Going EAST in Chemical Space: Challenges, Concepts, and Constraints

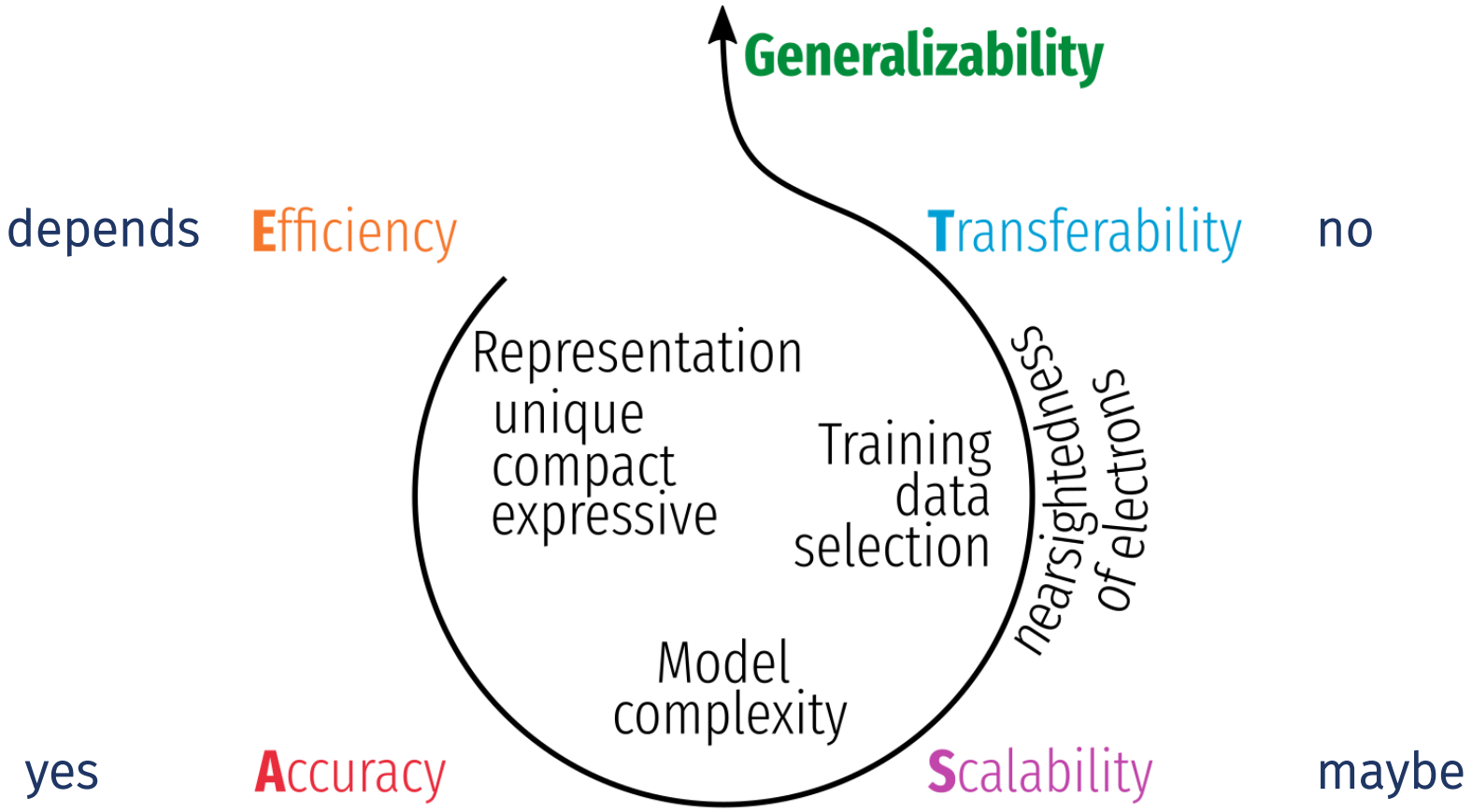
Guido Falk von Rudorff, University of Kassel


ferchault


@ferchault

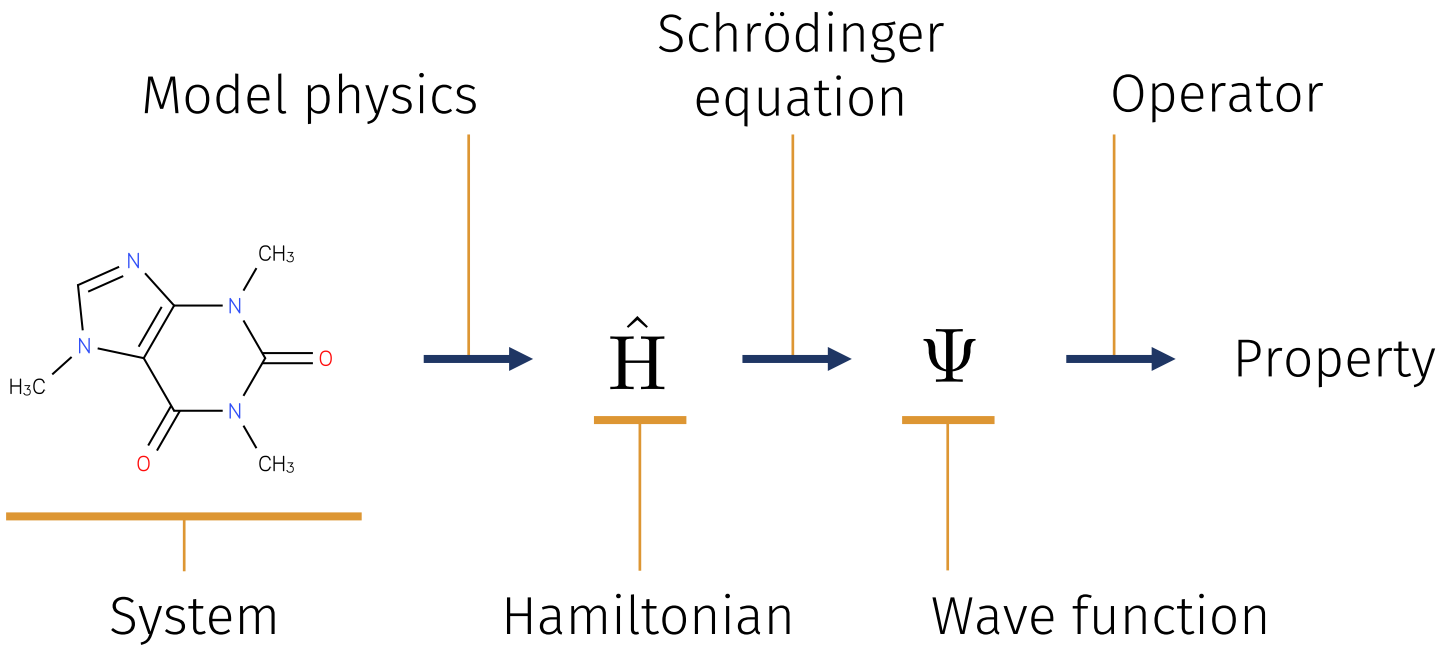

nablachem.org

EAST



B. Huang, **GFvR**, A. von Lilienfeld, *Science*, **2023** (accepted).

Solved?


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

Cost

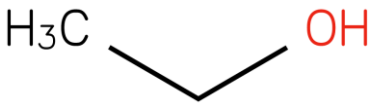
$$\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_n)$$

Methane

CH_4

N_2

Ethanol



Solved by approximations in computational chemistry?

