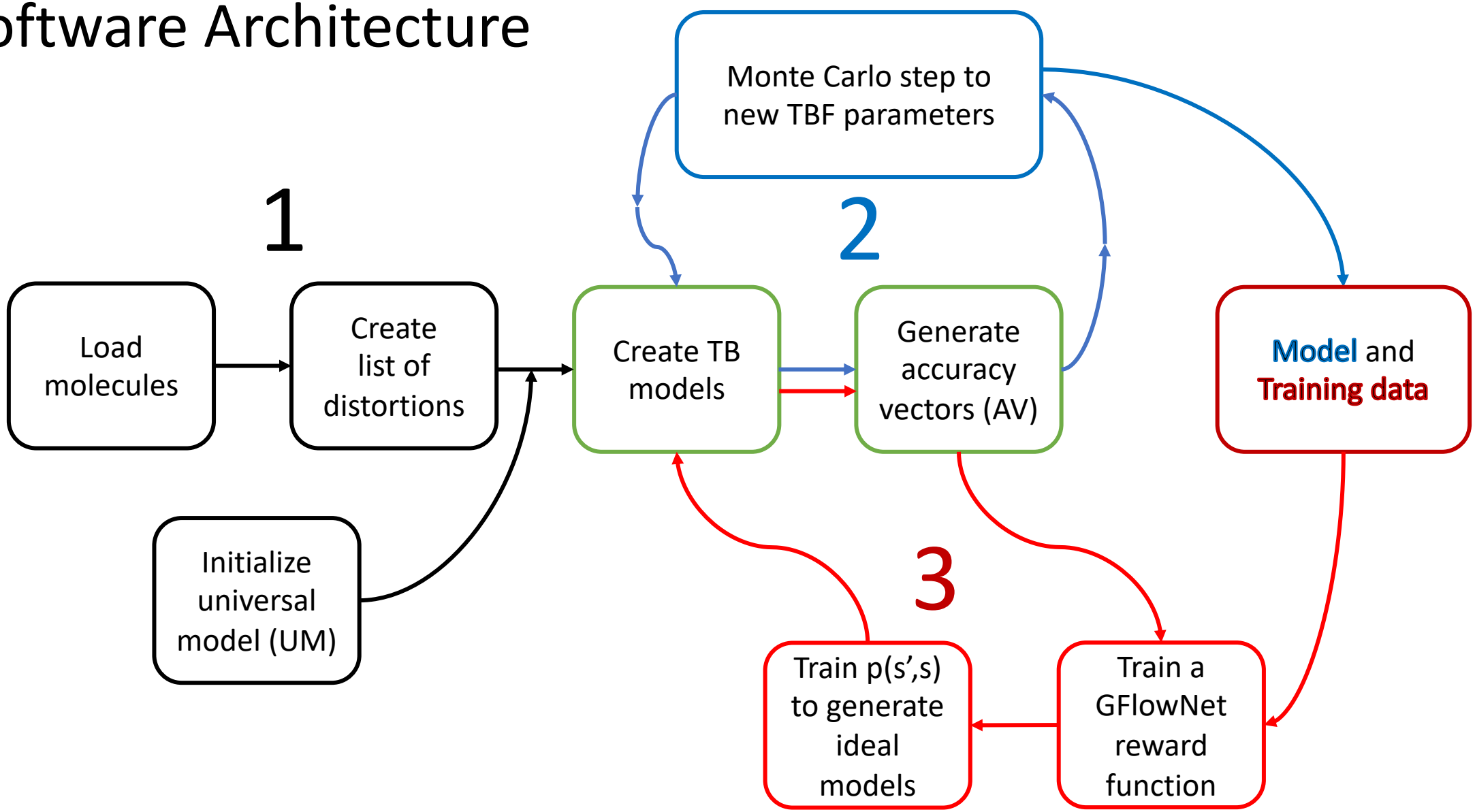
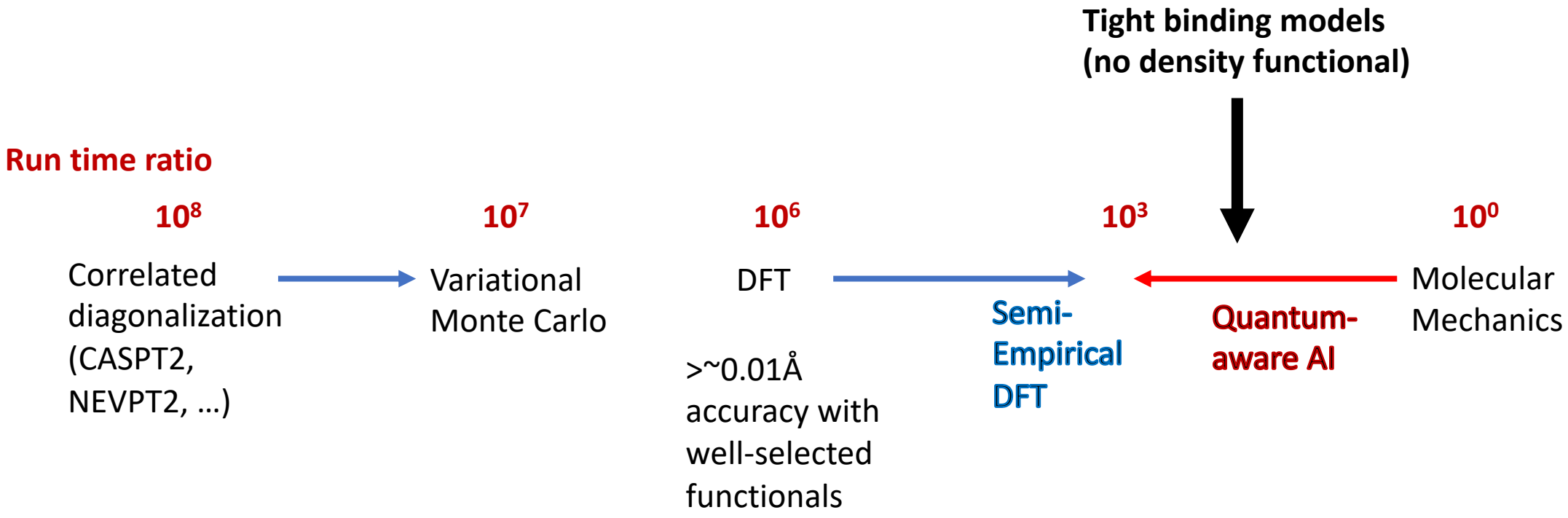
Software Architecture



