mi12091002

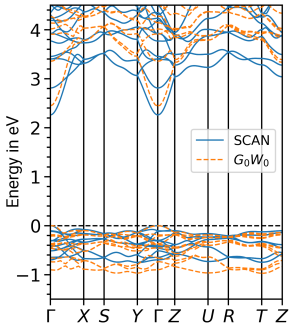
Carl von Ossietzky Universität Oldenburg

*Ab initio* methods like density functional theory (DFT) allow the calculation of material's properties without empirical parameters.

## Electronic Properties

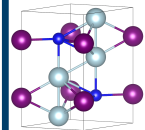


$Cs_2Te$
*Pnma* [62]

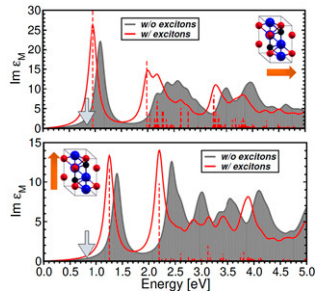



Cocchi and Saßnick, *Micromachines* **2021**, 12(9), 1002;
doi: 10.3390/mi12091002

## Optical Properties



$NaK_2Sb$
$P6_3/mmc$ [194]

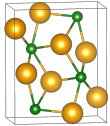



Amador, Saßnick and Cocchi, *J. Phys.: Condens. Matter* **2021**, 33, 365502;
doi: 10.1088/1361-648X/ac0e70

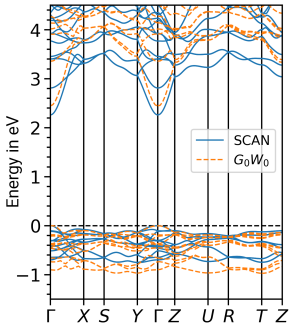
Carl von Ossietzky Universität Oldenburg

*Ab initio* methods like density functional theory (DFT) allow the calculation of material's properties without empirical parameters.

## Electronic Properties

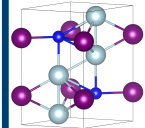


$Cs_2Te$
*Pnma* [62]

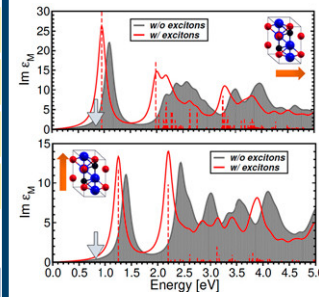



Cocchi and Saßnick, *Micromachines* **2021**, 12(9), 1002; doi: 10.3390/mi12091002

## Optical Properties

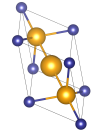


$NaK_2Sb$
$P6_3/mmc$ [194]

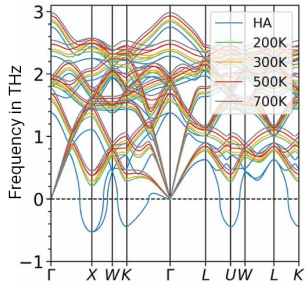



Amador, Saßnick and Cocchi, *J. Phys.: Condens. Matter* **2021**, 33, 365502; doi: 10.1088/1361-648X/ac0e70

## Vibrational Properties

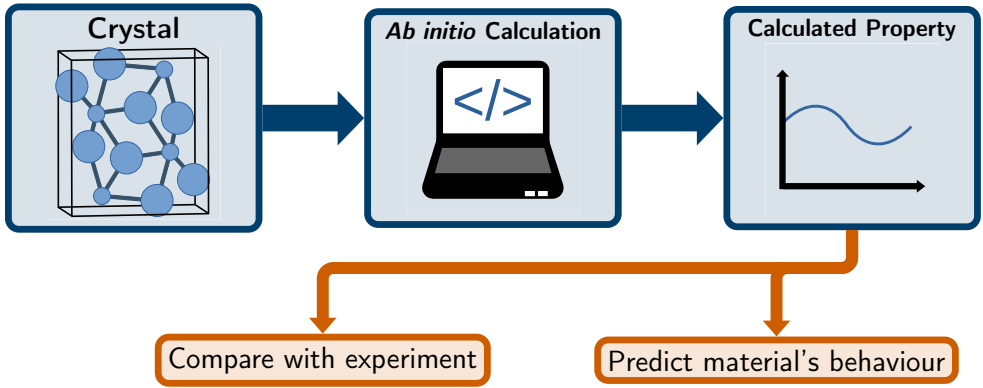


$Cs_3Sb$
$Fm\bar{3}m$ [225]