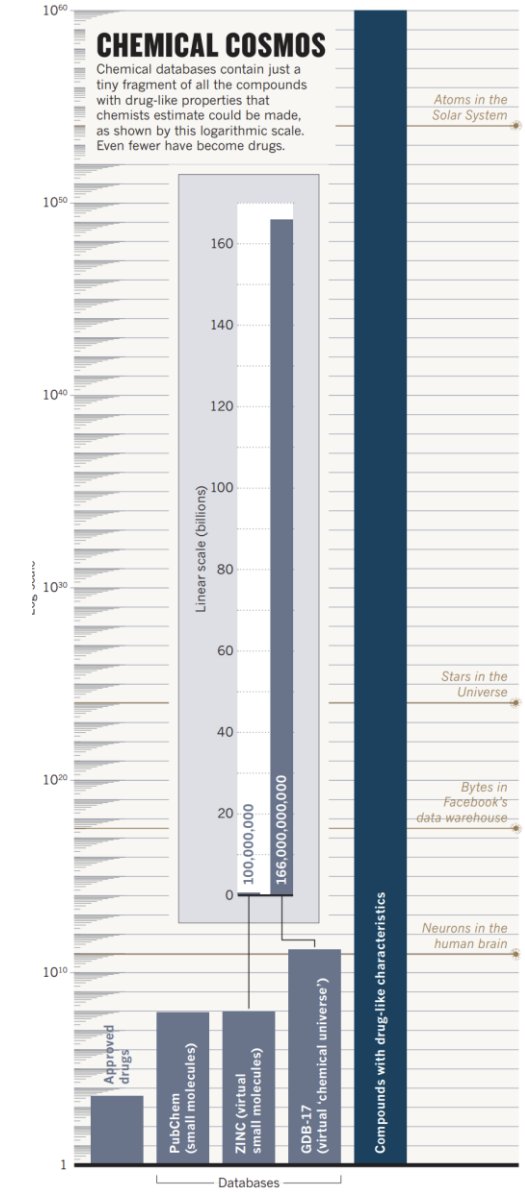
Scaling of Molecules

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.

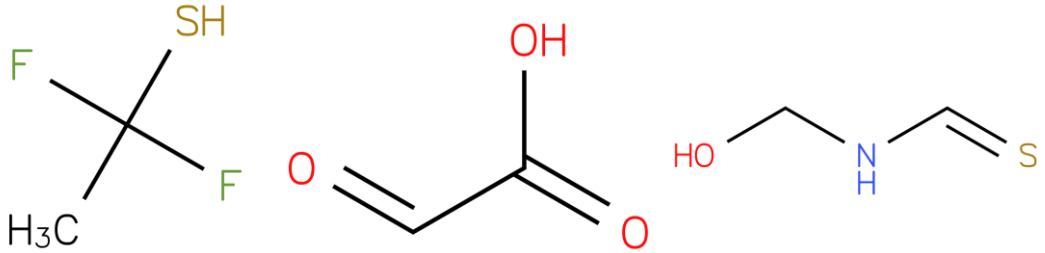
3D Geometries

Take some atoms from {C, O, N, F, S}, H-saturated

349:

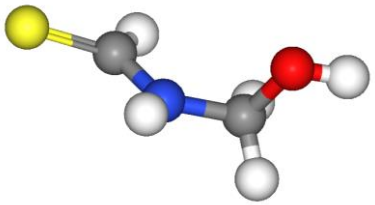
$C_2H_4F_2S$
$C_2H_2O_3$
C_2H_5NOS
...

9,917:



...

~52k:

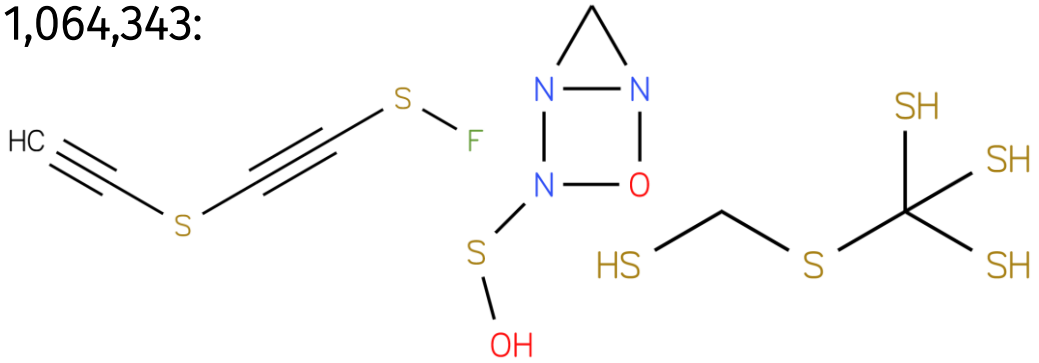


...

1,050:

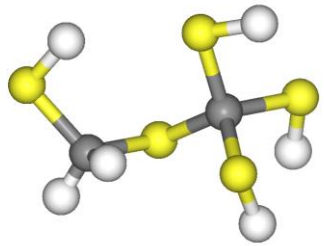
C_4HFS_2
$CH_3N_3O_2S$
$C_2H_6S_5$
...

1,064,343:



...

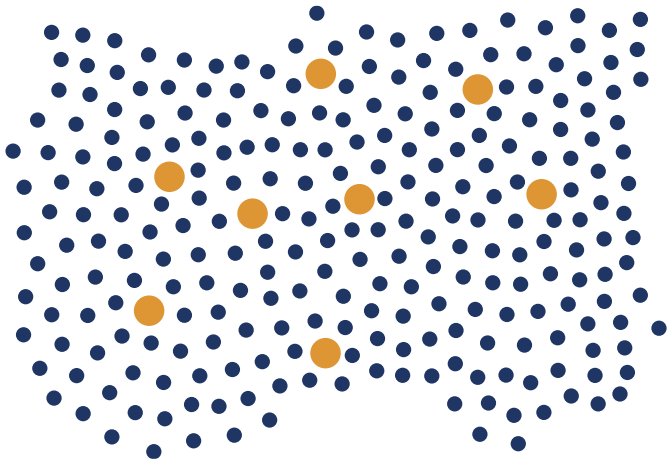
~23M:

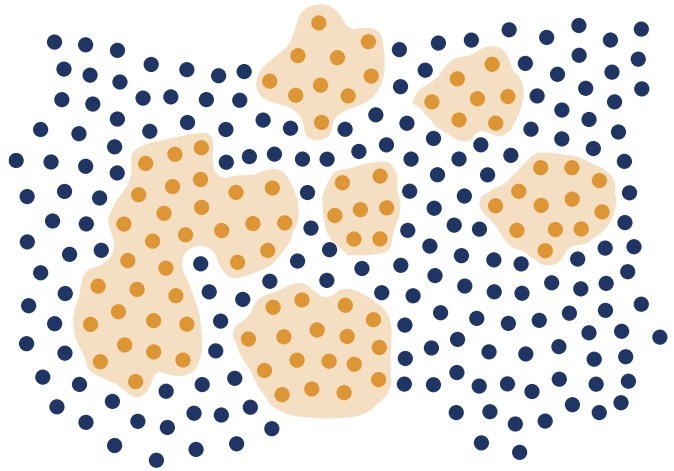


...

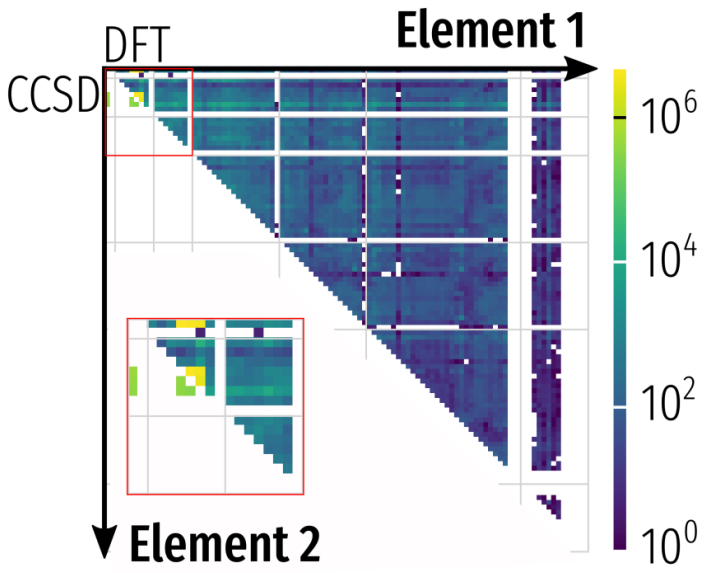
Constraint 1

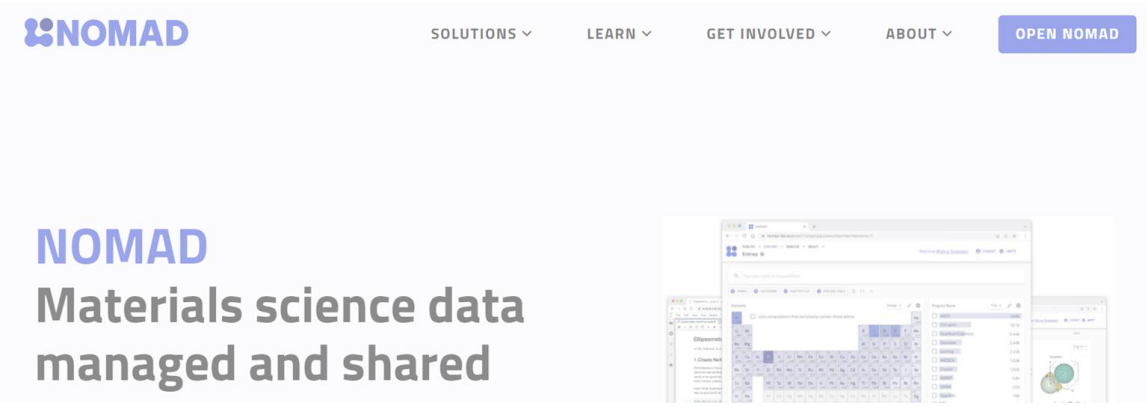
Speed does not matter:
even enumeration is impossible.

