SAND2022-10437C

DEPARTMENT OF CHEMISTRY & CHEMICAL BIOLOGY

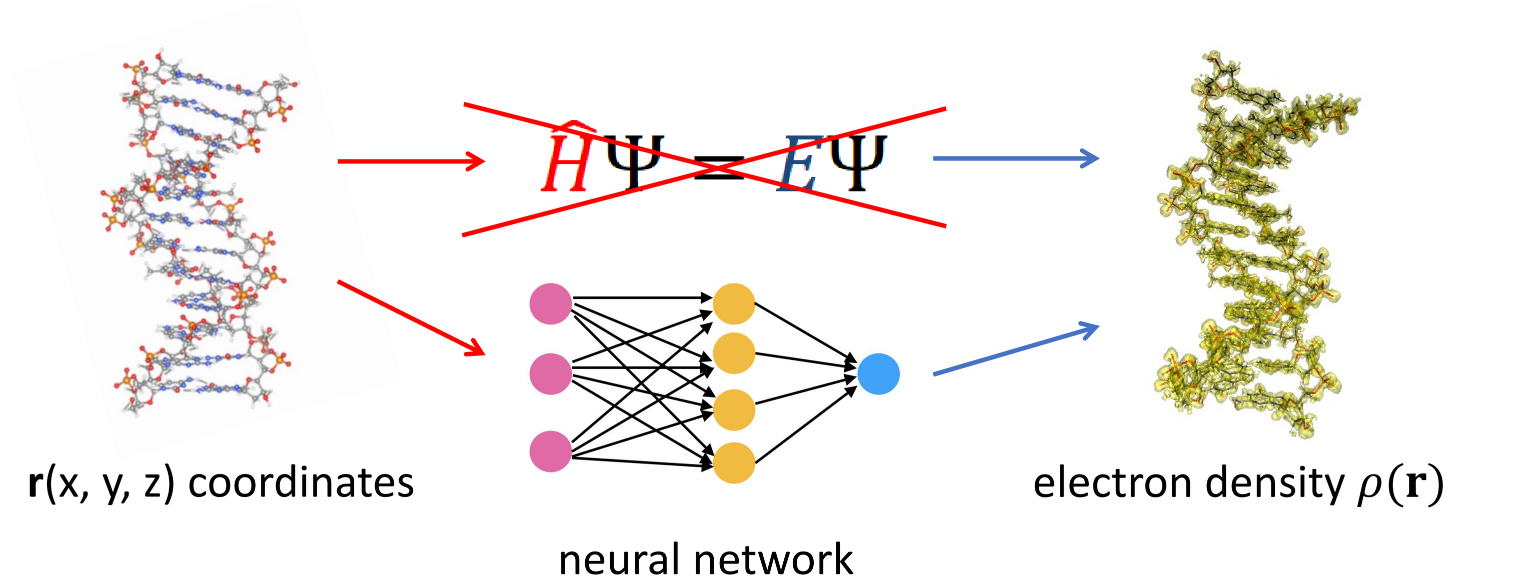
Predicting accurate *ab initio* DNA electron densities with equivariant neural networks

Alex J. Lee, Joshua A. Racker, William P. Bricker

Sandia National Laboratories

MOTIVATION

Quantum molecular modeling can predict a system's properties from **first principles** but is limited by **computational scaling**.

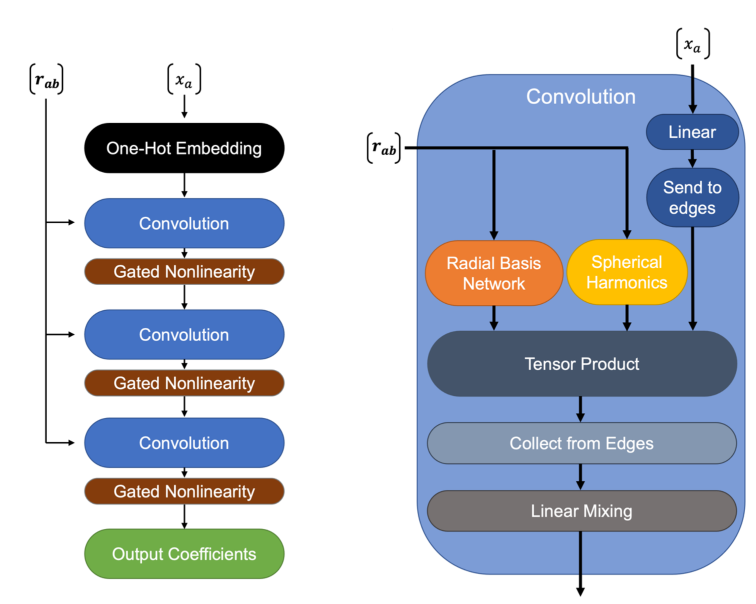


Machine learning **breaks traditional scaling barriers** and opens up the study of large biological macromolecules like DNA.

