

Compound Design with Alchemical Perturbations and Differentiable Quantum Chemistry

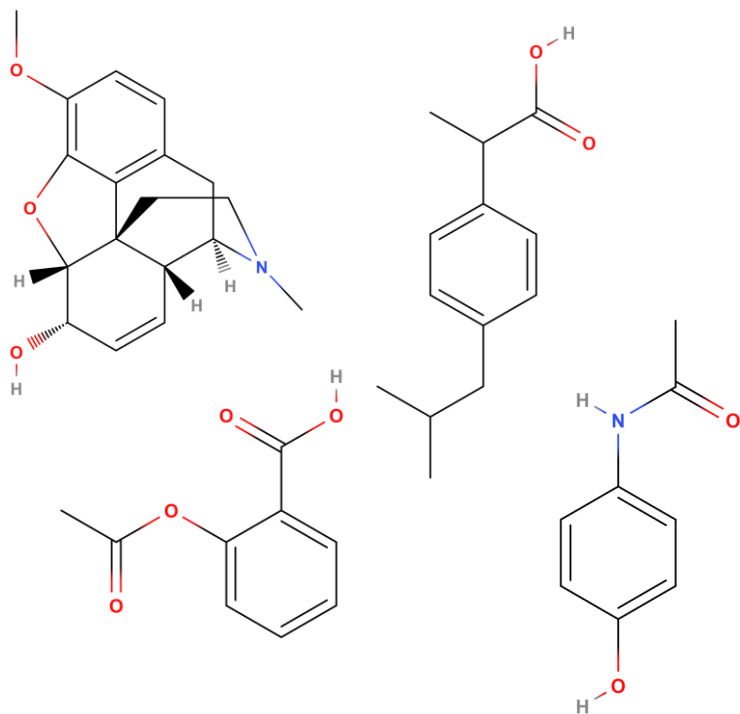
Guido Falk von Rudorff, University of Kassel

 ferchault

 @ferchault

 nablachem.org

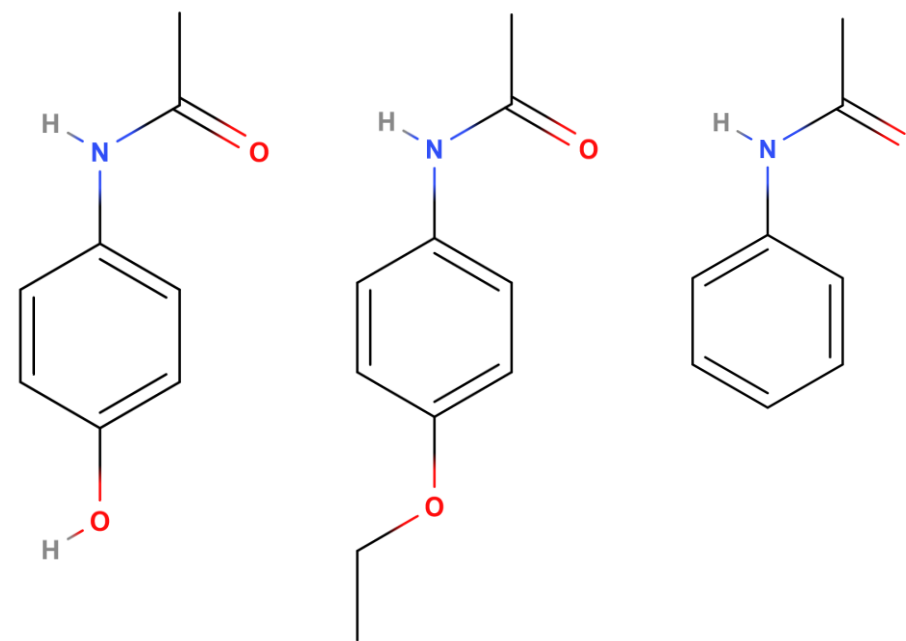
Design: sample by guided trial-and-error.



Global Search Problem

Which class of compounds?

Drug-like: 10^{60}

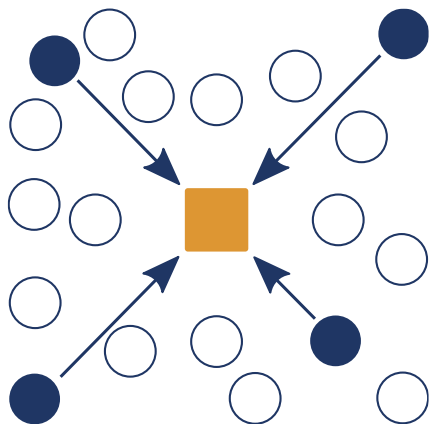


Local Search Problem

Which particular species within that class?

BN-doped 8x8 graphene: 10^{50}

Machine Learning



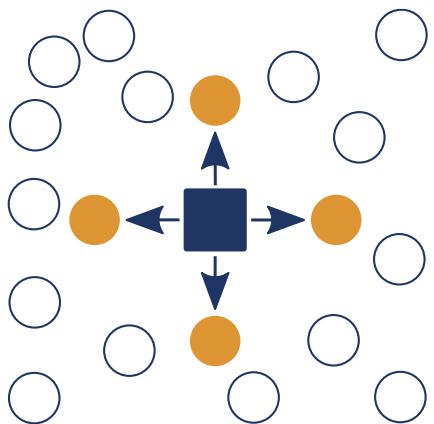
Foundations | Statistical modelling

Accuracy | Systematically improvable through data and training

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

Limitation | Requires training data, no black box

Quantum Alchemy



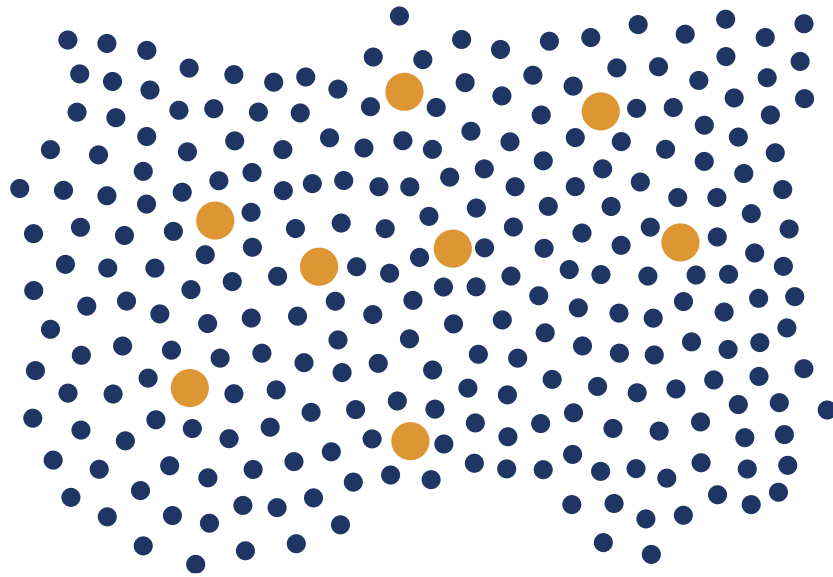
Foundations | Perturbation theory

Accuracy | Systematically improvable through higher orders terms

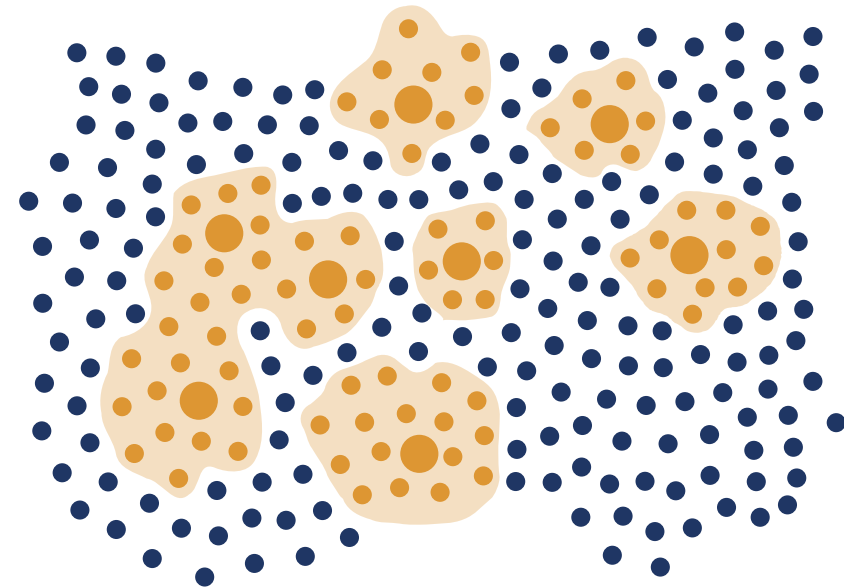
Specialty | Combinatorial scaling with chemical diversity

