Systematically Improvable Models From Alchemical Perturbations

Guido Falk von Rudorff, University of Vienna
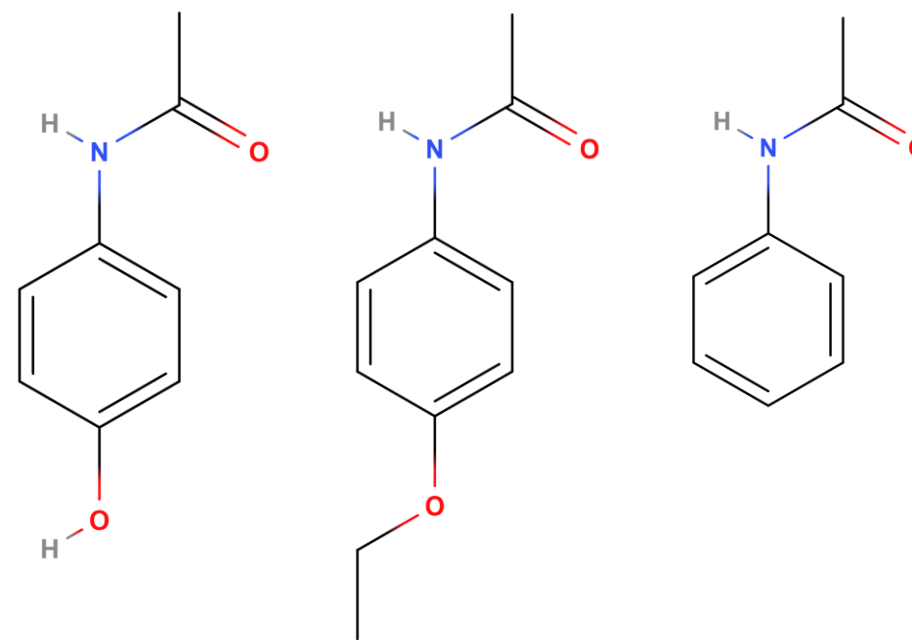
ferchault   @ferchault   guido.vonrudorff.de

Design: sample by guided trial-and-error.
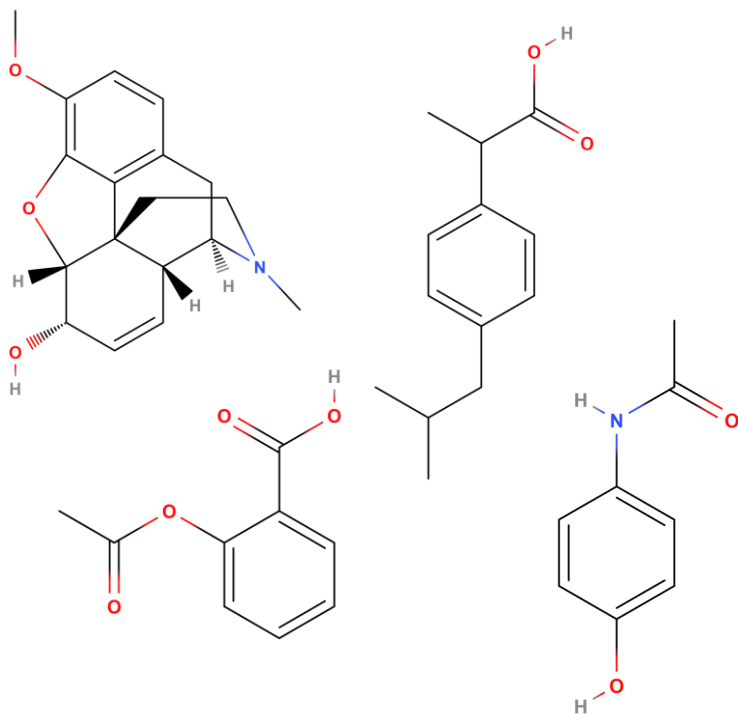
Global Search Problem
Which class of compounds?

Drug-like: 10^{60}

Local Search Problem
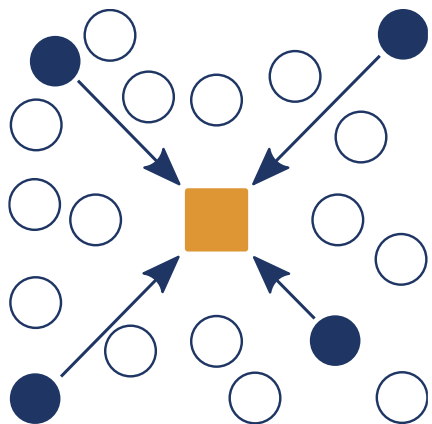Which particular species within that class?

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

G. Ceder, *Science* **1998**. A. Franceschetti, A. Zunger, *Nature*, **1999**. A. Mullard, *Nature*, **2017**.
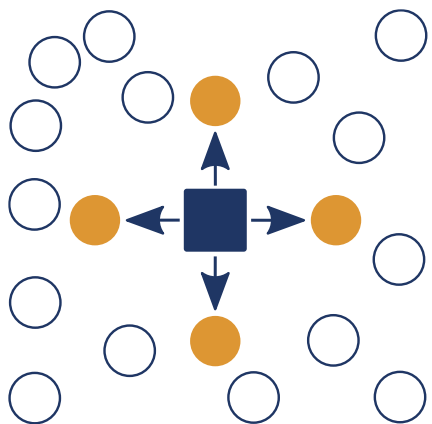
Machine Learning

Foundations **|** Statistical modelling

Accuracy **|** Systematically improvable through data and training

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

Limitation **|** Requires training data, no black box
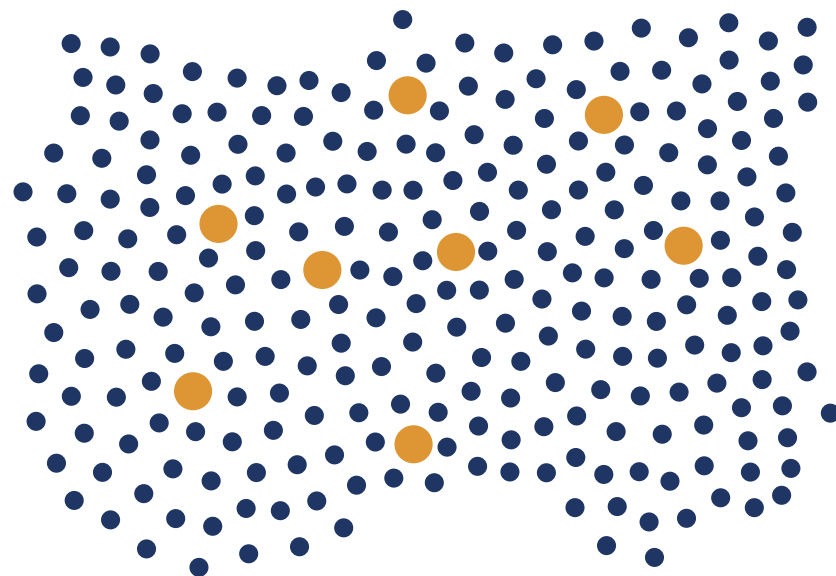
Quantum Alchemy

