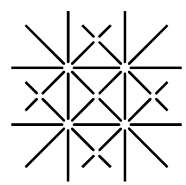
Quantum Mechanics / Machine Learning Models

Matthias Rupp

University of Basel
Department of Chemistry
matthias.rupp@unibas.ch

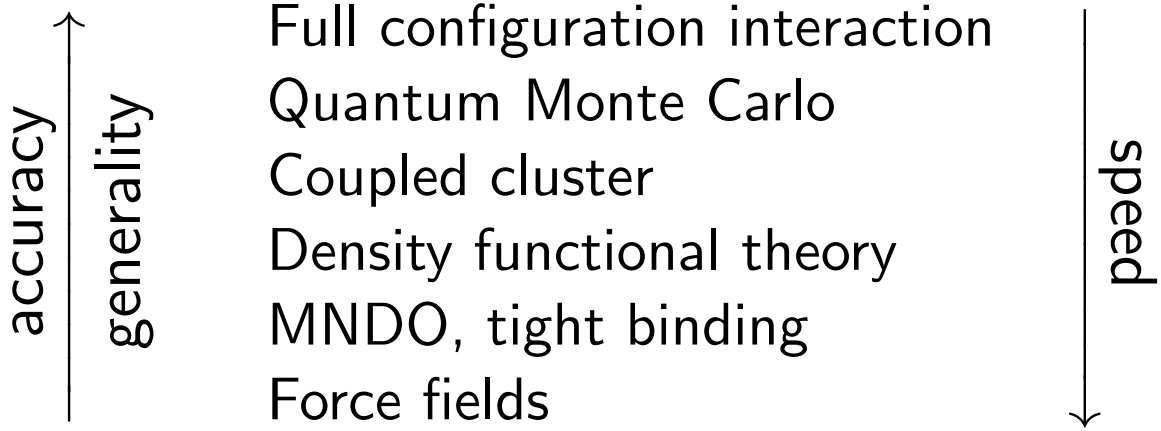
IPAM Summer School on Electronic Structure Theory,
Los Angeles, California, July 31, 2014



Outline

Introduction	What are QM/ML models?
Machine learning	How does ML work?
Applications	What can be done with them?
Pitfalls	What can go wrong?
Demonstration	Worked example

Approximations



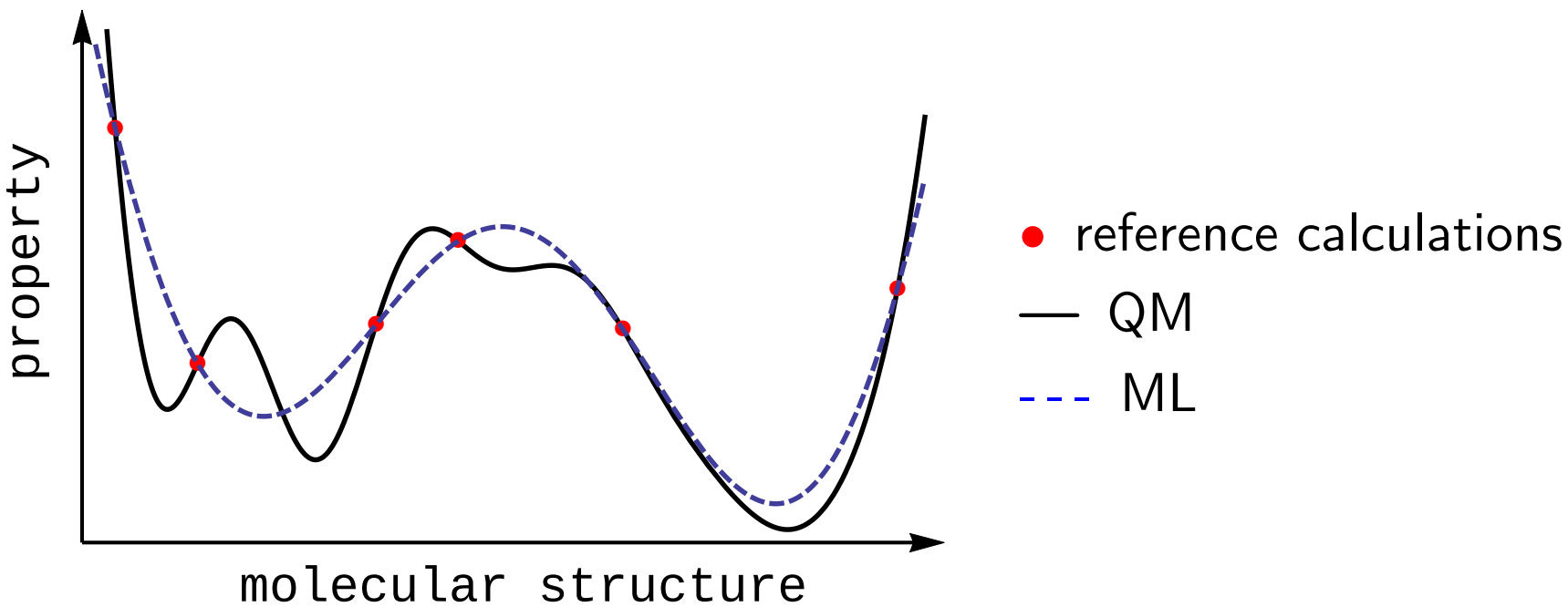
QM/ML models:

> The accuracy of quantum chemistry,
> at the speed of machine learning

QM/ML models

Exploit redundancy in a series of QM calculations

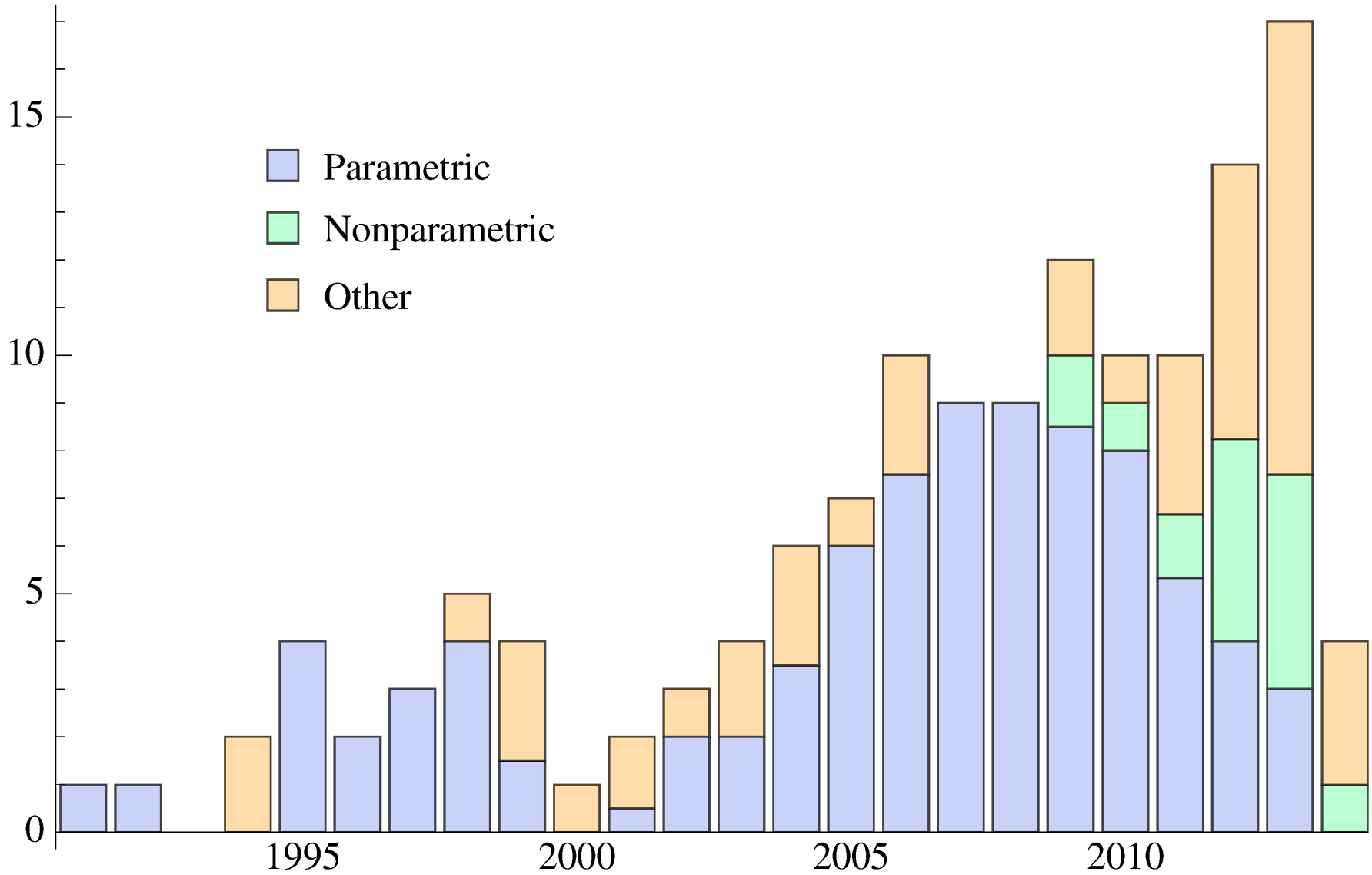
- QM/ML = quantum mechanics + machine learning
- Interpolate between QM calculations using ML
- Smoothness assumption (regularization)



