Predicting Molecular Electron Densities with Euclidean Neural Networks

MRS Tutorial; November 2021

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.

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

Who am I?

Physics and Political Philosophy

TEACH FOR AMERICA

High School Physics and Chemistry

Washington University in St. Louis

Biophysics Ph.D.

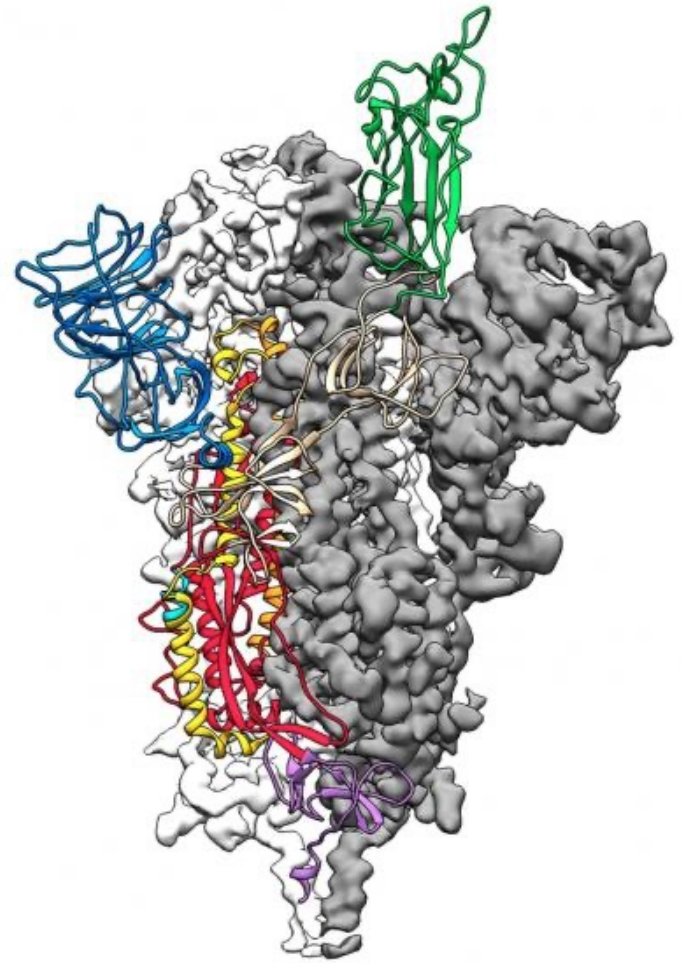
Truman Fellowship

The Grand Challenge of Molecular Science

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

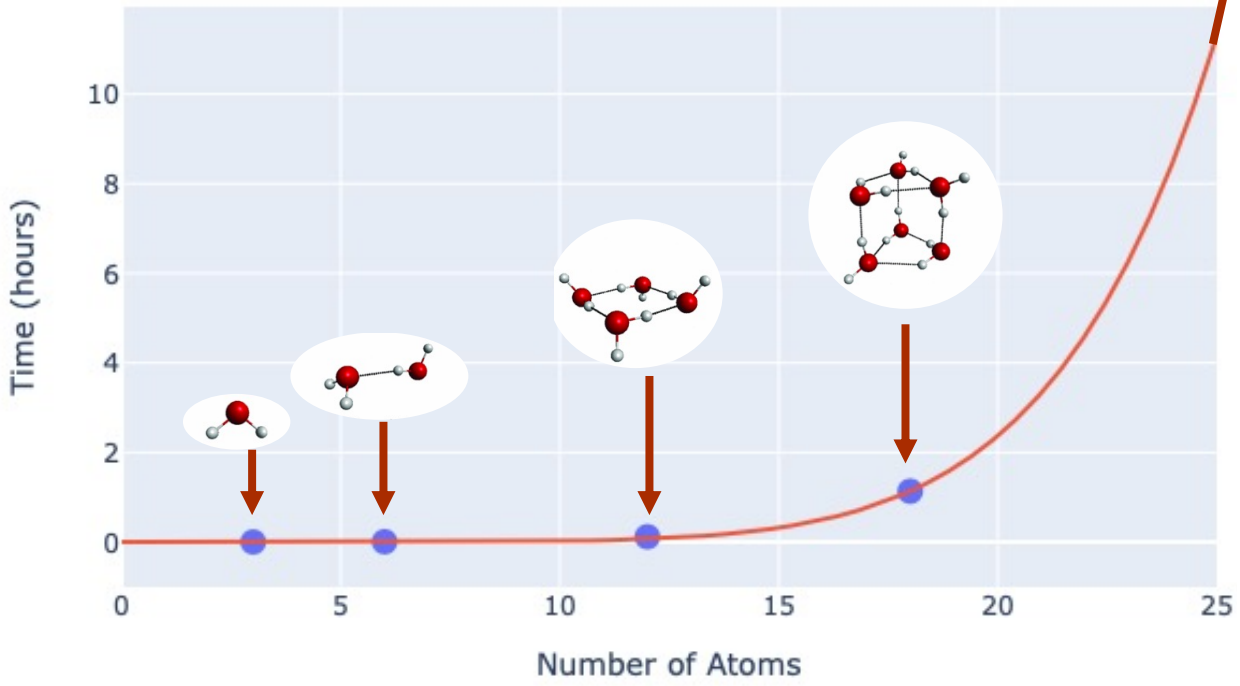
$$F = ma$$

$$H\Psi = E\Psi$$



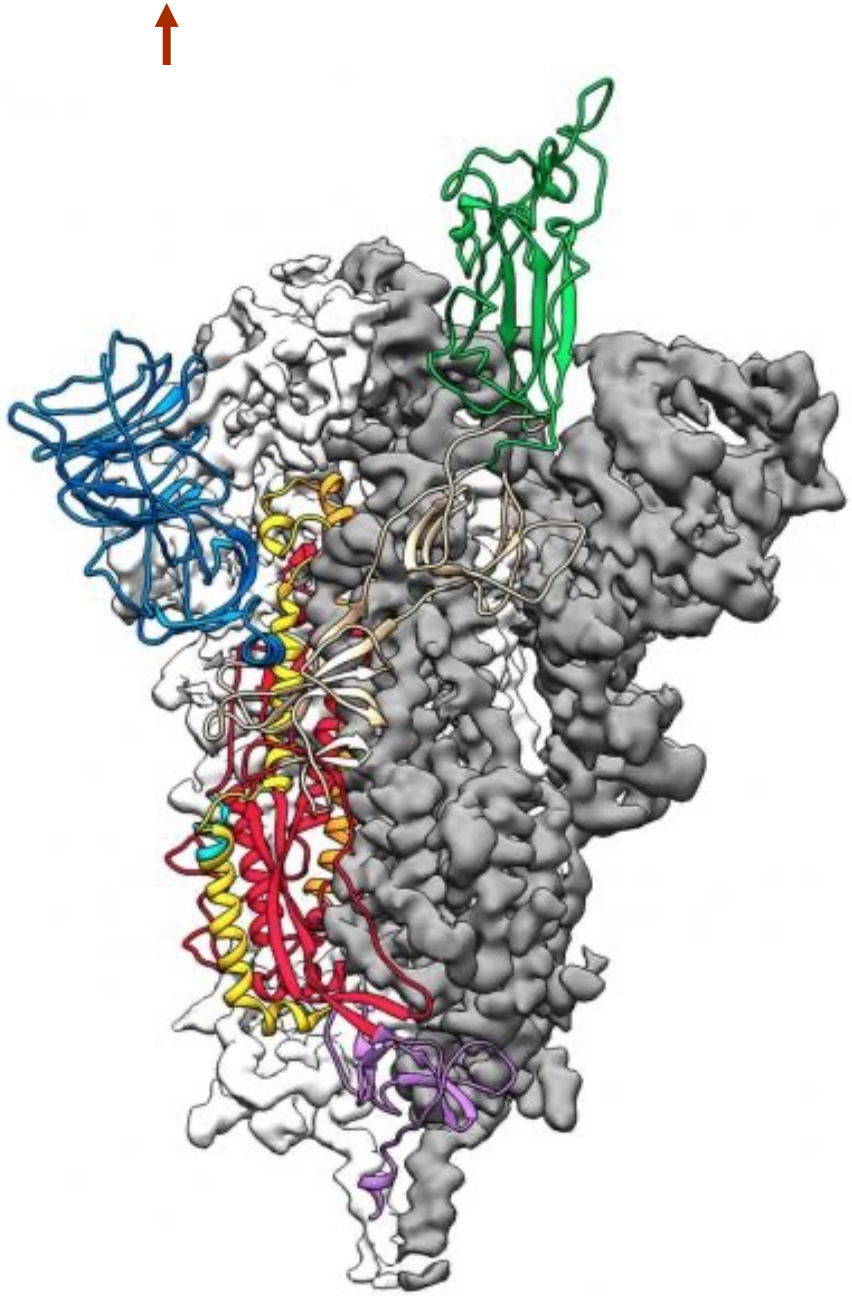
Why is this a hard problem?

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



For **50 atoms**, this

For **100 atom**



Why can't we just do Quantum Mech

No current, conventional Quantu
capable of simulating a large bi

My Laptop

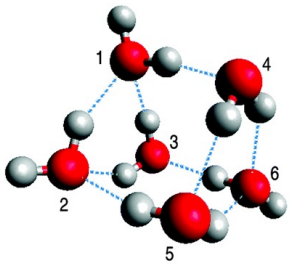
•

Summit Supercomputer

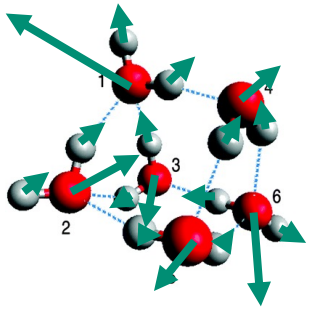
100 atoms? → **6 minutes**

100,000 atoms? → **42 Ages of the Universe**

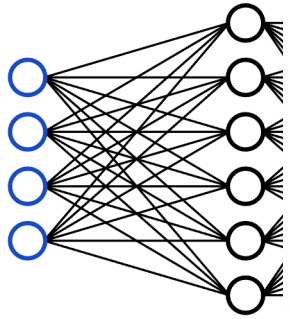
Let's use Machine Learning!