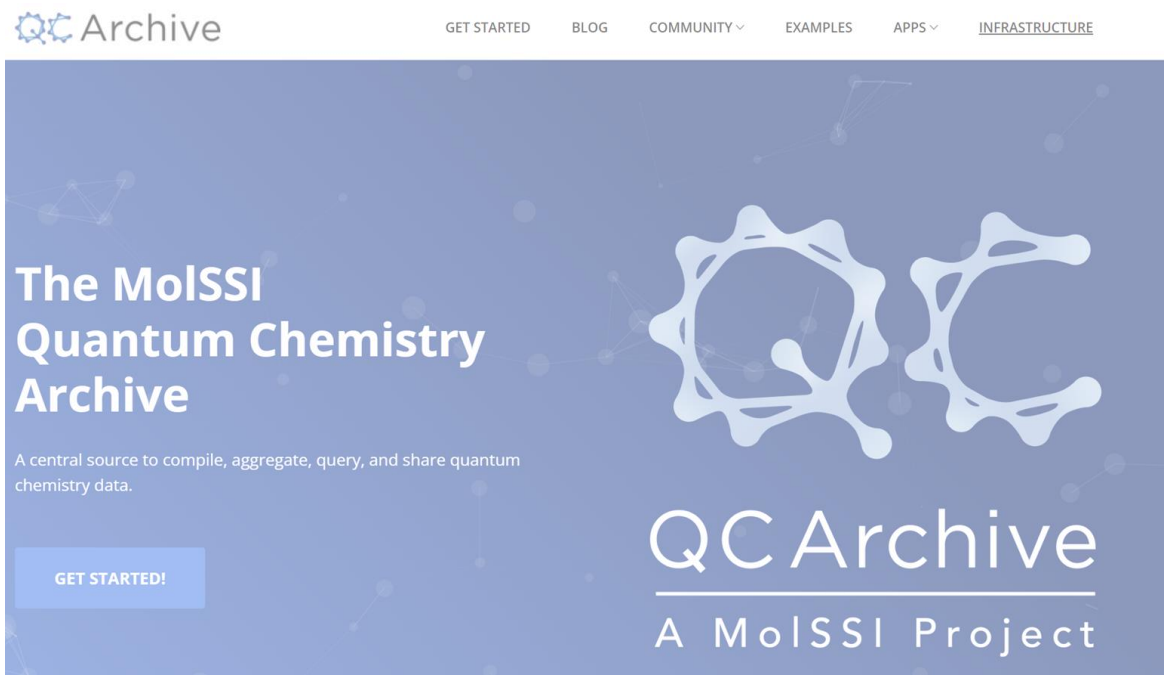




Data Availability



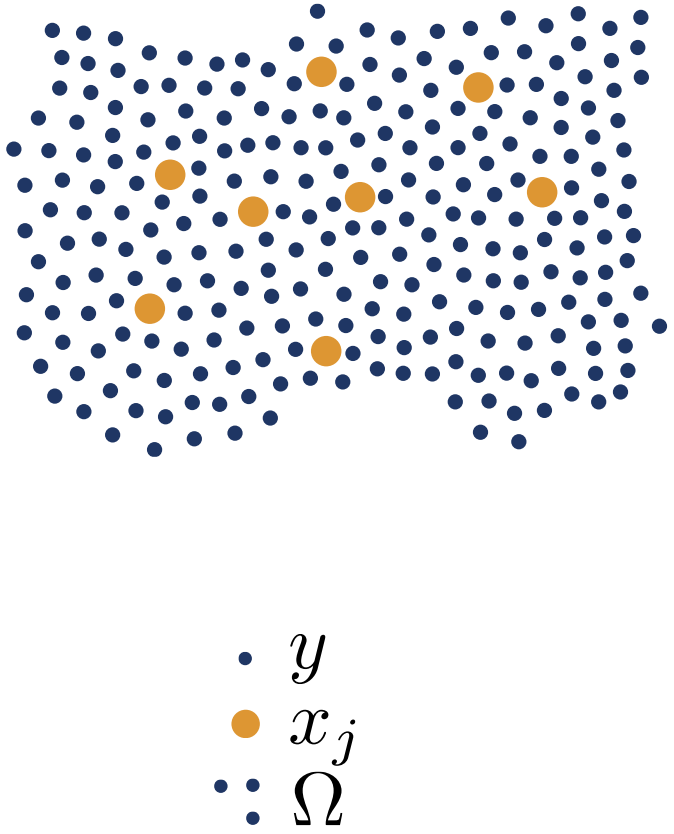




Fill Distance

$$h \equiv \sup_{y \in \Omega} \min_{x_j \in X} \|y - x_j\|_2$$

$$\text{Error} = a \exp(-c/h)$$

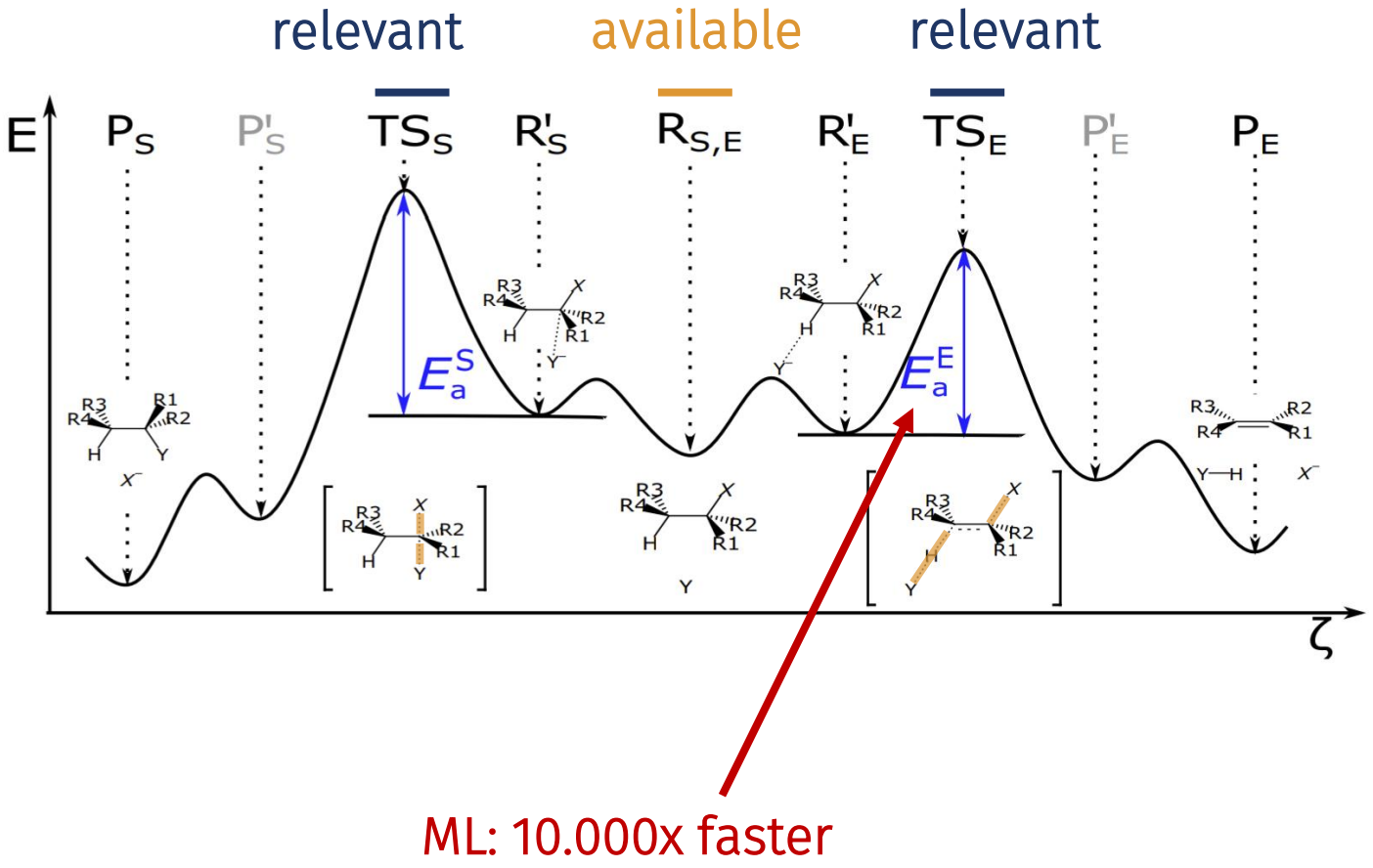


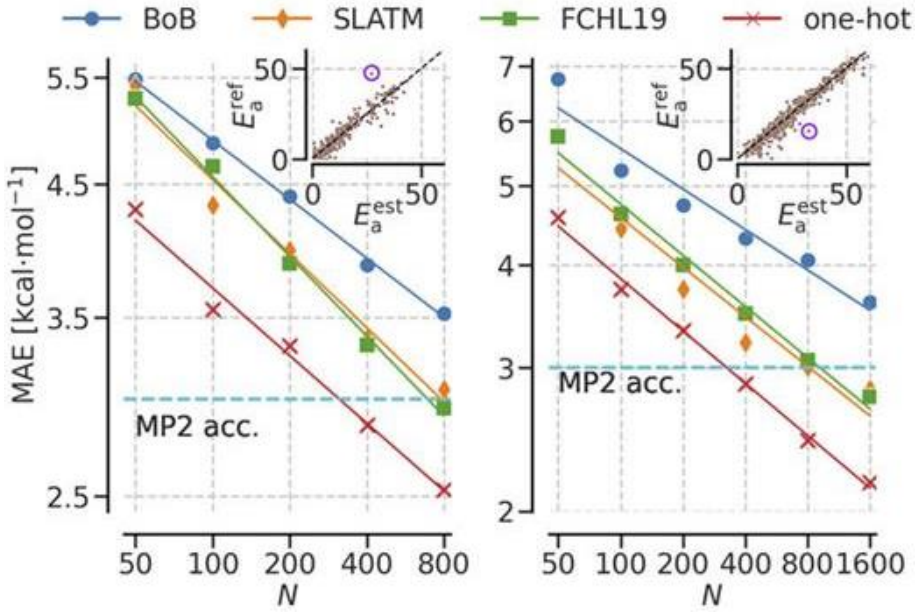
W.R. Madych, S.A. Nelson, *J Approx Theory*, 1992.

