Quantum Mechanics / Machine Learning Models

Matthias Rupp

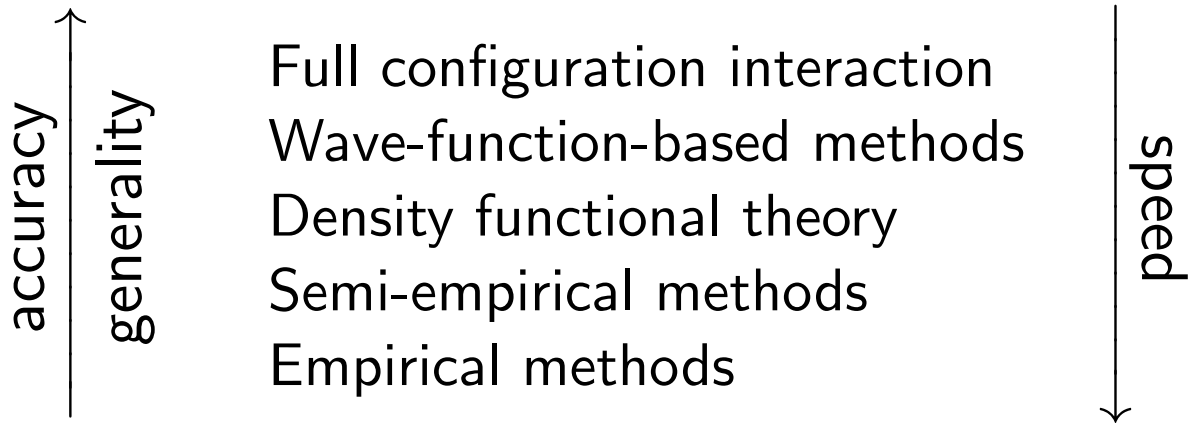
Fritz Haber Institute of the Max Planck Society
matthias.rupp@fhi-berlin.mpg.de

Hands-on Workshop Density-Functional Theory and Beyond
Berlin, Germany, July 13–23, 2015

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?
Tutorial	Worked example

Approximations



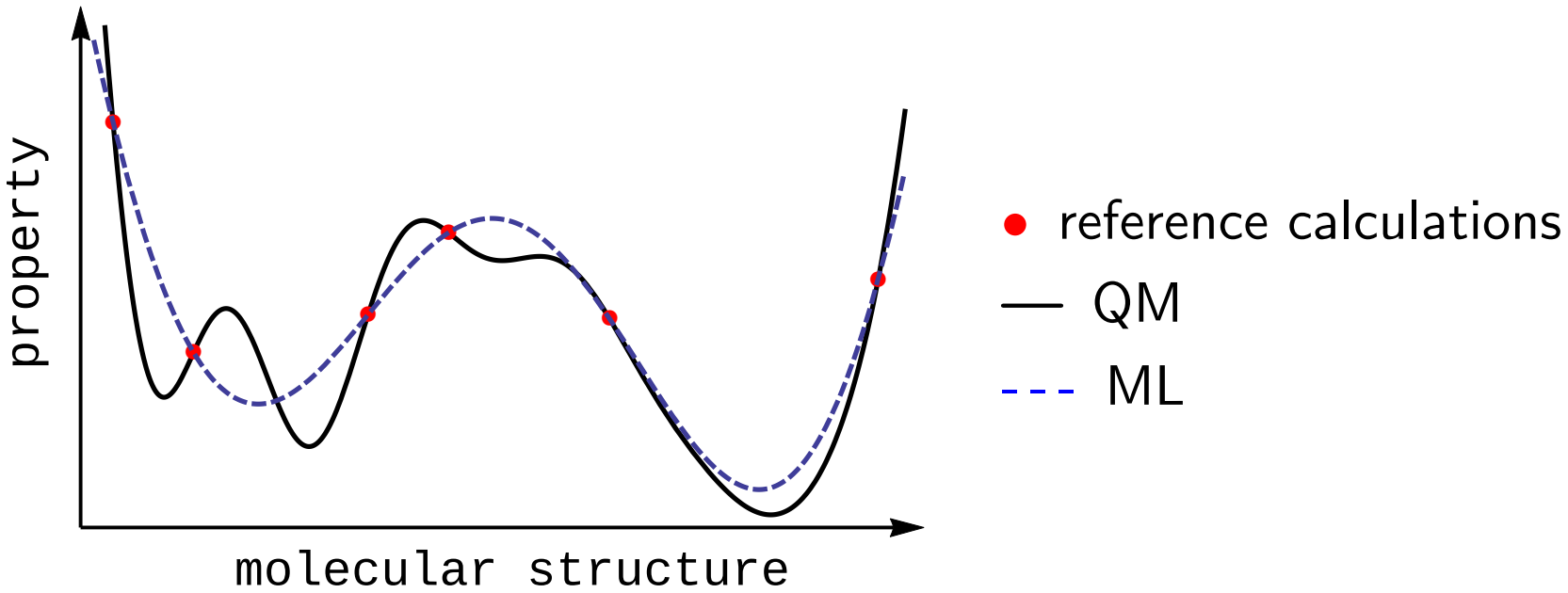
QM/ML models:

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

QM/ML models

Exploit redundancy in a series of QM calculations

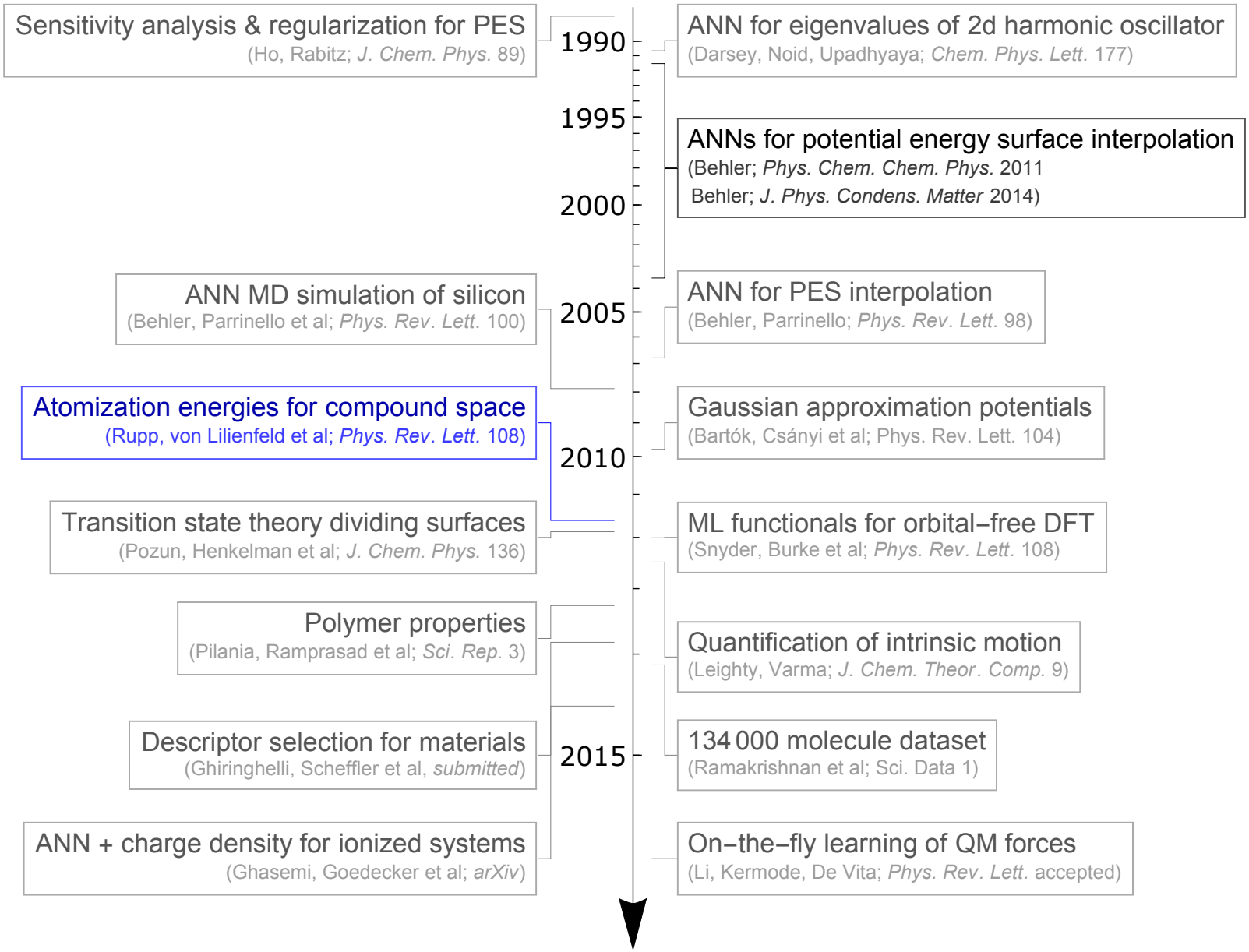
- QM/ML = quantum mechanics + machine learning
- Interpolate between QM calculations using ML
- Smoothness assumption (regularization)
- Large systems, long simulations, many systems