Related NN approaches are used for model optimization in high energy physics due to similar 'curse of dimensionality' challenges.

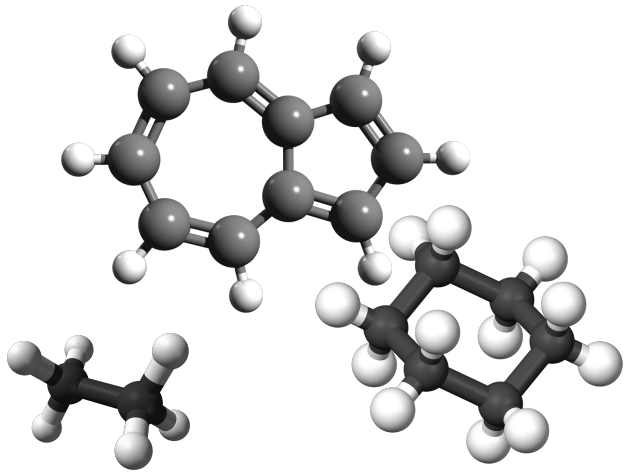
Context and Motivation



Density functional tight binding: values of semi-empirical methods in an ab initio era, Q. Cui and M. Elstner, Phys. Chem. Chem. Phys., 2014, 16, 14368–14377

Using Neural Network Force Fields to Ascertain the Quality of Ab Initio Simulations of Liquid Water, A Torres, LS Pedroza, M Fernandez-Serra, AR Rocha - The Journal of Physical Chemistry B 125 (38), 10772-10778

Tight binding field



A universal model mapping from **bond lengths** to **tight binding parameters**

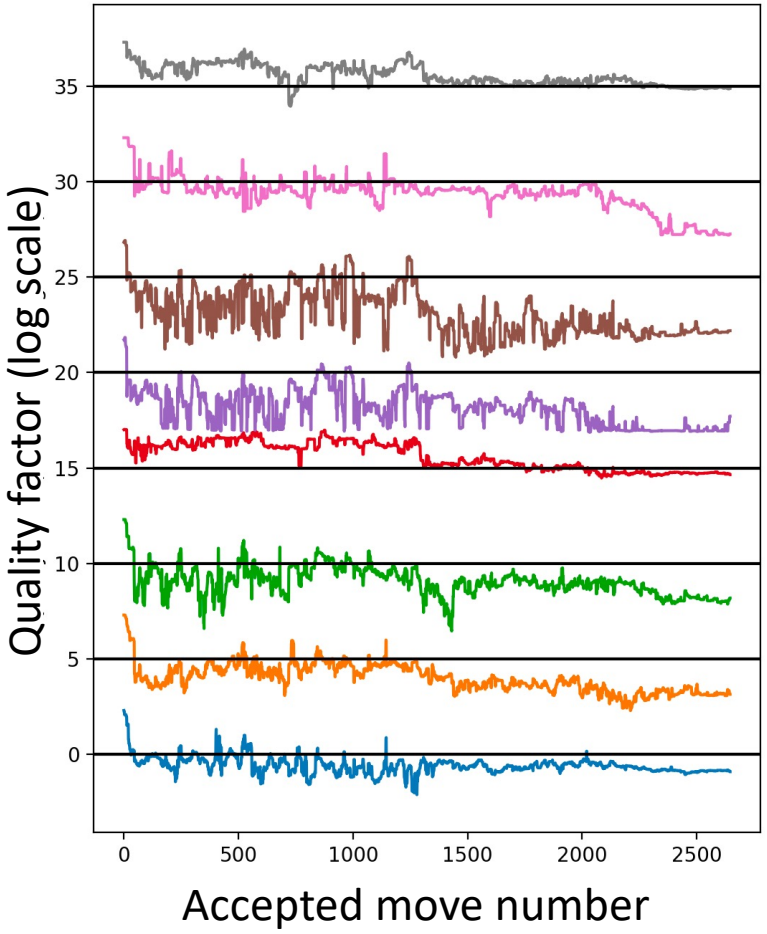
Demo case: s, p basis; C, H atoms; curves defined by 5 parameters → 87 dimensions