Foundations **|** Perturbation theory

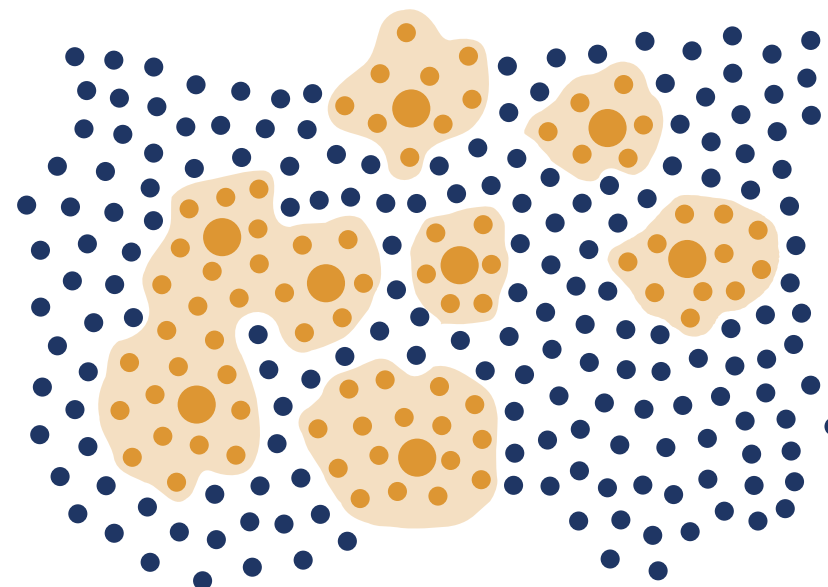Accuracy **|** Systematically improvable through higher orders terms

Specialty **|** Combinatorial scaling with chemical diversity

Limitation **|** Finite range in chemical space

Without Perturbation
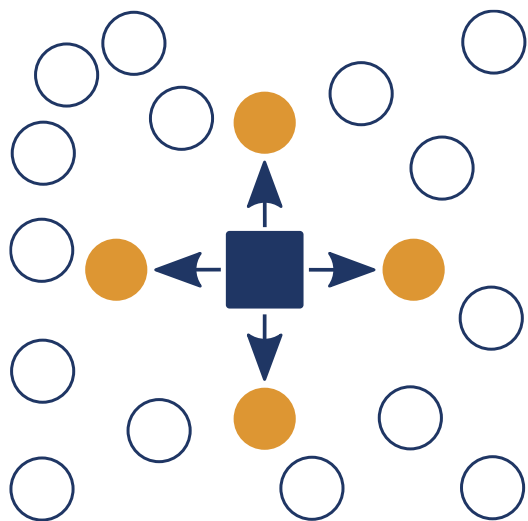
With Perturbation

Systems/Molecules
- Any
- Known
- Approximated

Paradigm shift
Few highly accurate calculations
instead of many intermediate ones

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

1D

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

Quantum Alchemy

Taylor expansion
- Energy function of
  - Geometry   Forces, Vibrations
  - Nuclear charges   Alchemical changes
- Idea: obtain dominant leading derivatives, predict many systems
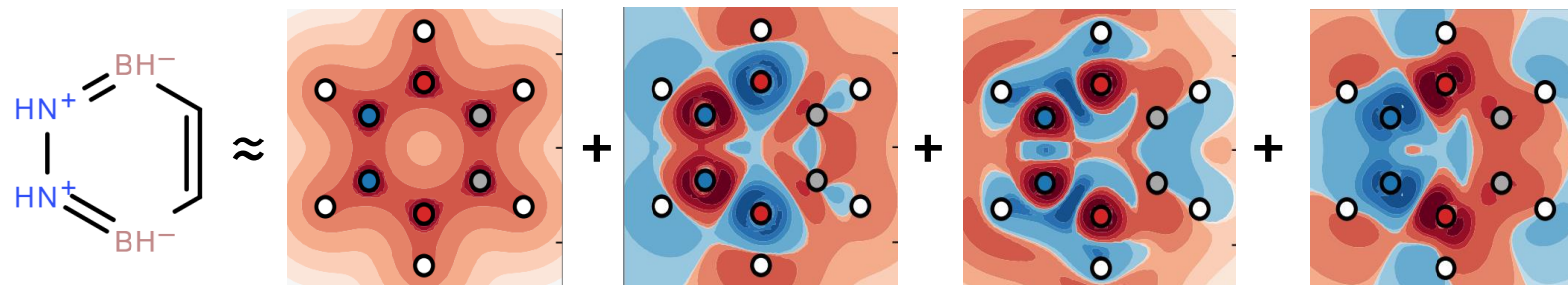
E. B. Wilson, *J. Chem. Phys.* **1962.**
GFvR, O. A. von Lilienfeld, *Phys. Rev. Res.*, **2020.**

Interpolate between molecular isoelectronic Hamiltonians

$$\hat{H}(\lambda) \equiv \lambda \hat{H}_{\mathrm{t}} + (1 - \lambda)\hat{H}_{\mathrm{r}} \qquad \lambda \in [0, 1]$$

Taylor expansion around reference molecule

$$E_{\mathrm{t}} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\partial^n}{\partial \lambda^n} \left\langle \psi_\lambda \left| \hat{H}(\lambda) \right| \psi_\lambda \right\rangle \Bigg|_{\lambda=0} = E_{\mathrm{r}} + \sum_{n=1}^{\infty} \frac{1}{n!} \frac{\partial^n E(\lambda)}{\partial \lambda^n} \Bigg|_{\lambda=0}$$

Hellmann-Feynman theorem

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

Alchemical Perturbation Density Functional Theory (APDFT)

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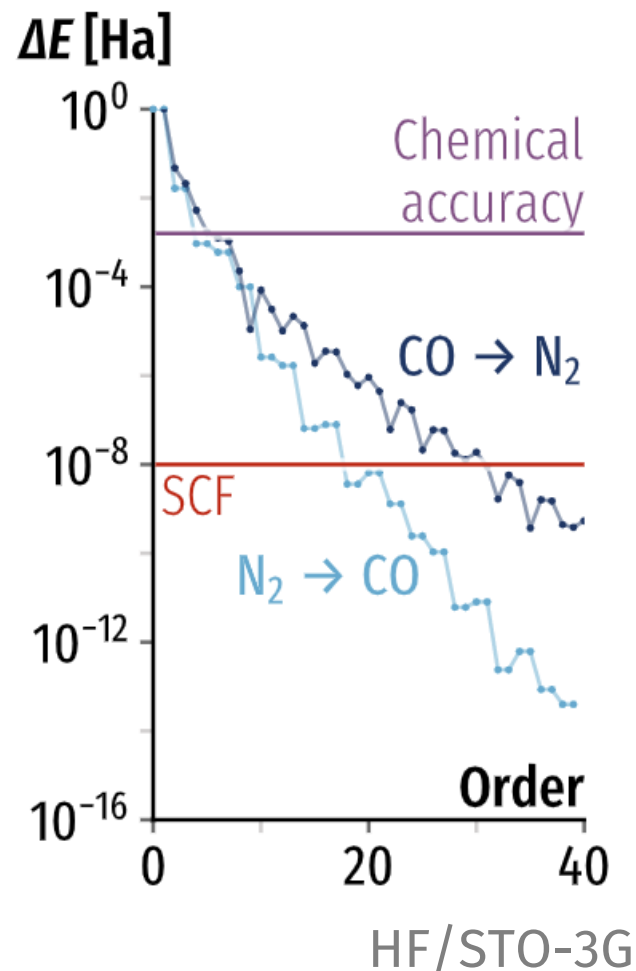
$$\rho_t = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\partial^n \rho}{\partial \lambda^n} \bigg|_{\lambda=0}$$