Relationship to other models

Quantum chemistry	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
Slow	Fast	In between
Small systems	Large systems	Large systems

History and literature



What is machine learning?

- Interpolation
- Algorithmic search for patterns in data
- Inference from known samples to new ones
- Regularity, information content
- Data-driven approach
- Empirical but principled

Hastie, Tibshirani, Friedman, The Elements of Statistical Learning, Springer, 2nd ed., 2009.
Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

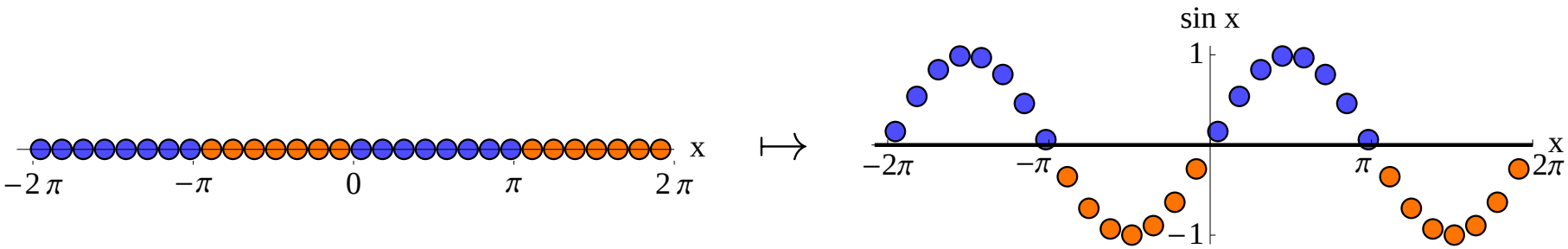
Machine learning algorithms

- Artificial neural networks (Haykin, 2008; Montavon et al (ed.), 2012)
- Kernel ridge regression (Hastie, Tibshirani, Friedman, 2009)
- Gaussian process regression (Rasmussen & Williams, 2006)
- Support vector machines (Cristianini & Shawe-Taylor, 2000)
- Principal component analysis (Jolliffe, 2004)
- Symbolic regression (Schmidt, Lipson, *Science*, 2009)
- Many others...

Kernel learning

Idea:

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



Input space $\mathcal{X}$ $\xrightarrow{\phi}$ Feature space $\mathcal{H}$

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

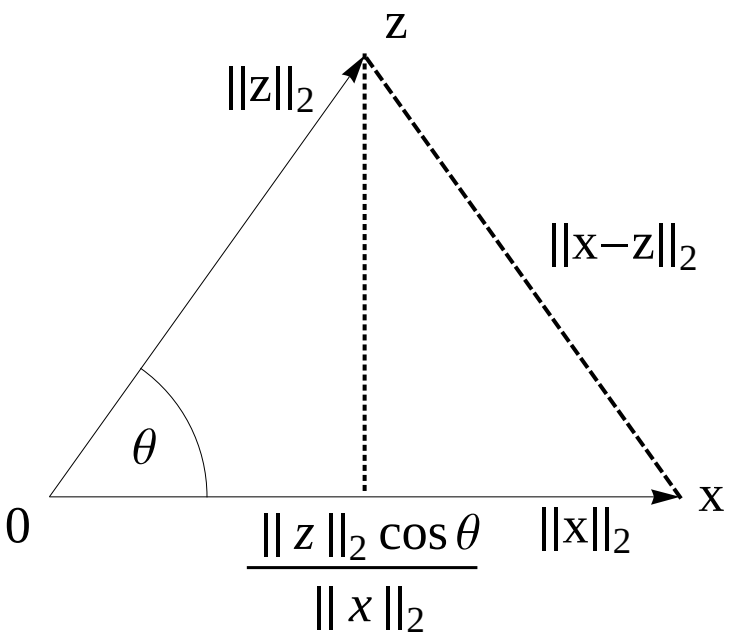
Schölkopf, Smola: Learning with Kernels, 2002; Hofmann et al.: *Ann. Stat.* 36, 1171, 2008.

Kernels

Kernels correspond to inner products.

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 $\phi : \mathcal{X} \to \mathcal{H}$.

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



- Well characterized function class
- Closure properties
- Access data only by $\mathbf{K}_{ij} = k(x_i, x_j)$
- $\mathcal{X}$ can be any non-empty set
- Examples:
  Linear kernel $\langle \mathbf{x}, \mathbf{z} \rangle$
  Gaussian kernel $\exp\left(-\frac{||\mathbf{x}-\mathbf{z}||^2}{2\sigma^2}\right)$

Kernel ridge regression

- Regularized form of ordinary regression
- Regularization prevents over-fitting by penalizing large coefficients
- Use of kernels for non-linearity

Solution has form

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

Coefficients α are obtained by solving

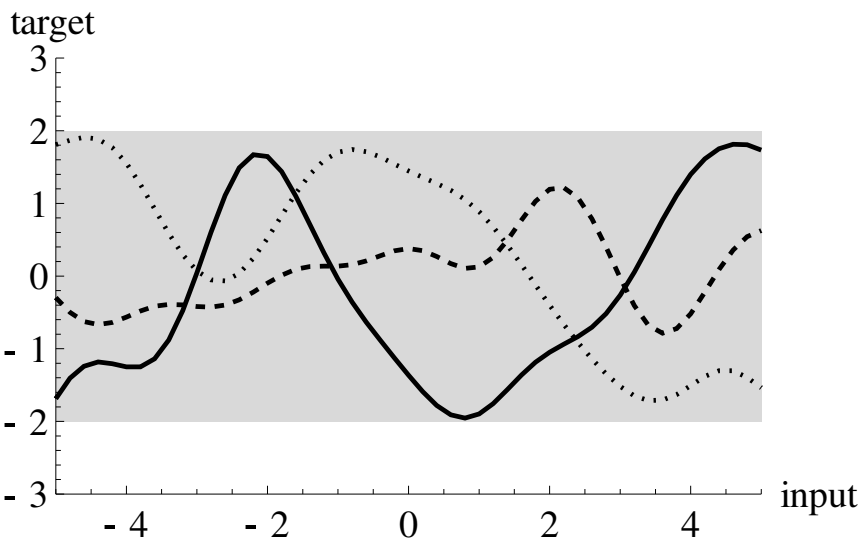
$$\sum_{i=1}^{n} \big(f(\mathbf{x}_i) - y_i\big)^2 + \lambda \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha},$$