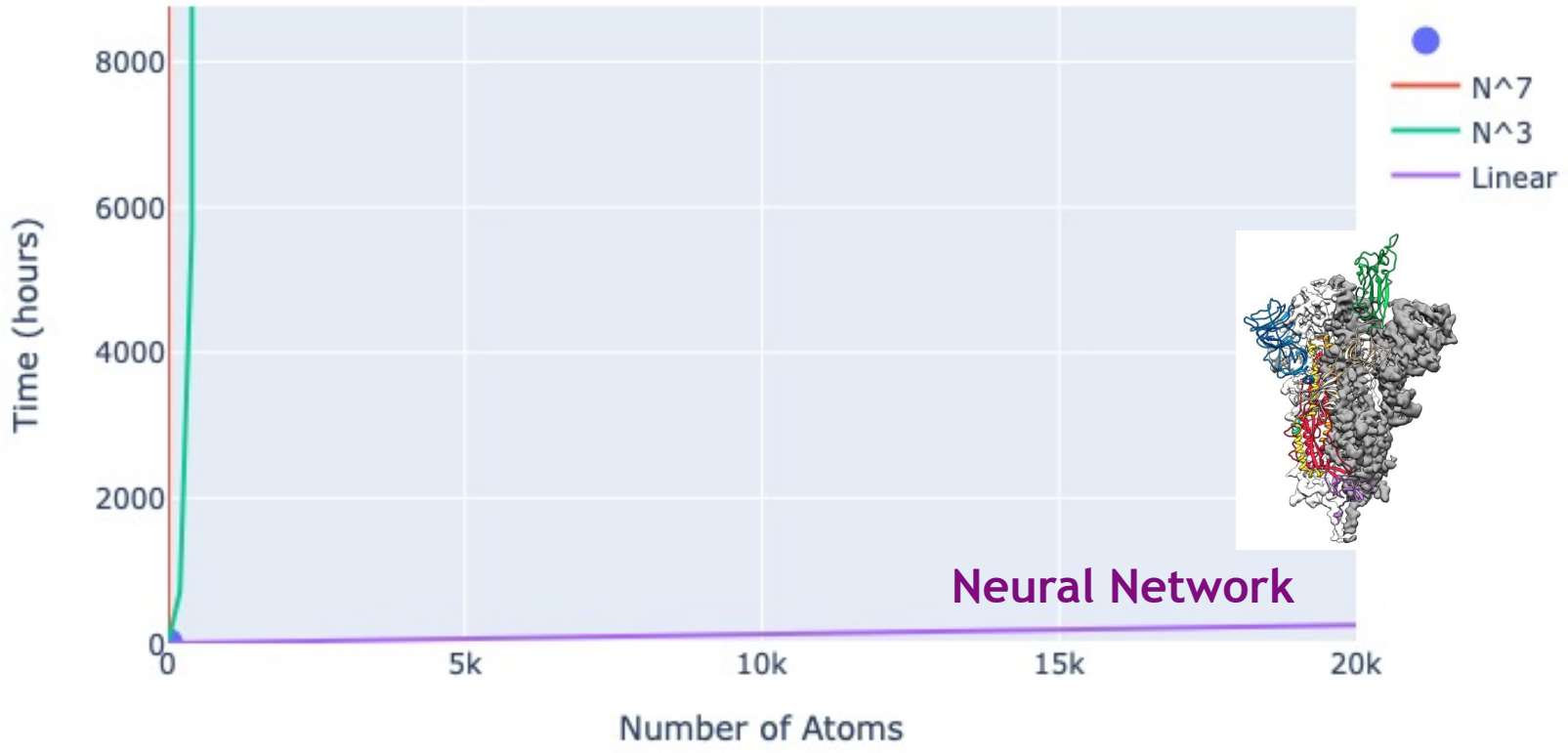


Quantum Mechanics



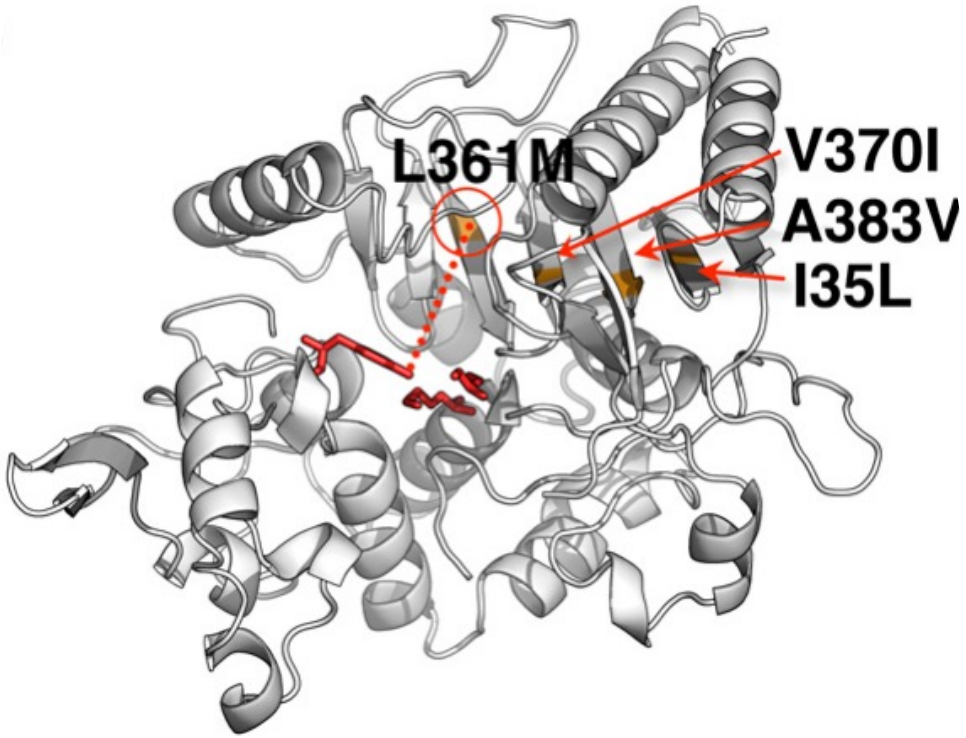
Neural Network



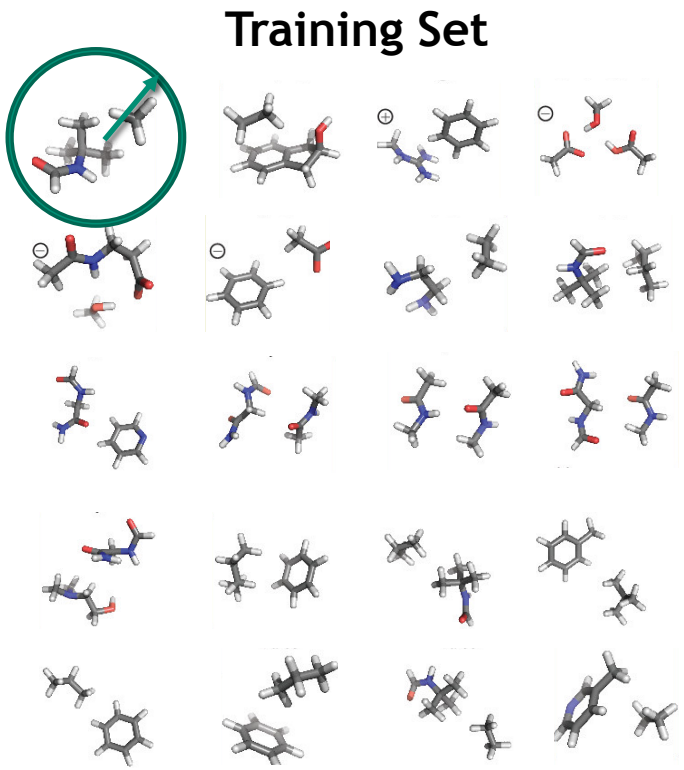