NEURAL NETWORK (e3nn)

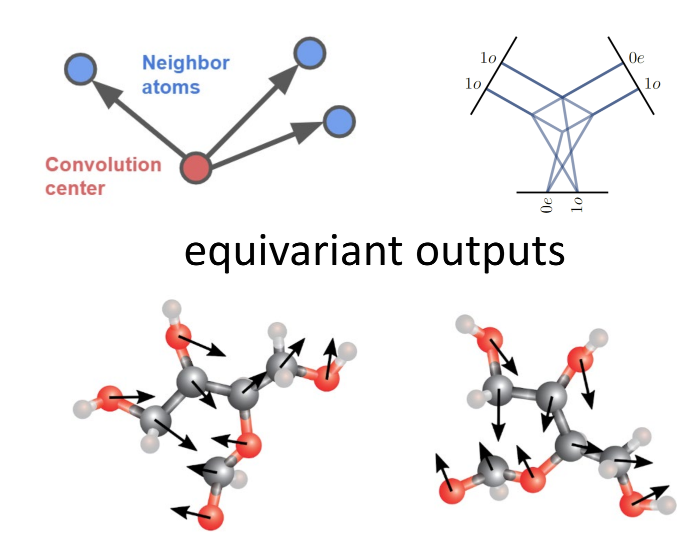
e3nn is a type of **graph convolutional neural network** that is **equivariant** to translations, rotations, and reflections.