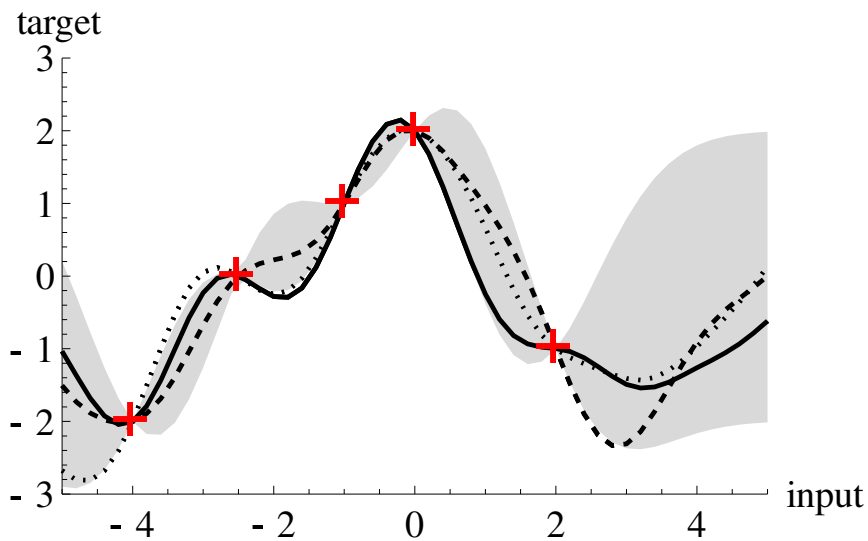
which has solution

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

Gaussian process regression

- Generalization of multivariate normal distribution to functions
- Determined by mean function and covariance function = kernel
- Conditioning of prior on training data yields posterior distribution
- Variance as confidence estimates for predictions





- Intuitively: Place a basis function on each training datum $\mathbf{x}_i$
- Solution has form $f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i k(\mathbf{x}_i, \mathbf{x})$

Rasmussen, Williams: Gaussian Processes for Machine Learning, MIT Press, 2006.

Applications

- Potential energy surfaces (Handley & Behler, *Eur. Phys. J. B* 152, 2014)
- Molecular and materials properties (Rupp et al, *PRL* 058301, 2012)
- Polarizabilities (Kandathil et al, *JCC* 1850, 2013)
- Density functional theory (Snyder et al, *JCP* 224104, 2013)
- Transition state theory dividing surfaces (Pozun et al, *JCP* 174101, 2012)
- Materials properties (Pilania et al, *Sci. Rep.* 2810, 2013; Ghiringhelli et al, 2014)
- Transmission coefficients (Lopez-Bezanilla & von Lilienfeld, *PRB* 235411, 2014)
- Collective variables (Rohrdanz et al, *JCP* 124116, 2011)
- Others (e.g., nuclear physics, cheminformatics)

Gaussian approximation potentials

- Gaussian process regression
- Molecular dynamics
- Partitioned energies

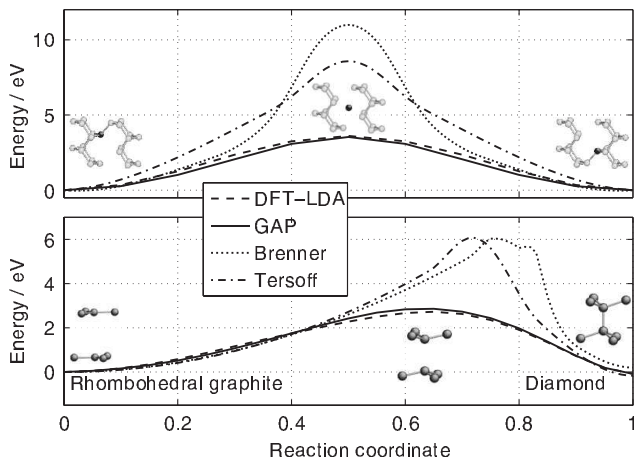
- Representation:

  Local density

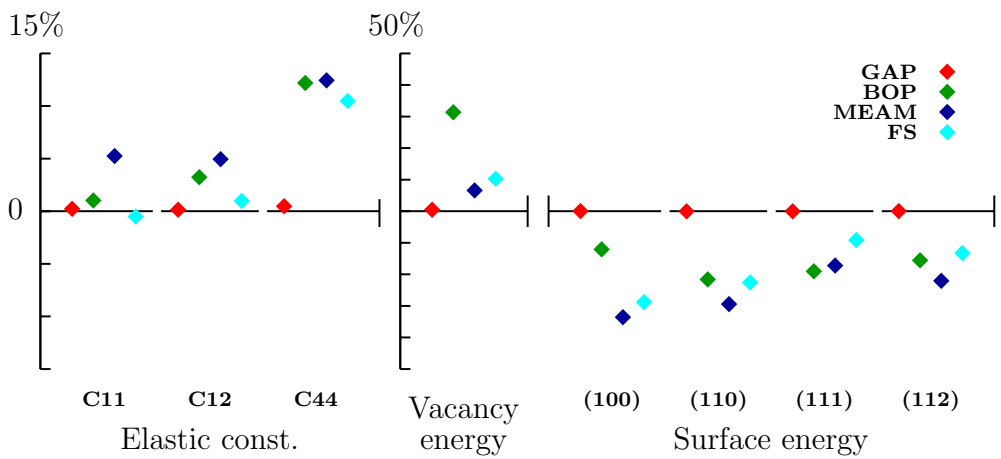
  Projection to 4d sphere

  Hyperspherical harmonics

  Bispectrum



Transition path energies



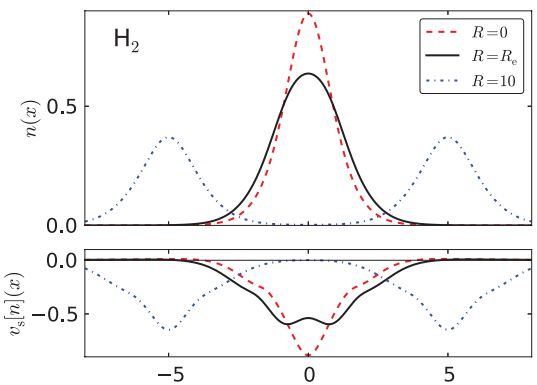
Errors on properties of Tungsten

Bartók, Csányi et al, *Phys Rev Lett* 104: 136403, 2010. Szlachta et al, *arXiv* 1405.4370, 2014.

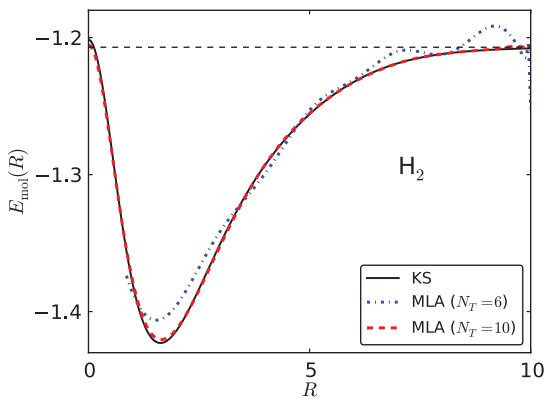
Density functional theory

Learning the map from electron density to kinetic energy

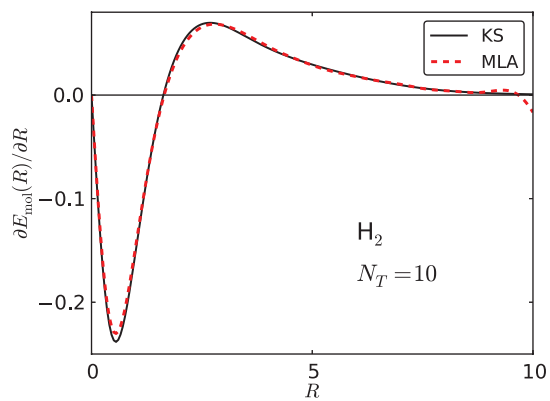
- Orbital-free DFT
- 1D toy system
- DFT/LDA as reference
- Error decays to zero
- Self-consistent densities
- Bond breaking and formation