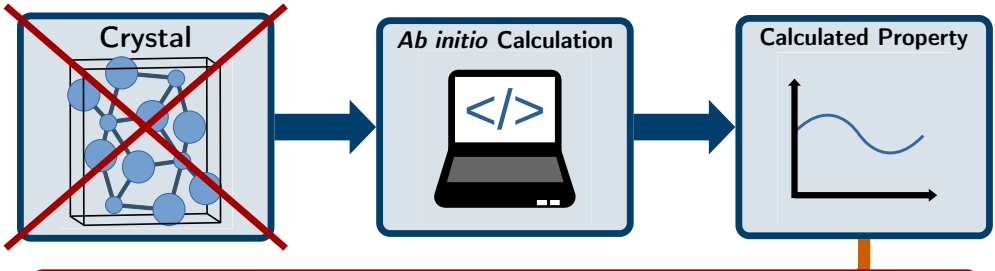



Santana-Andreo, Saßnick and Cocchi, *J. Phys. Mater.* **2024**, 7, 035004; doi: 10.1088/2515-7639/ad510b

# Structural information about the material is a required input.



# Structural information about the material is a required input.




- In some cases the atomistic structure of the studied material might not be known or is very complex.
- Can we find representative crystal structures without any experimental input?

# Categories of Crystal Structure Prediction Methods

## Methods with fixed stoichiometry:

- Minima hopping.

  Goedecker, *Modern Methods of Crystal Structure Prediction* **2010**, John Wiley & Sons, Ltd. Chap. 6, 131–145; doi: 10.1002/9783527632831.ch6

- Bayesian optimization.

  Yamashita et al., *Phys. Rev. Mater.* **2018**, 2(1), 013803; doi: 10.1103/PhysRevMaterials.2.013803

## Methods without stoichiometry constraints:

- Random structure search.

  Harper et al. *Johnson Matthey Technology Review* **2020**, 64.2, 103–118; doi: 10.1595/205651320X15742491027978

- Genetic algorithms.

  Lyakhov et al. *Modern Methods of Crystal Structure Prediction* **2010**, John Wiley & Sons, Ltd. Chap. 7, 147–180; doi: 10.1002/9783527632831.ch7

# Categories of Crystal Structure Prediction Methods

## Methods with fixed stoichiometry:

- Minima hopping.

  Goedecker, *Modern Methods of Crystal Structure Prediction* **2010**, John Wiley & Sons, Ltd. Chap. 6, 131–145; doi: 10.1002/9783527632831.ch6

- Bayesian optimization.

  Yamashita et al., *Phys. Rev. Mater.* **2018**, 2(1), 013803; doi: 10.1103/PhysRevMaterials.2.013803

## Methods without stoichiometry constraints:

- Random structure search.

  Harper et al. *Johnson Matthey Technology Review* **2020**, 64.2, 103–118; doi: 10.1595/205651320X15742491027978

- Genetic algorithms.

  Lyakhov et al. *Modern Methods of Crystal Structure Prediction* **2010**, John Wiley & Sons, Ltd. Chap. 7, 147–180; doi: 10.1002/9783527632831.ch7

All methods share large computational costs and/or are inflexible in its application.

# Categories of Crystal Structure Prediction Methods

## Methods with fixed stoichiometry:

- Minima hopping.

  Goedecker, *Modern Methods of Crystal Structure Prediction* **2010**, John Wiley & Sons, Ltd. Chap. 6, 131–145; doi: 10.1002/9783527632831.ch6

- Bayesian optimization.

  Yamashita et al., *Phys. Rev. Mater.* **2018**, 2(1), 013803; doi: 10.1103/PhysRevMaterials.2.013803

## Methods without stoichiometry constraints:

- Random structure search.

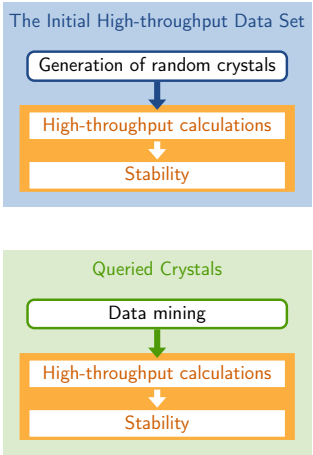
  Harper et al. *Johnson Matthey Technology Review* **2020**, 64.2, 103–118; doi: 10.1595/205651320X15742491027978

- Genetic algorithms.

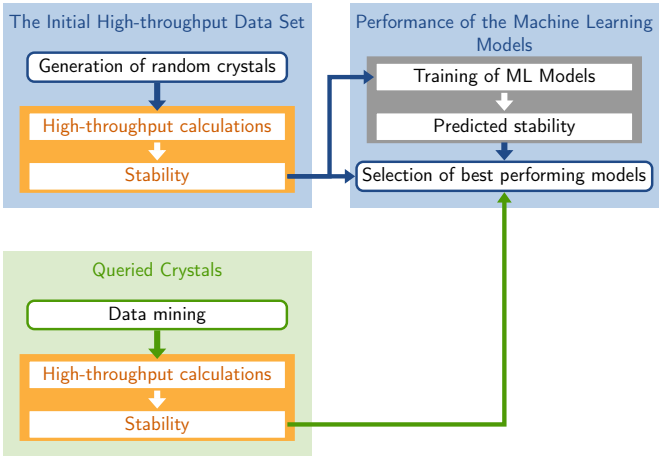
  Lyakhov et al. *Modern Methods of Crystal Structure Prediction* **2010**, John Wiley & Sons, Ltd. Chap. 7, 147–180; doi: 10.1002/9783527632831.ch7

This approach combines random structure search with machine learning to alleviate these issues.

