SAND2022-9105C

Sandia National Laboratories

Beyond the Black Box: The potential and problems of equivariant electron densities

Swiss Equivariant Learning Workshop

Lausanne, Switzerland – June 2022

Joshua A. Rackers, Truman Fellow

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Who am I?

ship

Who am I?

Physics and Political Philosophy

TEACH FOR AMERICA

High School Physics and Chemistry

Washington University in St. Louis

Biophysics Ph.D.

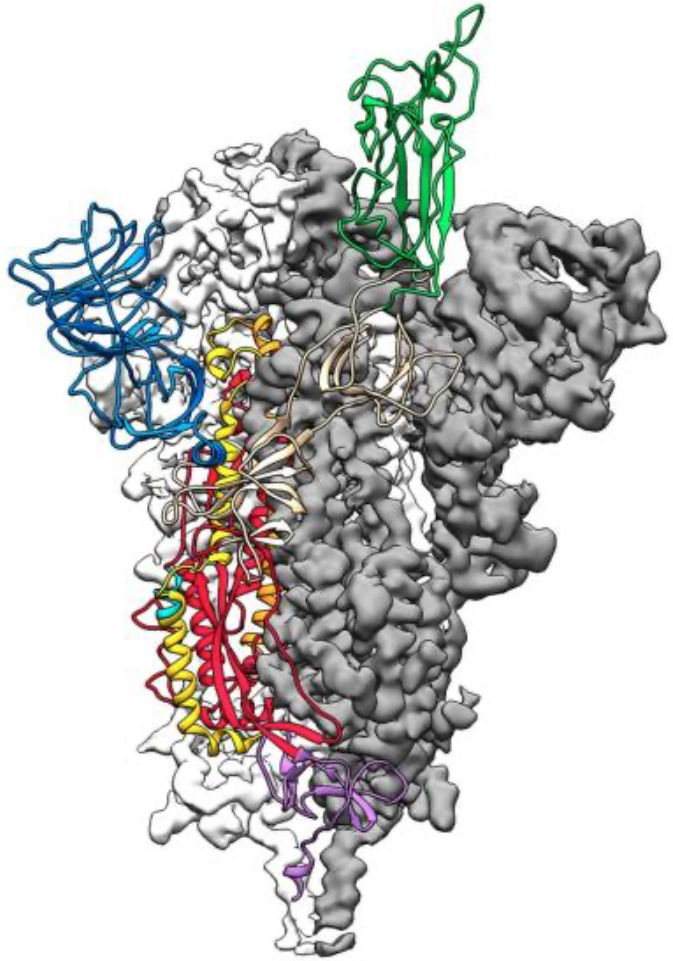
Truman Fellowship

The Grand Challenge of Molecular Modeling

Can we predict the behavior of molecules at the atomic scale?

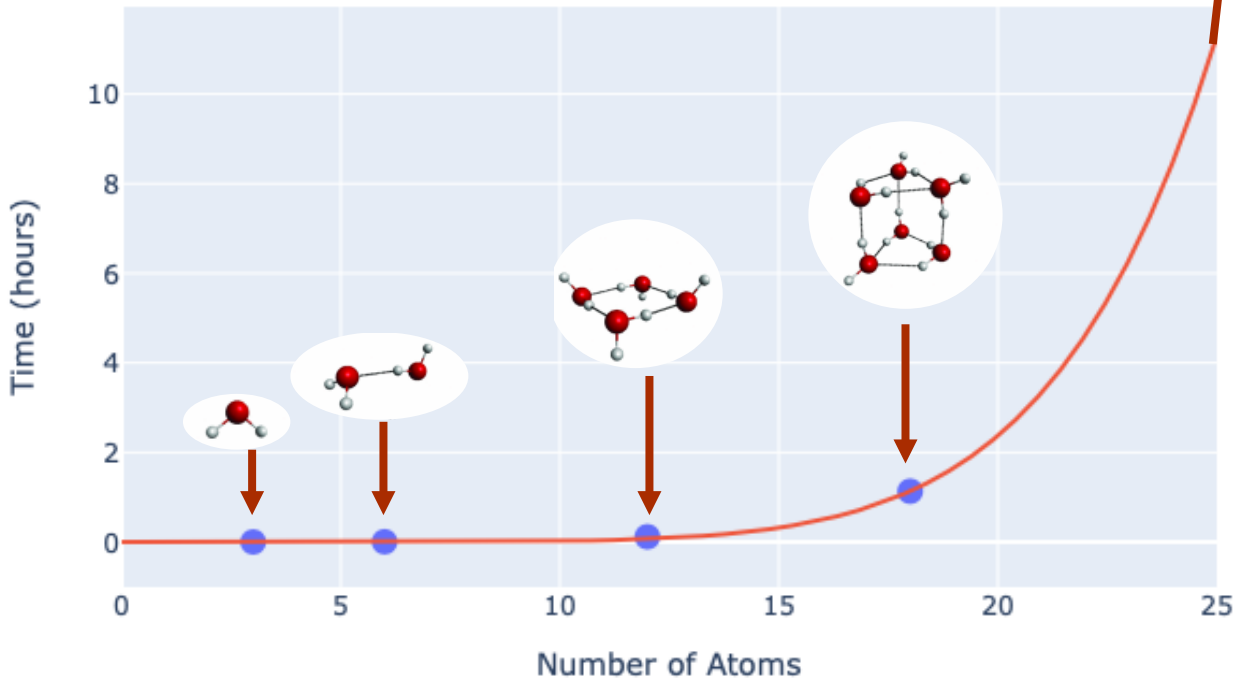
$$F = ma$$

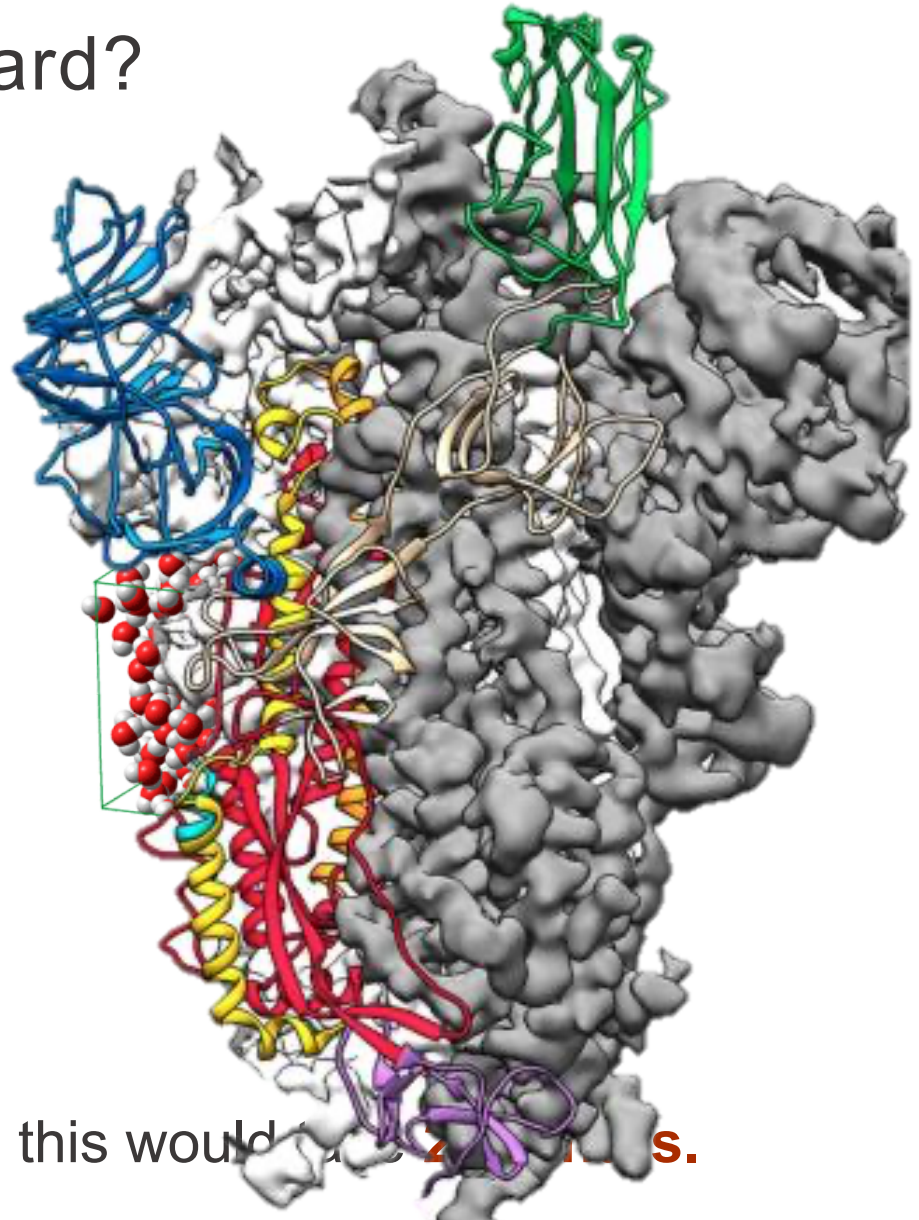
$$H\Psi = E\Psi$$



Why is doing Quantum Mechanics hard?

Solving $H\Psi = E\Psi$ scales horrendously!

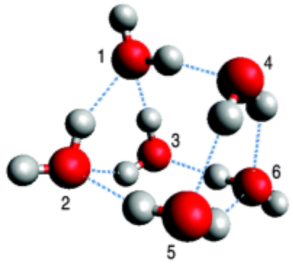




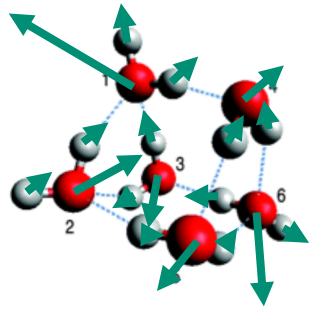
For **50 atoms**, this would [illegible] **s.**

For **100 atoms**, this would take **20**

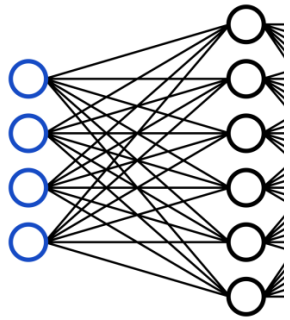
Let's use Machine Learning!