e3nn architecture

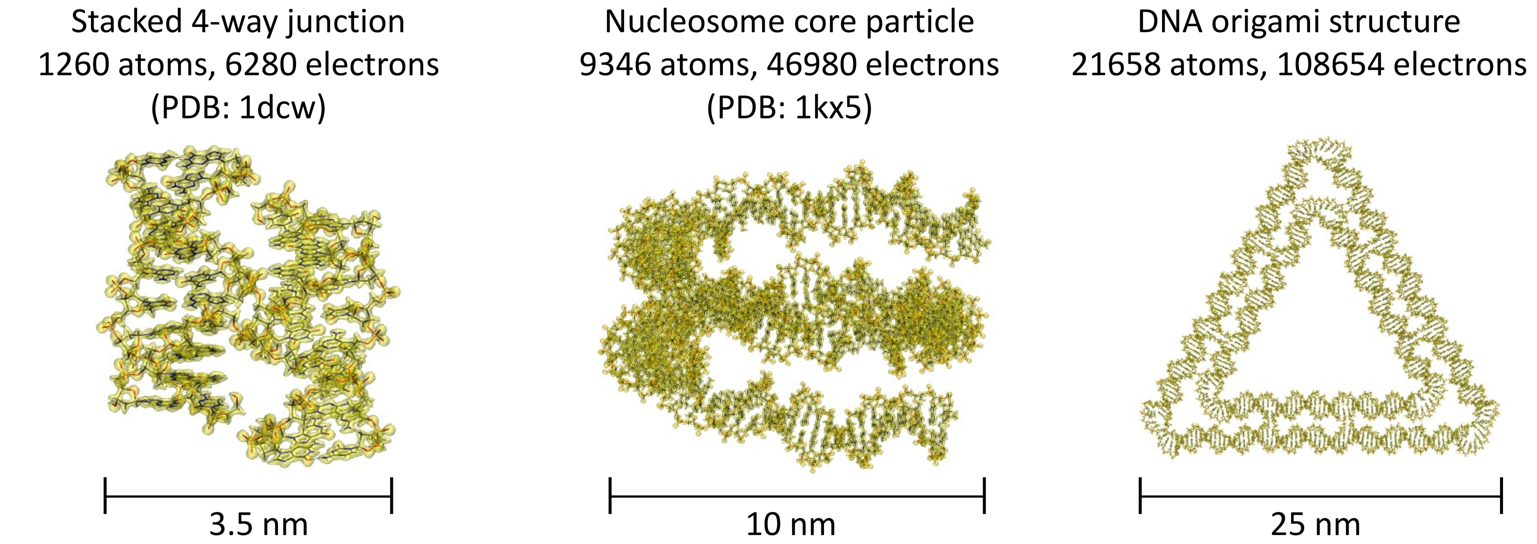


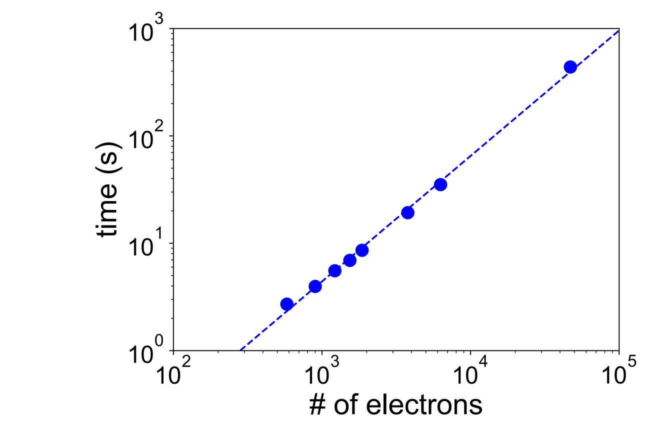
convolution tensor product

$$C = x_b \otimes R(r) Y_l^m(\hat{r})$$



RESULTS: LARGE DNA STRUCTURES



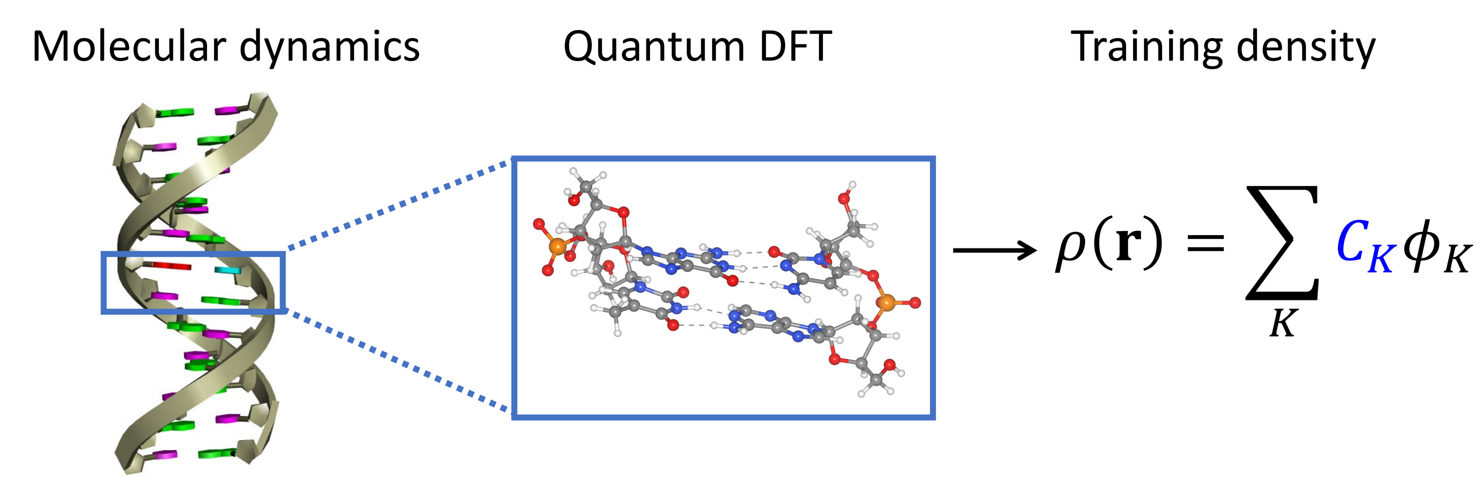


Computational scaling

Density functional theory	$\mathcal{O}(N^3)$
Hybrid functionals	$\mathcal{O}(N^4)$
MP2 and beyond	$\mathcal{O}(N^5 \text{ to } N^7)$
Machine learning	$\mathcal{O}(N)$

