# Accurate Energy Predictions via Machine Learning

Matthias Rupp

Fritz Haber Institute of the Max Planck Society, Berlin, Germany

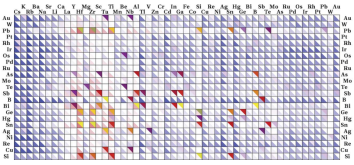
# Outline

1. Rationale
   ab initio simulations, machine learning
2. Kernel learning
   kernel trick, kernel ridge regression
3. Representations
   many-body tensor representation
4. Energy predictions
   energies of molecules and crystals

# Rationale

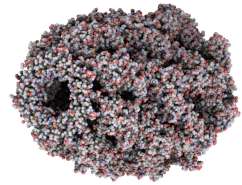
# Challenges in quantum mechanical simulations

## High-throughput screening



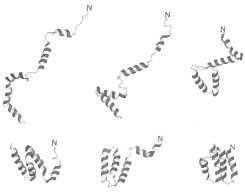
[Castelli et al, *Energy Environ Sci* 12, 2013]

## Large systems



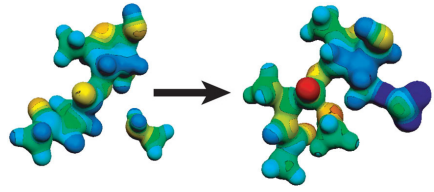
[Image: Tarini et al, *IEEE Trans Visual Comput Graph* 2006]

## Long simulations



[Liwo et al, *Proc Natl Acad Sci USA* 102: 2362, 2005]

## Quantum effects



[Image: Hiller et al, *Nature* 476: 236, 2011]

# Accurate simulations are limited by computational cost

accuracy ↑

| Method |
| --- |
| Full Configuration Interaction |
| Coupled Cluster |
| Møller-Plesset second order perturbation theory |
| Kohn-Sham Density Functional Theory |
| Tight Binding |
| Molecular Mechanics |

speed ↓

**Is it possible to be both accurate and fast?**
In high-throughput settings, machine learning offers a new trade-off

# Machine learning

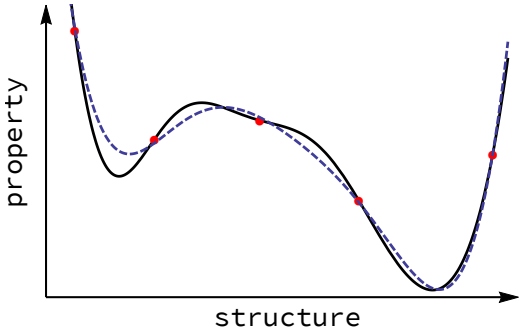
Machine learning (ML) studies algorithms whose performance *improves with data* ("learning from experience"). [Mitchell, McGraw Hill, 1997]

Data $\boldsymbol{X}$ $\rightarrow$  $\rightarrow$ Model $\hat{f}$

- No explicitly programmed problem-specific solution
- Systematic identification of regularity in data for prediction & analysis
- Interpolation in high-dimensional spaces

# Interpolating electronic structure calculations with machine learning

Correlated calculations can be rapidly and accurately interpolated




● reference calculations, — ground truth, - - - machine learning model

Rupp, Int. J. Quant. Chem., 2015

Assumes similar structure, similar property
Corresponding space or metric

Flexible functional form
Smoothness assumption

Physics in the examples
Flexibility versus physical constraints

# Relationship with other models

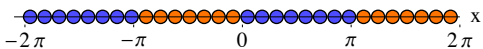
| ab initio | force fields | machine learning |
|---|---|---|
| generally applicable | limited domain | generally applicable |
| no or little fitting | fitting to one class | refitted to any dataset |
| form from physics | form from physics | form from statistics |
| deductive | mostly deductive | inductive |
| few or no parameters | some parameters | many parameters |
| high accuracy | limited accuracy | high accuracy |
| slow | fast | in-between |
| small systems | large systems | large systems |

# Kernel learning

# Learning with kernels

The kernel trick:

- Transform samples into higher-dimensional space
- Implicitly compute inner products there
- Rewrite linear algorithm using only inner products