Relationship to other models

Quantum mechanics	Molecular mechanics	Machine learning
Deductive	Mostly deductive	Inductive
Form from physics	Form from physics	Form from statistics
No or little fitting	Fitted to one class	Fitted per dataset
Few or no parameters	Some parameters	Many parameters
General applicability	Limited domain	General applicability
Slow	Fast	In between
Small systems	Large systems	Large systems

Overview


Sensitivity analysis & regularization for PES
(Ho, Rabitz; J. Chem. Phys. 89)
1990
1995
2000
2005
2010
2015
ANN for eigenvalues of 2d harmonic oscillator
(Darsey, Noid, Upadhyaya; Chem. Phys. Lett. 177)
ANNs for potential energy surface interpolation
(Behler; Phys. Chem. Chem. Phys. 2011
Behler; J. Phys. Condens. Matter 2014)
ANN MD simulation of silicon
(Behler, Parrinello et al; Phys. Rev. Lett. 100)
ANN for PES interpolation
(Behler, Parrinello; Phys. Rev. Lett. 98)
Atomization energies for compound space
(Rupp, von Lilienfeld et al; Phys. Rev. Lett. 108)
Gaussian approximation potentials
(Bartók, Csányi et al; Phys. Rev. Lett. 104)
Transition state theory dividing surfaces
(Pozun, Henkelman et al; J. Chem. Phys. 136)
ML functionals for orbital–free DFT
(Snyder, Burke et al; Phys. Rev. Lett. 108)
Polymer properties
(Pilania, Ramprasad et al; Sci. Rep. 3)
Quantification of intrinsic motion
(Leighty, Varma; J. Chem. Theor. Comp. 9)
Descriptor selection for materials
(Ghiringhelli, Scheffler et al, submitted)
134 000 molecule dataset
(Ramakrishnan et al; Sci. Data 1)
ANN + charge density for ionized systems
(Ghasemi, Goedecker et al; arXiv)
On–the–fly learning of QM forces
(Li, Kermode, De Vita; Phys. Rev. Lett. accepted)

What is machine learning?

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

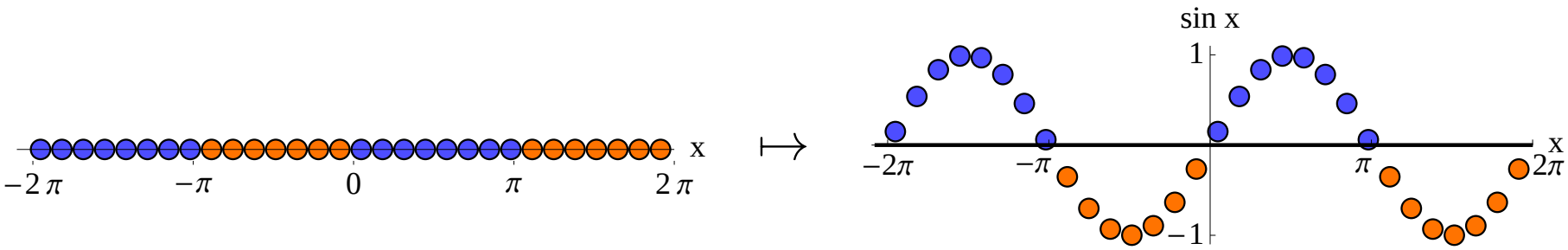
- Widely applied, many problems and algorithms
- Systematic identification of regularity in data for prediction & analysis
- Interpolation in high-dimensional spaces
- Inductive, data-driven; empirical in a principled way

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

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 $\mathcal{H}$

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

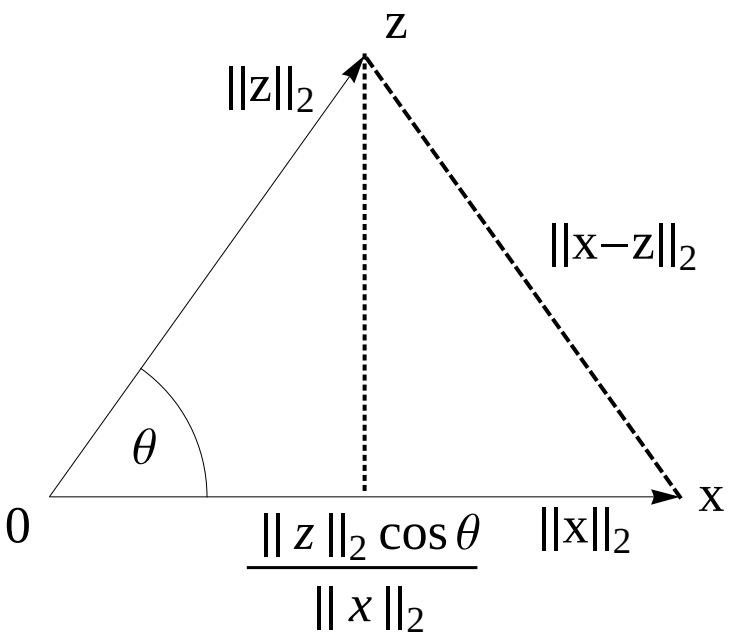
Schölkopf, Smola: Learning with Kernels, 2002; Hofmann et al., *Ann Stat* 36: 1171, 2008

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

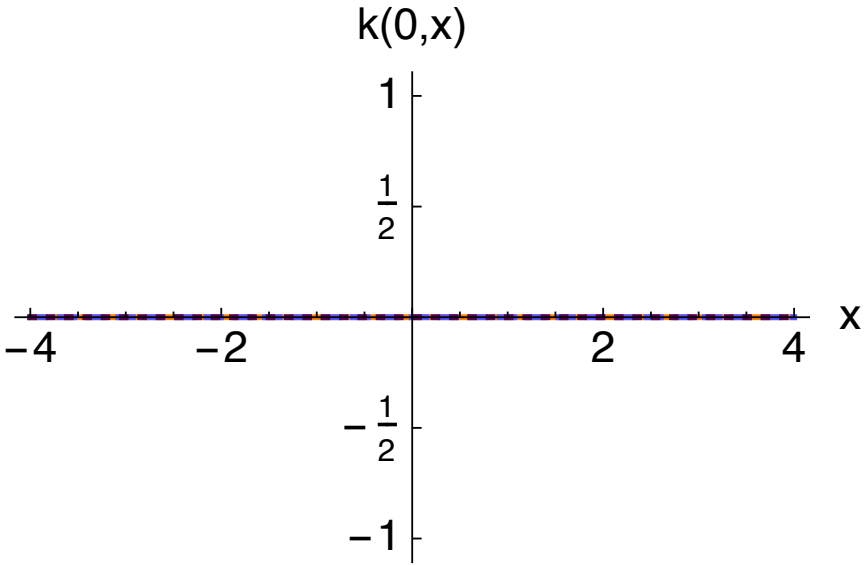
Inner products encode information about lengths and angles:
$||x - z||^2 = \langle x, x \rangle - 2 \langle x, z \rangle + \langle z, z \rangle \,, \qquad \cos\theta = \frac{\langle x, z \rangle}{||x||\,||z||} \,.$

