Predicting Molecular Orbital Energies with Deep Learning on Pytorch

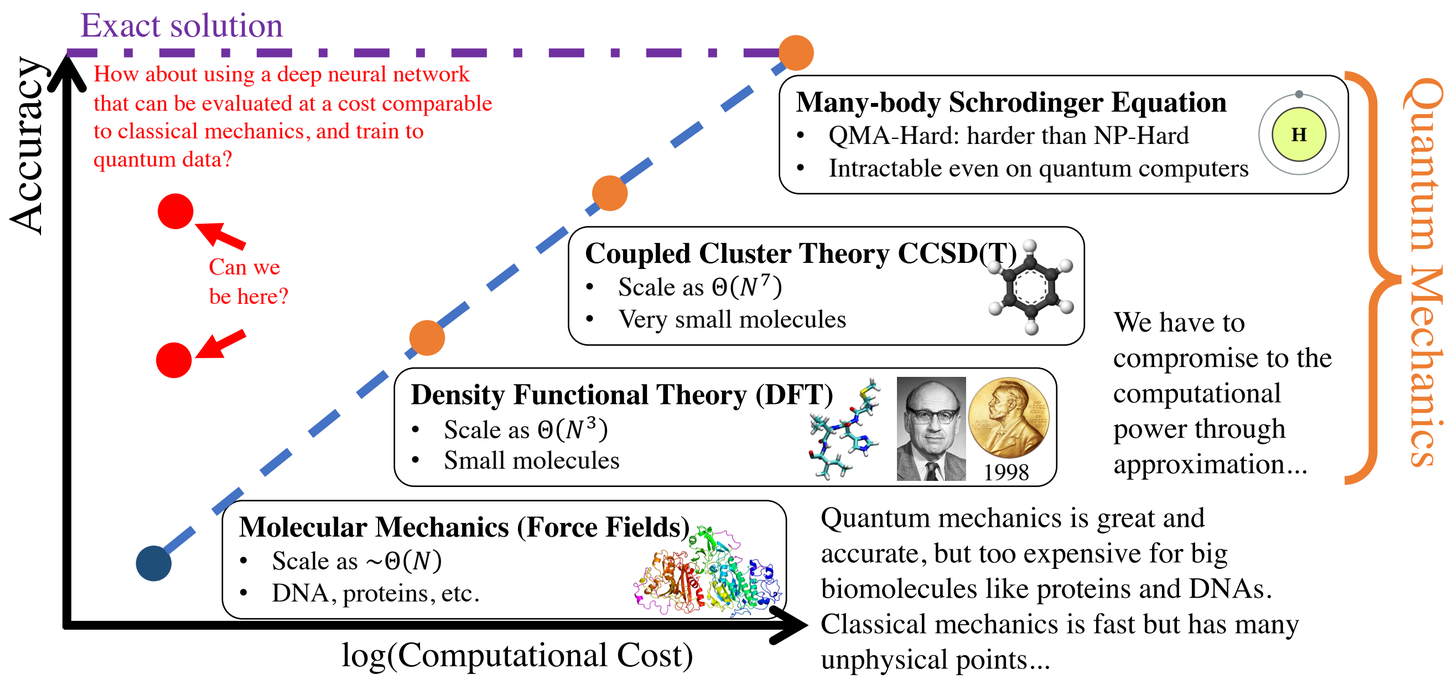
UF UNIVERSITY of FLORIDA

Richard (Jinze) Xue, Xiang Gao, Adrian E. Roitberg

Department of Chemistry, University of Florida

UF Roitberg Group Computational Chemistry

Neural Network Potential in Computational Chemistry



MO Energies vs Potential Energy

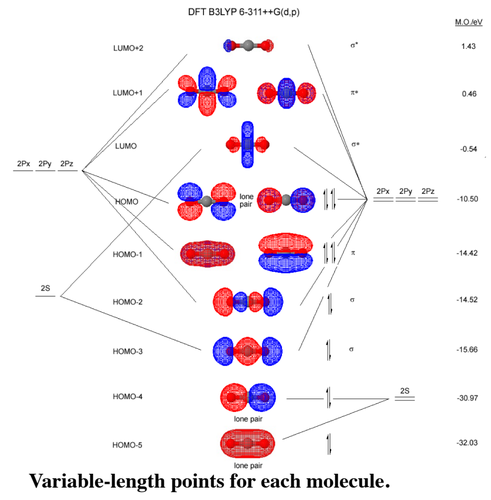
Potential Energy

TorchANI
Our previous work to predict Potential Energy

One point for each molecule, which is the total energy of molecule at 0 K

Application: Molecular Dynamics, since the derivative of potential to coordinates is force.
So it could be used to predict molecule movement.

