FRITZ-HABER-INSTITUT
MAX-PLANCK-GESELLSCHAFT

NOMAD
NOVEL MATERIALS DISCOVERY

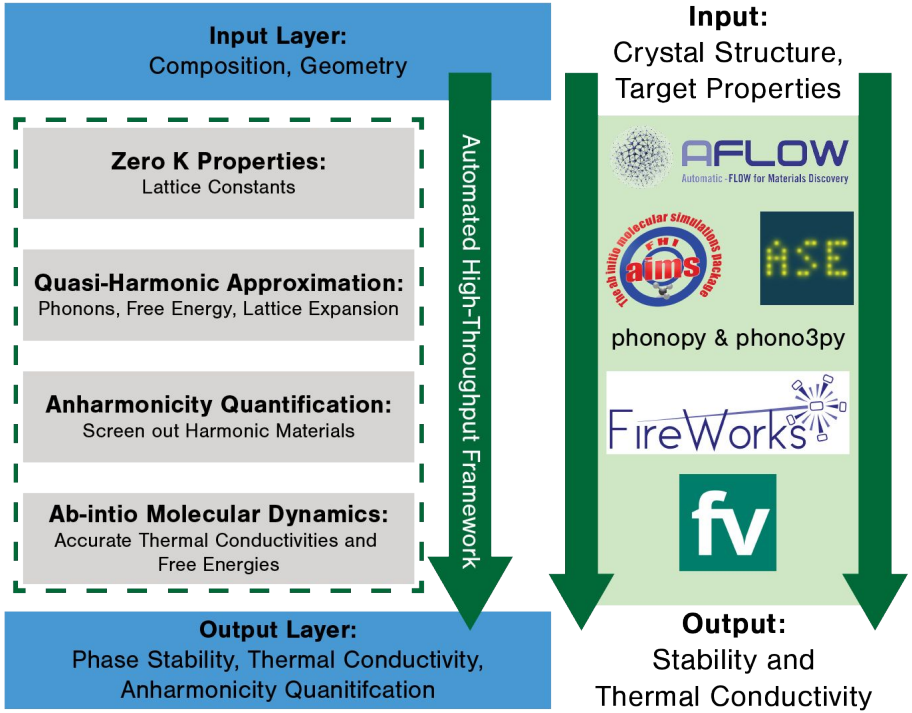
IRIS ADLERSHOF

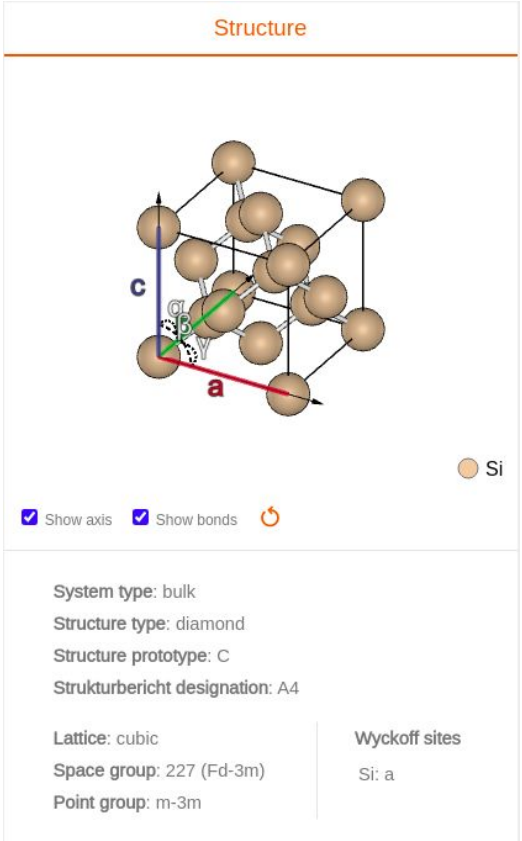
Alexander von Humboldt
Stiftung/Foundation

Accelerating High-Throughput Screening via Symbolic Regression: Uncovering the Relationship Between Thermal Conductivity and Anharmonicity

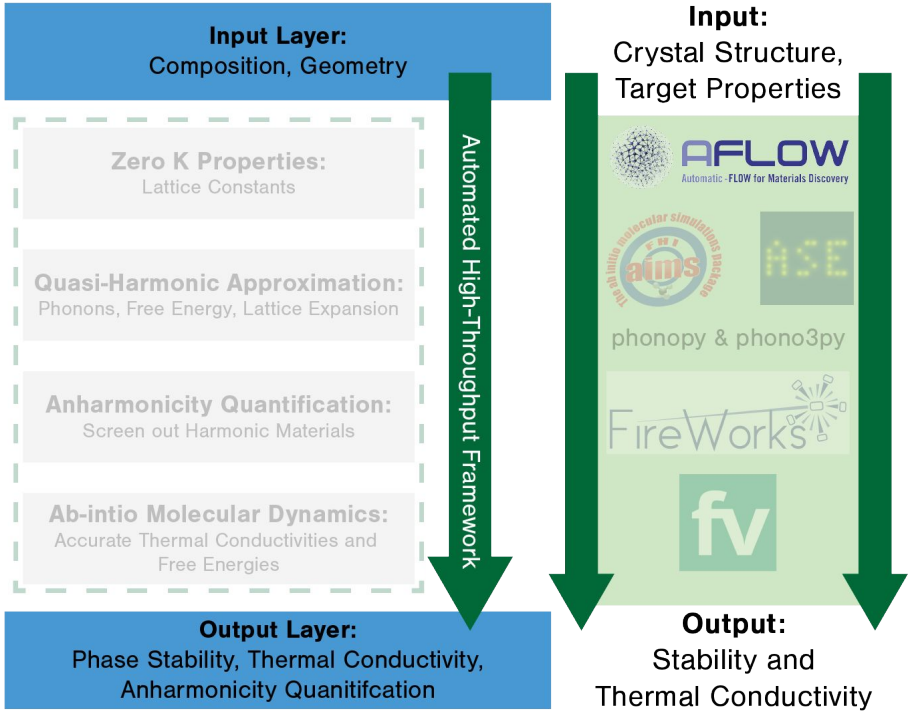
Thomas A. R. Purcell, Matthias Scheffler, Luca M. Ghiringhelli, Christian Carbogno
Nov 16, 2021

The HTC Framework





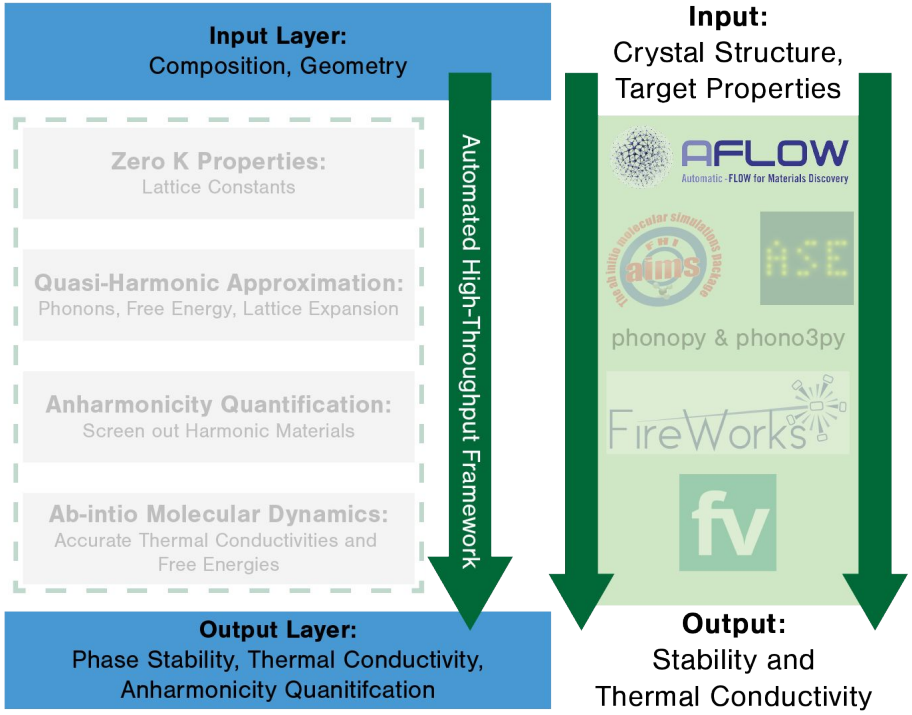
Use Aflow to Standardize Structures




```
u@c:~$ aflow --prototype < geometry.in
 AFLOW label    : A_cF8_227_a
 params         : a
 params values  : 5.406
```


Hicks, D.; et al..*npj Comput. Mater.* **7**, 30. (2021)

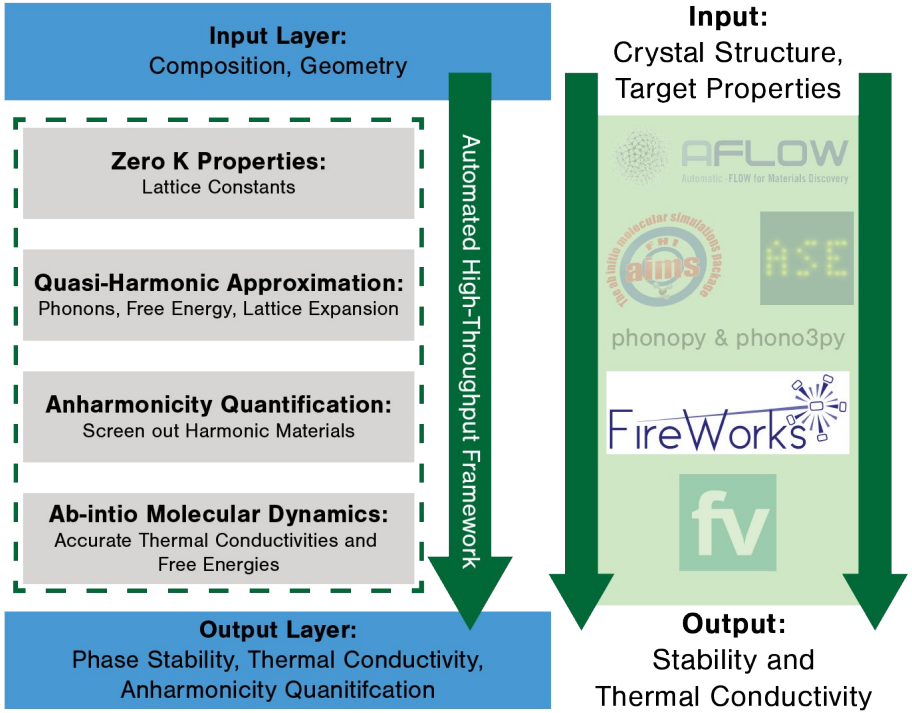
Generate Symmetrized Structures with Aflow

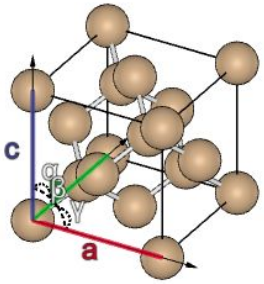



```
u@c:~$ aflow --prototype < geometry.in
 AFLOW label    : A_cF8_227_a
 params         : a
 params values  : 5.406
u@c:~$ aflow --proto=A_cF8_227_a:Si --params=5.406 --aims --add_equations
 # A/A_cF8_227_a.A params=5.406 SG=227 [ANRL doi: 10.1016/j.commats-
ci.2017.01.017 (part 1), doi: 10.1016/j.commatsci.2018.10.043 (part 2)]
 # AFLOW::AIMS BEGIN
 lattice_vector     0.00000000000000    2.70300000000000    2.70300000000000
 lattice_vector     2.70300000000000    0.00000000000000    2.70300000000000
 lattice_vector     2.70300000000000    2.70300000000000    0.00000000000000
 atom_frac     0.12500000000000     0.12500000000000     0.12500000000000  Si
 atom_frac     0.87500000000000     0.87500000000000     0.87500000000000  Si
 # format: symmetry_n_params [n n_lv n_fracpos]
 symmetry_n_params 1 1 0
 symmetry_params a
 symmetry_lv 0 , 0.5*a , 0.5*a
 symmetry_lv 0.5*a , 0 , 0.5*a
 symmetry_lv 0.5*a , 0.5*a , 0
 symmetry_frac 0.125 , 0.125 , 0.125
 symmetry_frac 0.875 , 0.875 , 0.875
 # AFLOW::AIMS END
```


