

# Exploring Regions in Chemical Space with Perturbations

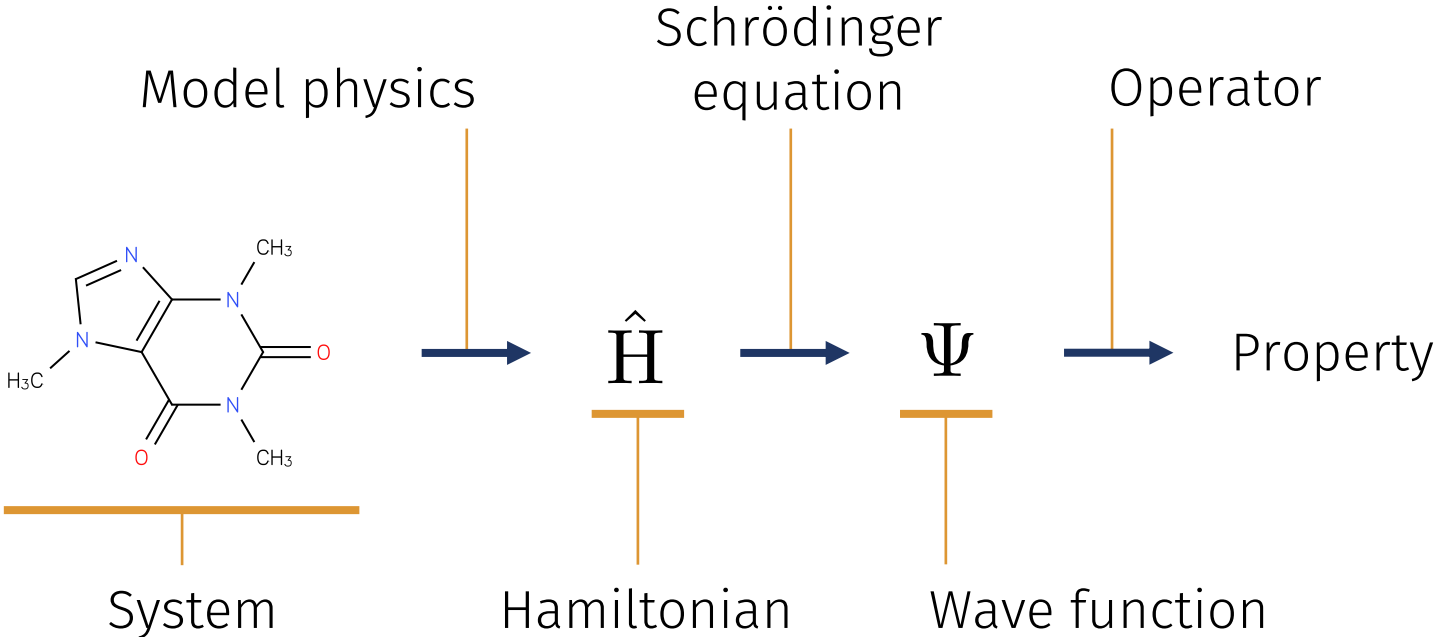
Guido Falk von Rudorff, University of Kassel

 [vonrudorff@uni-kassel.de](mailto:vonrudorff@uni-kassel.de)

 [nablachem.org/talks](https://nablachem.org/talks)

 [ferchault](https://github.com/ferchault)

 [@ferchault](https://twitter.com/ferchault)

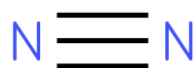


$$\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_n)$$

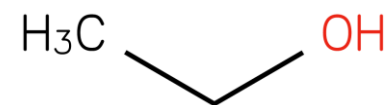
Methane



N<sub>2</sub>



Ethanol



Solved by approximations in computational chemistry?

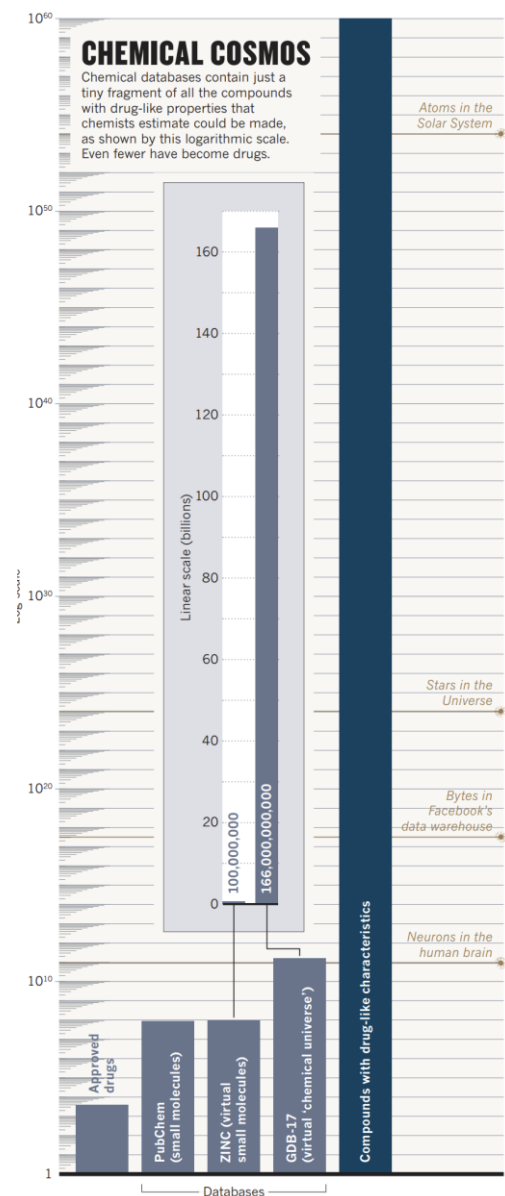


## Commercial databases

- 164 million molecules
- 15k added daily

## Scale

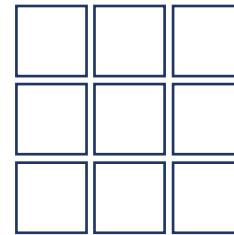
- One person: 1 million compounds/second
- 10 billion people on earth
- $10^{26}$  universe ages to go through