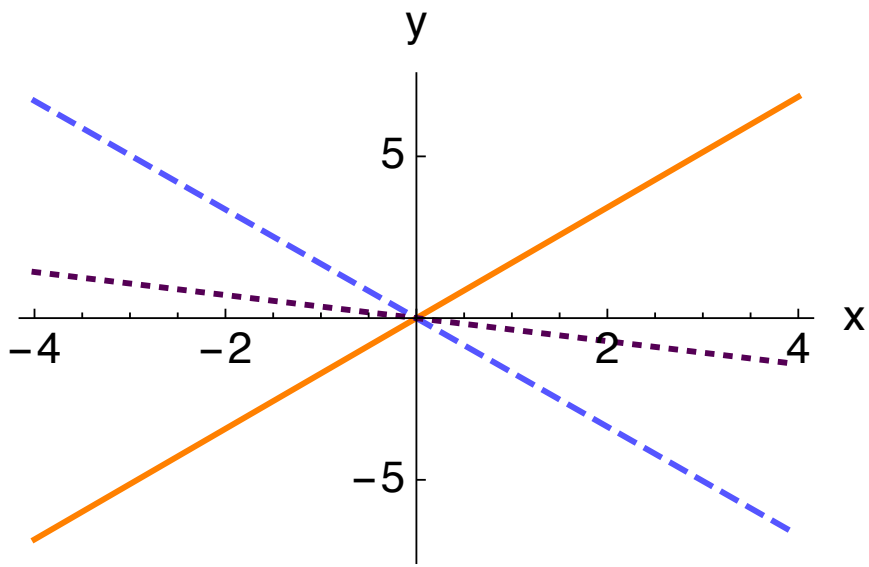


- Well characterized function class
- Closure properties
- Access data only by $\boldsymbol{K}_{ij} = k(x_i, x_j)$
- $\mathcal{X}$ can be any non-empty set
- Examples:
  Linear kernel $\langle \boldsymbol{x}, \boldsymbol{z} \rangle$
  Gaussian kernel $\exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{z}||^2}{2\sigma^2}\right)$

Examples of kernel functions

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

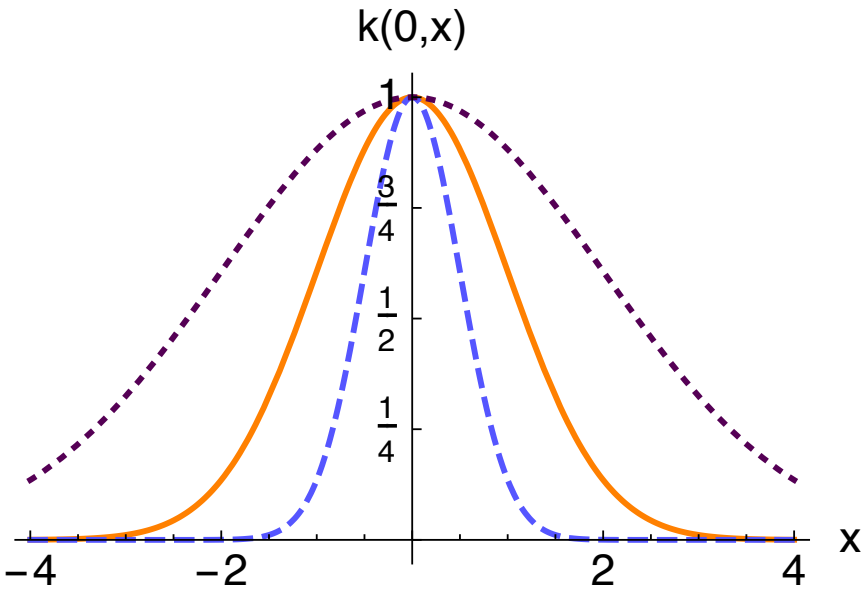


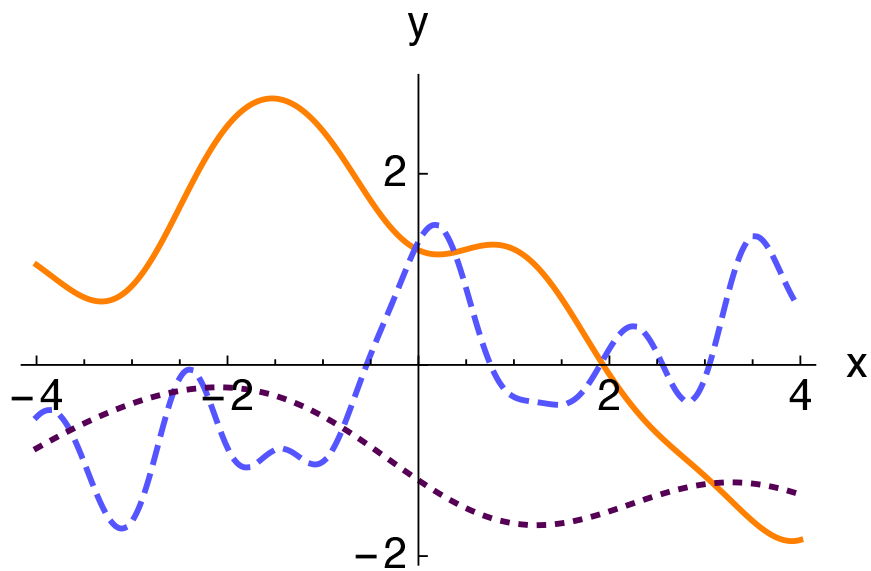


- Recovers original linear model

Examples of kernel functions

Gaussian kernel $k(\boldsymbol{x}, \boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{z}\|^2}{2\sigma^2}\right)$

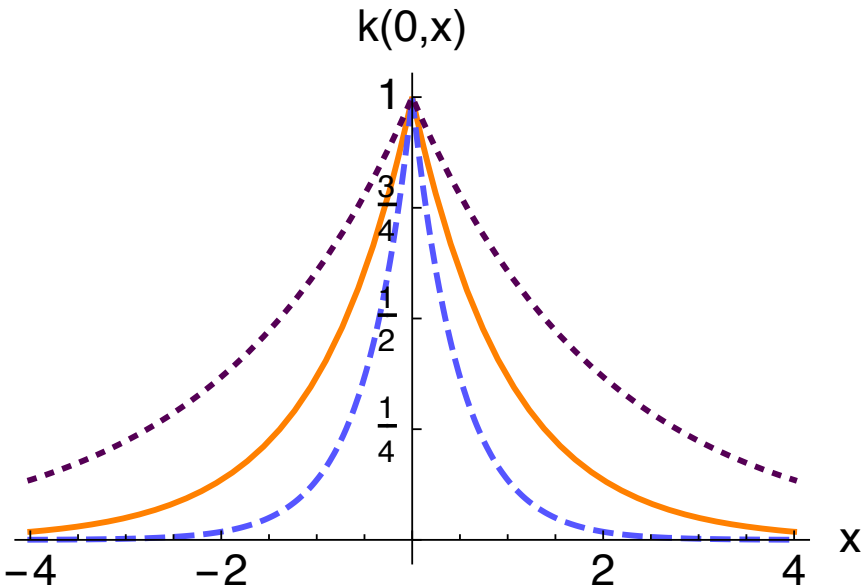


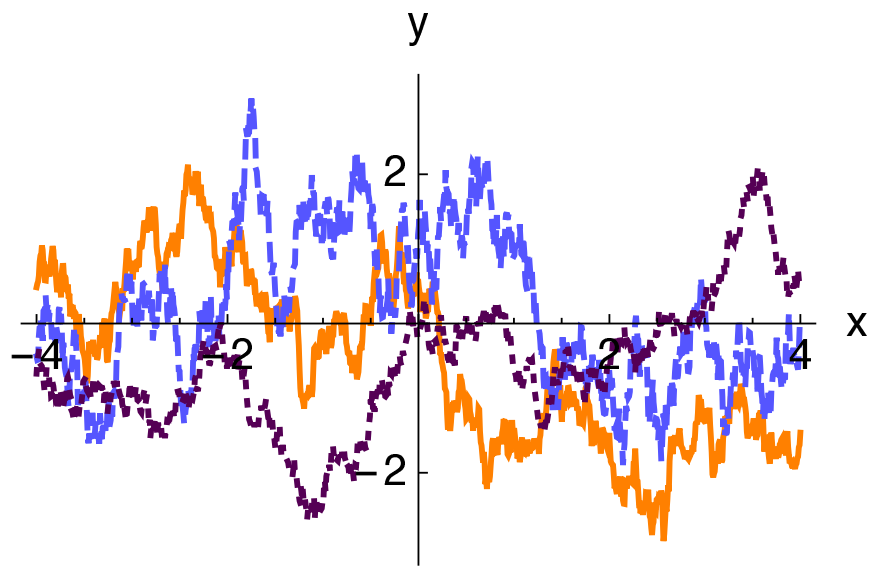


- Length scale σ
- Infinite dimensional feature space
- Universal local approximator

Examples of kernel functions

Laplacian kernel $k(\boldsymbol{x}, \boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{z}\|_1}{\sigma}\right)$





- Length scale σ

Representer theorem

Kernel models have form

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

due to the representer theorem:

Any function minimizing a regularized risk functional

$$\ell\Big(\big(\boldsymbol{x_i}, y_i, \hat{f}(\boldsymbol{x_i})\big)_{i=1}^{n}\Big) + g\big(\|\hat{f}\|\big)$$

admits to above representation.