Limitation | Finite range in chemical space

Without Perturbation



With Perturbation



Systems/Molecules

- Any
- Known
- Approximated

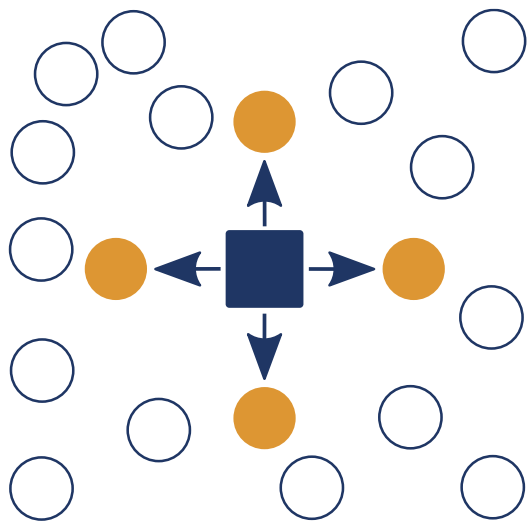
Perspective shift

Few highly accurate calculations
instead of many intermediate ones

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

4N 1D, close to $\sum_i Z_i$ 1D

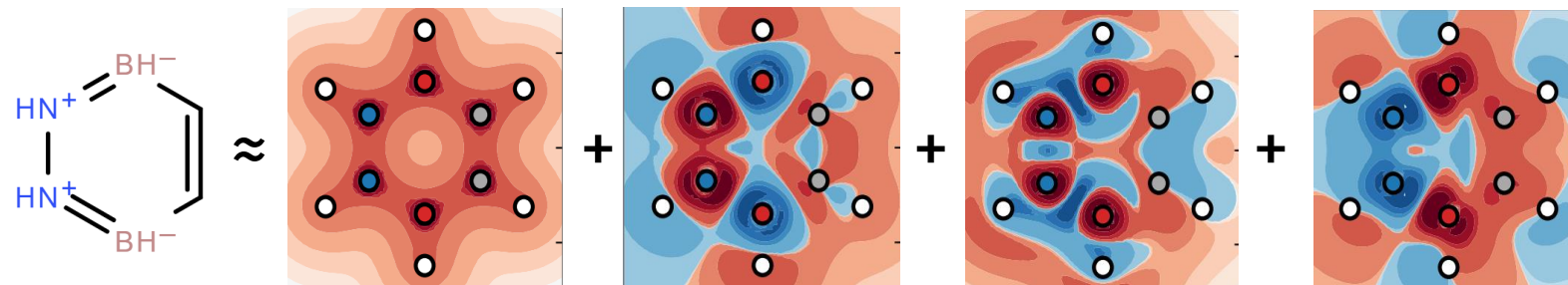
Quantum Alchemy



Taylor expansion

- Energy function of
 - Geometry
 - Nuclear charges
- Idea: obtain dominant leading derivatives, predict many systems

Forces, Vibrations
Alchemical changes



E. B. Wilson, *J. Chem. Phys.* 1962.

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

Interpolate between molecular isoelectronic Hamiltonians

$$\hat{H}(\lambda) \equiv \lambda \hat{H}_t + (1 - \lambda) \hat{H}_r \quad \lambda \in [0, 1]$$

Taylor expansion around reference molecule

$$E_t = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\partial^n}{\partial \lambda^n} \left\langle \psi_\lambda \left| \hat{H}(\lambda) \right| \psi_\lambda \right\rangle \Big|_{\lambda=0} = E_r + \sum_{n=1}^{\infty} \frac{1}{n!} \frac{\partial^n E(\lambda)}{\partial \lambda^n} \Big|_{\lambda=0}$$

Hellmann-Feynman theorem


$$\partial_\lambda E = \left\langle \psi_\lambda \left| \hat{H}_t - \hat{H}_r \right| \psi_\lambda \right\rangle = \Delta E^{\text{NN}} + \int_{\Omega} d\mathbf{r} \underbrace{(v_t(\mathbf{r}) - v_r(\mathbf{r}))}_{\equiv \Delta v} \rho_\lambda(\mathbf{r})$$

Alchemical Perturbation Density Functional Theory (APDFT)

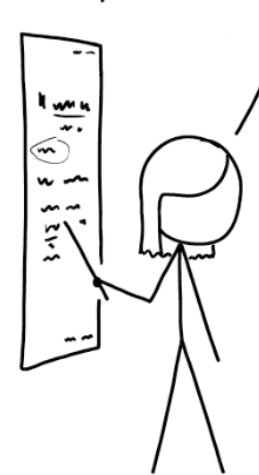
$$E_t = E_r + \Delta E^{\text{NN}} + \int_{\Omega} d\mathbf{r} \sum_{n=0}^{\infty} \frac{1}{(n+1)!} \Delta v \frac{\partial^n \rho_{\lambda}(\mathbf{r})}{\partial \lambda^n} \Big|_{\lambda=0}$$

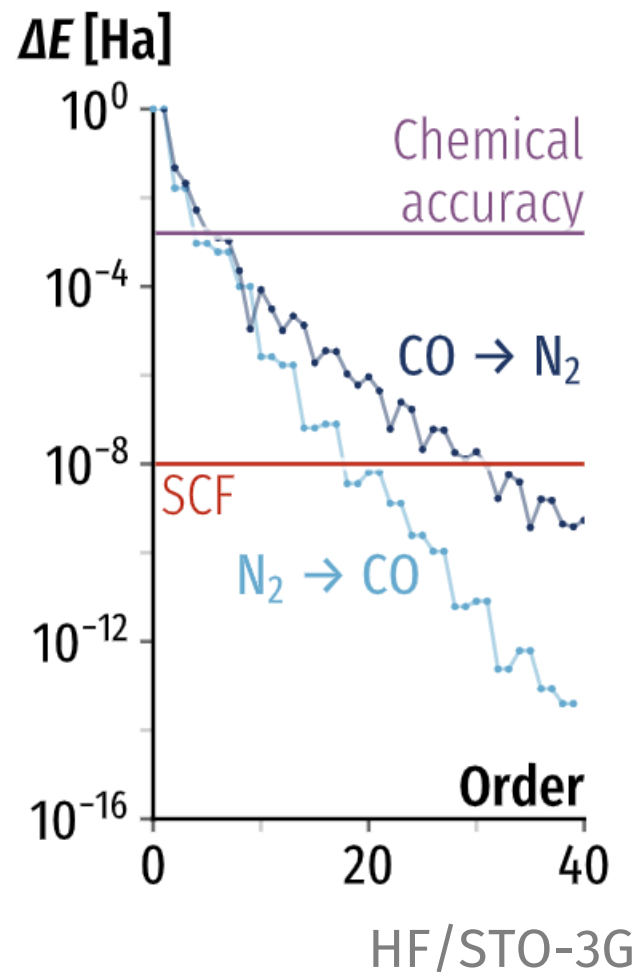
$$\rho_t = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\partial^n \rho}{\partial \lambda^n} \Big|_{\lambda=0}$$