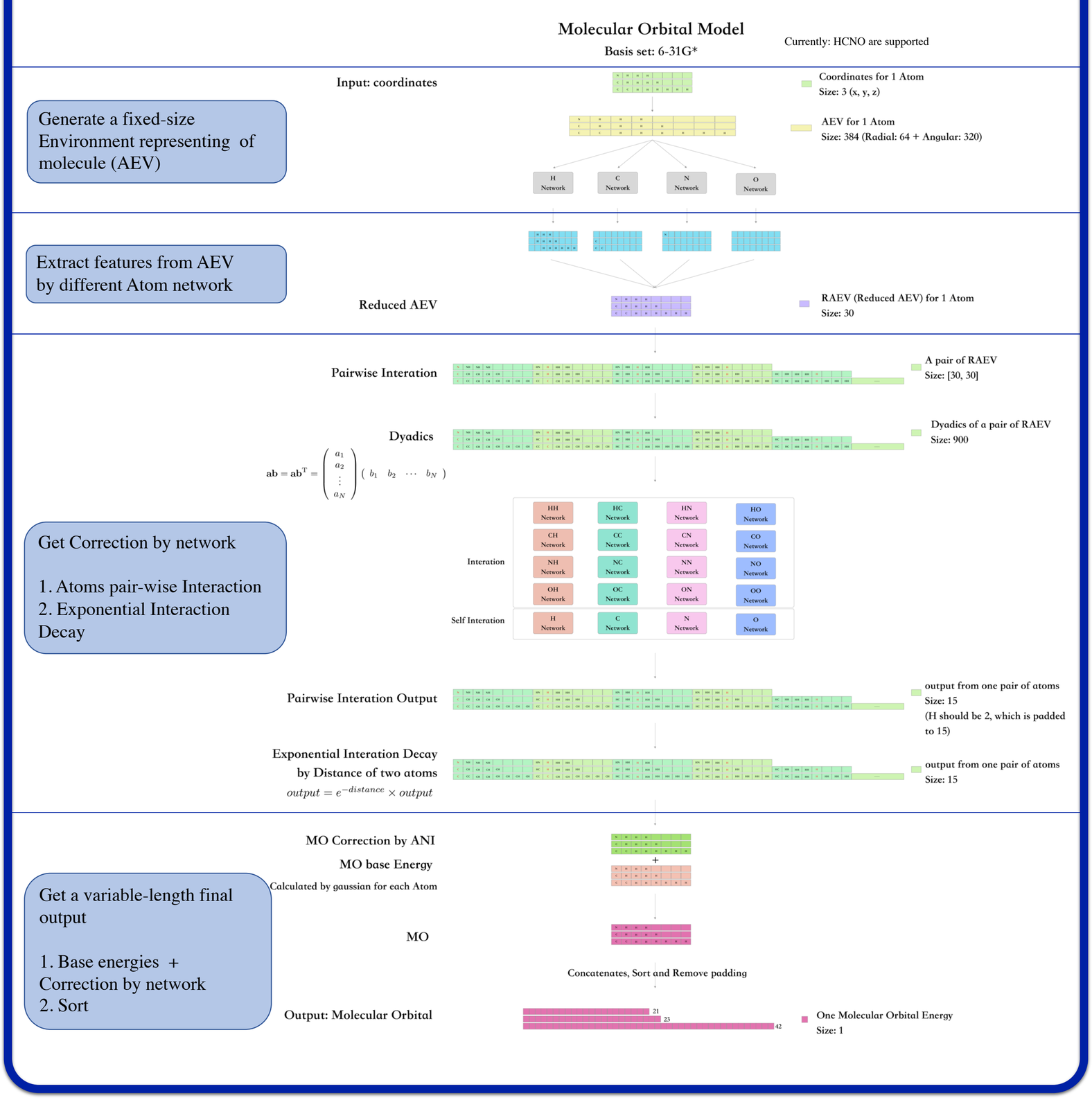
MO Energies



Variable-length points for each molecule.