Constraint 2

Transferable models require a compact chemical space.
Or ludicrous amounts of data.

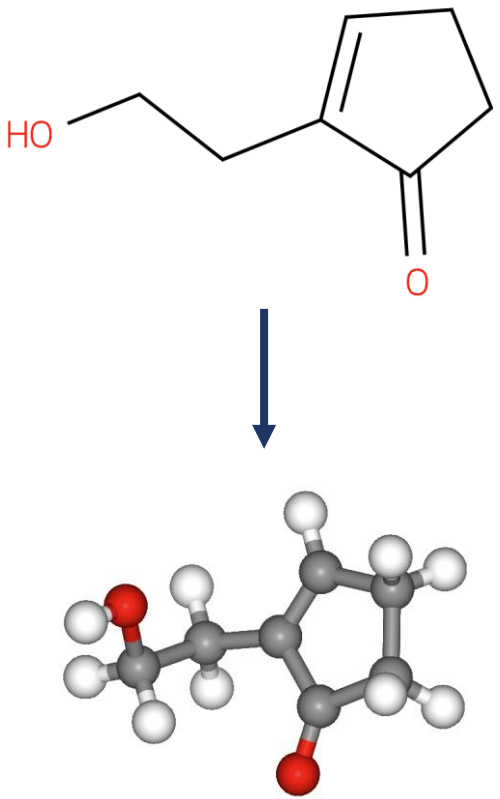
Concept: Avoid Knowing the PES



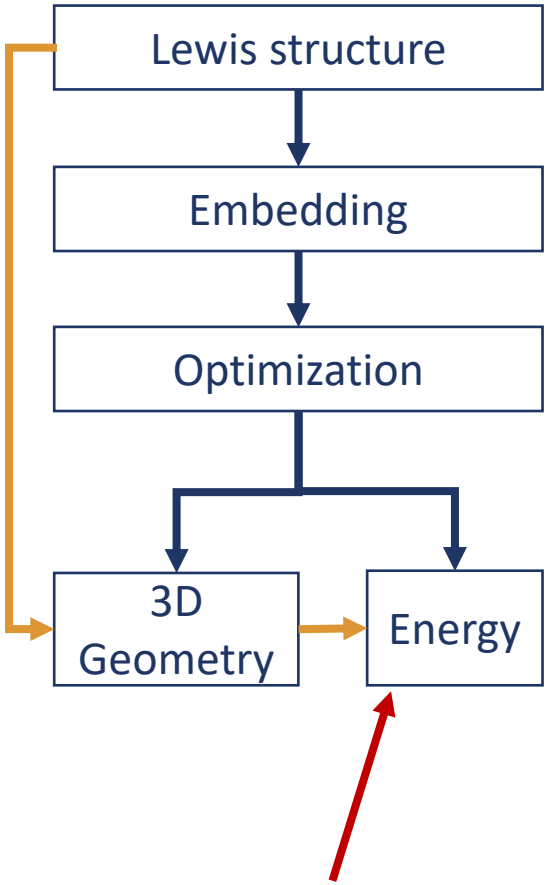


GFvR, S.N. Heinen, M. Bragato, O. A. von Lilienfeld, *Mach. Learn.: Sci. Technol.* 2020.
S.N. Heinen, GFvR, O. A. von Lilienfeld, *J. Chem. Phys.*, 2021.