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

 ferchault/APDFT

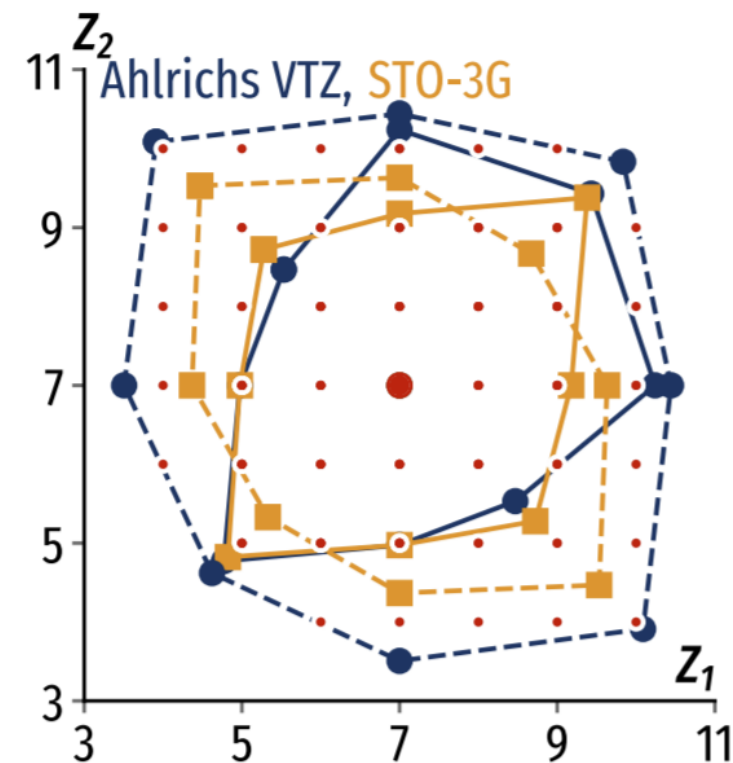
AT THIS POINT, YOU'RE PROBABLY THINKING, "I LOVE THIS EQUATION AND WISH IT WOULD NEVER END!"
WELL, GOOD NEWS!



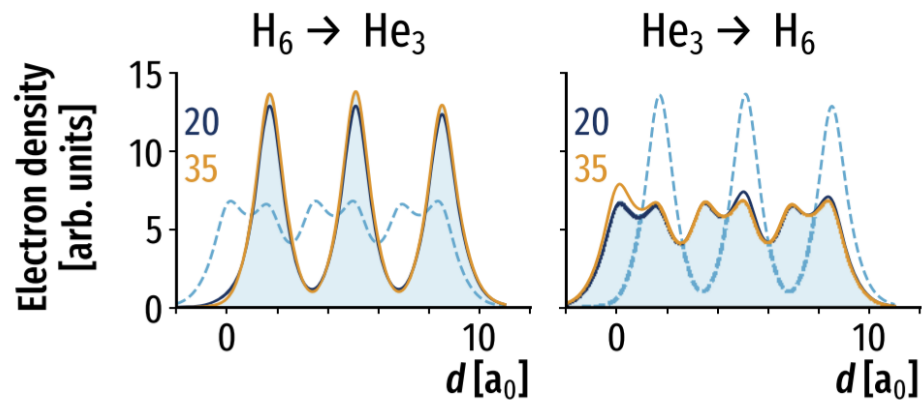
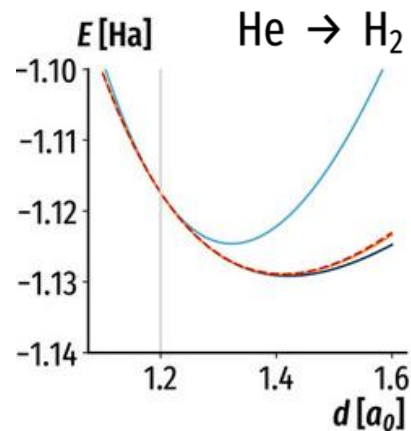
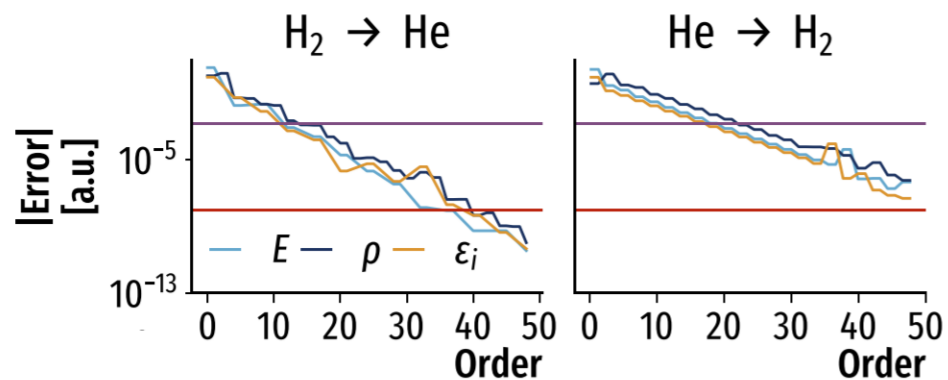


Taylor expansion

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



 ferchault/APHF

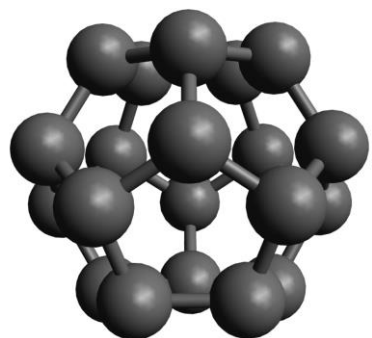


Taylor expansion

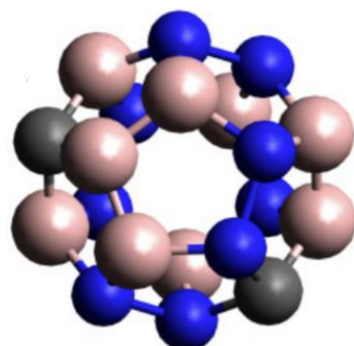
- Large changes still converge (more slowly)
- Geometric response can be recovered

Scaling with chemical space

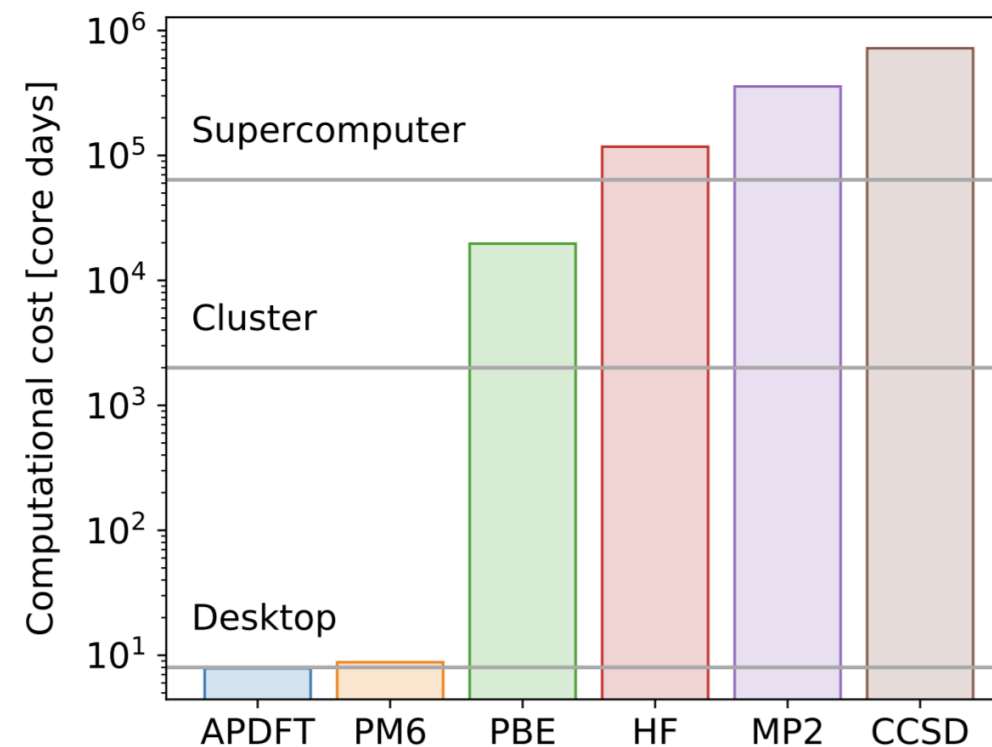
- 1 derivative for second order
- 5 derivatives for third order