Schölkopf, Herbrich & Smola, COLT 2001

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

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

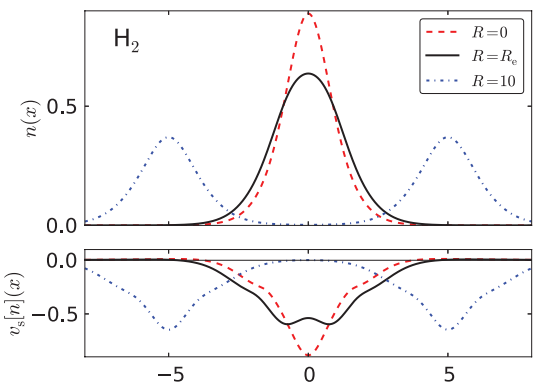
which has solution

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

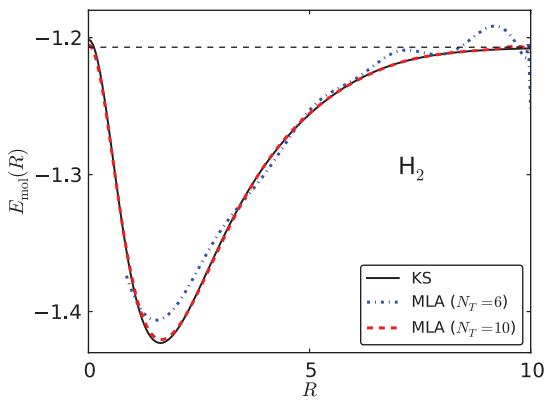
Density functional theory

Learning the map from electron density to kinetic energy

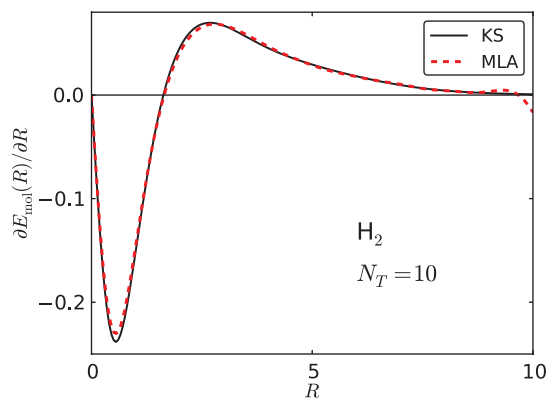
- Orbital-free DFT
- 1D toy system
- DFT/LDA as reference
- Error decays to zero
- Self-consistent densities
- Bond breaking and formation



H_2 potential



H_2 binding curve

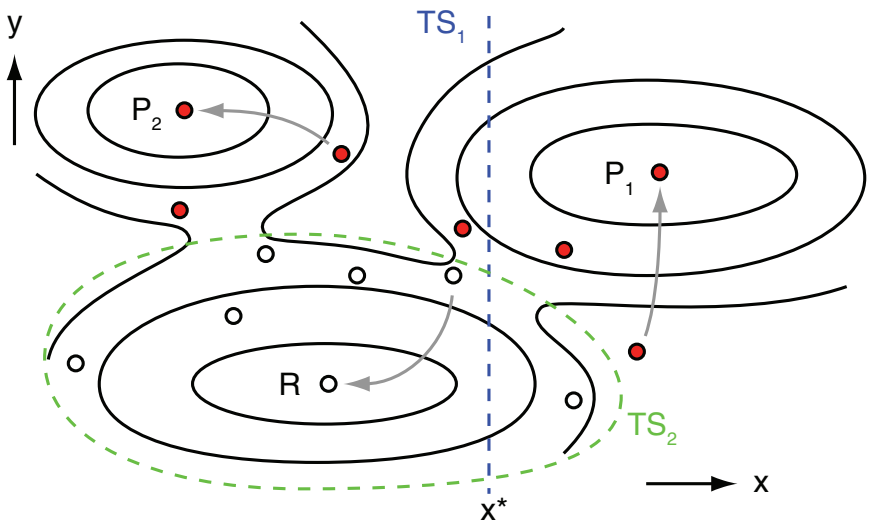


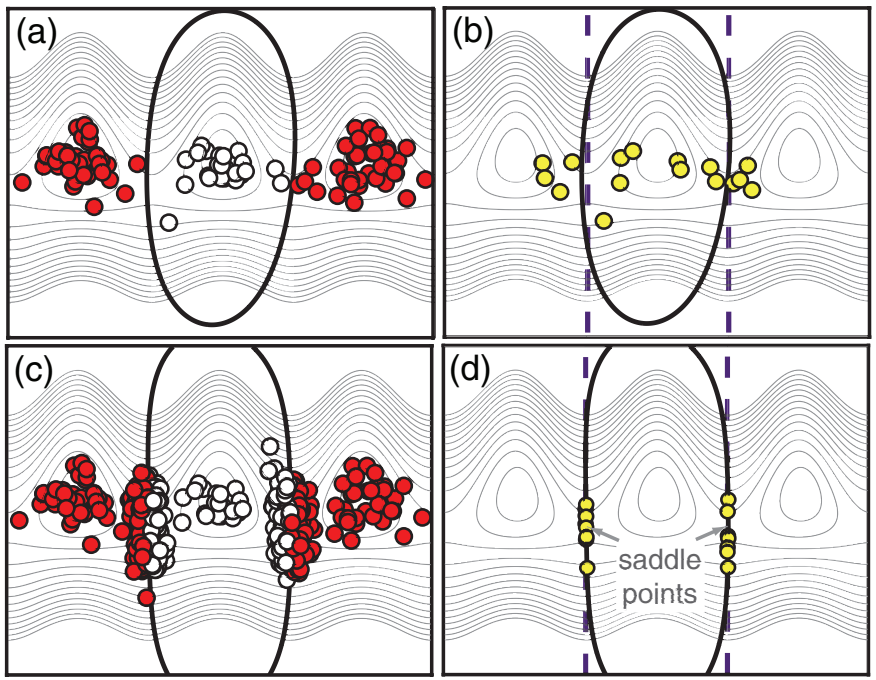
H_2 forces

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.

Gaussian approximation potentials

- Gaussian process regression
- Molecular dynamics
- Partitioned energies

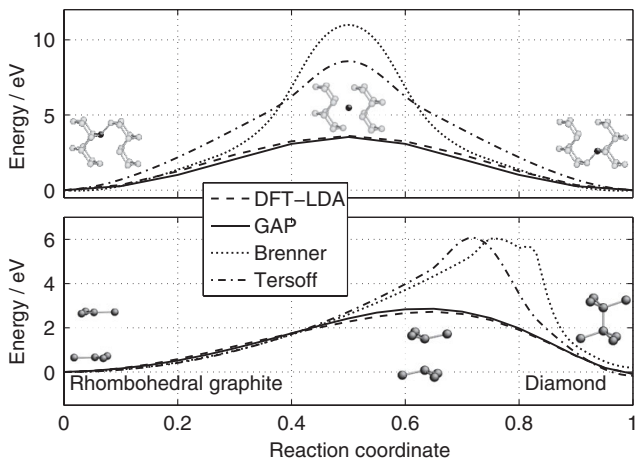
- Representation:

  Local density

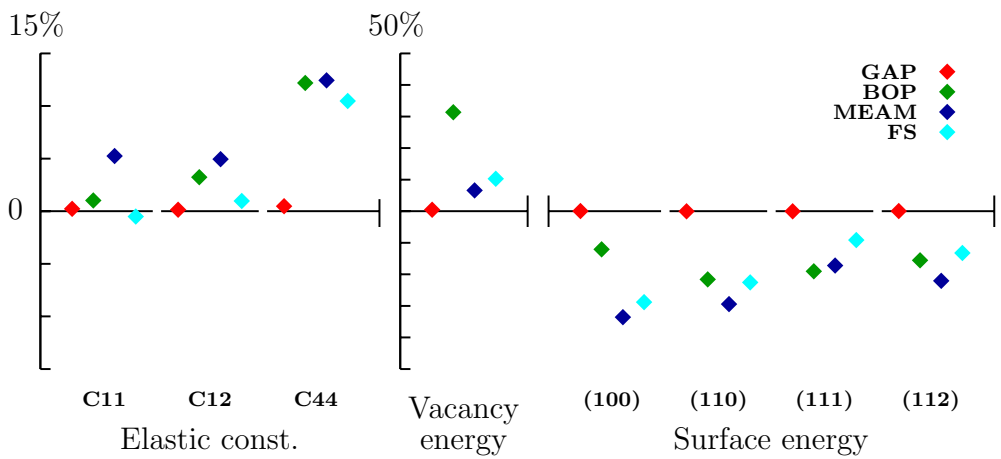
  Projection to 4d sphere

  Hyperspherical harmonics

  Bispectrum



Transition path energies



Errors on properties of Tungsten

Bartók, Csányi et al, *Phys Rev Lett* 104: 136403, 2010.
Szlachta et al, *Phys Rev B* 90: 104108, 2014