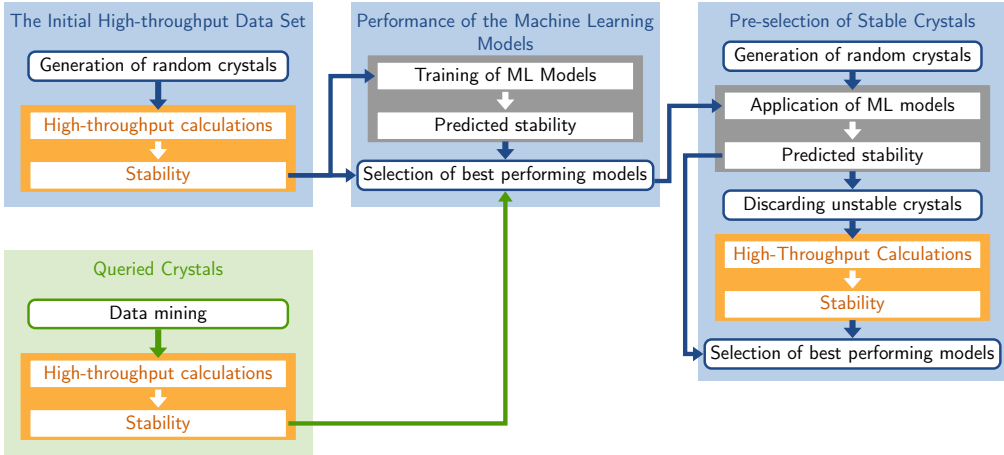
# Overview of the Crystal Structure Prediction Approach



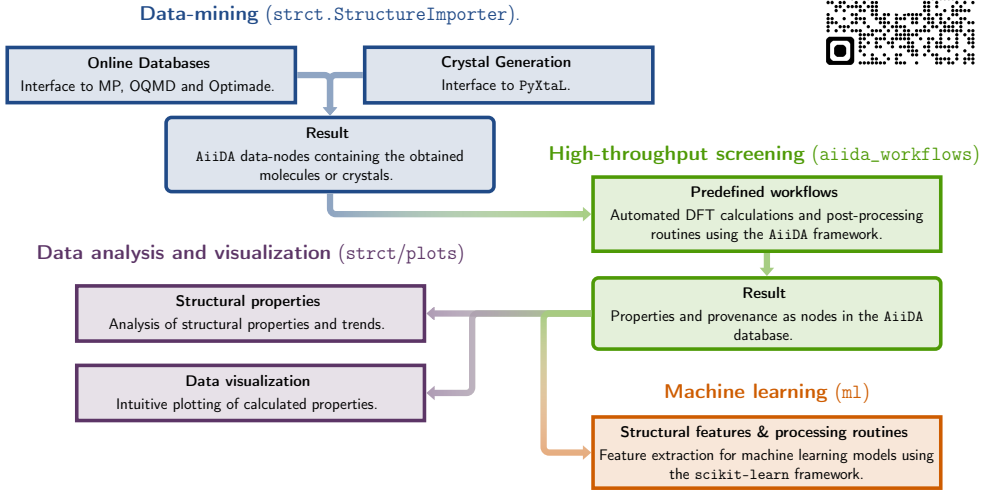
# Overview of the Crystal Structure Prediction Approach