Concept: Avoid Knowing the PES



Traditional / G2S



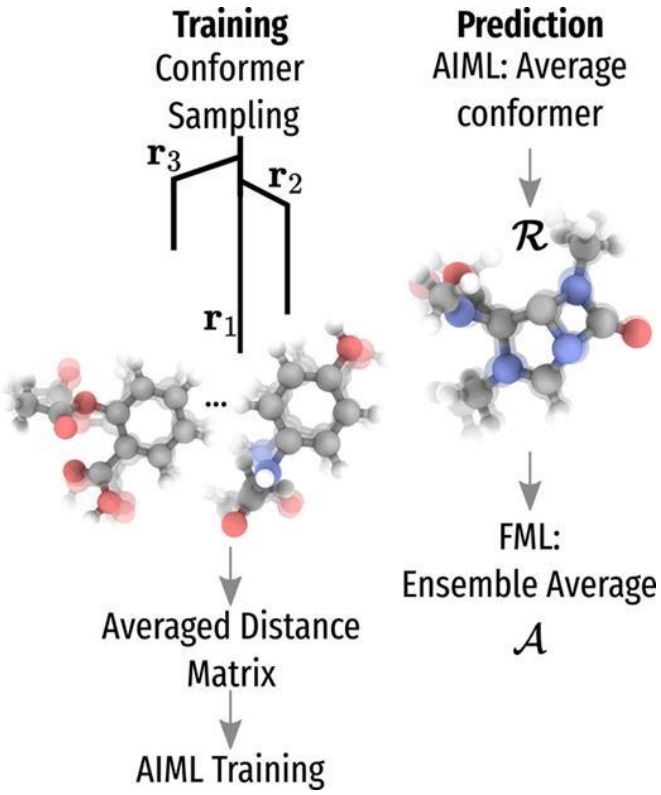
ML: 100.000x faster

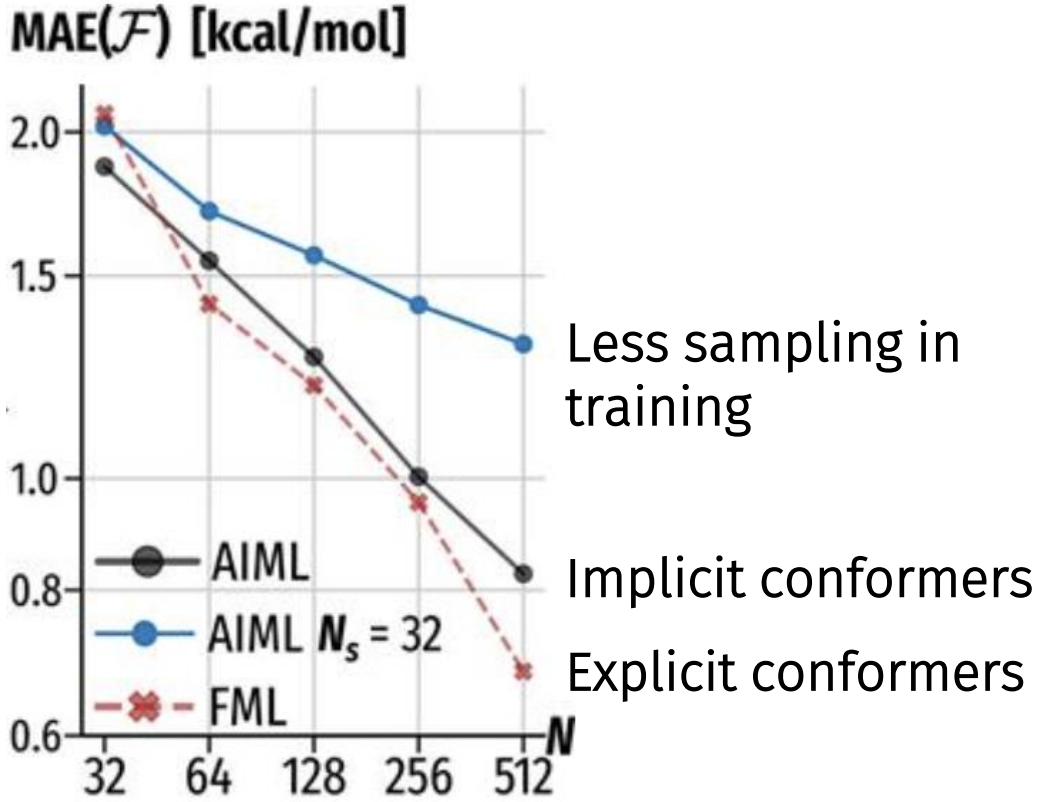
G2S

- Closer to DFT than common methods
  - Small molecules
- Applicable to complex chemical spaces
  - Transition state geometries
  - Carbenes
  - Elpasolite crystals

D. Lemm, **GFvR**, O. A. von Lilienfeld, *Nat. Commun.* **2021**.

Concept: Consider Ensembles





Less sampling in training

Implicit conformers

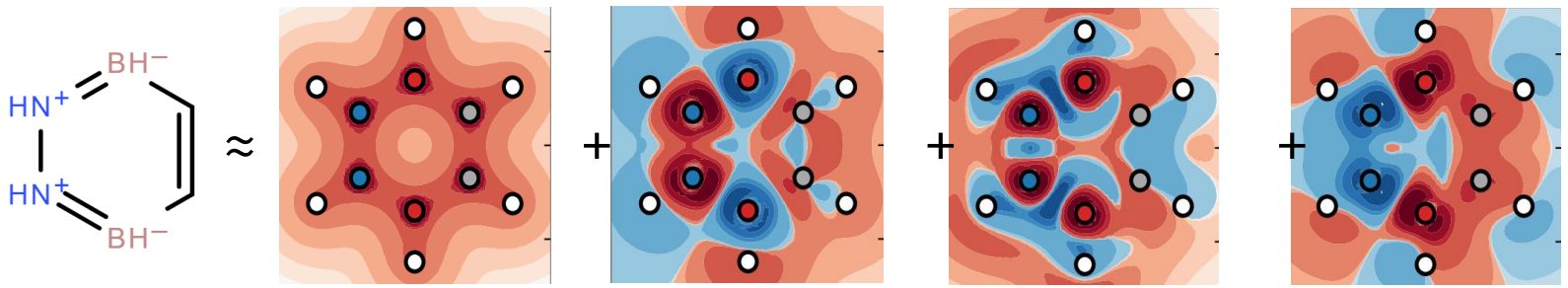
Explicit conformers

J Weinreich, D Lemm, GFvR, O. A. von Lilienfeld, J Chem Phys, 2022.

Quantum Alchemy

Taylor expansion

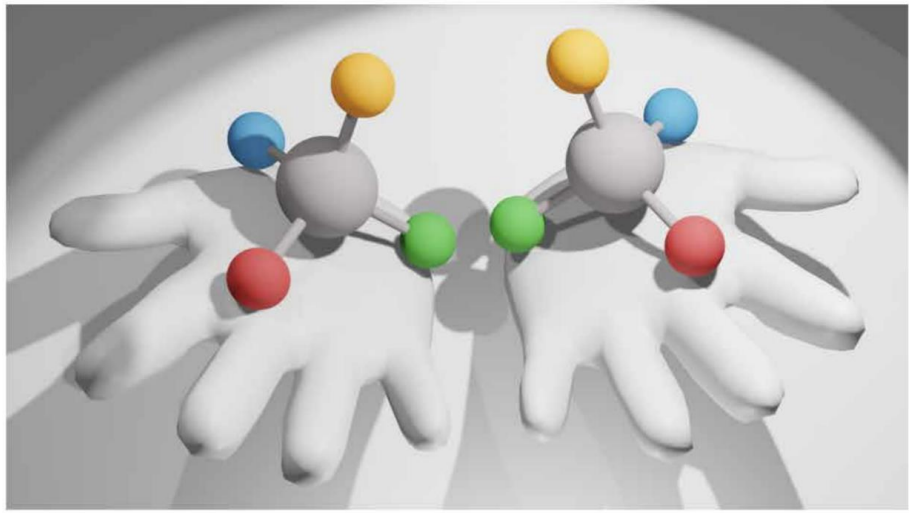
- Energy function of
  - Geometry — Forces, Vibrations
  - Nuclear charges — Alchemical changes
- Idea: obtain dominant leading derivatives, predict many systems