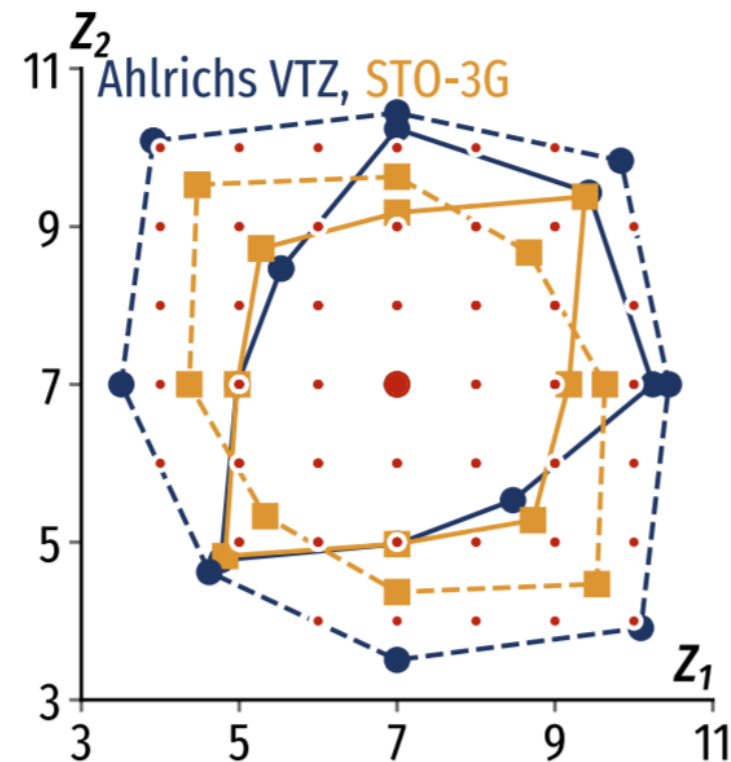- Gives consistent energies, densities, forces, …
- Uses the same derivatives for all predictions

ferchault/APDFT

GFvR, O. A. von Lilienfeld, *Phys. Rev. Res.*, **2020.**
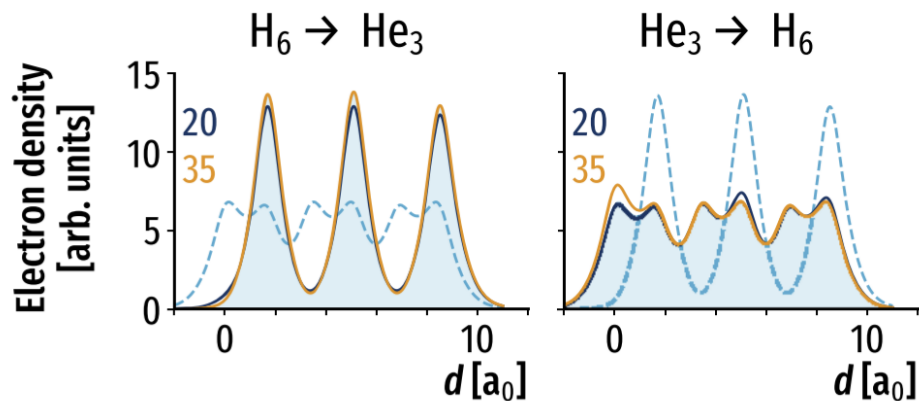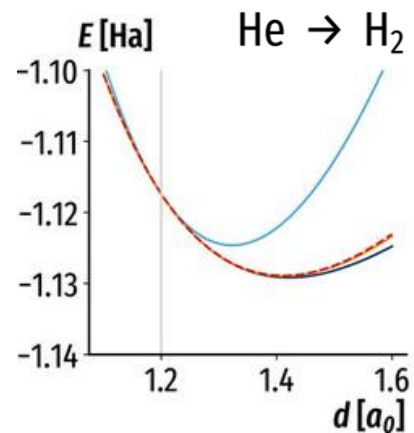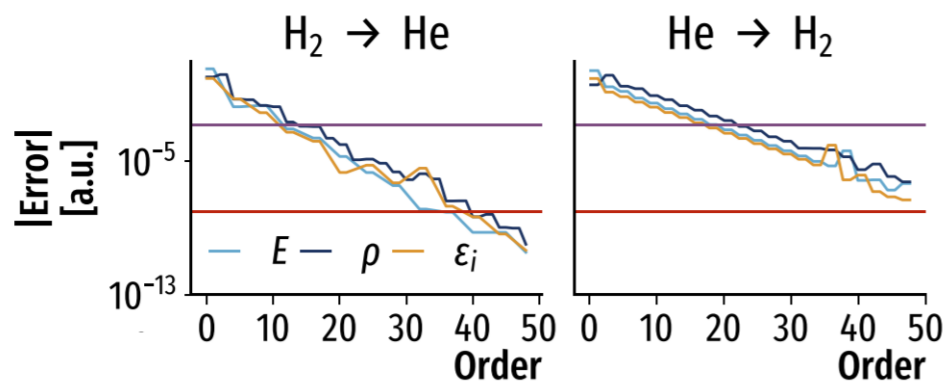
xkcd.com/2605

HF/STO-3G

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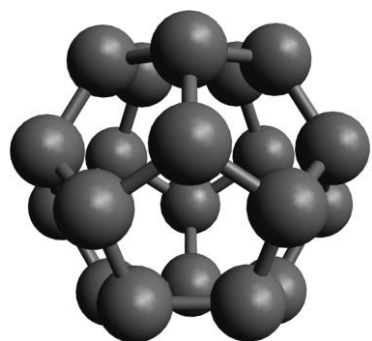

Ahlrichs VTZ, STO-3G
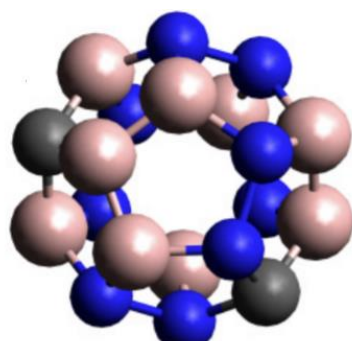
ferchault/APHF

Taylor expansion

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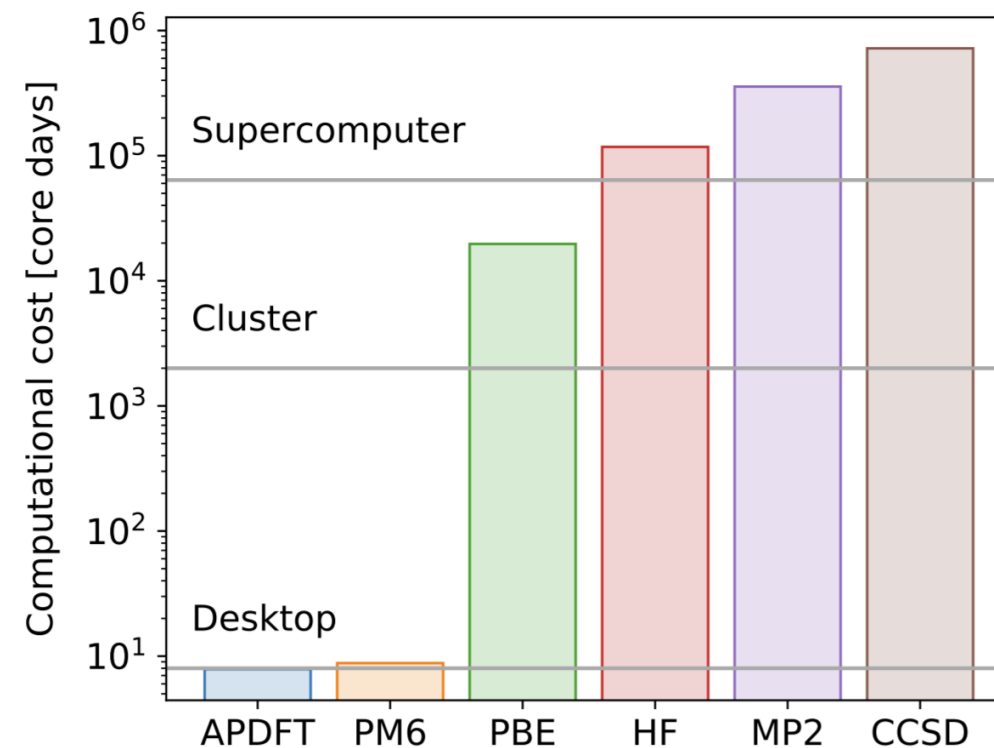- Geometric response can be recovered