What's the Problem? – Part 1: Training Data

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



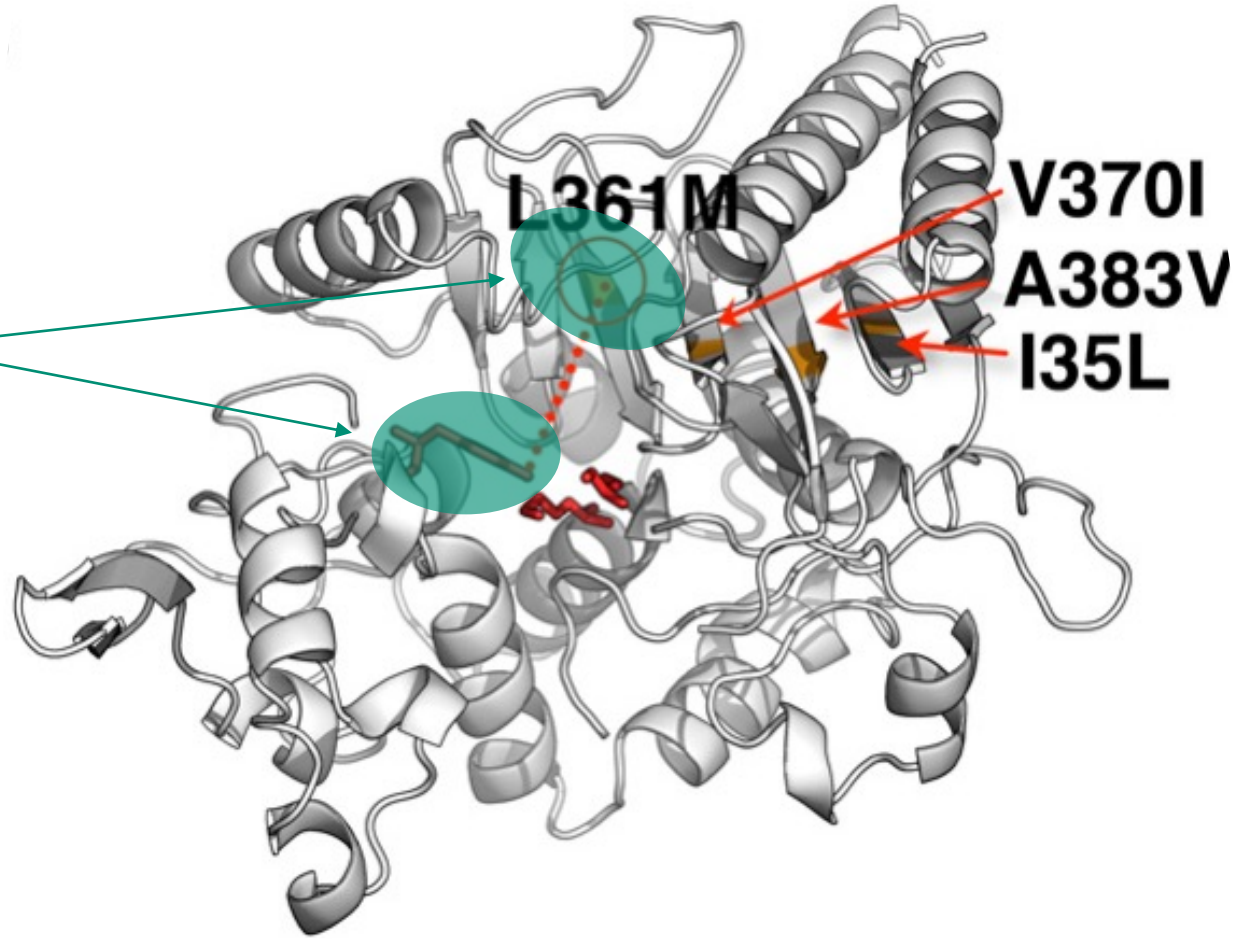
Gonzalo, Osuna, et al., *Nature chemical biology* (2014)