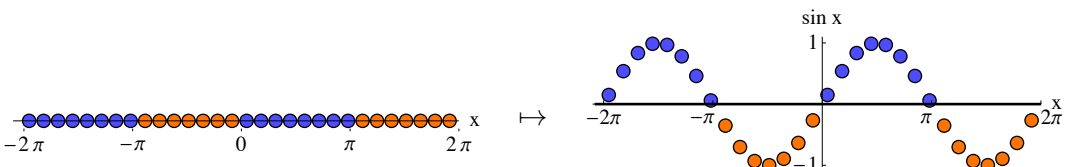

Input space $\mathcal{X}$

Schölkopf & Smola: Learning with Kernels, 2002; Hofmann et al., Ann. Stat., 2008

# Learning with kernels

The kernel trick:

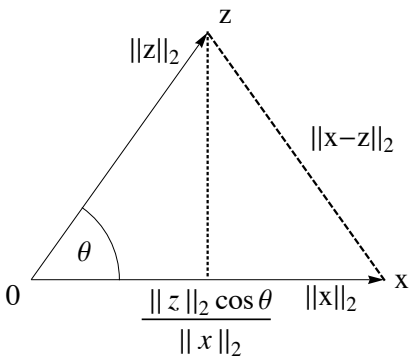
- Transform samples into higher-dimensional space
- Implicitly compute inner products there
- Rewrite linear algorithm using only inner products




$$k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}, \qquad k(x, z) = \langle \phi(x), \phi(z) \rangle$$

Schölkopf & Smola: Learning with Kernels, 2002; Hofmann et al., Ann. Stat., 2008

# Kernel functions




Kernels correspond to **inner products**

Encode information about lengths and angles:
$||x - z||^2 = \langle x, x\rangle - 2\langle x, z\rangle + \langle z, z\rangle$, $\cos\theta = \frac{\langle x,z\rangle}{||x||\,||z||}$

If $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is symmetric positive semi-definite, then $k(x, z) = \langle \phi(x), \phi(z)\rangle$ for some $\phi : \mathcal{X} \to \mathcal{H}$

Useful properties

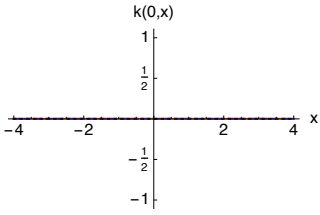
Closed convex cone structure

Natural interface $\boldsymbol{K}_{ij} = k(x_i, x_j)$

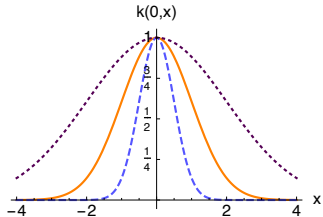
$\mathcal{X}$ can be any non-empty set

# Examples of kernel functions

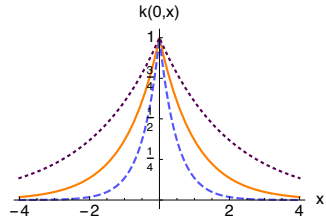
Linear kernel
$k(\boldsymbol{x}, \boldsymbol{z}) = \langle \boldsymbol{x}, \boldsymbol{z} \rangle$




Gaussian kernel
$\exp(-\|\boldsymbol{x} - \boldsymbol{z}\|^2 / 2\sigma^2)$




Laplacian kernel
$\exp(-\|\boldsymbol{x} - \boldsymbol{z}\|_1 / \sigma)$



# Regression with kernels

**Linear ridge regression**

For models

$$f(\boldsymbol{x}) = \sum_{i=1}^{d} \beta_i \boldsymbol{x}_i$$

minimizing

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^d} \sum_{i=1}^{n} \big(f(\boldsymbol{x_i}) - y_i\big)^2 + \lambda ||\boldsymbol{\beta}||^2$$

yields

$$\boldsymbol{\beta} = \big(\boldsymbol{X}^T \boldsymbol{X} + \lambda \boldsymbol{I}\big)^{-1} \boldsymbol{X}^T \boldsymbol{y}$$

**Kernel ridge regression**

For models

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i k(\boldsymbol{x_i}, \boldsymbol{x})$$

minimizing

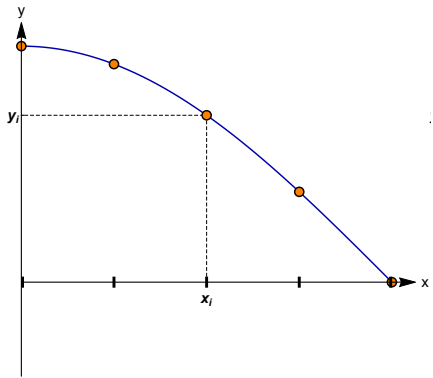
$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^n} \sum_{i=1}^{n} \big(f(\boldsymbol{x_i}) - y_i\big)^2 + \lambda ||\boldsymbol{\alpha}||^2_{\mathcal{H}}$$

yields

$$\boldsymbol{\alpha} = \big(\boldsymbol{K} + \lambda \boldsymbol{I}\big)^{-1} \boldsymbol{y}$$