Molecular properties

Data

7 165 small organic molecules

DFT PBE0 atomization energies

Representation

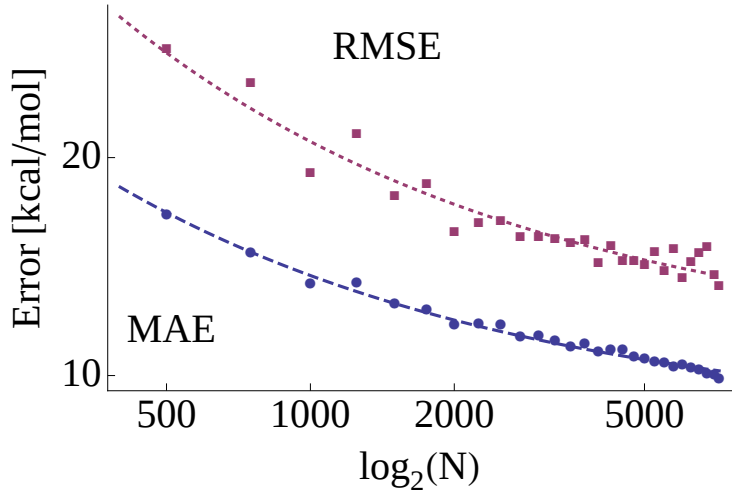
$$\boldsymbol{M}_{IJ} = \begin{cases} \frac{1}{2} Z_I^{2.4} & \text{if } I = J \\ \frac{Z_I Z_J}{\|\boldsymbol{R}_I - \boldsymbol{R}_J\|} & \text{if } I \neq J \end{cases}$$

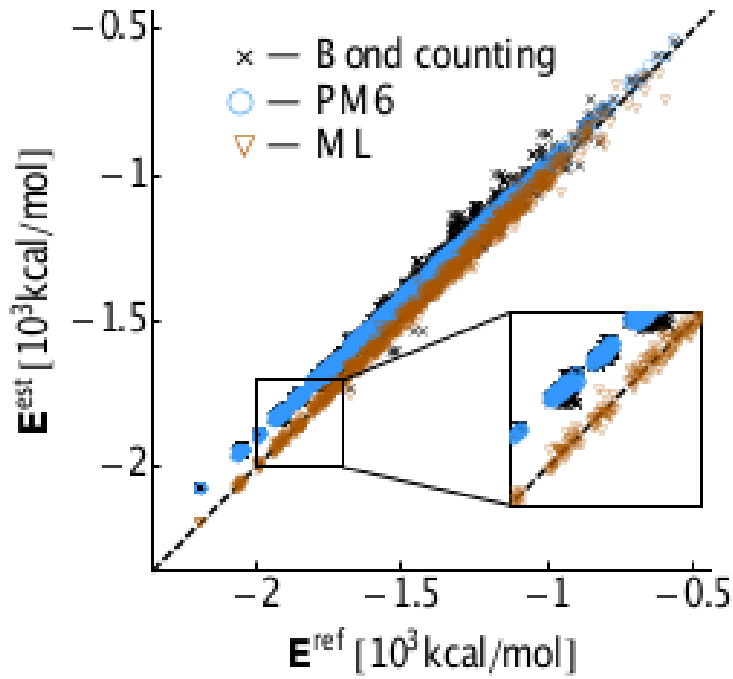
Model

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

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

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



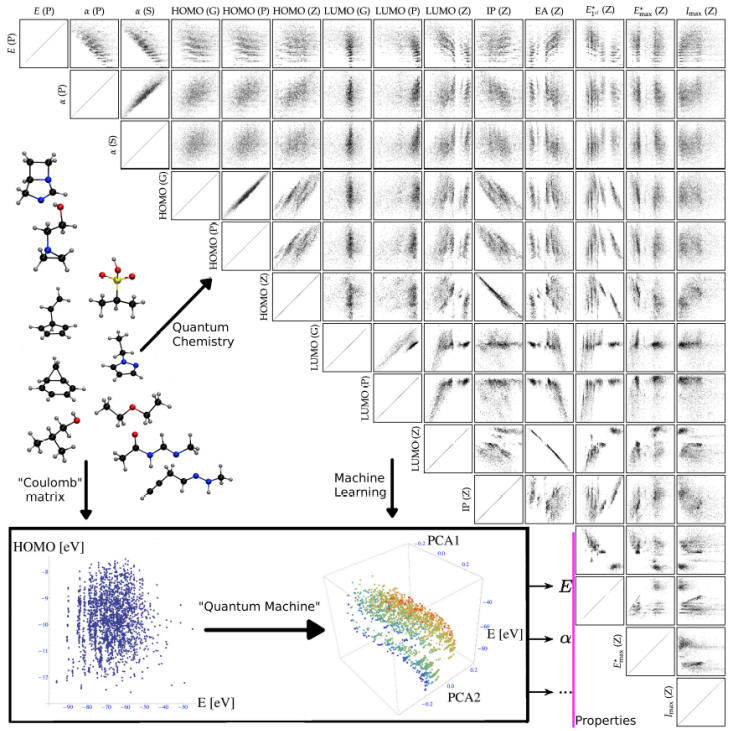


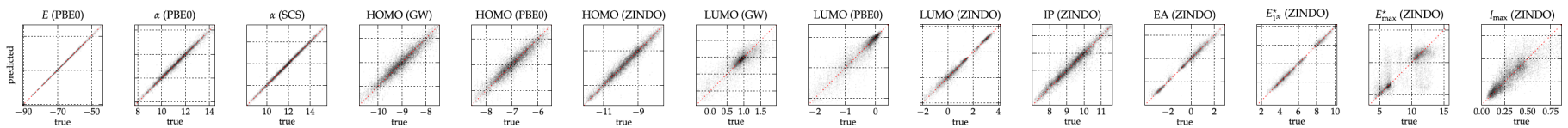
Rupp et al, *Phys Rev Lett* 108: 058301, 2012

Extension to other properties

Learning the map from molecular structure to molecular properties

- Various properties
- Various levels of theory
- Small organic molecules
- Coulomb matrix representations
- Kernel learning,
  deep neural networks
- For 5k training molecules, errors
  are comparable to the reference



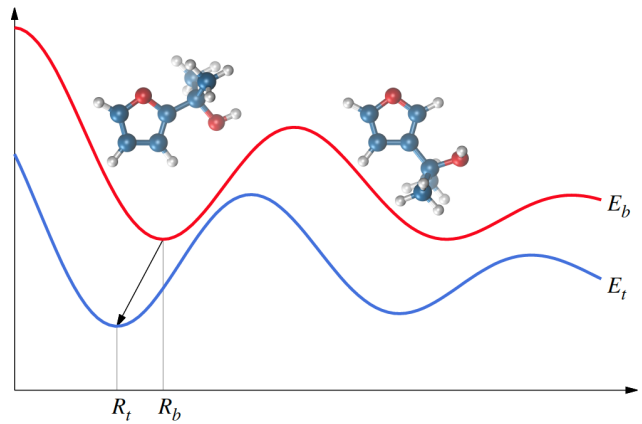


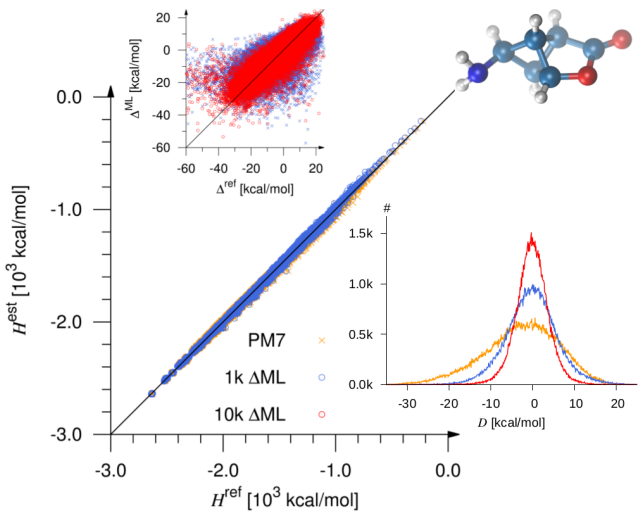
Montavon et al, *New J Phys* 15: 095003, 2013. Hansen et al, *J Chem Theor Comput* 9: 3404, 2013