Scaling with chemical space
- 1 derivative for second order
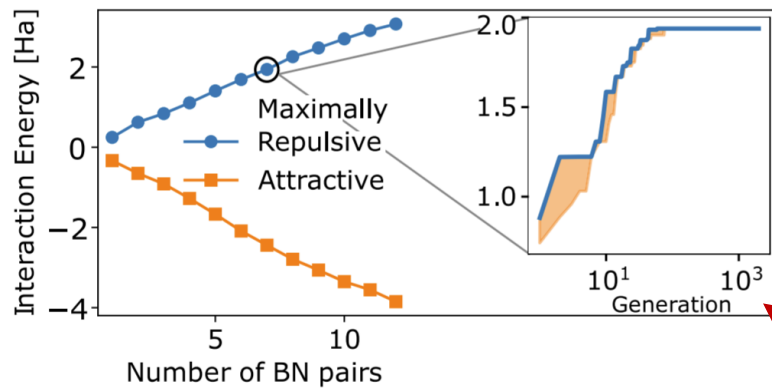- 5 derivatives for third order
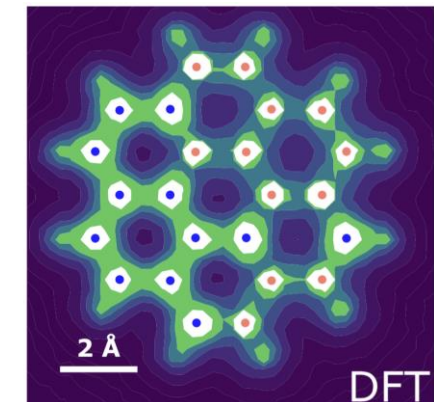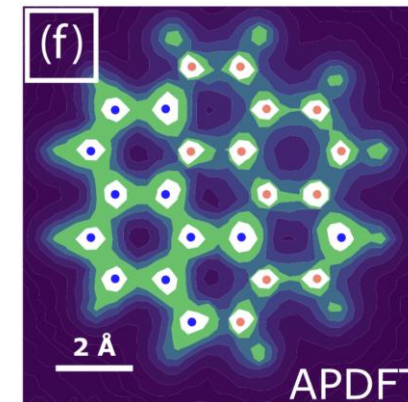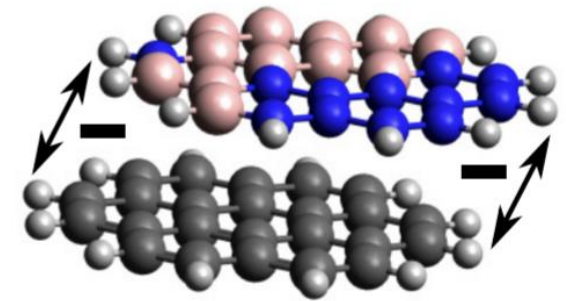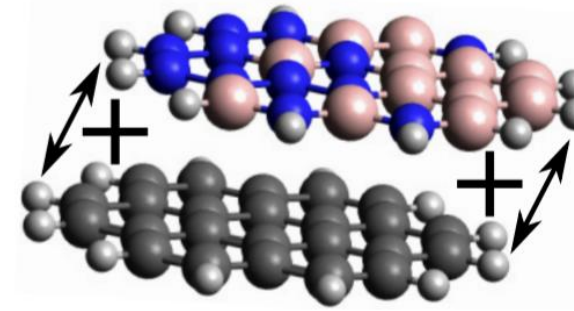

C_{20}

$3.1 \cdot 10^6$ targets

QA: 80.000x faster

BN-doped coronene dimer
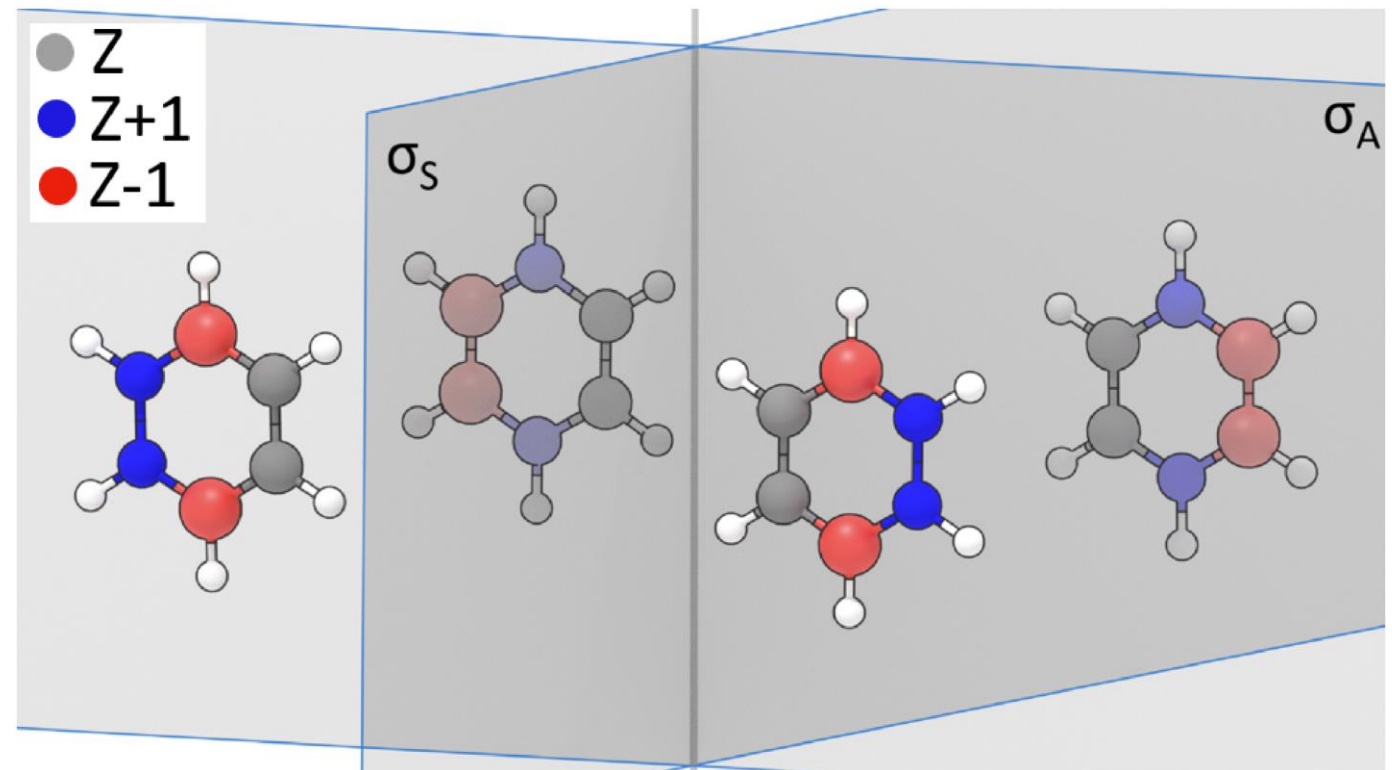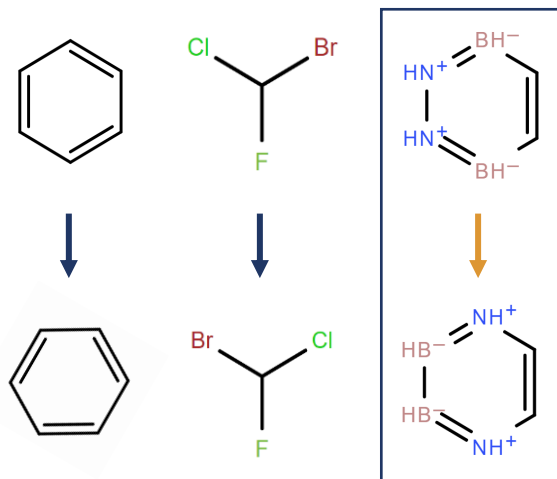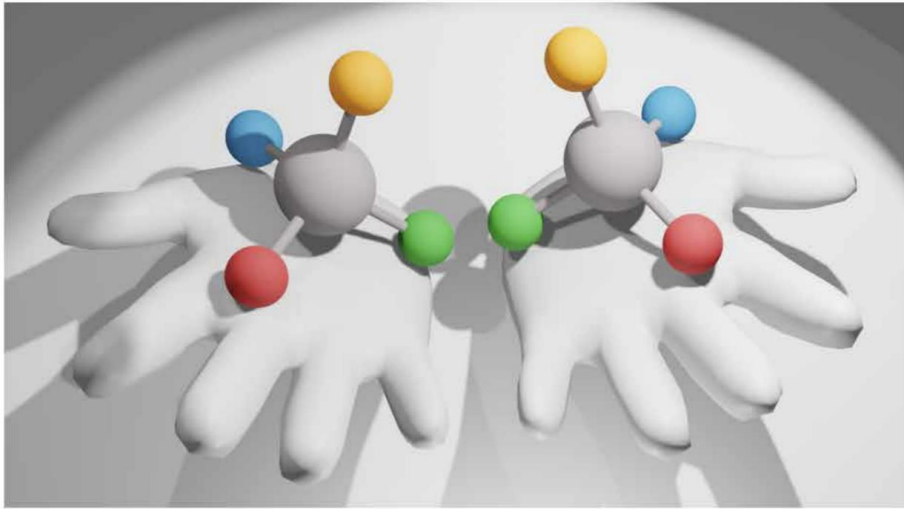- Identify most/least attractive doping pattern
- Design case