Effective Radius
10 Å

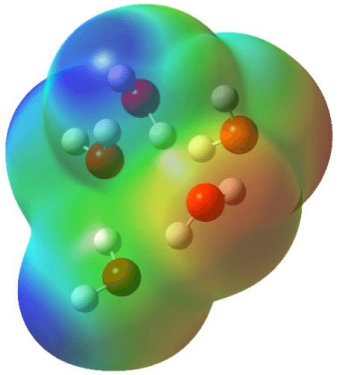
Solution: Learn the **electron density**

Electron Densities can interact!

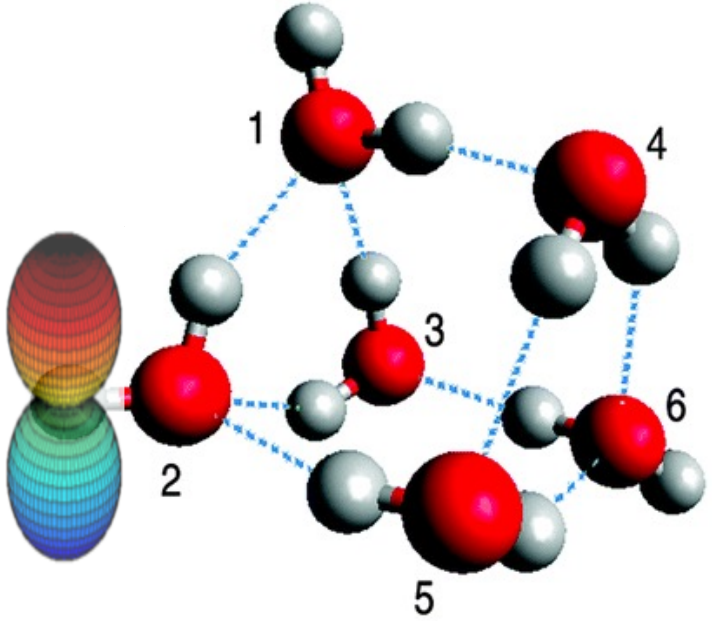


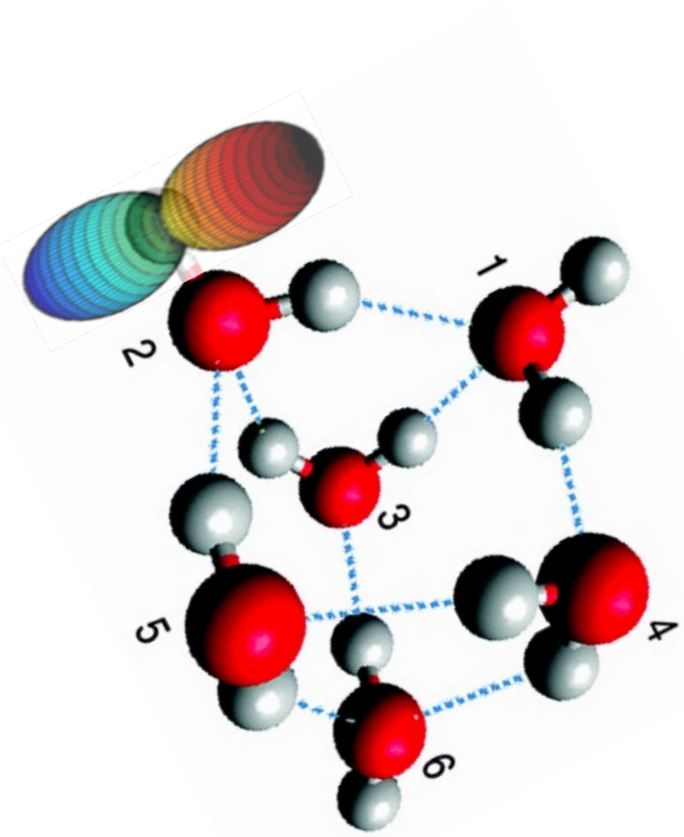
What's the Problem? – Part 2: Learning in 3D Space

The electron density is a 3D object.