C_{20}



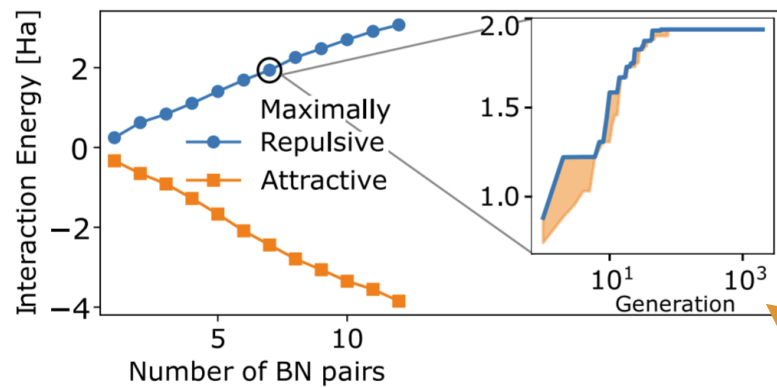
$3.1 \cdot 10^6$
targets



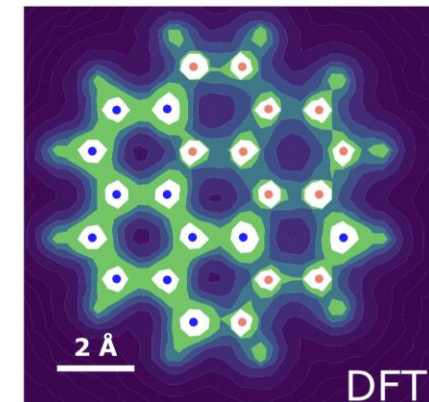
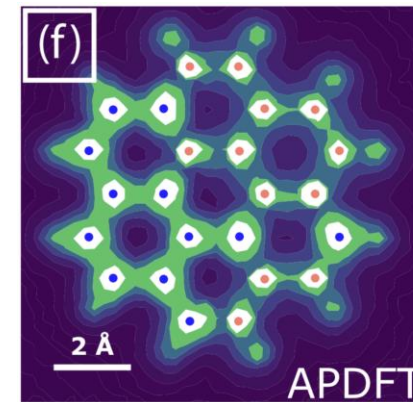
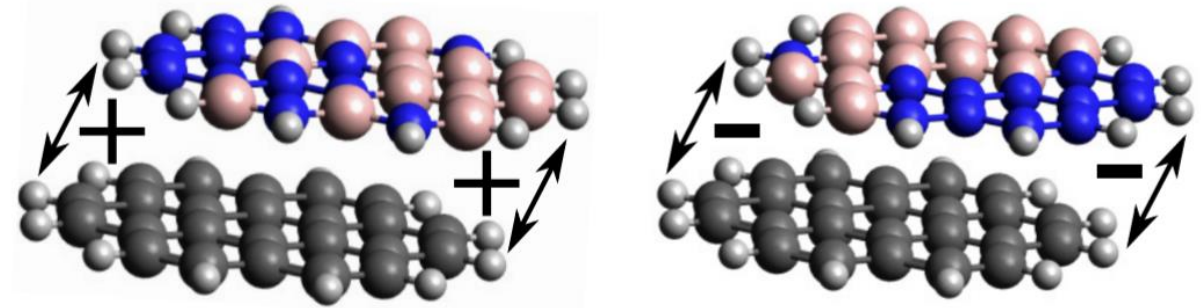
QA: 80.000x faster

BN-doped coronene dimer

- Identify most/least attractive doping pattern
- Design case



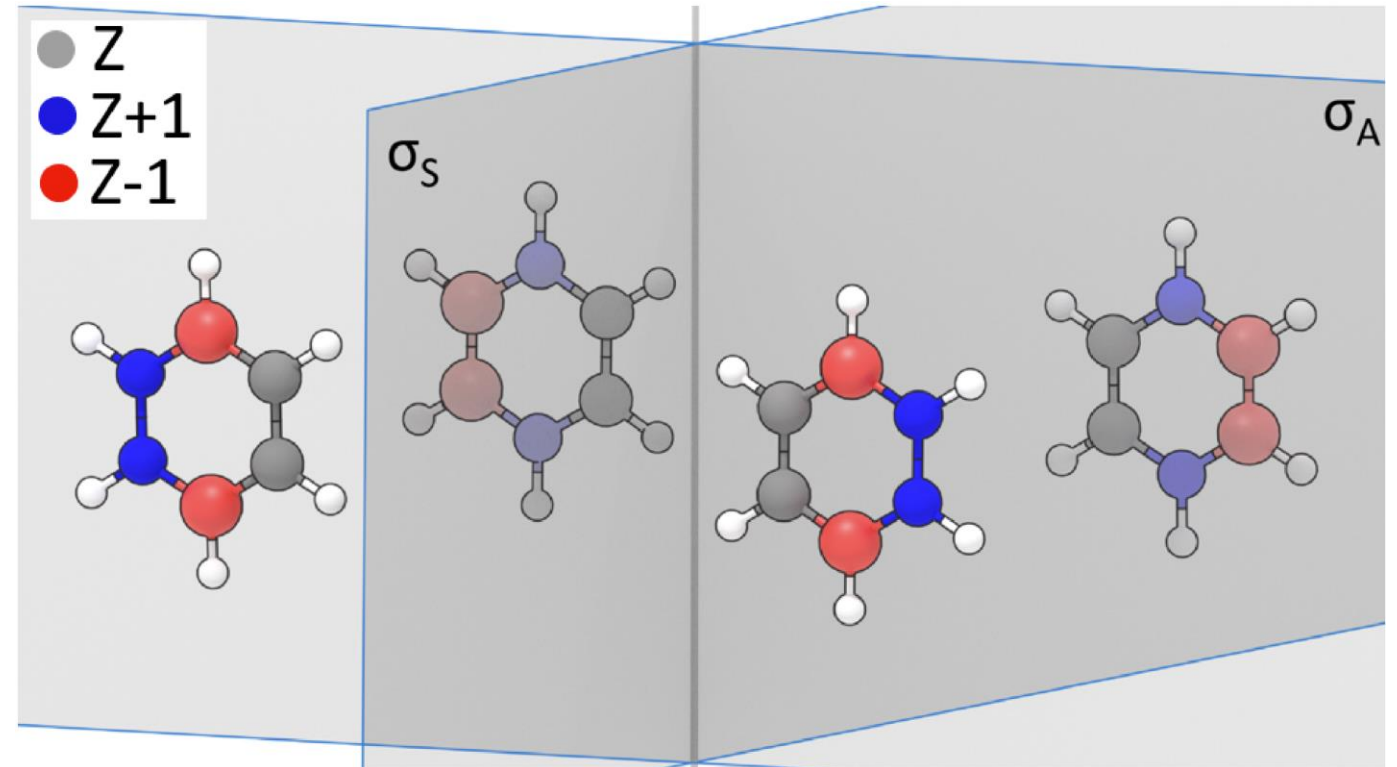
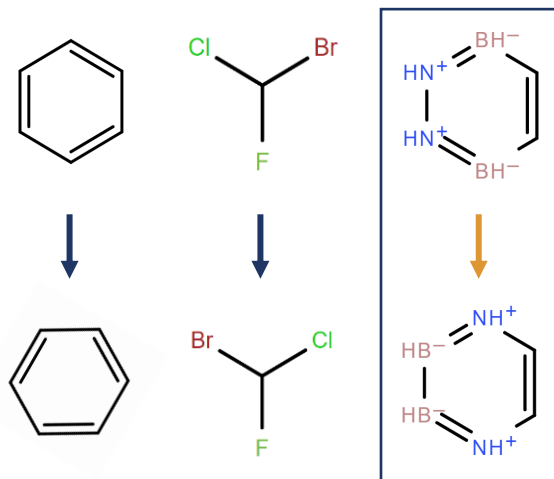
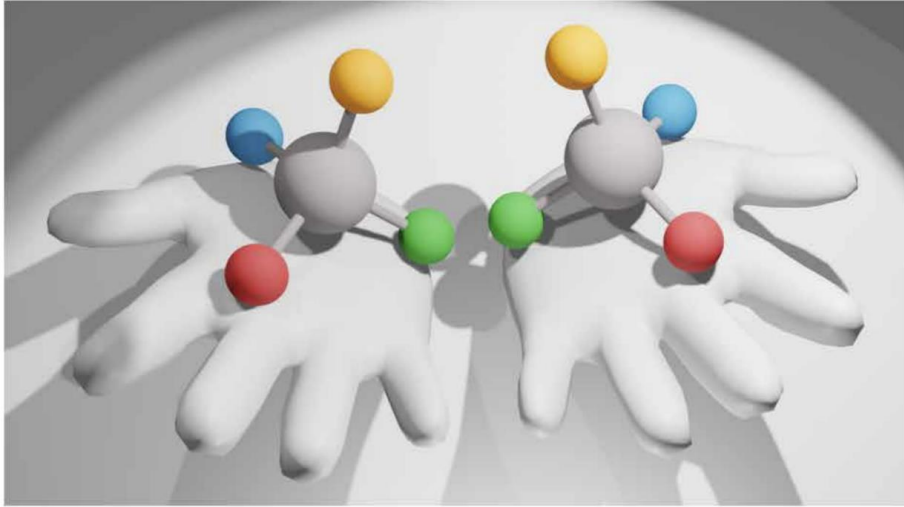
QA: 20.000x faster