Mehl, M.; et al. *Comput. Mater. Sci.* **136**, S1. (2017)
Hicks, D.; et al. *Comput. Mater. Sci.* **161**, S1. (2018)
Lenz, M.-O.; et al. *npj Comput. Mater.* **5**, 123. (2019)

Initiate the Workflow



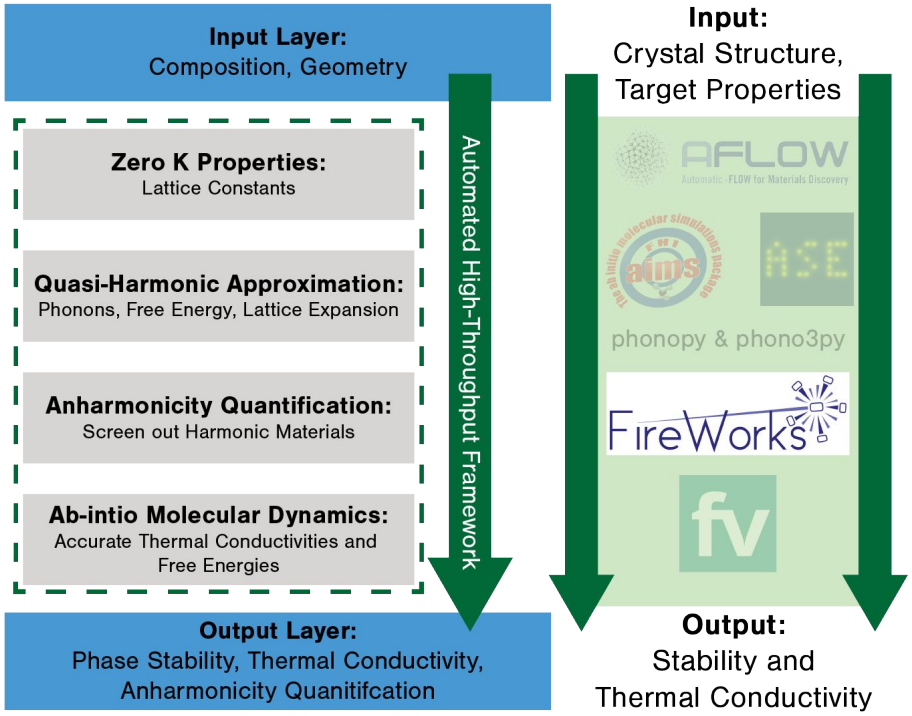


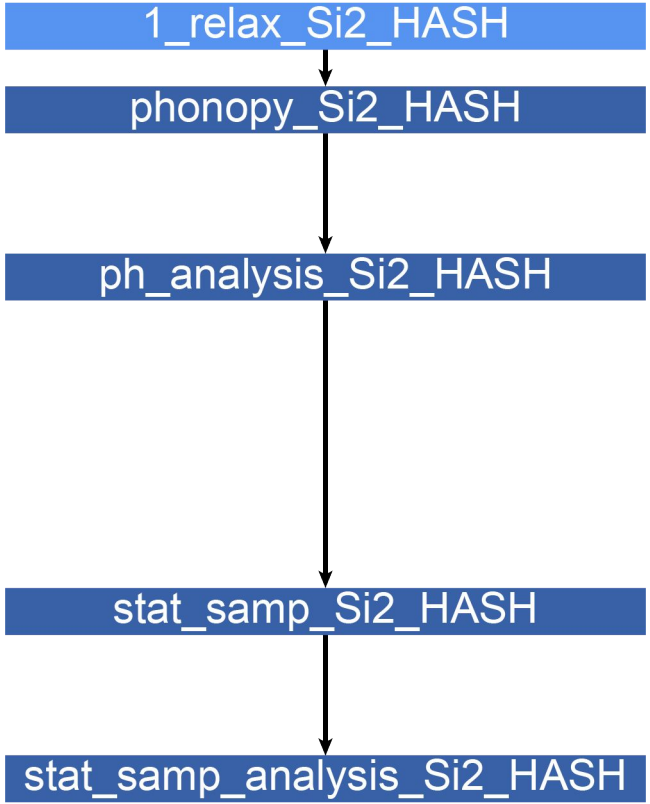

```
{
    "cell": [
        [0.000, 2.703, 2.703],
        [2.703, 0.000, 2.703],
        [2.703, 2.703, 0.000]
    ],
    "positions":[
        [0.67575, 0.67575, 0.67575],
        [4.73025, 4.73025, 4.73025]
     ],
    "symbols": ["Si", "Si"],
    "masses": [28.0855, 28.0855],
    "info": {},
    "constraints": [...]
}
```

```
[calculator]
...
[relaxation]
...
[phonopy]
...
[statistical_sampling]
...
```

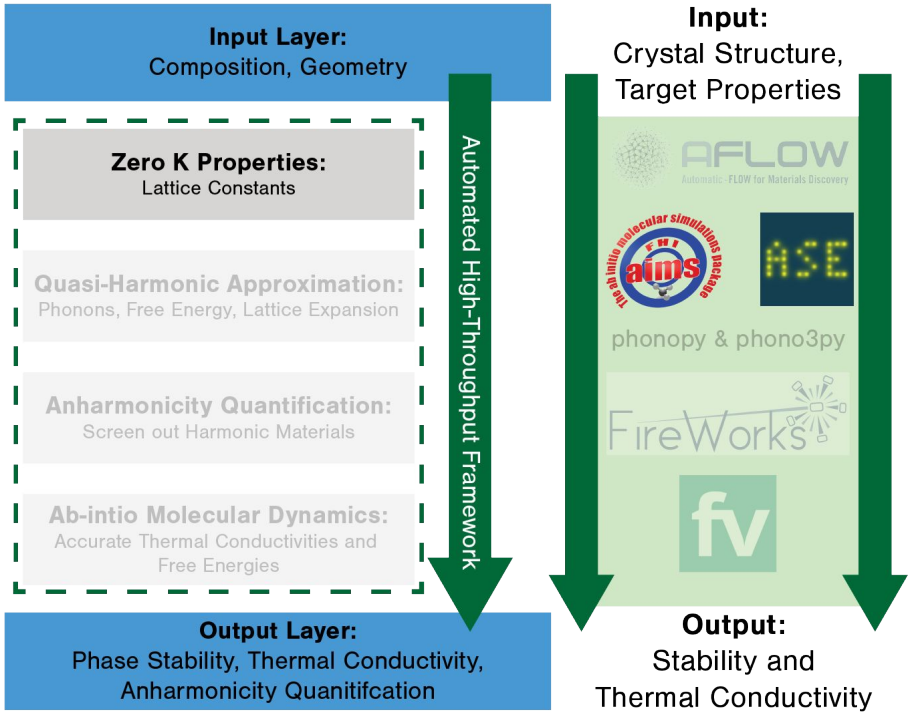

Knoop, F.; Purcell, T. A. R.; Scheffler, M. Carbogno, C. *J. Open Source Softw.* **5**, 2671. (2020)

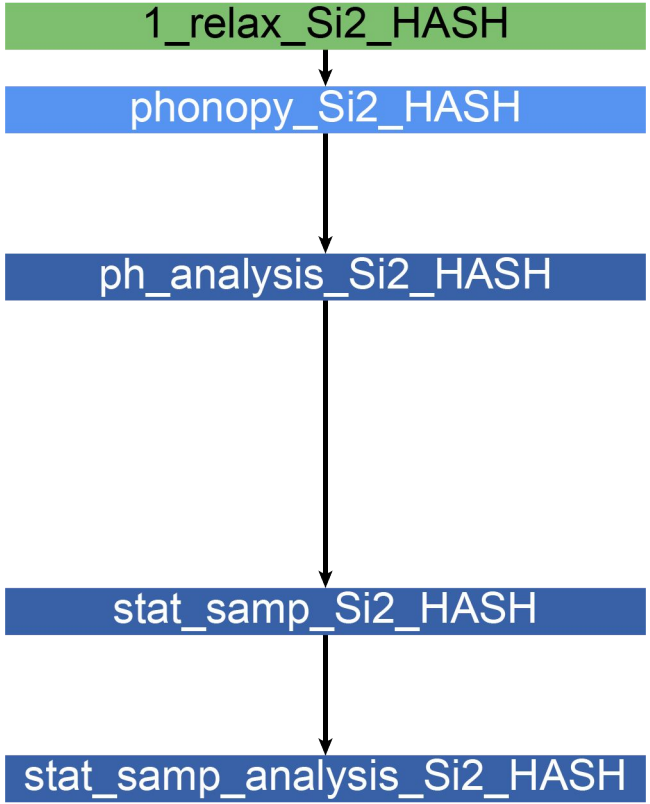
Initiate the Workflow



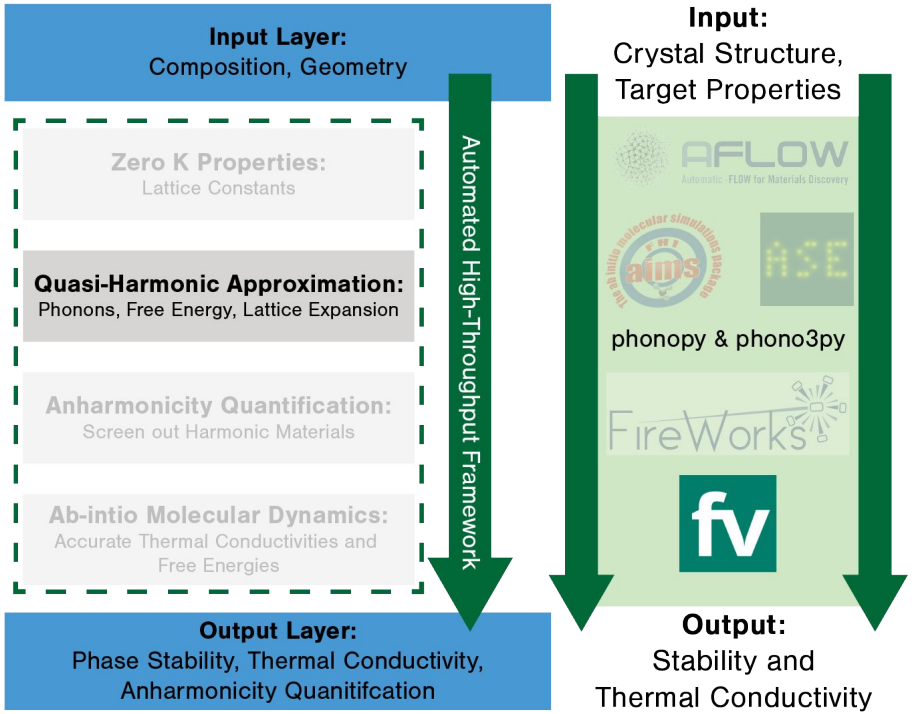


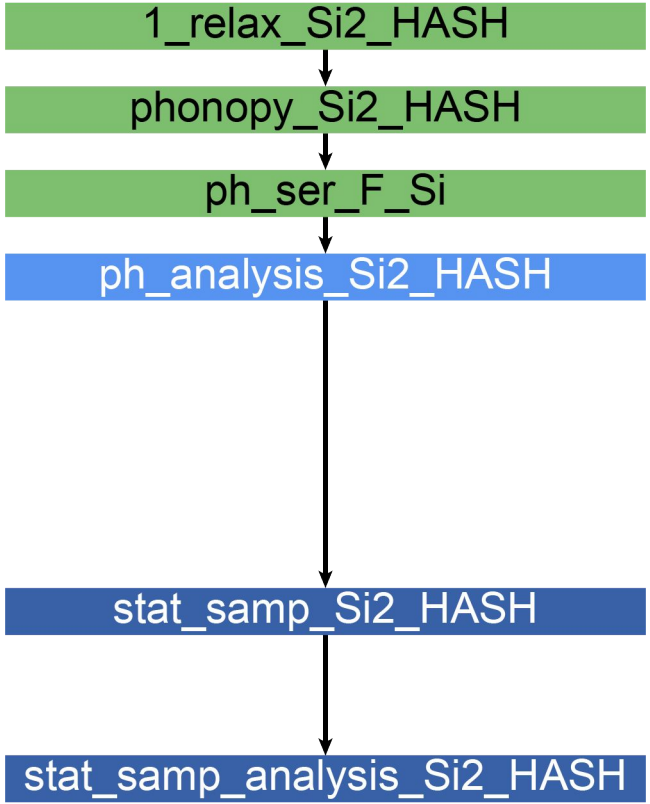
Relax the Structures with Parametric Constraints