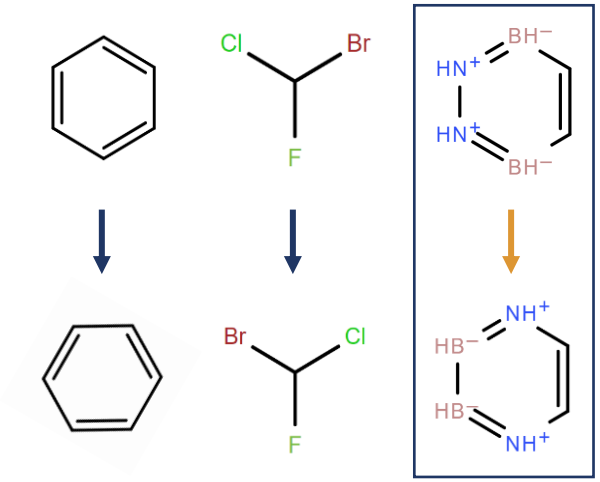


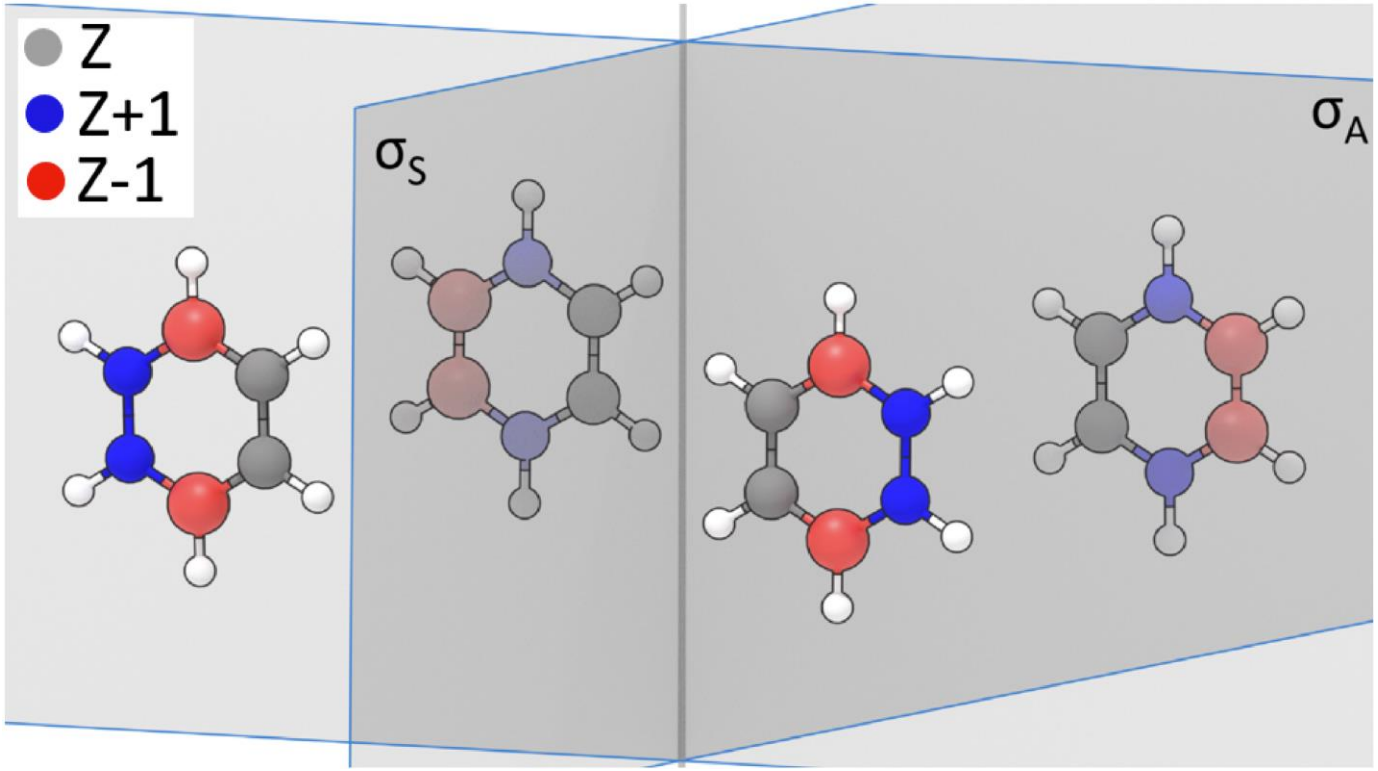
$$\hat{H}(\lambda) \equiv \lambda \hat{H}_{\mathrm{t}} + (1-\lambda)\hat{H}_{\mathrm{r}}$$

E. B. Wilson, *J. Chem. Phys.* 1962.
GFvR, O. A. von Lilienfeld, *Phys. Rev. Res.*, 2020.

Concept: Use Symmetries





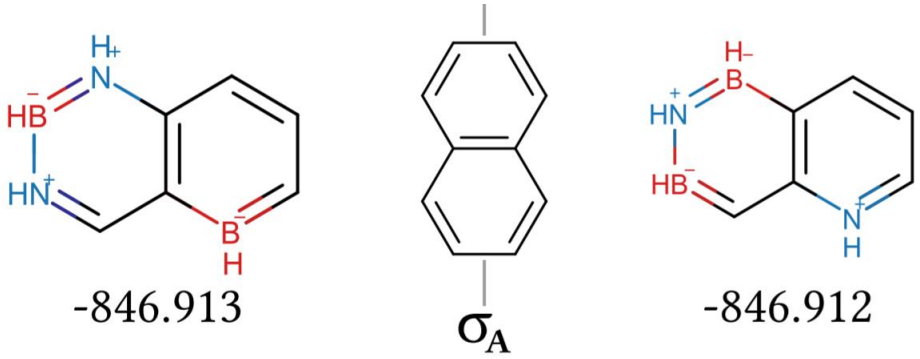


GFvR, O. A. von Lilienfeld, *Science Adv.* 2021.