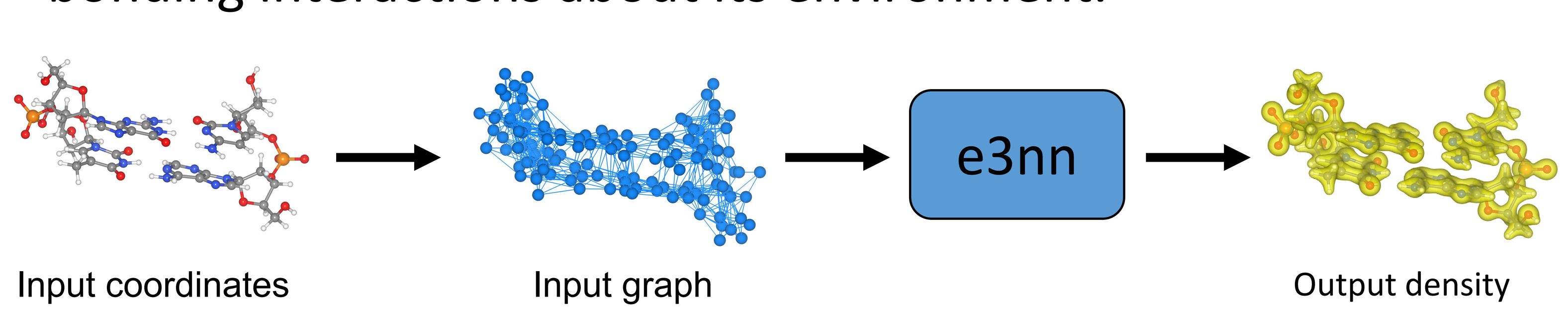
TRAINING THE MODEL

- Configurations sampled from molecular dynamics
- Smallest meaningful unit – **base pair step**
- Machine learning target is the **electron density** $\rho(\mathbf{r})$



$$\rightarrow \rho(\mathbf{r}) = \sum_K C_K \phi_K$$

During training, the model learns geometric features such as bonding interactions about its environment.

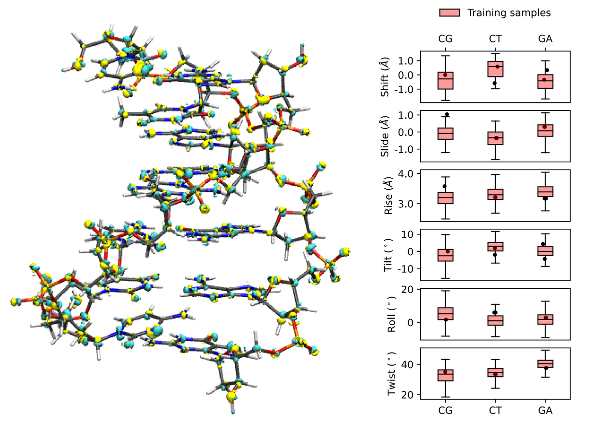


RESULTS: PREDICTION ACCURACY

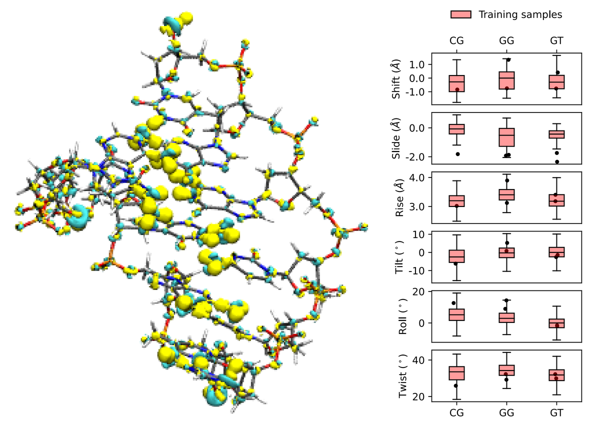
- Density prediction errors (%) for a model **trained only on B-DNA**