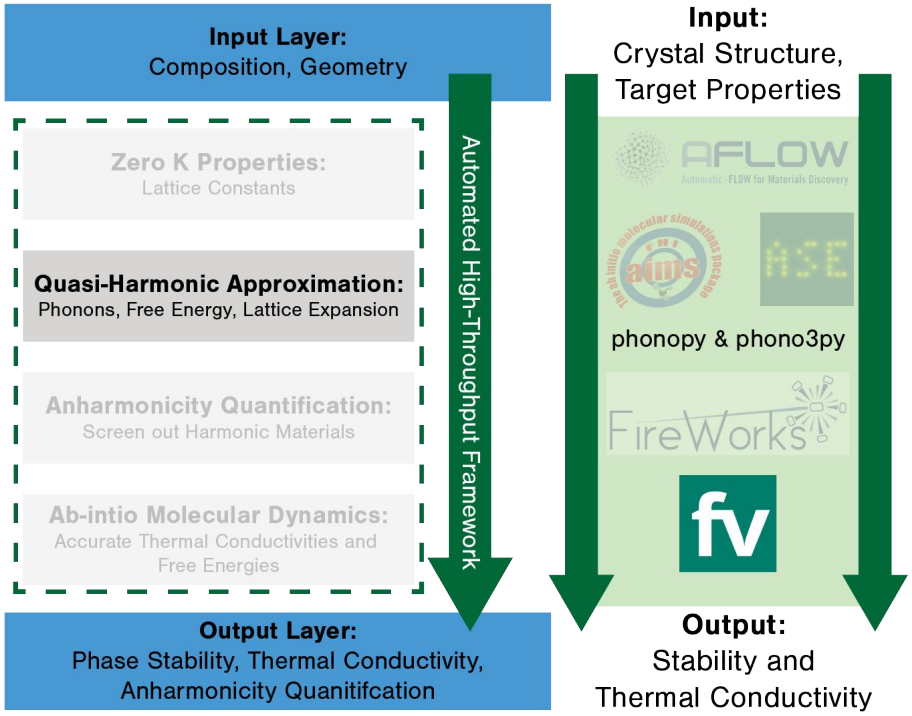


Build a Harmonic Model

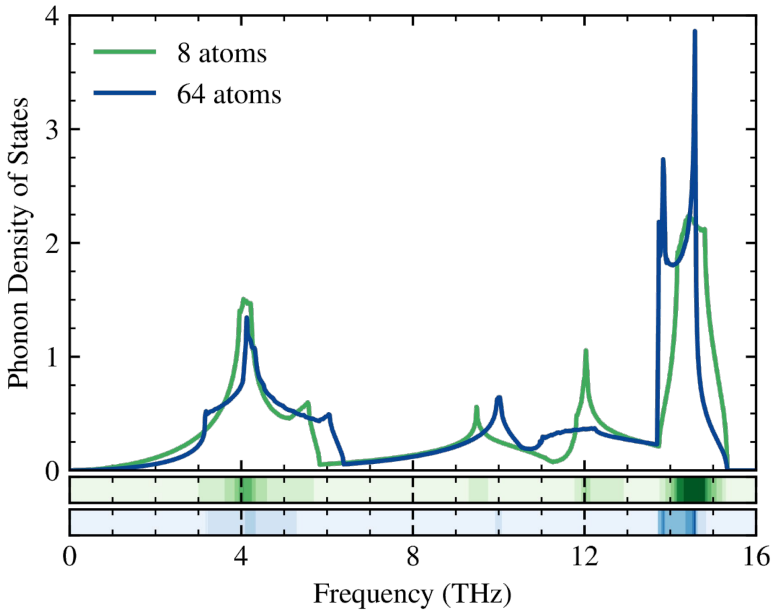




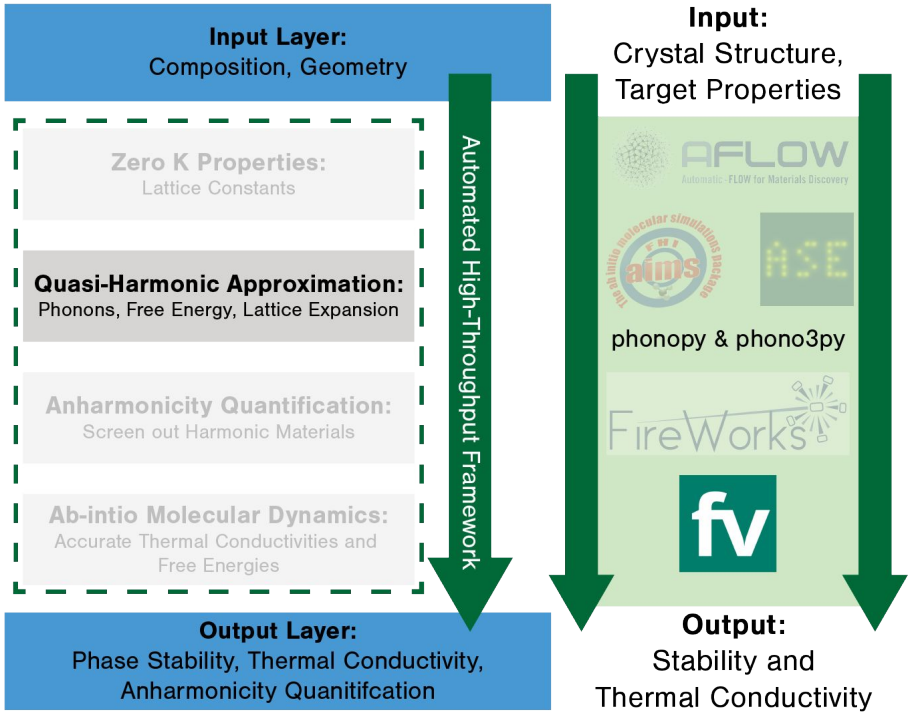
Check the Convergence of the Harmonic Model