Fields respect Slater-Koster symmetry and are:

1. Coulomb: atom-atom
2. Coulomb: atom-orbital
3. Coulomb: orbital-orbital
4. Coulomb: atom-2 orbital (allowing hybridized sp orbitals)
5. Hopping (all Slater-Koster–allowed orbital combinations)

Non-field parameters:
Orbital energy

3-step Monte Carlo convergence for 8 molecules

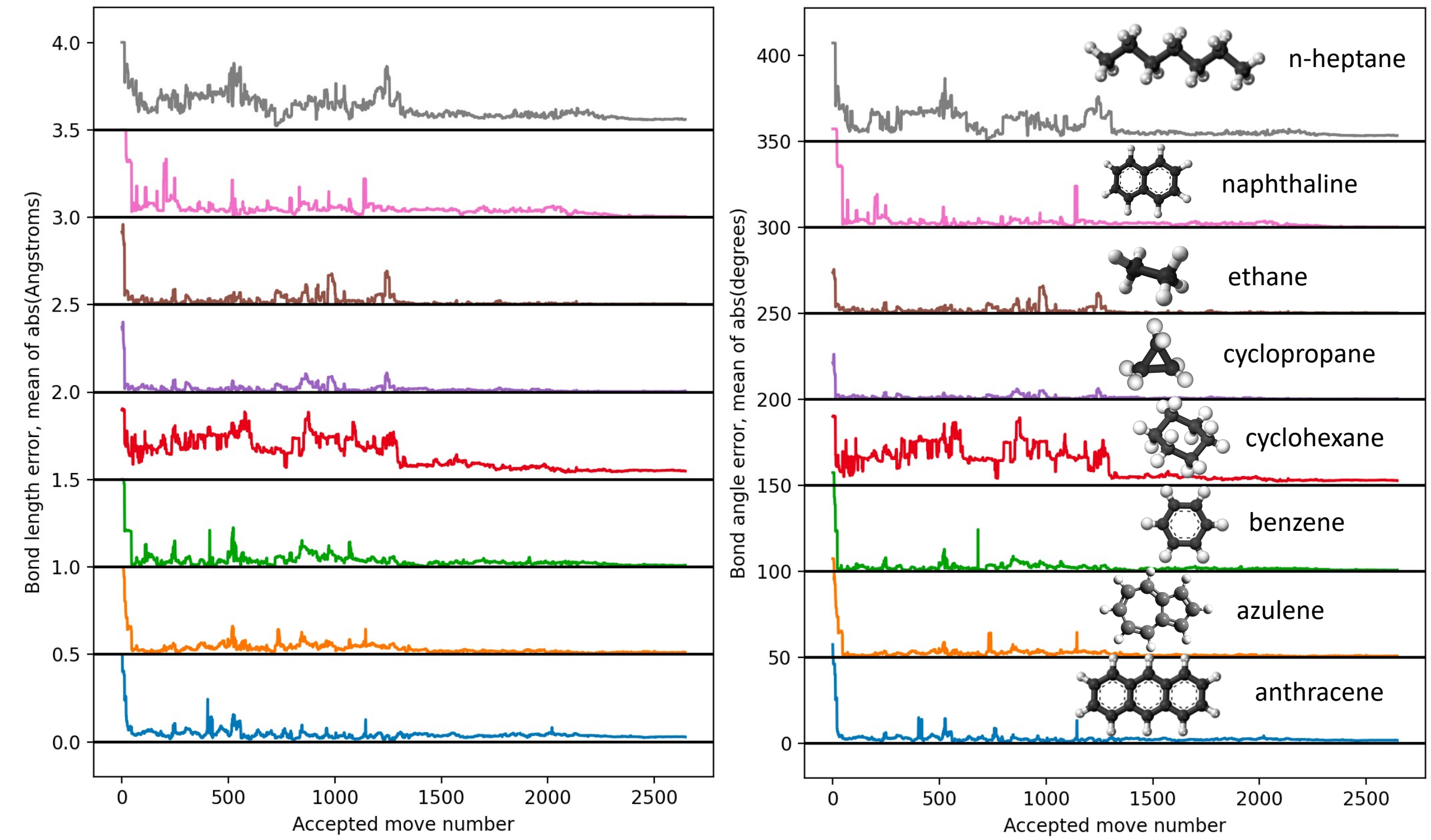


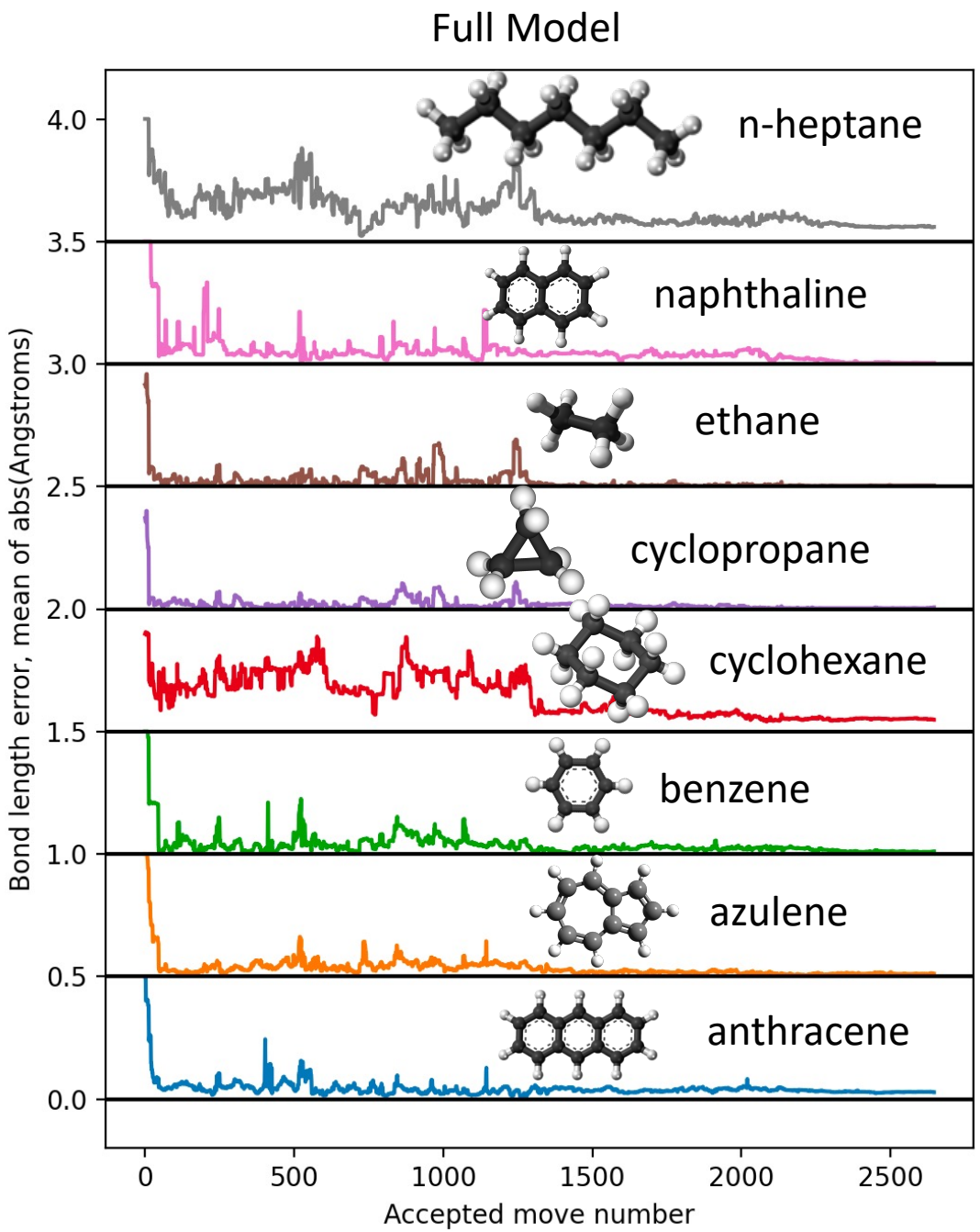

```
Current temperature is: 2
Current temperature is: 0.2
Current temperature is: 0.020000000000000004

Monte Carlo runtime is 260.2200481891632 s, or 0.0578266773753696s per attempted step.
```