H_2 potential



H_2 binding curve

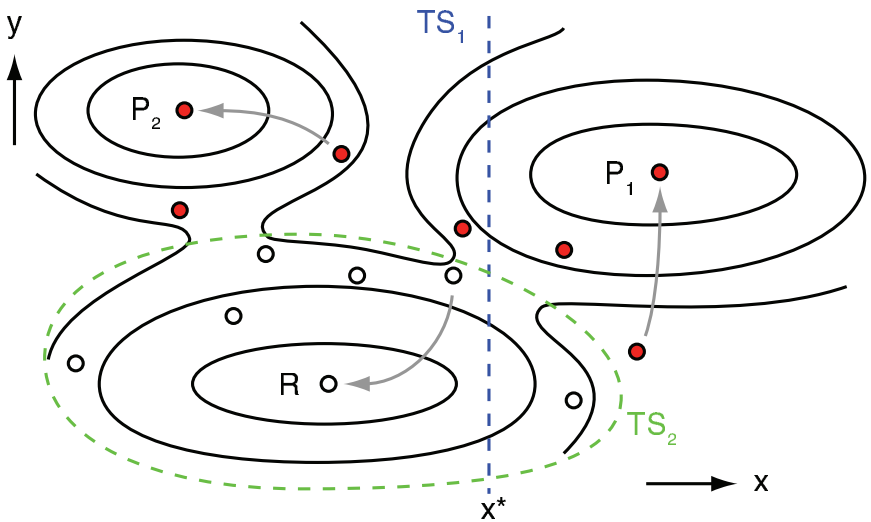


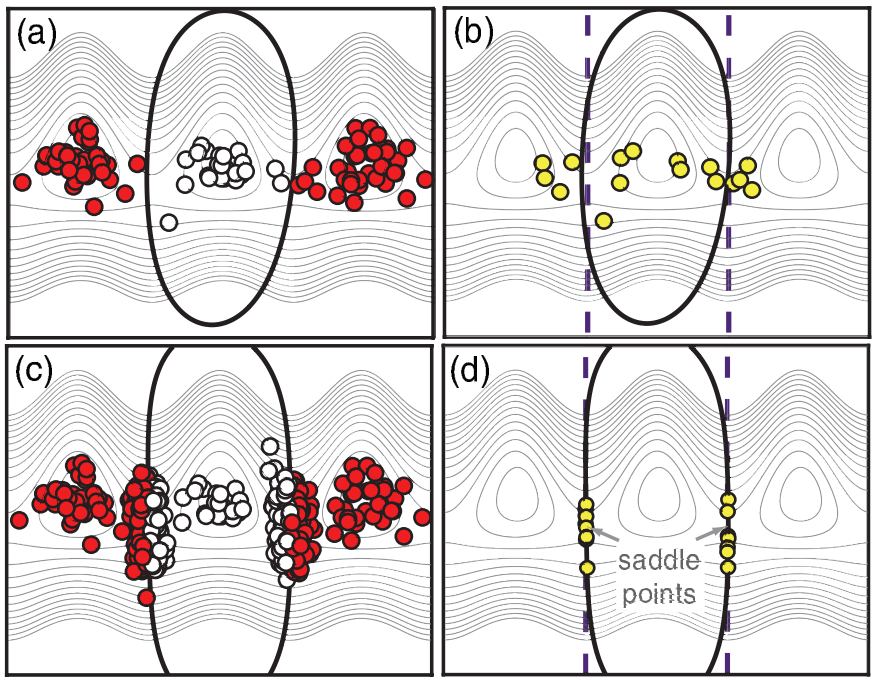
H_2 forces

Snyder et al, *Phys Rev Lett* 108: 253002, 2012. Snyder et al, *J Chem Phys* 139: 224104, 2013.

Transition state theory

- Characterization of dividing surfaces
- Support vector machines
- No prior information required
- Iteratively refined by biased sampling along dividing surface





Pozun et al, *J. Chem. Phys.* 136: 174101, 2012.

Molecular properties

Data

7 165 small organic molecules

DFT PBE0 atomization energies

Representation

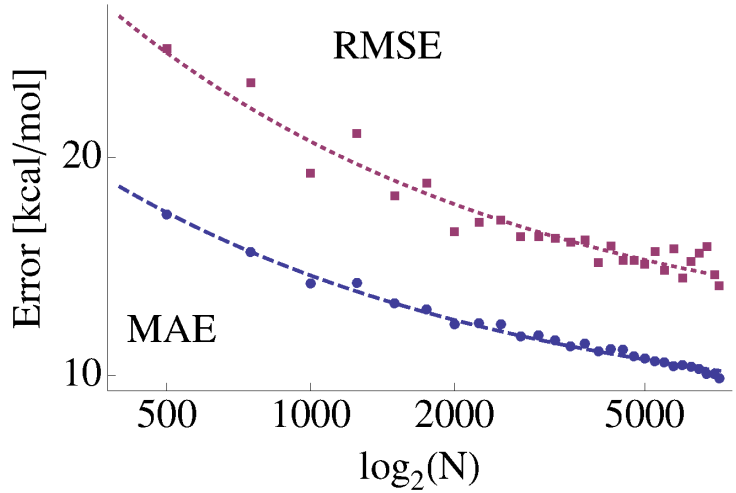
$$\mathbf{M}_{IJ} = \begin{cases} \frac{1}{2} Z_I^{2.4} & \text{if } I = J \\ \frac{Z_I Z_J}{\|\mathbf{R}_I - \mathbf{R}_J\|} & \text{if } I \neq J \end{cases}$$

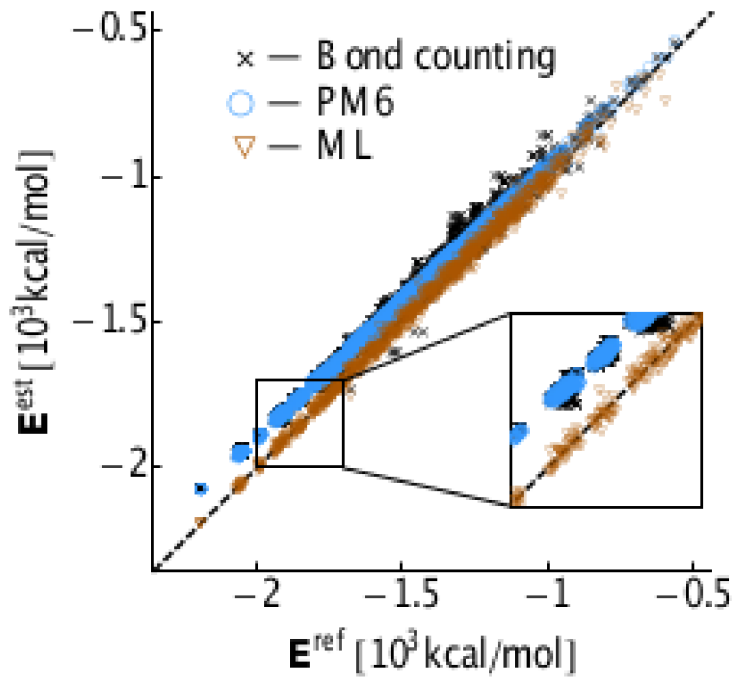
Model

$$E^{\mathrm{ML}}(\mathbf{M}) = \sum_{I=1}^{N} \alpha_i k(\mathbf{M}_i, \mathbf{M})$$

$$k(\mathbf{M}_i, \mathbf{M}) = \exp\left(-\frac{\|\mathbf{M}_i - \mathbf{M}\|^2}{2\sigma^2}\right)$$

$$\boldsymbol{\alpha} = (\mathbf{K} - \lambda \mathbf{I})^{-1} \mathbf{E}^{\mathrm{QM}}$$