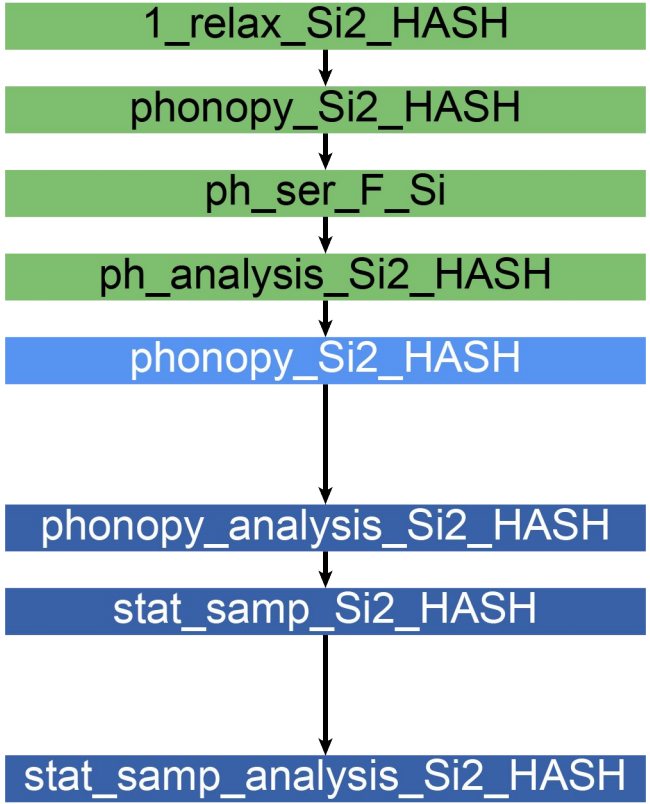


$$f(A,B) = \frac{A \cdot B}{|A|^2 + |B|^2 - A \cdot B}$$

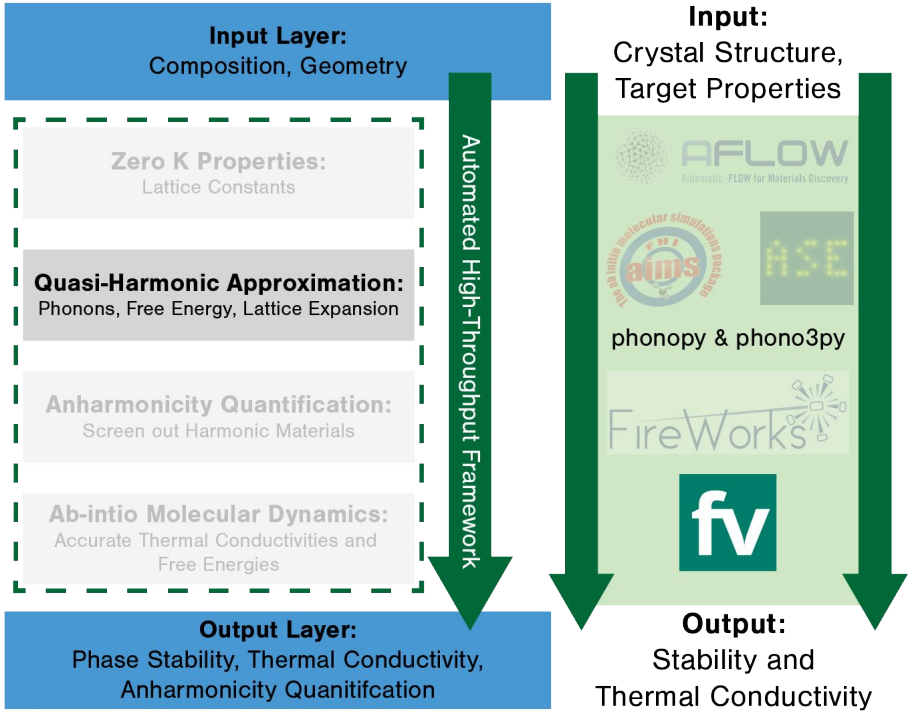


Automatically Generate New Tasks with Larger Supercells

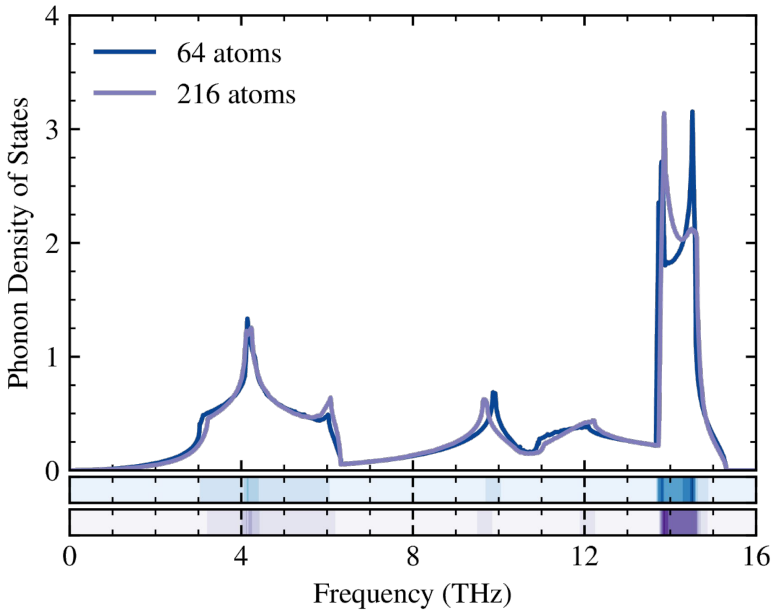




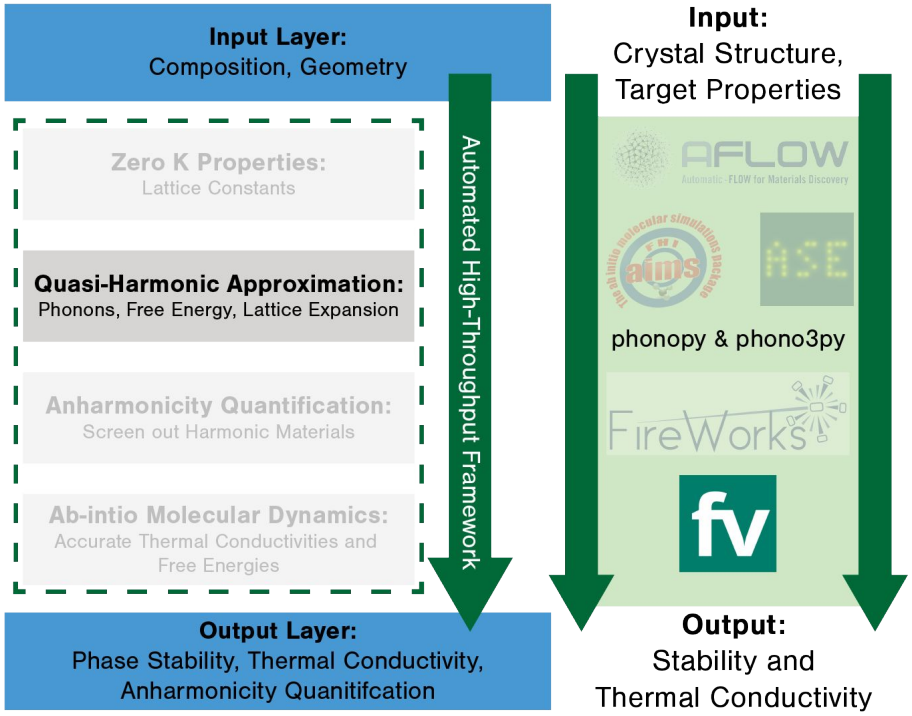
Harmonic Model Convergence Reached

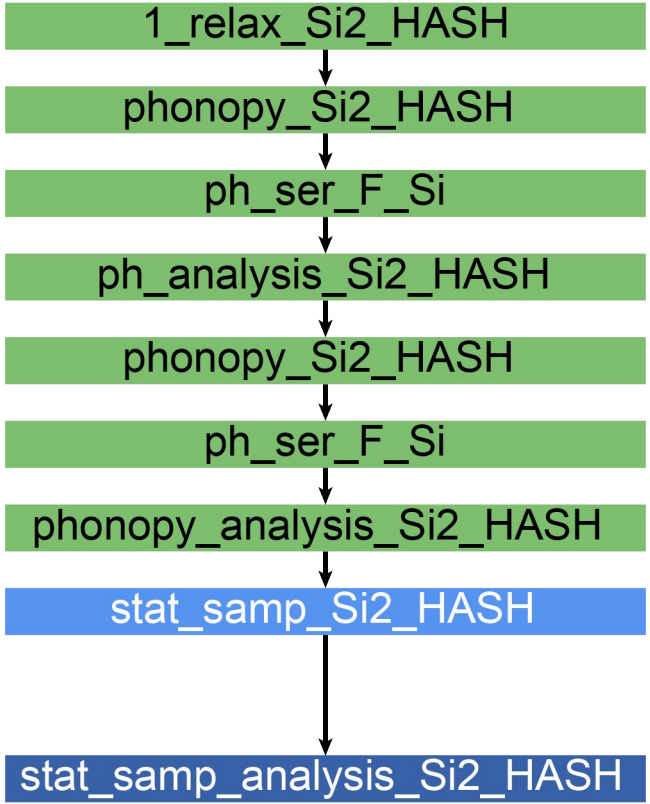


$$f(A,B) = \frac{A \cdot B}{|A|^2 + |B|^2 - A \cdot B}$$

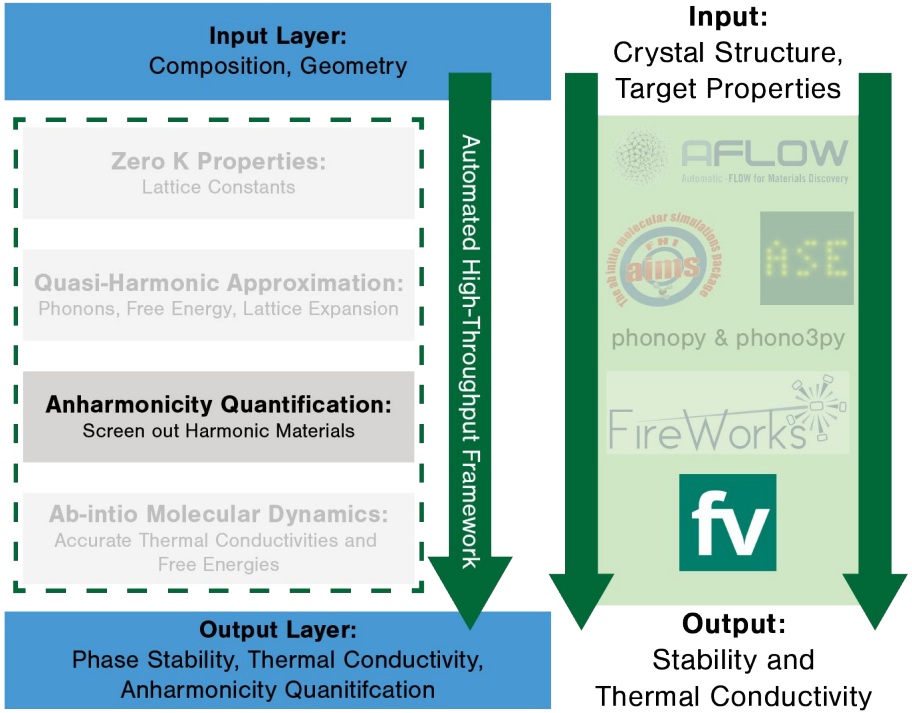


Harmonic Model Convergence Reached





Use the One-Shot Approximation to Quickly Calculate σ^A



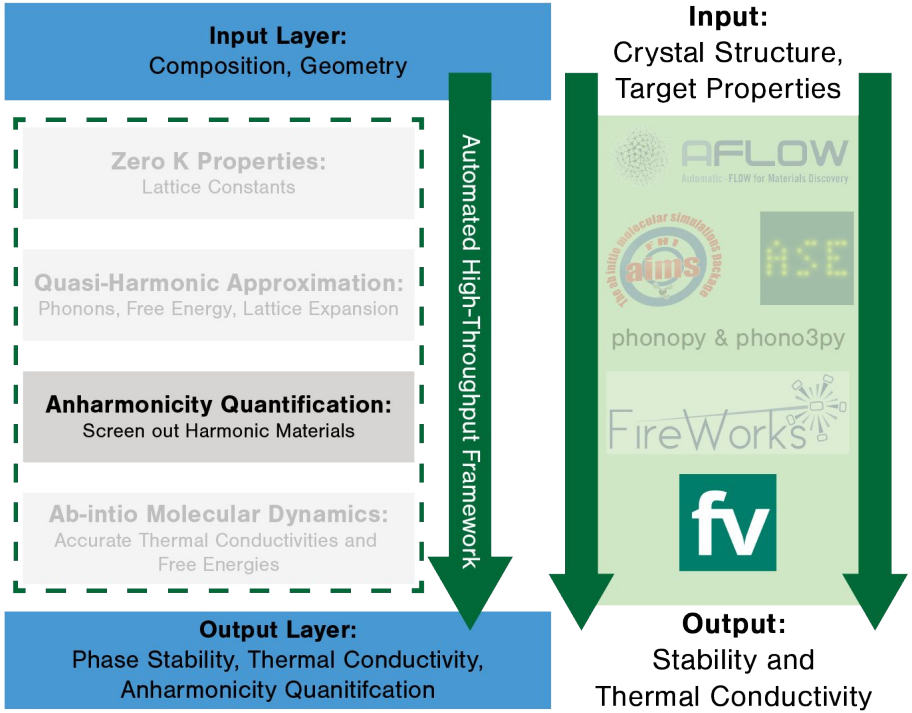
$$D(\Gamma)\left[\nu_s(\Gamma)\right] = \omega_s^2(\Gamma)[\nu(\Gamma)]$$

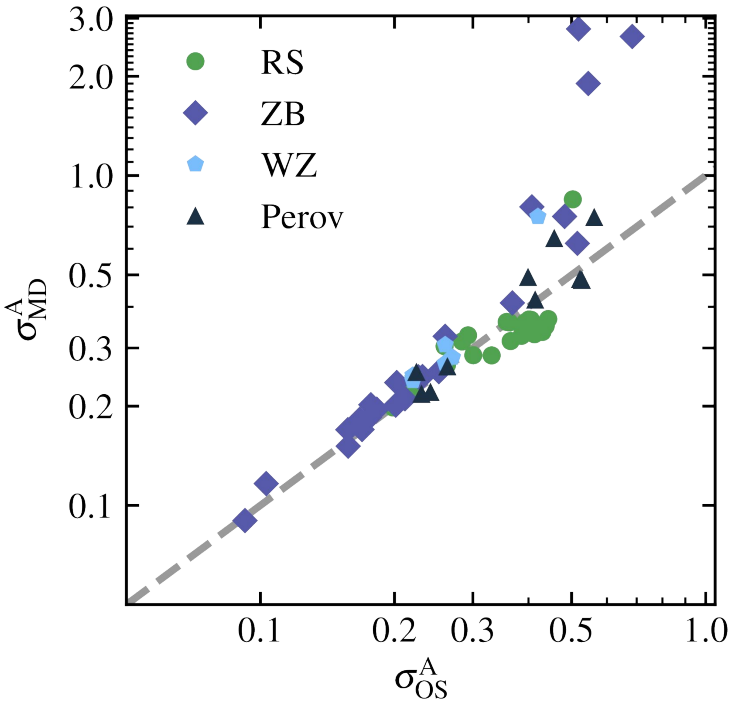
$$A_s = \frac{\sqrt{k_B T}}{\omega_s}$$

$$d_s = \sum_{s=3}^{n_w} -1^s A_s \nu_s(\Gamma)$$

M. Zacharias and F. Giustino, *Phys. Rev. B.* **94** 75125 (2016)

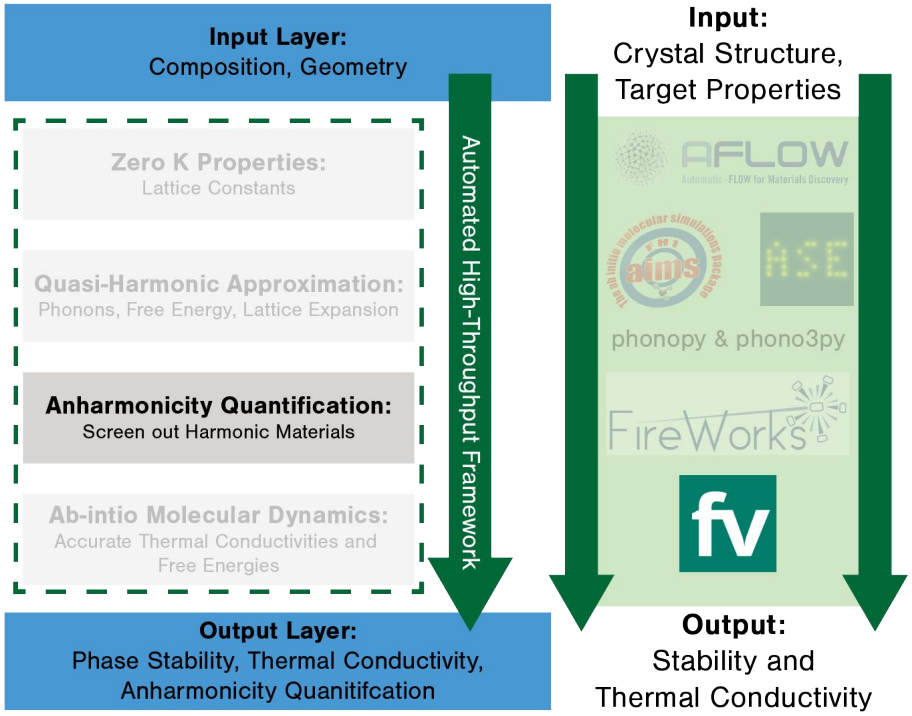
The One-Shot Approximation is Accurate Up to a Point