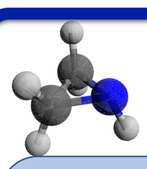
Application: Semiconductors, Photon-chemistry and Organometallic Chemistry, and dive deeper of the molecule

Architecture



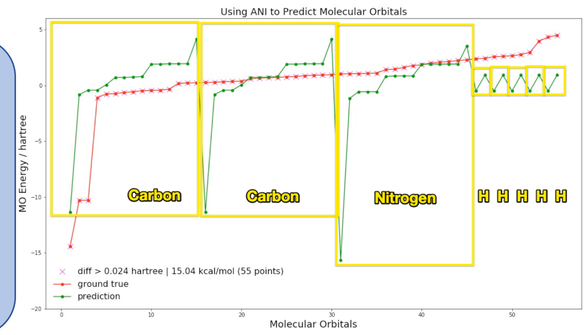
Example

Example of test molecule C_2NH_5

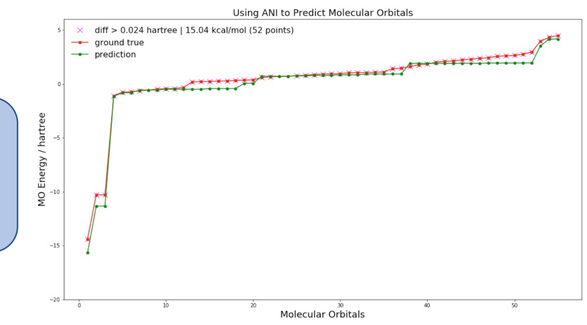


Base MO Energy by putting atom's Atomic Energies together.

Atomic Energy got by running Gaussian for each single Atom

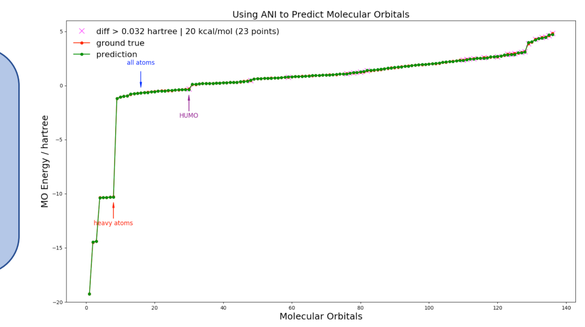


Sort the Atomic Energies to get good base MO Energies

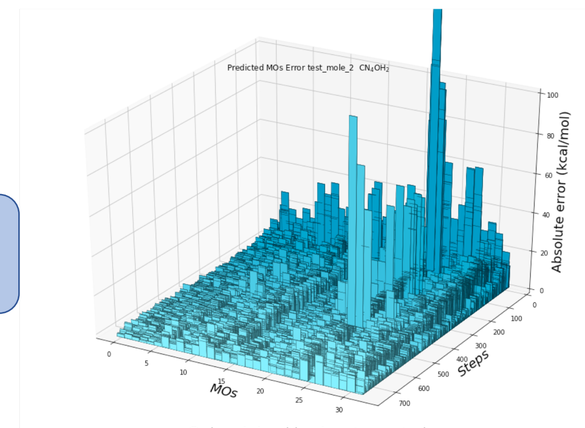


Final output

Base MO Energies + Correction by network.



Absolute error changes with training iterations



Only minimal basis set counted

Parameters and Results

- Dataset:
  MO Energies are extracted from part of (1/4) checkpoints files of ANI-1x datasets
  0.9 M datapoints for training
  0.1 M datapoints for validation
- Batch Size: 1000
- Optimizer: Adam
- Training Process:
  - First 1000 epochs are trained for all MO energies
  - After 1000 epochs, Loss function is changed to optimize the minimal basis set of MO Energies, since 6-31G(d) calculated too many virtual orbitals. And every 10 iterations, loss function will change to all basis set for one iteration.
- Result (minimal basis counted):
  - After 1000 epochs: training _rmse is 8.0111 kcal/mol , val_rmse is 8.0221kcal/mol
  - After 4000 epochs: training _rmse is 5.5680 kcal/mol, val_rmse is 6.4262 kcal/mol

Acknowledgements

HiPerGator 2.0 The University of Florida Supercomputer

TACC TEXAS ADVANCED COMPUTING CENTER

NSF

NVIDIA

Reference

- https://github.com/yueyericardo/ANI_D/ (have not public)
- https://github.com/aiqm/torchani
- https://github.com/isayev/ase_ani

- ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost
  J. S. Smith, O. Isayev, A. E. Roitberg. Chemical Science, 8:3192-3203 (2017)
- ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules
  Justin S. Smith, Olexandr Isayev, Adrian E. Roitberg. Scientific Data, 4:170193 (2017)
- Less is more: Sampling chemical space with active learning.
  Justin S. Smith, Ben Nebgen, Nicholas Lubbers, Olexandr Isayev, Adrian E. Roitberg.
  The Journal of Chemical Physics, 148:241733 (2018)
- Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. Justin S. Smith, Benjamin T. Nebgen, Roman Zubatyuk, Nicholas Lubbers, Christian Devereux, Kipton Barros, Sergei Tretiak, Olexandr Isayev, Adrian E. Roitberg.
  Nature Communications, 10:2903 (2019)