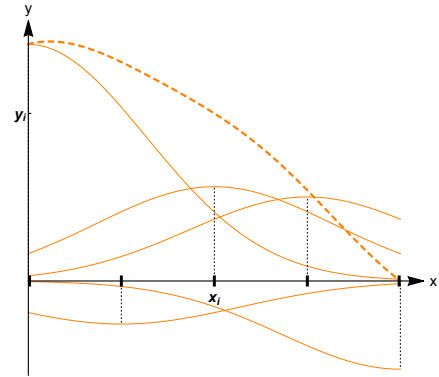
# A basis function picture of kernel regression

Weighted basis functions placed on training samples $x_i$




— learned $f(x)$

• training samples $(x_i, y_i)$

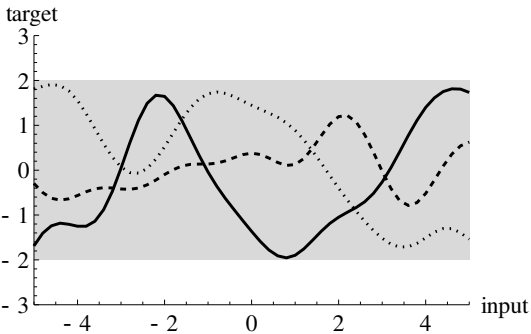


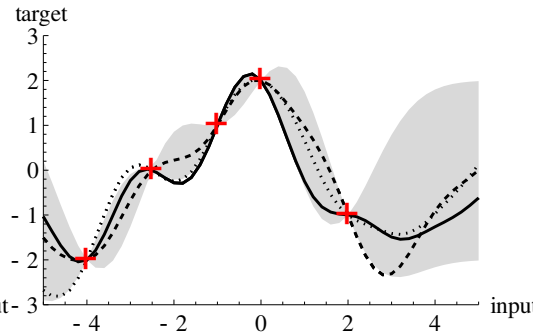

— basis functions $k(x_i, x)$

- - - prediction $\hat{f}(x) = \sum_i \alpha_i k(x_i, x)$

Vu et al., Int. J. Quant. Chem., 2015; Rupp, Int. J. Quant. Chem., 2015

# The Gaussian process view




- Generalization of normal distribution
- Covariance = kernel

- Condition prior on data for posterior
- Variance as uncertainty estimate

# Representations

# Representations and Hilbert spaces

Numerical **representation** for arbitrary poly-atomic systems required

Corresponding Hilbert space as feature space

No unique solution

Example:

$(a, b) \mapsto (a^2, \sqrt{2}ab, b^2)$

$(a, b) \mapsto (a^2, ab, ab, b^2)$

Several state-of-the-art representations

symmetry functions

bispectrum, smooth overlap of atomic positions

moment tensor potentials

**many-body-tensor** and distribution-based representations

# Hilbert spaces for arbitrary atomistic systems

Representation = numerical encoding of atomistic system for accurate interpolation

**Requirements**:

(i) **Invariance** against transformations preserving the property
translation, rotation, homonuclear permutations

(ii) **Uniqueness**: different in property ⇒ different in representation
reconstructability

(iii) **Smoothness**: continuous, ideally differentiable

(iv) **Generality**: work with any atomistic system
molecules, periodic systems

(v) **Efficiency:** fast to compute, require few reference calculations

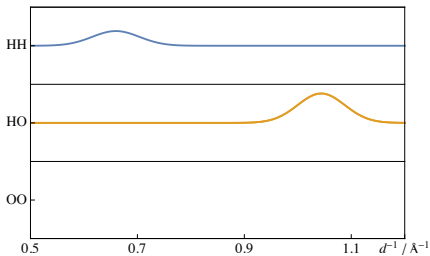
(vi) **Simplicity:** conceptually straightforward

