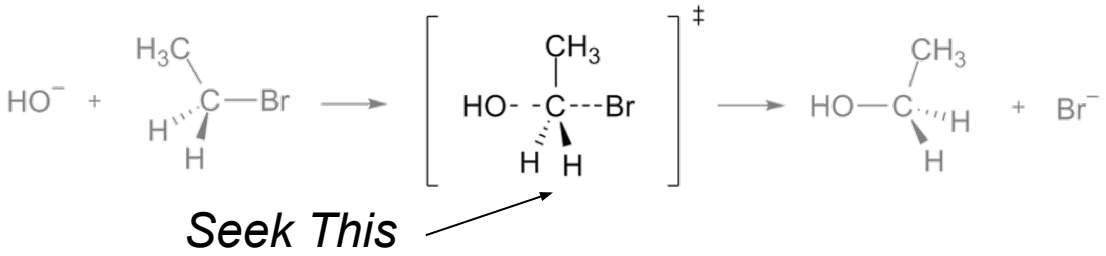
XTBTSScreener.jl

Screening Likely Transition States with Julia & ML

Jackson Burns
MIT 18.337 - Spring 2023

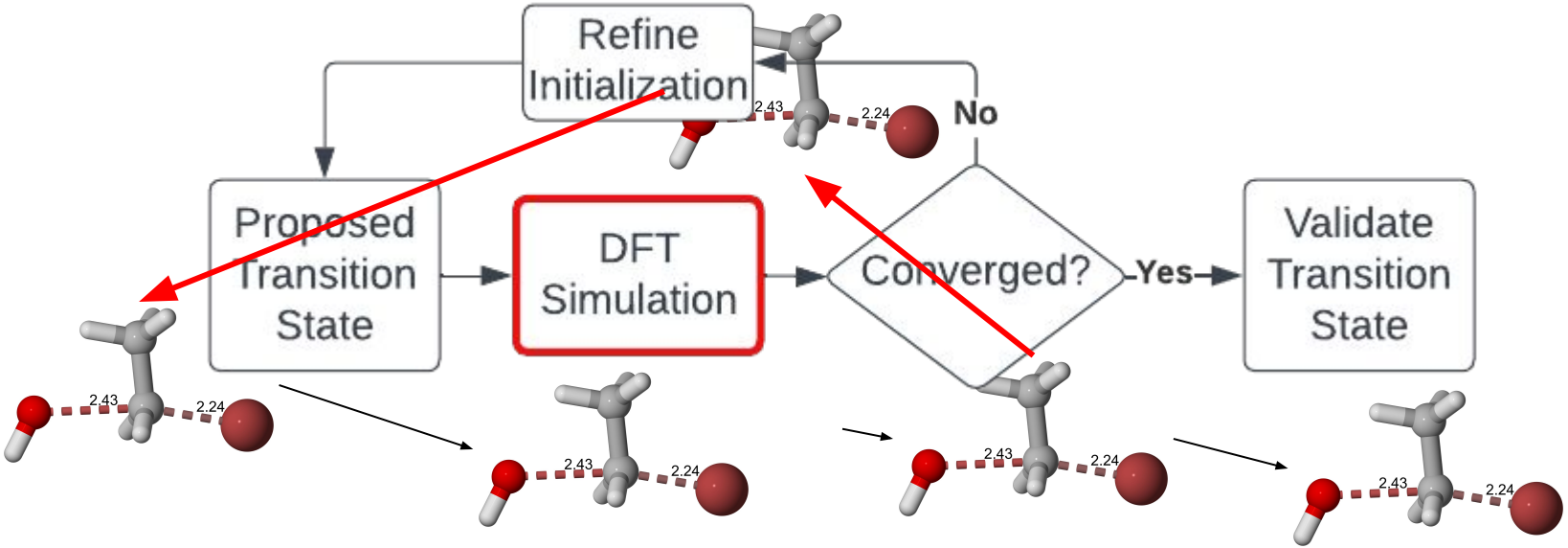
Massachusetts Institute of Technology

Background - Transition States

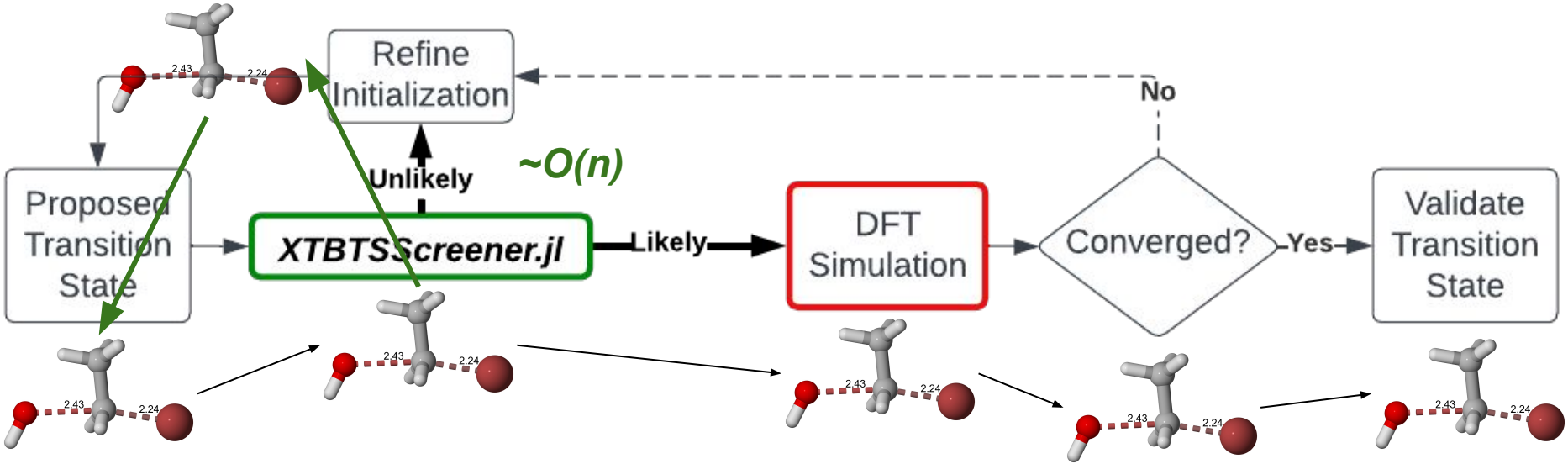


- Chemical kinetics uses transition states to estimate reaction parameters
  - speed of reaction → correlation with temperature → control scheme
- Identifying them is computationally difficult
  - approximations of the hamiltonian
  - scales ***O(n^3)*** to ***O(n^6)*** for *n* electrons (depending on level of theory)

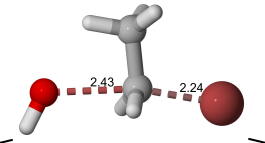
Current Workflow



Enhanced Workflow



Embedding & Architecture



[[Gibbs free energy]
[E_0+ZPE]
[# proposal steps]
[|
- atomic coordinates -
|]]

Software:

Long Short-Term Memory Recurrent Neural Network (LSTM)