QA: 20.000x faster

$2.8 \cdot 10^{10}$ targets

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

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.
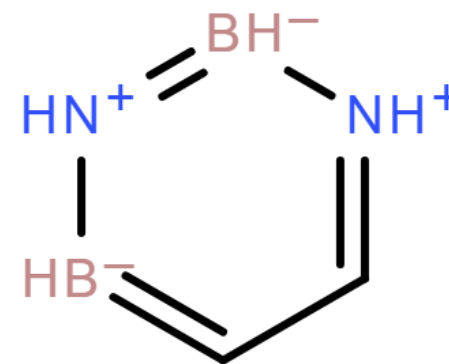
These are no alchemical enantiomers!

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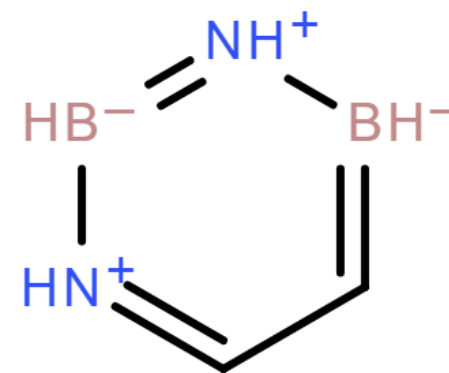
Alchemical enantiomers are
-   two spatially non-superimposable,
-   alchemically coupled,
-   **and iso-electronic compounds with the same formal charge,**
-   where each transmutating atom is assigned to exactly one subset within each of which averaging of nuclear charges results in identical chemical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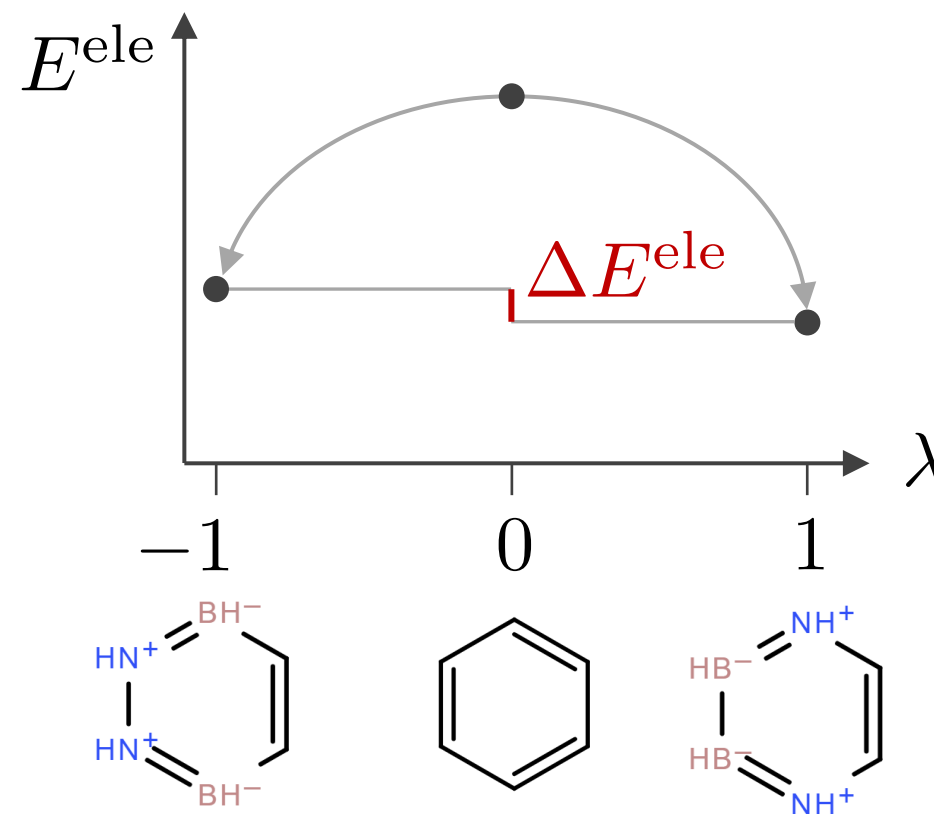
Alchemical enantiomers are
- two spatially non-superimposable,
- alchemically coupled,
- and iso-electronic compounds with the same formal charge,
- where each transmutating atom is assigned to exactly one subset within each of which averaging of nuclear charges results in identical chemical environments.
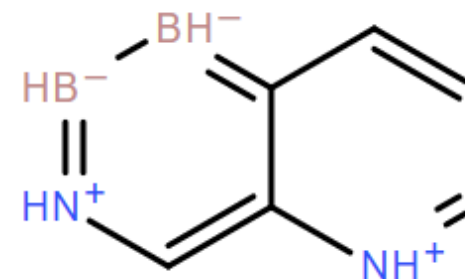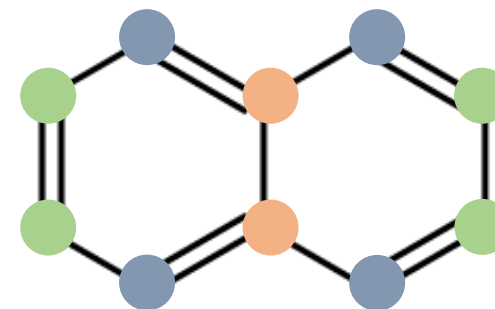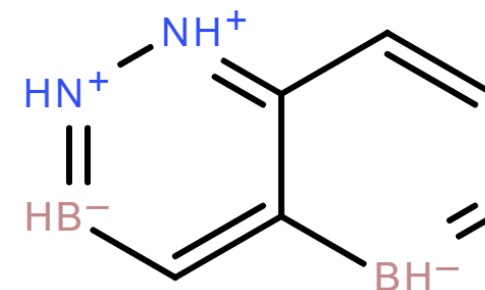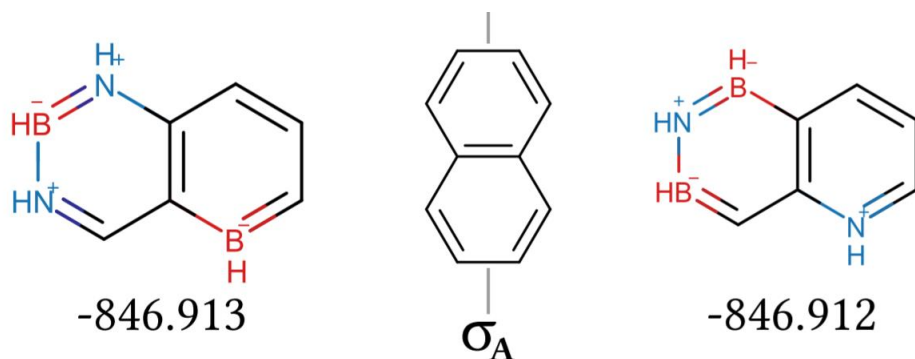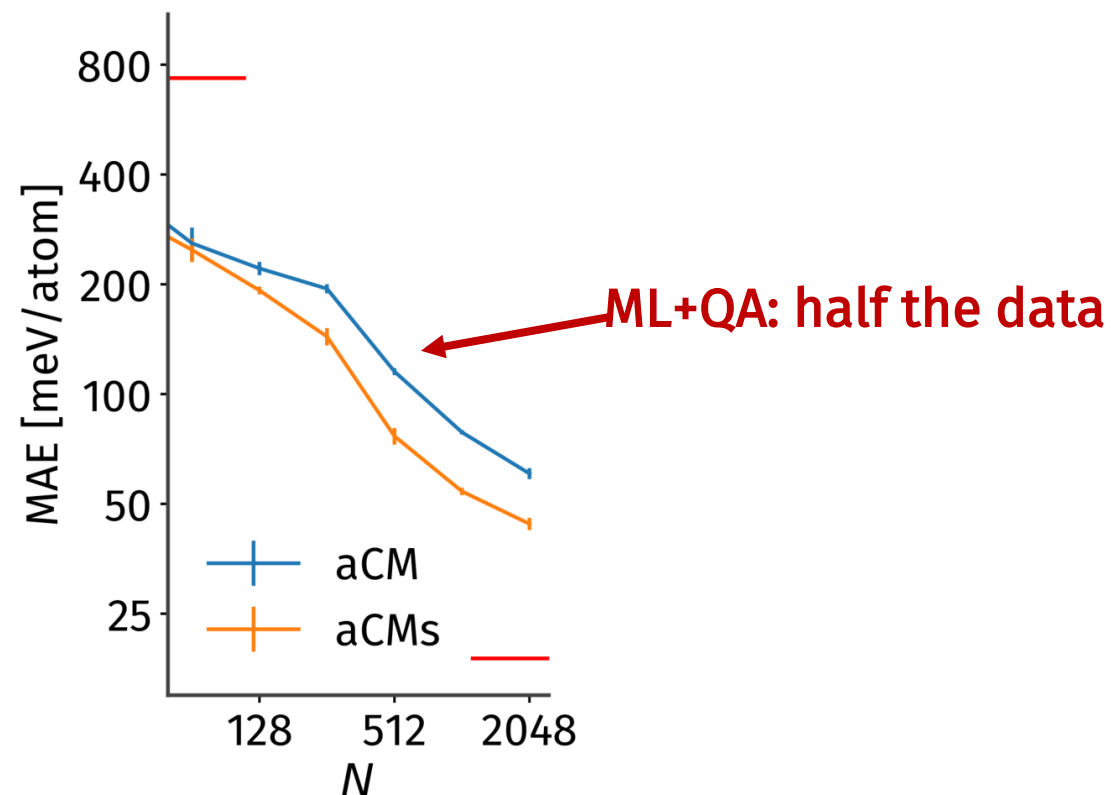
Fundamentally new symmetry
Electronic energy only