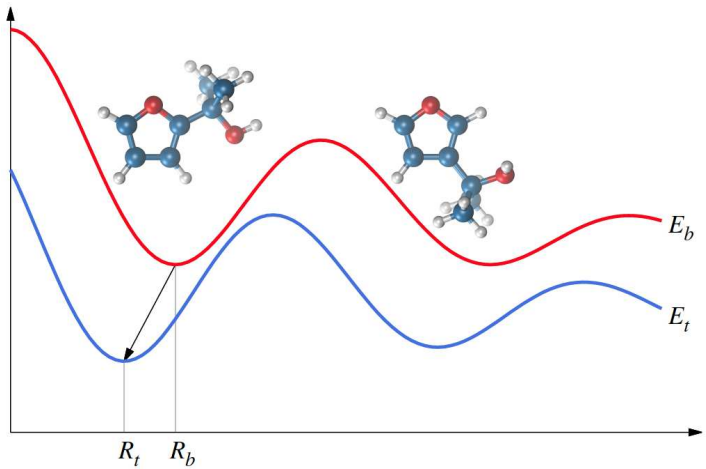


Rupp et al, *Phys Rev Lett* 108: 058301, 2012

Δ-learning: Setup

Learning the error between different levels of theory

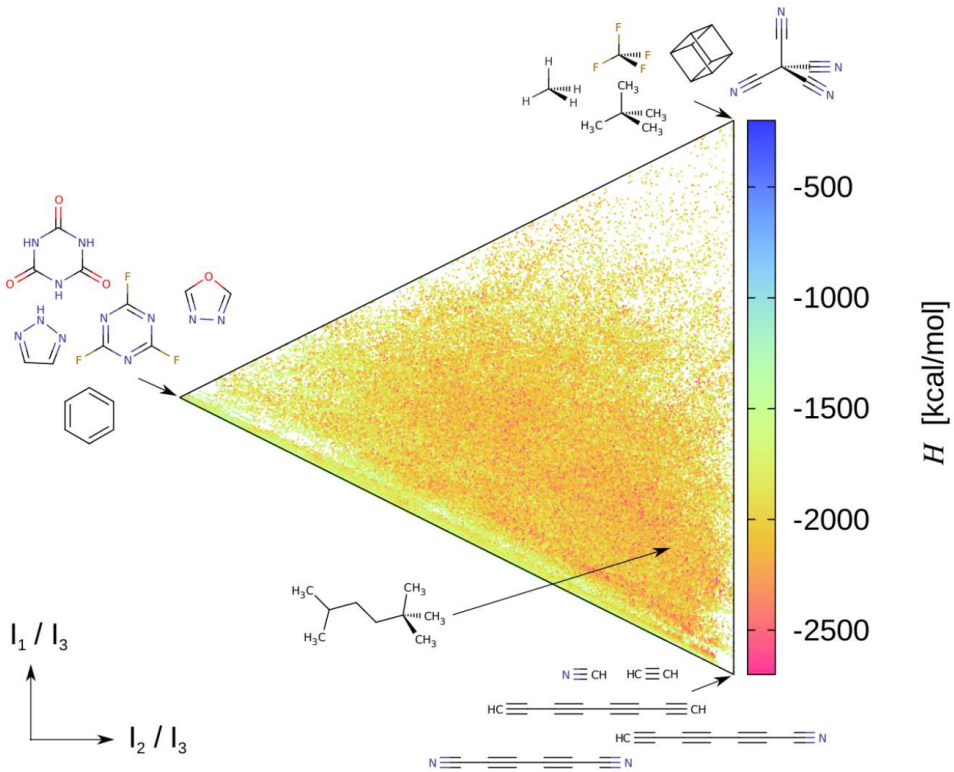
- Learn corrections to a baseline method (Δ = reference - baseline)
- Augmenting legacy QM methods
- Puts physics into QM/ML model
- Examples: $\Delta^{\text{B3LYP}}_{\text{PM7}}$, $\Delta^{\text{G4MP2}}_{\text{PM7}}$, $\Delta^{\text{CCSD(T)}}_{\text{HF}}$

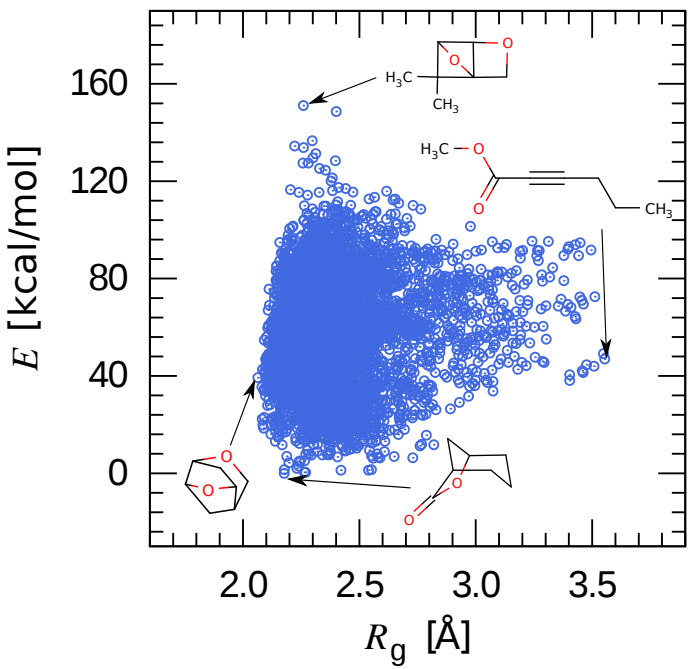


Ramakrishnan, Dral, Rupp, von Lilienfeld, *submitted*, 2014.

Δ-learning: Data

Learning the error between different levels of theory





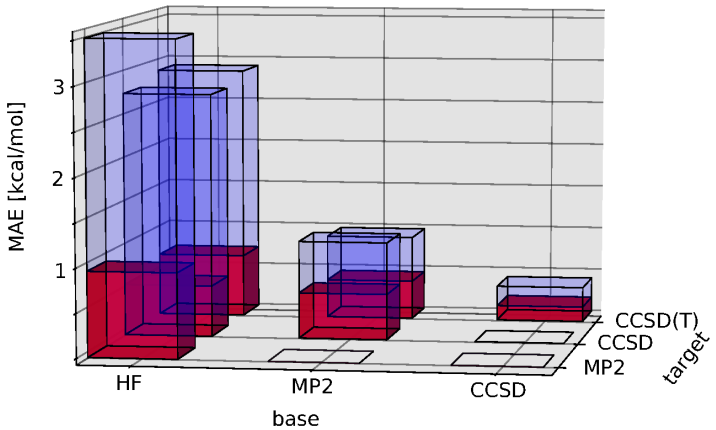
134 k small organic molecules
PM7, DFT B3LYP

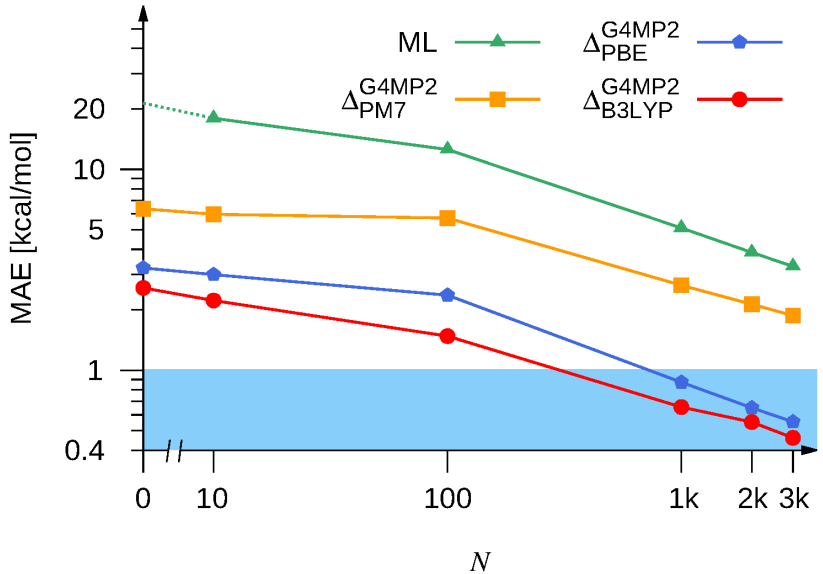
6 k const. isomers of $C_7H_{10}O_2$
PM7, G4MP2; HF, MP2, CCSD(T)

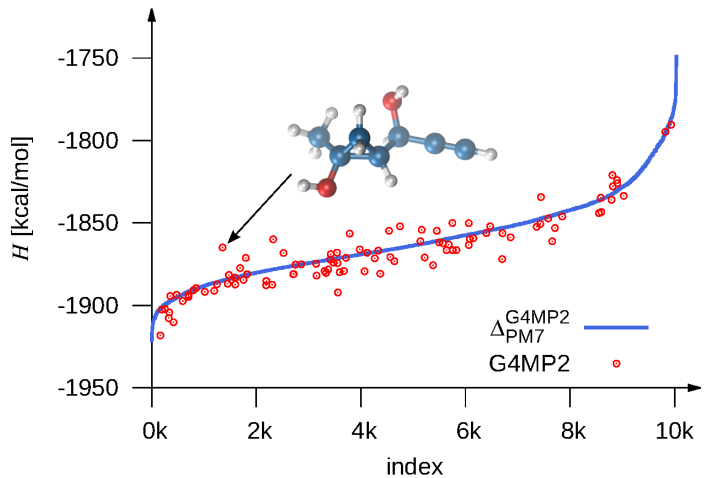
Ramakrishnan et al, *submitted*, 2014. Ramakrishnan et al, *Nat Scientific Data*, accepted, 2014.

