

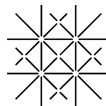
Quantum Mechanical Properties of Atoms in Molecules via Machine Learning

Matthias Rupp

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

Joint work with Raghunathan Ramakrishnan and O. Anatole von Lilienfeld
University of Basel, Switzerland

Ψ_k 2015 Conference
September 6–10, San Sebastian, Spain



Overview

Problem

Computational cost of numerical approximations
limits uses of electronic structure theory

Goal

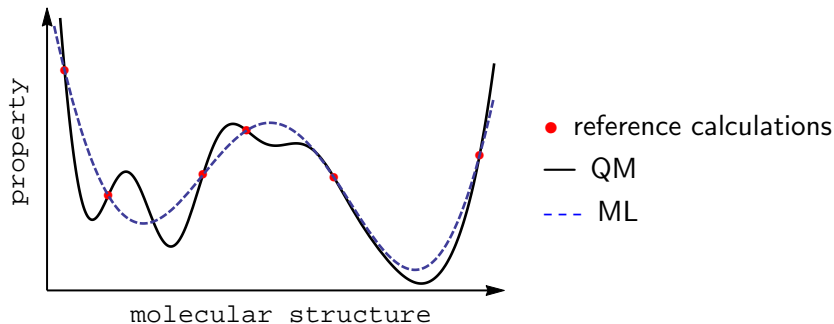
Combining quantum mechanics with machine learning to handle
larger systems, longer simulations, more systems, and higher accuracy

Approach

Interpolation between reference calculations
across geometries and compositions

Idea of QM/ML models

- QM/ML = quantum mechanics + machine learning
- exploit redundancy in a series of QM calculations
- interpolate between QM calculations using ML



Kernel ridge regression

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

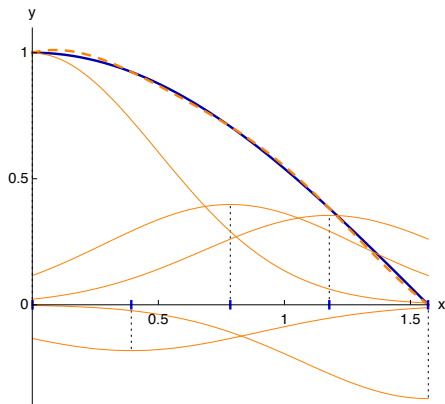
$$\text{optimization problem } \underset{\alpha \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^n (\hat{f}(\mathbf{x}_i) - y_i)^2 + \lambda \alpha^T \mathbf{K} \alpha$$

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

with k positive definite, $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, regularization strength $\lambda \in \mathbb{R}$.

Kernel ridge regression example

Weighted basis functions placed on training samples \mathbf{x}_i



Example:

— $f(x) = \cos(x)$

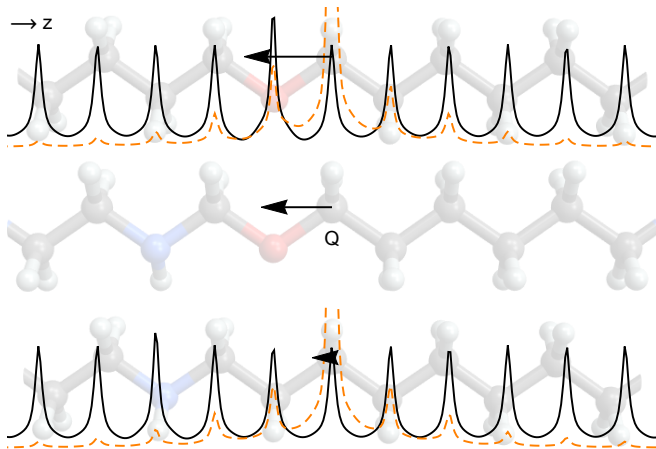
■ training samples

— Gaussian basis functions

- - prediction \hat{f}

Local environments

Local properties of atoms in molecules



Rupp et al., *J Phys Chem Lett* 6(16): 3309, 2015.

Local environments

Local atom-centered coordinate systems.

atomic Coulomb matrix

$$\mathbf{M}_{I,J}^{(Q)} = \begin{cases} \frac{1}{2}Z_I^{2.4} & I = J \\ \frac{Z_I Z_J}{\|\mathbf{R}_I - \mathbf{R}_J\|} & I \neq J \end{cases}$$

principal component coordinates

$$\left(\frac{1}{n}\mathbf{X}^T\mathbf{X}\right)\mathbf{v}_\ell = \lambda_\ell\mathbf{v}_\ell$$

$$\mathbf{X}\mathbf{V}^T$$

augmented by Z_I

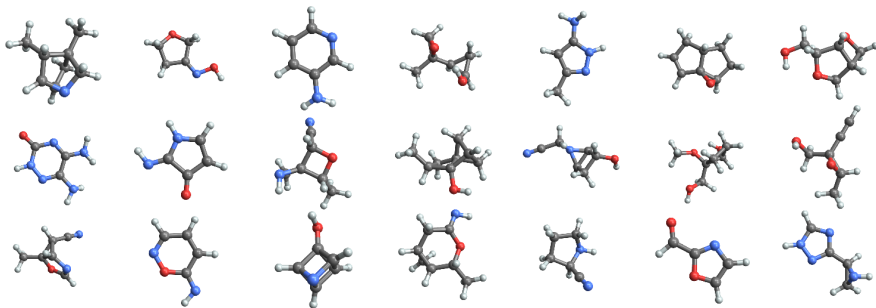
Representations sorted by distance to atom Q .