$2.8 \cdot 10^{10}$ targets

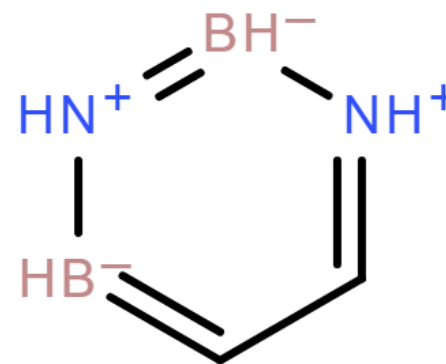
Alchemical enantiomers

Quasi-degeneracy for systems if this symmetry applies to them.

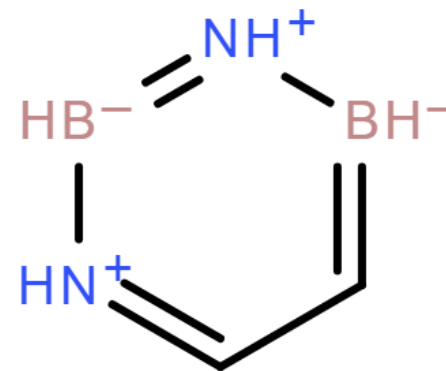


Alchemical enantiomers are

- **two spatially non-superimposable,**
- alchemically coupled,
- and iso-electronic compounds with the same formal charge,
- where each transmutating atom is assigned to exactly one subset within each of which averaging of nuclear charges results in identical chemical environments.

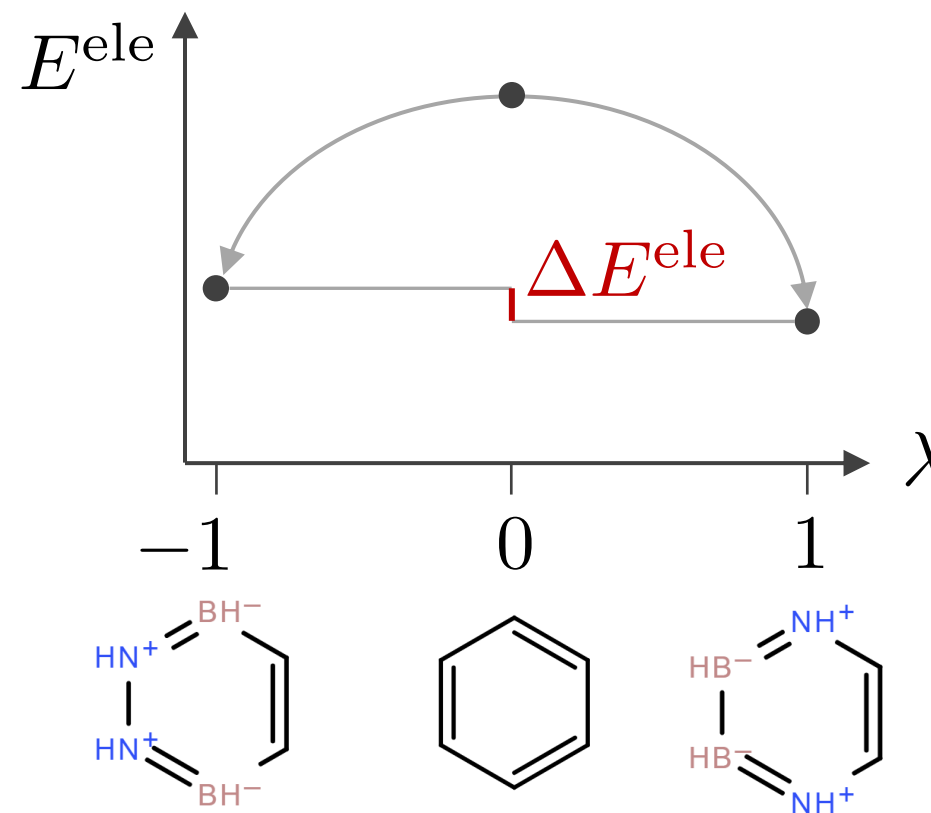


These are no alchemical enantiomers!



Alchemical enantiomers are

- two spatially non-superimposable,
- **alchemically coupled**,
- and iso-electronic compounds with the same formal charge,
- where each transmutating atom is assigned to exactly one subset within each of which averaging of nuclear charges results in identical chemical environments.

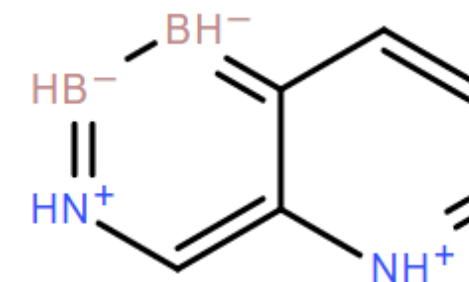
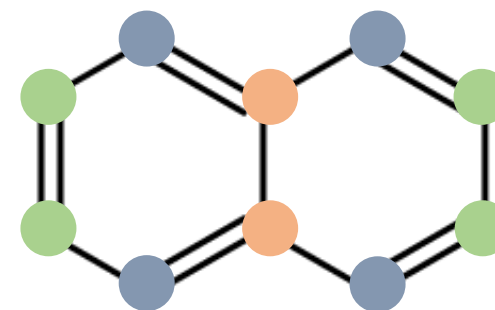
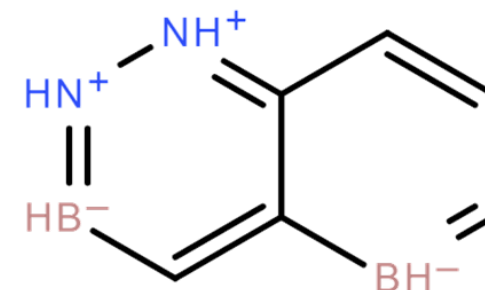


Alchemical enantiomers are

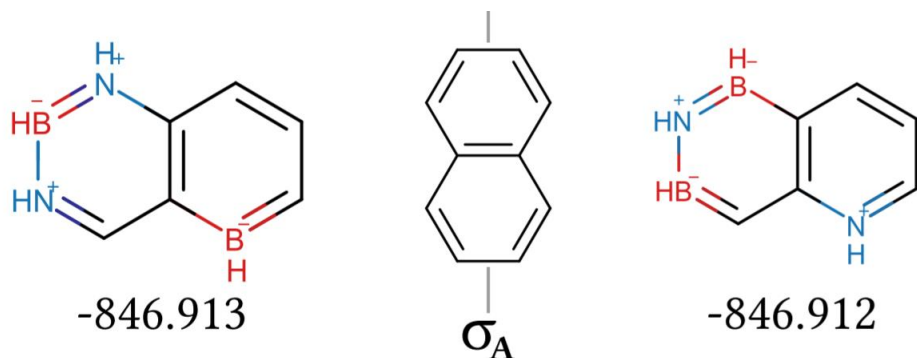
- two spatially non-superimposable,
- alchemically coupled,
- **and iso-electronic compounds with the same formal charge,**
- where each transmutating atom is assigned to exactly one subset within each of which averaging of nuclear charges results in identical chemical environments.

