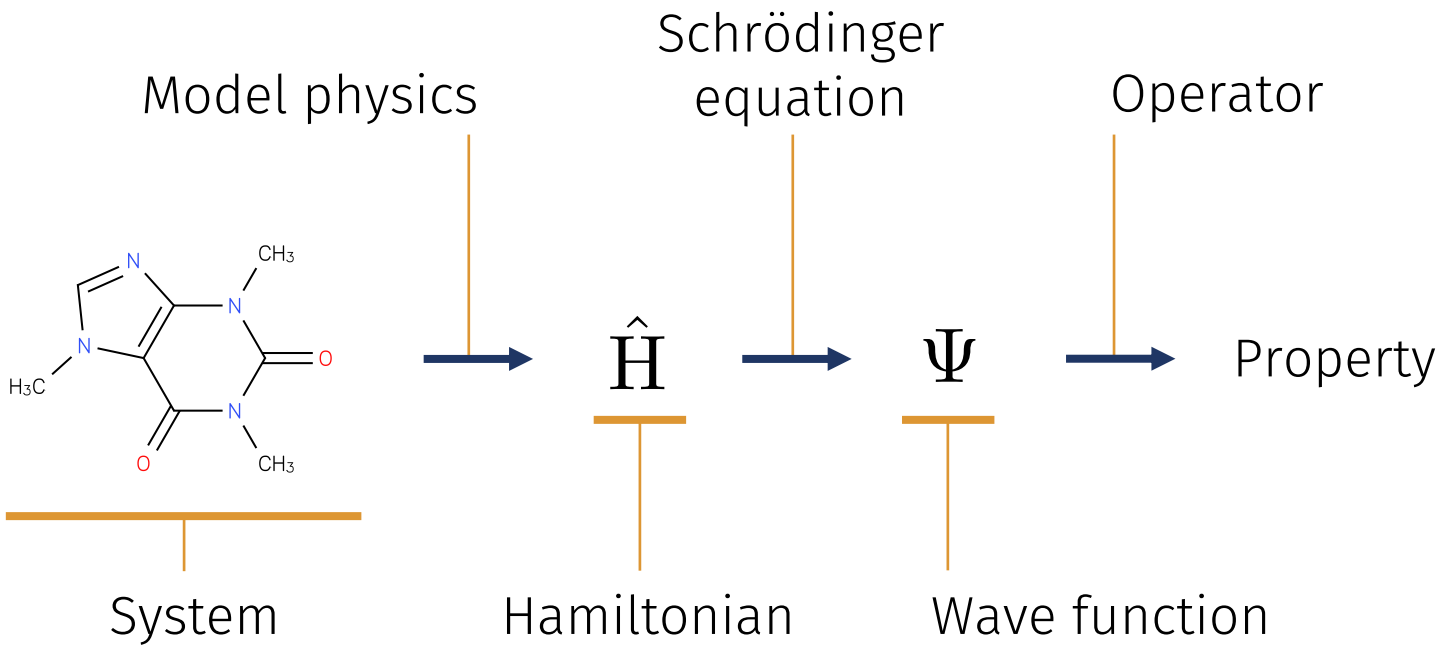
Quantum Alchemy: Treating Chemical Space with Perturbations

Guido Falk von Rudorff, University of Kassel

vonrudorff@uni-kassel.de nablachem.org/talks ferchault @ferchault

Solved?


Model physics
Schrödinger equation
Operator
$\hat{H}$
Ψ
Property
System
Hamiltonian
Wave function

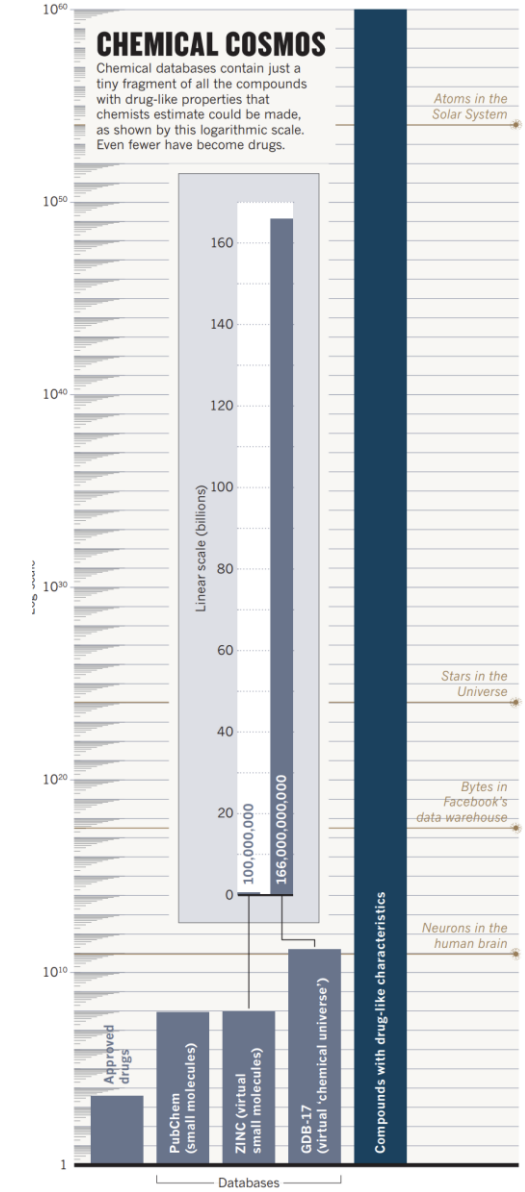
Scaling of Chemical Space

Commercial databases

- 164 million molecules
- 15k added daily

Scale

- One person: 1 million compounds/second
- 10 billion people on earth
- 10^{26} universe ages to go through



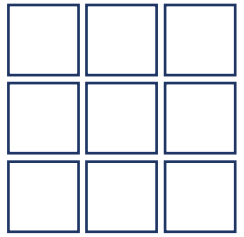
A. Mullard, *Nature*, 2017.