Face centered cubic and 70 elements only

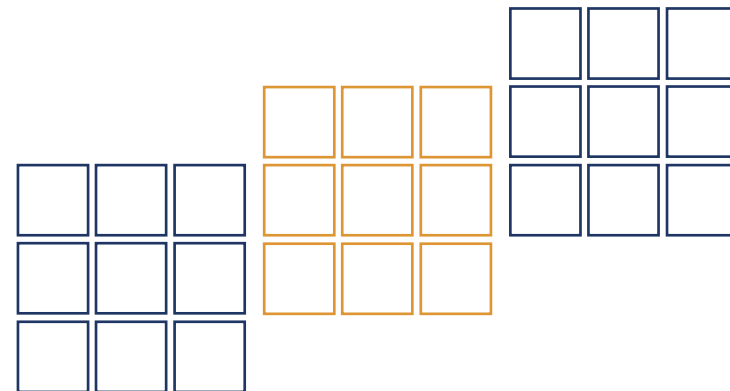
## 9 primitive cells

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

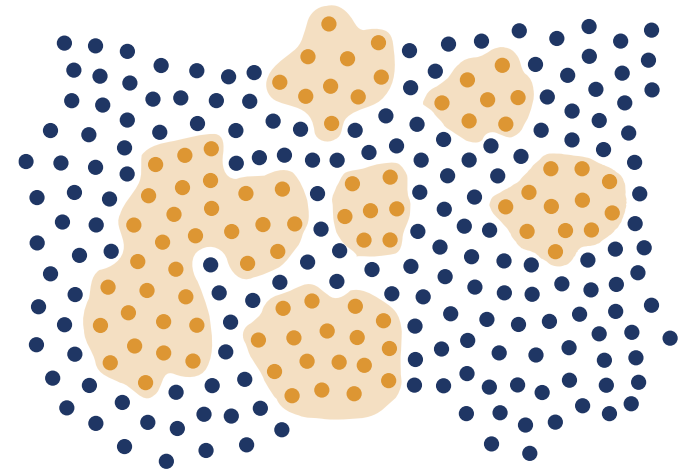
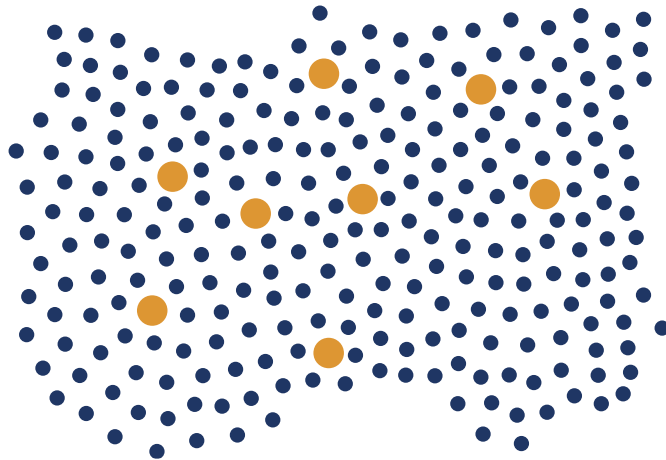


## 27 primitive cells

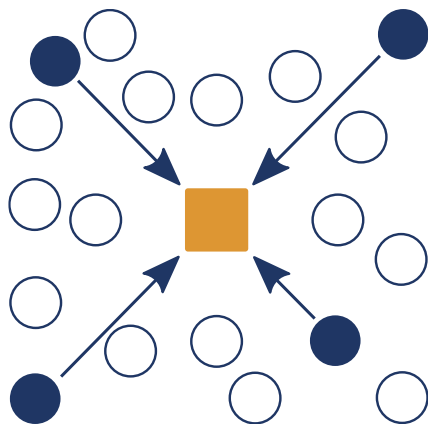
- Binary:  $\sim 10^{17}$
- Ternary:  $\sim 10^{29}$
- Quaternary:  $\sim 10^{36}$



Speed does not matter:  
even enumeration is impossible.



## Machine Learning



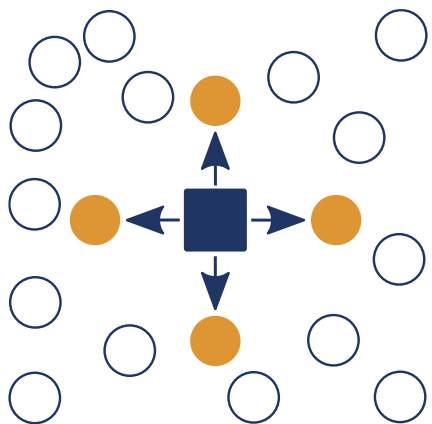
Foundations | Statistical modelling

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

## Idea

Treat system changes perturbatively<sup>[1,2]</sup>

Build a Taylor/Padé approximant<sup>[3]</sup>: often 100.000 times faster

## Steps

Choose system



Alter system, calculate property response functions



Predict many modified systems



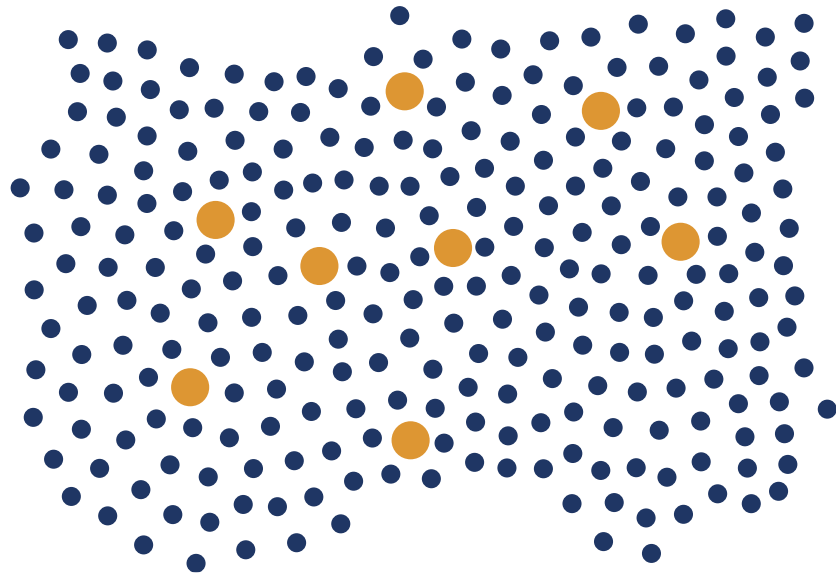
■ Forwards  
■ Backwards



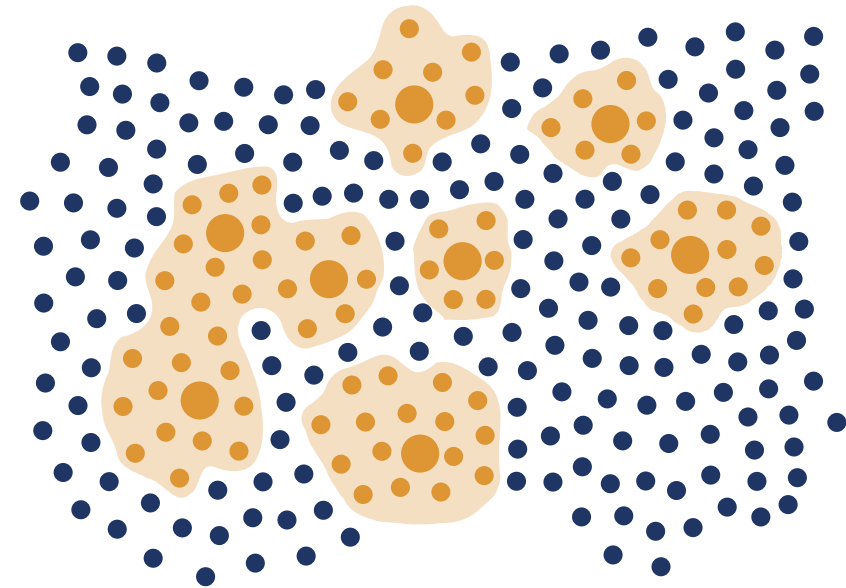
Few highly accurate calculations  
instead of many intermediate ones

$$\hat{H} = \hat{H}(\underbrace{Z_i}_{4N}, \underbrace{\mathbf{R}_i}_{1D, \text{ close to } \sum_i Z_i}, \underbrace{N_e}_{1D}, \sigma)$$

Without Perturbation



With Perturbation



Systems

- Any
- Known
- Approximated

## Differentiable / Analytic + Converge quickly

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

## In progress

NMR spectra

Tested in: Gaussian, Psi4, PySCF, MRCC, ORCA, cp2k, CPMD

Tested with: HF, DFT, CCSD

**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.

Interpolate between molecular isoelectronic Hamiltonians

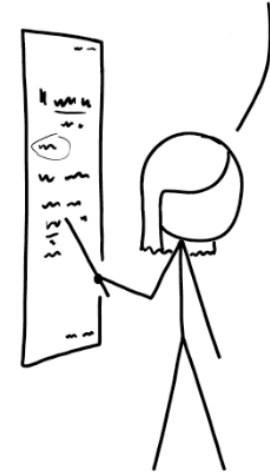
$$\hat{H}(\lambda) \equiv \lambda \hat{H}_t + (1 - \lambda) \hat{H}_r$$