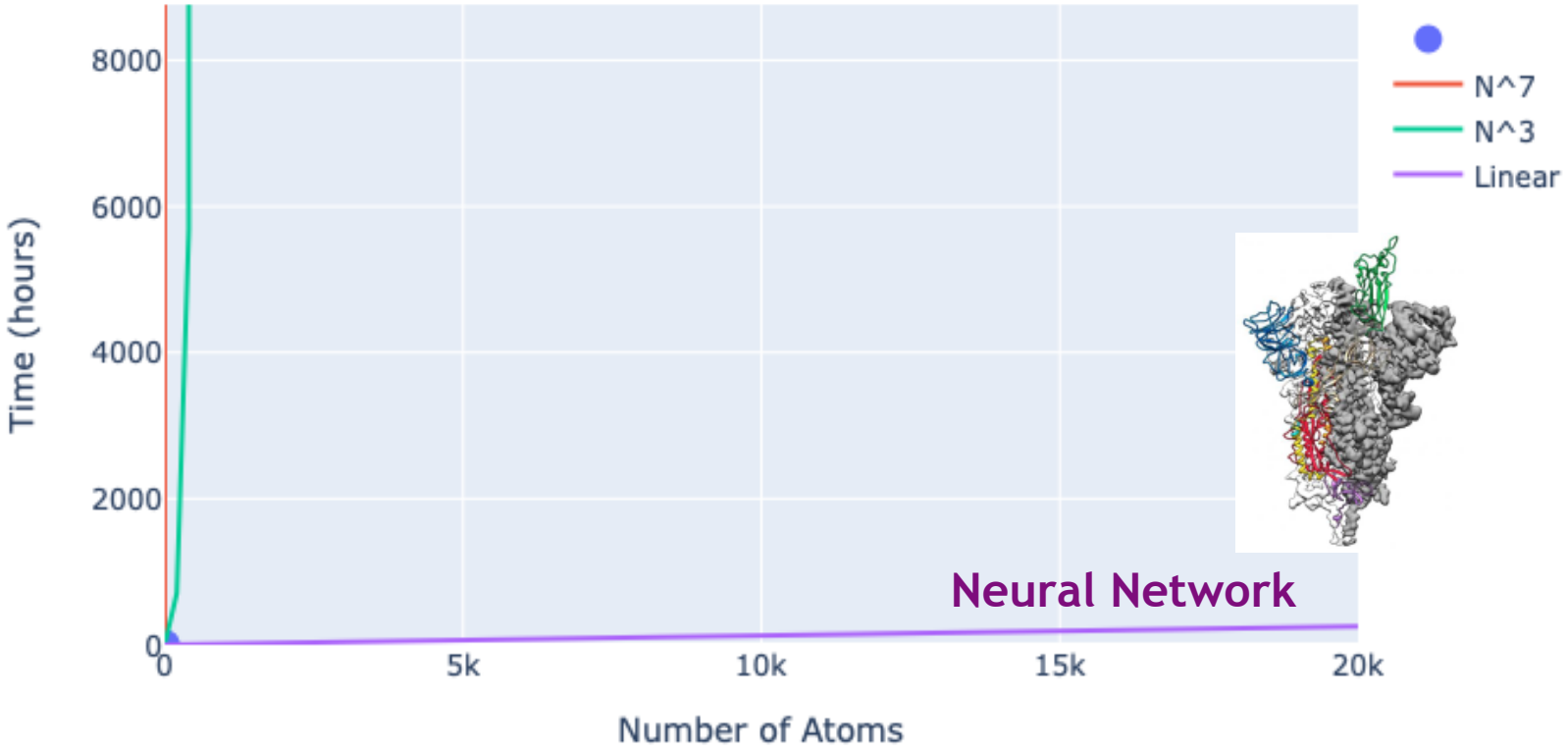


Quantum Mechanics



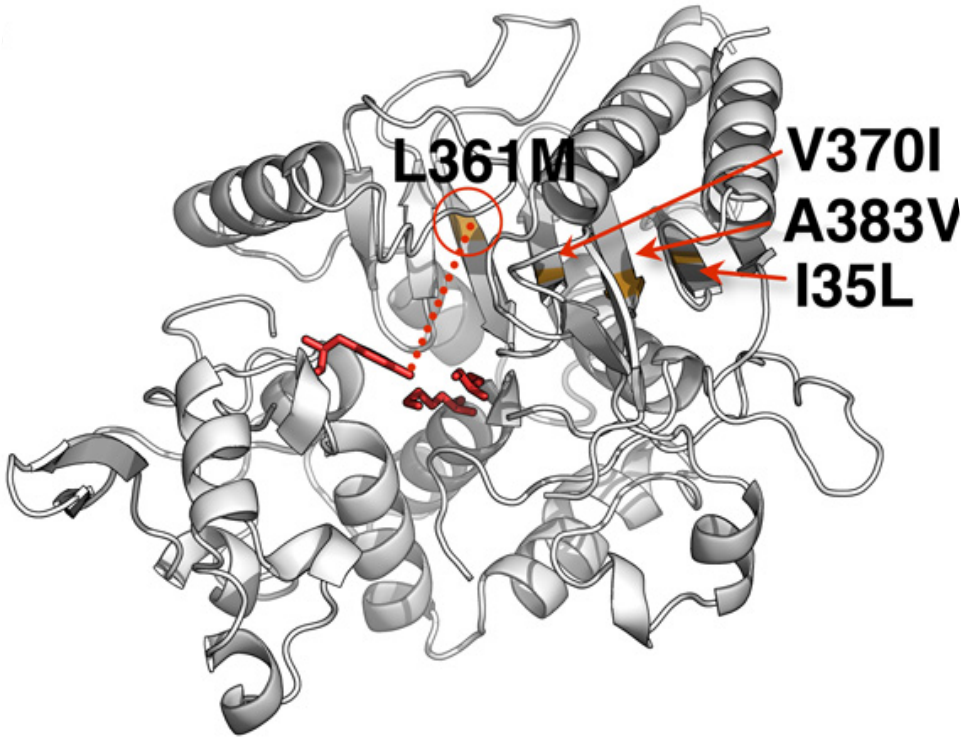
Neural Network



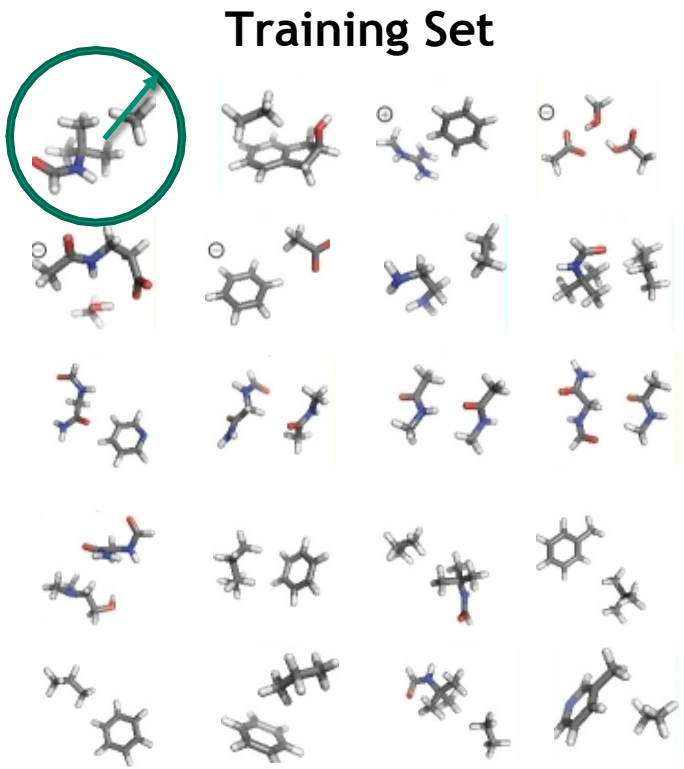


What's the Problem? – Part 1: Training Data

We want to use ML on large molecules, but we can't train on things that big!



Gonzalo, Osuna, et al., *Nature chemical biology* (2014)



Effective Radius
10 Å

What's the Problem? – Part 1: Training Data

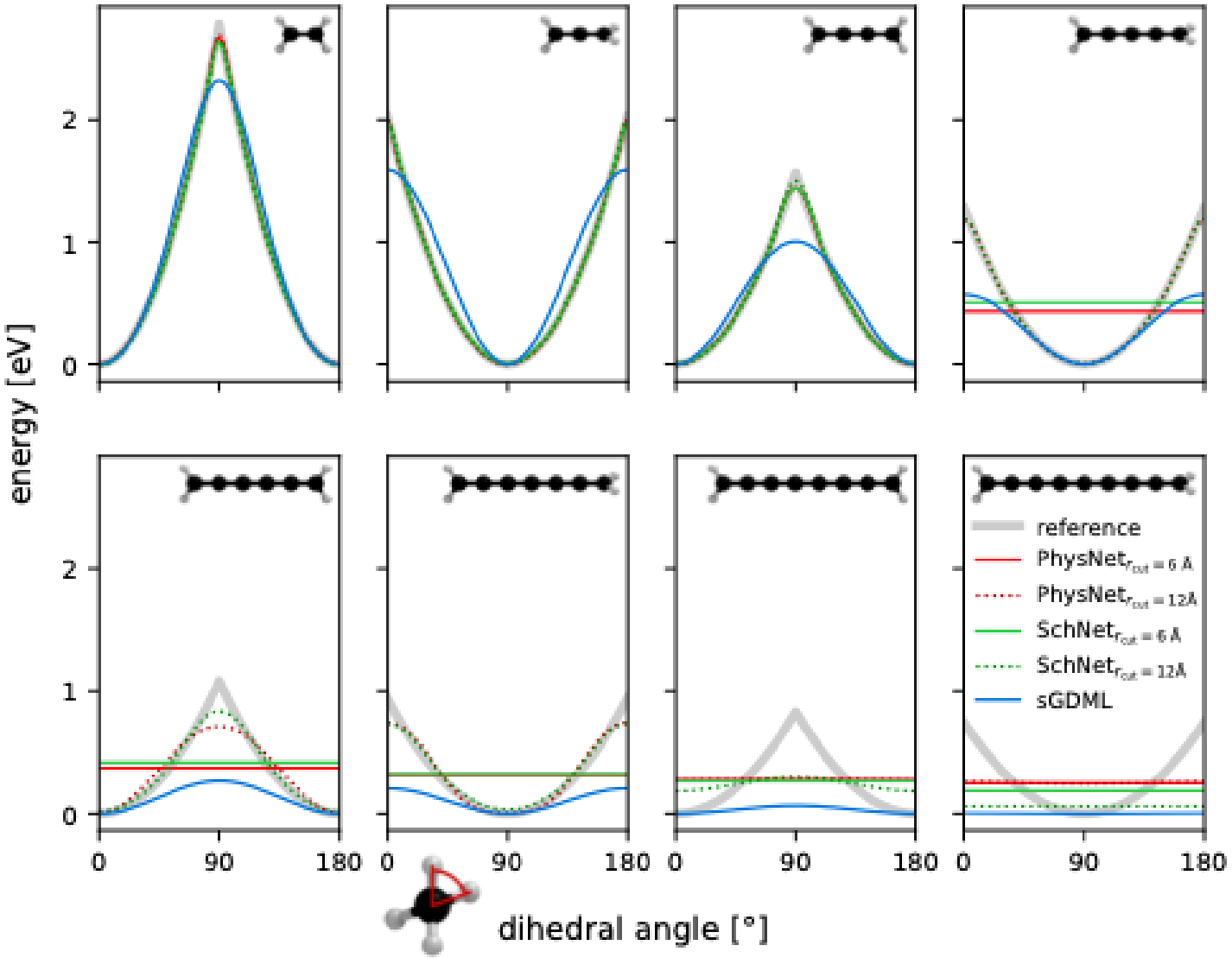
CHEMICAL REVIEWS

pubs.acs.org/CR

Review

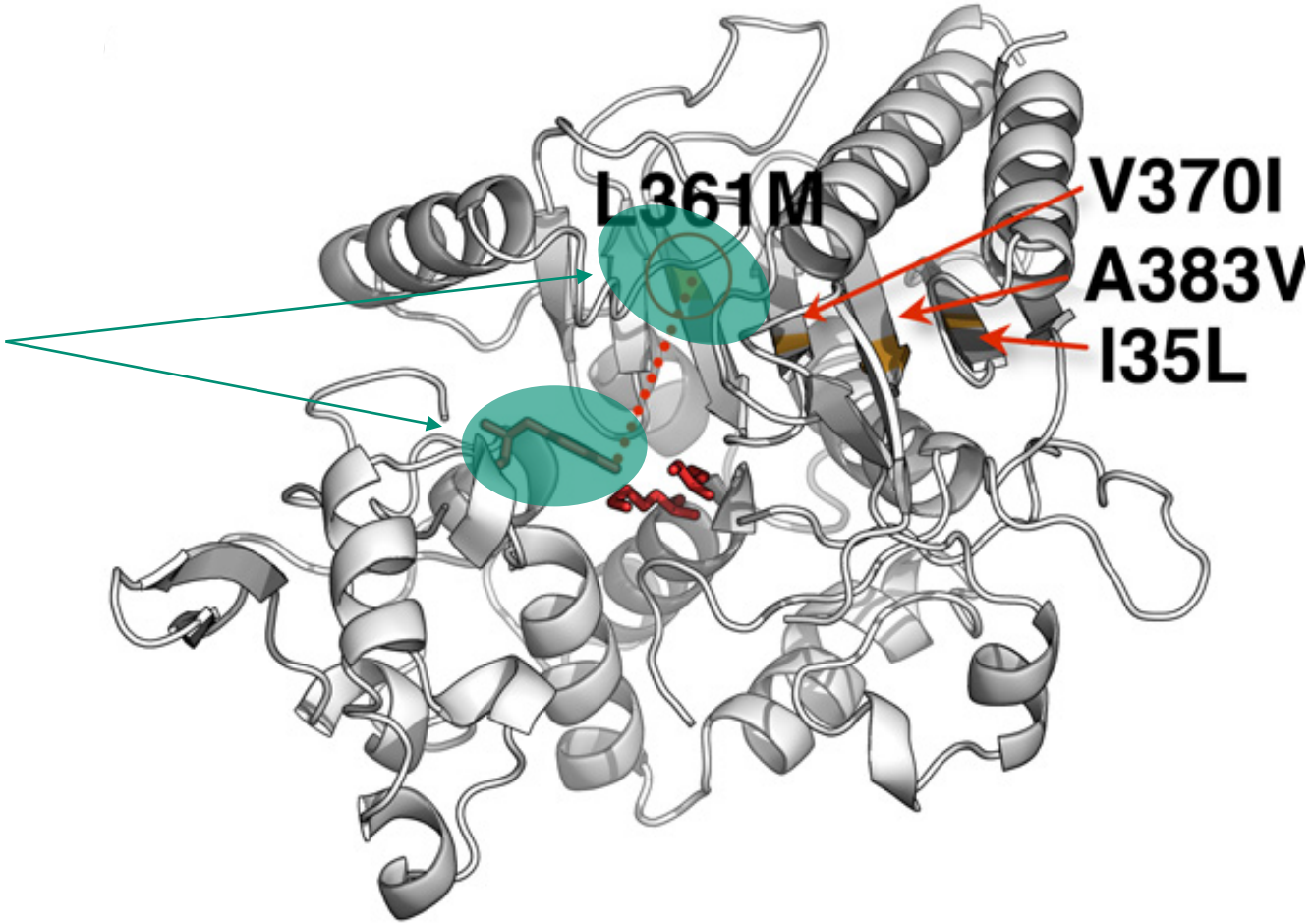
Machine Learning Force Fields

Oliver T. Unke,[△] Stefan Chmiela,[△] Huziel E. Sauceda, Michael Gastegger, Igor Poltavsky, Kristof T. Schütt, Alexandre Tkatchenko,[*] and Klaus-Robert Müller[*]



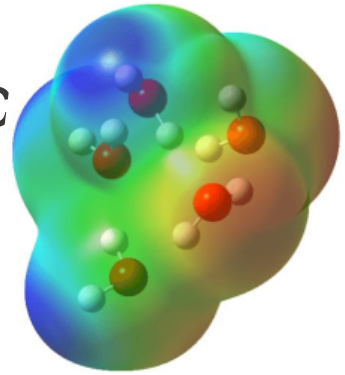
Solution: Learn the **electron density**

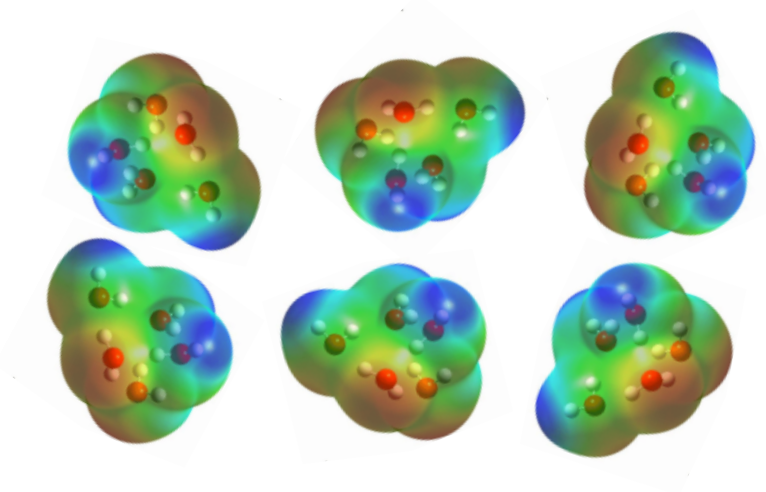
Electron Densities can interact!