(L) Charge density difference plots
(R) Base step structure parameters

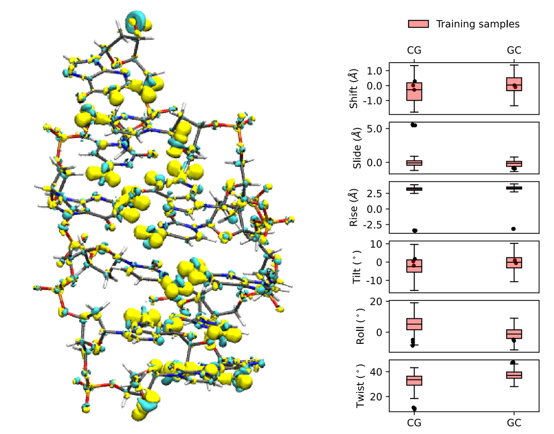
B-DNA, PDB: 251d
Density error: 1.15%



A-DNA, PDB: 1zf8
Density error: 1.69%



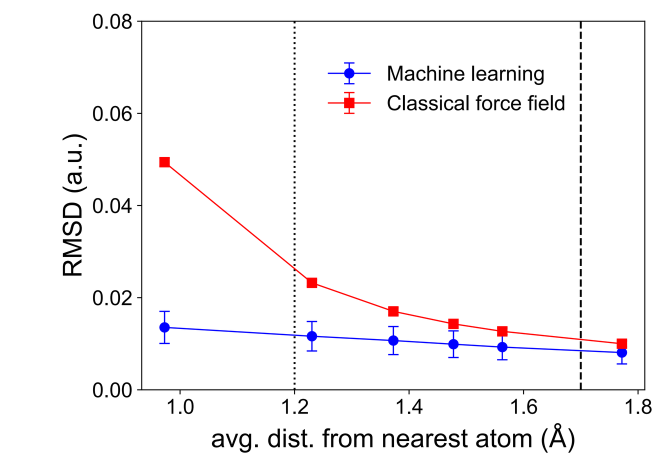
Z-DNA, PDB: 4fs5
Density error: 1.77%

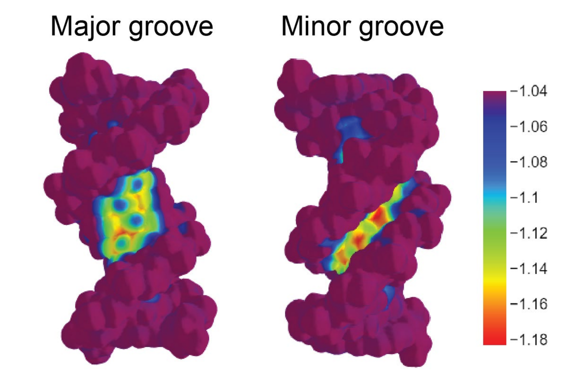


Preprint available on chemRxiv
DOI: [10.26434/chemrxiv-2022-pmrg8](https://doi.org/10.26434/chemrxiv-2022-pmrg8)

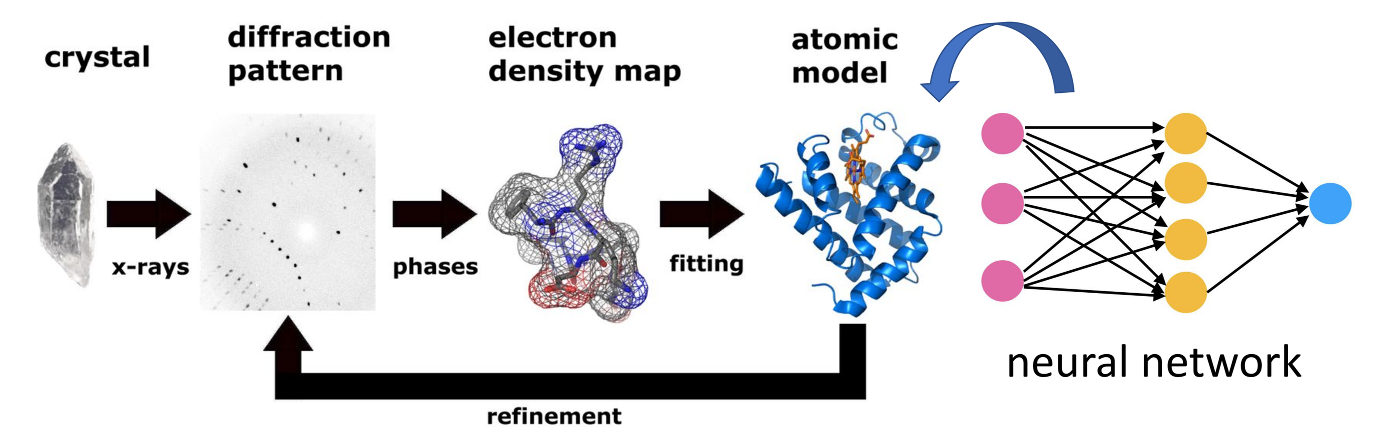
FUTURE APPLICATIONS

- Machine-learning derived electrostatic potentials





- High-resolution X-ray crystal structure refinement



- Hellman-Feynman forces for *ab initio* molecular dynamics

$$\vec{F} = -\left\langle \psi_R \left| \frac{d\hat{H}_R}{dR} \right| \psi_R \right\rangle$$

Only needs density predicted by ML

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. Unlimited Release SAND2021-xxxxxx

NM Partnerships

U.S. DEPARTMENT OF ENERGY
National Nuclear Security Administration