Scaling of Material Space

Face centered cubic and 70 elements only

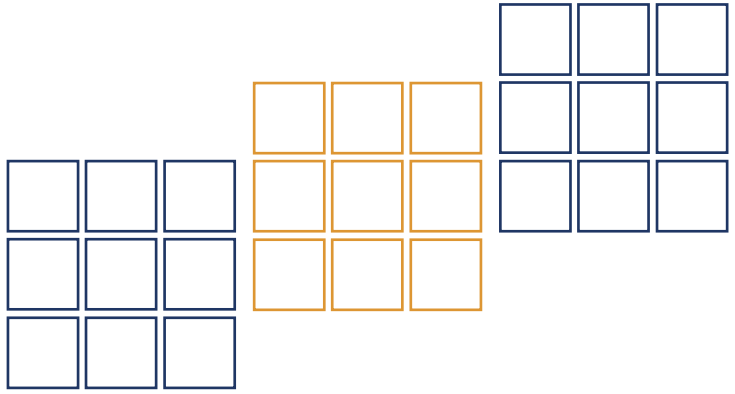
9 primitive cells

- Binary 10^{7}
- Ternary: 10^{13}
- Quaternary: 10^{15}



27 primitive cells

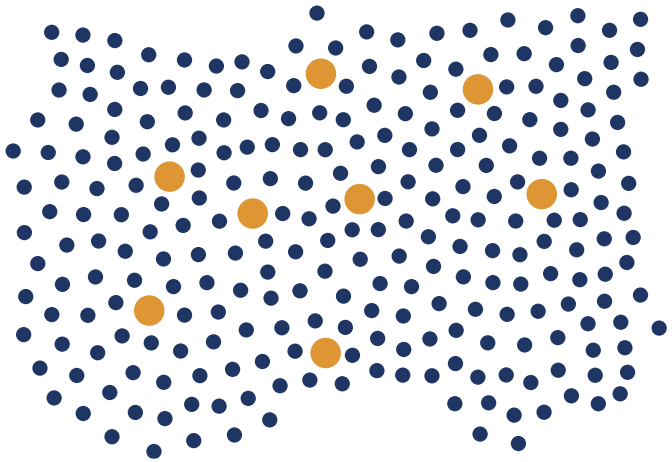
- Binary: ~ 10^{17}
- Ternary: ~ 10^{29}
- Quaternary: ~ 10^{36}

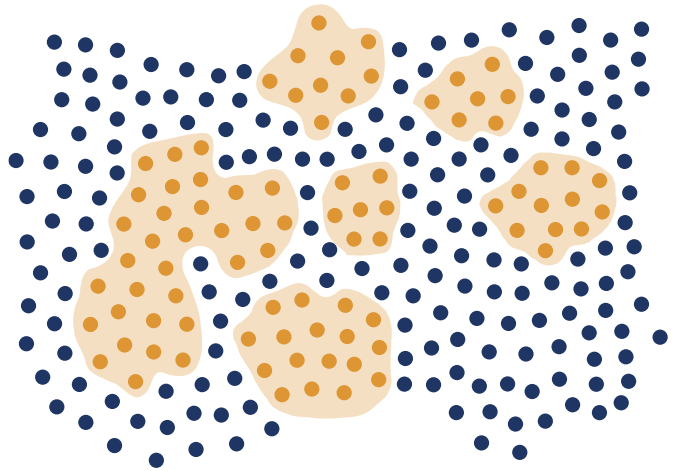


K. Shinohara, A. Seko, T. Horiyama, M. Ishihata, J. Honda, I. Tanaka, *J Chem Phys*, 2020.

Properties of domains?

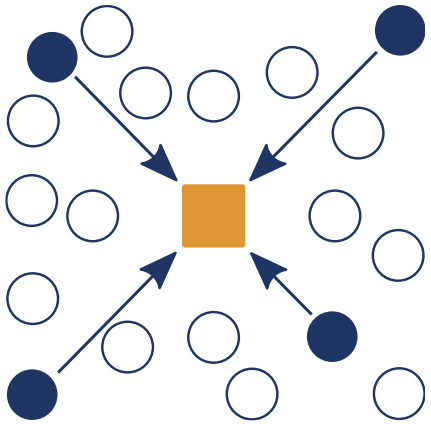
Speed does not matter:
even enumeration is impossible.





Options

Machine Learning



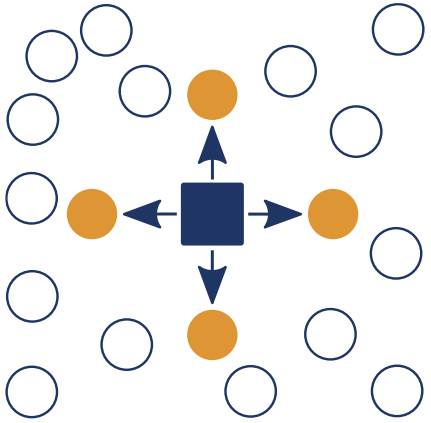
Foundations | Statistical modelling

Accuracy | Systematically improvable through data and training

Specialty | Universal, scale-bridging, data-driven approach

Limitation | Requires training data, no black box

Quantum Alchemy



Foundations | Perturbation theory

Accuracy | Systematically improvable through higher orders terms

Specialty | Combinatorial scaling with chemical diversity

Limitation | Finite range in chemical space

Quantum Alchemy

Idea

Treat system changes perturbatively[1,2,3]
Build a Taylor/Padé approximant[4] : **often 100.000 times faster**

Steps

Choose system

Alter system, calculate property response functions

Predict many modified systems

1 | L. L. Foldy, *Phys Rev.* 1951. 2 | E. B. Wilson, *J. Chem. Phys.* 1962.
3 | GFvR, O. A. von Lilienfeld, *Phys. Rev. Res.*, 2020. 4 | GFvR, *J. Chem. Phys.*, 2021.
Also: Ayers, Cárdenas, Keith, Geerlings, Politzer, ...