# Overview of the Crystal Structure Prediction Approach

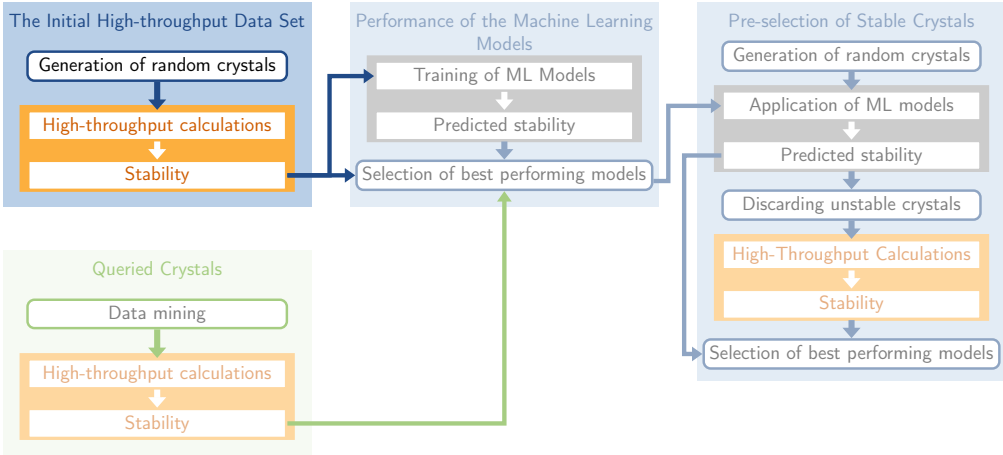


# Implementation in aim$^2$dat




https://aim2dat.github.io/aim2dat/

# Overview of the Crystal Structure Prediction Approach



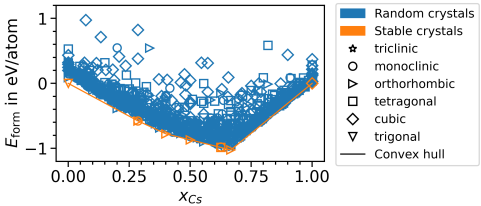
# The Initial High-Throughput Data Set

- 1073 crystal structures successfully converged.
- Formation energy:

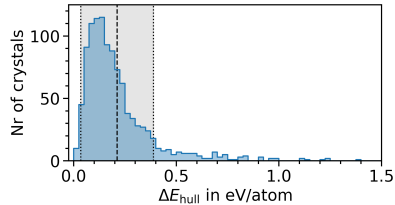
$$E_{\mathrm{form}}(\mathrm{Cs}_x\mathrm{Te}_{1-x}) = E(\mathrm{Cs}_x\mathrm{Te}_{1-x}) - [xE(\mathrm{Cs}) + (1-x)E(\mathrm{Te})]$$

- The stability ($\Delta E_{\mathrm{hull}}$) of each crystal is given by the distance to the convex hull.
- Average stability: 0.212 eV/atom.

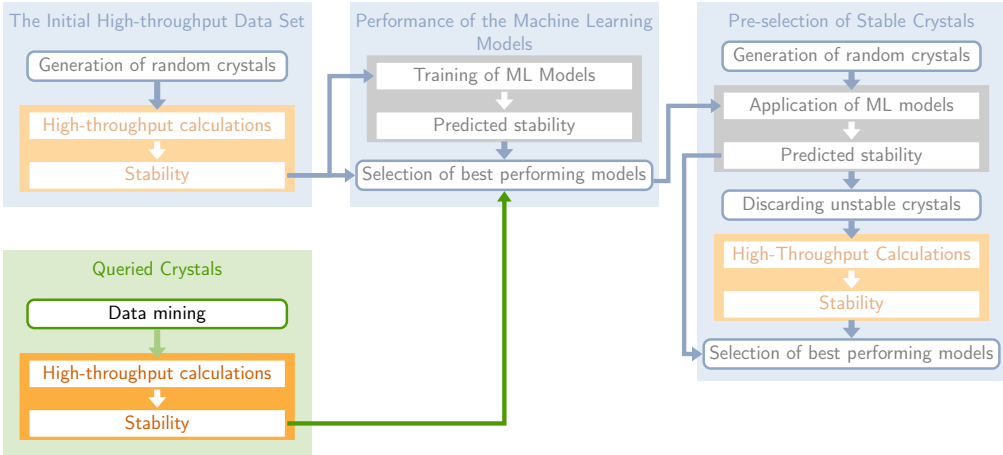
Formation energy:




Stability distribution:



# Overview of the Crystal Structure Prediction Approach

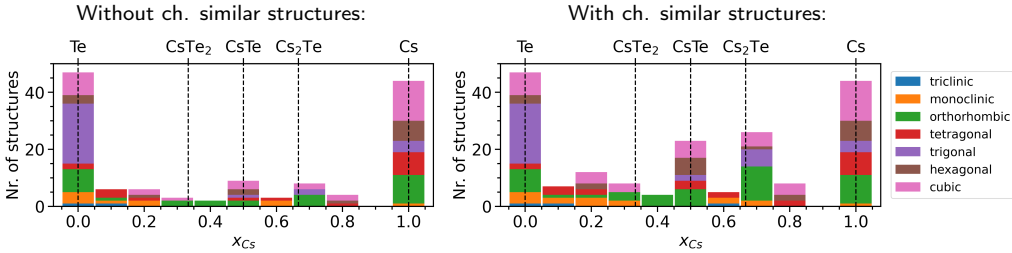


# Data Mining of Cs-Te Bulk Crystals

**Materials project (MP)**
Jain, Ong, *et al.*, *APL Materials* **2013**, 1, 011002; doi: 10.1063/1.4812323

**Open quantum materials database (oqmd)**
Saal, Kirklin, *et al.*, *JOM* **2013**, 65, 1501–1509; doi: 10.1007/s11837-013-0755-4

- Mined from online databases and enriched by substituting chemically similar elements.
- Distribution of the queried crystal structures:




Saßnick and Cocchi, *J. Chem. Phys.* **2022**, 156, 104108; doi: 10.1063/5.0082710