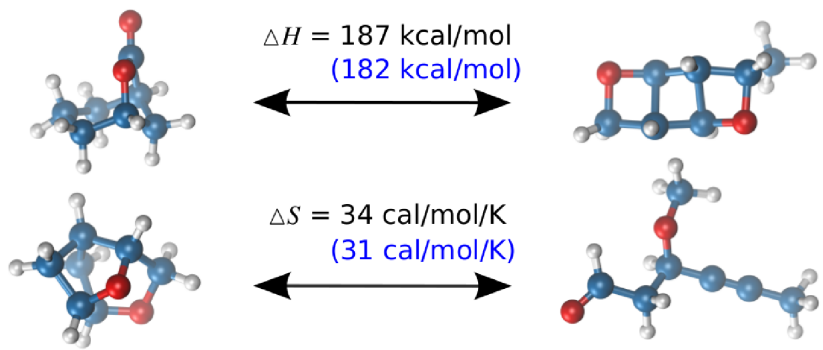
Δ-learning: Results

Learning the error between different levels of theory





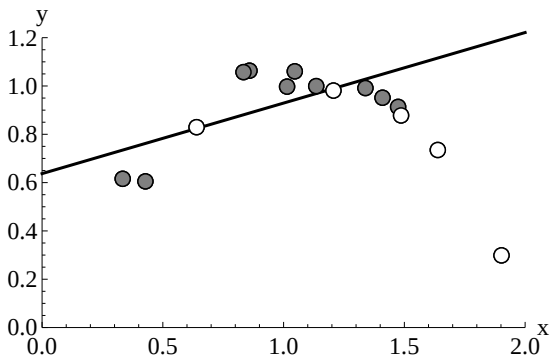




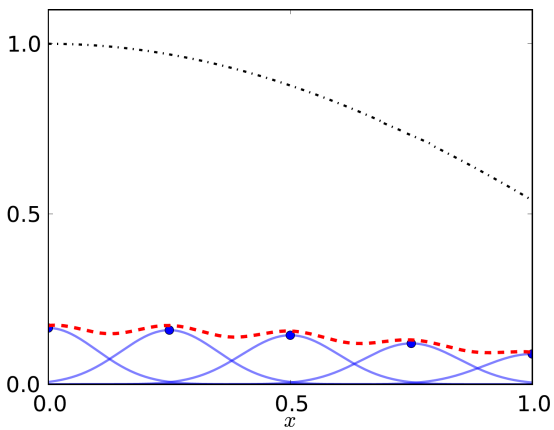
Ramakrishnan, Dral, Rupp, von Lilienfeld, *submitted*, 2014