Perspective shift

Few highly accurate calculations
instead of many intermediate ones

$$\hat{\mathrm{H}} = \hat{\mathrm{H}}(Z_i, \mathbf{R}_i, N_e, \sigma)$$

4N

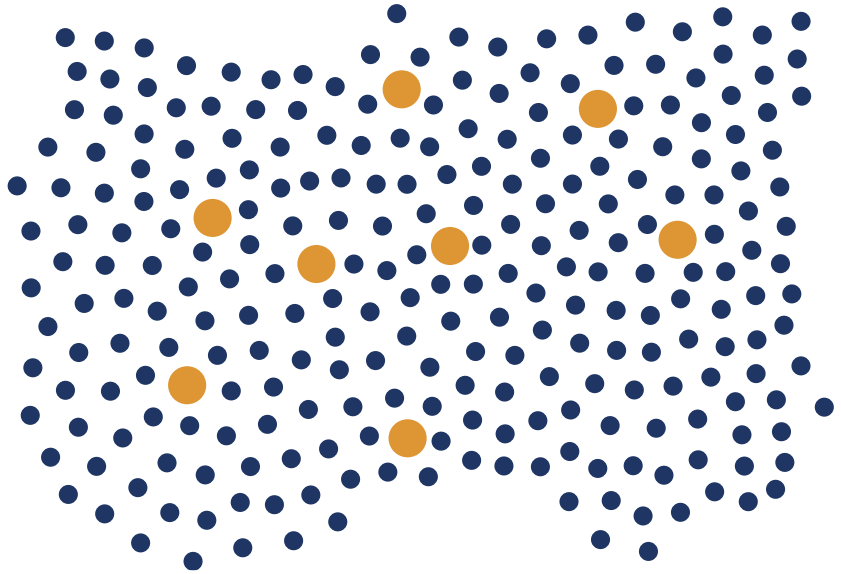
1D, close to $\sum_i Z_i$

1D

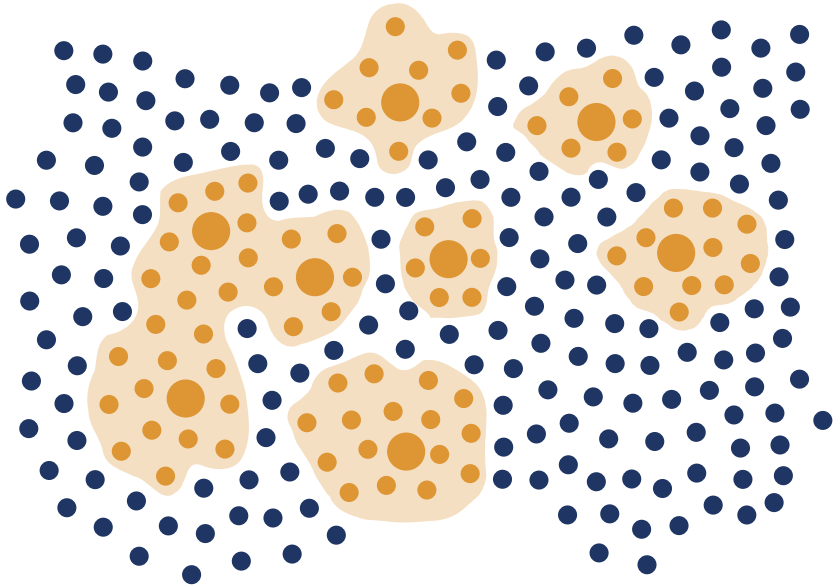
Joseph Wright, 1771

Motivation

Without Perturbation



With Perturbation



Systems
- Any
- Known
- Approximated

Quantum Alchemy

Two systems: Interpolate between molecular isoelectronic Hamiltonians

$$\hat{H}(\lambda) \equiv \lambda \hat{H}_\mathrm{t} + (1-\lambda)\hat{H}_\mathrm{r}$$

Taylor expansion around reference molecule

$$E_\mathrm{t} = E_\mathrm{r} + \Delta E^\mathrm{NN} + \int_\Omega d\mathbf{r} \sum_{n=0}^{\infty} \frac{1}{(n+1)!} \Delta v \frac{\partial^n \rho_\lambda(\mathbf{r})}{\partial \lambda^n}\bigg|_{\lambda=0}$$

(Alchemical Perturbation Density Functional Theory, APDFT)

- Gives consistent energies, densities, forces, ...
- Uses the same derivatives for all predictions

xkcd.com/2605

Requirements

Differentiable / Analytic + Converge quickly

- ✓ Total Energy [1,2]
- ✓ Dipole moments [2]
- ✓ Deprotonation energies [3]
- ✓ Photoelectron circular dichroism [5]
- ✓ Electron density [1,2]
- ✓ Non-covalent interactions [1]
- ✓ Ionisation Energy [4]
- ✓ Orbital eigenvalues [2]
- ✓ Binding energies [1,2]
- ✓ Electron Affinity [4]

In progress

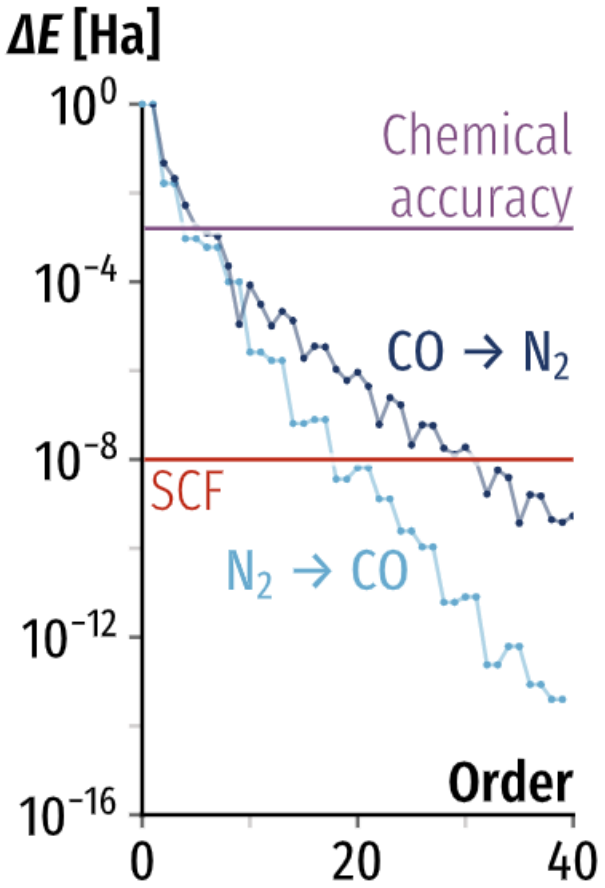
NMR spectra, QMC energies

Tested in: Gaussian, Psi4, PySCF, PySCF-AD, dqc, ORCA, MRCC, cp2k, CPMD
Tested with: HF, KS-DFT, CCSD

NablaChem/nablachem
`pip install nablachem`

1 | GFvR, O. A. von Lilienfeld, *Phys. Rev. Res.*, 2020. 2 | GFvR, *J. Chem. Phys.*, 2021. 3 | GFvR, O. A. von Lilienfeld, *Phys. Chem. Chem. Phys.*, 2020. 4 | E Eikey, A Maldonado, C Griego, GFvR, J Keith, *J. Chem. Phys.*, 2022. 5 | GFvR, A Artemyev, B Lagutin, P Demekhin, *J. Chem. Phys.*, 2024.

Convergence



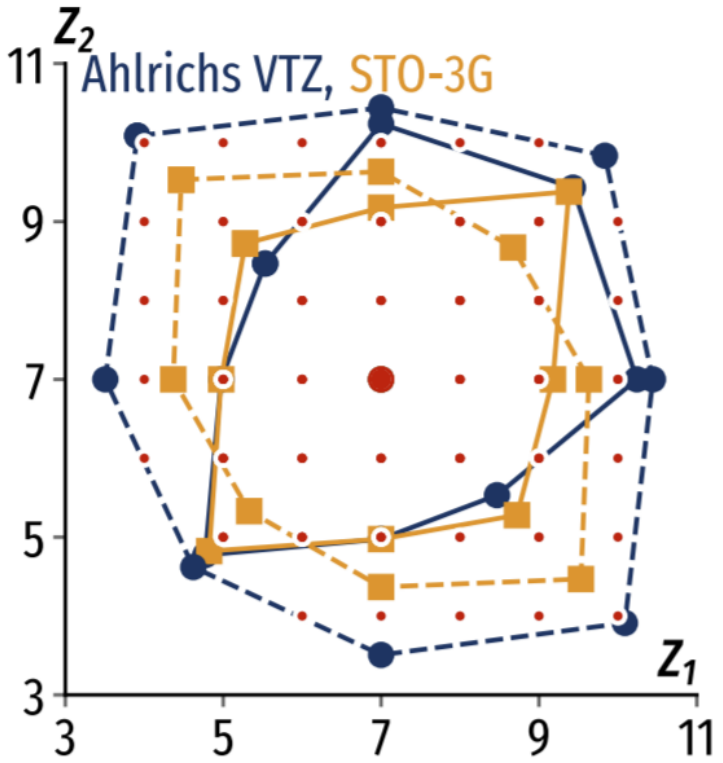
HF/STO-3G

Taylor expansion

- First terms accurate enough
  - Truncate early
- Converges to the right value
- Large convergence radius
- Scales with chemical space