Data set and properties

- 9 k small organic molecules
- C, N, O, H; 7–9 non-H atoms
- subset of GDB9
- forces: 100 conformations for each of 168 $C_7H_{10}O_2$ isomers

- nuclear chemical shifts
- core level excitations
- forces

Calculations at DFT/PBE0/
def2TZVP level using Gaussian

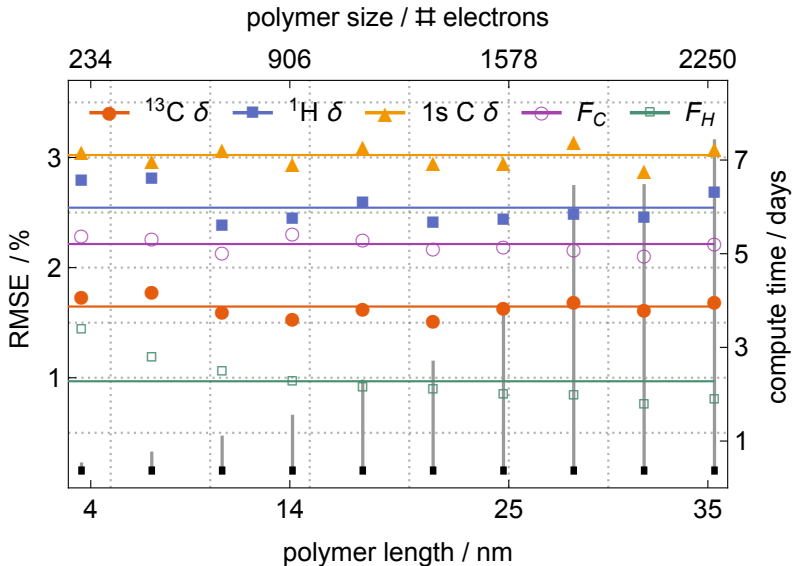


Results

Property	Ref.	Range	MAE	%	R^2
^{13}C δ /ppm	2.4	6 – 211	3.9 ± 0.28	1.9	0.988 ± 0.001
^1H δ /ppm	0.11	0 – 10	0.28 ± 0.01	2.8	0.954 ± 0.005
1s C δ /mE _h	7.5	-165 – -2	4.9 ± 0.12	3.0	0.971 ± 0.002
$F_{\text{C}}/mE_{\text{h}} a_0^{-1}$	1	-99 – 96	3.6 ± 0.10	1.8	0.983 ± 0.002
$F_{\text{H}}/mE_{\text{h}} a_0^{-1}$	1	-43 – 43	0.8 ± 0.02	0.9	0.996 ± 0.003

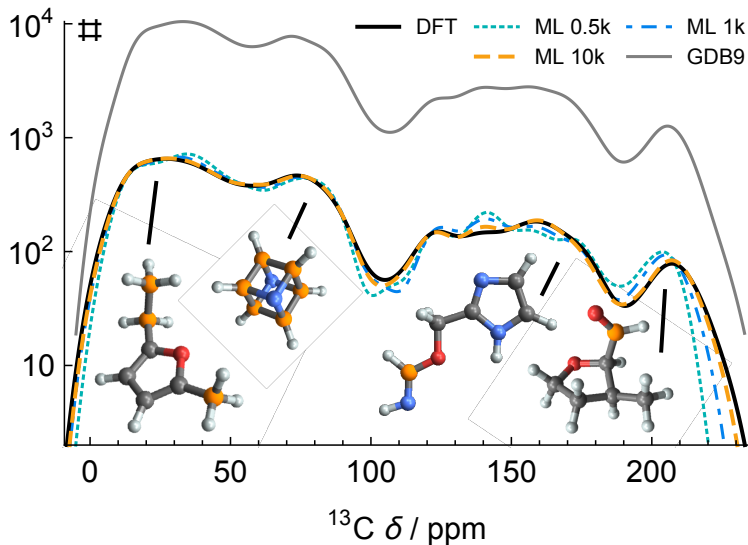
MAE = mean absolute error, R = correlation coefficient

Linear scaling



Rupp et al., *J Phys Chem Lett* 6(16): 3309, 2015.

Prediction of chemical shifts



Rupp et al., *J Phys Chem Lett* 6(16): 3309, 2015.

Conclusions

Machine learning can be used to interpolate between quantum mechanical reference calculations to enable larger systems, longer simulations, more systems, higher accuracy.