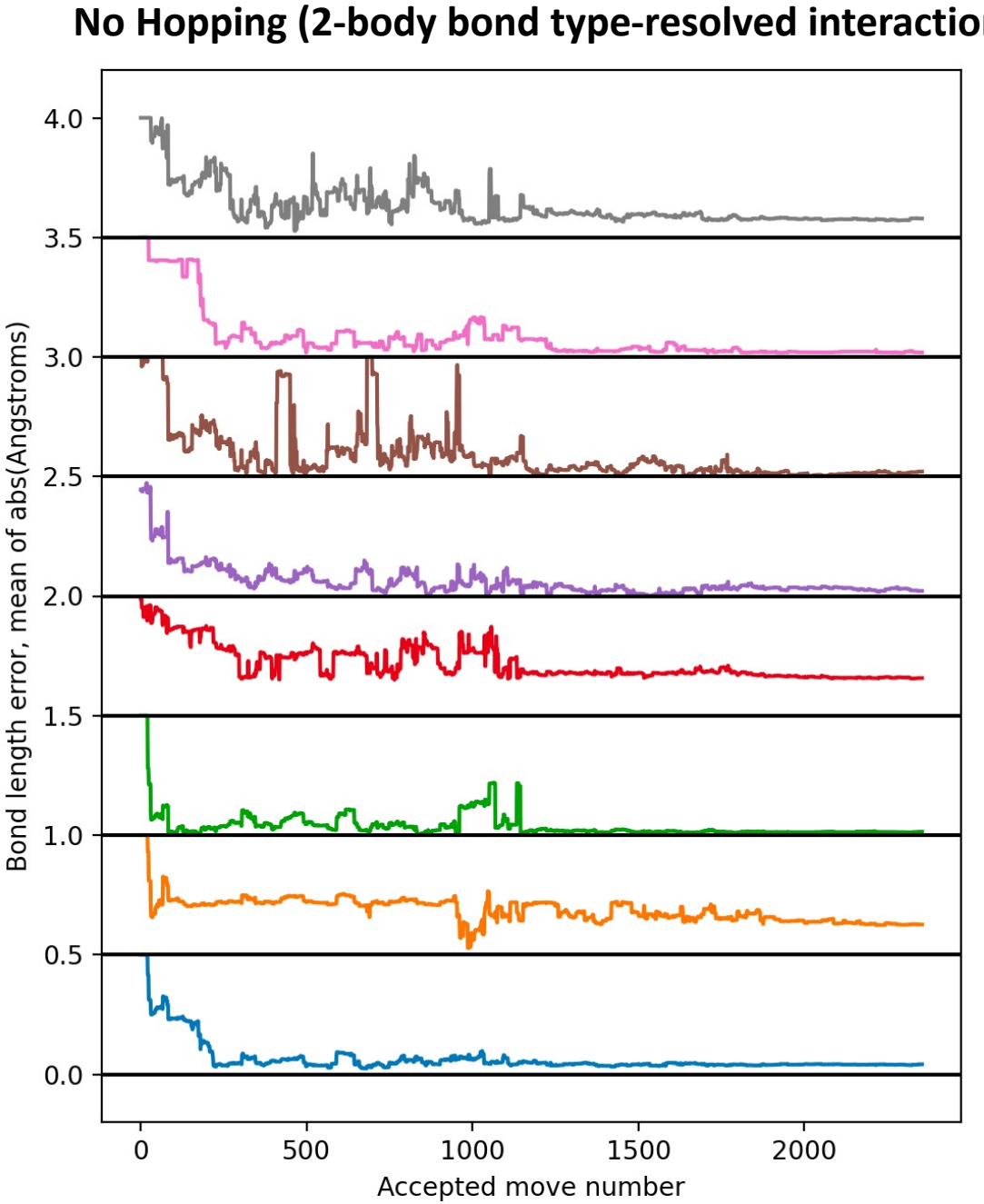
Each step involves diagonalizing for 40 configurations

Lines represent ~0.1 Å/bond

Molecular structures selected from the Cambridge Structural Database (CSD) hosted by the Cambridge Crystallographic Data Centre (CCDC)