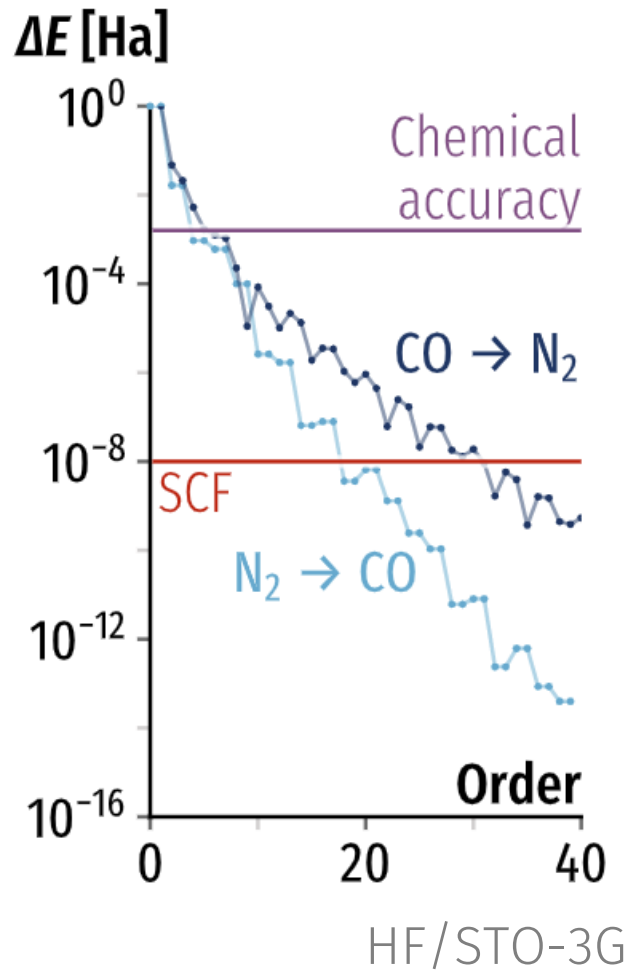
Taylor expansion around reference molecule

$$E_t = E_r + \Delta E^{\text{NN}} + \int_{\Omega} d\mathbf{r} \sum_{n=0}^{\infty} \frac{1}{(n+1)!} \Delta v \frac{\partial^n \rho_{\lambda}(\mathbf{r})}{\partial \lambda^n} \Big|_{\lambda=0}$$

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

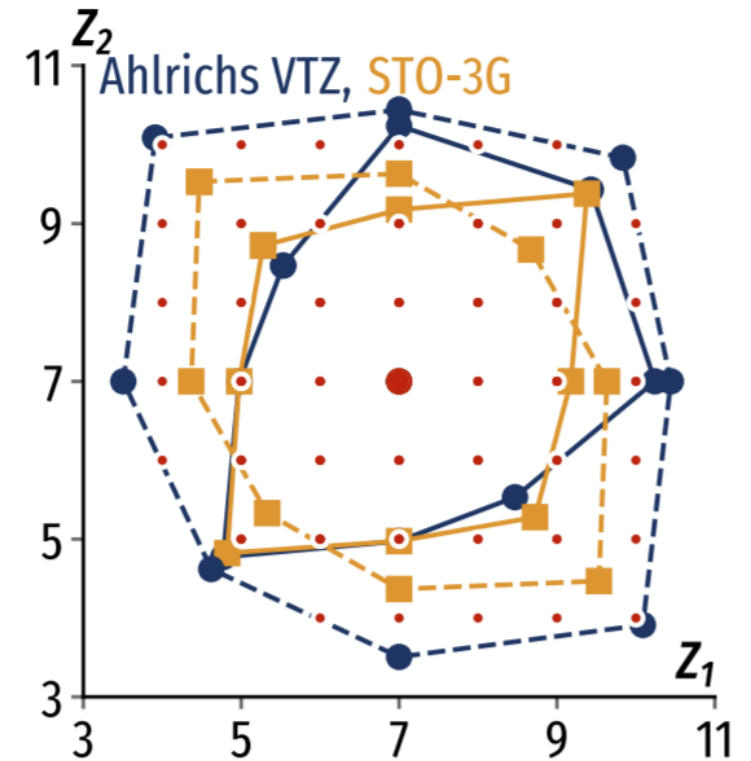
AT THIS POINT, YOU'RE PROBABLY THINKING, "I LOVE THIS EQUATION AND WISH IT WOULD NEVER END!"  
WELL, GOOD NEWS!

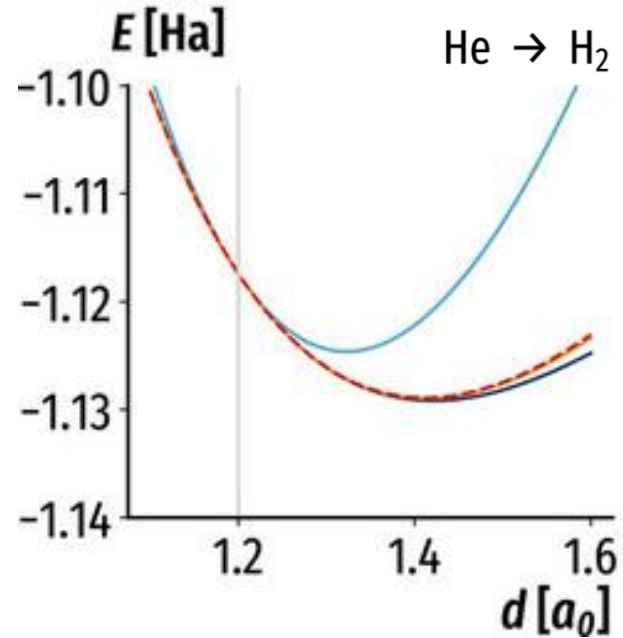




## Taylor expansion

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



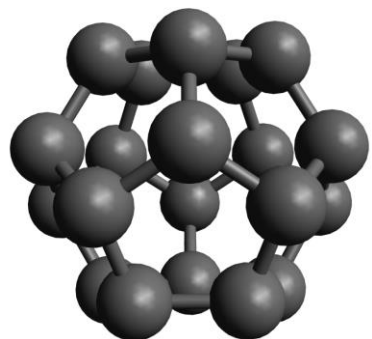


## Taylor expansion

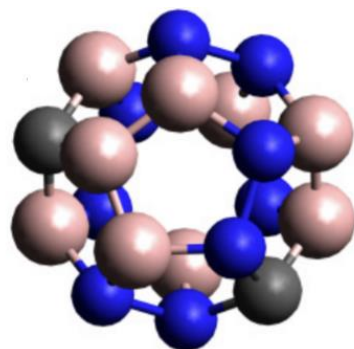
- Large changes still converge (more slowly)
- Geometric response can be recovered