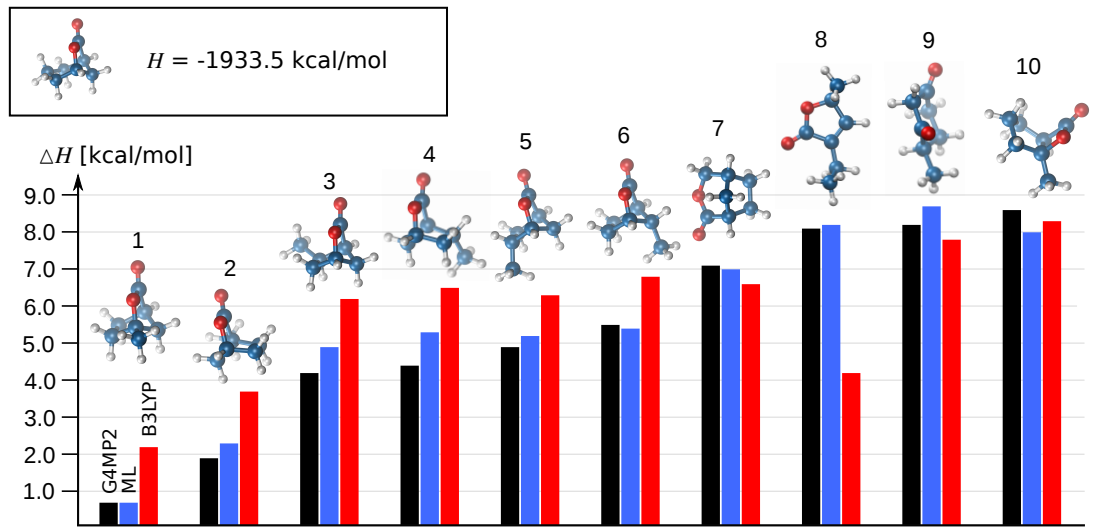
Δ-learning

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: $\Delta^{\text{B3LYP}}_{\text{PM7}}$, $\Delta^{\text{G4MP2}}_{\text{PM7}}$, $\Delta^{\text{CCSD(T)}}_{\text{HF}}$

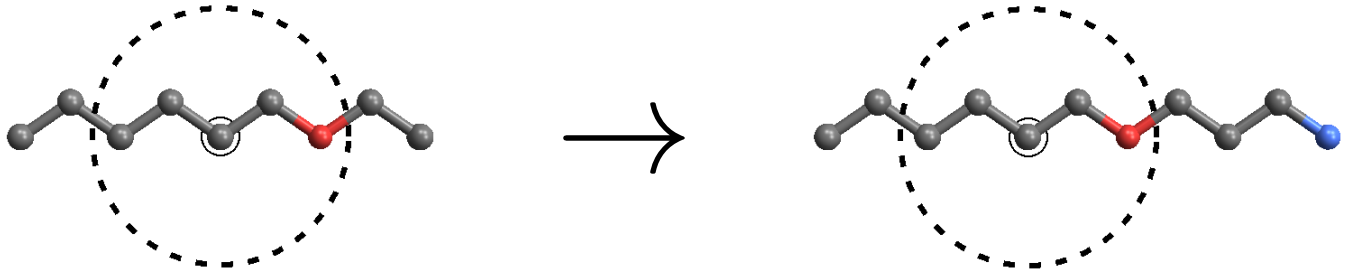




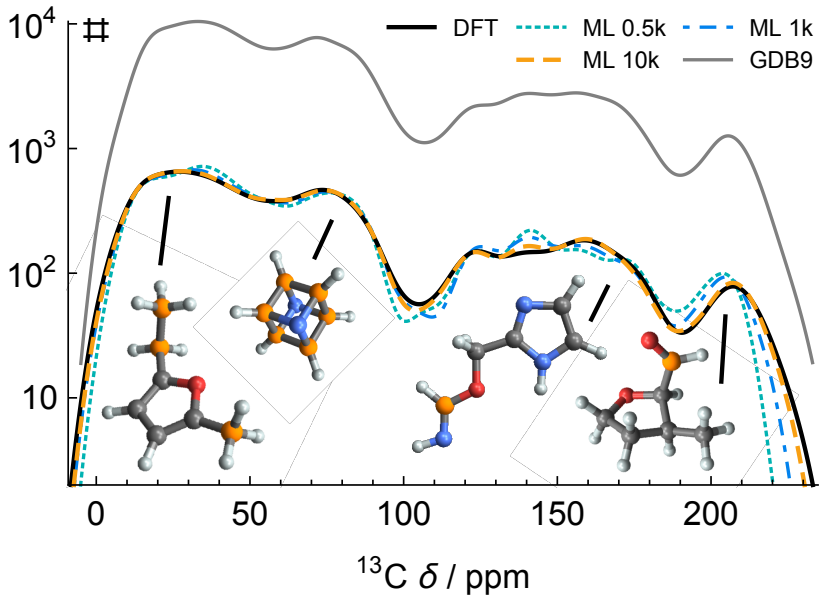


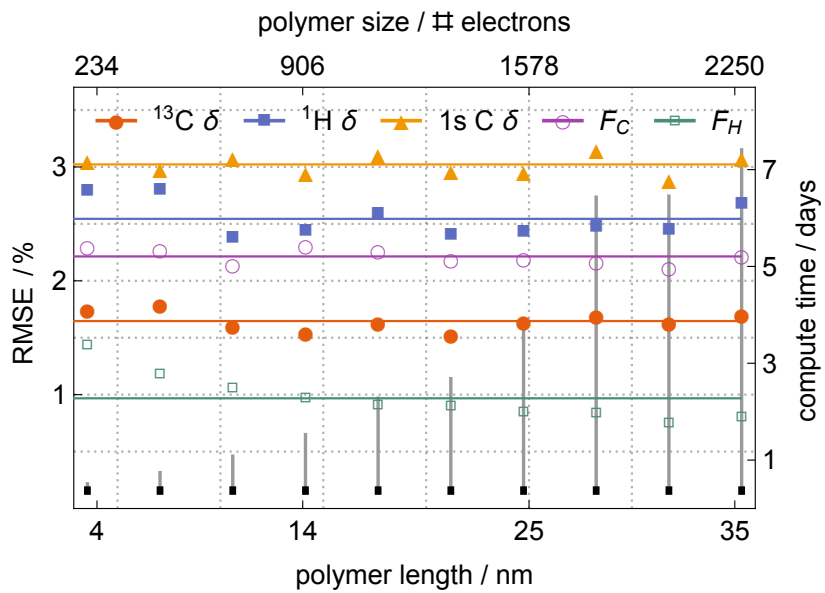
Ramakrishnan, Dral, Rupp, von Lilienfeld, *J Chem Theor Comput* 11: 2087, 2015.

Properties of atoms in molecules



- Local interpolation is global extrapolation
- Linear scaling due to locality

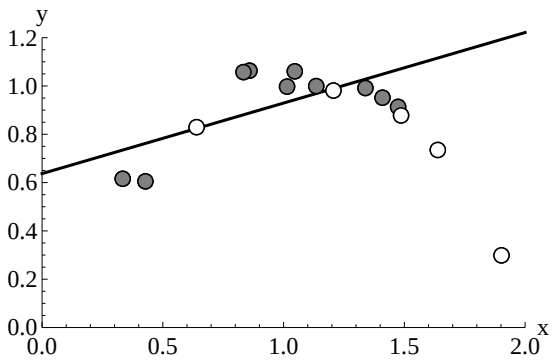




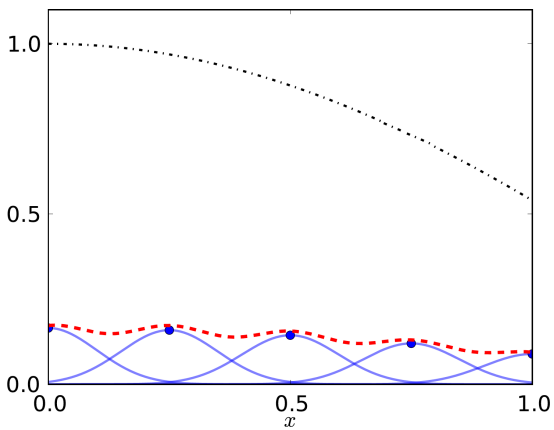
Rupp, Ramakrishnan, von Lilienfeld, *arxiv* 1505.00350, 2015