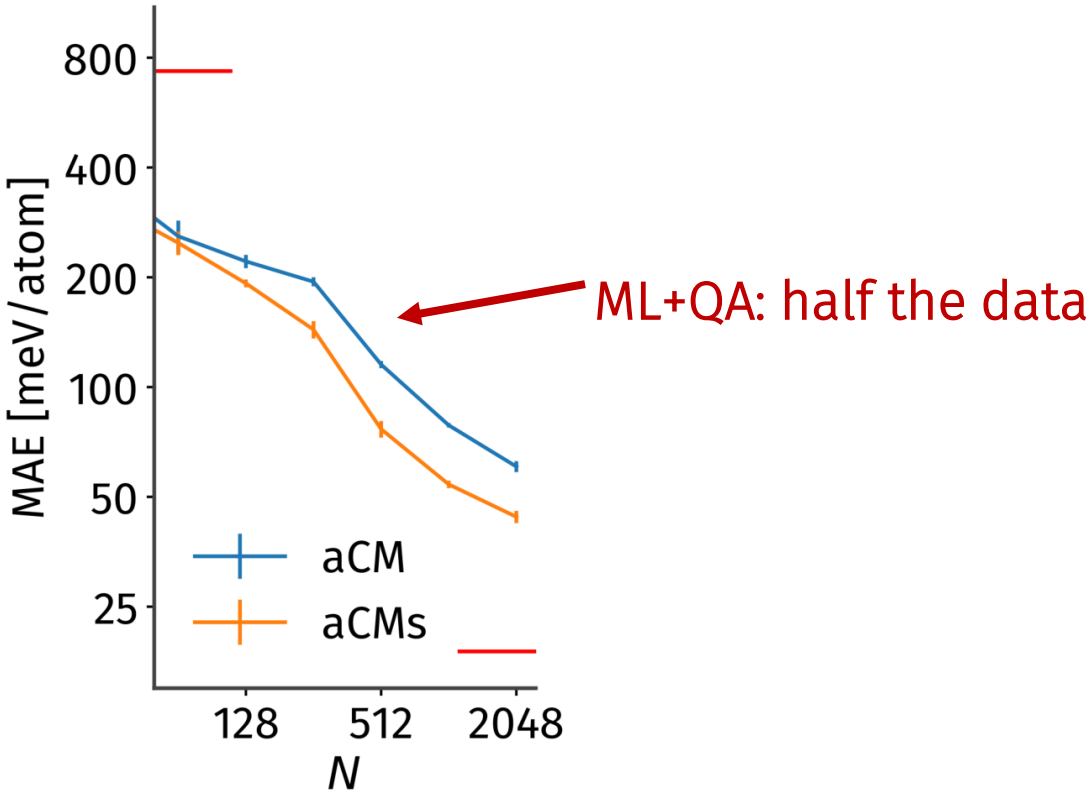
Concept: Use Symmetries

Alchemical Enantiomers

Electronic energy only

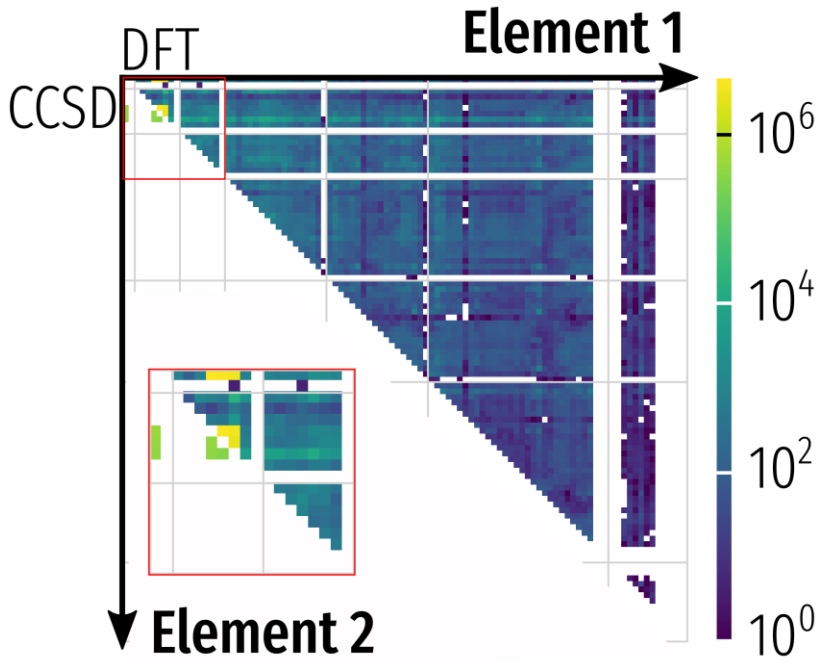


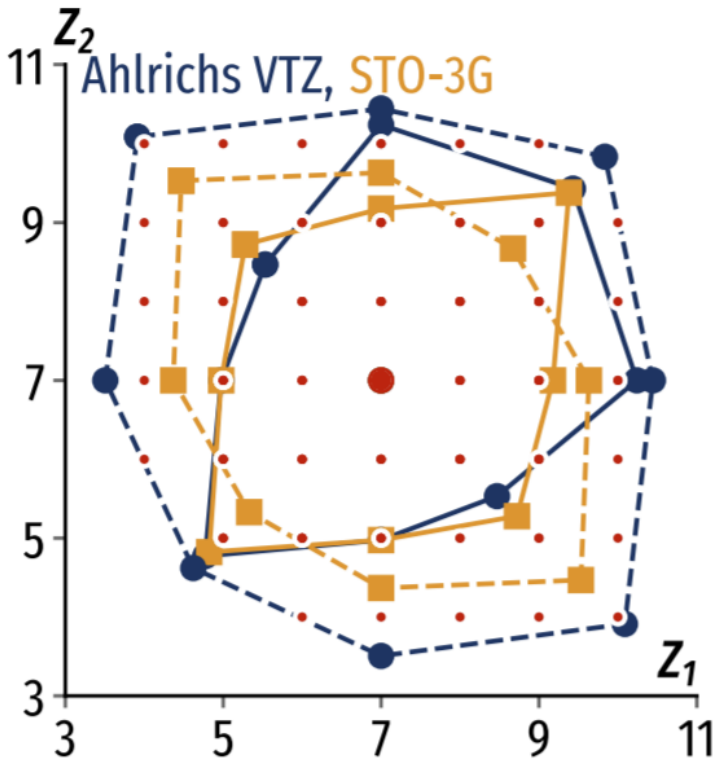
Speed up machine learning



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

Challenge: Element Diversity

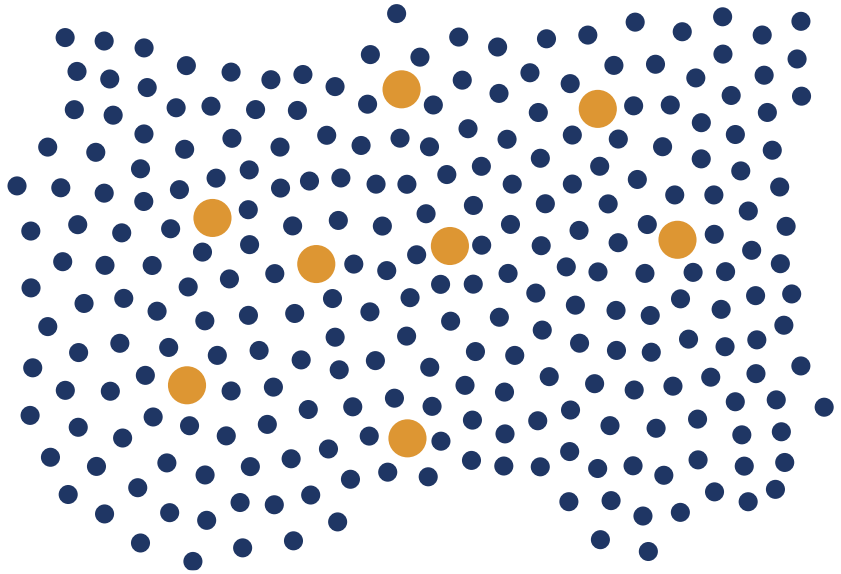




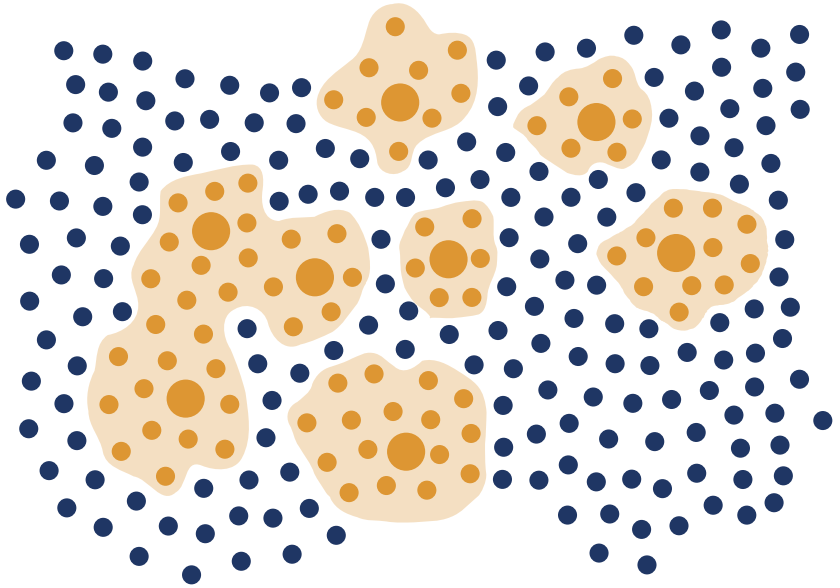
GFvR, *J. Chem. Phys.* 2021.