## Scaling with chemical space

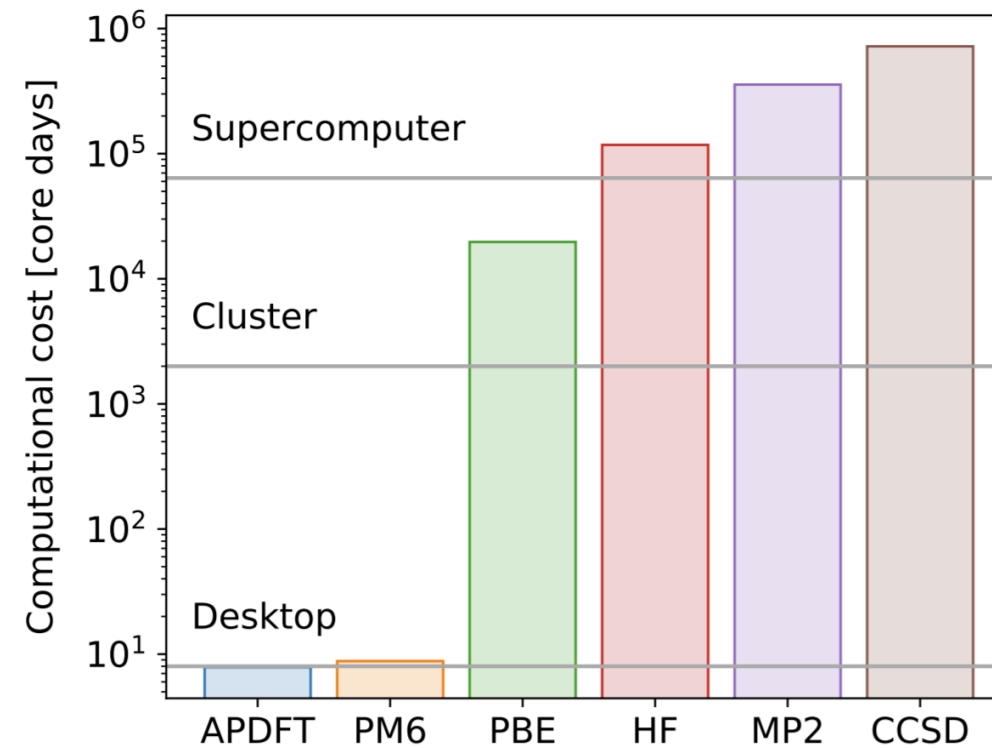
- 1 derivative for second order
- 5 derivatives for third order



$C_{20}$



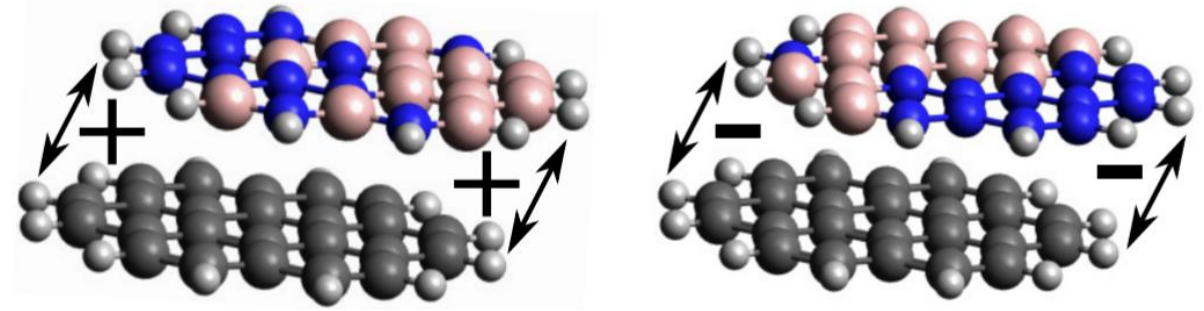
$3.1 \cdot 10^6$   
targets



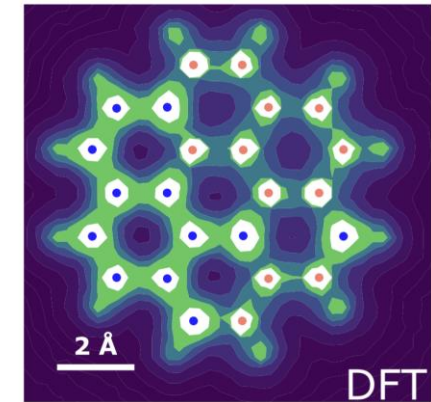
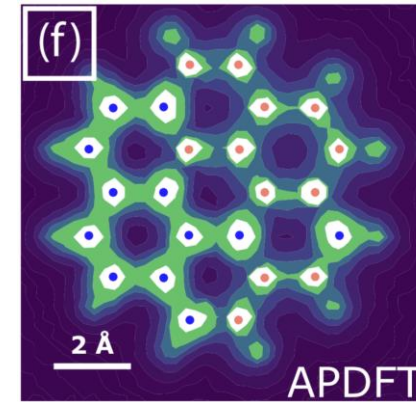
QA: 80.000x faster