Overfitting: Model complexity and generalization error

Underfitting

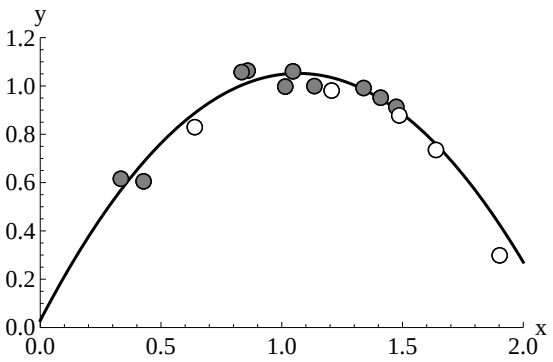


0.123 / 0.443

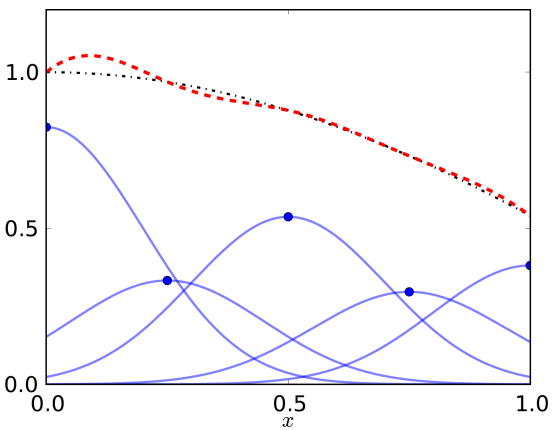


λ too large

Fitting

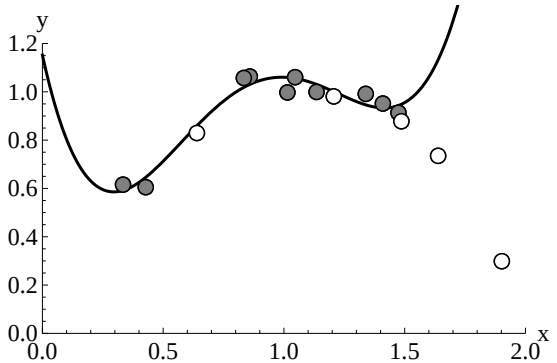


0.044 / 0.068

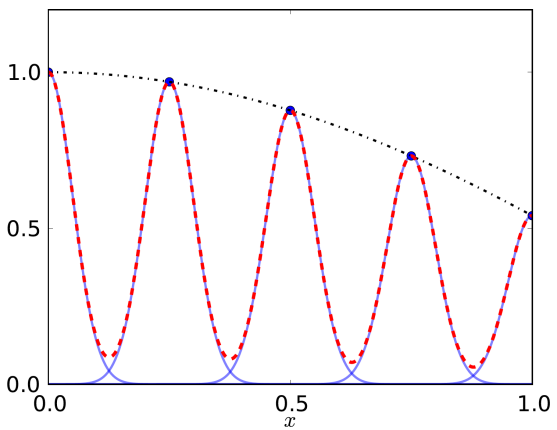


λ right

Overfitting



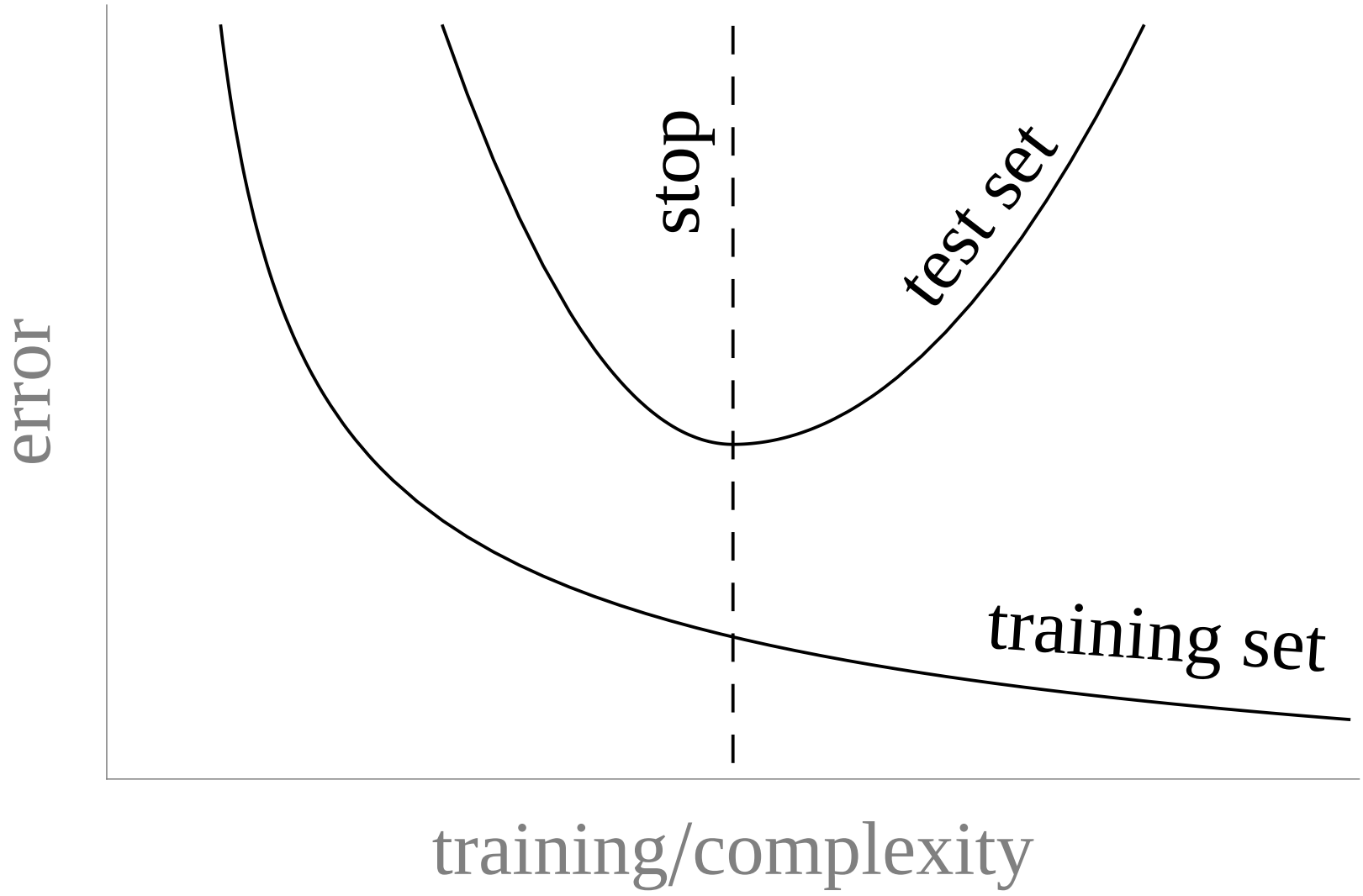
0.036 / 0.939



λ too small

Rupp, PhD thesis, 2009. Vu, Snyder et al, *Int. J. Quant. Chem.* 115: 1115, 2015

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

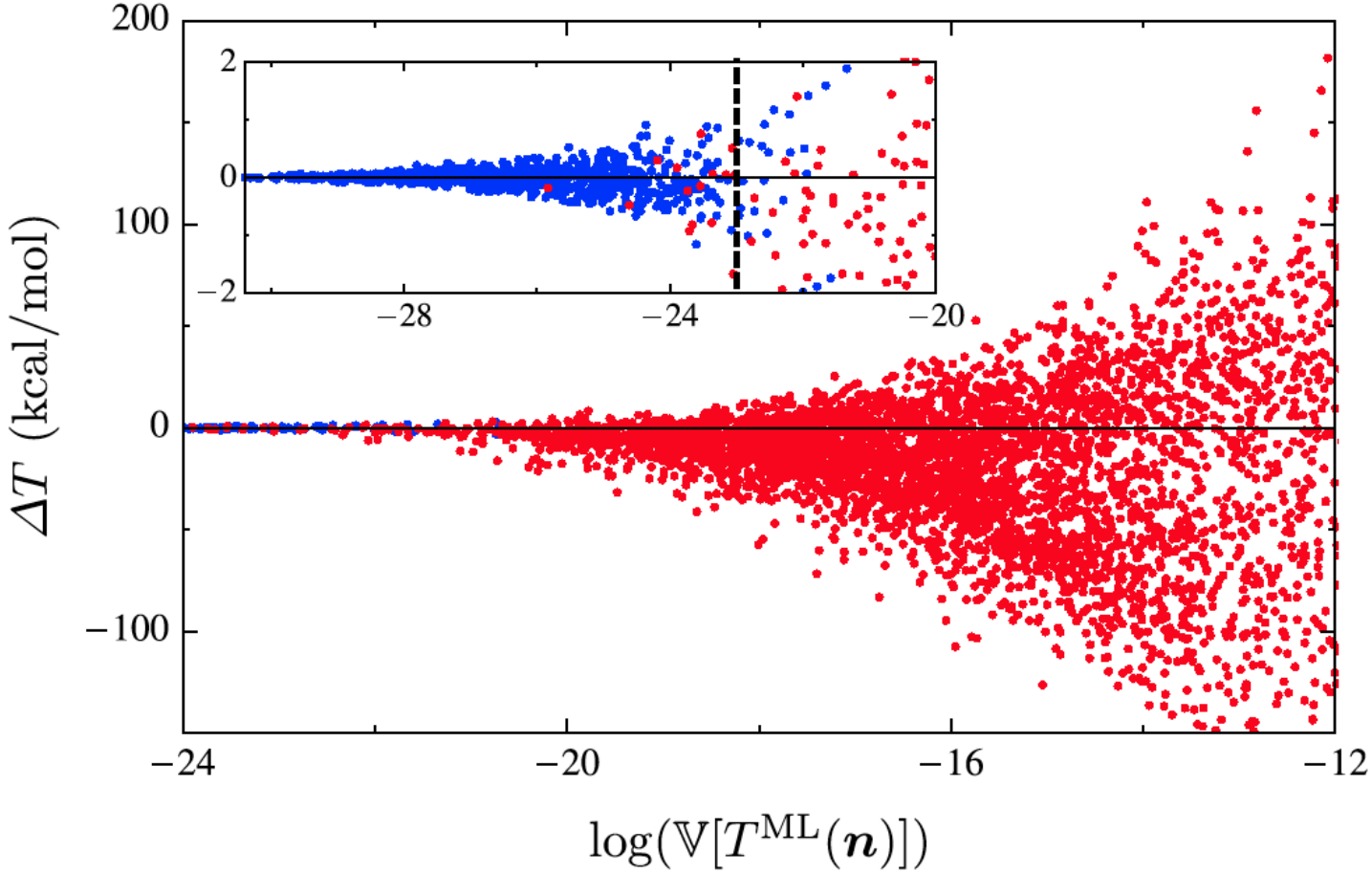
✓ split data, center training set, train, center test set, predict

Example 3: cross-validation with feature selection

× feature selection, cross-validation

✓ feature selection for each split of cross-validation

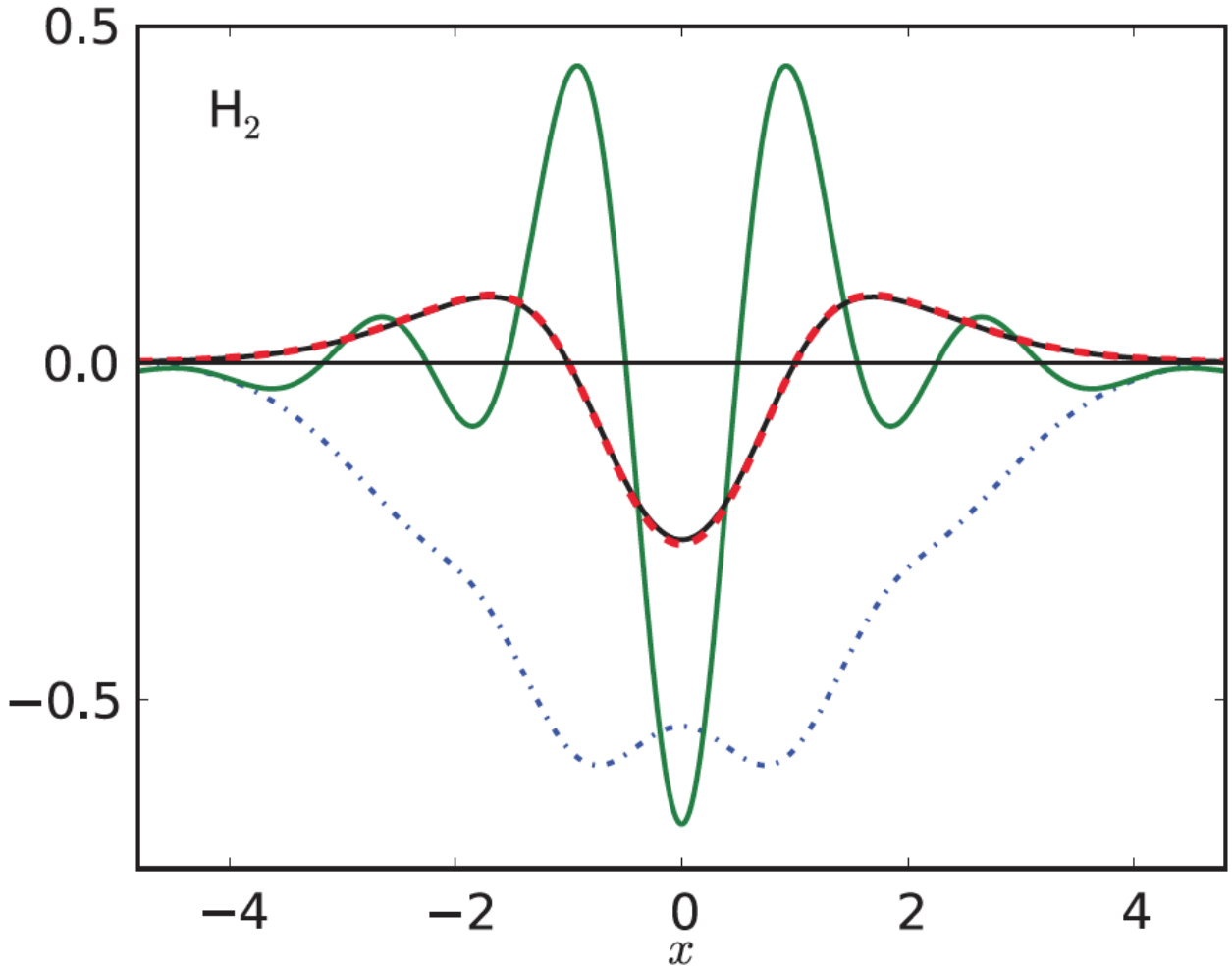
Reliability of predictions



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 and allows investigation of larger systems, longer timescales, and more systems

Tutorial

International Journal of Quantum Chemistry 115(16), 2015
Special issue on *Quantum Chemistry and Machine Learning*

TUTORIAL REVIEWS WWW.Q-CHEM.ORG International Journal of QUANTUM CHEMISTRY

Machine Learning for Quantum Mechanics in a Nutshell