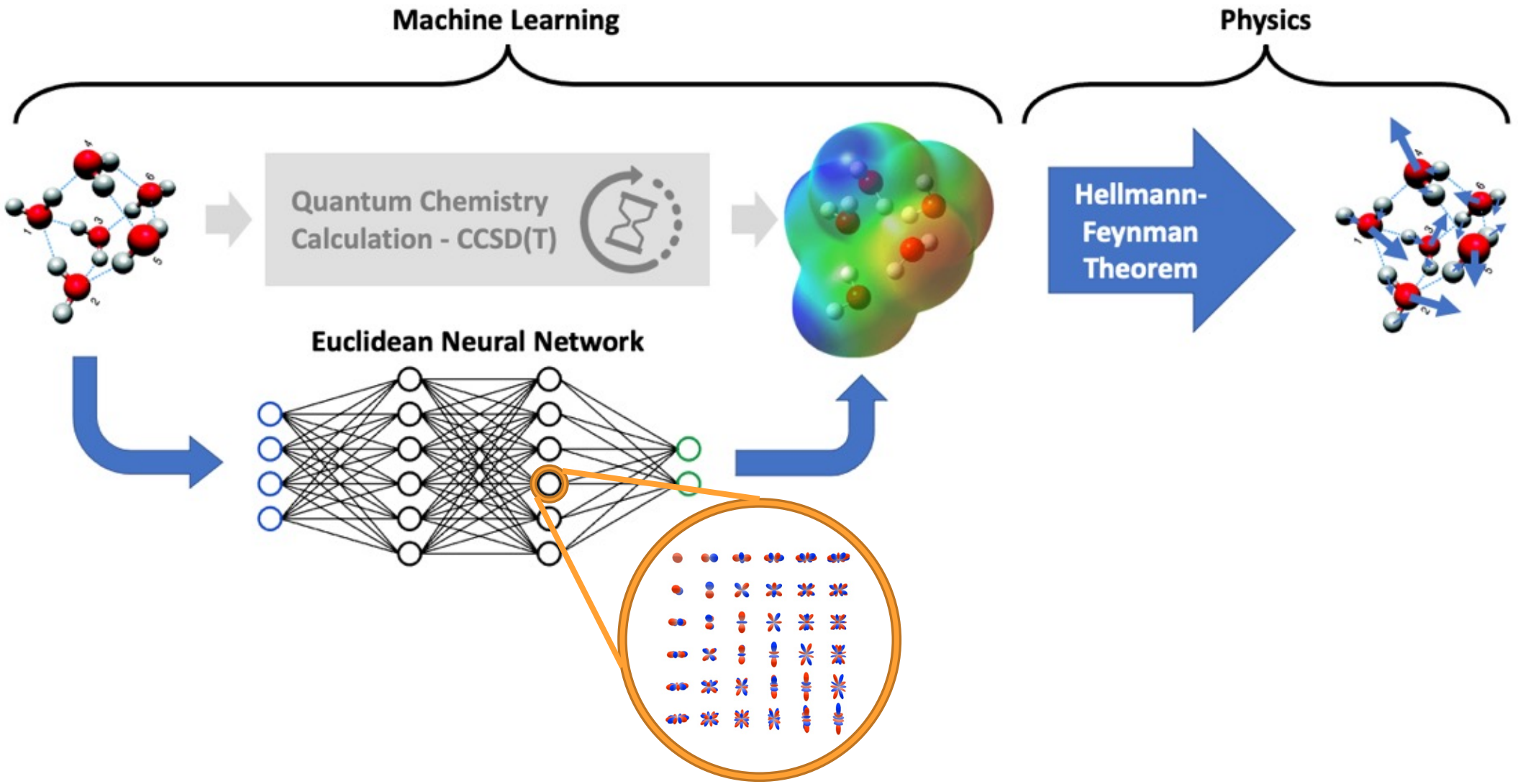


We want equivariance.





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



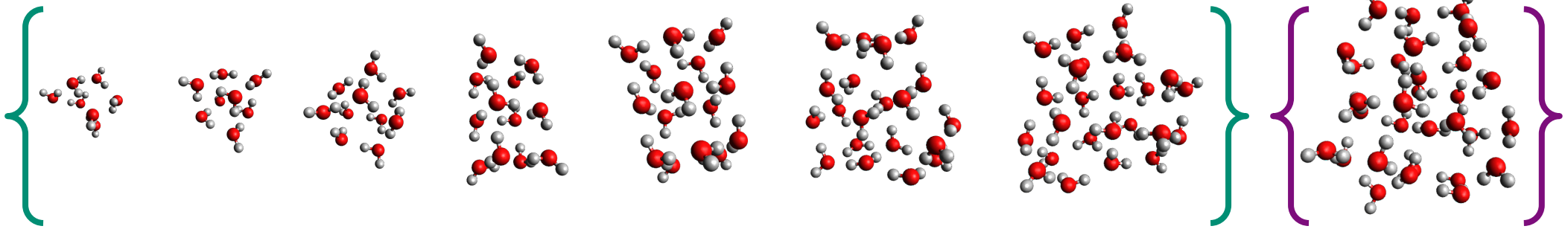
Break: Colab Notebook!

https://colab.research.google.com/drive/1ryOQ6hXxCidM_mGN0Yrf4BbjUtpyCxgy#scrollTo=X7IIA9wwvJQP

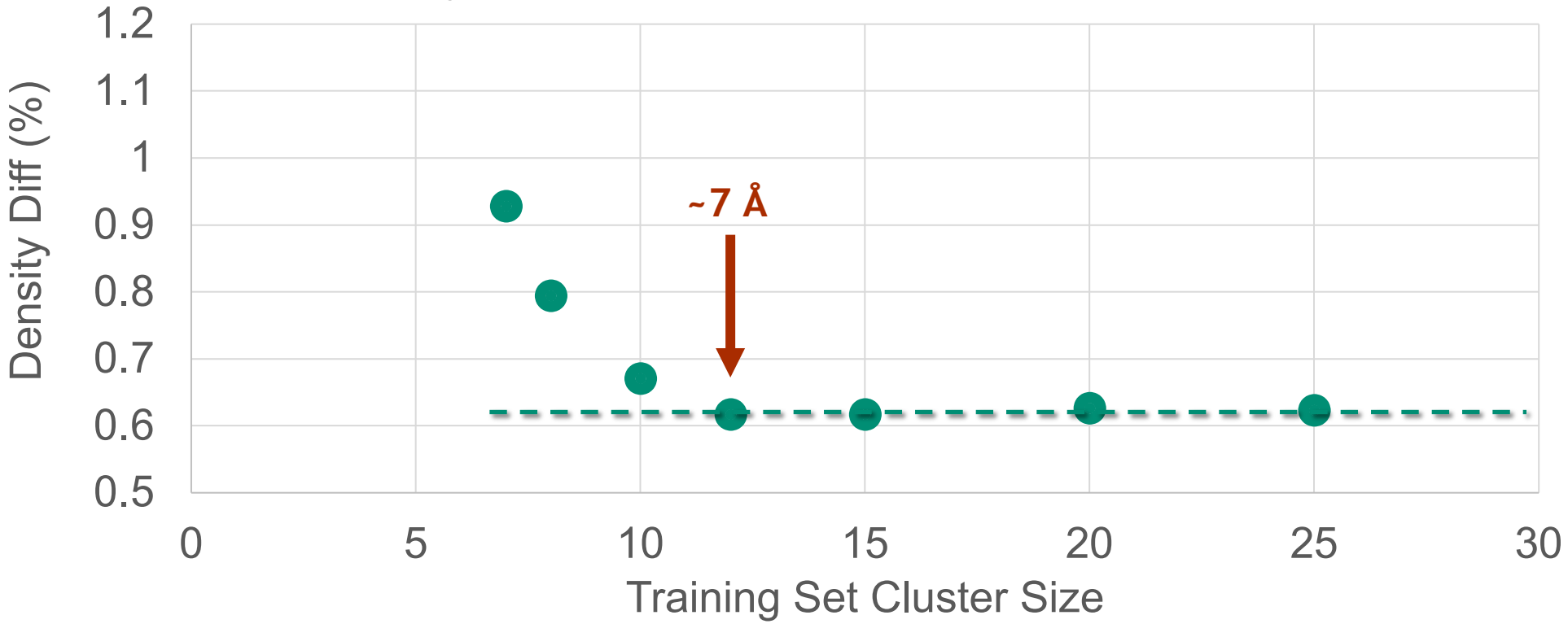
Water Experiment 4: Systematic Extrapolation