Alchemical enantiomers are

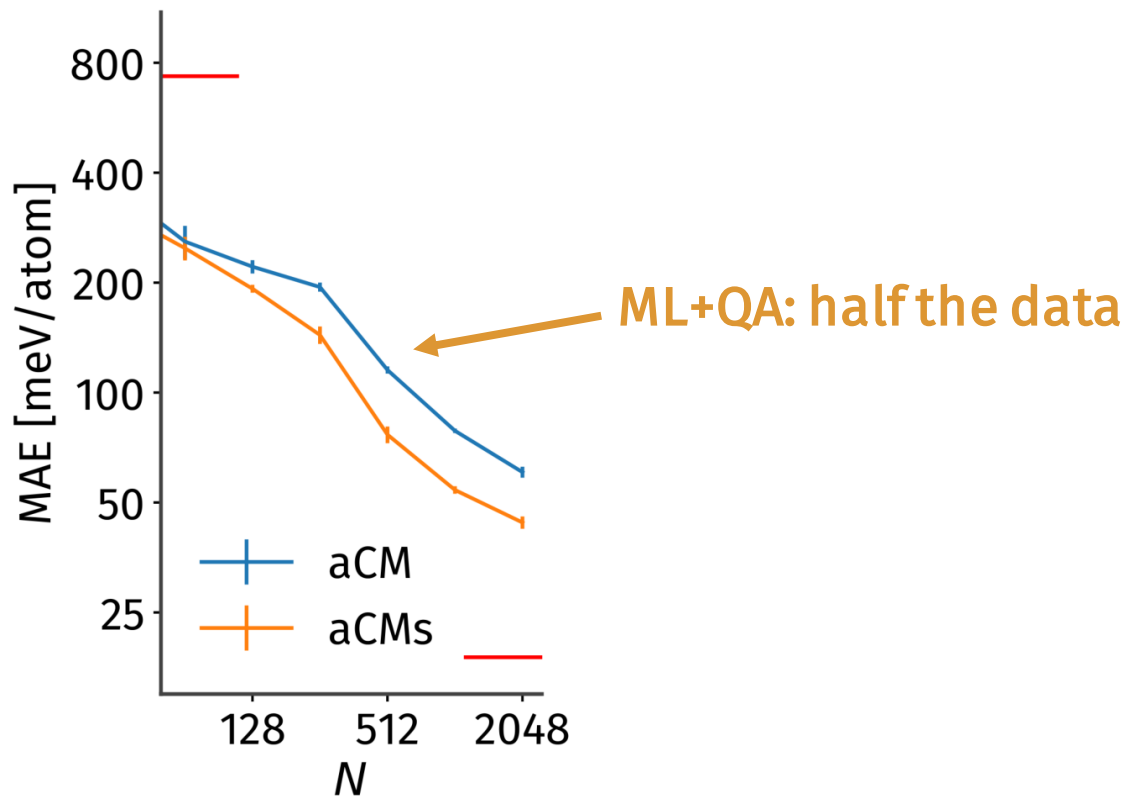
- two spatially non-superimposable,
- alchemically coupled,
- and iso-electronic compounds with the same formal charge,
- **where each transmutating atom is assigned to exactly one subset within each of which averaging of nuclear charges results in identical chemical environments.**

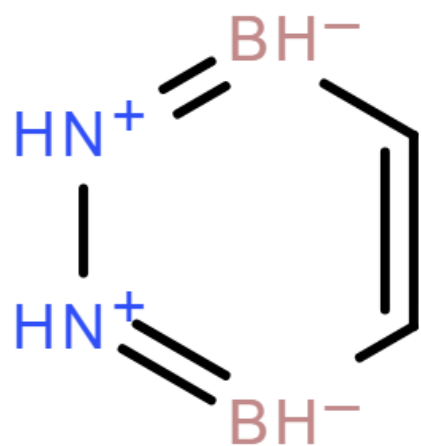


Fundamentally new symmetry
Electronic energy only

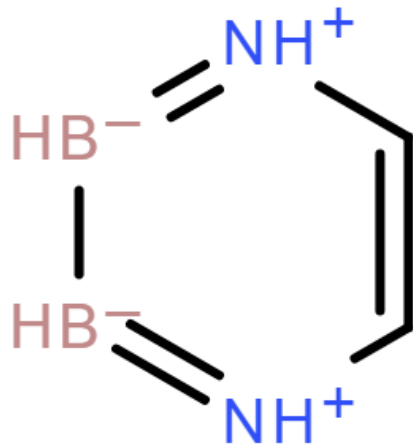


Speed up machine learning





CC
2BC
2BN
NN



CC
2NC
2BN
BB

Consecutive Elements

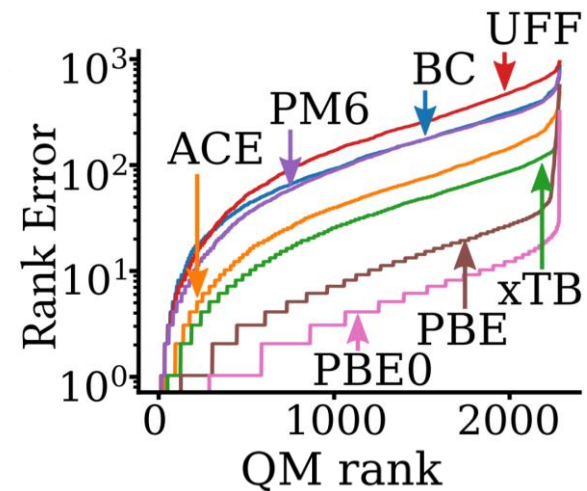
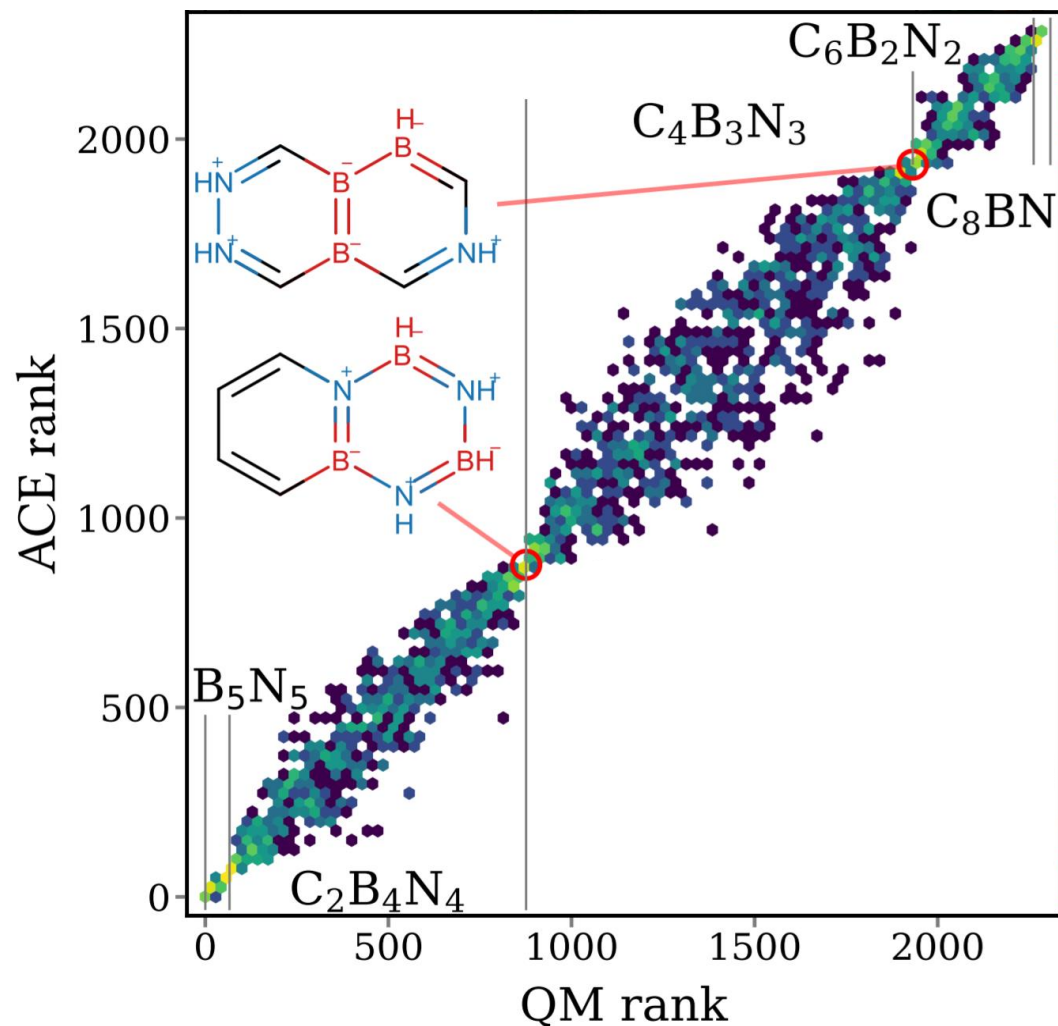
Q R S