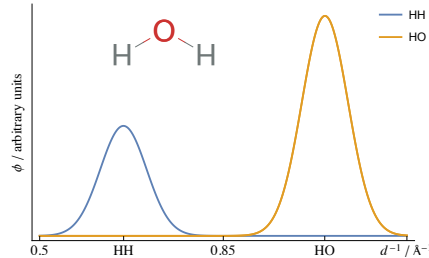
[Bartók et al, Phys Rev B, 2013; Shapeev, Multiscale Model Simul, 2016; Behler, J Chem Phys, 2011]

# The many-body tensor representation

Distribution of $k$-body terms stratified by chemical elements:

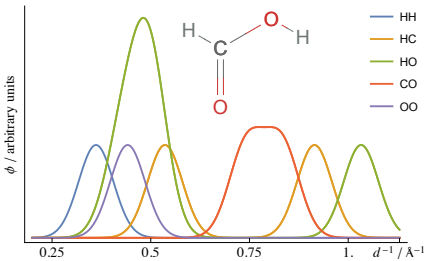
$$f_k(x, \boldsymbol{z}) = \sum_{\boldsymbol{i}=1}^{N_a} w_k(\boldsymbol{i})\, \mathcal{D}\big(x, g_k(\boldsymbol{i})\big) \prod_{j=1}^{k} C_{z_j, Z_{i_j}}$$

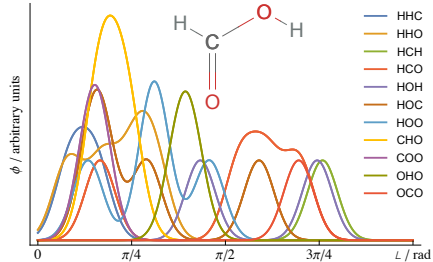


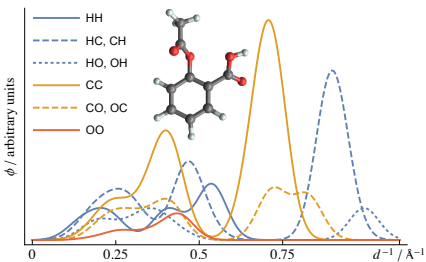



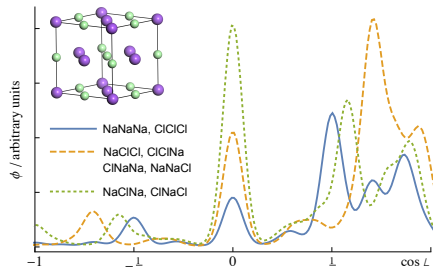

[Huo & Rupp, arXiv, 2017]