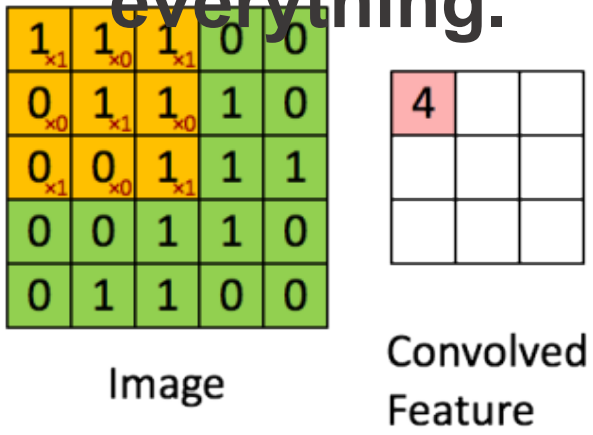
What's the Problem? – Part 2: Learning in 3D Space

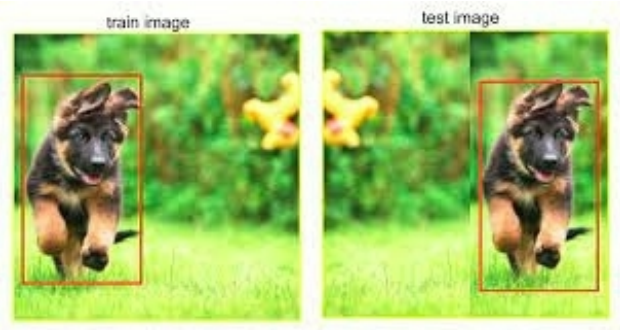
The electron density is a 3D objec



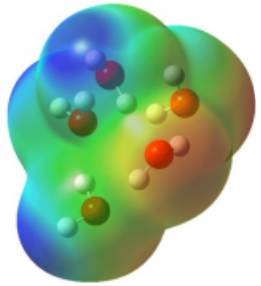


Convolution changed **everything.**

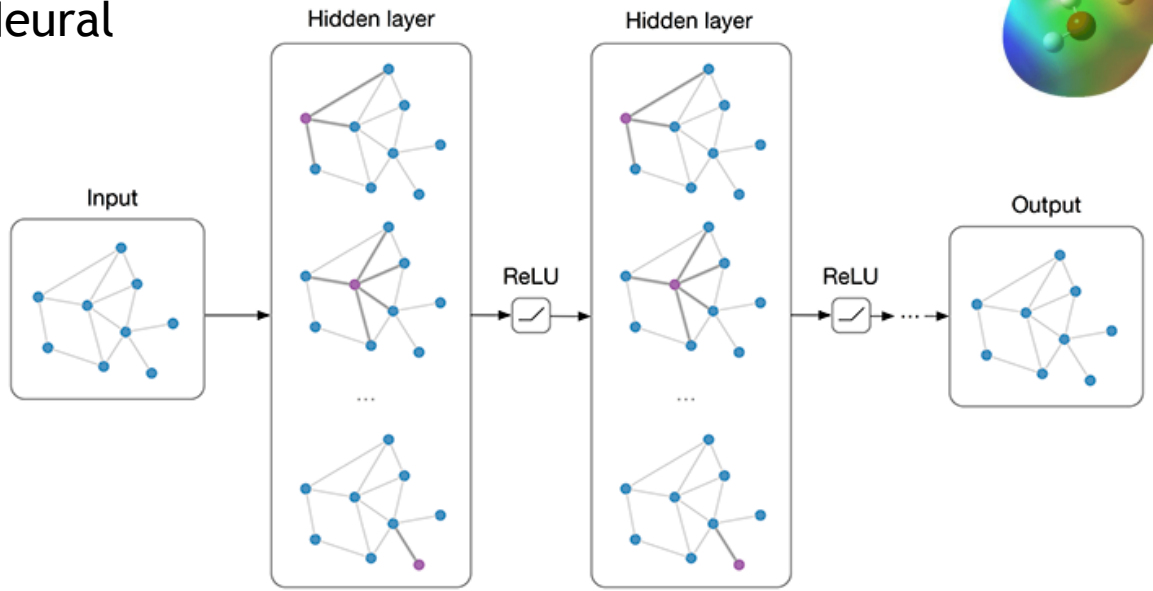




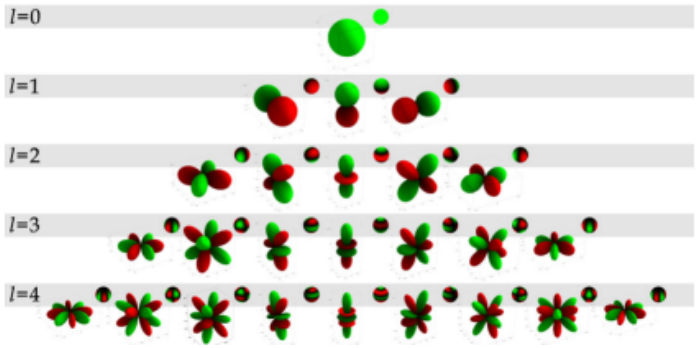
Solution: Euclidean Neural Networks



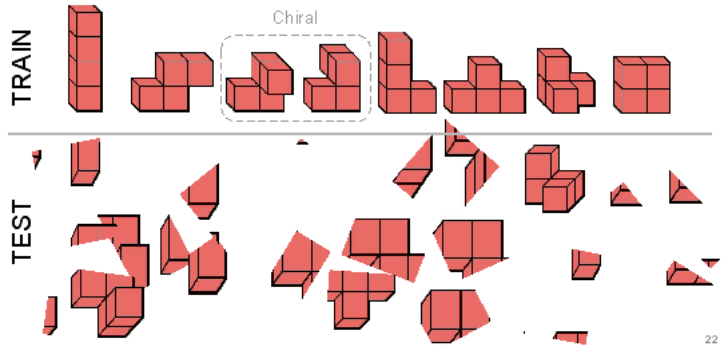
Graph Convolutional Neural Network:



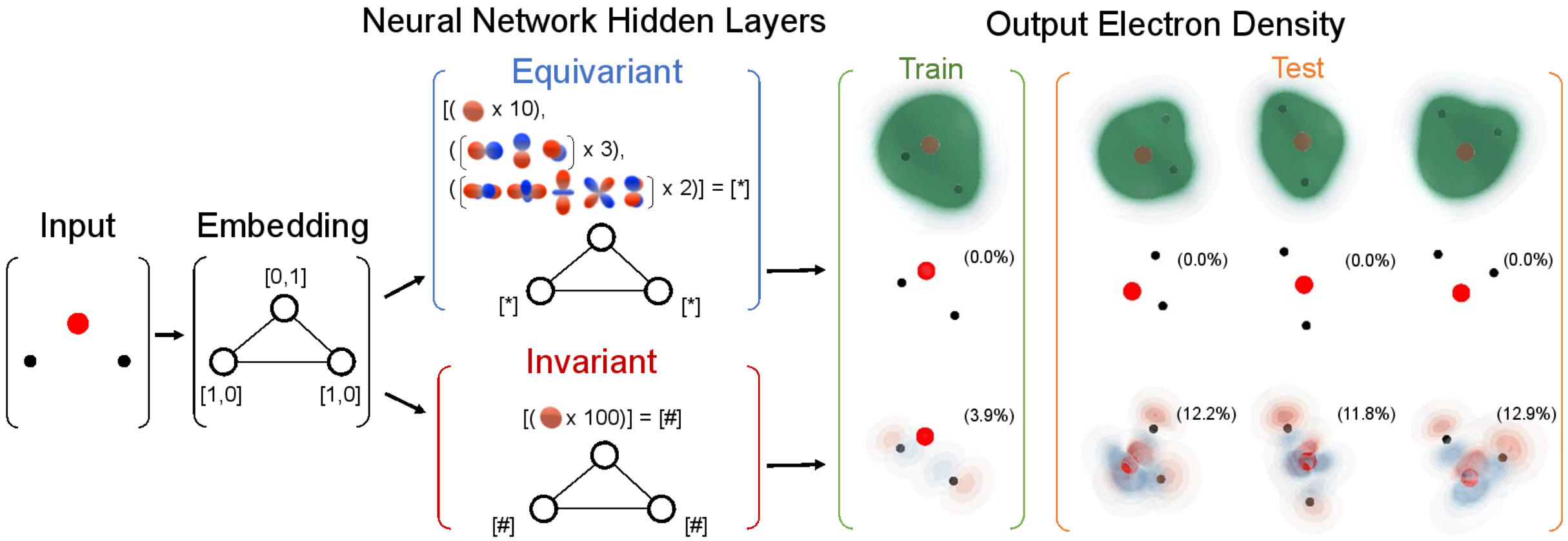
e3nn features are combinations of spherical harmonics



e3nn is equivariant

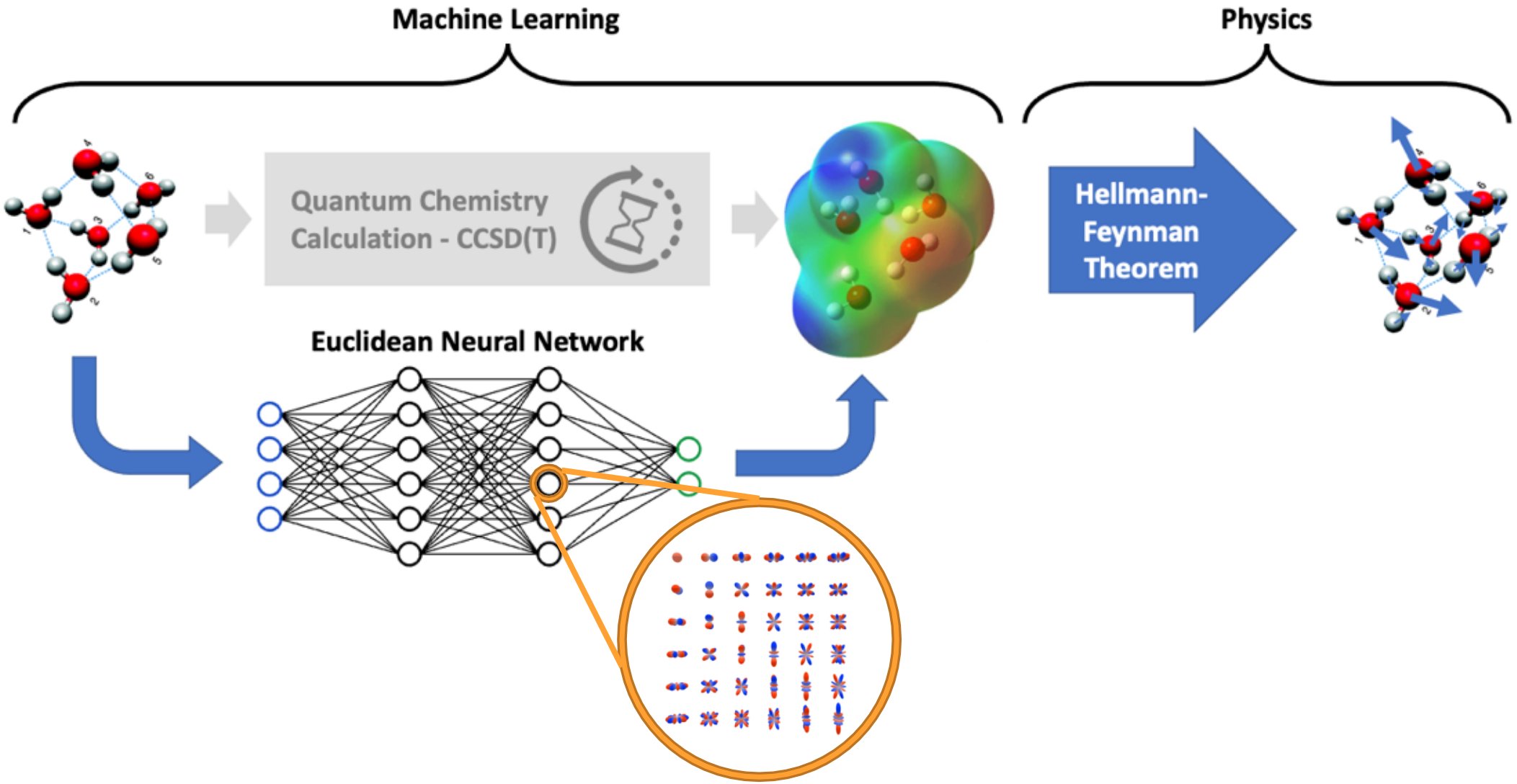


Solution: Euclidean Neural Networks



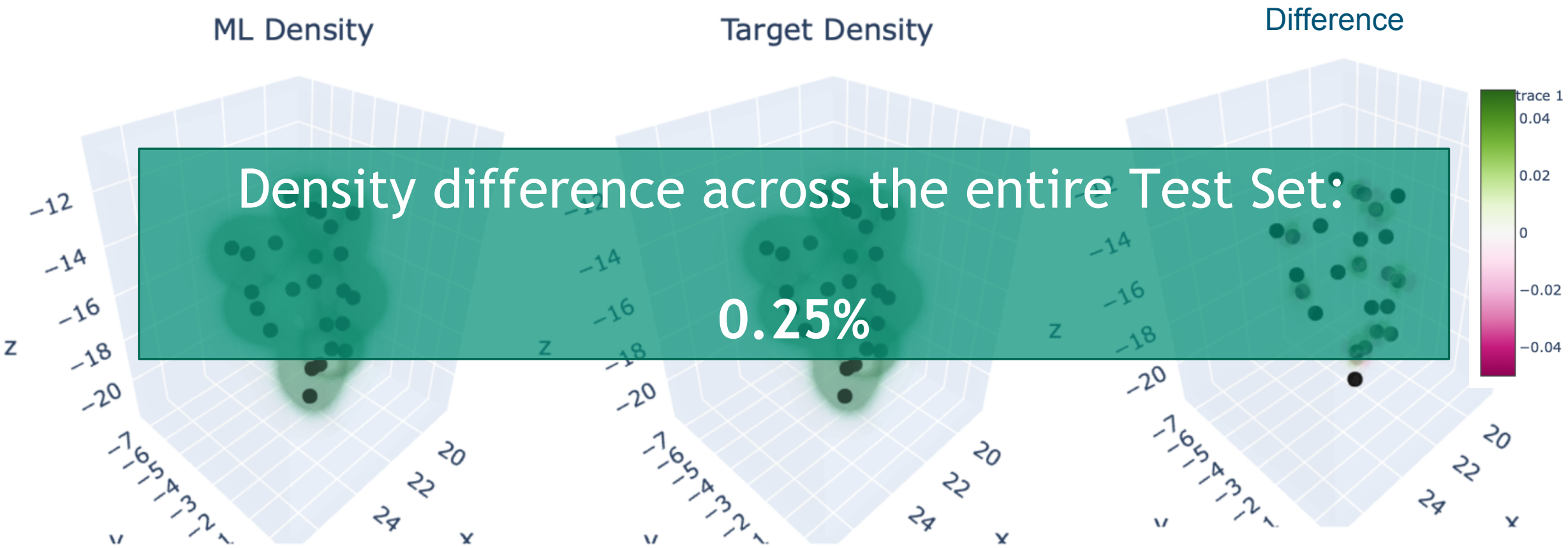
Rackers, Tecot, Geiger, Smidt, *arXiv*, 2022