Challenge: Element Diversity

Without Perturbation



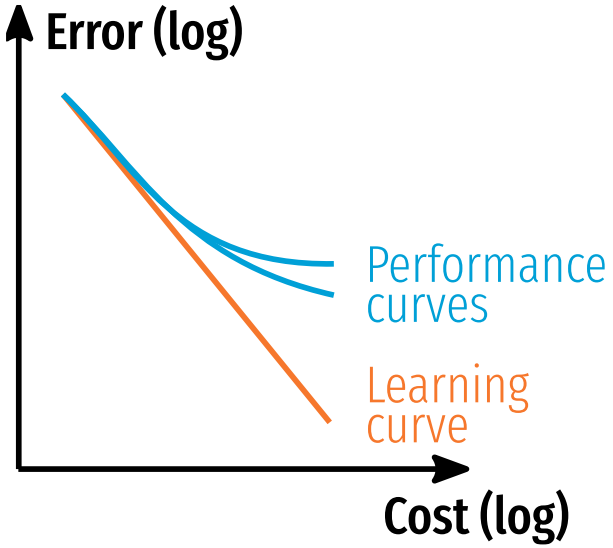
With Perturbation

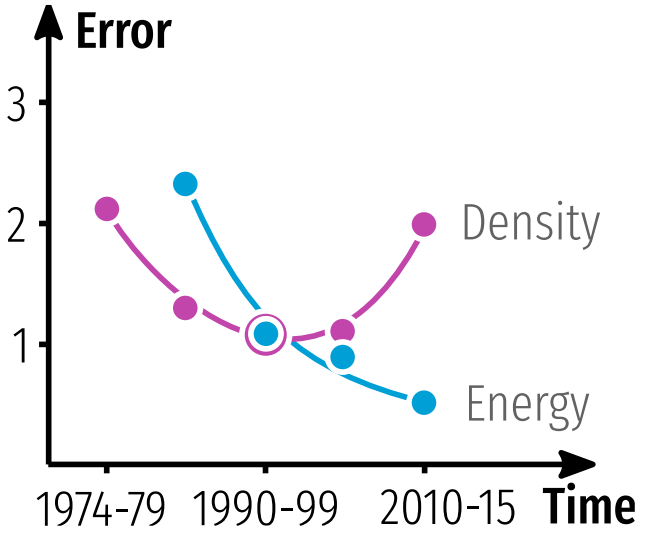


Systems/Molecules

- Any
- Known
- Approximated

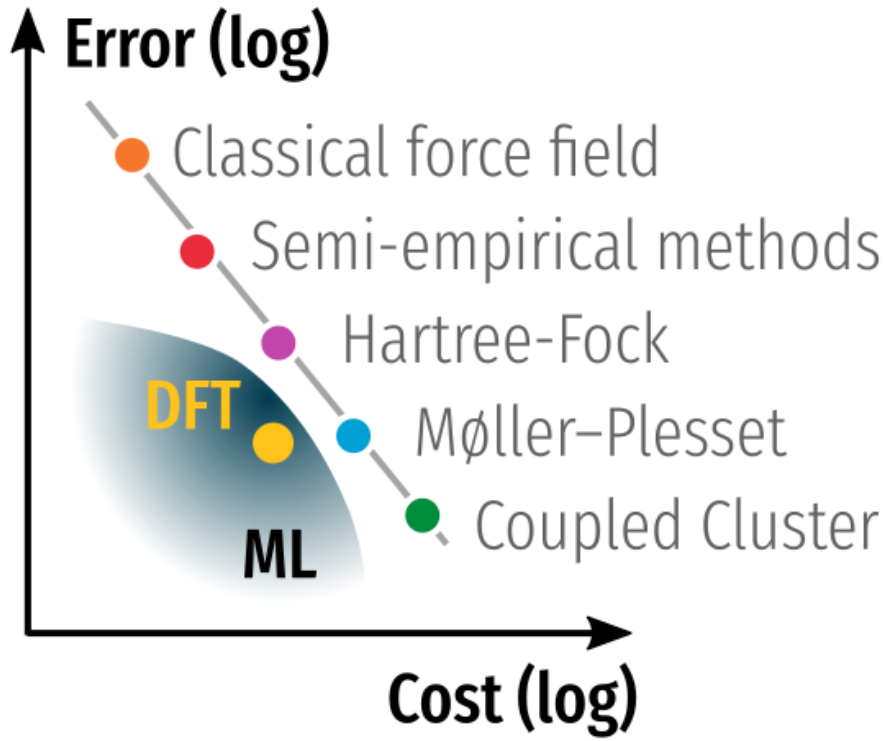
Challenge: Multi-Property Training





M.G. Medvedev et al., *Science*, 2017.

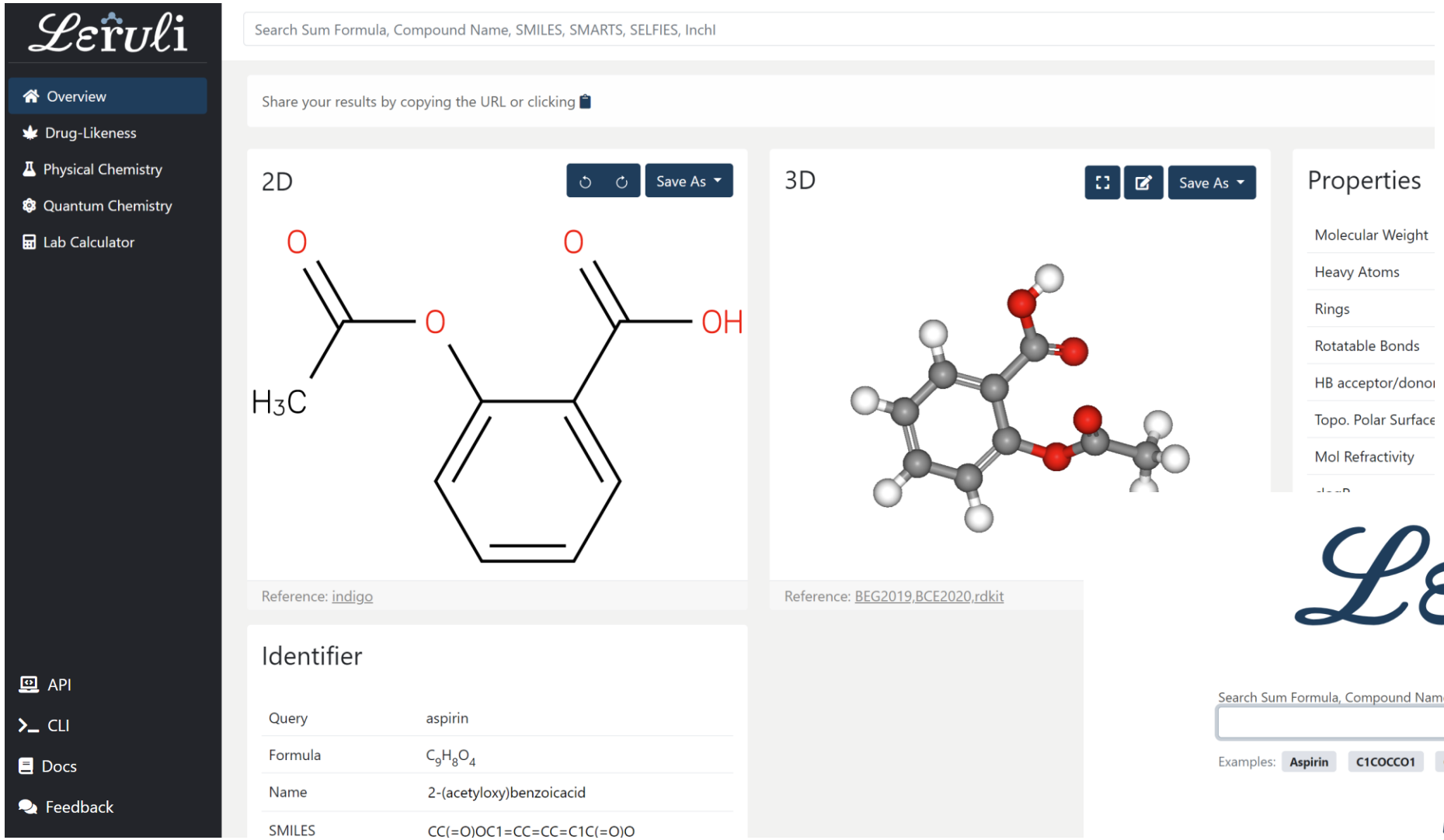
Challenge: Prediction Performance



How accurate can we be given
- an application and
- a compute budget?

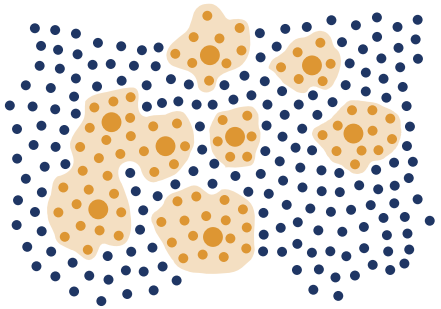
B. Huang, **GFvR**, A. von Lilienfeld, arXiv, **2023**.

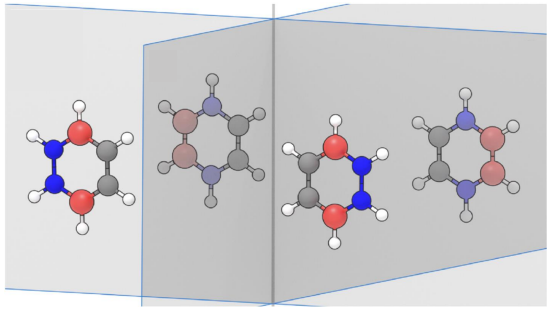
Challenge: Access to Models



Try it:
leruli.com

Summary





Concept | Avoid knowing the PES

Concept | Consider ensembles

Concept | Quantum Alchemy: derivatives and symmetry

Challenge | Element diversity

Challenge | Multi-property training

Challenge | Prediction performance

Challenge | Access to models

ferchault

@ferchault

nablachem.org
vonrudorff@uni-kassel.de

Thanks

Marco Bragato
Giorgio Domenichini
Emily Eikey
Stefan Heinen
Chasz Griego
Konstantin Karandashev
John Keith
Mario Krenn
Simon Krug
Dominik Lemm
Anatole von Lilienfeld
Alex Maldonado
Michael Sahre
Max Schwilk
Enrico Tapavicza
Jan Weinreich

We're hiring!