# Examples: molecules and bulk crystals

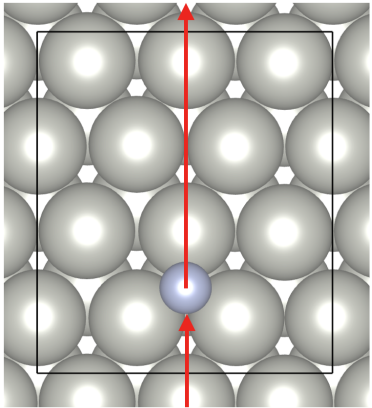


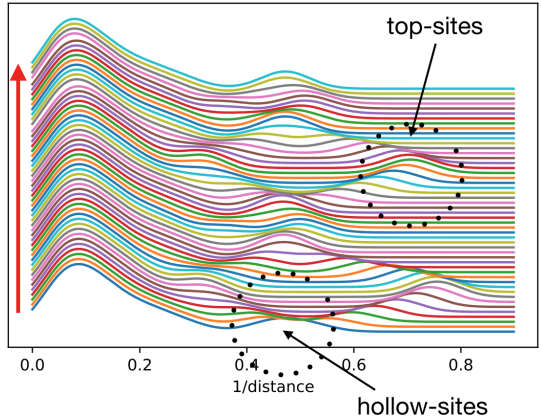



# Example: surfaces



**N atom on Pd(111) surface**

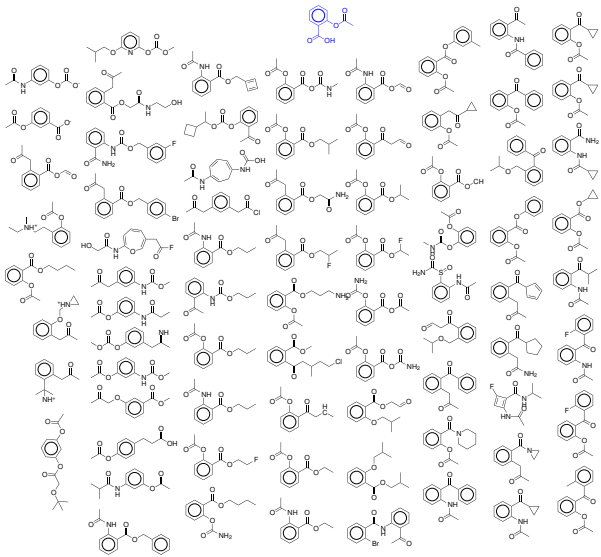



**MBTR of 2nd order (N-Pd)**

Yasunobu Ando, work in progress, 2018

# Energy predictions

# The combinatorial nature of chemical and materials space



Molecule space
Graph theory

Materials space
Group theory

Combinatorial explosion
Machine learning

Left: aspirin derivatives

# Learning potential energy surfaces

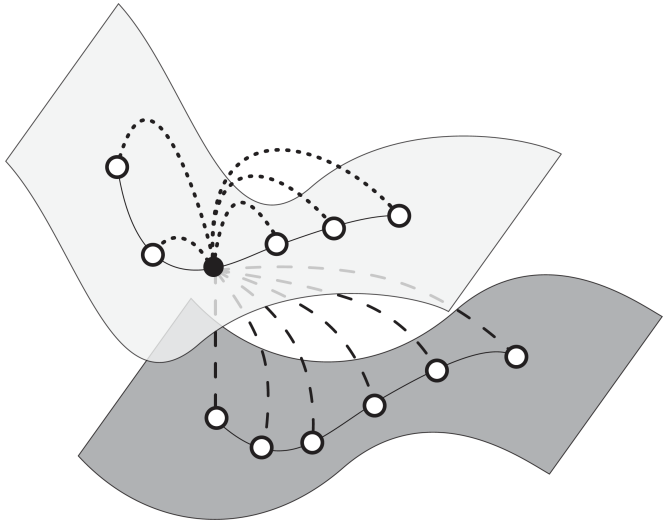
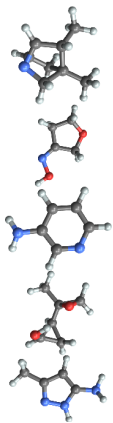


Figure: Chang & von Lilienfeld, CHIMIA, 2014

# Enthalpies of formation of small organic molecules



7 211 small organic molecules; C, N, O, S, Cl, saturated with H
Relaxed geometries, formation energies at DFT/PBE level of theory

| | | E / kcal mol$^{-1}$ | | $\alpha$ / Å$^3$ | |
|---|---|---|---|---|---|
| Representation | Kernel | RMSE | MAE | RMSE | MAE |
| CM [1] | Laplacian | 4.76 | 3.47 | 0.17 | 0.13 |
| BoB [2] | Laplacian | 2.86 | 1.79 | 0.12 | 0.09 |
| BAML [3] | Laplacian | 2.54 | 1.15 | 0.12 | 0.07 |
| SOAP [4] | REMatch | 1.61 | 0.92 | 0.07 | 0.05 |
| MBTR | Linear | 1.14 | 0.74 | 0.10 | 0.07 |
| MBTR | Gaussian | **0.97** | **0.60** | **0.06** | **0.04** |

[1] Rupp *et al.*, Phys. Rev. Lett., 2012; [2] Hansen *et al.*, J. Phys. Chem. Lett., 2015;
[3] Huang & von Lilienfeld, J. Chem. Phys., 2016; [4] De *et al.*, Phys. Chem. Chem. Phys., 2016

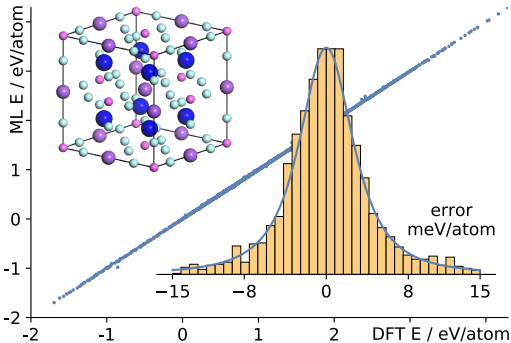
# Potential energy surfaces of organic molecules