Train

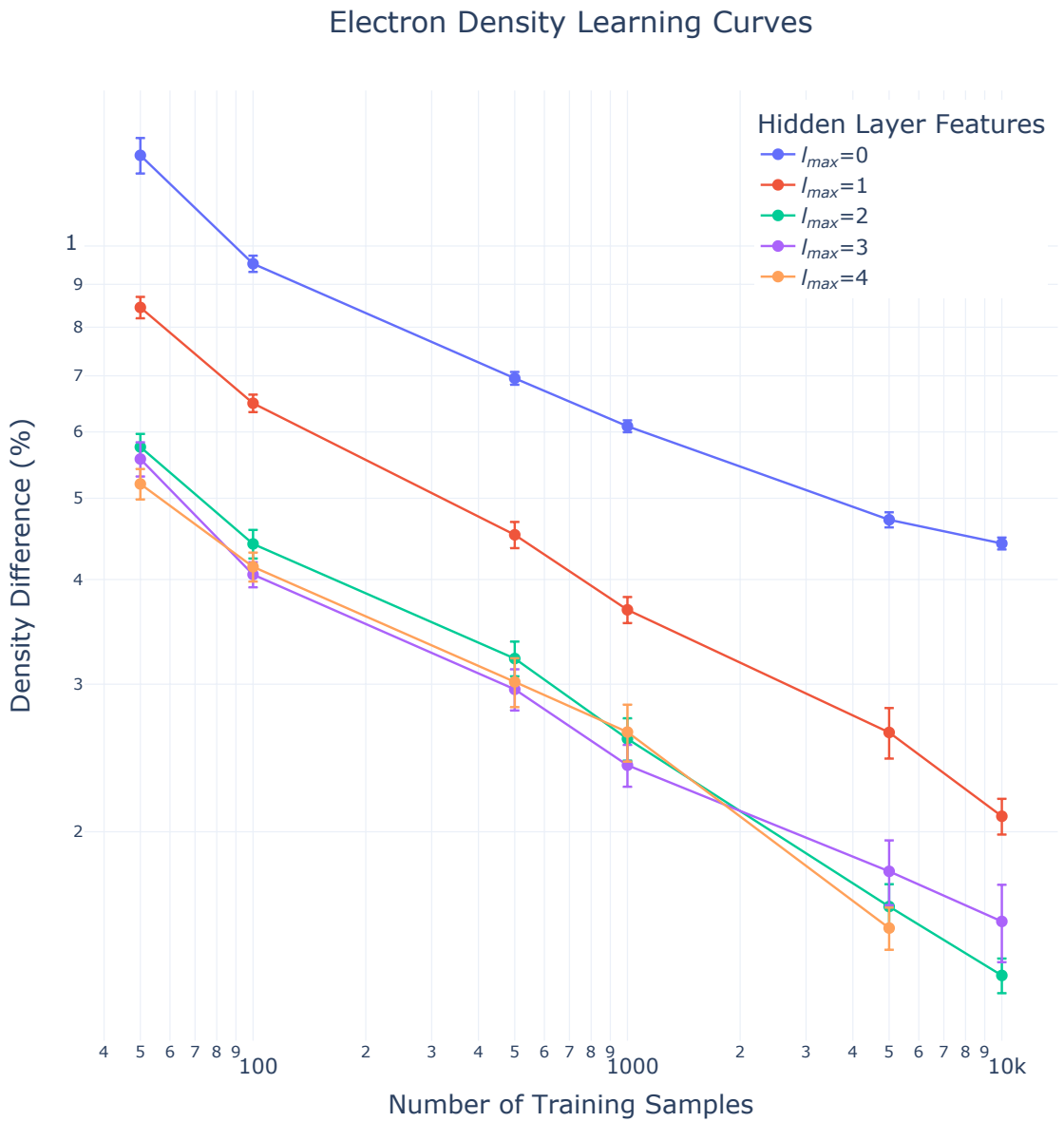
Test



Density Difference: 30 water molecule test set

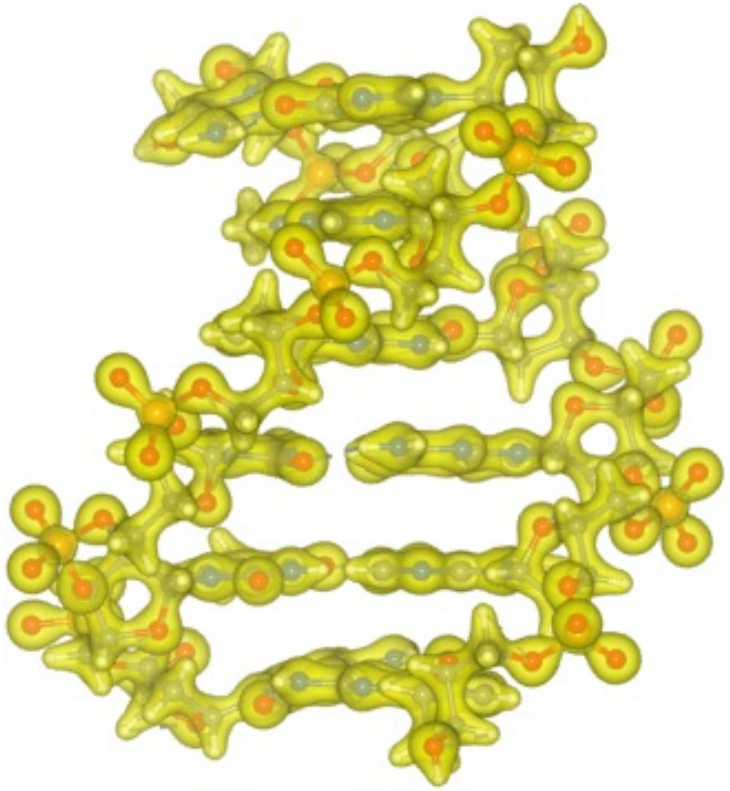


Water Experiment 2: The importance of L>1 features

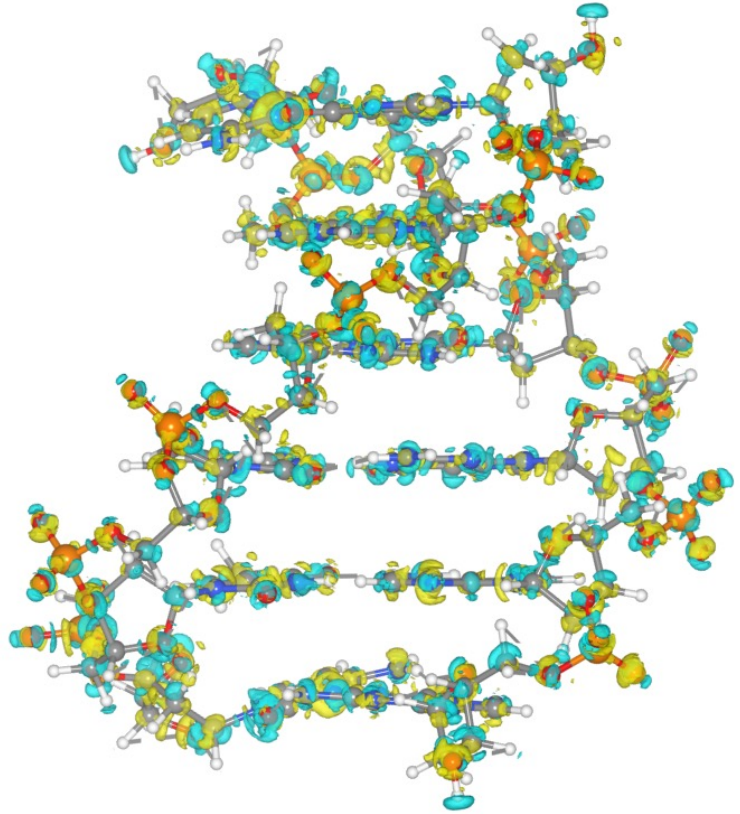


DNA – A real, experimental structure! (PDB code 251d)