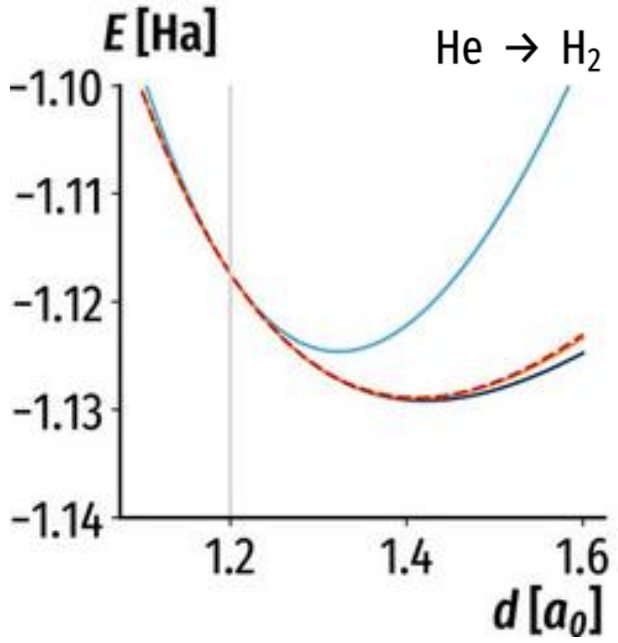
Arbitrary precision Hartree-Fock

NablaChem/APHF



GFvR, *J. Chem. Phys.* 2021.

Geometry Relaxation



Taylor expansion

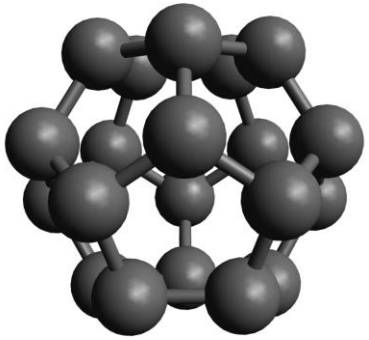
- Large changes still converge (more slowly)
- Geometric response can be recovered

GFvR, *J. Chem. Phys.* 2021.

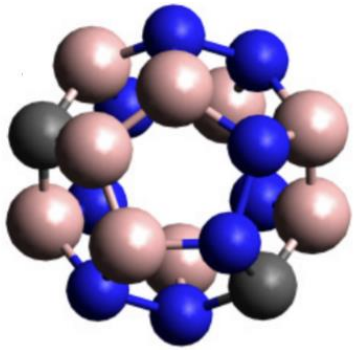
Covalent Interactions

Scaling with chemical space

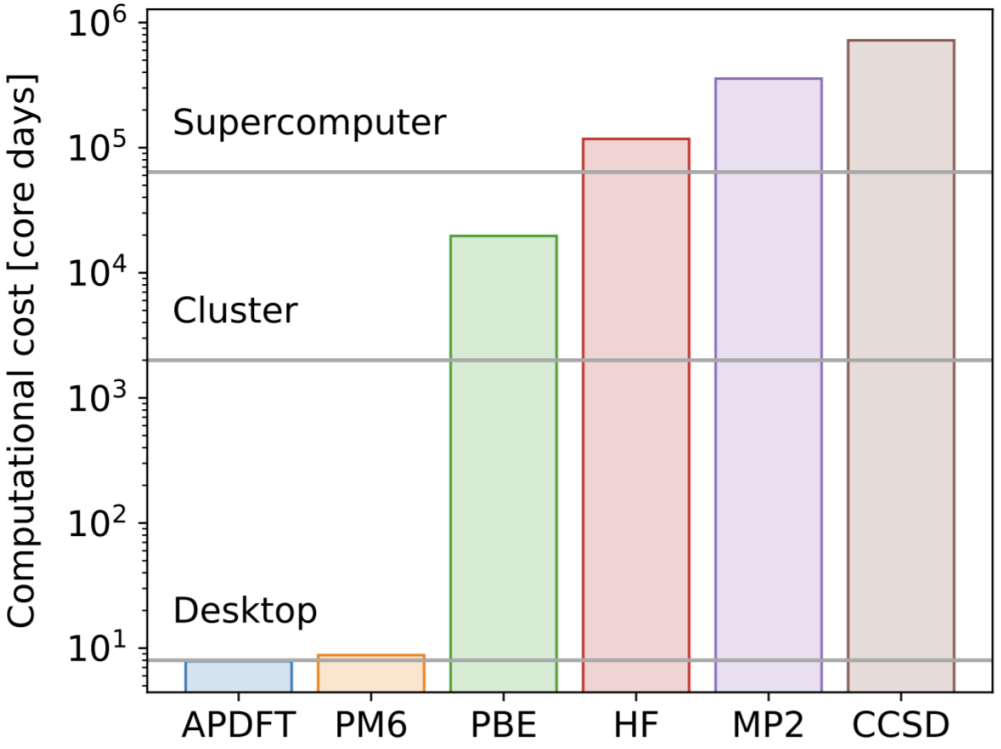
- 1 derivative for second order
- 5 derivatives for third order