Ab initio molecular dynamics; DFT/PBE, Tkatchenko-Scheffler van der Waals forces

| Kernel | DTNN | GDML Matérn | MBTR linear | | MBTR Gaussian | | |
|---|---|---|---|---|---|---|---|
| Molecule | MAE | MAE | RMSE | MAE | RMSE | MAE | [kcal/mol] |
| benzene | **0.04** | 0.07 | 0.06 | 0.05 | 0.05 | **0.04** | |
| uracil | – | 0.11 | 0.14 | 0.10 | 0.06 | **0.04** | |
| naphthalene | – | 0.12 | 0.15 | **0.11** | 0.15 | **0.11** | |
| aspirin | – | 0.27 | 0.26 | **0.18** | 0.32 | 0.22 | |
| salicylic acid | 0.50 | 0.12 | 0.17 | 0.12 | 0.11 | **0.08** | |
| malonaldehyde | 0.19 | 0.16 | 0.28 | 0.21 | 0.13 | **0.10** | |
| ethanol | – | 0.15 | 0.22 | 0.16 | 0.10 | **0.07** | |
| toluene | 0.18 | 0.12 | 0.16 | 0.12 | 0.15 | **0.11** | |

DTNN = Deep Tensor Neural Network, Schütt et al., Nat. Comm, 2017;
GDML = Gradient Domain Machine Learning, Chmiela et al., Sci. Adv., 2017

# Different decorations in elpasolite crystals




Elpasolites $ABC_2D_6$

Density functional theory

Relaxed geometry

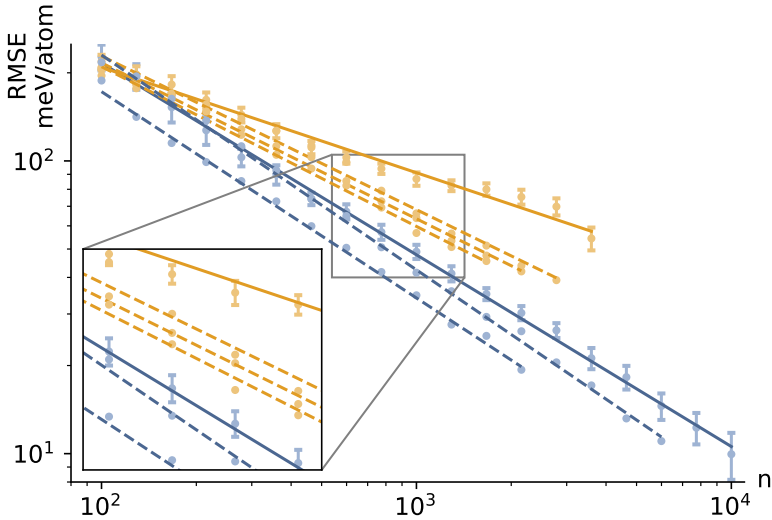
12 k materials, 12 elements

[Faber et al, Phys Rev Lett, 2016]

Prediction error 8 meV/atom (0.1 % of range)

Convergence threshold 10 meV/atom

# Learning curves

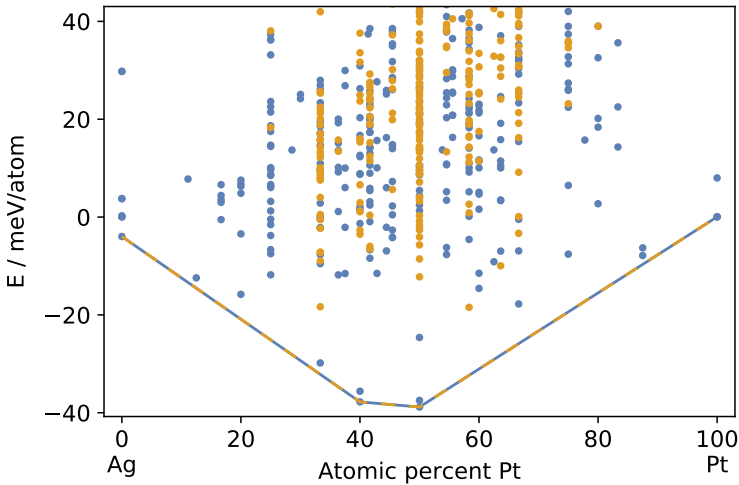



11 k $ABC_2D_6$ structures
12 chemical elements
DFT/PBE level of theory

4 611 $ABC_2$ structures
22 chemical elements
DFT/PBE level of theory

--- 1,2,3 chemical elements removed

# Phase diagrams of Pt-group/transition metal binary alloys




Identification of low-energy compositions in convex hull

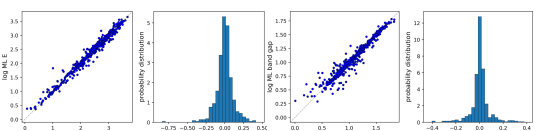
Active learning to exclude high-energy structures