-846.913 σ_A -846.912

Speed up machine learning

ML+QA: half the data

MAE [meV/atom]

N

aCM
aCMs
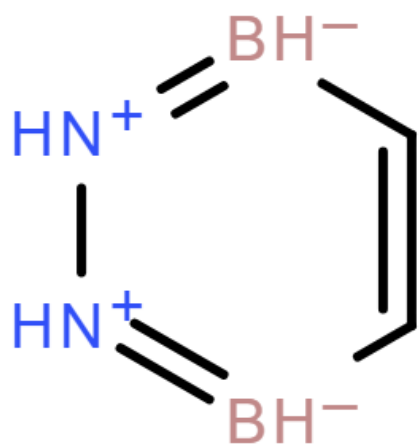
CC
2BC
2BN
NN
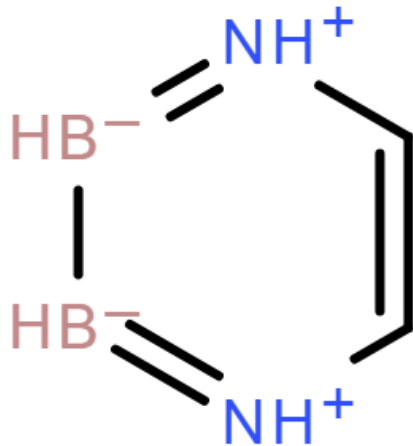
CC
2NC
2BN
BB

Consecutive Elements
Q R S
B C N

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

Other skeletons and all substitution patterns
- More such rules
- No violations
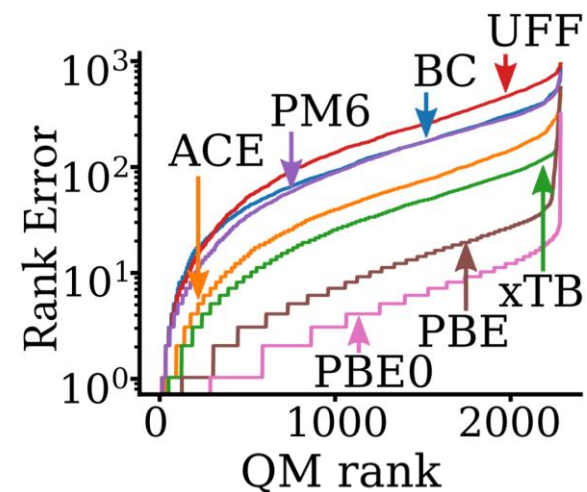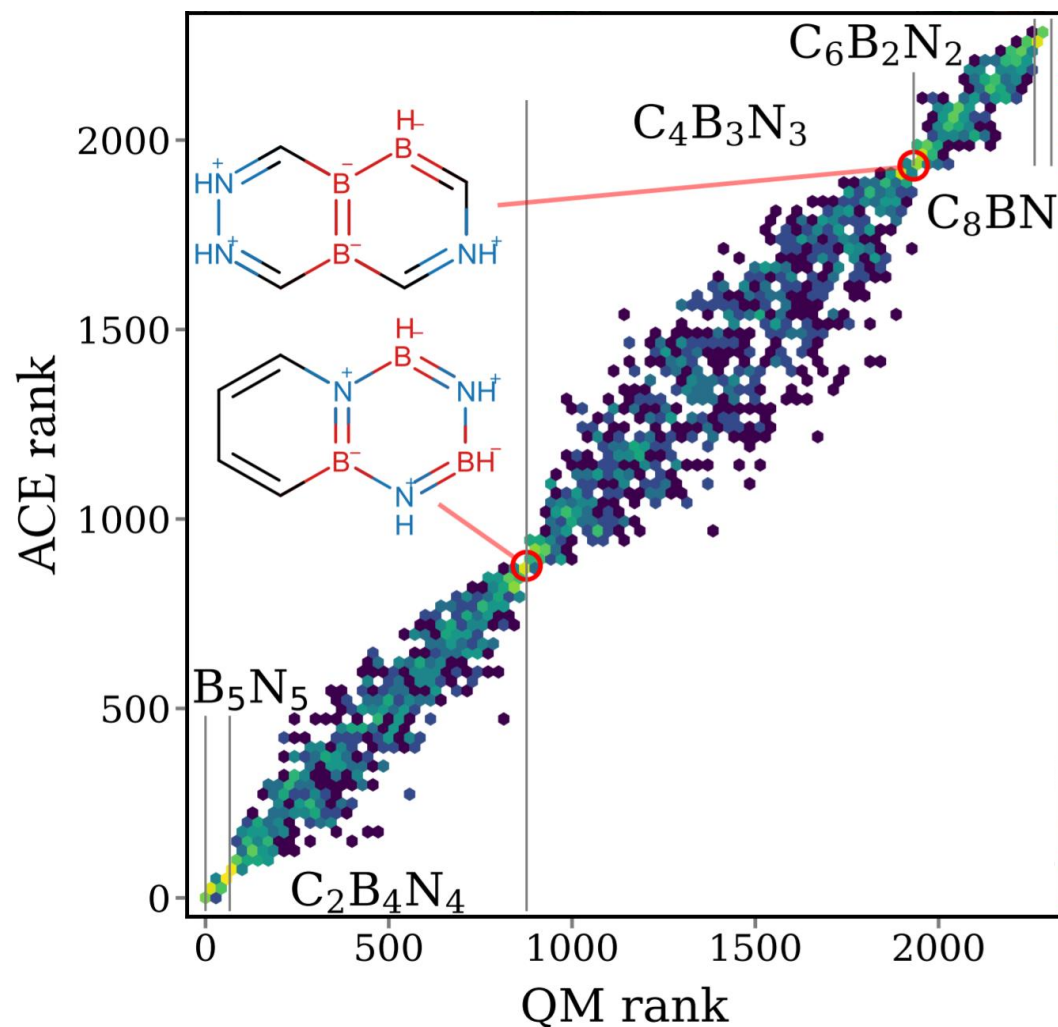
▲ BN-doped naphthalene
CCSD/cc-pVDZ
Molpro/MRCC/xTB-GFN2/mopac/OpenBabel