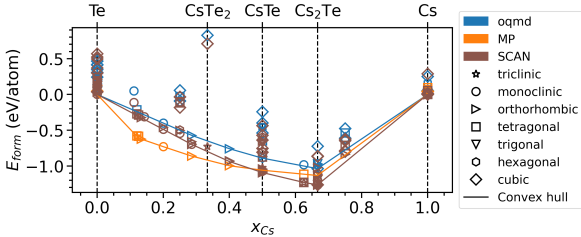
# Distribution and Characteristics

- Formation energy:

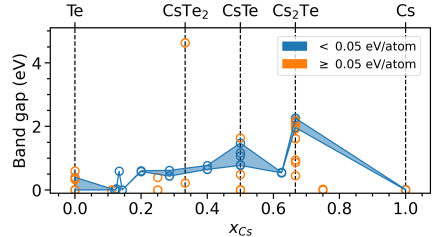
$$E_{\mathrm{form}}(\mathrm{Cs}_x\mathrm{Te}_{1-x}) = E(\mathrm{Cs}_x\mathrm{Te}_{1-x}) - [xE(\mathrm{Cs}) + (1-x)E(\mathrm{Te})]$$

- The stability ($\Delta E_{\mathrm{hull}}$) of each crystal is given by the distance to the convex hull.
- Average stability: 0.136 eV/atom.
- The electronic band gap correlates with the composition for stable compounds ($\Delta E_{\mathrm{hull}} < 0.05$ eV/atom).

Formation energy:

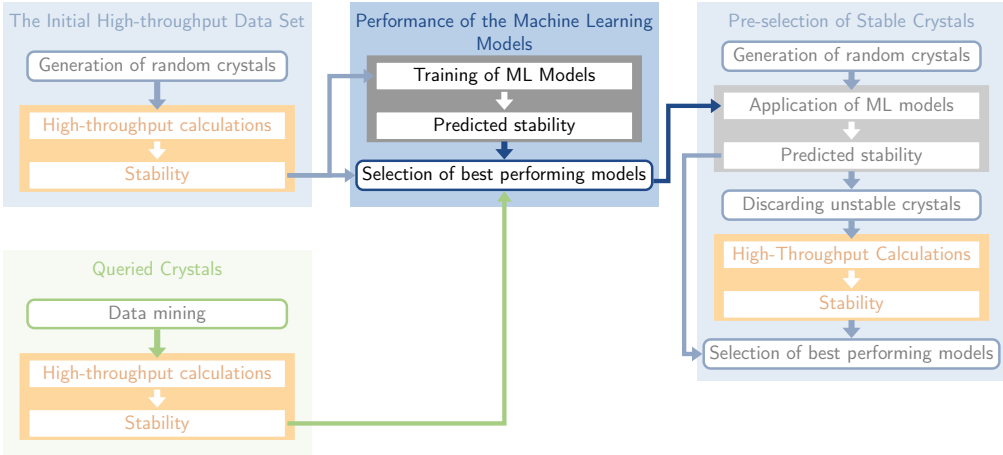



Band gap:

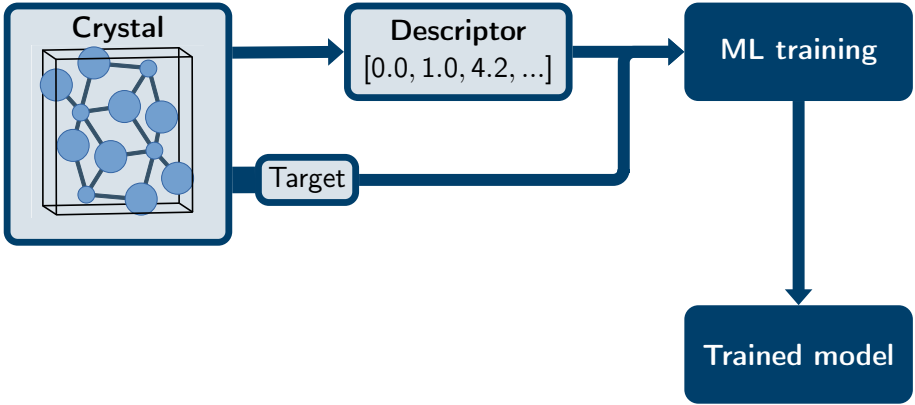



Saßnick and Cocchi, *J. Chem. Phys.* **2022**, 156, 104108; doi: 10.1063/5.0082710