The Big Idea: **Learn the electron density with Euclidean Neural Networks**

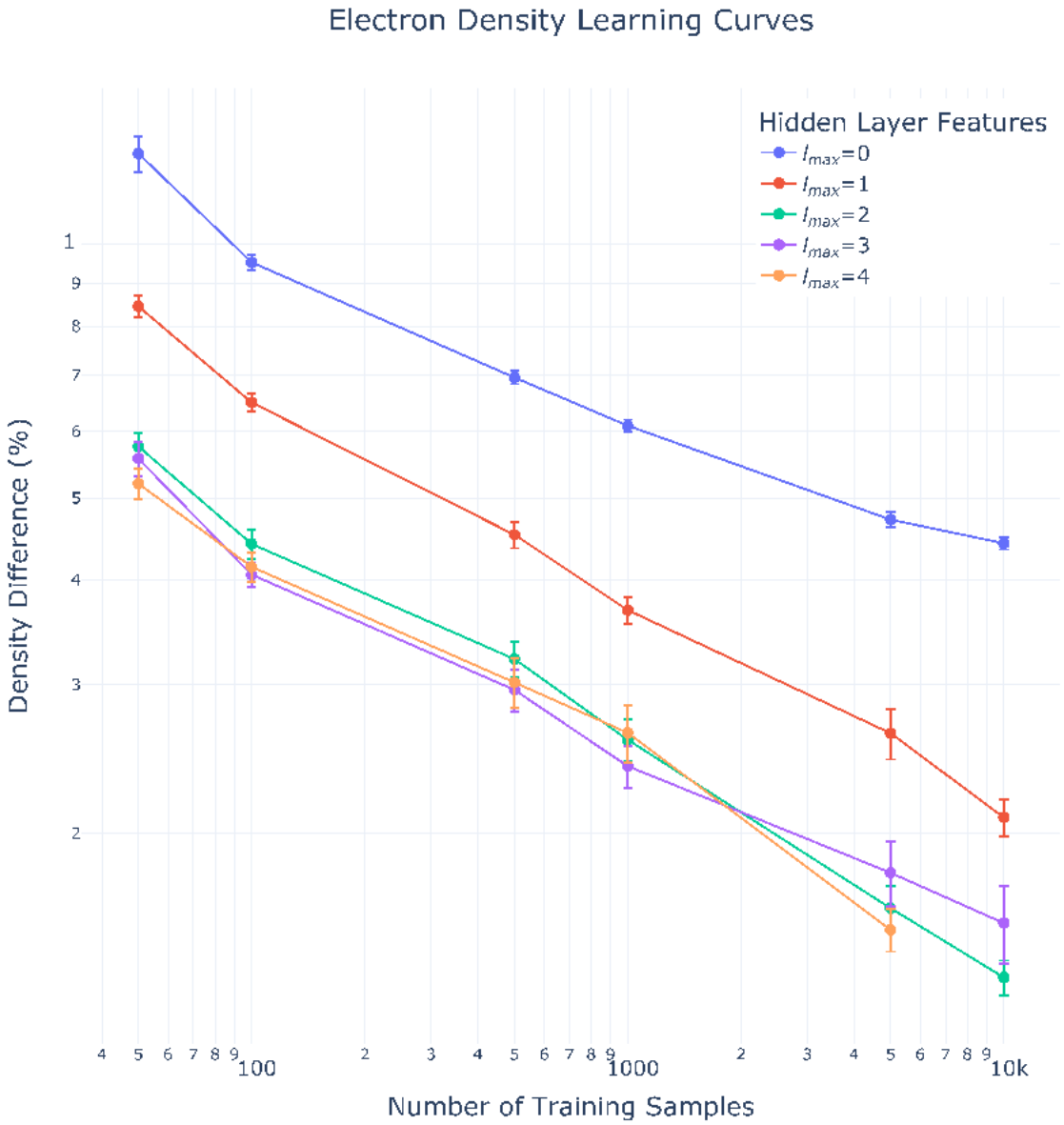


Experiment 1: Learning the e^- density

Train on model on 500 samples of 8 water molecule clusters (DFT) (For reference, ImageNet has 14 million images.)



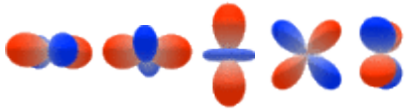
An aside: The importance of equivariance



l_{max}=0 : x 500

l_{max}=1 : x 250 + x 83

l_{max}=2 : x 167 + x 56 +



x 33

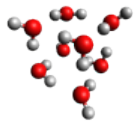
⋮

(Total number of 500 features in every model)

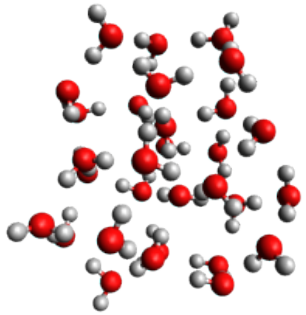
Experiment 2: Extrapolation

Train on 8 water clusters. Test on 30 water clusters.

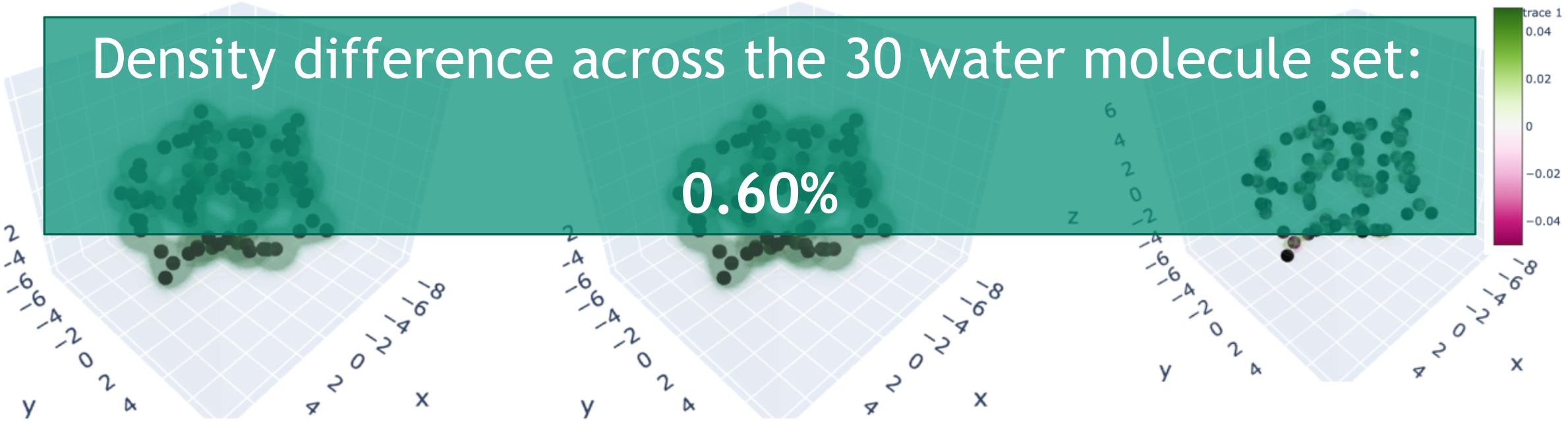
Train



Test



ML Density | Target Density | Difference



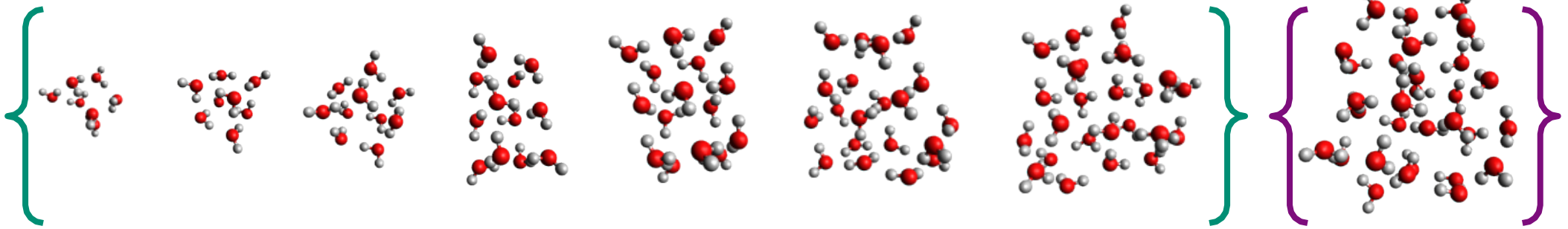
Density difference across the 30 water molecule set:

0.60%

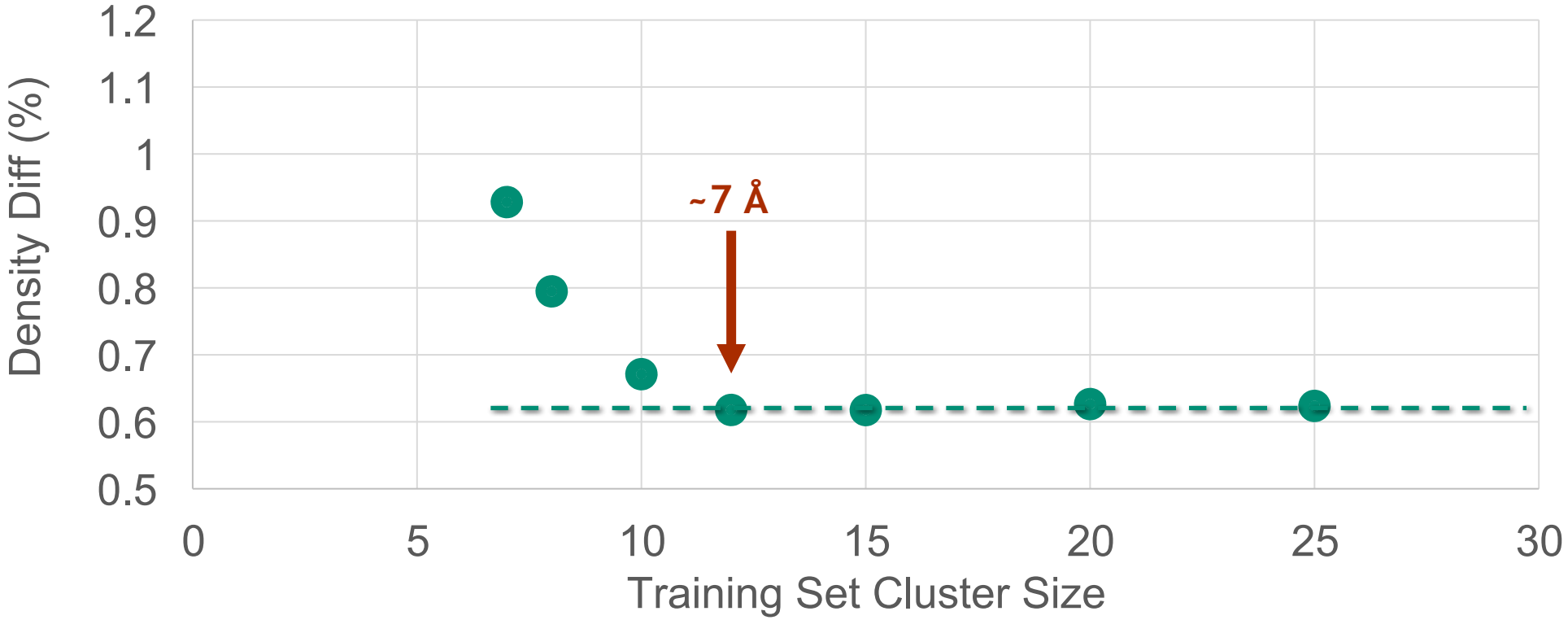
Experiment 3: Systematic Extrapolation

Train

Test



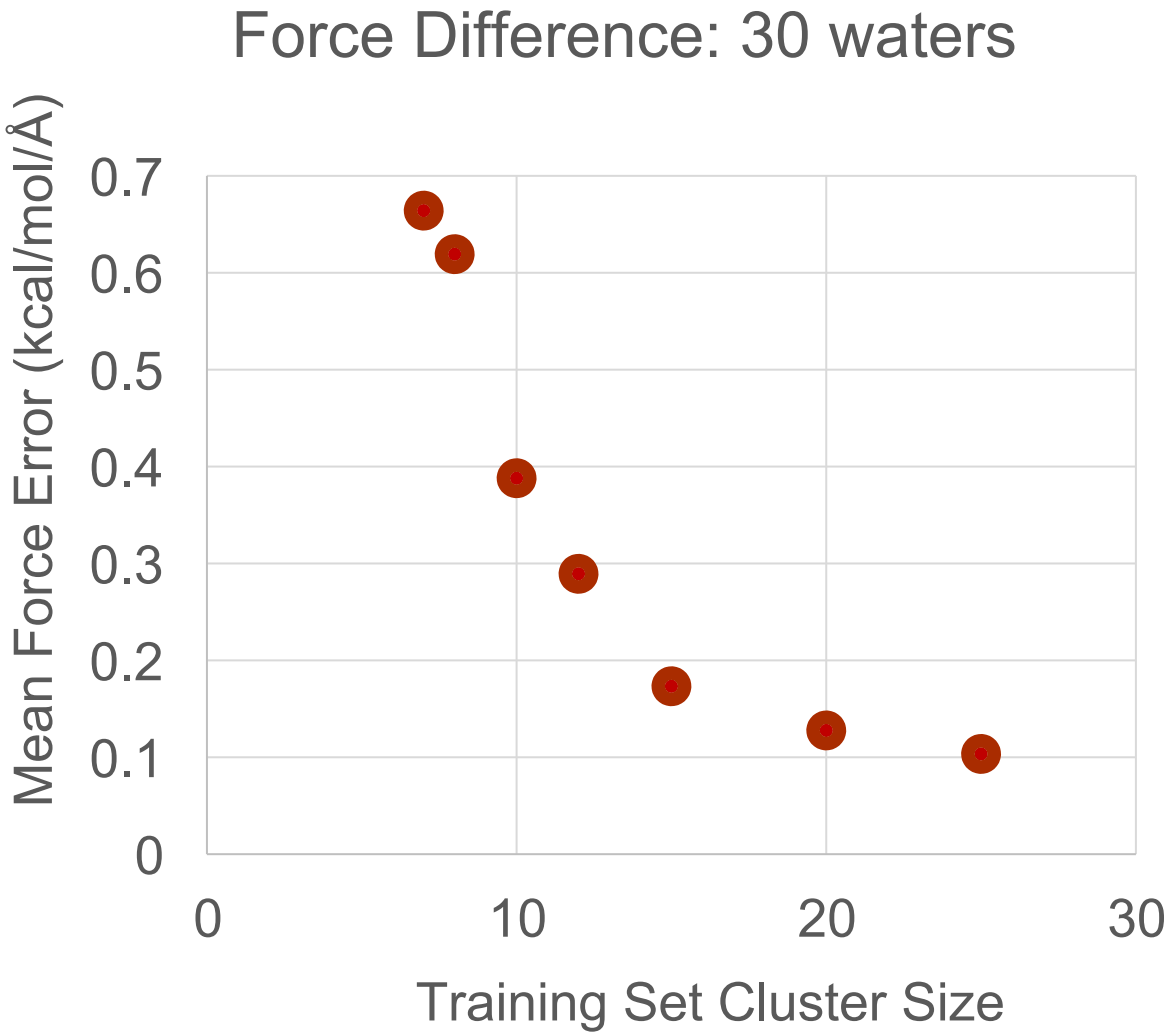
Density Difference: 30 water molecule test set

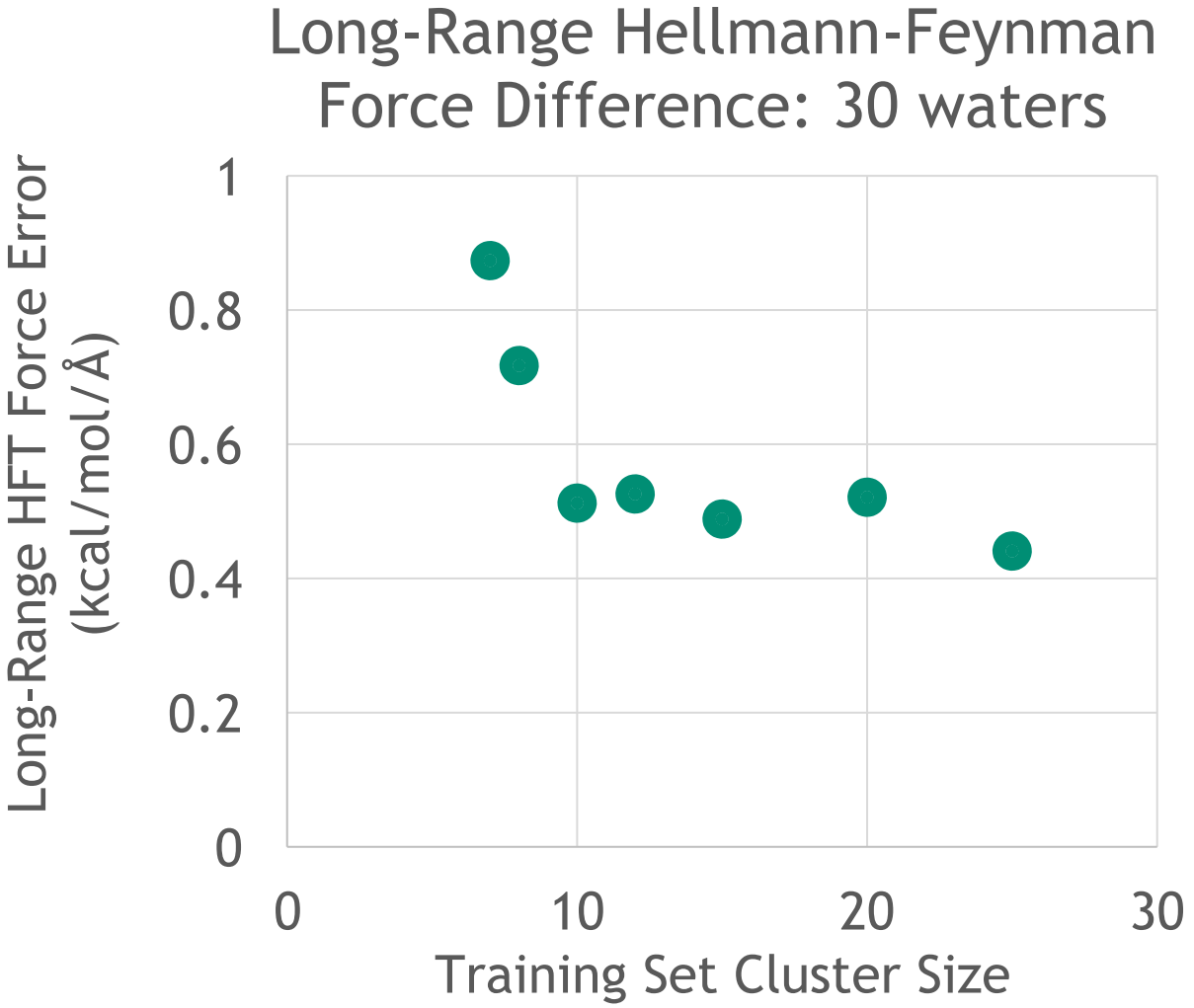


Rackers, Tecot, Geiger, Smidt, *arXiv*, 2022

Experiment 4: Comparison with forces

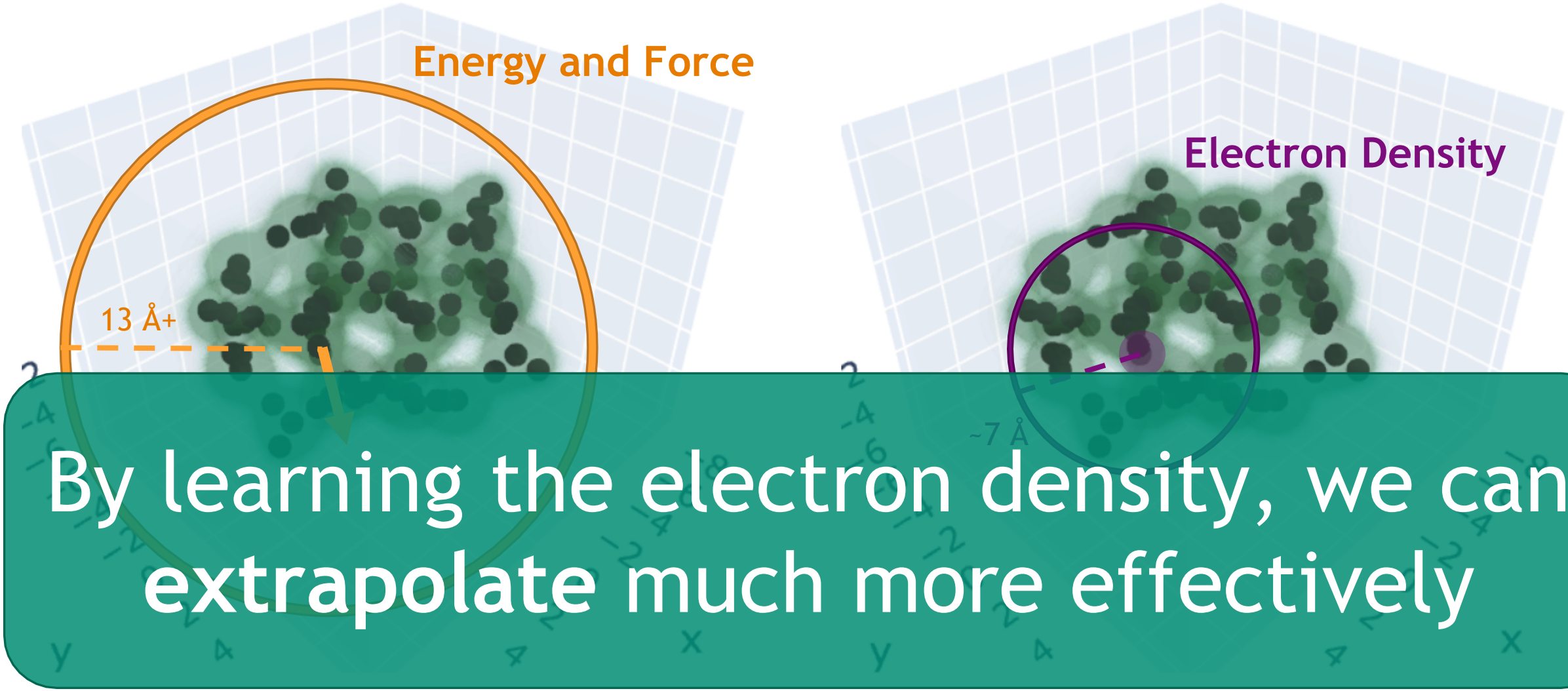
(ML models from previous slide)





Rackers, Tecot, Geiger, Smidt, *arXiv*, 2022

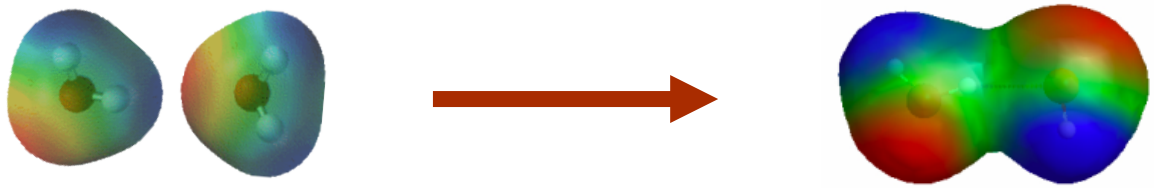
Long-range interactions affect forces more than density



By learning the electron density, we can **extrapolate** much more effectively

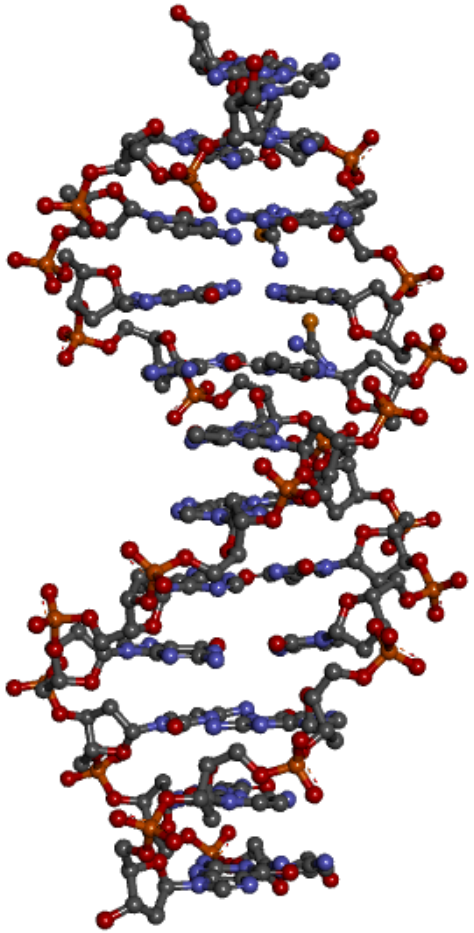
Long-range interactions affect forces more than density

This is consistent with our understanding of perturbation theory.