◀ BN-doped naphthalene
CCSD/cc-pVDZ
Molpro/MRCC

x 414 M

QA: Millions at once!
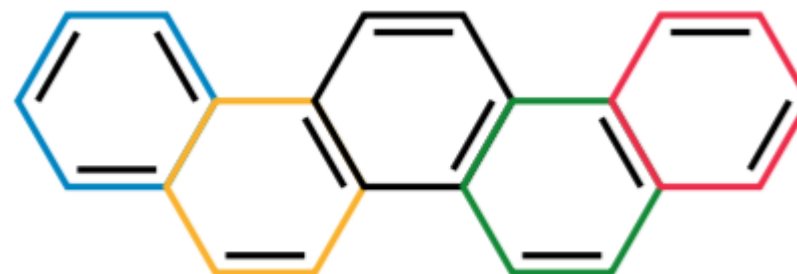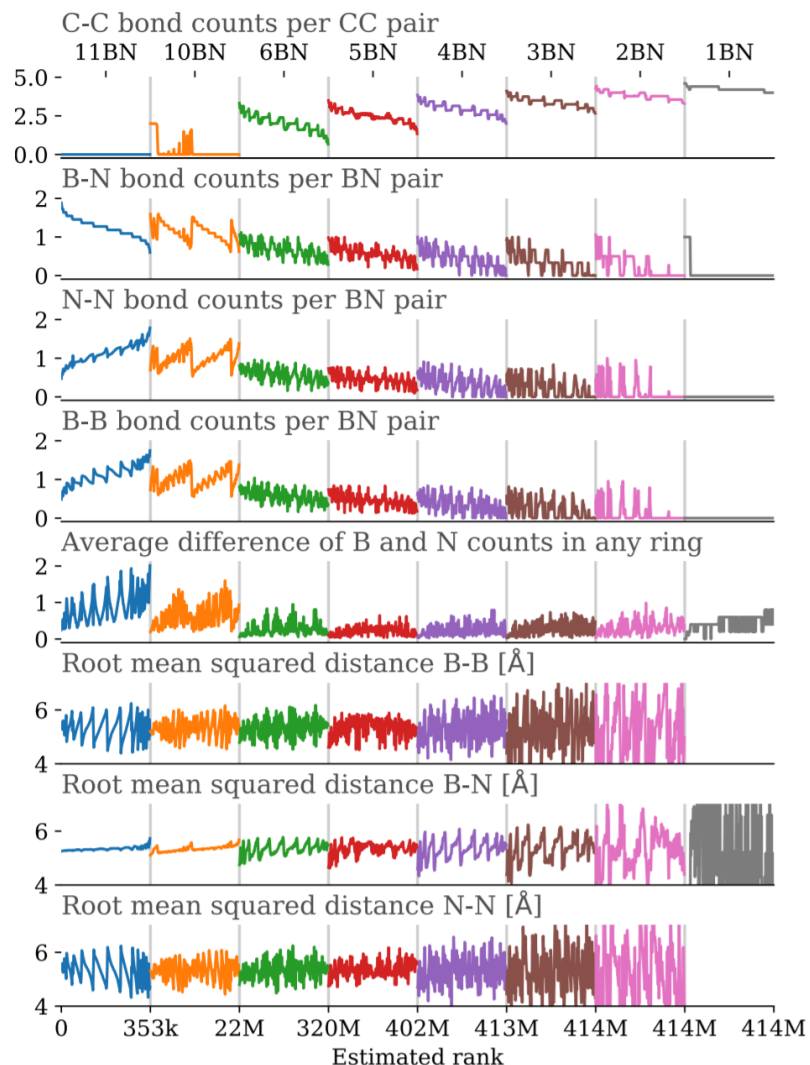
Design rules in order of decreasing strength
- Add BN pairs
- Maximize CC bonds
- Substitute sites shared between rings
- Maximize BN bonds
- Avoid N substitutions on rings sharing a larger amount of bonds with other rings
- Balance BN substitutions in each ring

Not a single QM calculation required!

Derivatives without electronic perturbations

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

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

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

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

SL Krug, **GFvR**, O. A. von Lilienfeld, *arXiv, 2203.13794*, **2022.**

Derivatives without electronic perturbations

$$E_{\text{t}} = E_{\text{r}} + \Delta E^{\text{NN}} + \int_{\Omega} d\mathbf{r} \sum_{n=0}^{\infty} \frac{1}{(n+1)!} \Delta v \frac{\partial^n \rho_\lambda(\mathbf{r})}{\partial \lambda^n}\bigg|_{\lambda=0}$$
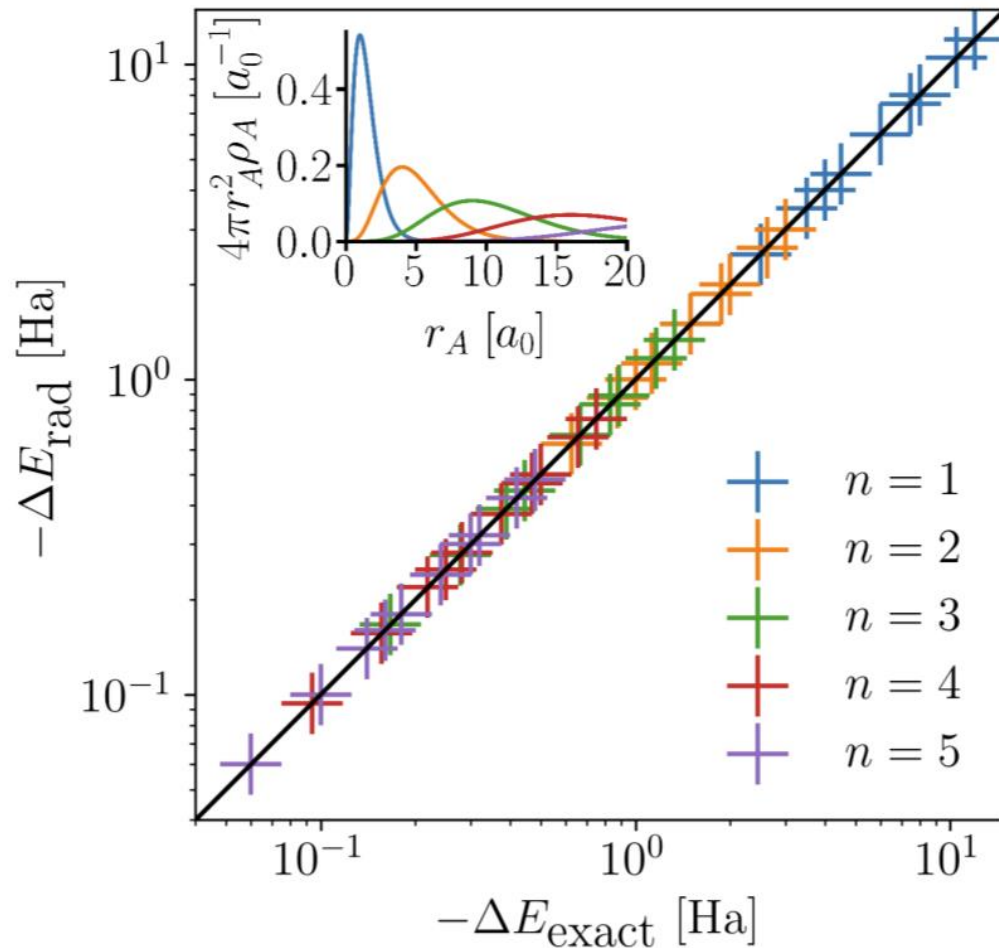
3D, arbitrary order $2n + 1 \to 0$

$$\Delta E = \int_{\Omega} d\mathbf{r} K(\mathbf{r}, v_i, v_f) \rho(v_i)$$

Ugly, but analytical

Only one!