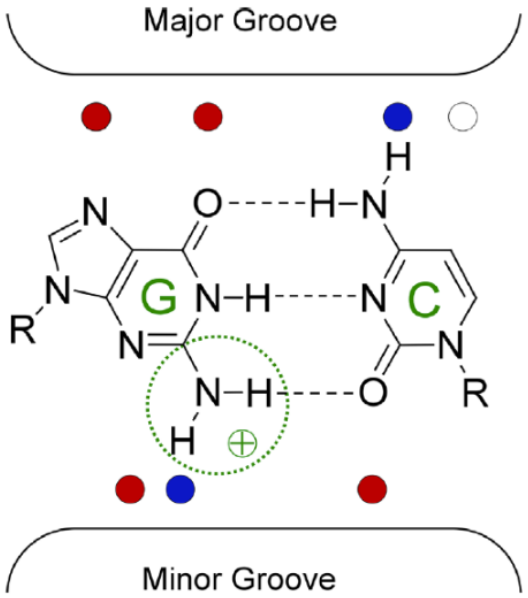
ML prediction



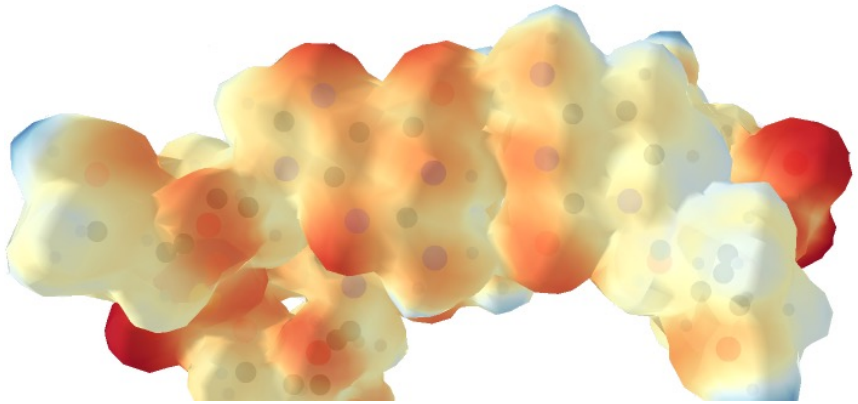
Density difference: ML - target



DNA – Electrostatic Potential



ML predicted potential

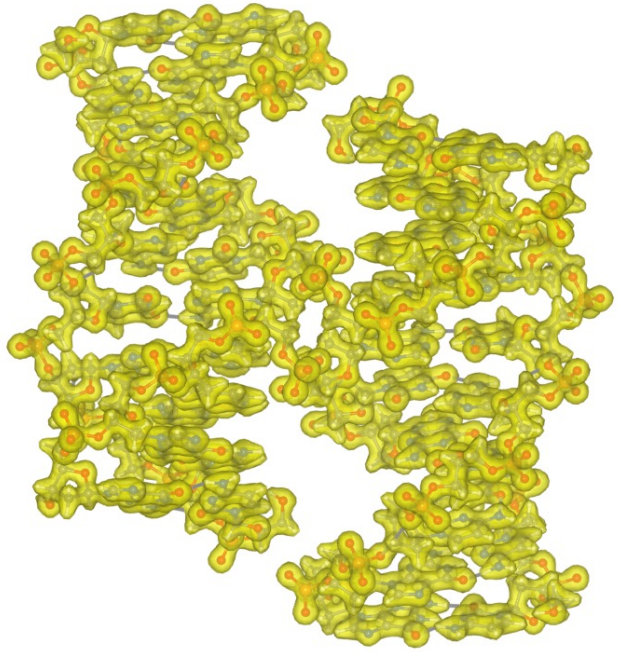


RMSD	
ML	0.017
AMBER	0.049

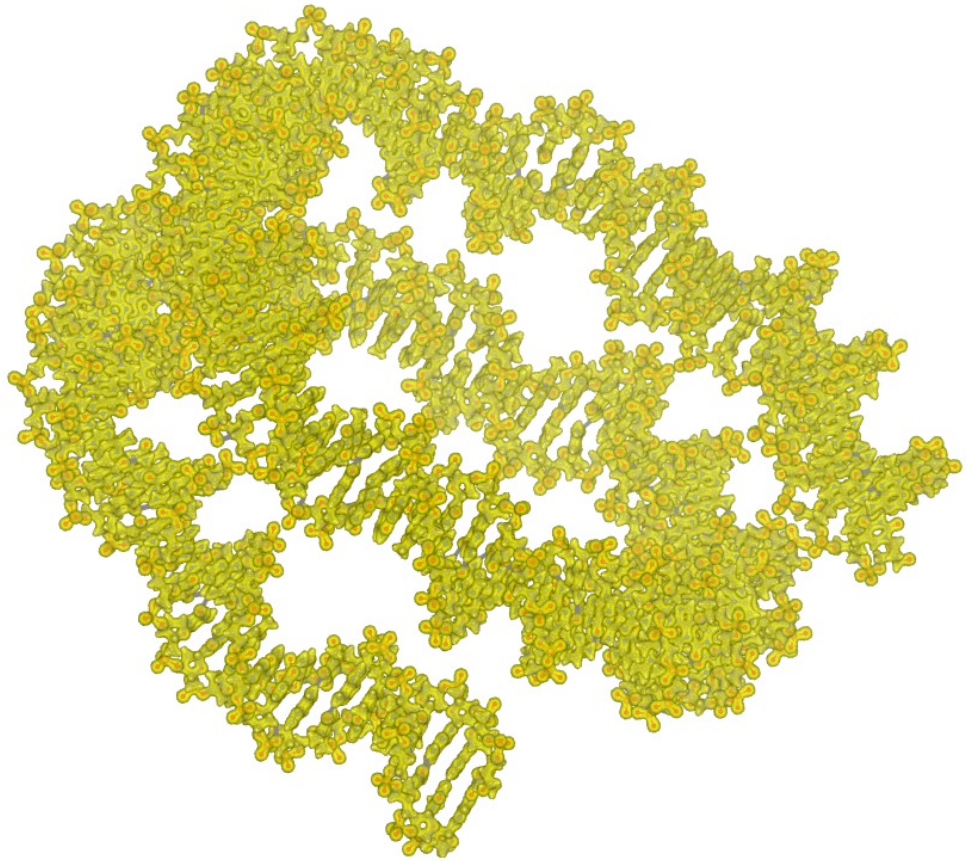
DNA – Predictions

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)

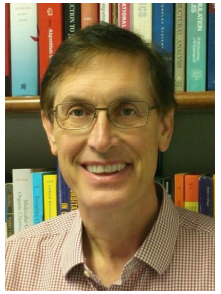


Acknowledgements

Steve Plimpton

Aidan Thompson

Susan Rempe



Jay Ponder
(Washington University in St. Louis)

Lucas Tecot
(UCLA)

Tess Smidt
(MIT)

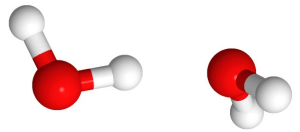
Mario Geiger
(EPFL)

Will Bricker and Alex Lee
(UNM)