B C N

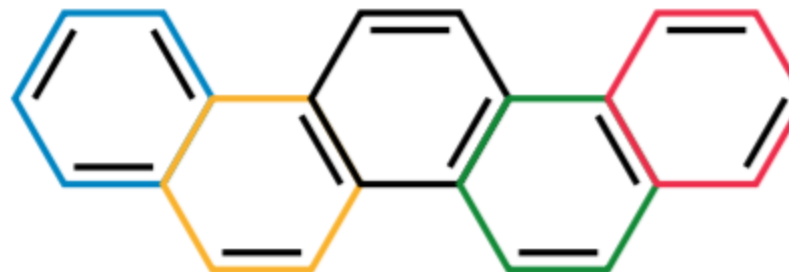
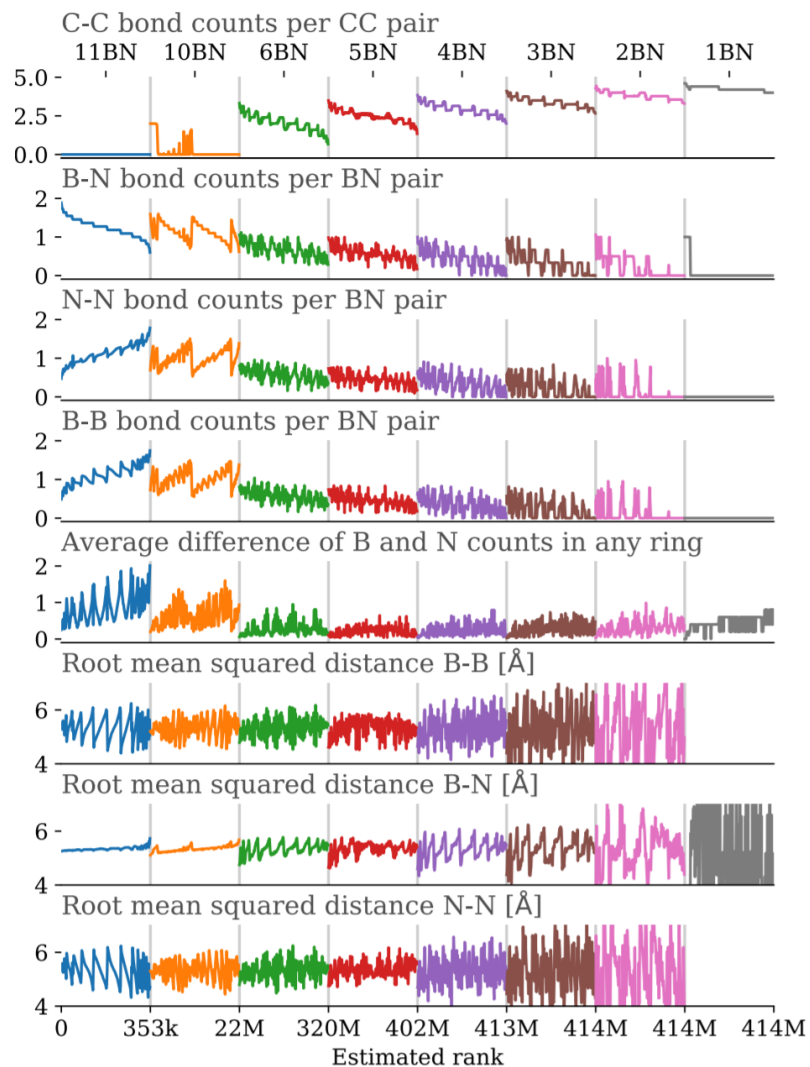
$$E_{QR} \simeq E_{SR} + 0.5(E_{QQ} - E_{SS})$$

Other skeletons and all substitution patterns

- More such rules
- No violations



- ▲ **BN-doped naphthalene**
CCSD/cc-pVDZ
Molpro/MRCC/xTB-GFN2/mopac/OpenBabel
- ◀ **BN-doped naphthalene**
CCSD/cc-pVDZ
Molpro/MRCC



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!

Derivatives without electronic perturbations

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

$\mathbf{r} \rightarrow \mathbf{r}(\lambda)$ + chain rule + partial integration e.g. for second order in 1D

$$\Delta E^{(2)} = \frac{1}{2} \int_{\Omega} dr \Delta v(r(\lambda)) \frac{\partial \rho(r(\lambda))}{\partial r} \frac{\partial r}{\partial \lambda} \Big|_{\lambda=0}$$

$$= -\frac{1}{2} \int_{\Omega} dr \rho(r(\lambda)) \frac{\partial \Delta v(r(\lambda))}{\partial r} \frac{\partial r}{\partial \lambda} \Big|_{\lambda=0}$$

Derivatives without electronic perturbations

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

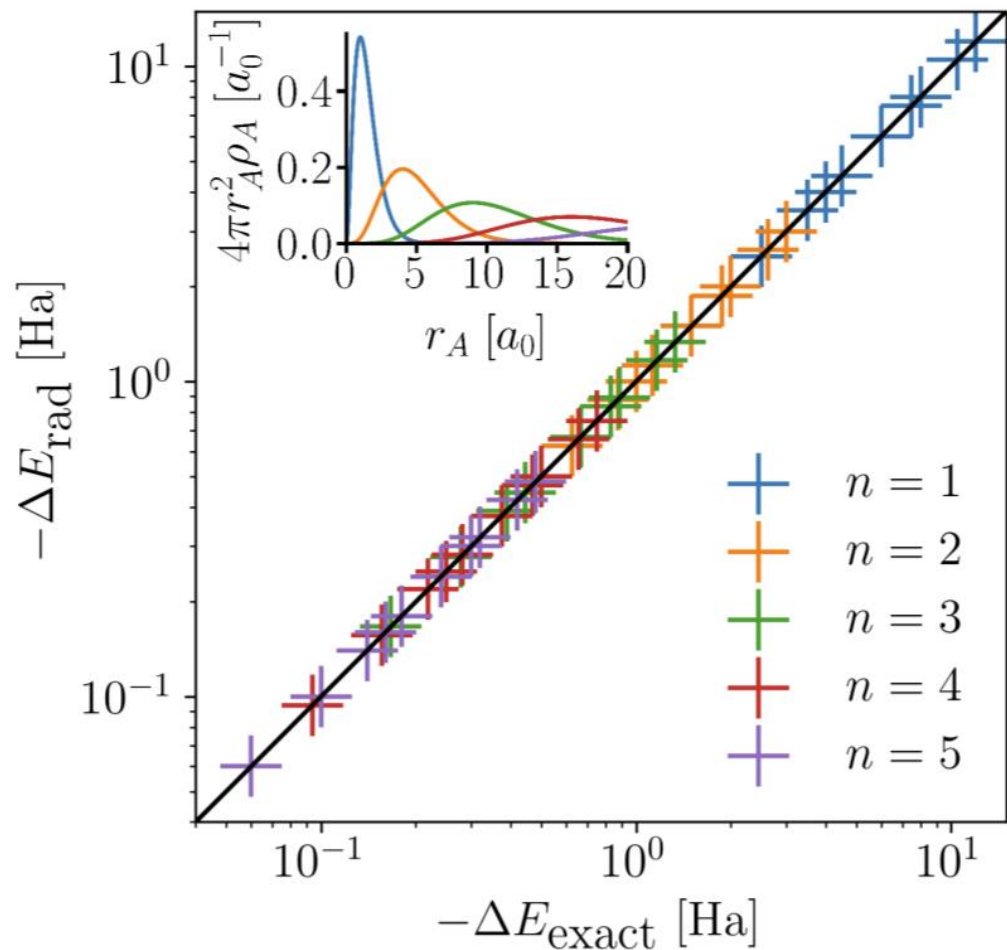
3D, arbitrary order $2n + 1 \rightarrow 0$

$$\Delta E = \int_{\Omega} d\mathbf{r} \underbrace{K(\mathbf{r}, v_i, v_f)}_{\text{Ugly, but analytical}} \underbrace{\rho(v_i)}_{\text{Only one!}}$$