## BN-doped coronene dimer

- Identify most/least attractive doping pattern
- Design case

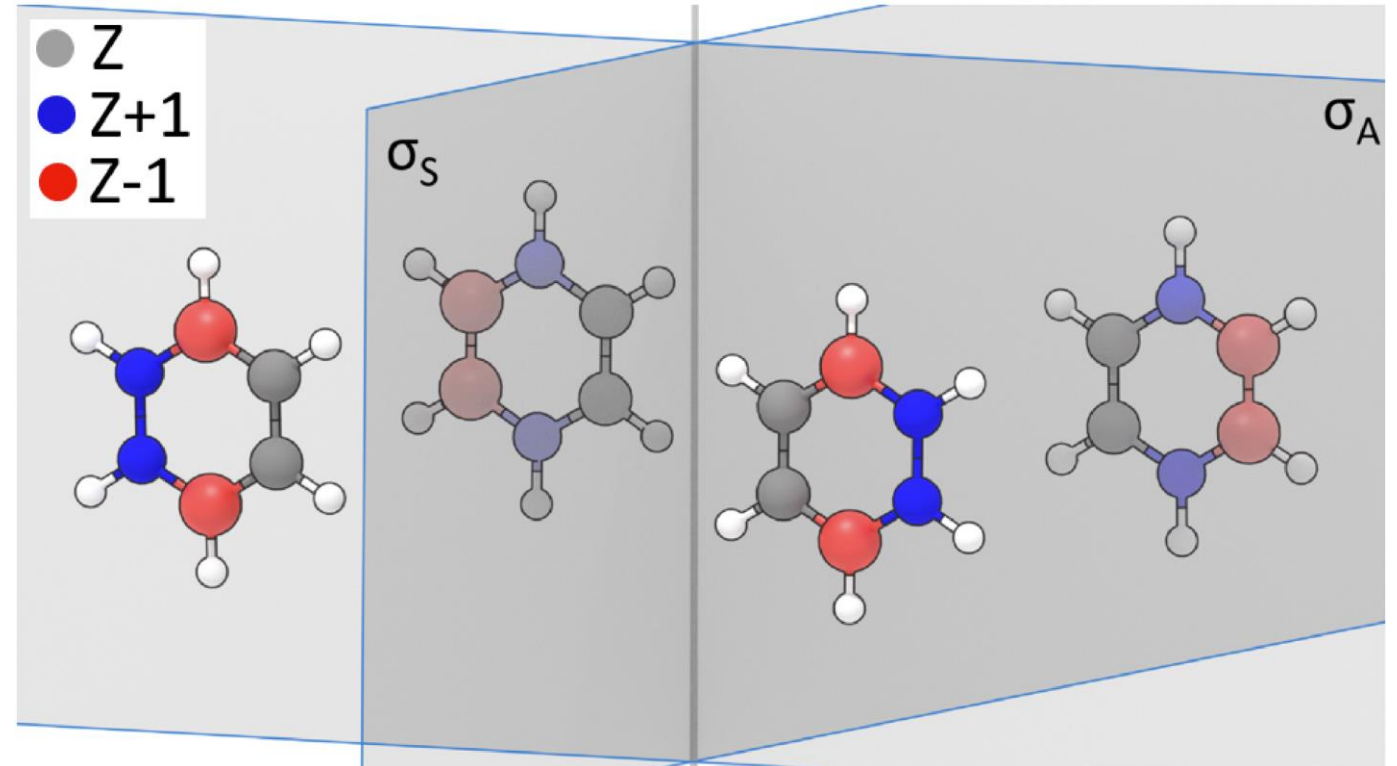
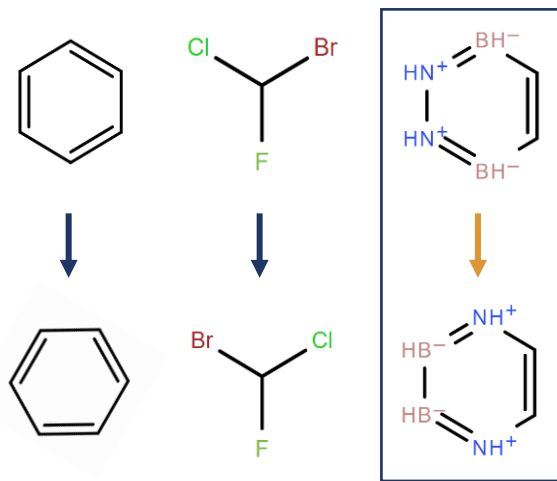
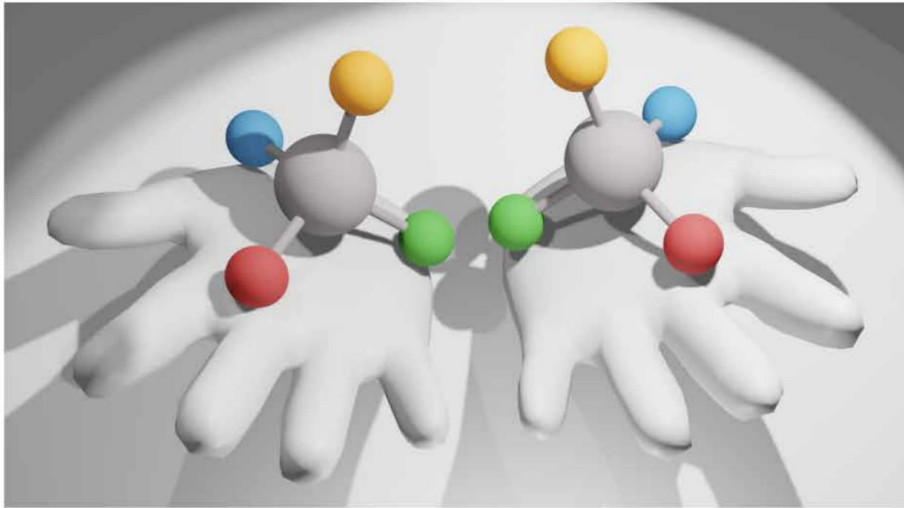