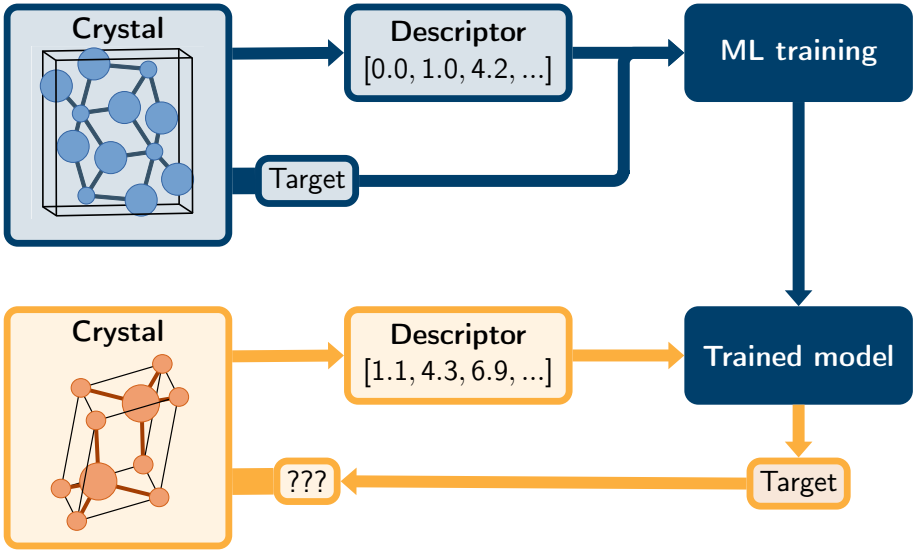
# Overview of the Crystal Structure Prediction Approach



# Machine Learning Structure-Property Relationships



# Machine Learning Structure-Property Relationships



Carl von Ossietzky Universität Oldenburg

# Used Descriptors and ML Algorithms

# Used Descriptors and ML Algorithms

- *Coord.:* Statistical quantities of the coordination environment of each atomic site and composition.
  - *ET:* Extremely randomized trees
  - *KRR:* Kernel-ridge regression

# Used Descriptors and ML Algorithms

- *Coord.:* Statistical quantities of the coordination environment of each atomic site and composition.
  - *ET:* Extremely randomized trees
  - *KRR:* Kernel-ridge regression
- *F-Fprint:* Averaged radial distribution function for each element-pair.
  - *KRR:* Kernel-ridge regression

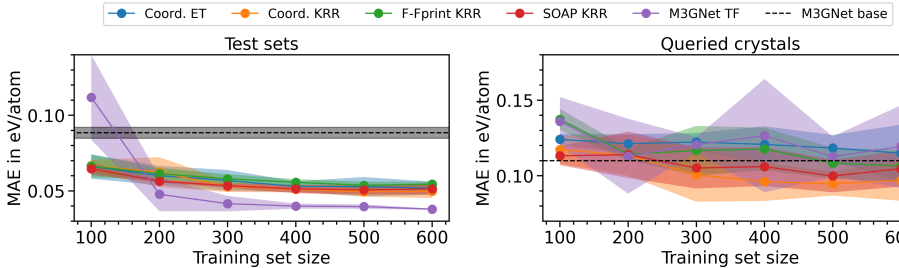
# Used Descriptors and ML Algorithms

- *Coord.:* Statistical quantities of the coordination environment of each atomic site and composition.
  - *ET:* Extremely randomized trees
  - *KRR:* Kernel-ridge regression
- *F-Fprint:* Averaged radial distribution function for each element-pair.
  - *KRR:* Kernel-ridge regression
- *SOAP:* Power-spectrum of the overlap matrix of a set of atom-centered Gaussian functions.
  - *KRR:* Kernel-ridge regression

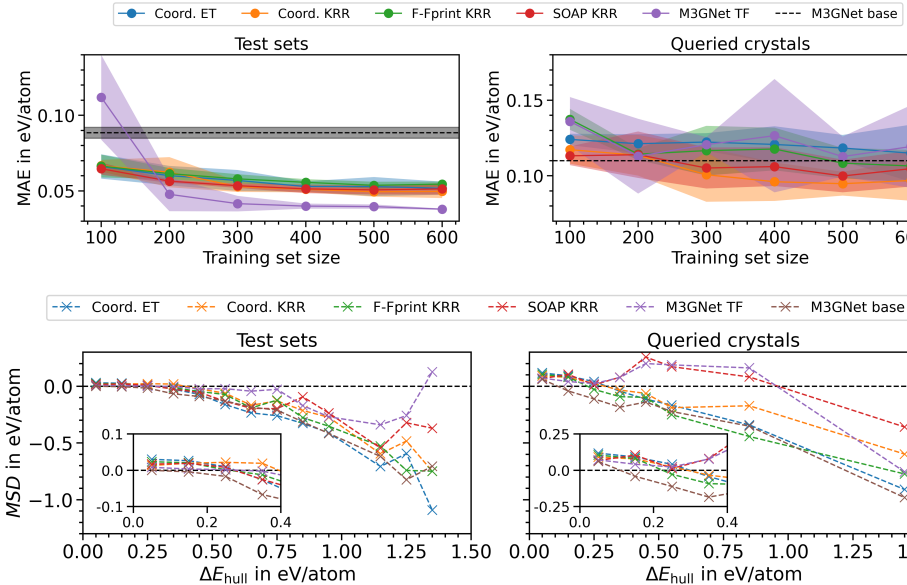
# Used Descriptors and ML Algorithms

- *Coord.:* Statistical quantities of the coordination environment of each atomic site and composition.
  - *ET:* Extremely randomized trees
  - *KRR:* Kernel-ridge regression
- *F-Fprint:* Averaged radial distribution function for each element-pair.
  - *KRR:* Kernel-ridge regression
- *SOAP:* Power-spectrum of the overlap matrix of a set of atom-centered Gaussian functions.
  - *KRR:* Kernel-ridge regression
- *M3GNet:* Graph representation with atomic sites as nodes and chemical bonds as edges.
  - Graph neural network