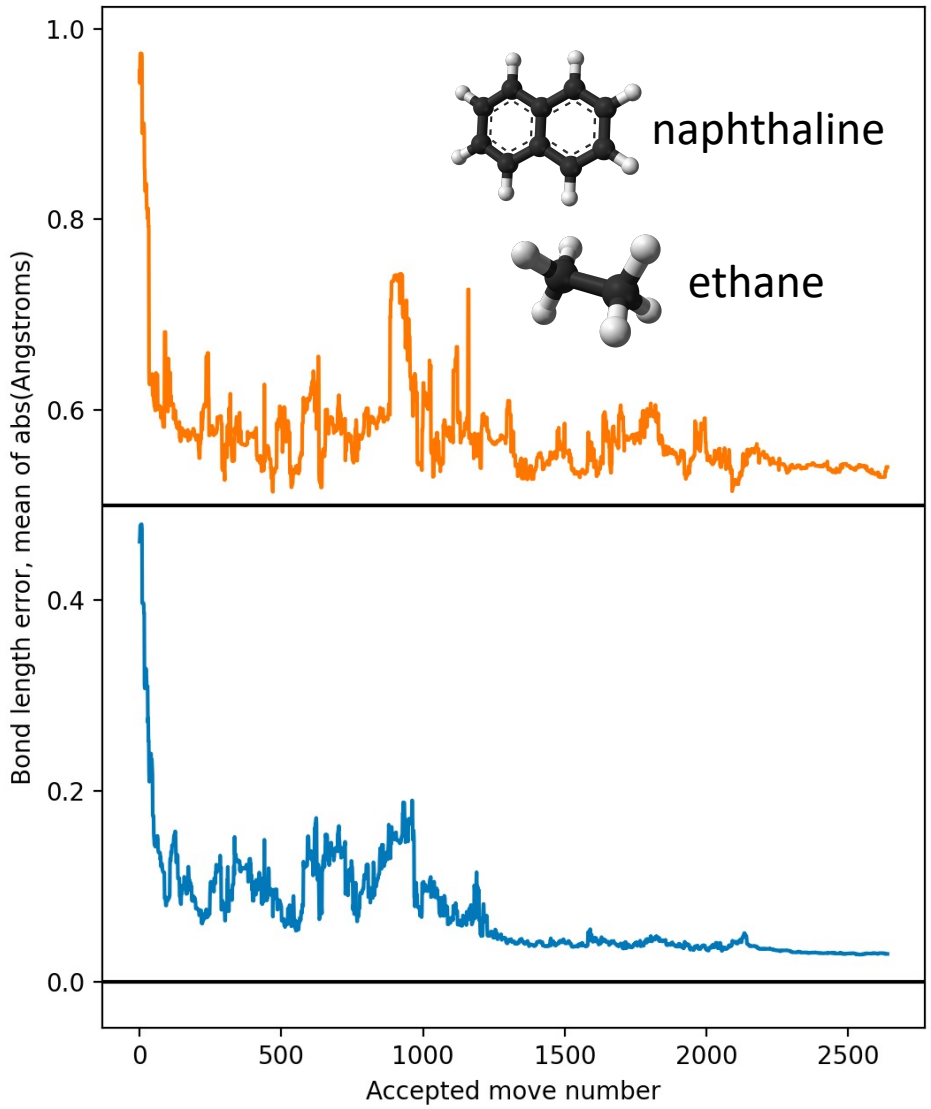

Bond length error, mean of abs(Angstroms)
Bond angle error, mean of abs(degrees)
Accepted move number
Accepted move number
n-heptane
naphthaline
ethane
cyclopropane
cyclohexane
benzene
azulene
anthracene


Full Model
n-heptane
naphthaline
ethane
cyclopropane
cyclohexane
benzene
azulene
anthracene
Bond length error, mean of abs(Angstroms)
Accepted move number
0
500
1000
1500
2000
2500
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