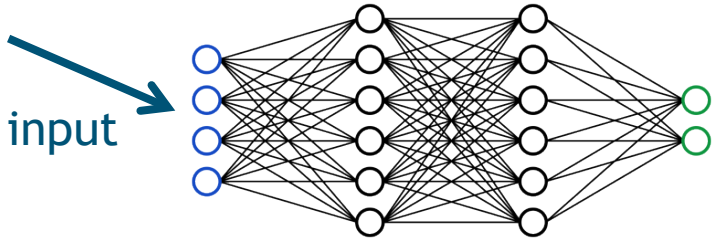


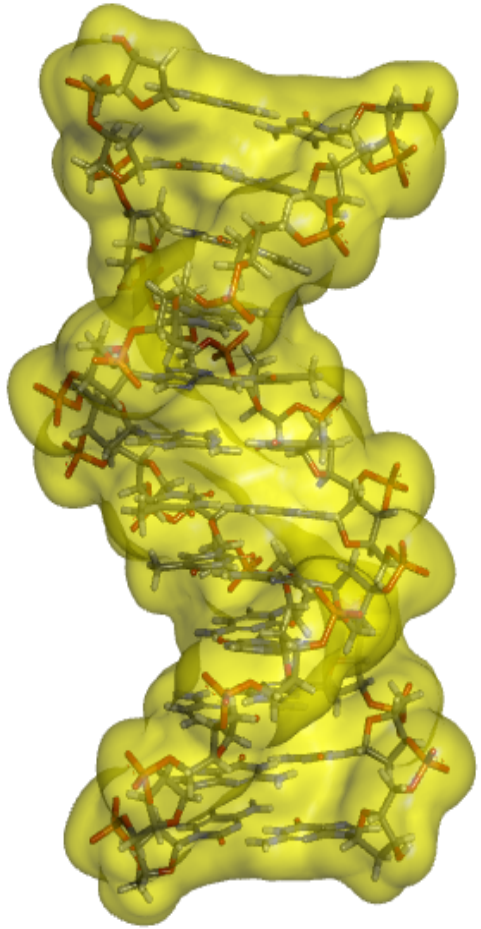
$$E_{PT} = E^{(0)} + \underbrace{E^{(1)}_{elst}}_{\text{Electrostatics}} + \underbrace{E^{(1)}_{exch} + E^{(2)} + \cdots}_{\text{Polarization, Exchange-Repulsion, Dispersion}}$$

DNA – ML electron densities at the biomolecular scale



$$\hat{H}\Psi = E\Psi$$

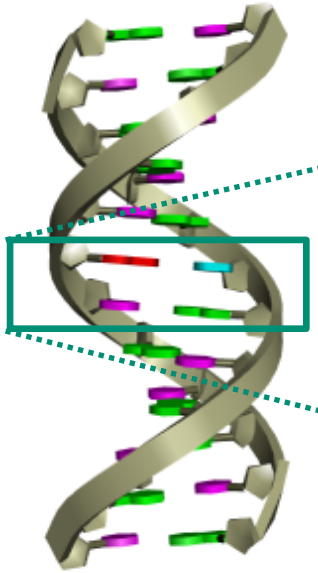


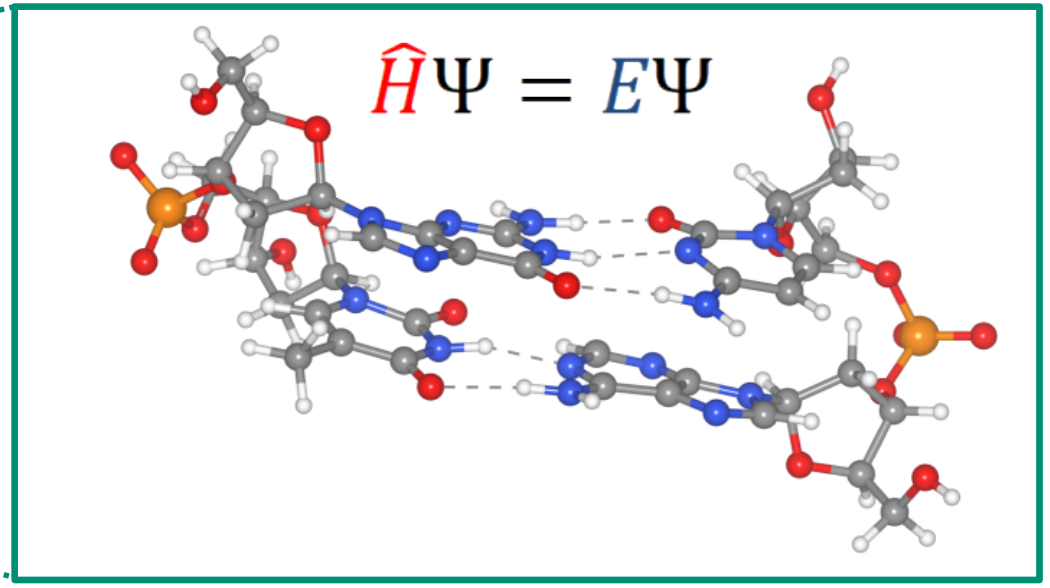


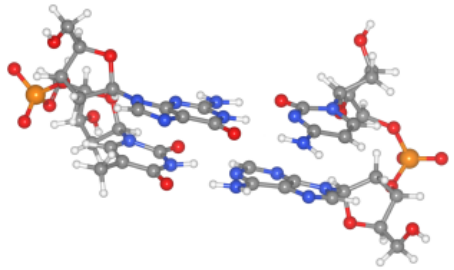
DNA – Learning the electron density

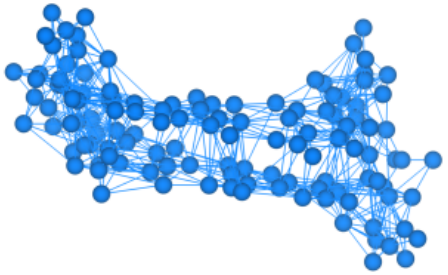
Molecular dynamics

Quantum

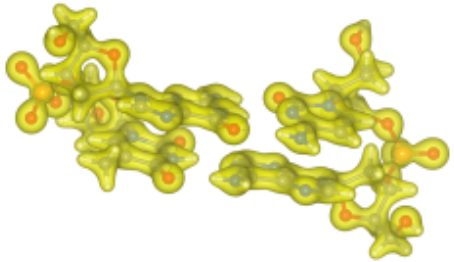








Euclidean Neural Network



Atomic coordinates

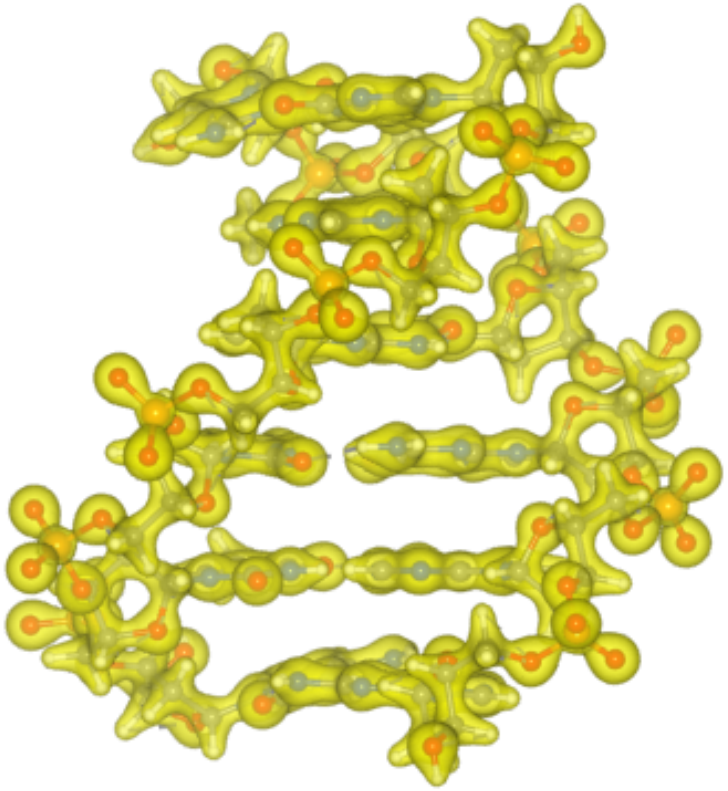
Input layer

Hidden layers

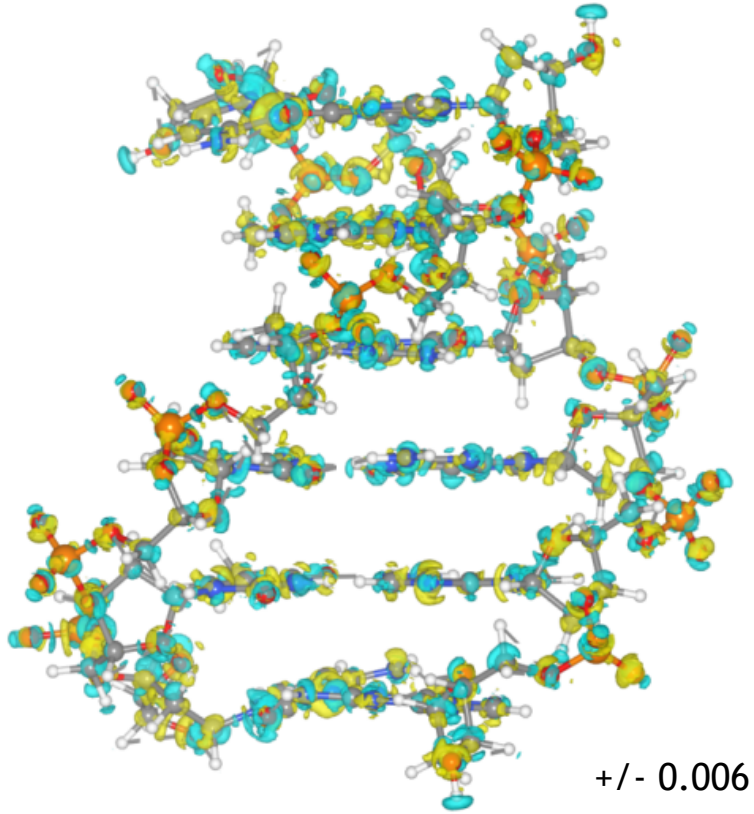
Output layer

A real, experimental structure! (PDB code 251d)