C_{20}



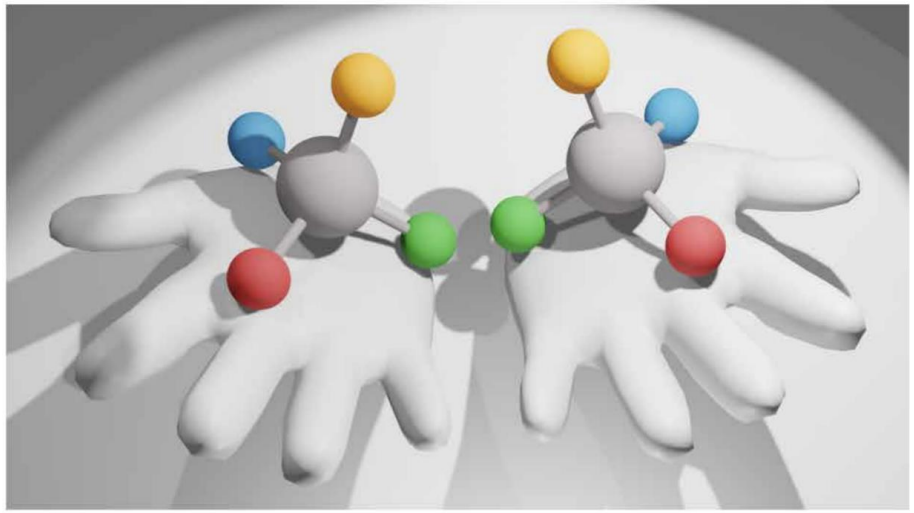
$3.1 \cdot 10^6$ targets

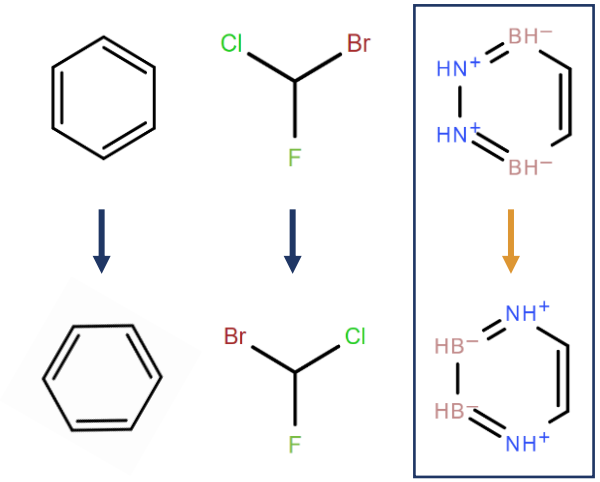


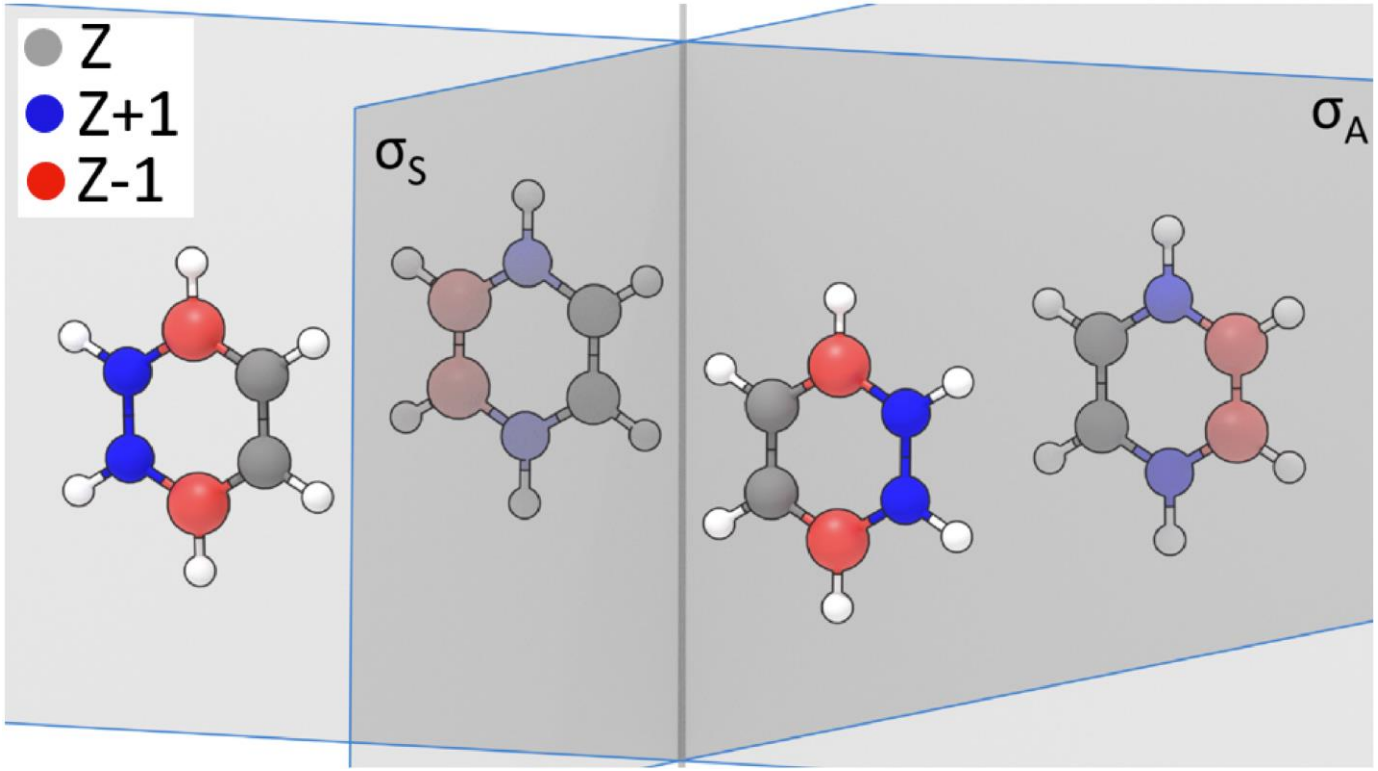
QA: 80.000x faster

GFvR, O. A. von Lilienfeld, *Phys. Rev. Res.*, 2020.

Alchemical Enantiomers





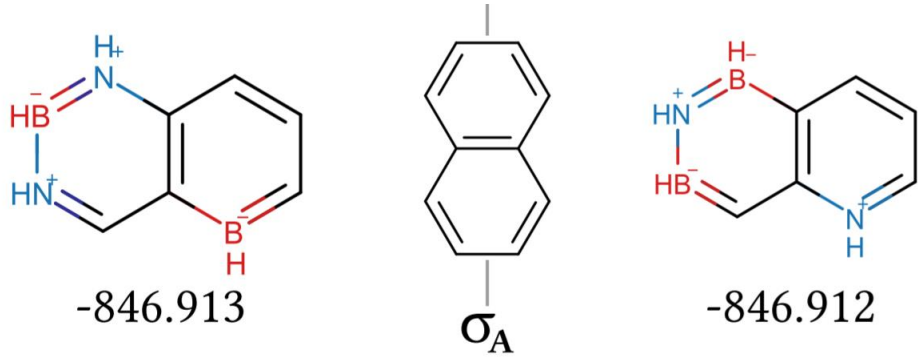


GFvR, O. A. von Lilienfeld, *Science Adv.* 2021.