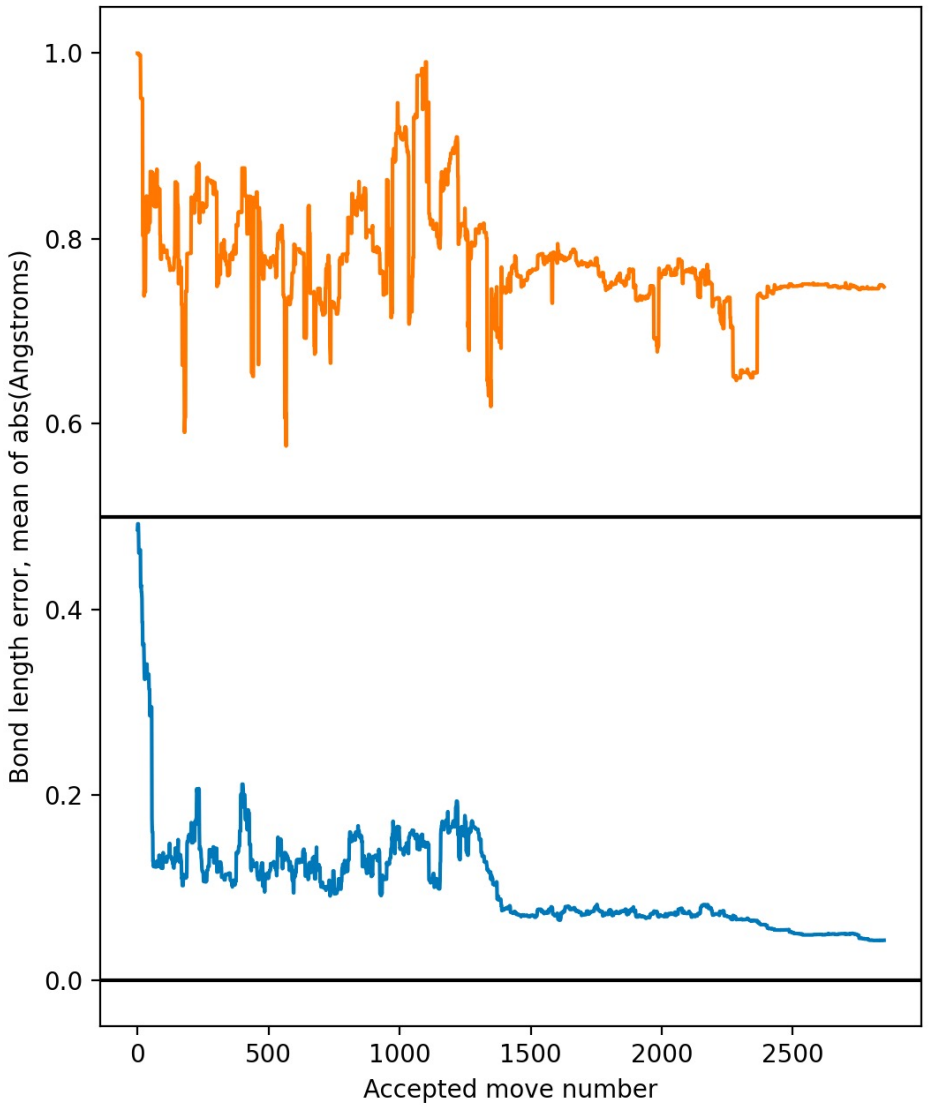

No Hopping (2-body bond type-resolved interactions)
Bond length error, mean of abs(Angstroms)
Accepted move number
0
500
1000
1500
2000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