Overfitting: Model complexity and generalization error

Underfitting

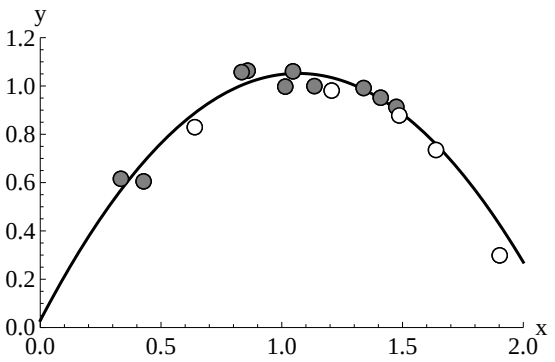


0.123 / 0.443

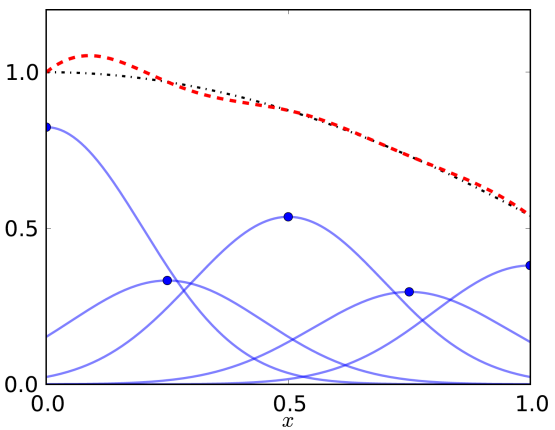


λ too large

Fitting

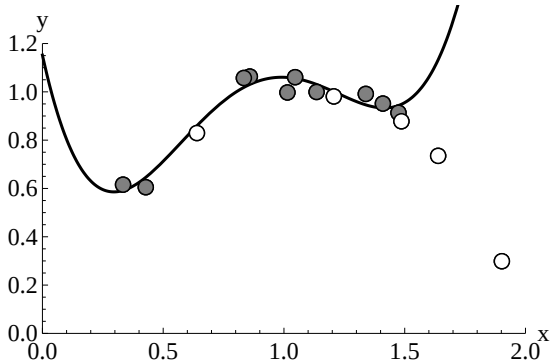


0.044 / 0.068

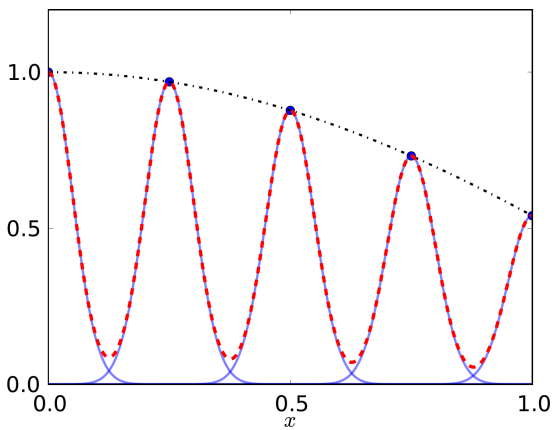


λ right

Overfitting



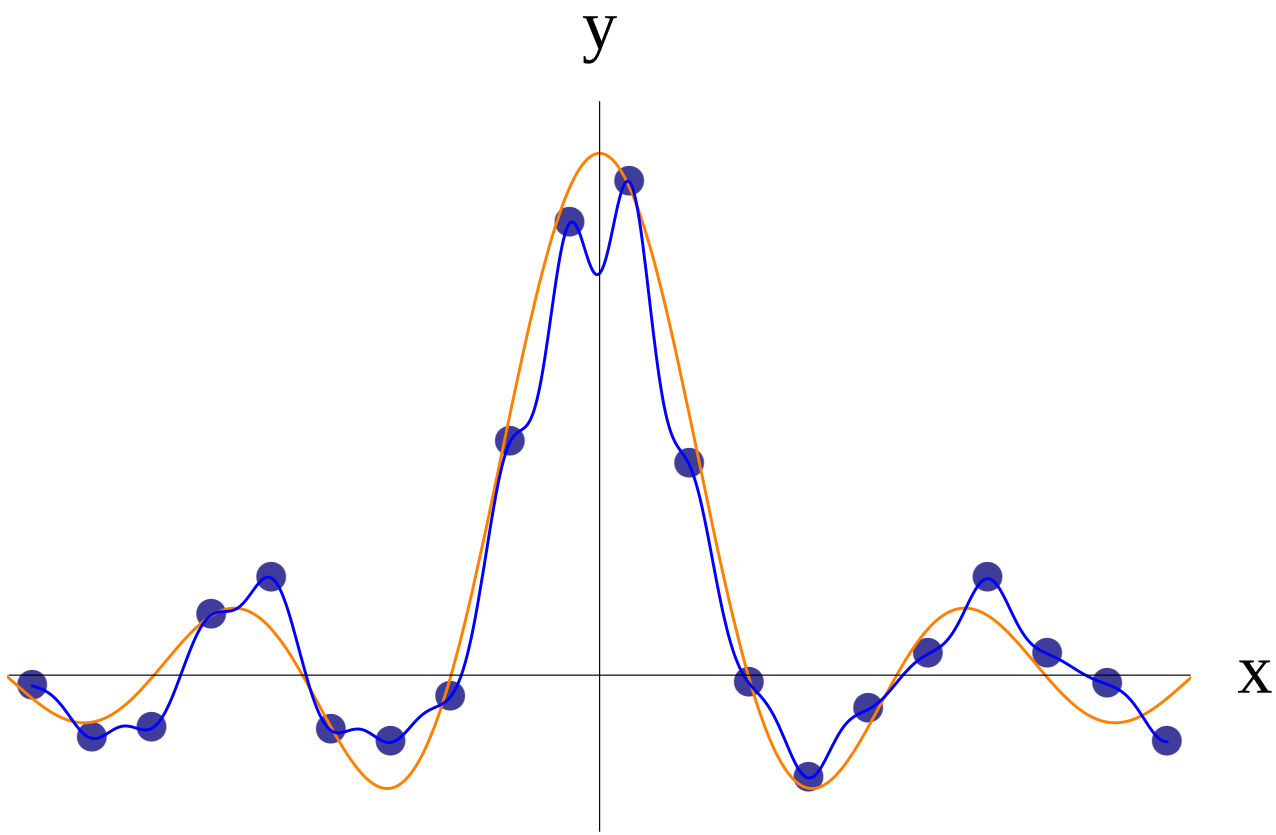
0.036 / 0.939



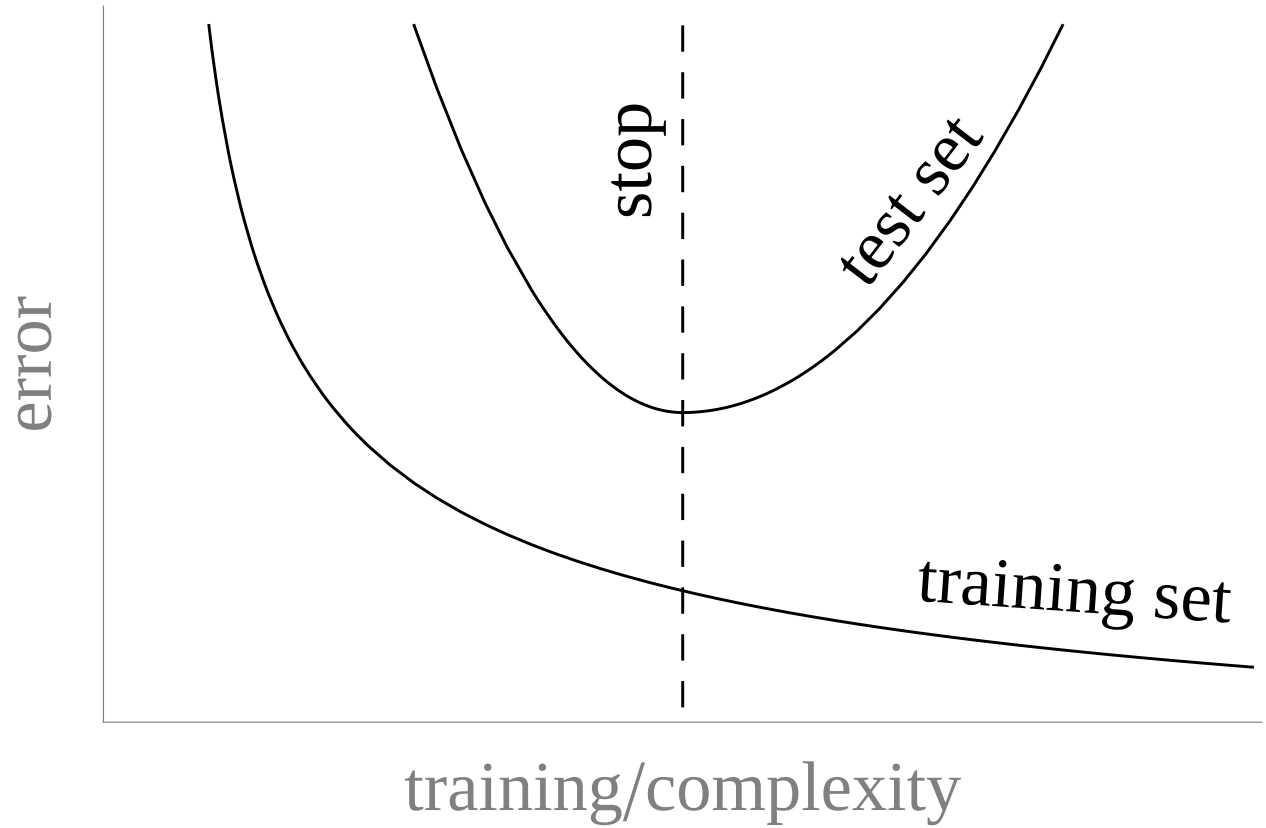
λ too small

Rupp, PhD thesis, 2009. Li Li et al, *submitted*.

Overfitting: Another example



Overfitting: Early stopping rule



Validation

Golden rule

Training must never use validation data

Example 1: overfitting

× train on all data, predict all data

✓ split data, train, predict

Example 2: centering

× center data, split data, train & predict

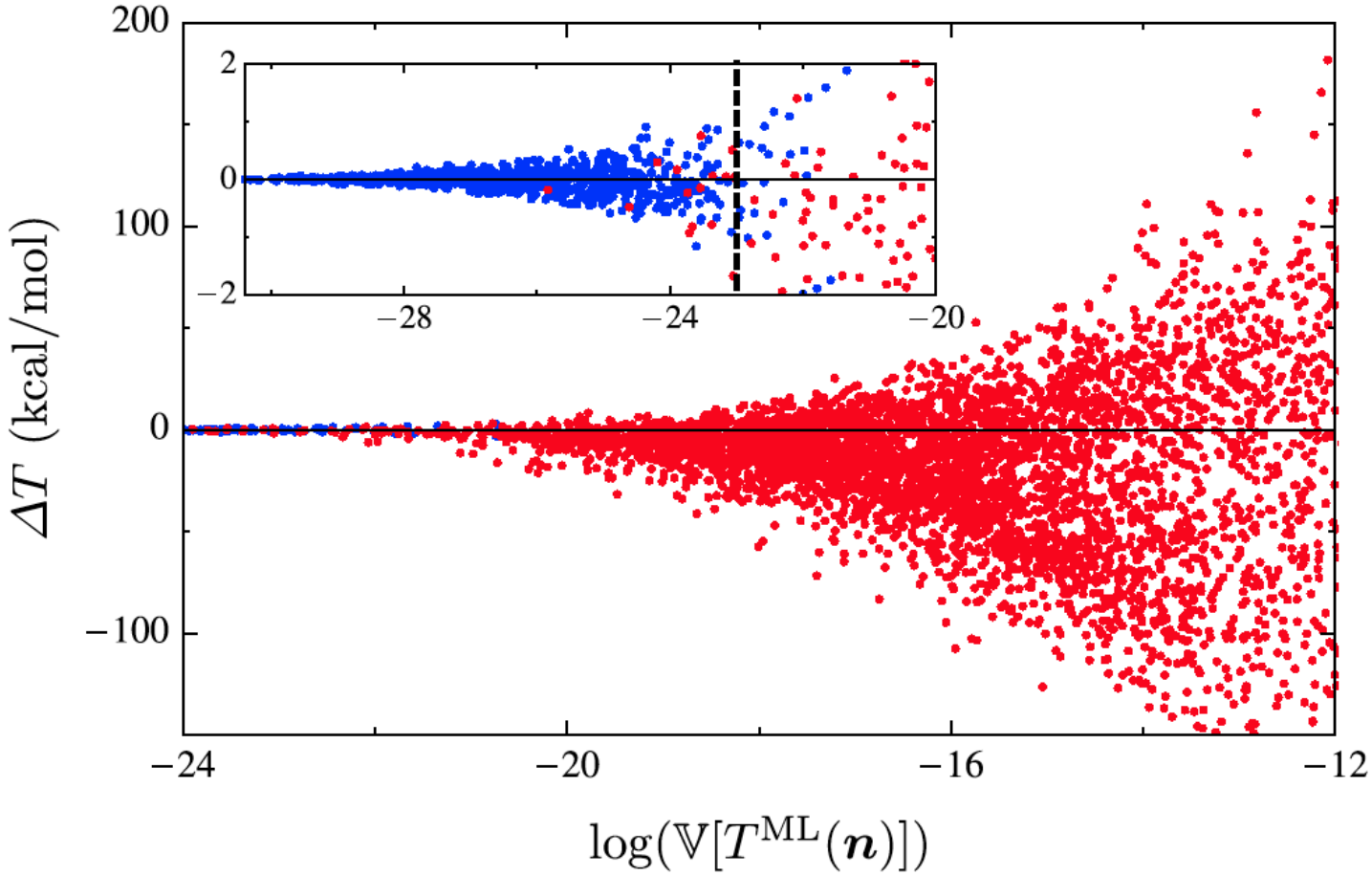
✓ split data, center training set, train, center test set, predict