SL Krug, **GFvR**, O. A. von Lilienfeld, *arXiv, 2203.13794*, **2022.**

Hydrogen-like atom

Multi-electron atom

SL Krug, **GFvR**, O. A. von Lilienfeld, *arXiv, 2203.13794*, **2022.**

Limitations
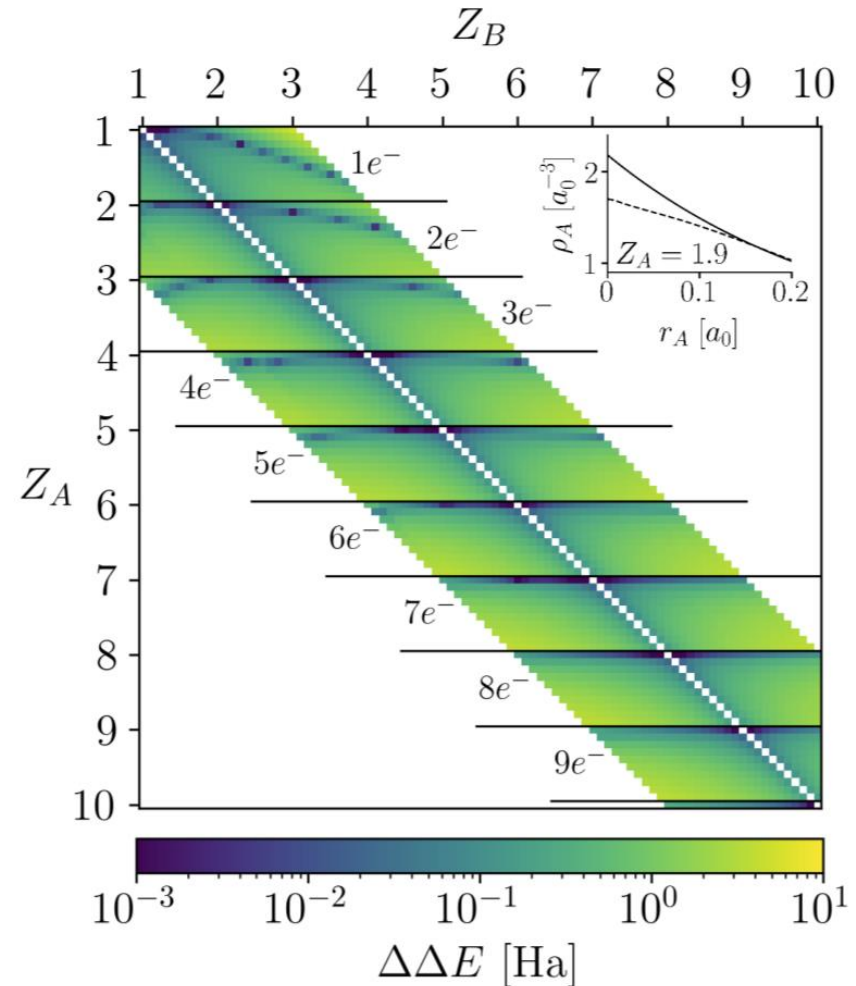
Both energy and density derivatives are hard
- Finite differences   expensive, numerical instabilities
- Hellmann-Feynman   finite order, basis set inaccurate
- Coupled-perturbed   finite order, tedious
- Conceptual DFT   memory hungry, no post-HF
- Automatic differentiation   niche: DiffiQult, quax, dqc, …

ferchault/APDFT   ferchault/APHF

aspuru-guzik-group/DiffiQult

CCQC/Quax

diffqc/dqc   Alchemy!

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
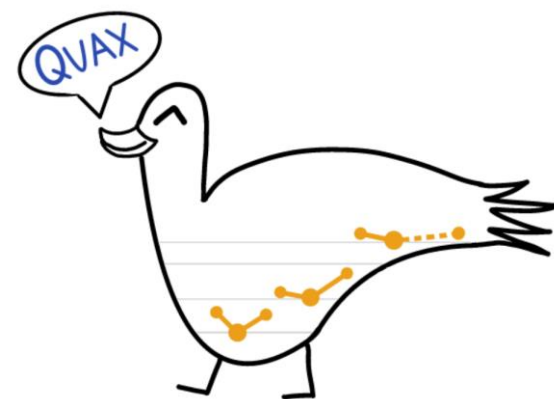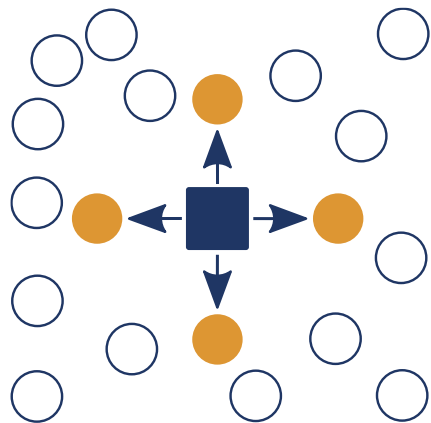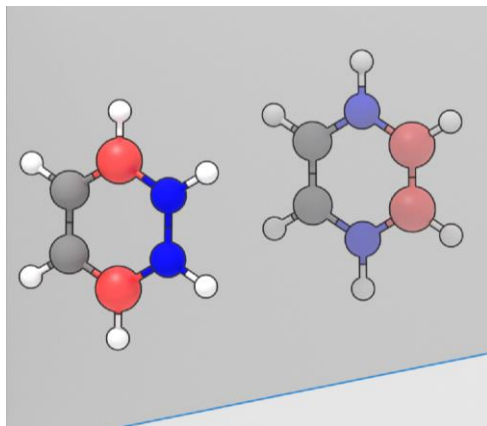- Unless complete basis set limit: Pulay terms

Convergence
- Finite radius
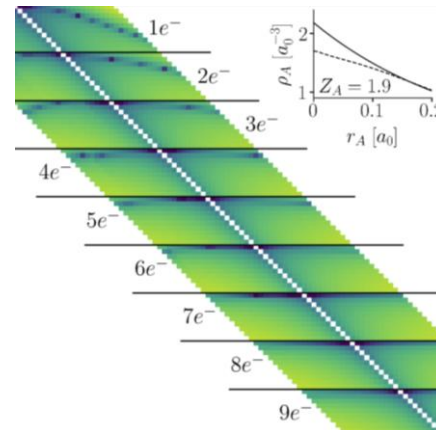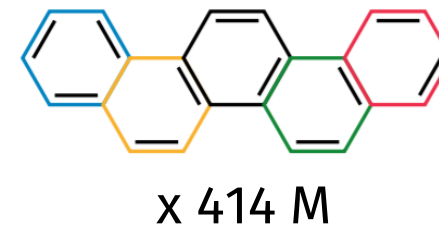- Not all systems are made equally

Quantum Alchemy yields systematically improvable results.

Closed expressions reveal structure of chemical space.

Efficient ways to obtain derivatives in progress.

x 414 M

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 | arXiv 2022, 2203.13794.

ferchault   @ferchault   guido.vonrudorff.de

Simon Krug

Anatole von Lilienfeld