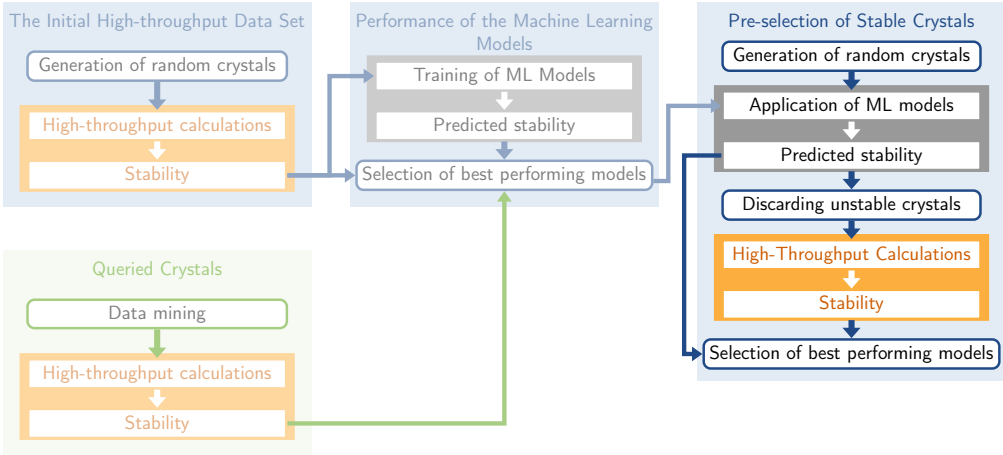
# Performance of the Machine Learning Models



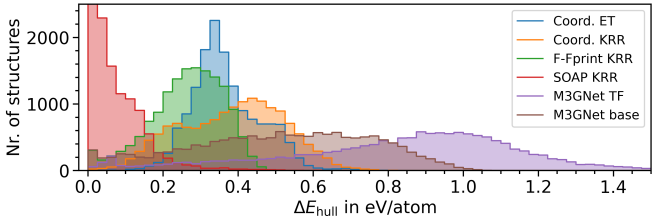
# Performance of the Machine Learning Models



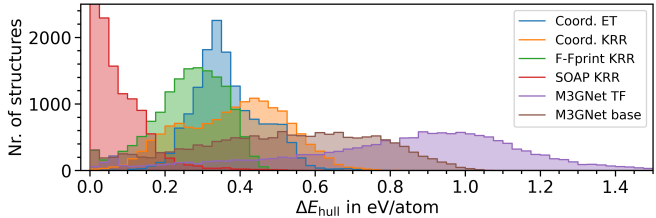
# Overview of the Crystal Structure Prediction Approach



# Pre-selection of Stable Crystals

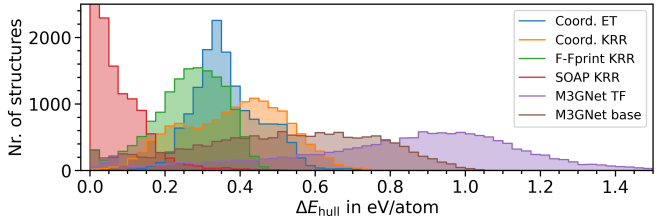


# Pre-selection of Stable Crystals



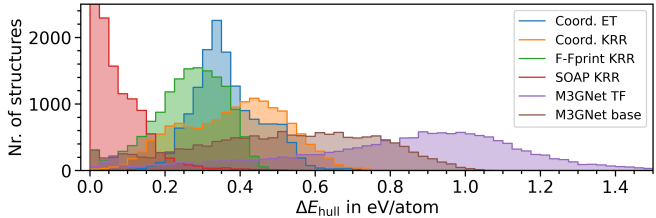

| | $\Delta E_{\text{Hull}}$ in eV/at. | | MAE in eV/at. | S.R. in % | Stable phases in % |
|---|---|---|---|---|---|
| | Mean | SD | | | Te/mixed/Cs |
| *Initial pool* | 0.212 | 0.177 | – | 65.8 | 0.0/2.1/1.3 |
| Coord. ET | 0.178 | 0.109 | **0.081** | **90.0** | **0.0/6.0/5.0** |
| Coord. KRR | 0.164 | 0.098 | 0.101 | 80.0 | 0.0/5.0/5.0 |
| F-Fprint KRR | 0.378 | 0.308 | 0.344 | 34.0 | 0.0/0.0/0.0 |
| SOAP KRR | **0.158** | 0.124 | 0.157 | 63.0 | 0.0/0.0/3.0 |
| M3GNet TF | 0.195 | 0.106 | 0.110 | **90.0** | 1.0/5.0/1.0 |
| M3GNet base | 0.193 | 0.126 | 0.156 | 77.0 | 0.0/4.0/3.0 |