QA: 20.000x faster



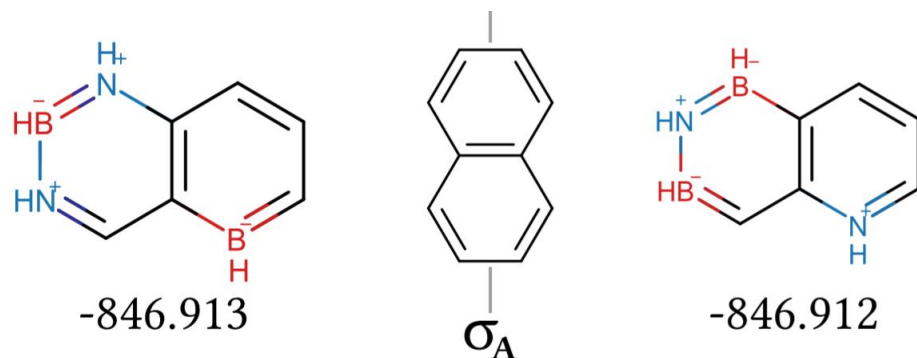
$2.8 \cdot 10^{10}$  targets





Fundamentally new symmetry

Electronic energy only



Bond energy rules

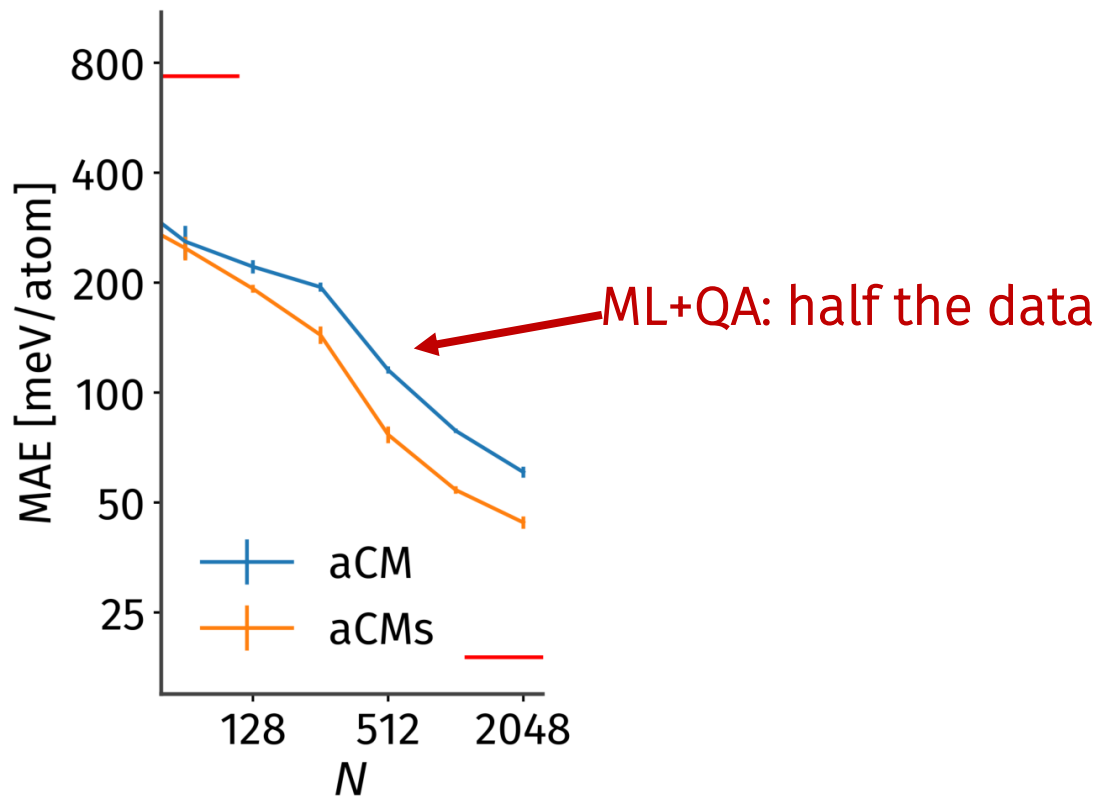
Consecutive Elements

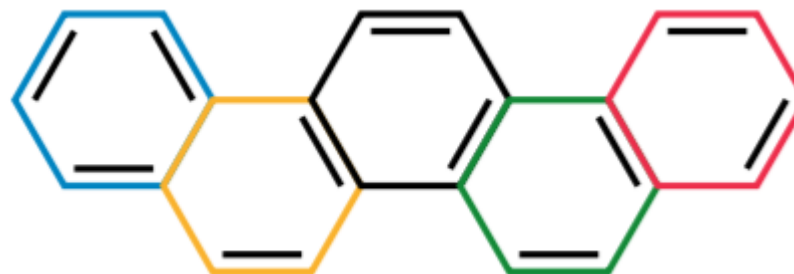
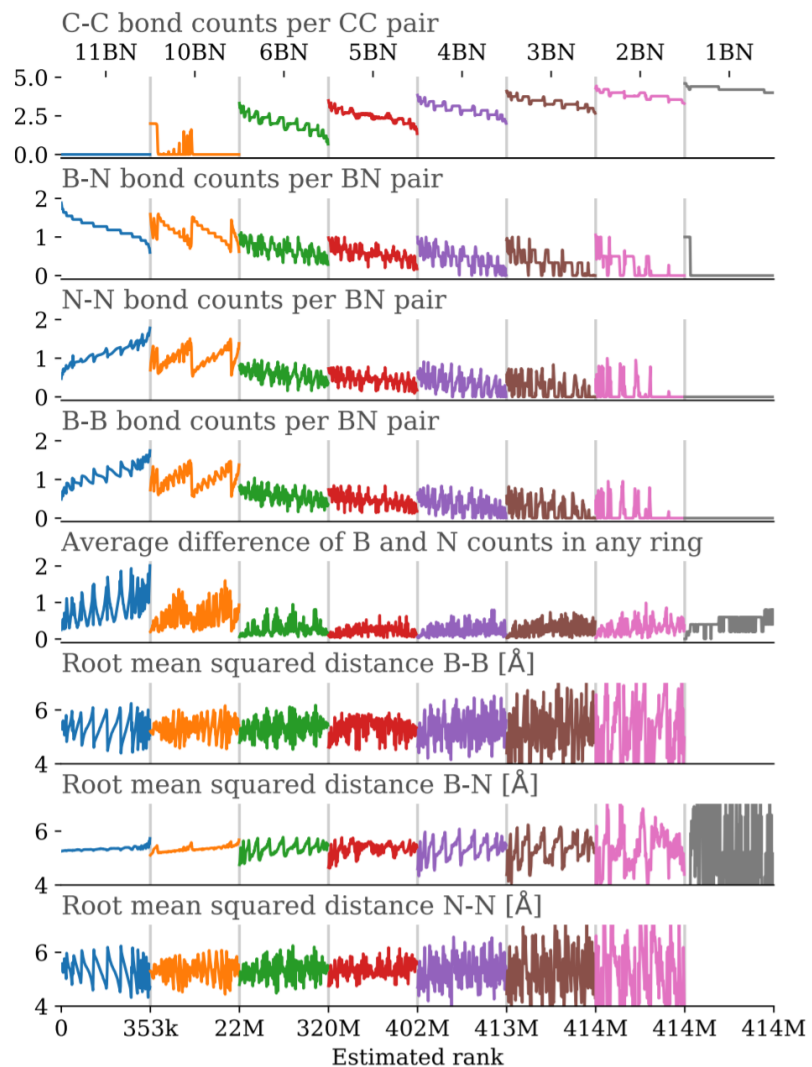
Q R S

B C N

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

Speed up machine learning





x 414 M

QA: Millions at once!

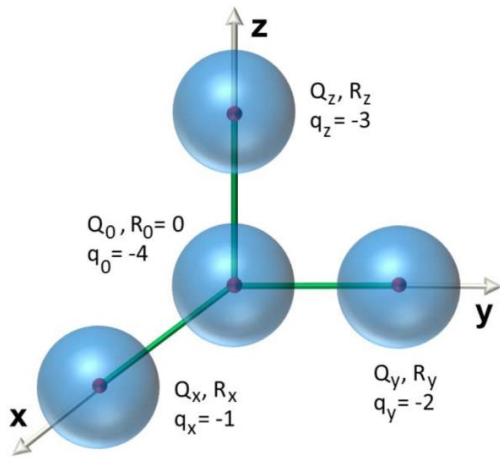
Design rules in order of decreasing strength

- Add BN pairs
- Maximize CC bonds
- Substitute sites shared between rings
- Maximize BN bonds
- Avoid N substitutions on rings sharing a larger amount of bonds with other rings
- Balance BN substitutions in each ring

Not a single QM calculation required!

## Angular emission

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



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

## Data

- 85 pairs of  $\beta_i$
- Center:  $Q_{xyz}=2.5$ ,  $R_{xyz} = 3$ , photoelectron = 6 eV

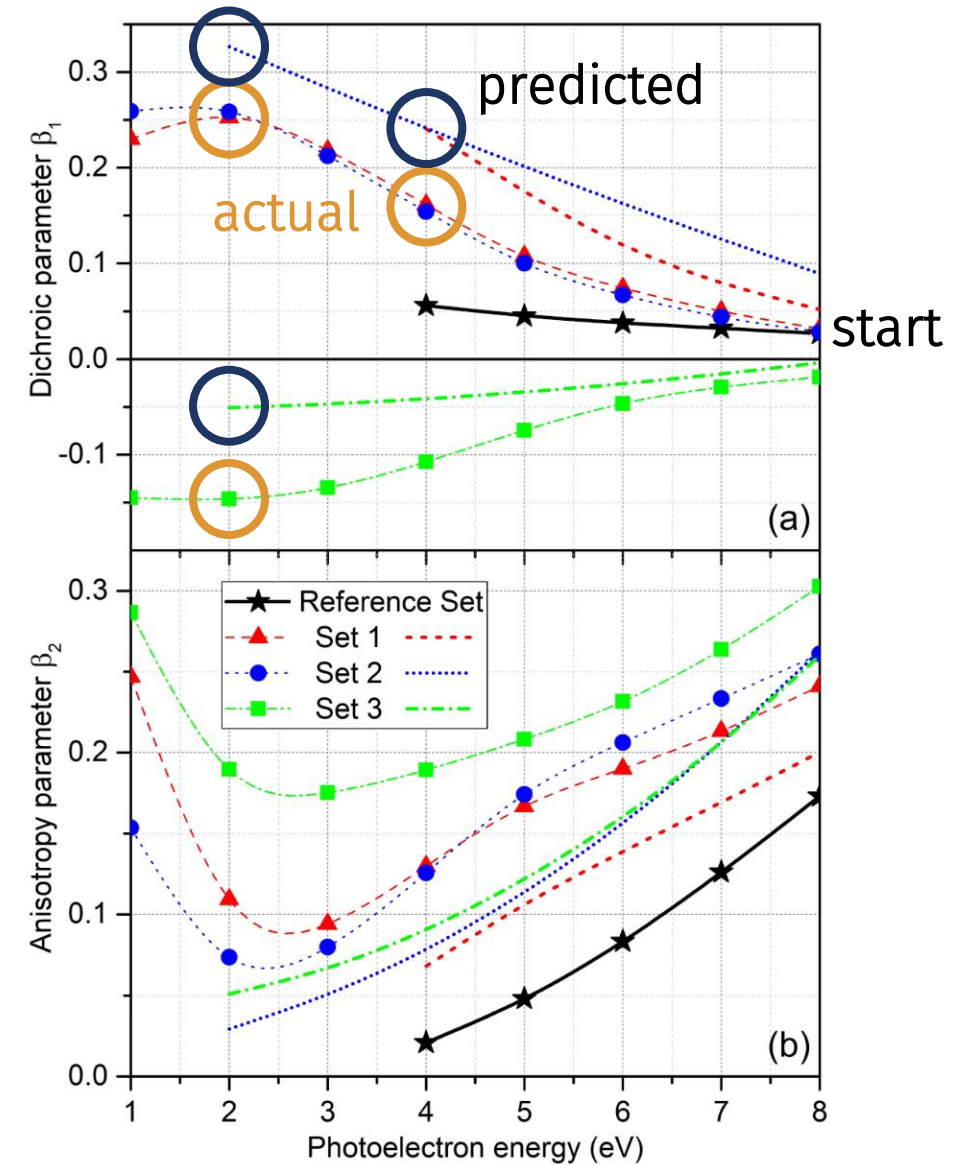
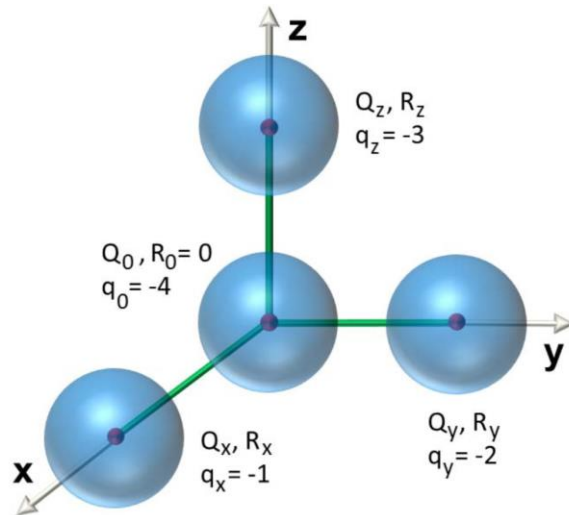
$$\beta_i(\mathbf{x}) \simeq \sum_{|\alpha| \leq k} \frac{\partial^{|\alpha|} \beta_i(\mathbf{a})}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}} \frac{(\mathbf{x} - \mathbf{a})^\alpha}{\alpha!}$$