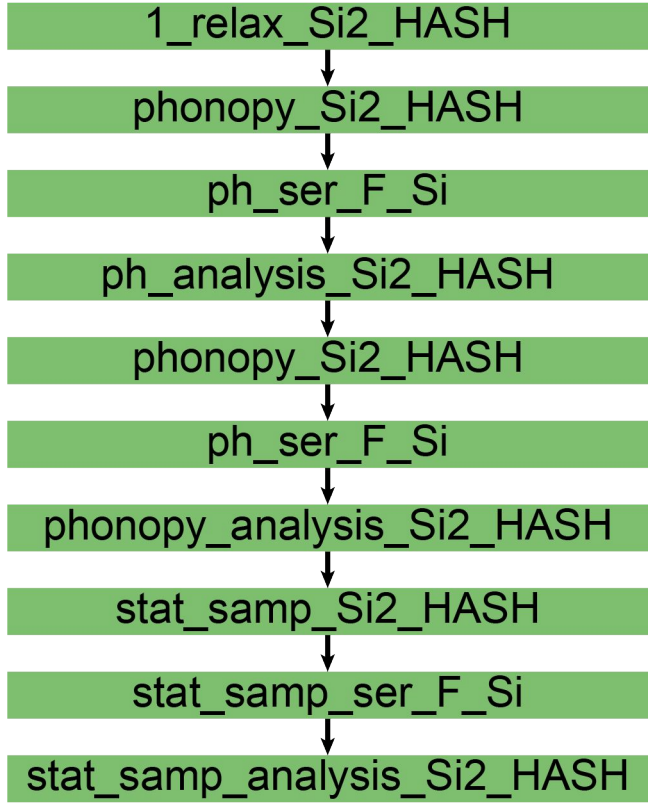




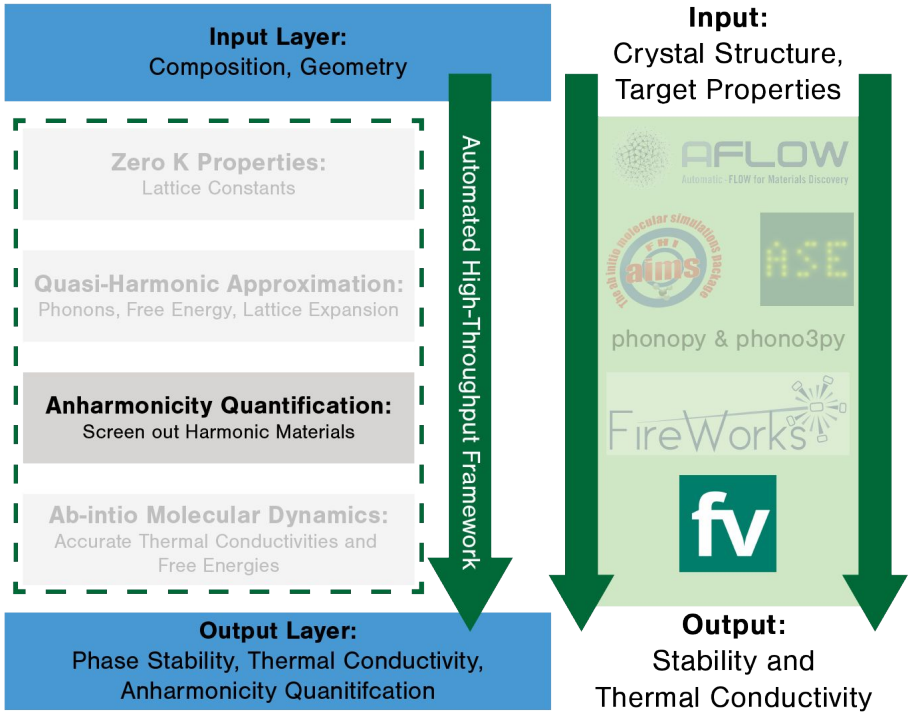
Knoop, F.; Purcell, T. A. R.; Scheffler, M. Carbogno, C. *Phys. Rev. Mater.* **4**, 083809. (2020)

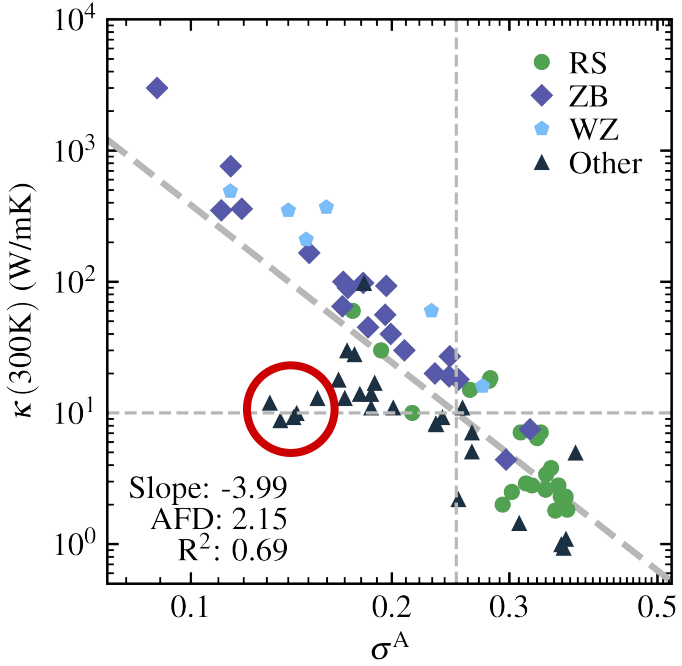
Workflow is Complete: Should we do aiMD?





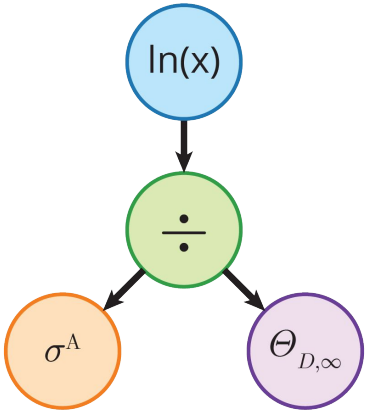
Probably Not, but Why the Outliers?

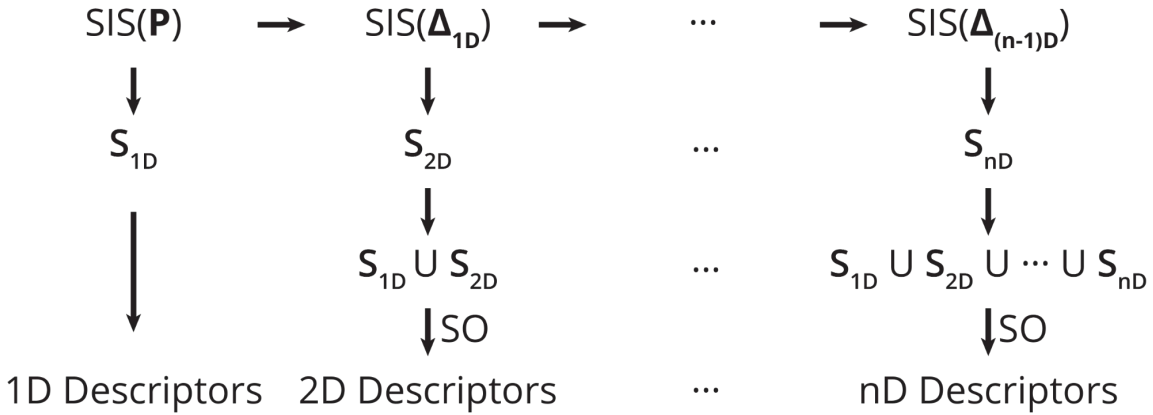




Knoop, F.; Purcell, T. A. R.; Scheffler, M.; Carbogno, C. *Phys. Rev. Mater.* **4**, 083809 (2020)
Chen, L; Tran, H.; Batra,R. Kim, C.; Ramprasad, R.. *Comput. Mater. Sci.* **170.** 109155 (2019)

Sure-Independence Screening and Sparsify Operator (SISSO)

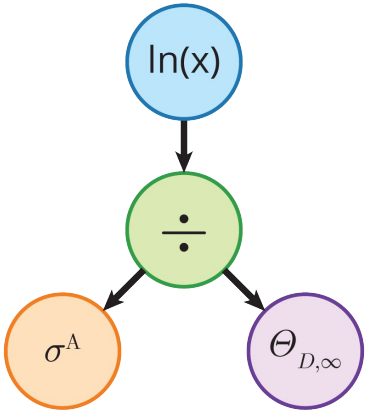




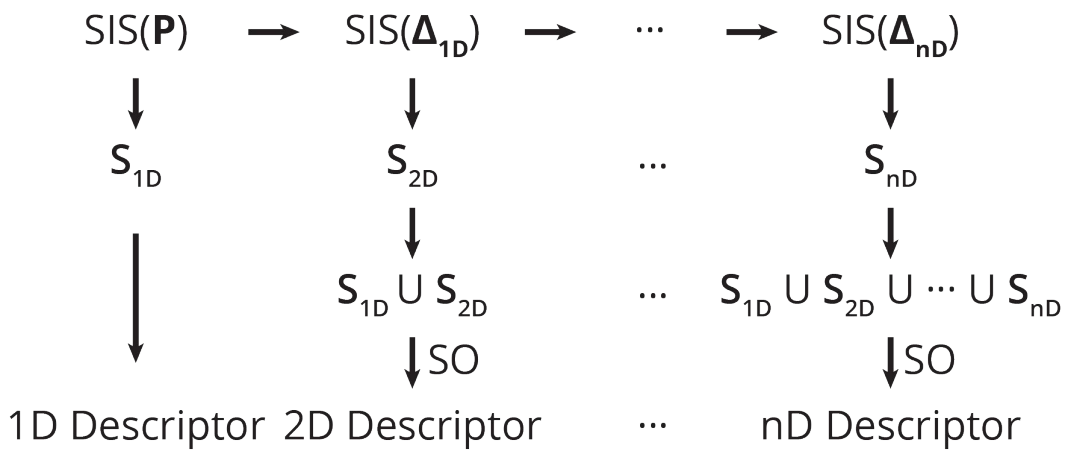
Purcell, T. A. R.; Scheffler, M. Carbogno, C.; Ghiringhelli, L. M. *submitted to J. Open Source Softw.*

Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M. Ghiringhelli, L. M. *Phys. Rev. Mater.* **2**, 083802. (2018)

Improve Feature Creation Step with Non-Linear Optimization



$$\ln\left(\alpha_1 \frac{\sigma^A}{\Theta_{D,\infty} + \beta_2} + \beta_1\right)$$



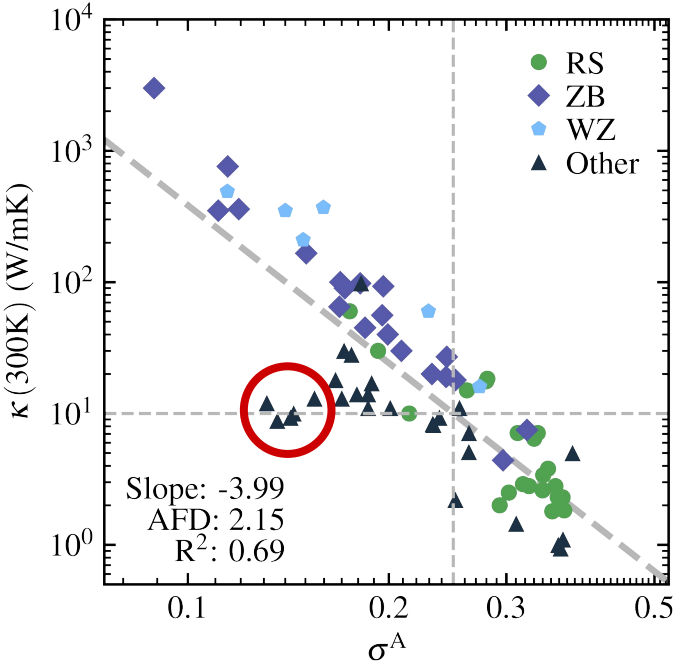
Purcell, T. A. R.; Scheffler, M. Carbogno, C.; Ghiringhelli, L. M. *submitted to J. Open Source Softw.*

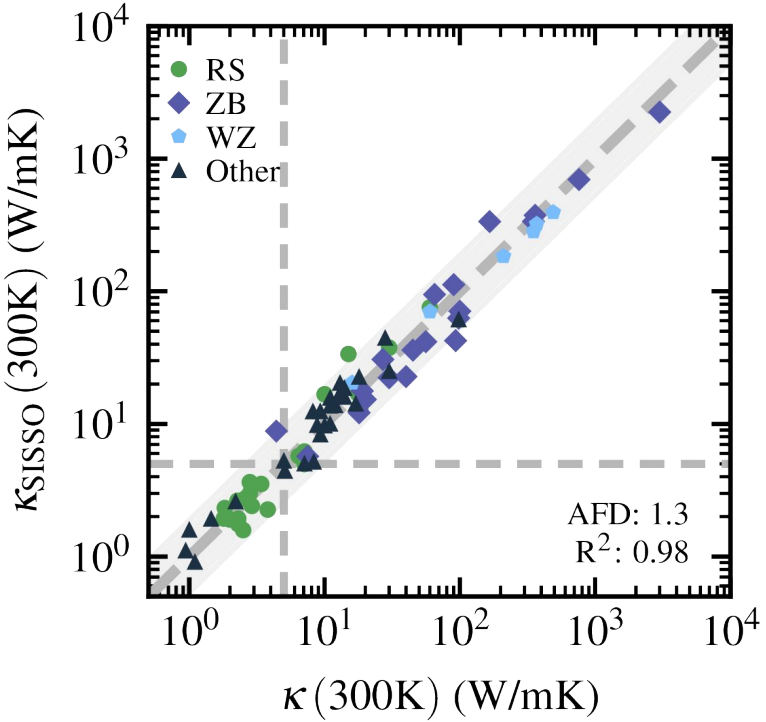
Steven G. Johnson, The NLopt nonlinear-optimization package, http://github.com/stevengj/nlopt