ML prediction



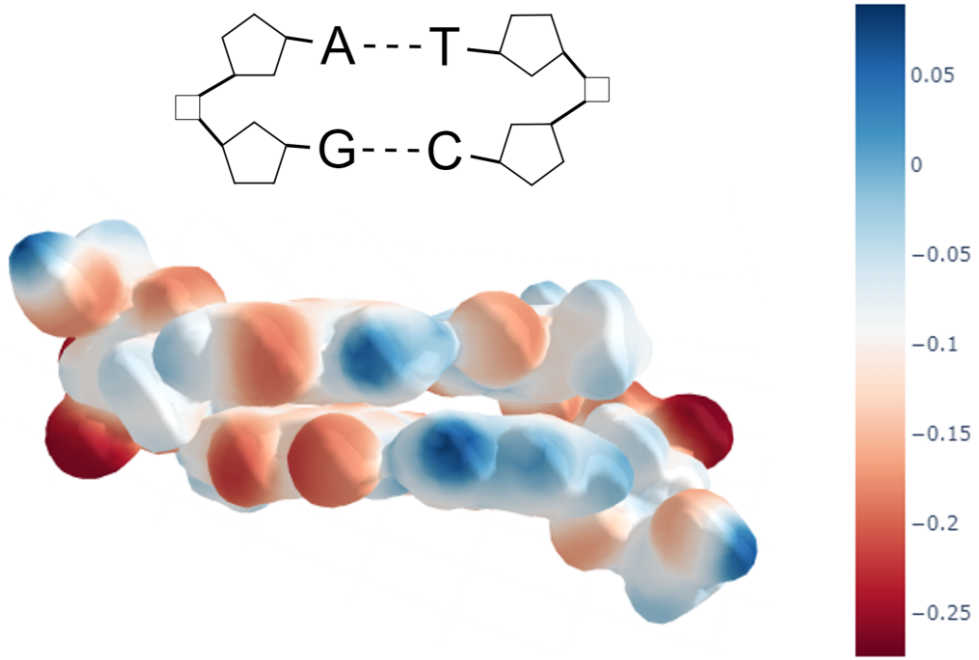
Density difference: ML - target

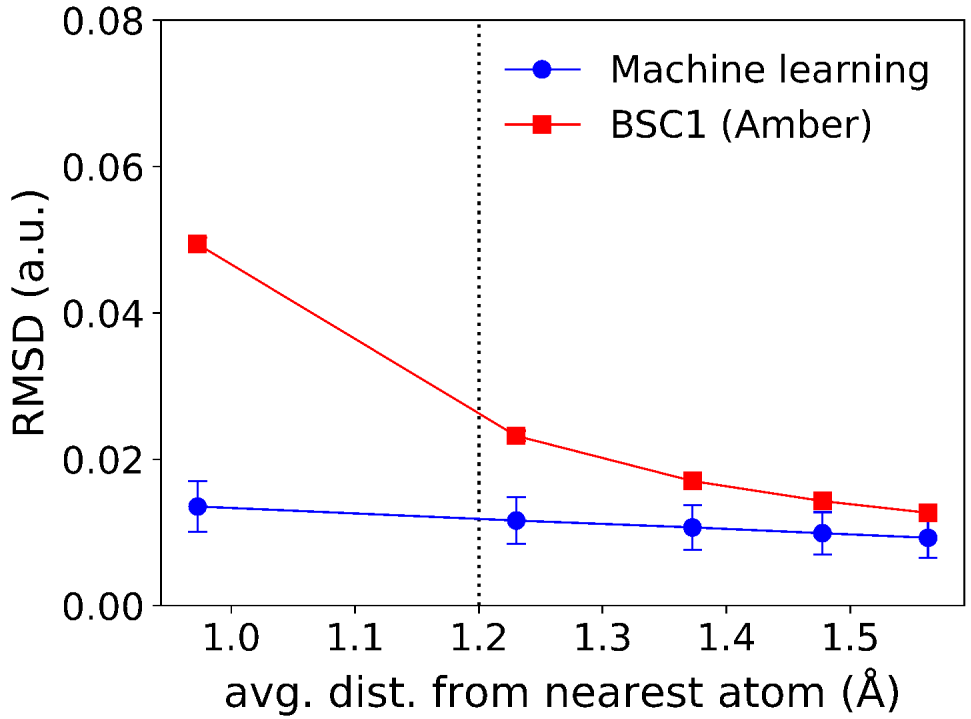


+/- 0.006 $e^-/bohr^3$

Lee, Rackers & Bricker, *ChemrXiv*, 2022

Electrostatic Potential

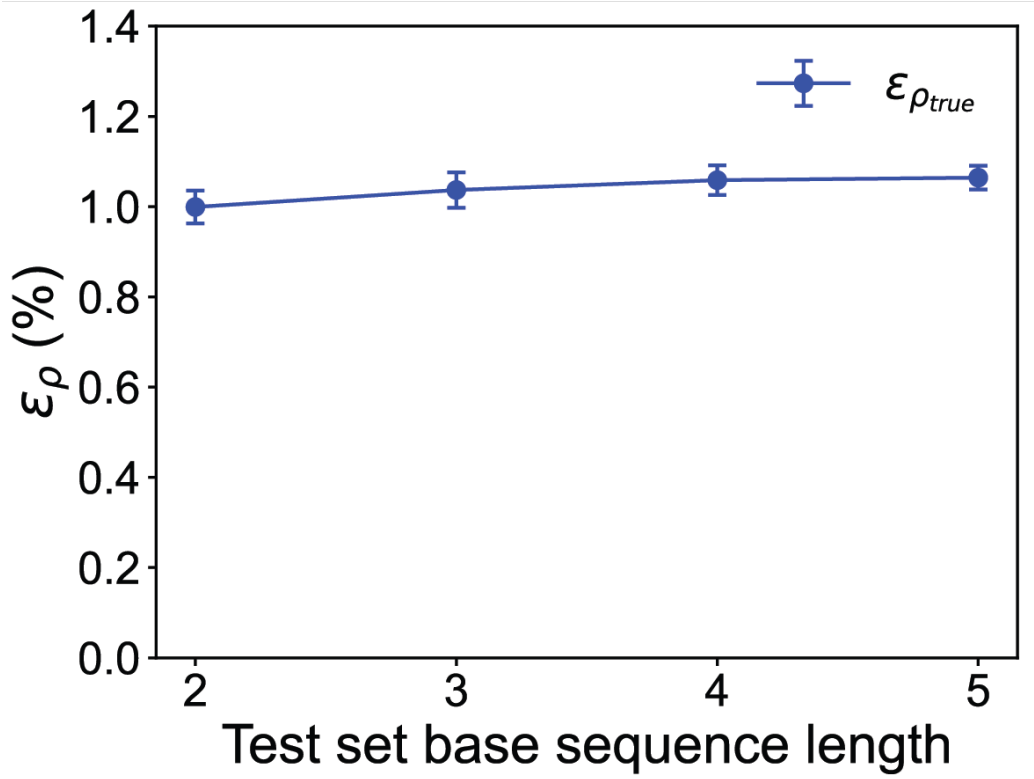




Lee, Rackers & Bricker, *ChemrXiv*, 2022

Systematic Extrapolation

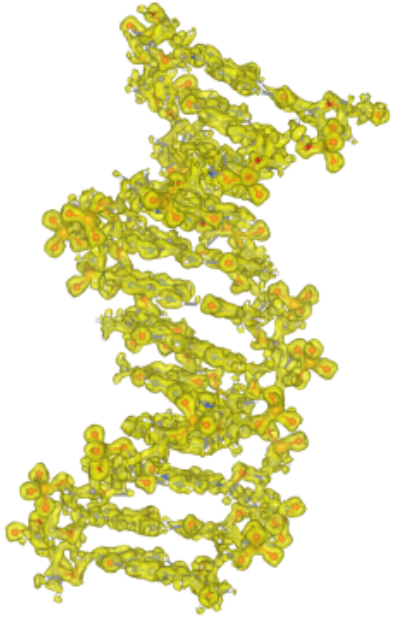
Sequence Length Size Dependence



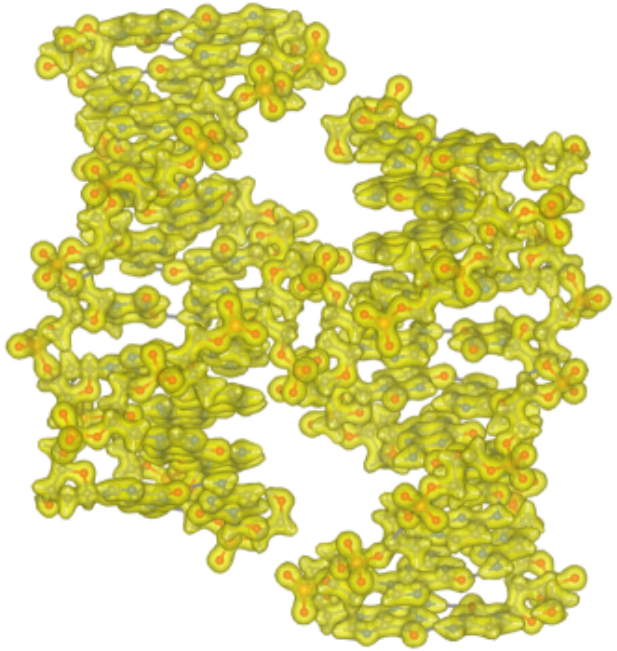
Lee, Rackers & Bricker, *ChemrXiv*, 2022

We can predict very large structures!

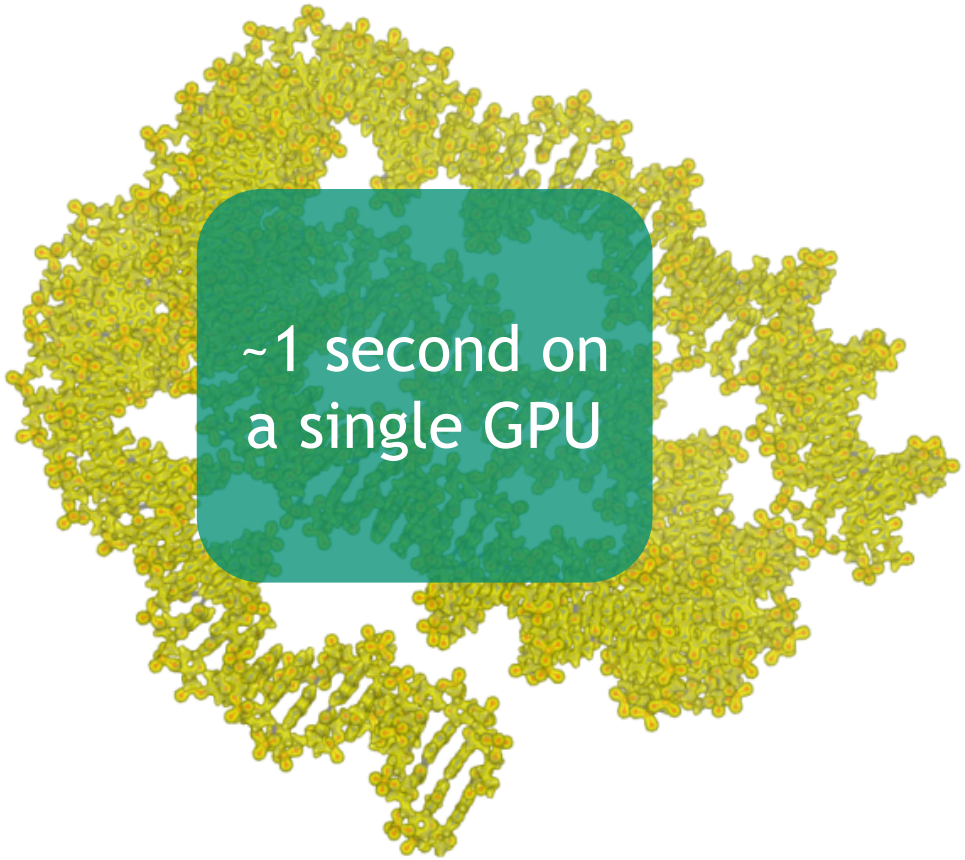
Drew-Dickerson dodecamer
758 atoms, 3780 electrons
(PDB code 4c64)



Stacked Holliday junction
1260 atoms, 6280 electrons
(PDB code 1dcw)



Nucleosome core particle
147 bps, 9346 atoms, 46980 electrons
(PDB code 1kx5)



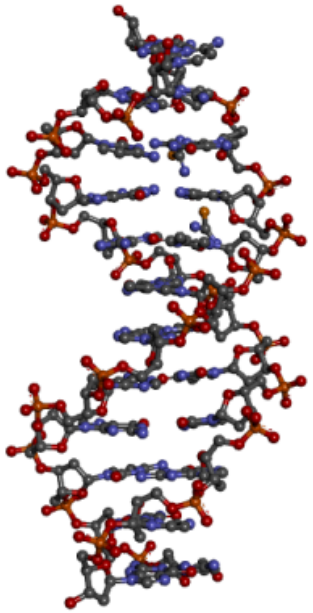
Lee, Rackers & Bricker, *ChemrXiv*, 2022

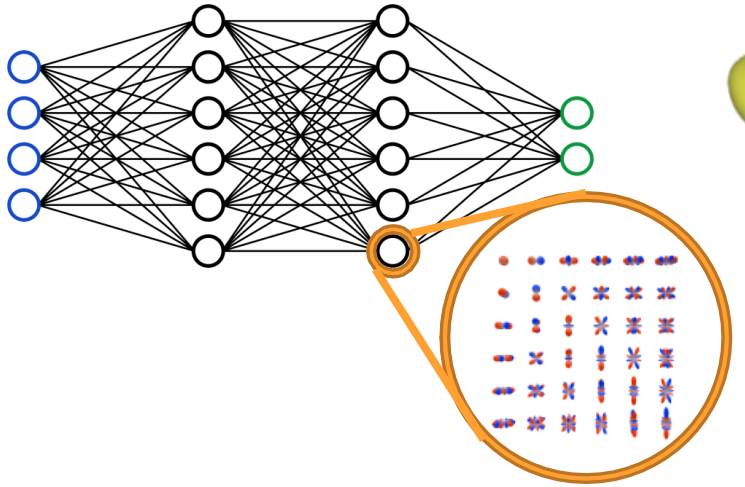
Take-Aways

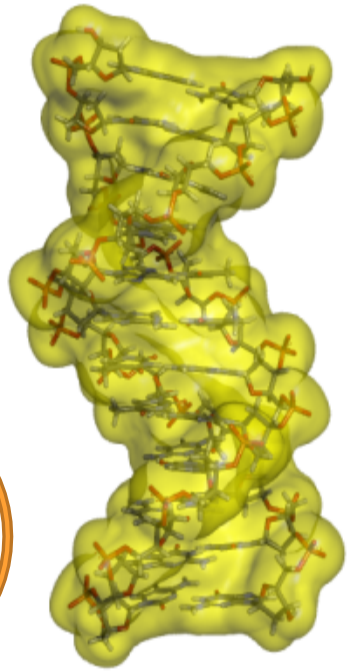
A. Any serious model has to be able to "train on small, run on big".

B. Euclidean Neural Networks + physics is the tool for the job

C. This w[illegible]ater and DNA!







Acknowledgements

Steve Plimpton

Aidan Thompson

Susan Rempe

Lucas Tecot
(UCLA)

Tess Smidt
(MIT)

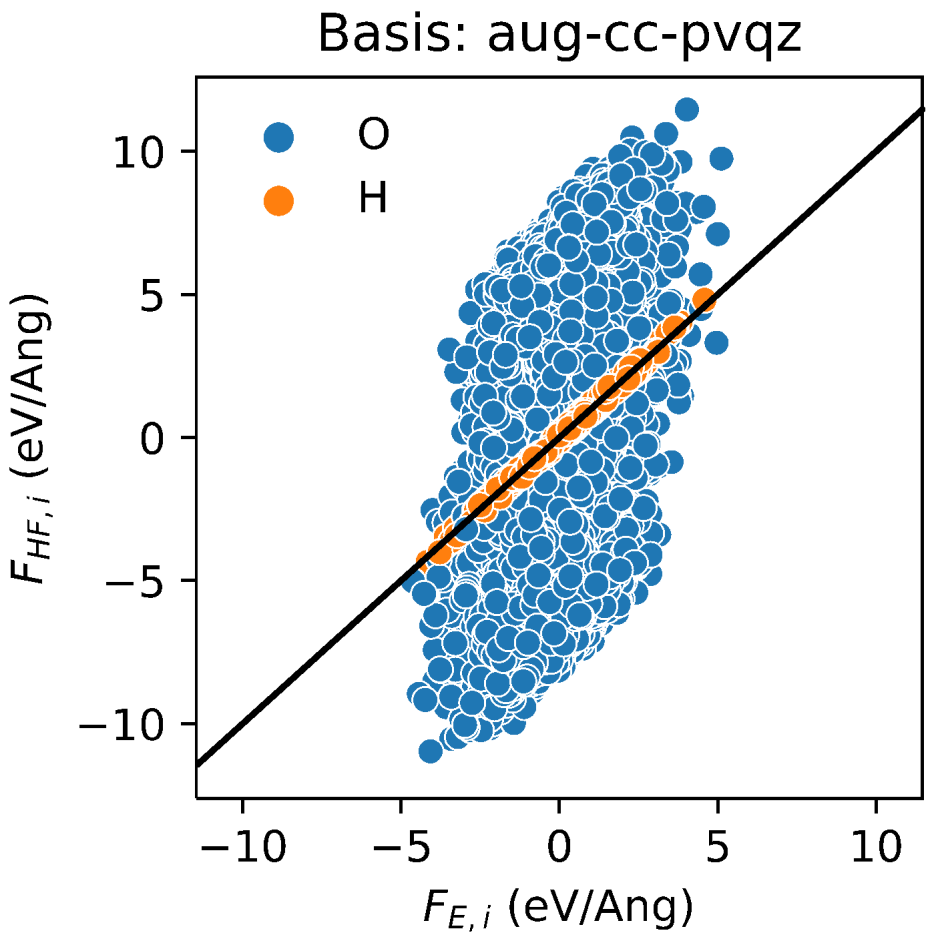
Shivesh Pathak
(Sandia)