## Objective

- Modify **all**  $Q_i, R_i$ , photoelectron energy
- Find extremal  $\beta_1$ , predict matching  $\beta_2$

## Results

- max  $\beta_1$ : 5.6%  $\rightarrow$  25.8%
- Min  $\beta_1$ : 5.6%  $\rightarrow$  -14.7%



## Group terms

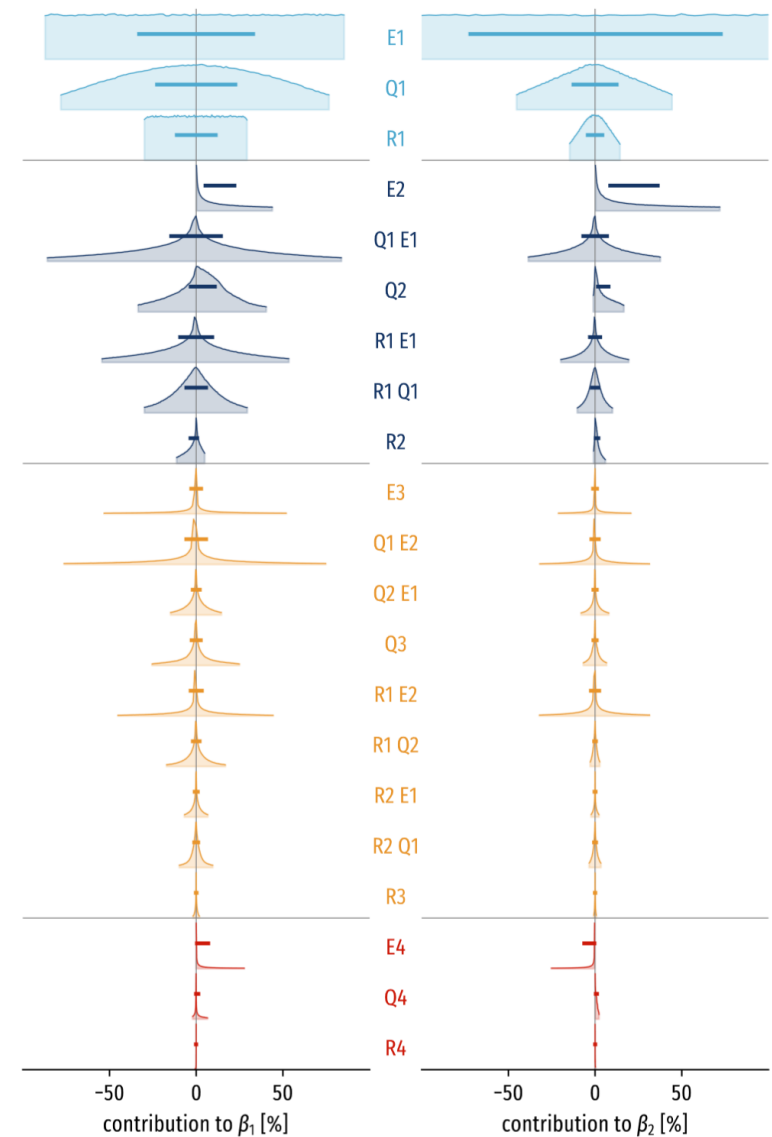
- $Q_i$ : Q
- $R_i$ : R
- photoelectron energy: E
- Mixed: count groups

$$\frac{\partial \beta_i}{\partial R_x} \quad \frac{\partial^2 \beta_i}{\partial Q_y^2} \quad \frac{\partial^2 \beta_i}{\partial R_x \partial Q_y}$$

∈ R1                      ∈ Q2                      ∈ R1 Q1

## Results

- Converges quickly
- Few third order terms contribute
- Almost no spatial-spatial coupling