The Primary Features

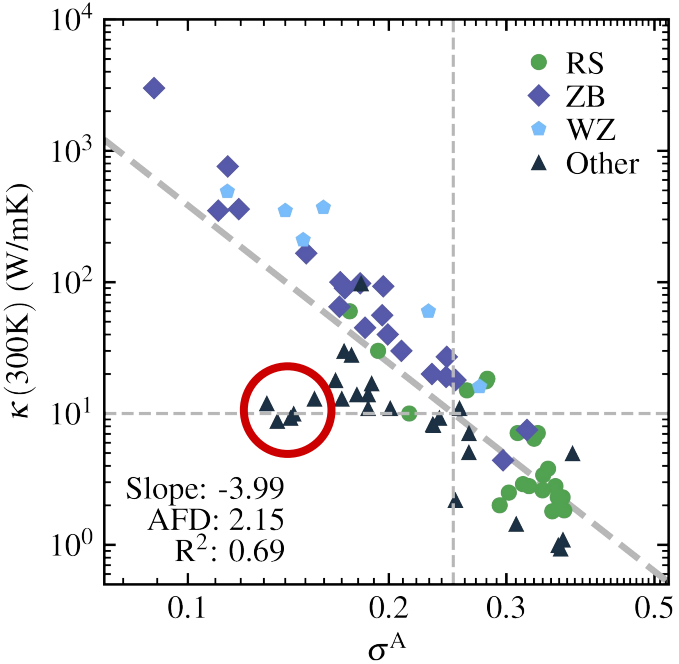
Property	Symbol	Units
Anharmonicity	σ^{A}, $\sigma^{\mathrm{A}}_{\mathrm{OS}}$	Unitless
Debye Temperature Approximations	$\Theta^{\infty}_{\mathrm{D}}$, Θ_{P}	K
Max Phonon Frequency	$\omega_{\Gamma,\mathrm{max}}$	THz
Heat Capacity	C_{V}	J mol^{-1} K^{-1}
Speed of Sound	v_{s}	m s^{-1}
Lattice	V_{m}, L_{min}, L_{max}, L_{avg}	Å
Mass	μ, m_{min}, m_{max}, m_{avg}	Da
Number of Atoms	n_{at}	Unitless
Density	ρ	Da Å^{-3}

The SISSO Model Removes the Outliers





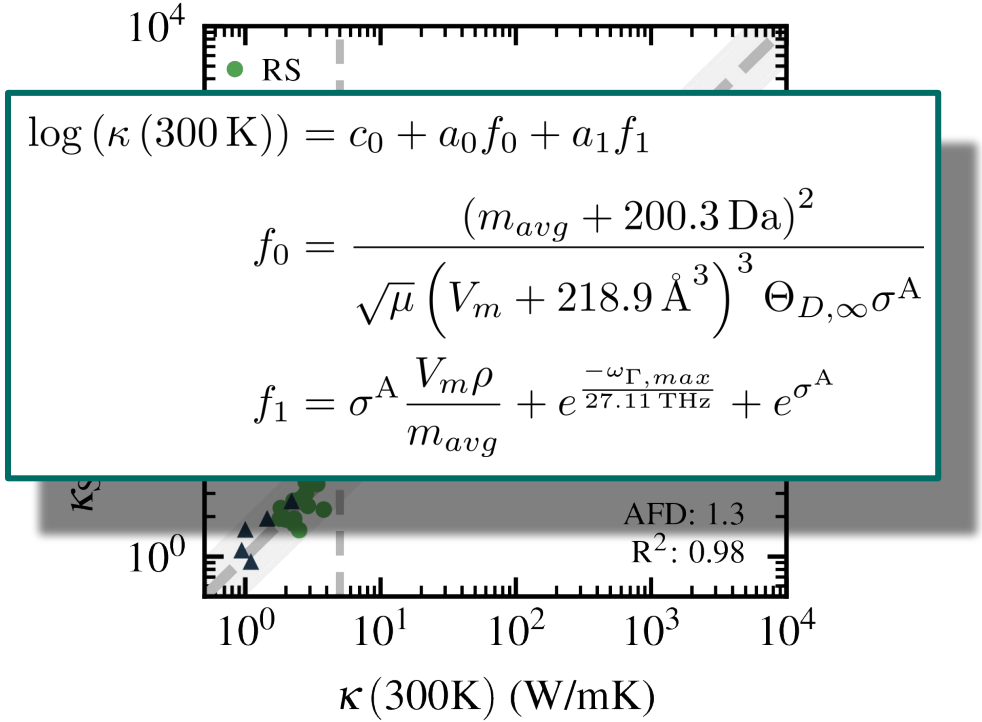
The SISSO Model Removes the Outliers



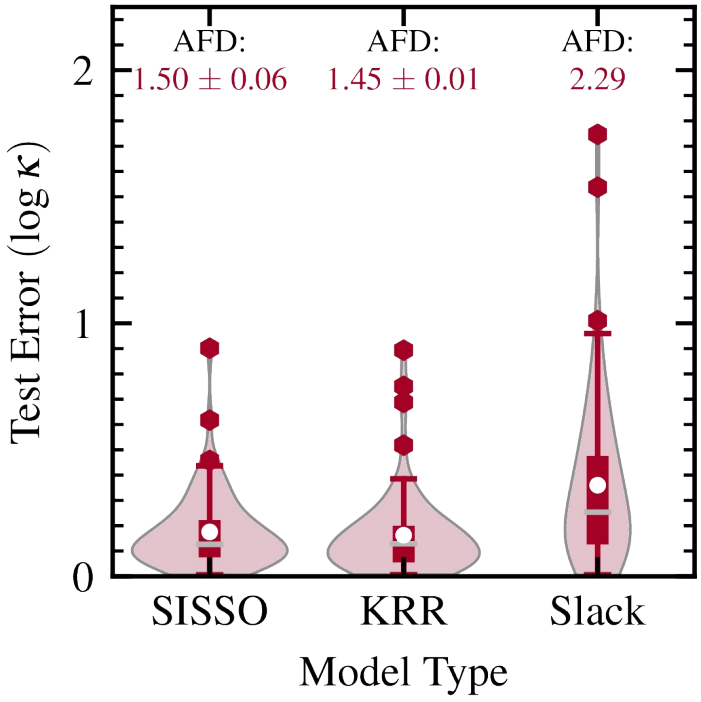
$$\log\left(\kappa\left(300\,\mathrm{K}\right)\right) = c_0 + a_0 f_0 + a_1 f_1$$

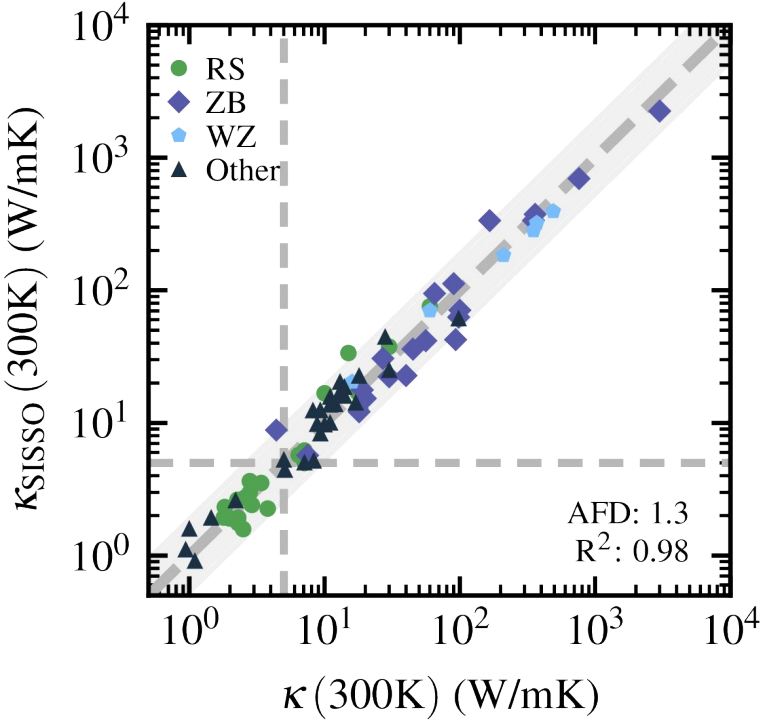
$$f_0 = \frac{\left(m_{avg} + 200.3\,\mathrm{Da}\right)^2}{\sqrt{\mu}\left(V_m + 218.9\,\mathrm{\AA}^3\right)^3 \Theta_{D,\infty}\sigma^A}$$

$$f_1 = \sigma^A \frac{V_m \rho}{m_{avg}} + e^{\frac{-\omega_{\Gamma,max}}{27.11\,\mathrm{THz}}} + e^{\sigma^A}$$