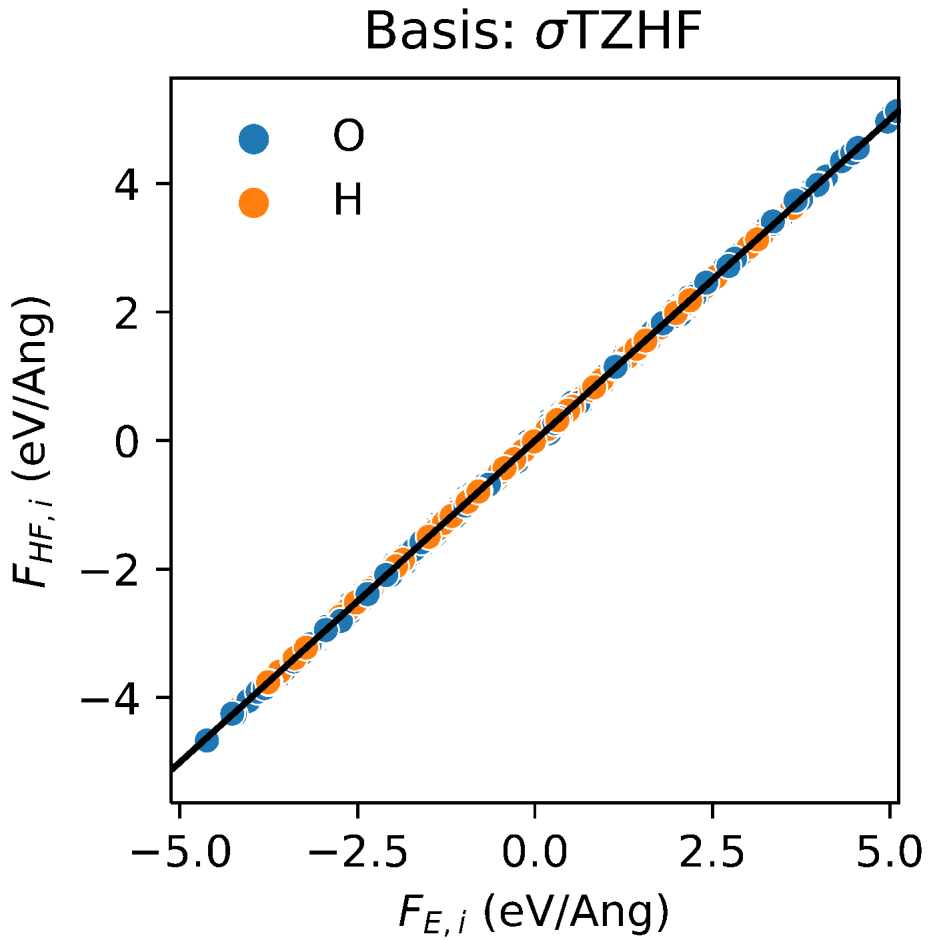
Will Bricker and Alex Lee
(UNM)

Preside
Nationa

LDRD LABORATORY DIRECTED RESEARCH & DEVELOPMENT

Hellmann-Feynman Forces





Solution: Learn the **electron density**

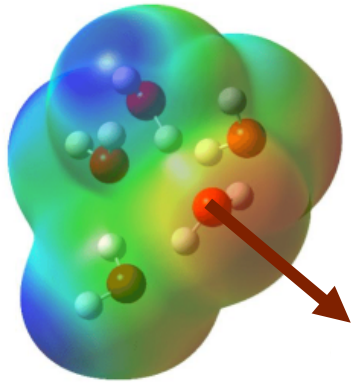
The Hellmann-Feynman Theorem:

The force on any atom can be exactly calculated from the surrounding electron density **using only classical electrostatics.**

Machine Learning

$$H\Psi = E\Psi$$

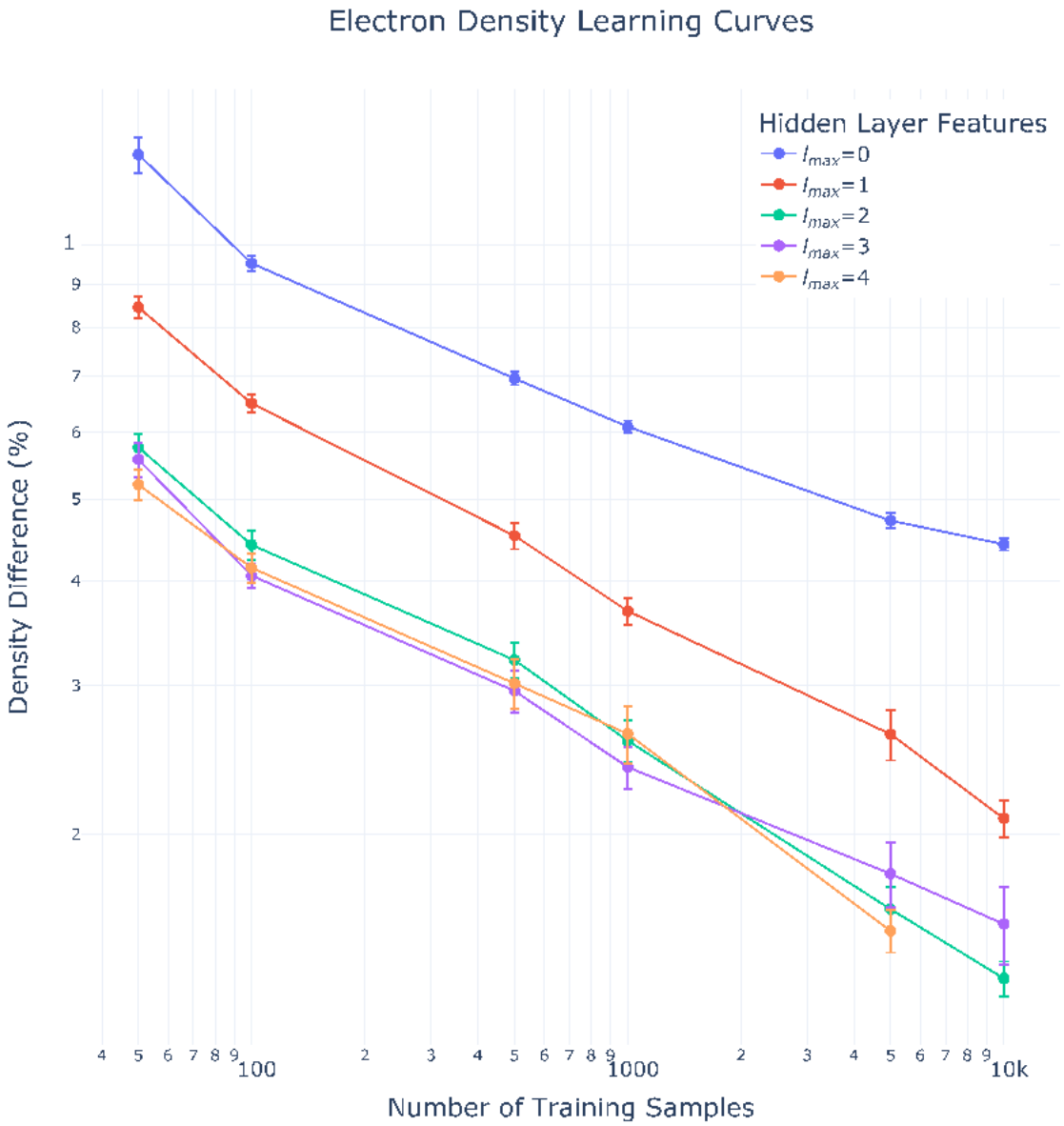
$$\rho = \Psi^*\Psi$$



Physics

$$\overrightarrow{F_i} = \sum_{j}^{N_{nuclei}} \frac{Z_i Z_j}{R_{ij}^3} - Z_i \int \frac{\rho}{r^3} dv$$

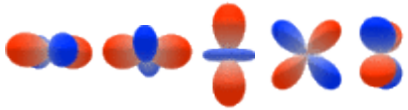
Experiment 2: The importance of higher-order features



$l_{max}=0$: x 500

$l_{max}=1$: x 250 + x 83

$l_{max}=2$: x 167 + x 56 +



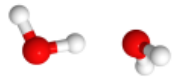
x 33

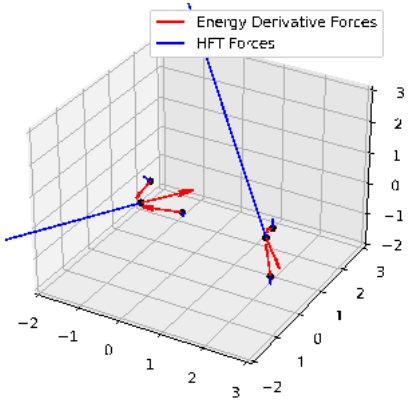
⋮

(Total number of 500 features in every model)

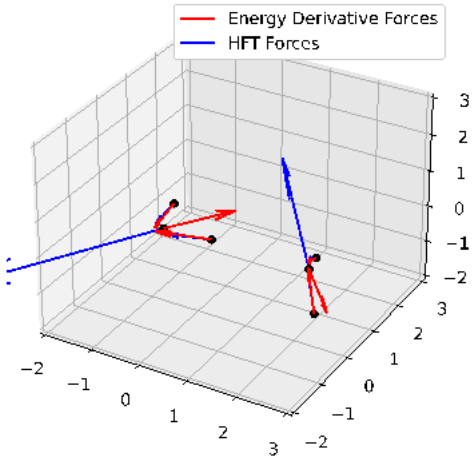
Hellmann-Feynman Forces

aug-cc-pvdz
82 basis functions

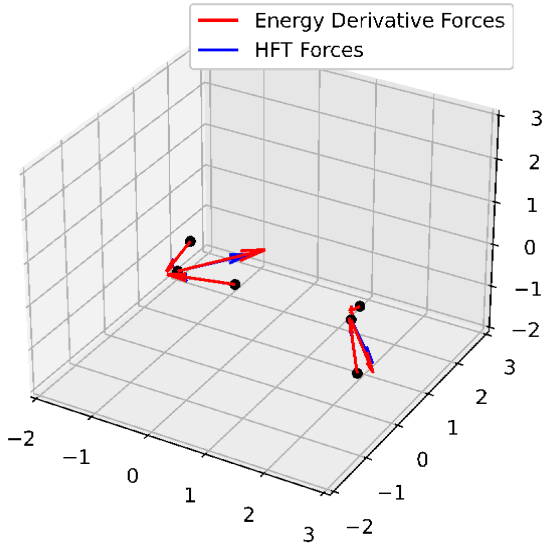




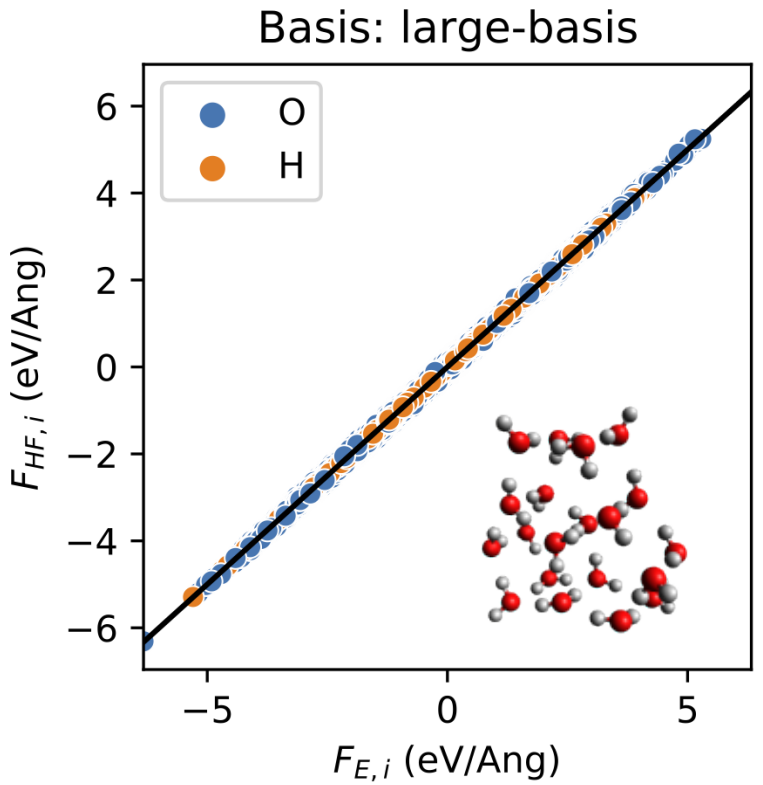
aug-cc-pvqz
344 basis functions



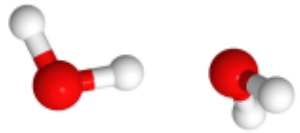
Hellmann-Feynman Basis
154 basis functions



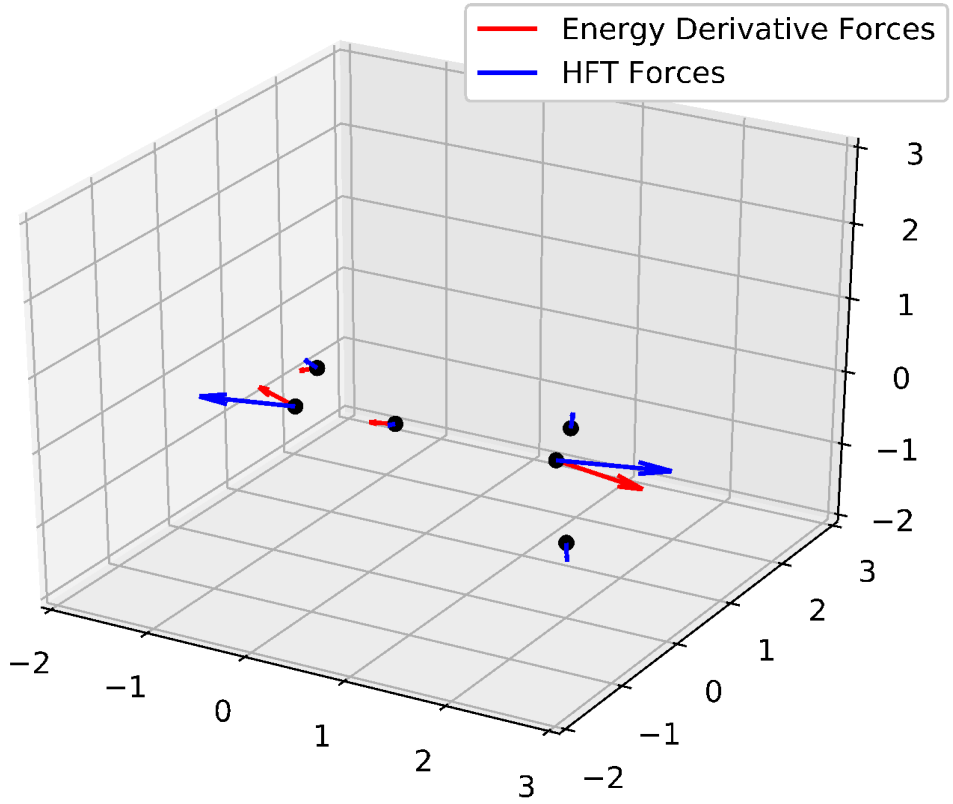
Hellmann-Feynman Force Error: 20 water molecule clusters



HFT forces with CCSD(T) densities



hft-basis-MEDIUM
154 basis functions



hft-basis-LARGE
334 basis functions