Evaluating test data

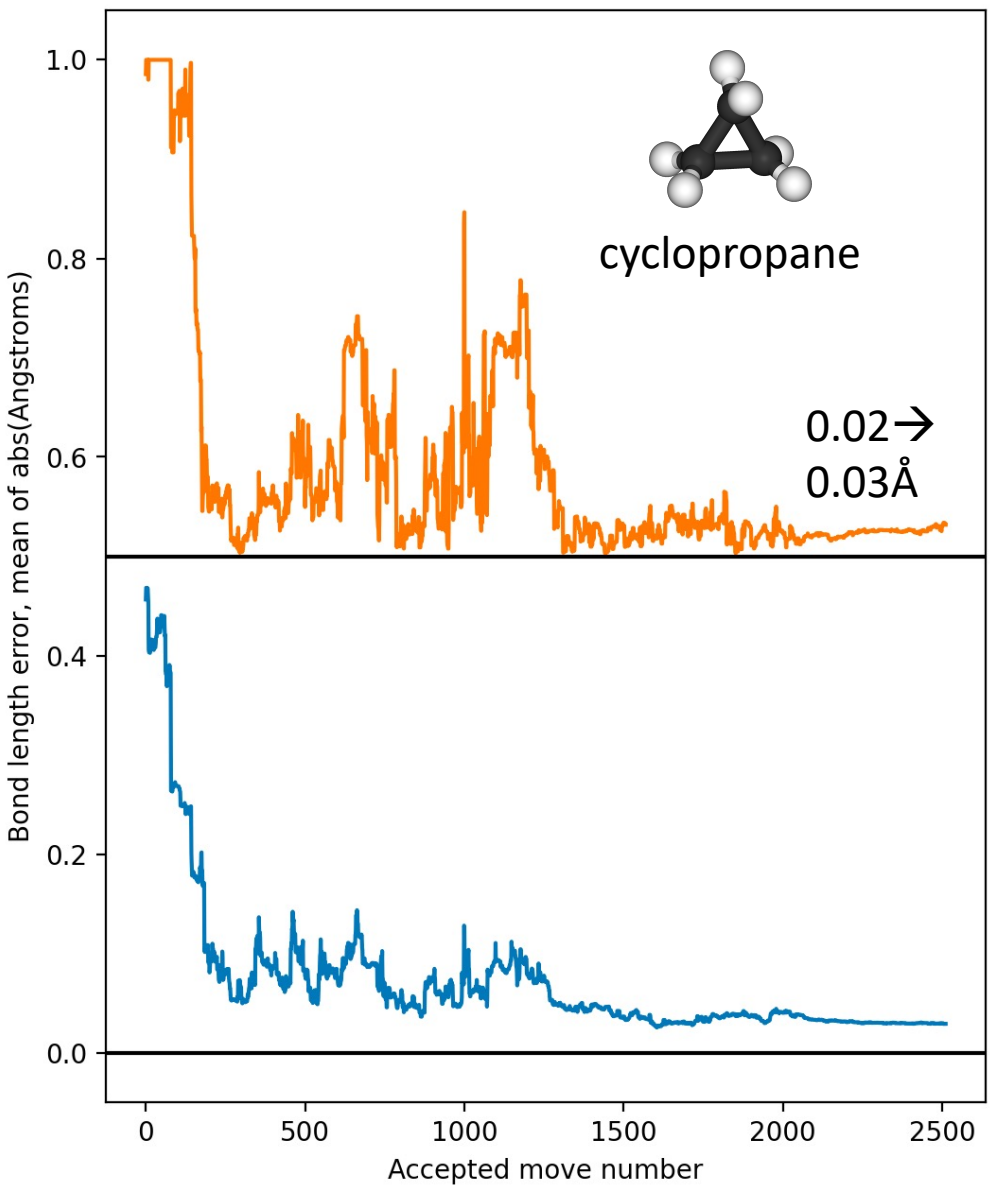
Full Model

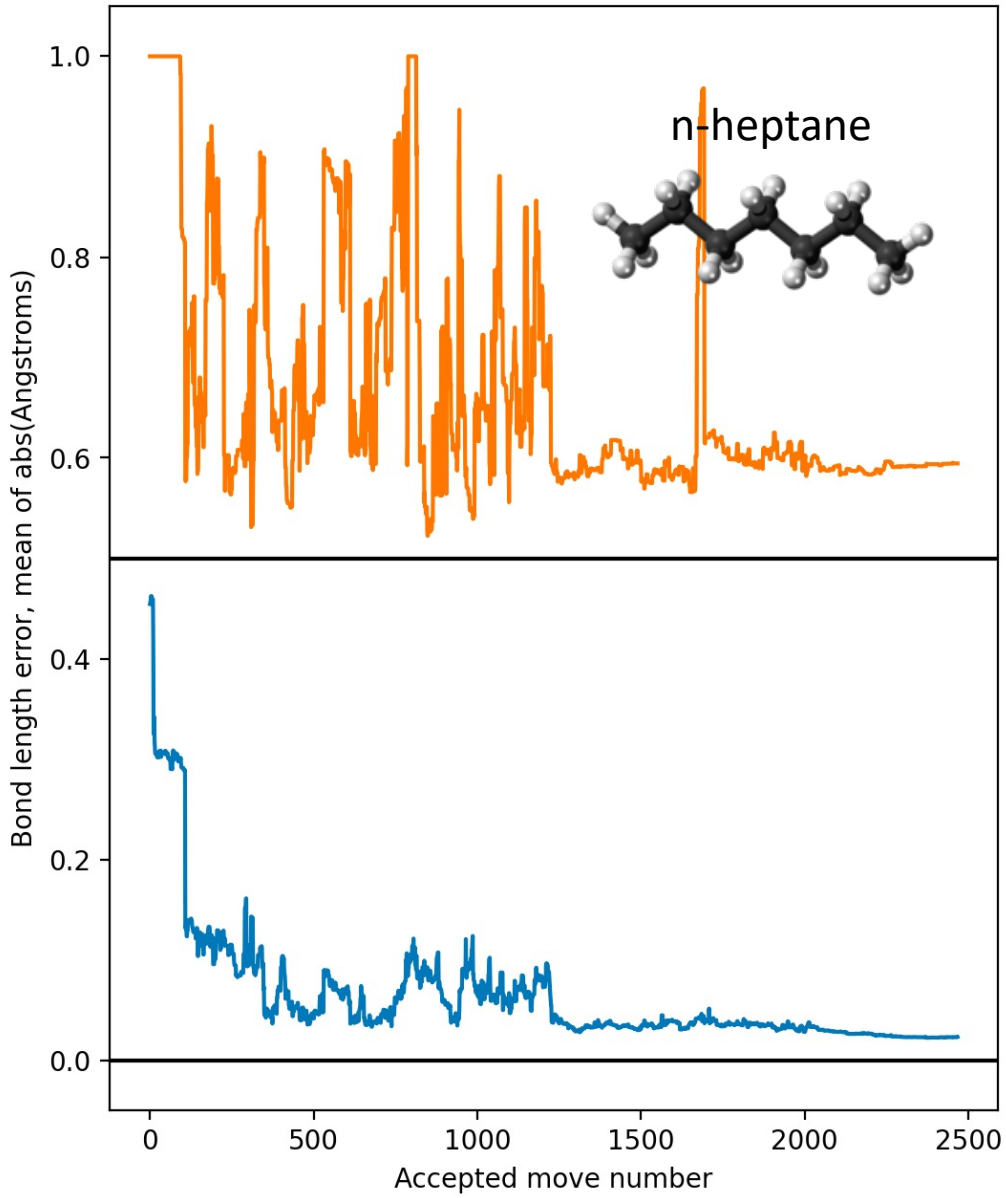


No Hopping



Overfitting the training data





Advantages

1. Can handle delocalized wavefunctions (rings, chains).
2. Speed of **~5 ms/config** for a 20 atom molecule. (time~n^3)
3. Error can be comparable to DFT (~1% of bond length) following Monte Carlo optimization, and possibly superior.
4. If successful scenarios are identified, it can plug into existing use cases for better-than-MM optimization.

Development frontiers/Limits

1. The quantum model is partly missing charge density dependence. Density terms can be added at a cost of ~1.2-10 times the computational time.
2. Fields should depend on whether there is another atom in the way (3-body terms). This is currently mitigated by making the hopping field short range. Limited 3-body terms can be added with negligible computational overhead.
3. Converged TBF parameter sets from different runs can look quite different, but give similar performance. A random forest approach may increase accuracy at the cost of time.

4. As for AI-based approaches, model dimensionality includes terms that grow as the factorial of the number of elements. Adding O and N multiplies dimensionality by 4.6. Going beyond ~6 elements will likely require sequential training for cycles with different groups of elements, even after the AI optimization loop is implemented.

5. Cluster physics is missing. Accuracy may hit a DFT-like limit.

Immediate next steps

1. Full implementation of the neural network loop.

2. Close characterization of model performance, including more manipulations (twists, full structure relaxation.

3. Characterization for nitrogen- and oxygen-bearing molecules.