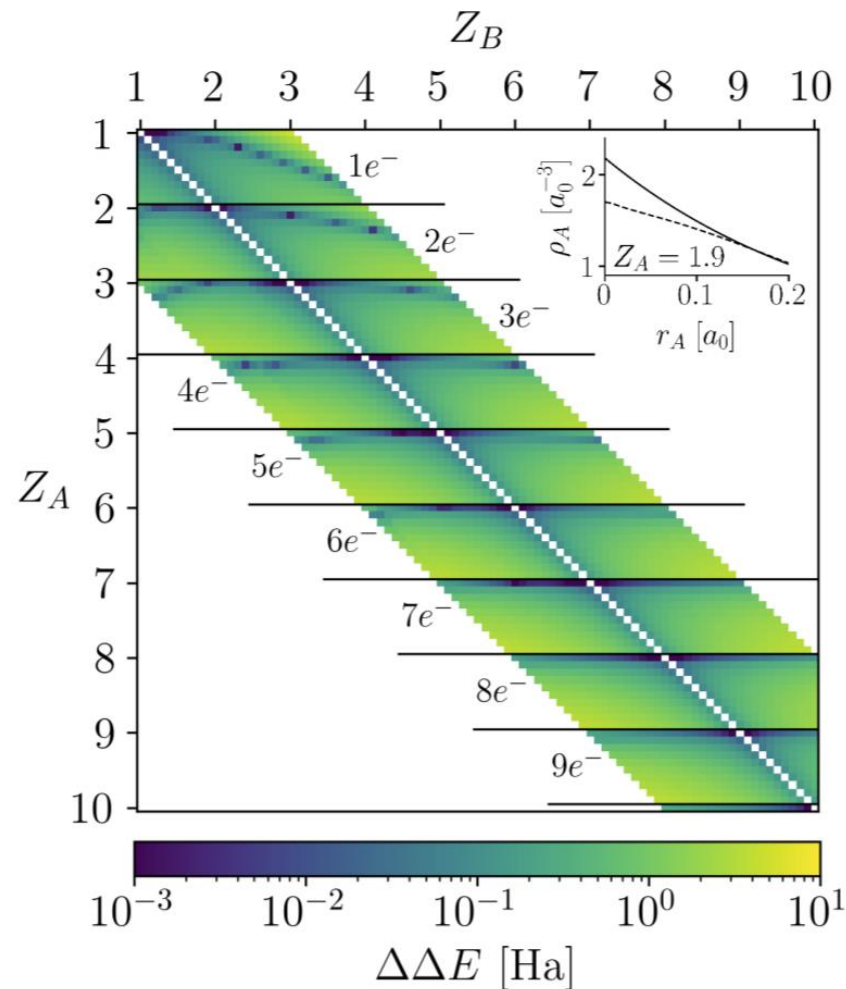
Ugly, but analytical

Only one!

Hydrogen-like atom



Multi-electron atom



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, ...

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

Convergence

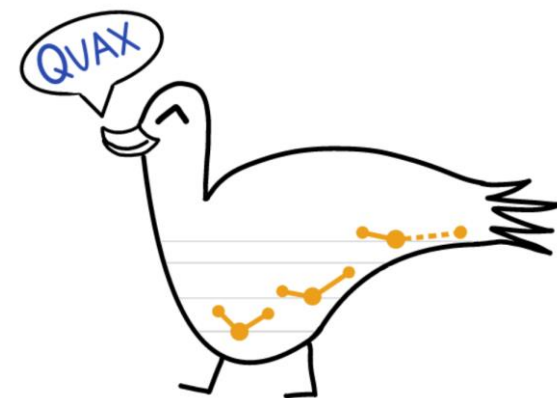
- Finite radius
- Not all systems are made equally

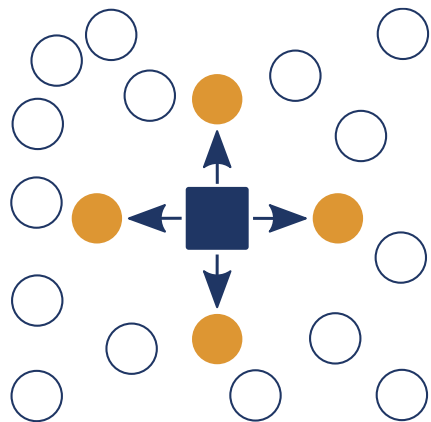
 ferchault/APDFT  ferchault/APHF

 aspuru-guzik-group/DiffiQult

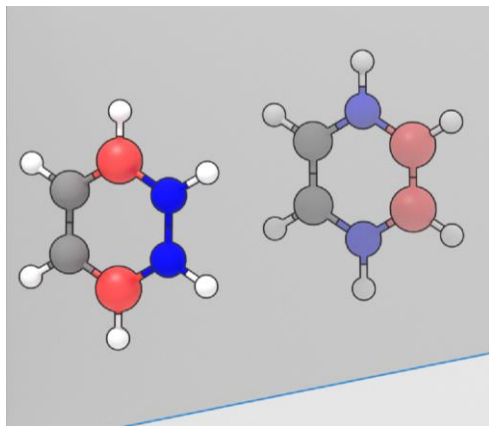
 CCQC/Quax

 diffqc/dqc **Alchemy!**

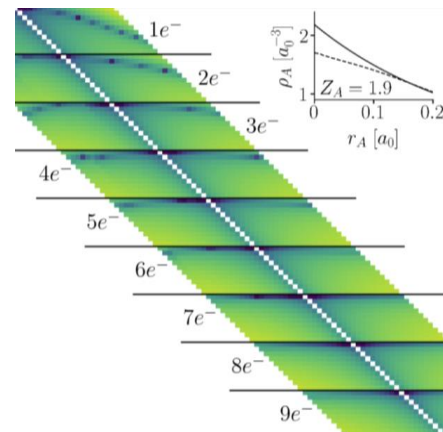




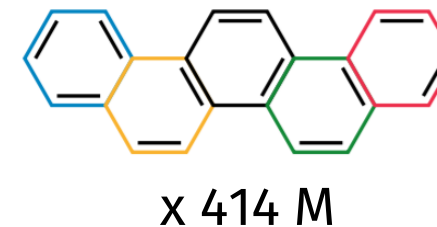
Quantum Alchemy yields systematically improvable results.



Closed expressions reveal structure of chemical space.



Efficient ways to obtain derivatives in progress.



Combinatorial scaling with size of system.

Quantum Alchemy | Phys. Rev. Res. 2020, 2, 023220.

Convergence | J. Chem. Phys. 2021, 155(22), 224103.

Alchemical Chirality | Sci. Adv. 2021, 7, eabf1173.

Integral Transform | J. Chem. Phys. 2022, 157(16), 164109.



Simon
Krug



Anatole
von Lilienfeld

$$\Delta E_{ij}^{\text{ele}} = E_0 - E_0 + \int_{\Omega} d\mathbf{r} \sum_{n=0}^{\infty} \frac{\Delta v_i}{(n+1)!} \left[\frac{\partial^n \rho}{\partial \lambda_i^n} + \frac{\partial^n \rho}{\partial \lambda_j^n} \right]$$

$$\Delta E_{(0)}^{\text{ele}} = E_0 - E_0 = 0$$

$$\Delta E_{(1)}^{\text{ele}} = 2 \int_{\Omega} \Delta v \rho = \int_{\Omega} e \cdot o = 0$$

$$\begin{aligned} \Delta E_{(2)}^{\text{ele}} &= \int_{\Omega} \Delta v \left[\frac{\partial \rho}{\partial \lambda_i} + \frac{\partial \rho}{\partial \lambda_j} \right] \\ &= \int_{\Omega} \Delta v \left[\sum_I \frac{\partial \rho}{\partial Z_I} \frac{\partial Z_I}{\partial \lambda_i} + \frac{\partial \rho}{\partial Z_I} \frac{\partial Z_I}{\partial \lambda_j} \right] = 0 \end{aligned}$$