Saves up to half the calculations
MAE 39 meV/atom

Hart, Curtarolo, Massalski, Levy, Phys Rev X, 2013

# The NOMAD 2018 Kaggle challenge

Christopher Sutton (Fritz Haber Institute, Max Planck Society)


Al-Ga-In sesquioxides
DFT, 3 k structures

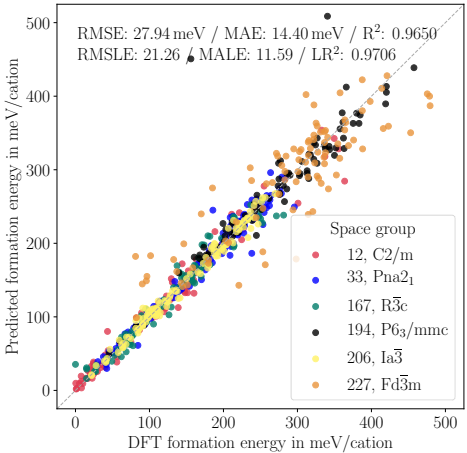
Δ-learning
Geometry from Vegard's rule
Enthalpy of formation and band gaps relaxed

Errors:
27 meV/atom (energy)
180 meV (band gap)

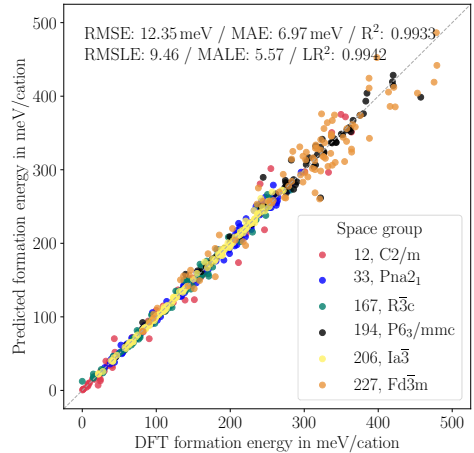
www.kaggle.com/c/nomad2018-predict-transparent-conductors

# The NOMAD 2018 Kaggle challenge

Christopher Sutton (Fritz Haber Institute, Max Planck Society)



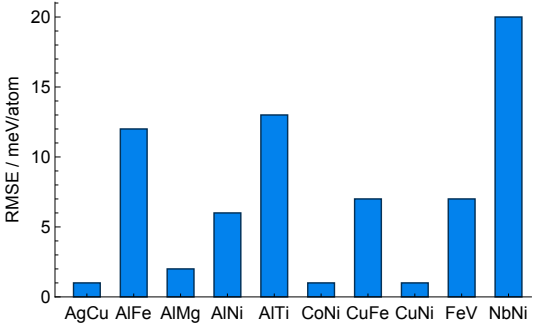

Vegard's rule geometries

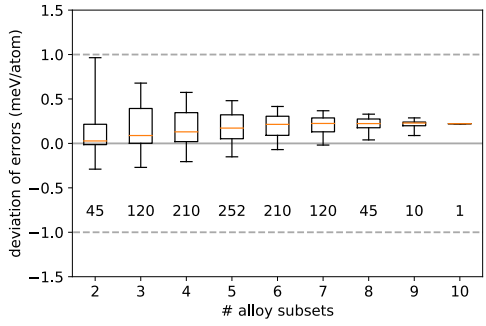



Relaxed geometries

# Binary alloys

With Gus Hart, Chandramouli Nyshadham, Brayden Bekker (Brigham Young University)




Binary alloys

$10 \times 1\,\mathrm{k} + 0.6\,\mathrm{k}$ structures

DFT formation enthalpy

Vegard's rule geometries

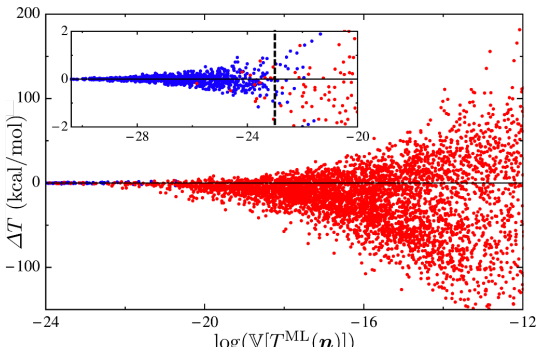
fcc, bcc, hcp lattices

all structures up to 8 atoms

# Reliability

*"It is not the estimate that matters so much, as the degree of confidence with the opinion"*