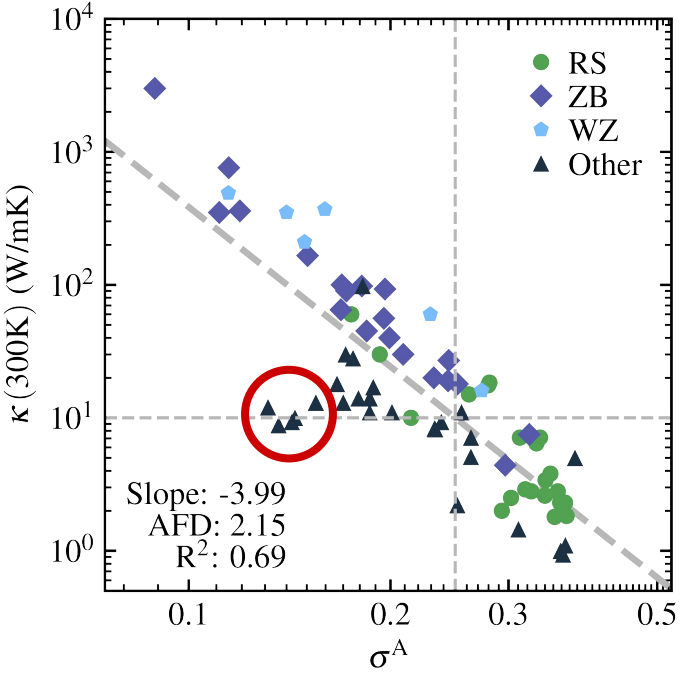


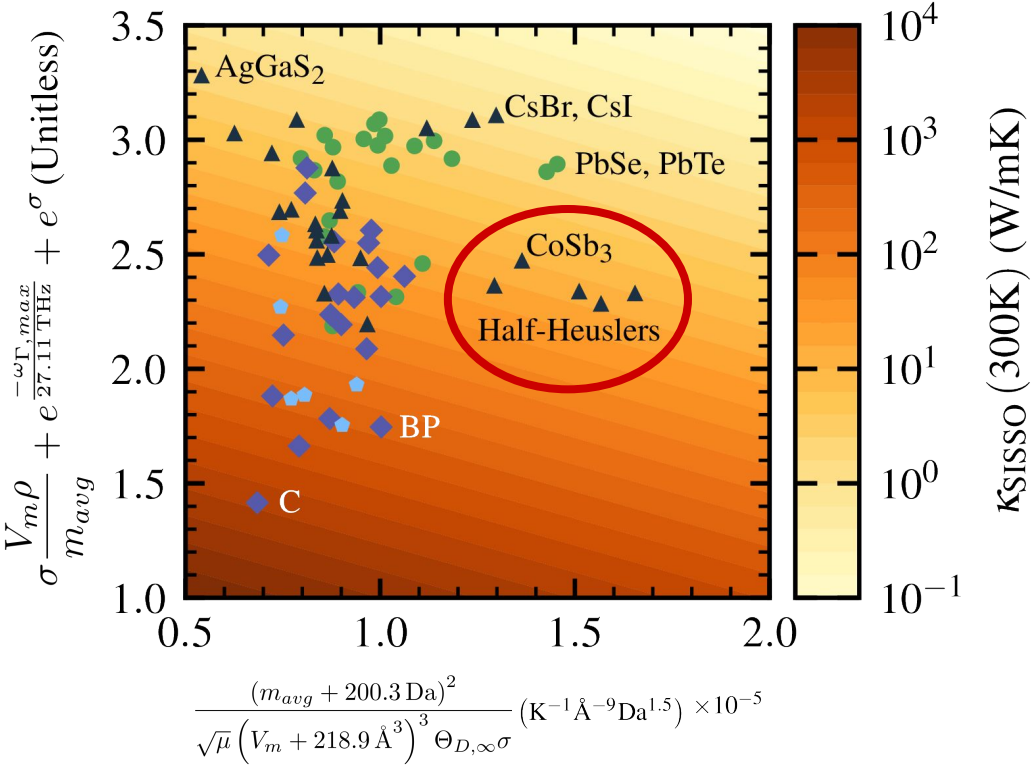
The Model is Predictive and Descriptive



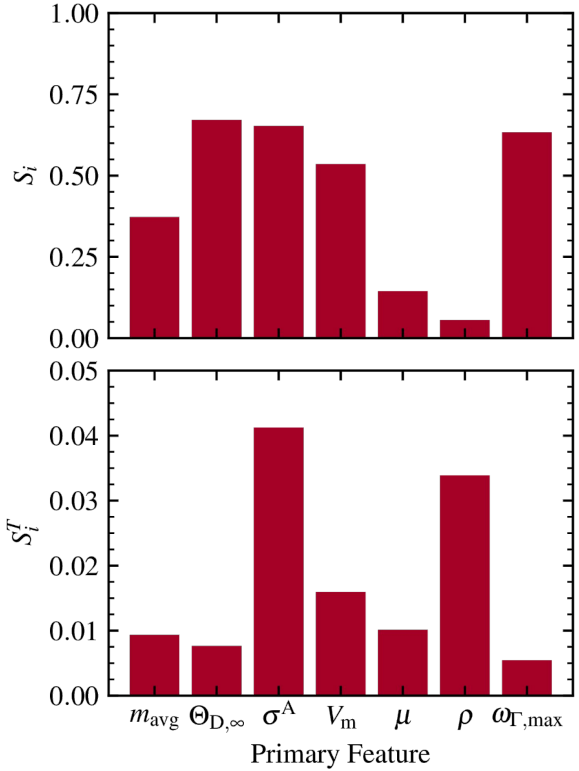


The Origins of the Outliers Found





Which Features Best Describe κ?

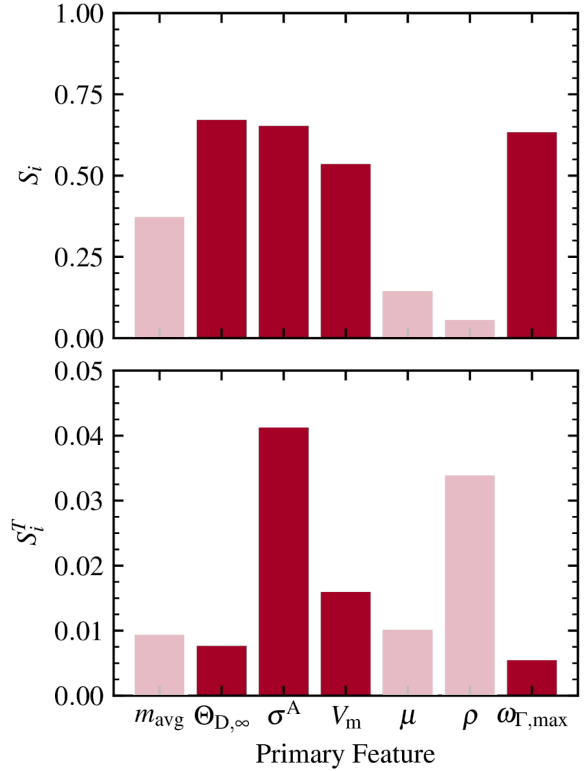


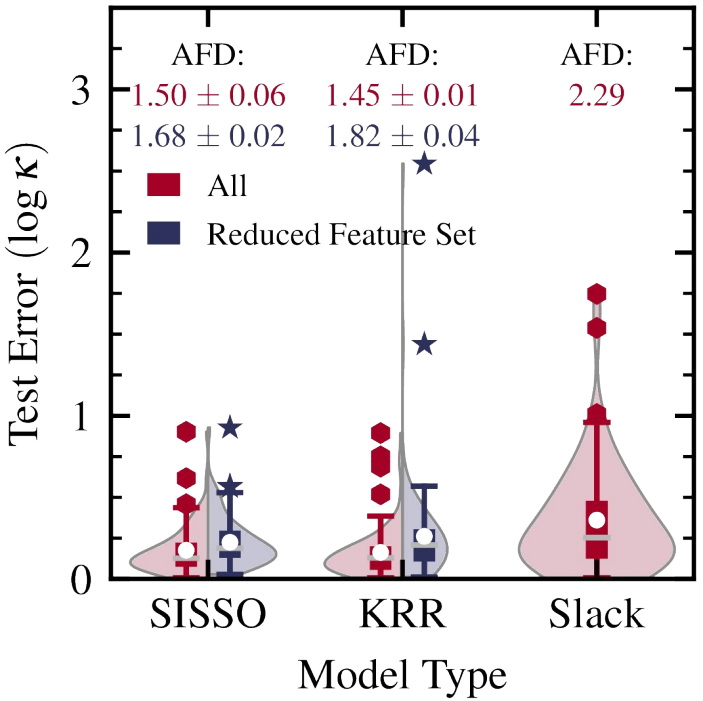
$$S_i = \frac{\mathrm{Var}_{\hat{x}_i}\left(E_{\widetilde{\mathcal{X}}_i}\left(Y|\hat{x}_i\right)\right)}{\mathrm{Var}\left(Y\right)}$$

$$S_i^T = 1 - \frac{\mathrm{Var}_{\widetilde{\mathcal{X}}_i}\left(E_{\hat{x}_i}\left(Y|\widetilde{\mathcal{X}}_i\right)\right)}{\mathrm{Var}\left(Y\right)}$$

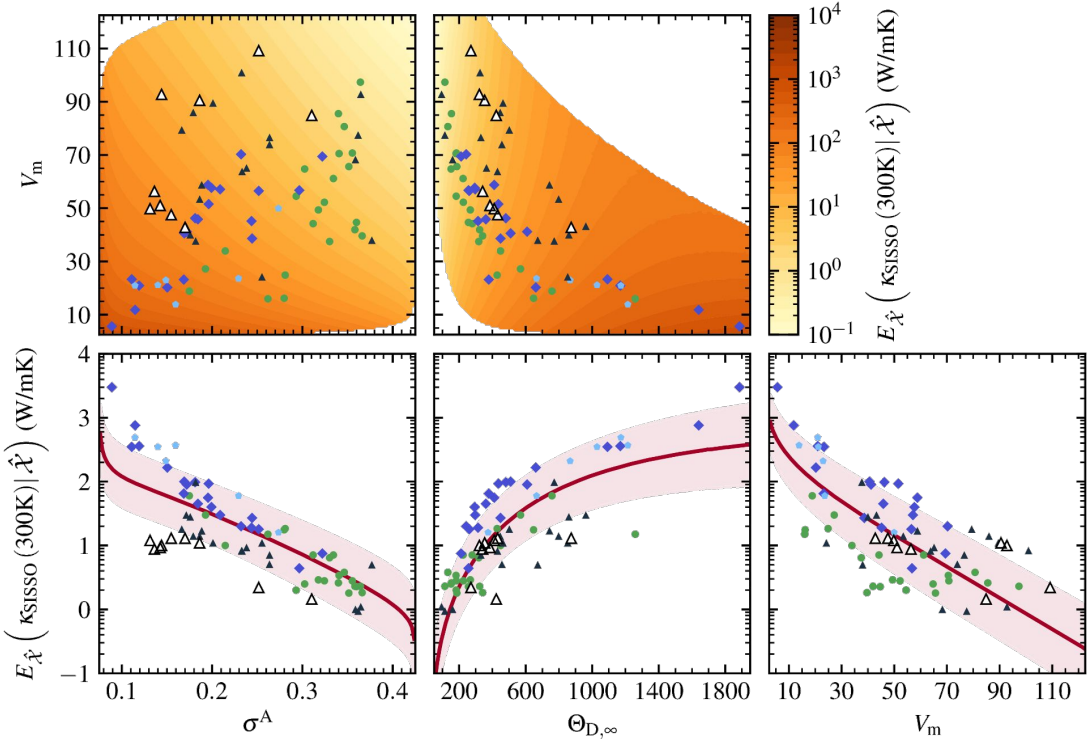
P. Wiederkehr, *Global Sensitivity Analysis with Dependent Inputs*, Ph.D. thesis (2018)
Mara, T.; Tarantola, P.; Annoni, P. *Environ. Model. Softw.* **72.** 173. (2015)
Kucherenko, S., T.; Tarantola, P.; Annoni, P. *Comput. Phys. Commun.* **183.** 937. (2012)