# Pre-selection of Stable Crystals



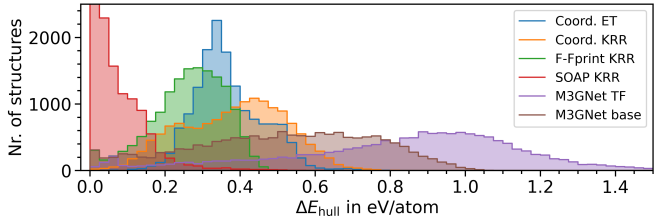

| | $\Delta E_{\text{Hull}}$ in eV/at. | | MAE in eV/at. | S.R. in % | Stable phases in % |
|---|---|---|---|---|---|
| | Mean | SD | | | Te/mixed/Cs |
| *Initial pool* | 0.212 | 0.177 | – | 65.8 | 0.0/2.1/1.3 |
| Coord. ET | 0.178 | 0.109 | **0.081** | **90.0** | **0.0/6.0/5.0** |
| Coord. KRR | 0.164 | 0.098 | 0.101 | 80.0 | 0.0/5.0/5.0 |
| F-Fprint KRR | 0.378 | 0.308 | 0.344 | 34.0 | 0.0/0.0/0.0 |
| SOAP KRR | **0.158** | 0.124 | 0.157 | 63.0 | 0.0/0.0/3.0 |
| M3GNet TF | 0.195 | 0.106 | 0.110 | **90.0** | 1.0/5.0/1.0 |
| M3GNet base | 0.193 | 0.126 | 0.156 | 77.0 | 0.0/4.0/3.0 |

# Pre-selection of Stable Crystals




| | $\Delta E_{\text{Hull}}$ in eV/at. | | MAE in eV/at. | S.R. in % | Stable phases in % |
|---|---|---|---|---|---|
| | Mean | SD | | | Te/mixed/Cs |
| *Initial pool* | 0.212 | 0.177 | – | 65.8 | 0.0/2.1/1.3 |
| Coord. ET | 0.178 | 0.109 | **0.081** | **90.0** | **0.0/6.0/5.0** |
| Coord. KRR | 0.164 | 0.098 | 0.101 | 80.0 | 0.0/5.0/5.0 |
| F-Fprint KRR | 0.378 | 0.308 | 0.344 | 34.0 | 0.0/0.0/0.0 |
| SOAP KRR | **0.158** | 0.124 | 0.157 | 63.0 | 0.0/0.0/3.0 |
| M3GNet TF | 0.195 | 0.106 | 0.110 | **90.0** | 1.0/5.0/1.0 |
| M3GNet base | 0.193 | 0.126 | 0.156 | 77.0 | 0.0/4.0/3.0 |

# Pre-selection of Stable Crystals




| | $\Delta E_{\text{Hull}}$ in eV/at. | | MAE in eV/at. | S.R. in % | Stable phases in % |
|---|---|---|---|---|---|
| | Mean | SD | | | Te/mixed/Cs |
| *Initial pool* | 0.212 | 0.177 | – | 65.8 | 0.0/2.1/1.3 |
| Coord. ET | 0.178 | 0.109 | **0.081** | **90.0** | **0.0/6.0/5.0** |
| Coord. KRR | 0.164 | 0.098 | 0.101 | 80.0 | 0.0/5.0/5.0 |
| F-Fprint KRR | 0.378 | 0.308 | 0.344 | 34.0 | 0.0/0.0/0.0 |
| SOAP KRR | **0.158** | 0.124 | 0.157 | 63.0 | 0.0/0.0/3.0 |
| M3GNet TF | 0.195 | 0.106 | 0.110 | **90.0** | 1.0/5.0/1.0 |
| M3GNet base | 0.193 | 0.126 | 0.156 | 77.0 | 0.0/4.0/3.0 |

# Conclusions

- The presented approach offers a straight-forward method to explore unknown materials.
  - The pre-selection step greatly improves over a random structure search.
  - The *Coord. ET* and the *M3GNet* models show the best performance.
- The implementation can be readily applied to any other material-system, e.g. ternary antimonides.
- The source code is published and freely available (https://aim2dat.github.io/aim2dat/).

## Experimental Collaborators

HZB Helmholtz Zentrum Berlin — Prof. Dr. Thorsten Kamps, and team.

## Sponsors

Bundesministerium für Bildung und Forschung

Niedersächsisches Ministerium für Wissenschaft und Kultur

DFG Project-id: 490940284

HLRN Project-id: nic00069, nic00076