N. Taleb, 2004




Uncertainty estimates

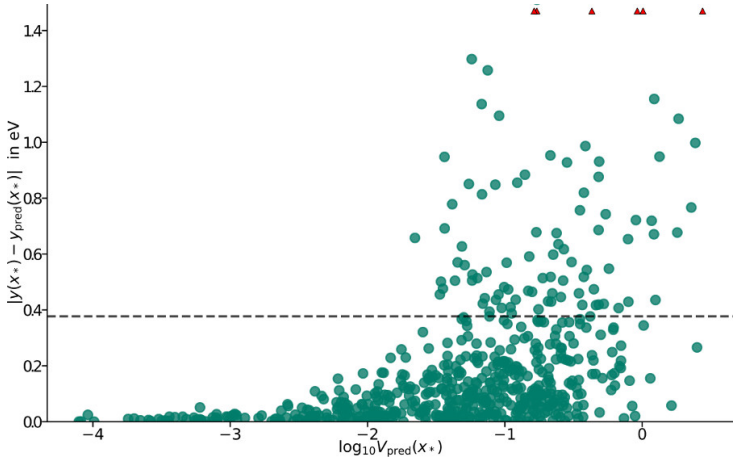
Extrapolation versus interpolation

Domain of applicability

Left: Logarithmized predictive variance versus signed error for energy predictions

[Snyder et al., Phys. Rev. Lett., 2012]

# Reliability




RMSE 380 meV, MAE 200 meV, $R^2 = 0.9916$
RMSE 12 meV/atom, MAE 6 meV/atom, $R^2 = 0.9854$

# Interpretability


Explain individual predictions

Right for the right reason?

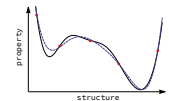
Sensitivity analysis

Local gradients explain variation, but not the "why"

How to quantify interpretability?

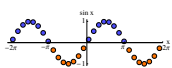
How to interpret basis set expansions?

# Summary



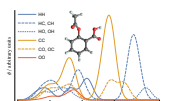

Interpolation of ab initio simulations

Accurate and fast, via machine learning



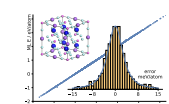

Learning with kernels

Kernel trick: implicit transformations




Representations

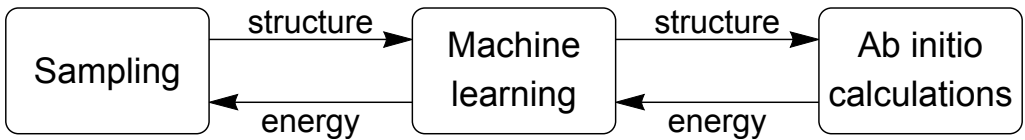
Many-body tensor representation




Energy predictions

Molecules, crystals, surfaces

# Outlook: towards crystal structure prediction

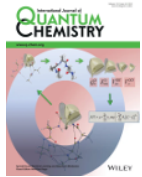



comprehensive fast accurate

## Tutorial

M. Rupp: Machine Learning for Quantum Mechanics in a Nutshell, *Int. J. Quant. Chem.* 15(16): 1058–1073, 2015.

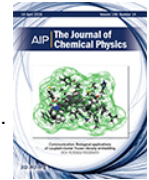
DOI 10.1002/qua.24954




## Special Issue

Rupp, von Lilienfeld, Burke: Guest Editorial: Special Topic on Data-enabled Theoretical Chemistry, *J. Chem. Phys.* 148(24): 241401, 2018.

DOI 10.1063/1.5043213




## Code

Efficient C++ implementation with Python bindings available as part of my `qmmlpack` package.

To appear this year.


## Datasets

Public datasets of ab initio calculations for molecules, solids and liquids available.

qmml.org
nomad-coe.eu

# Acknowledgements


**Group**

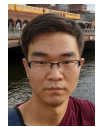
Current

Marcel Langer (PhD student)

Alex Goeßmann (PhD student)

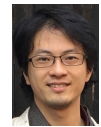
Fabio Hernandez-Hernandez (visitor)

Former



Haoyan Huo (visitor)

Lucas Deecke (student)



Yasunobu Ando (visitor)

**Funding**