The Reduced Feature Space Still Generates Accurate Models

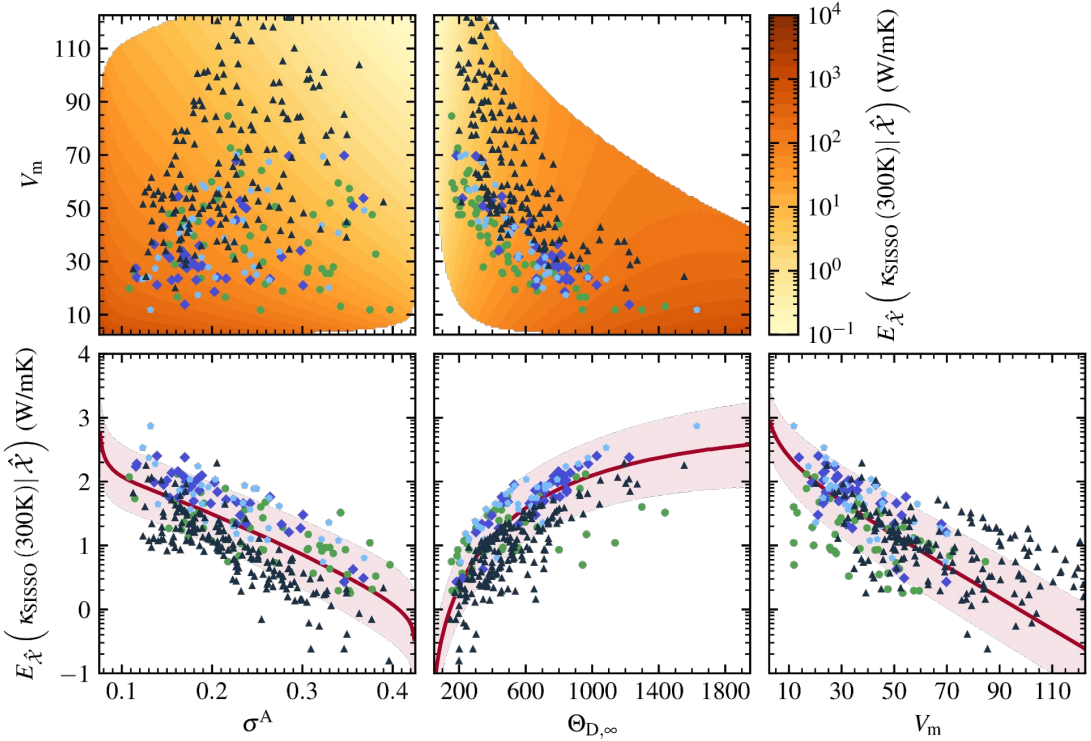


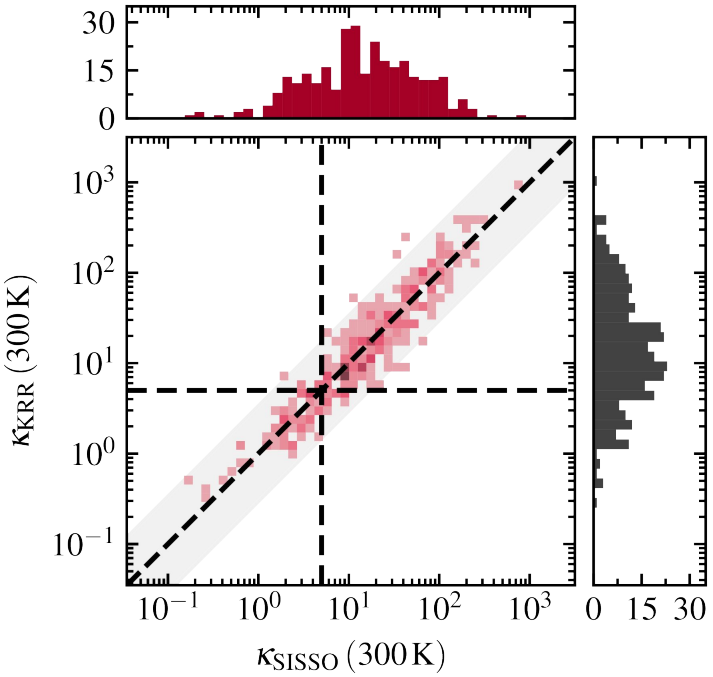


Increased Complexity and Soft Phonons Are Responsible for the Outliers

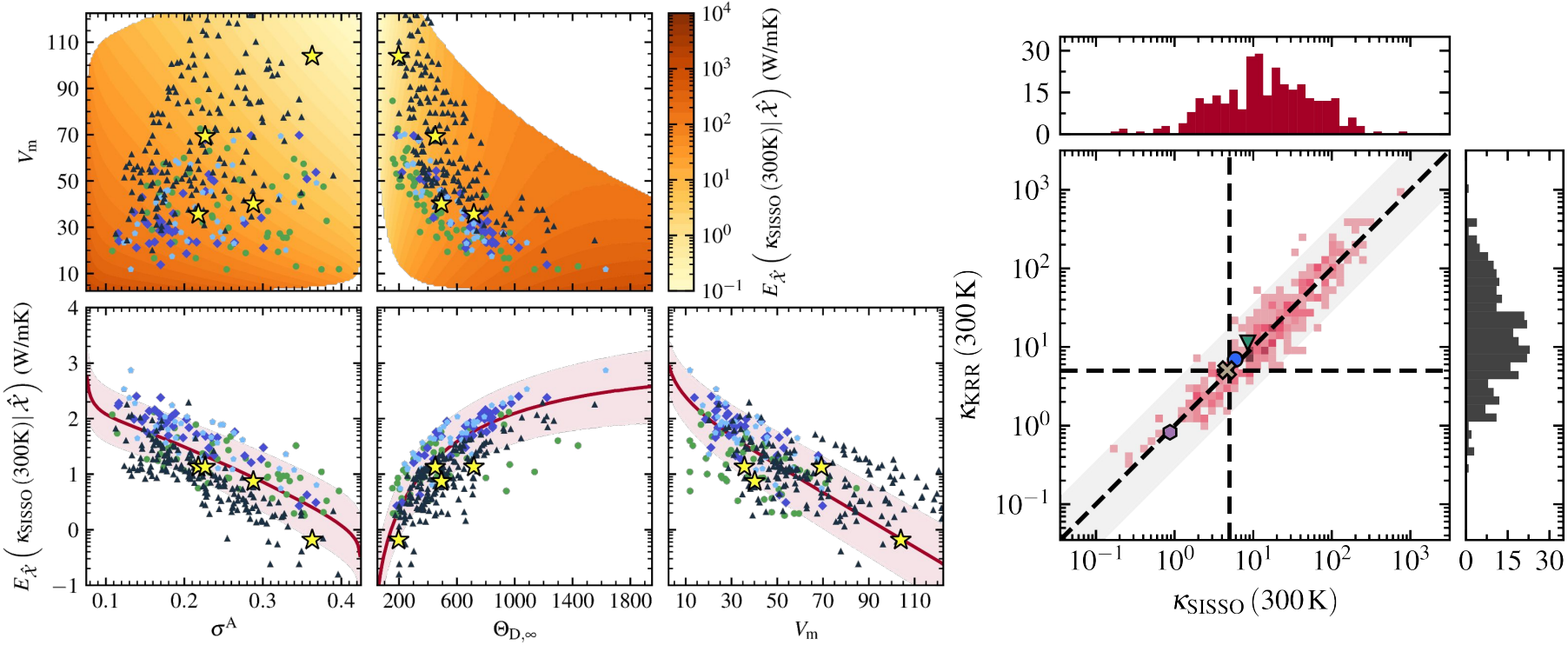


Models Can Predict New Thermal Insulators

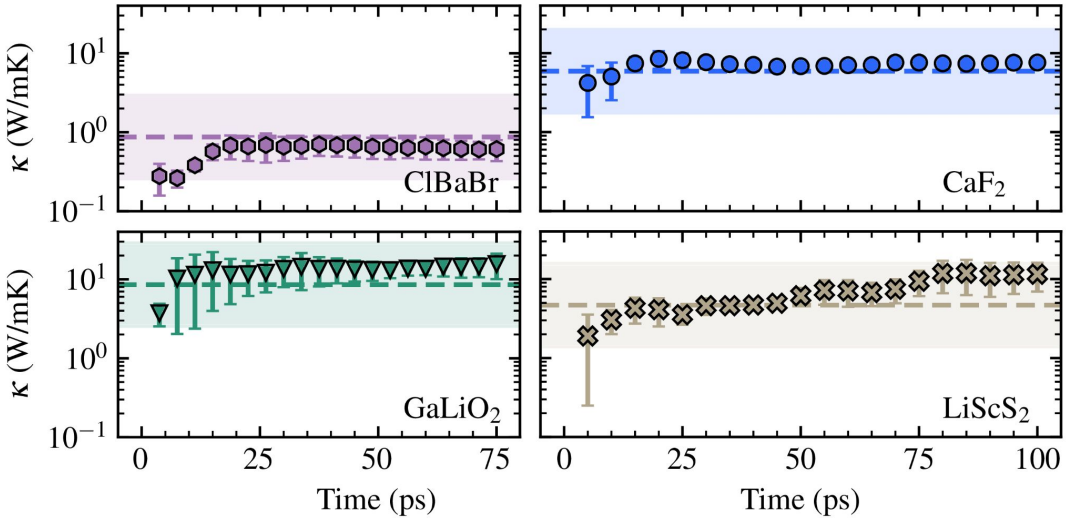


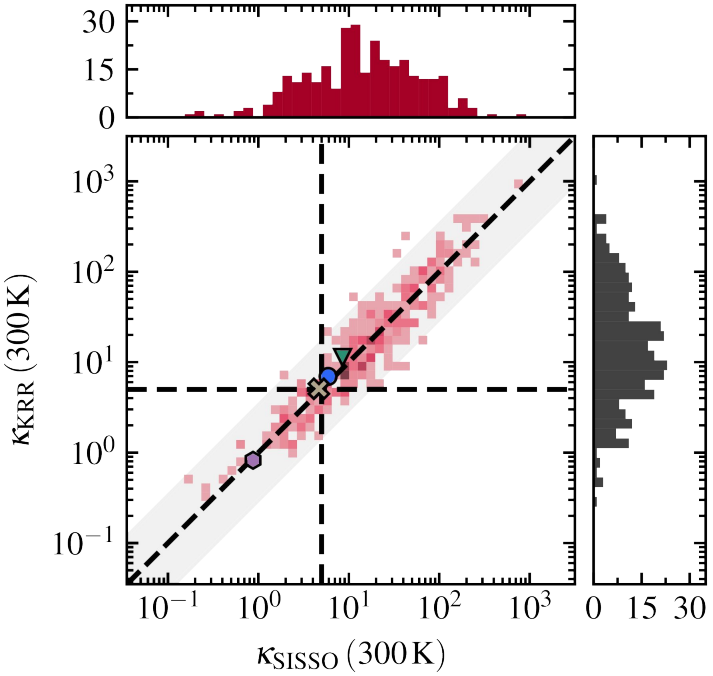


Models Can Predict New Thermal Insulators

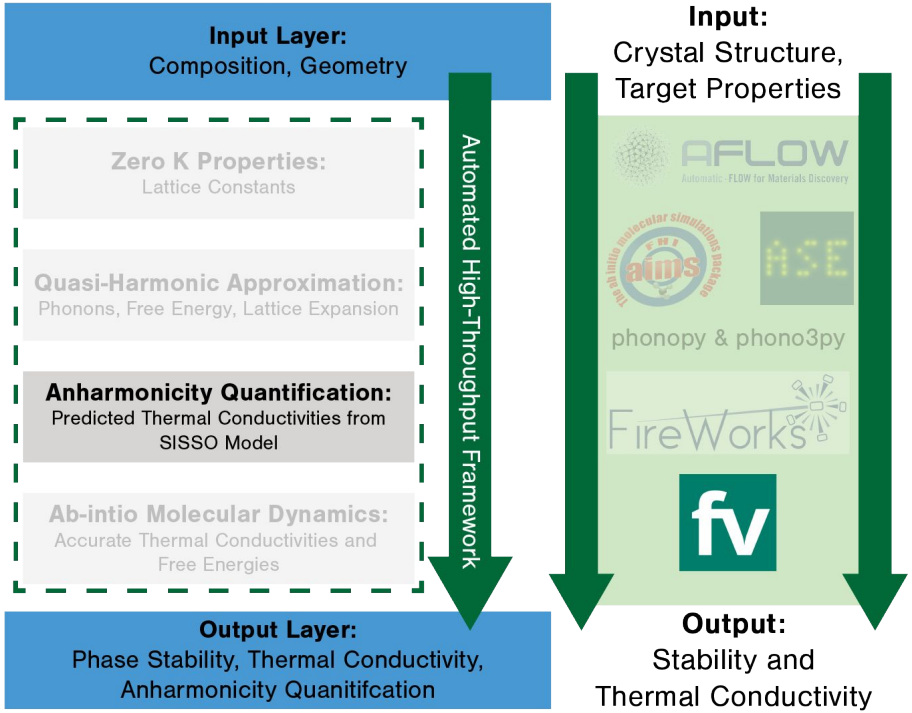


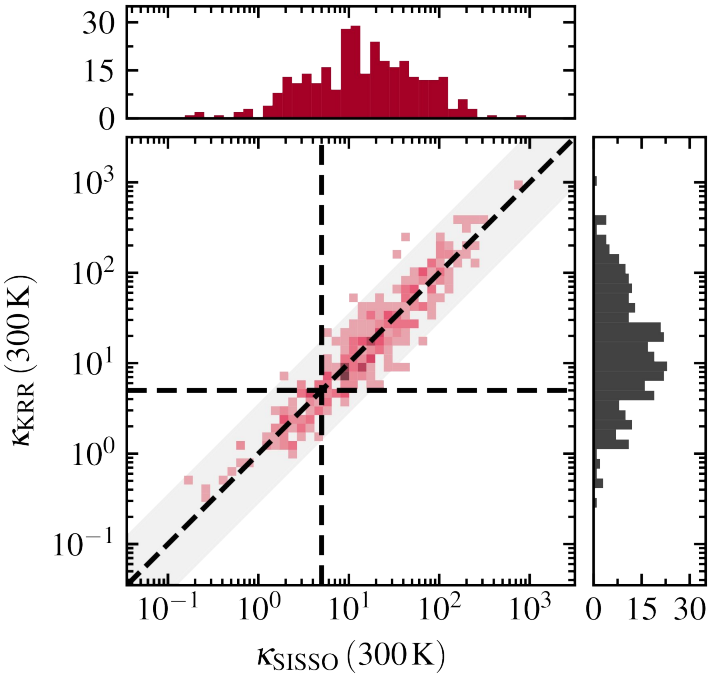
The SISSO Model Can Predict Accurate Thermal Conductivities





The Incorporation of SISSO Models: A Path to Active Workflows





Accessing the Software

SISSO++

https://tinyurl.com/sissopp

Purcell, T. A. R.; Scheffler, M. Carbogno, C.; Ghiringhelli, L. M. *submitted to J. Open Source Softw.*
Knoop, F.; Purcell, T. A. R.; Scheffler, M. Carbogno, C. *J. Open Source Softw.* **5**, 2671. (2020)

Acknowledgements

Florian Knoop

Lucas Foppa

European Research Council
Established by the European Commission

Alexander von Humboldt
Stiftung/Foundation

MPCDF Max Planck Computing & Data Facility

Questions?