Alchemical Enantiomers

Fundamentally new symmetry

Electronic energy only



Bond energy rules

Consecutive Elements

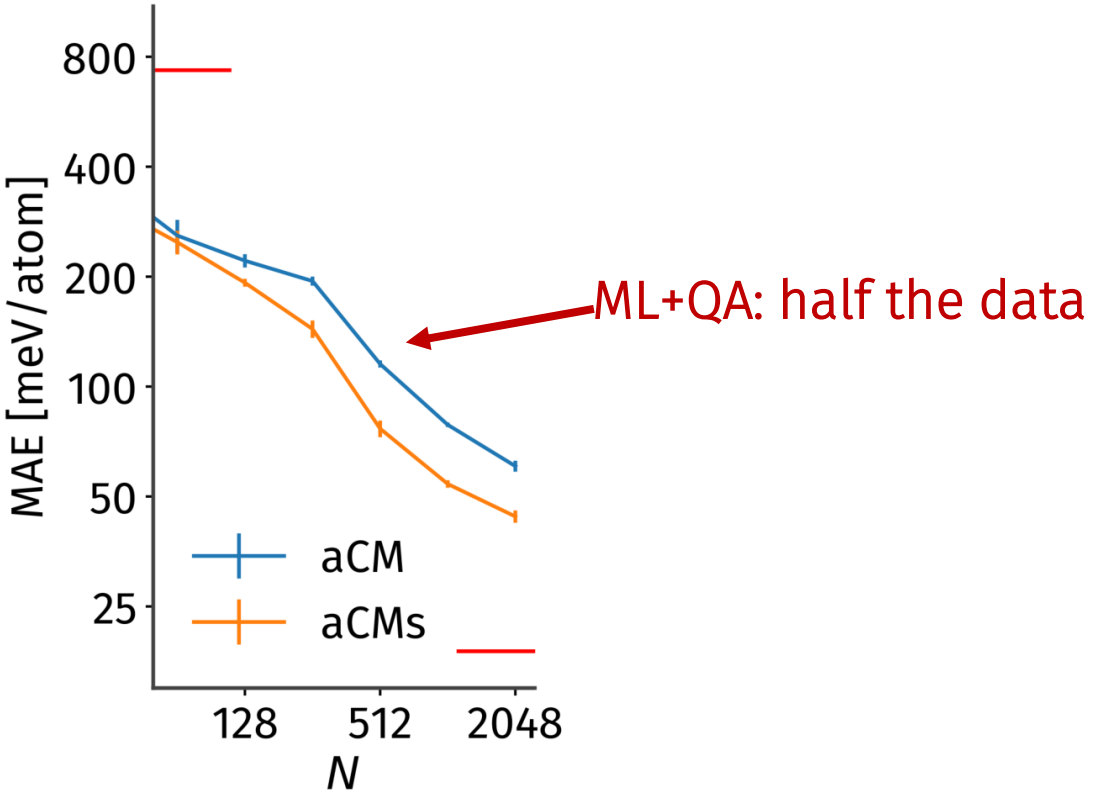
Q R S

B C N

$$E_{\mathrm{QR}} \simeq E_{\mathrm{SR}} + 0.5(E_{\mathrm{QQ}} - E_{\mathrm{SS}})$$

+ more rules for angles

Speed up machine learning



GFvR, O. A. von Lilienfeld, *Science Adv.* 2021.

PECD Angular Emission

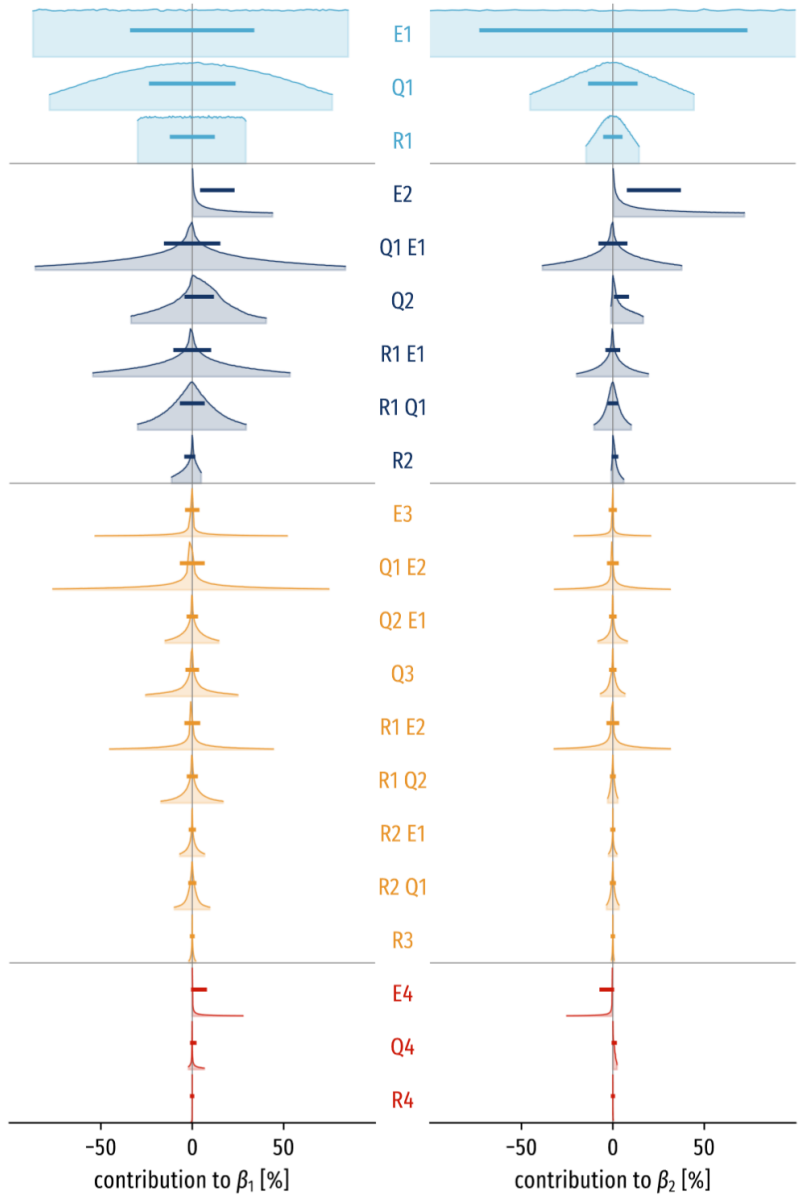
Angular emission

- Expensive to calculate
- Highly coupled degrees of freedom: multidimensional expansion

$$\frac{\mathrm{d}\sigma^{\pm}}{\mathrm{d}\Omega} = \frac{\sigma}{4\pi}\left[1 \pm \beta_1 P_1(\cos\theta) - \frac{1}{2}\beta_2 P_2(\cos\theta)\right]$$

dichroic parameter

anisotropy parameter



GFvR, A Artemyev, B Lagutin, P Demekhin, *J. Chem. Phys.*, 2024.

Quantifying Dimensions

Definition Intrinsic Dimension

Minimal number of degrees of freedom to describe a property.