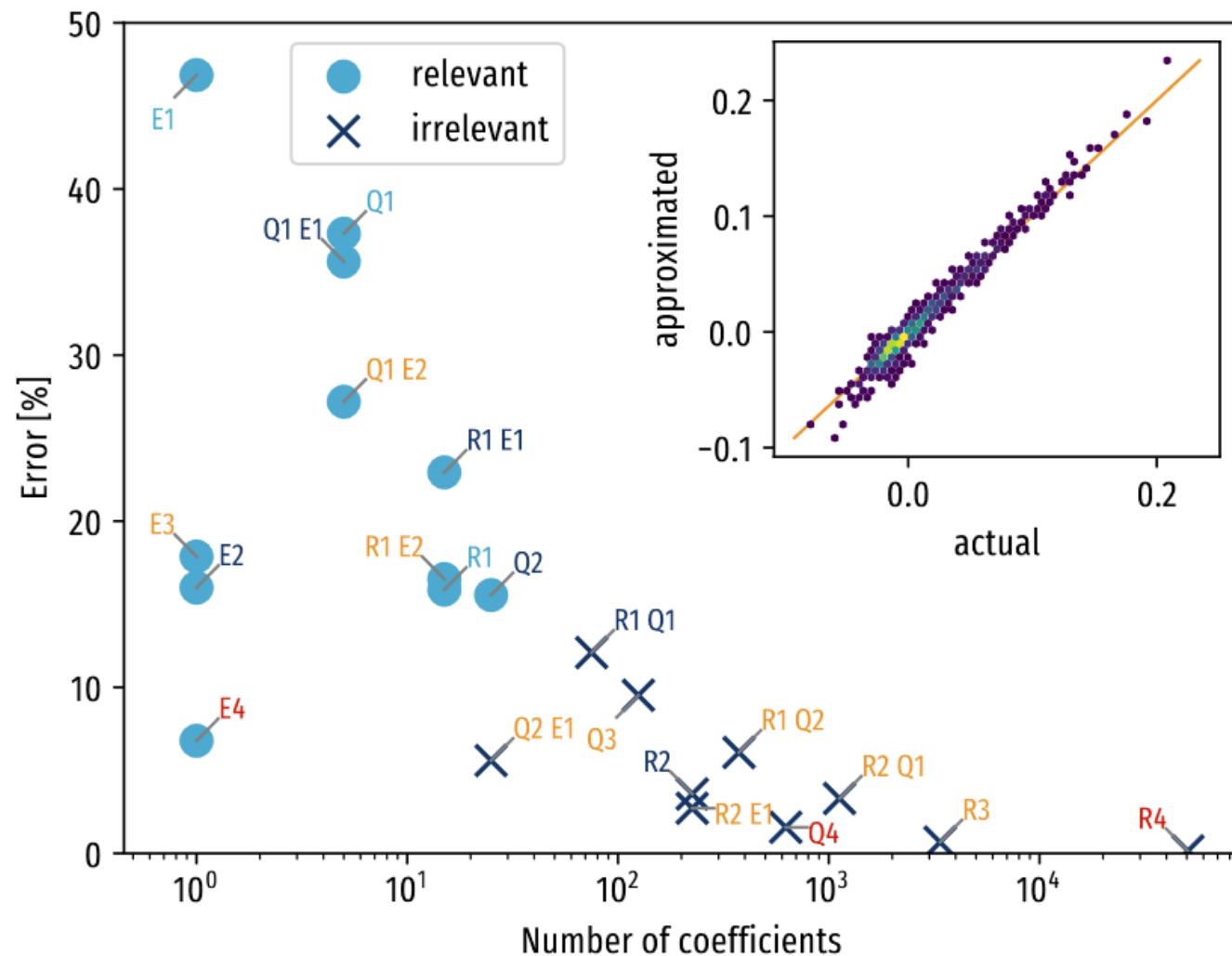


## Relate to cost

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

## Estimated speedup

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






## Both energy and density derivatives are hard

- Finite differences expensive, numerical instabilities
- Hellmann-Feynman finite order, basis set inaccurate
- Coupled-perturbed finite order, tedious
- Conceptual DFT memory hungry, no post-HF
- Automatic differentiation niche: DiffiQult, quax, dqc, ...

 ferchault/APDFT  ferchault/APHF

## Gaussian basis sets not overly cooperative

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

 aspuru-guzik-group/DiffiQult

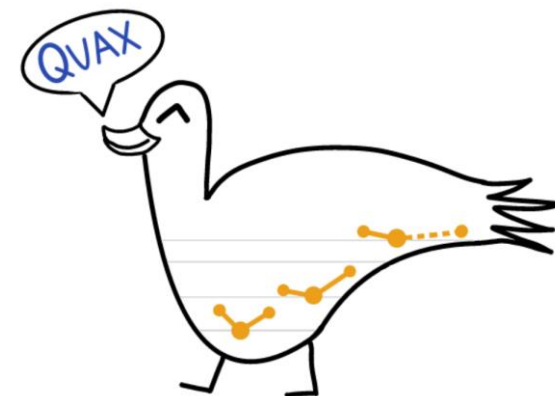
 CCQC/Quax

 diffqc/dqc **Alchemy!**

 fishjojo/pyscfad

## Convergence

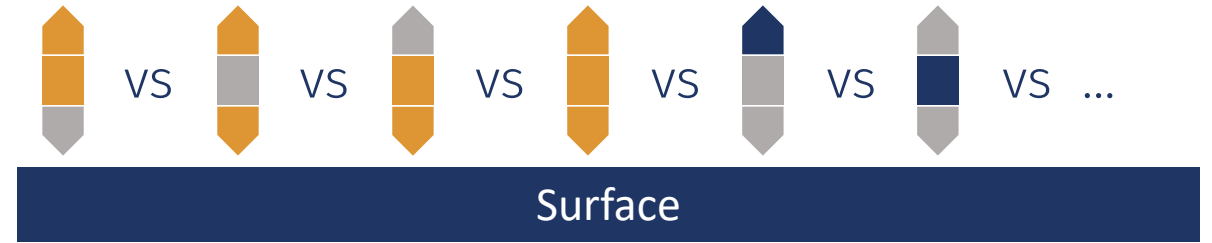
- Finite radius
- Not all systems are made equally





## Design space of similar systems

- Surface functionalization
- Self-assembly building blocks
- Binding motif

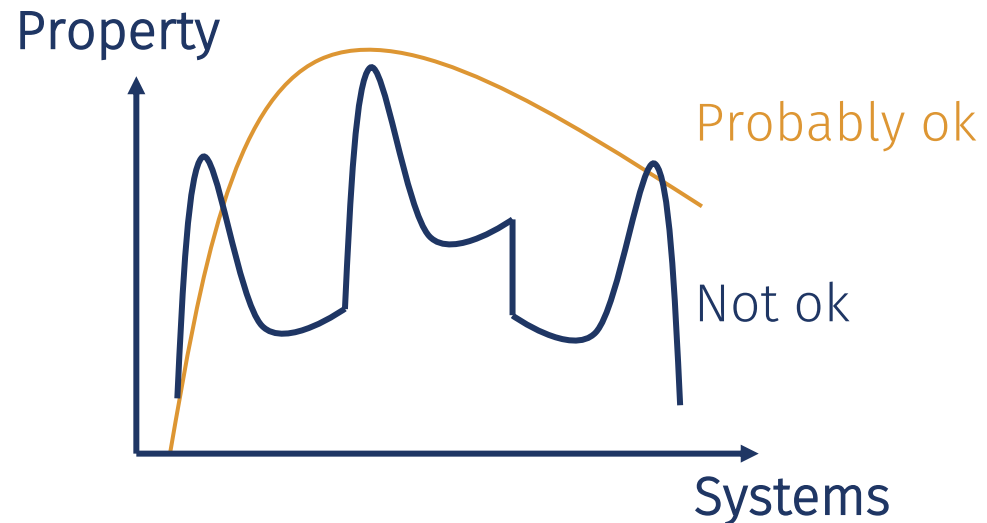


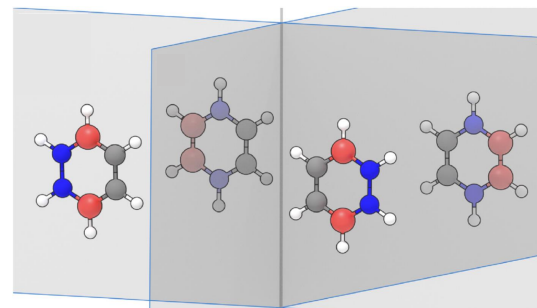
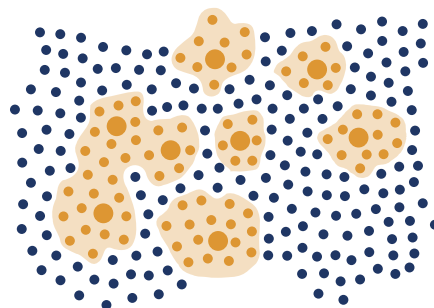
## Decompositions

- Global property
- Atomic or group contributions
- Explain models or *ab initio*

## Requirements

- Computationally accessible
- Smooth response





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

**Symmetries** | Reducing (“folding”) search space

**Physics-driven** | Predictive power

**Differentiable Chemistry** | Arbitrary derivatives of many properties

**Closed-Form** | Explainable and shows structure

**Fast Design** | Can assess whole regions at once

## Thanks

Anton Artemyev

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