Example 3: cross-validation with feature selection

× feature selection, cross-validation

✓ feature selection for each split of cross-validation

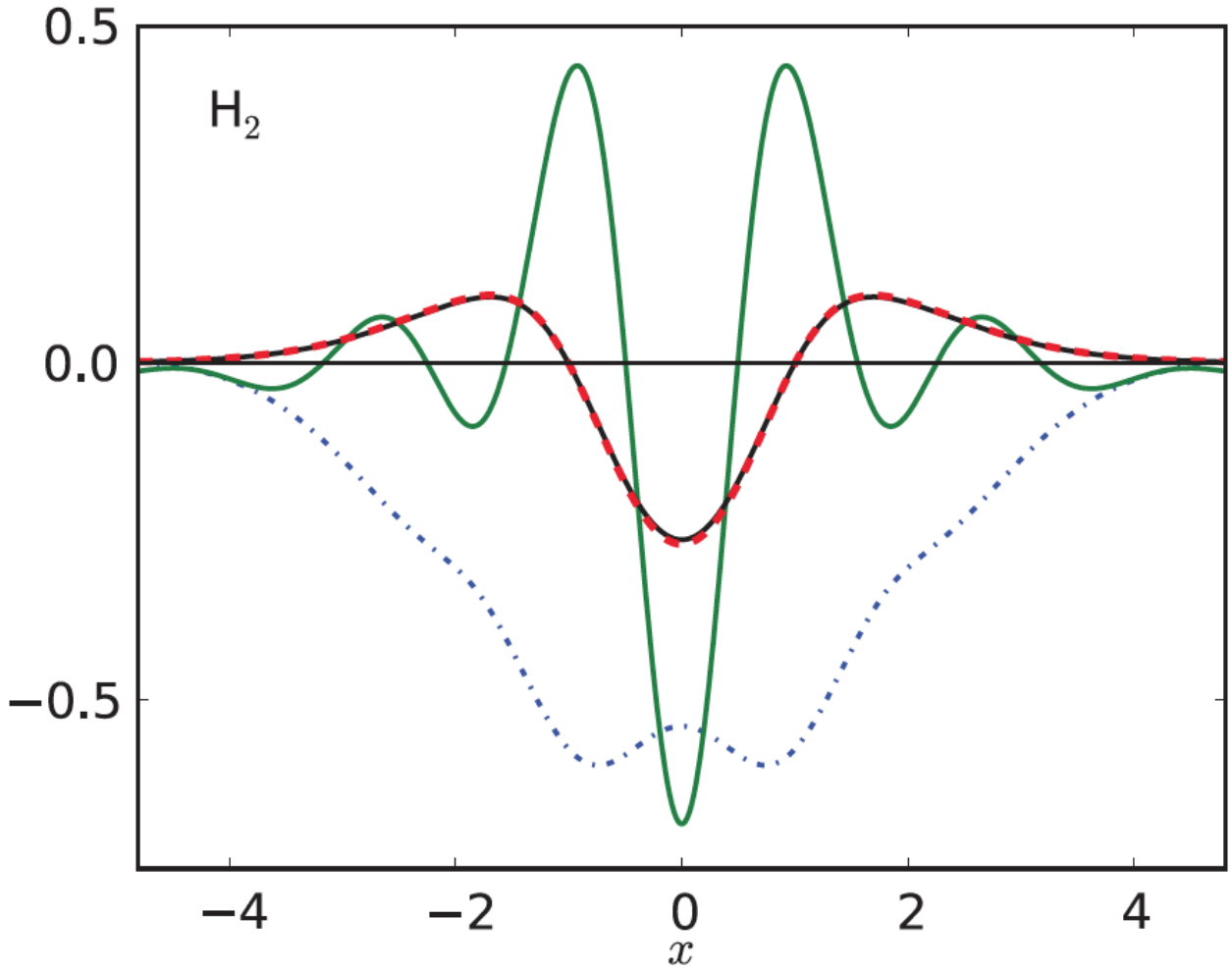
Reliability of predictions



Predictive variance of Gaussian process regression model

Snyder et al, *Phys. Rev. Lett.* 108: 253002, 2012.

Gradients



Functional derivative of model as-is and projected on training data

Snyder et al, *J. Chem. Phys.* 139: 224104, 2013.

Summary

- QM/ML models combine quantum chemistry with machine learning by interpolating between reference QM calculations
- The concept is broadly applicable

Live demonstration

PRL **108,** 058301 (2012) PHYSICAL REVIEW LETTERS week ending 3 FEBRUARY 2012

Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning

Matthias Rupp,[1,2] Alexandre Tkatchenko,[3,2] Klaus-Robert Müller,[1,2] and O. Anatole von Lilienfeld[4,2,*]

[1]*Machine Learning Group, Technical University of Berlin, Franklinstr 28/29, 10587 Berlin, Germany*
[2]*Institute of Pure and Applied Mathematics, University of California Los Angeles, Los Angeles, California 90095, USA*
[3]*Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin, Germany*
[4]*Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, Illinois 60439, USA*
(Received 15 June 2011; published 31 January 2012)

We introduce a machine learning model to predict atomization energies of a diverse set of organic molecules, based on nuclear charges and atomic positions only. The problem of solving the molecular Schrödinger equation is mapped onto a nonlinear statistical regression problem of reduced complexity. Regression models are trained on and compared to atomization energies computed with hybrid density-functional theory. Cross validation over more than seven thousand organic molecules yields a mean absolute error of ~10 kcal/mol. Applicability is demonstrated for the prediction of molecular atomization potential energy curves.

Acknowledgements

The Basel team


von Lilienfeld


Ramakrishnan


Chang

Collaborators

M.R. Bauer, F. Biegler, L. Blooston, F.M. Boeckler, F. Brockherde, K. Burke, P. Dral, S. Fazli, G. Folkers, V. Gobre, K. Hansen, G. Henkelman, J. Huang, A. Knoll, A. Lange, L. Li, A. Lopez-Bezanilla, G. Montavon, K.-R. Müller, I.M. Pelaschier, Z. Pozun, M. Reutlinger, M. Scheffler, G. Schneider, D. Sheppard, J.C. Snyder, A. Tkatchenko, S. Varma, A. Vazquez-Mayagoitia, R. Wilcken, A. Ziehe

Institutions

IPAM * **EU FP7** * **DFG** * **SNSF**