≠ intrinsic dimension of a point cloud!

$$f(x, y) = x + y$$

Example: Dimers

- Energy: 3 dimensions
- Net charge: 1 dimension

$$E(\mathbf{R}_I, \mathbf{Z}_I)$$

Distance

Net charge Z_1+Z_2
Asymmetry Z_1-Z_2

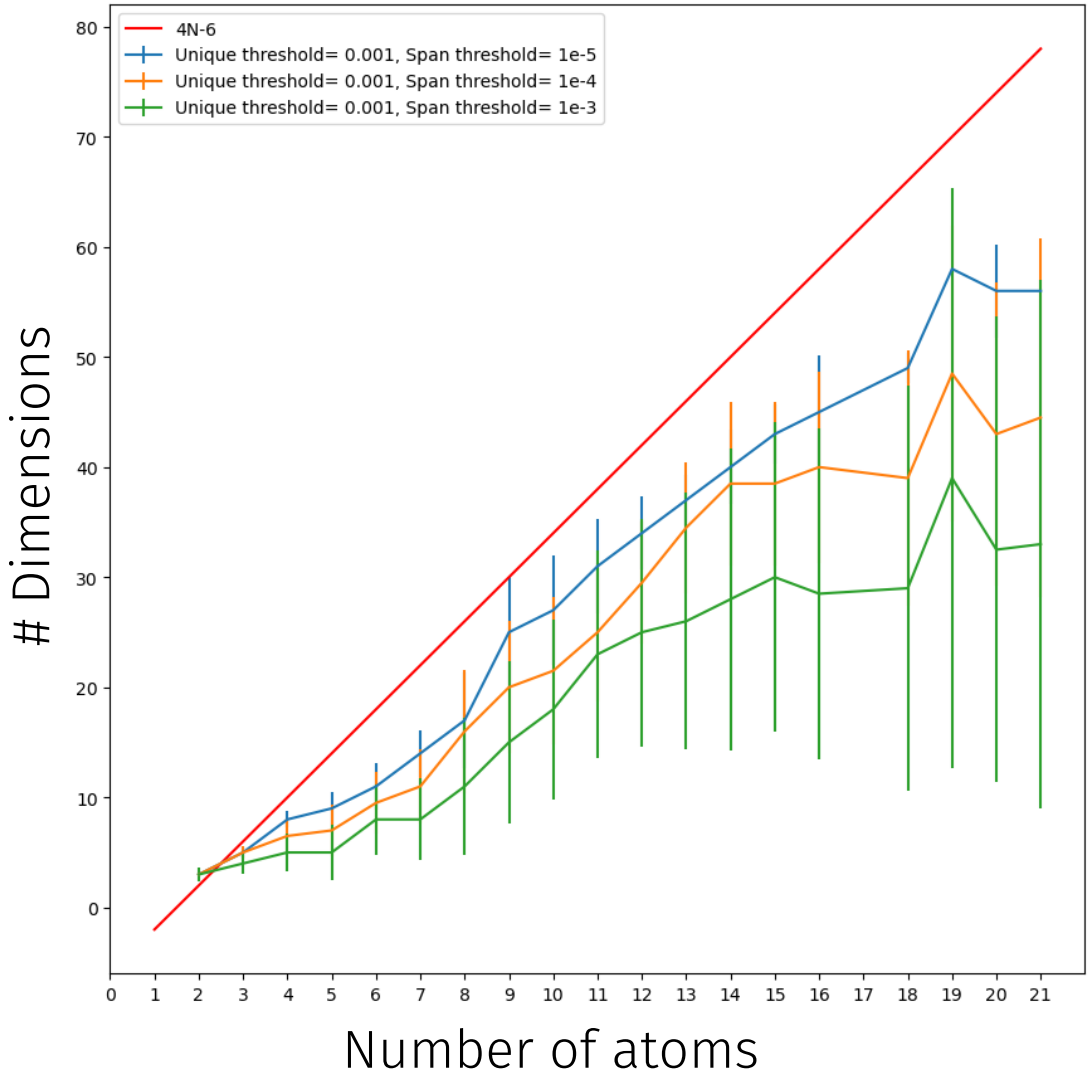
Example: Atom

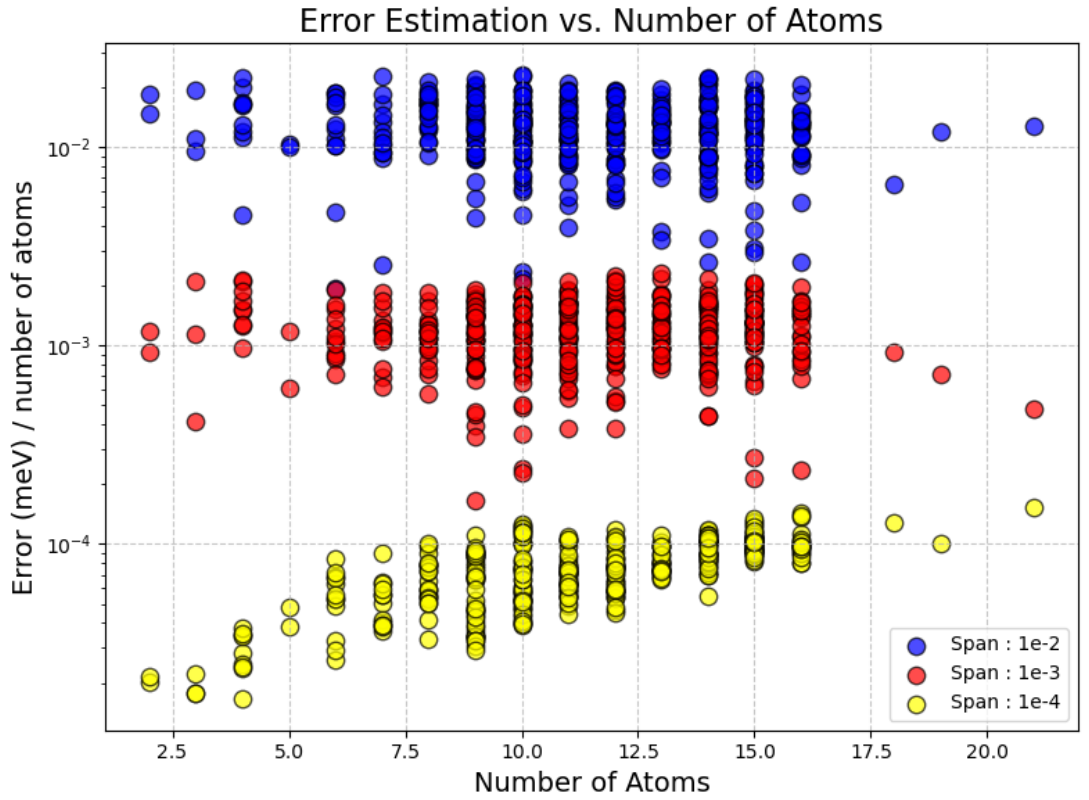
1 dimension
1 dimension

Ali Banjafar

A Banjafar, **GFvR**, *in preparation.*

Quantifying Dimensions





A Banjafar, **GFvR**, *in preparation.*

Limitations

Both energy and density derivatives are hard

- Finite differences — expensive, numerical instabilities
- Hellmann-Feynman — finite order, basis set inaccurate
- Coupled-perturbed — finite order, tedious
- Automatic differentiation — niche: DiffiQult, quax, dqc, …

ferchault/APDFT

NablaChem/APHF

aspuru-guzik-group/DiffiQult

CCQC/Quax

diffqc/dqc Alchemy!