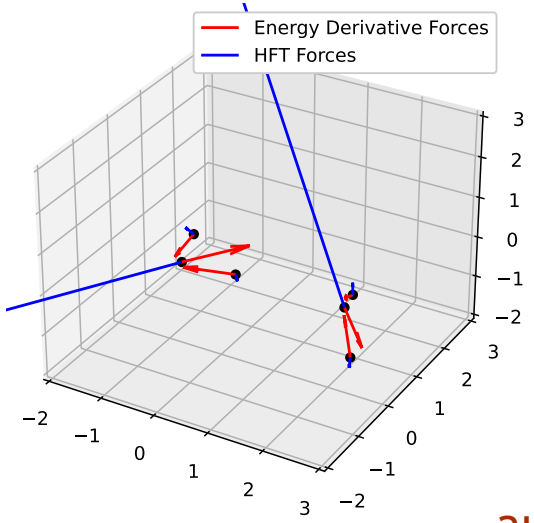
LDRD LABORATORY DIRECTED RESEARCH & DEVELOPMENT

SANDIA NATIONAL LABORATORIES THE PRESIDENT HARRY S. TRUMAN DISTINGUISHED LECTURES

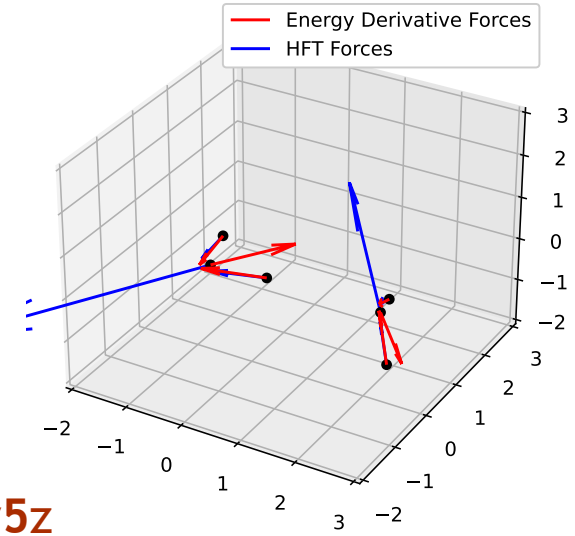
HFT forces with HFT optimized basis sets



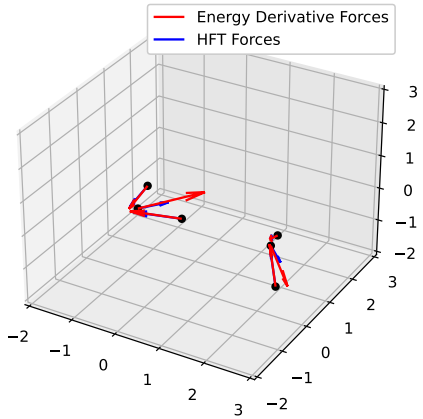
aug-cc-pvdz
82 basis functions



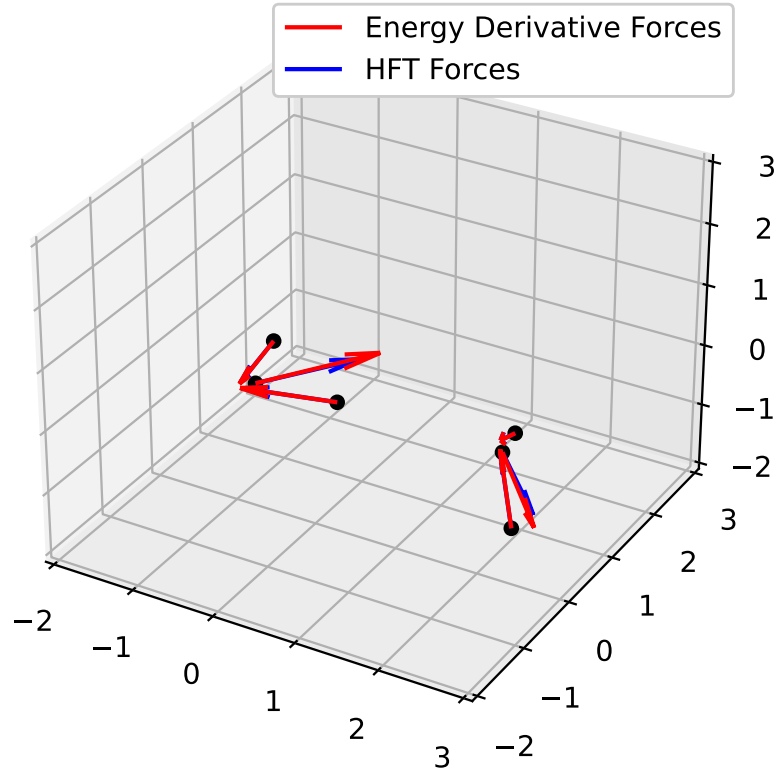
aug-cc-pvqz
344 basis functions



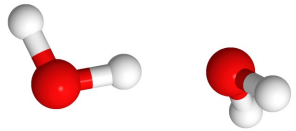
aug-cc-pv5z
574 basis functions



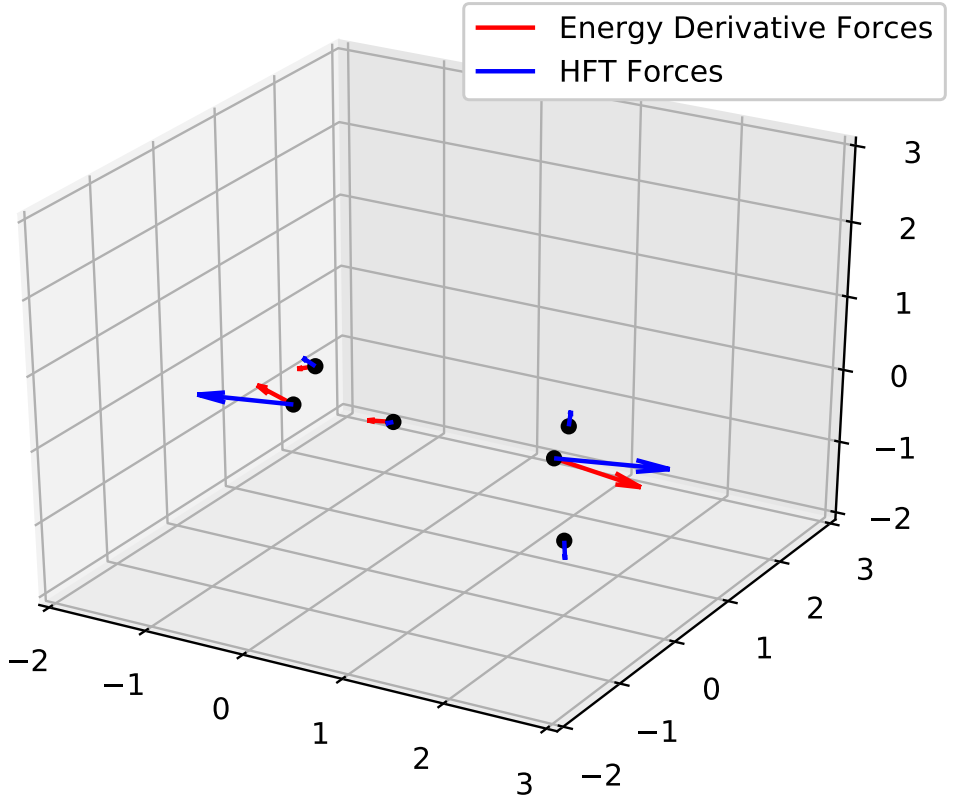
hft-basis
154 basis functions



HFT forces with CCSD(T) densities



hft-basis-MEDIUM
154 basis functions



hft-basis-LARGE
334 basis functions