Matthias Rupp*

Models that combine quantum mechanics (QM) with machine learning (ML) promise to deliver the accuracy of QM at the speed of ML. This hands-on tutorial introduces the reader to QM/ML models based on kernel learning, an elegant, systematically nonlinear form of ML. Pseudocode and a reference implementation are provided, enabling the reader to reproduce results from recent publications where atomization energies of small organic molecules are predicted using kernel ridge regression. © 2015 Wiley Periodicals, Inc.

DOI: 10.1002/qua.24954

Rupp, *Int J Quant Chem* 115: 1058, 2015.

Acknowledgements

Collaborators

O A von Lilienfeld, R Ramakrishnan	U Basel
M Scheffler, A Tkatchenko, V Gobre	FHI
K-R Müller, A Ziehe, F Biegler, F Brockherde, G Montavon	TU Berlin
J C Snyder, K Hansen, S Fazli	TU Berlin
K Burke, I M Pelaschier, J Huang, L Blooston, L Li	UCI
G Schneider, G Folkers, M Reutlinger	ETHZ
G Henkelman, D Sheppard, Z Pozun	U Austin
F M Boeckler, A Lange, M R Bauer, R Wilcken	U Tübingen
A Knoll, A Lopez-Bezanilla, A Vazquez-Mayagoitia, P O Dral	

Institutions

- IPAM
- EU IEF (273039)
- DFG (FOR1406TP4)
- SNF (205321-134783, PP00P2_138932)
- sciCORE, U Basel
- FHI, Max Planck Society

www.mrupp.info