fishjojo/pyscfad

Gaussian basis sets not overly cooperative

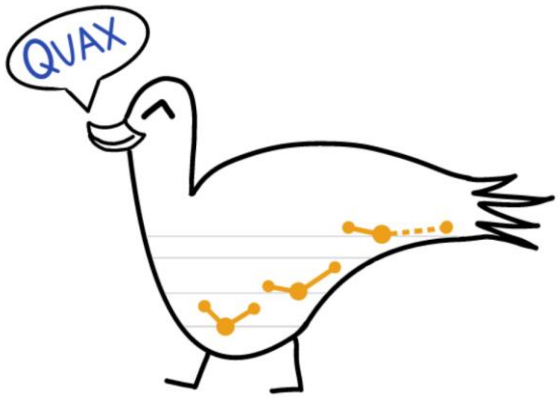
- Elements are discrete, derivatives are not
- Density converges more slowly with basis set quality than energy
- Unless complete basis set limit: Pulay terms

Pseudopotentials

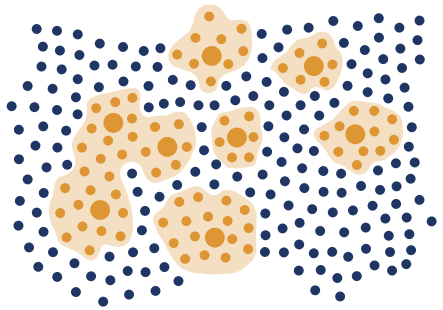
- Discrete Elements, parameters not differentiable

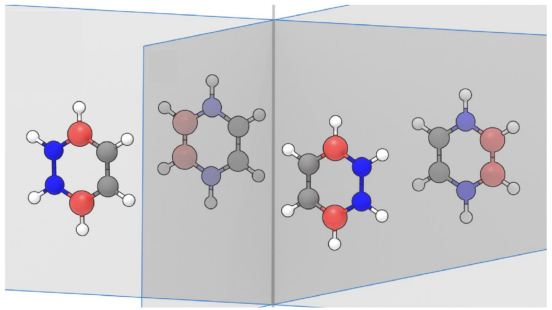
Convergence

- Finite radius
- Not all systems are made equally



Summary





Efficient | Re-use knowledge, no one-by-one

Symmetries | Reducing (“folding”) search space

Physics-driven | Predictive power, hard constraint

Differentiable Chemistry | Arbitrary derivatives of many properties

Closed-Form | Explainable and shows structure

Fast Design | Can assess whole regions at once

Thanks

Anton Artemyev
Ali Banjafar
Marco Bragato
Philipp Demekhin
Giorgio Domenichini
Emily Eikey
Chasz Griego
Nicolas Grimblat
John Keith
Simon Krug
Boris Lagutin
Anatole von Lilienfeld
Alex Maldonado
Michael Sahre

vonrudorff@uni-kassel.de
nablachem.org/talks
ferchault
@ferchault

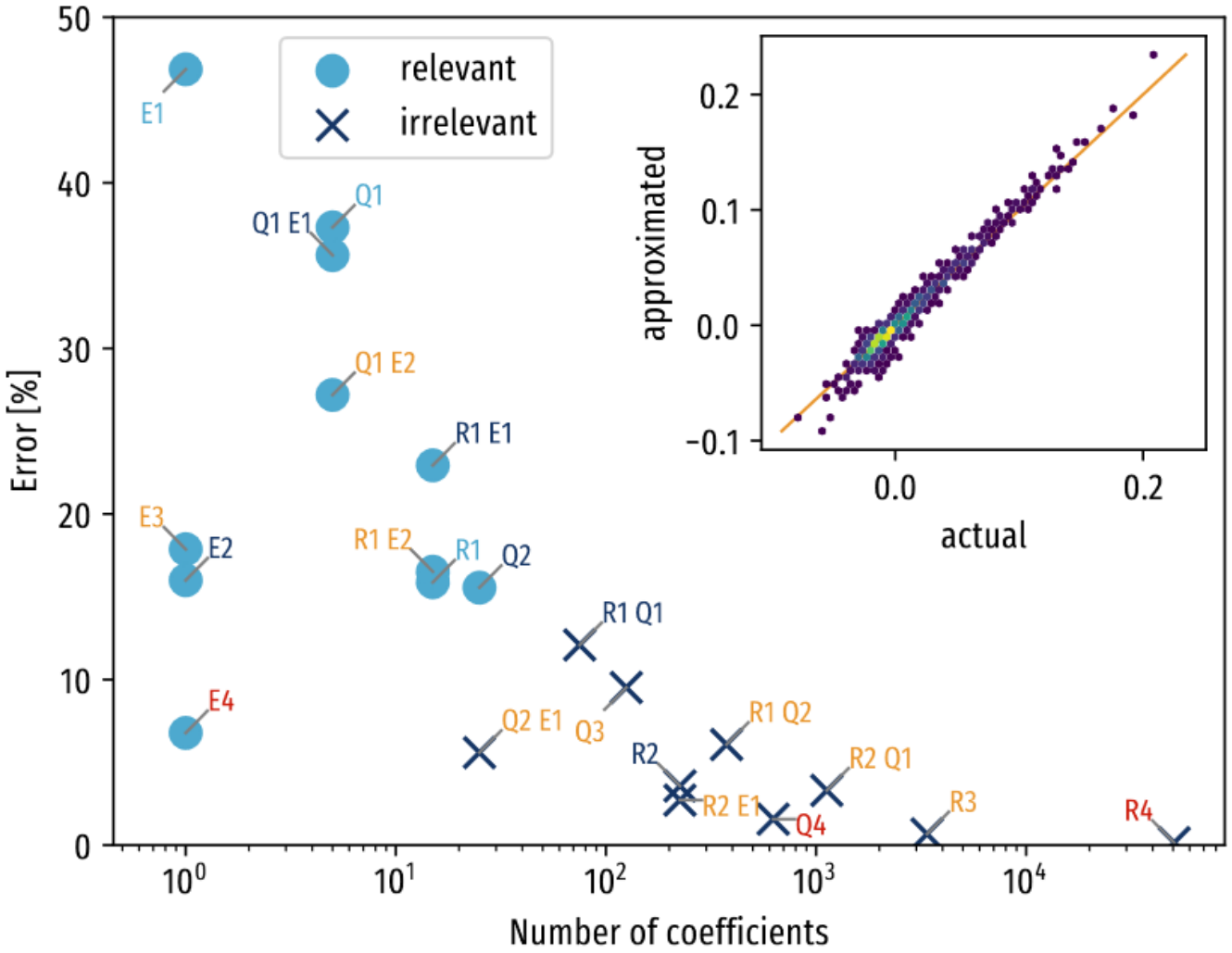
Towards Molecules

Relate to cost

- Expensive terms also irrelevant
- For molecules, subset only
- Allows for efficient stencil design

Estimated speedup

$$\frac{3^N}{N^2 + 12N - 1}$$



GFvR, A Artemyev, B Lagutin, P Demekhin, *J. Chem. Phys.*, 2024.