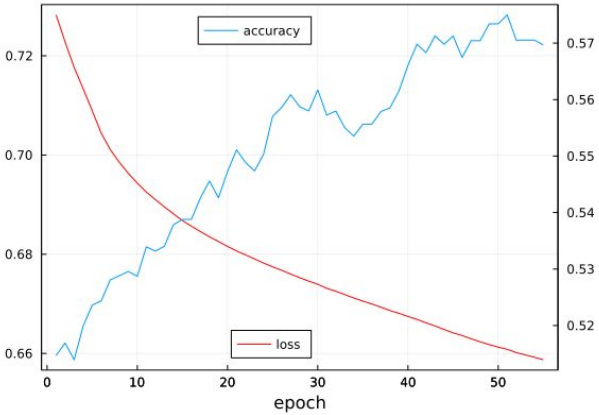
Hyperparameters: learning rate *0.0001*, batch size *64*, *60* input dimensions, *6* hidden dimensions, sigmoid activation, binary cross-entropy loss

Hardware:

MIT SuperCloud Xeon-P8 Partition

- 48 cores & 192 GB RAM

Results



Combined Accuracy & Loss Curves for the LSTM

- Achieved +7% accuracy over mean
- Needs improvement, but:
  - Rejecting highly unlikely guesses outright saves *days*
  - proof-of-concept value
- Simple augmented coordinates can inform a NN

Future Work

- Generate better embeddings
  - Recombinant Neural Network for variable length input
- Alternative network architectures
  - `GraphNeuralNetworks.jl` – established in cheminformatics
- Data preprocessing is embarrassingly parallel
  - mapreduce archetype
- Continue expanding dataset to deal with imbalance
  - especially *negative* examples

Code Walkthrough - Follow Along

- HTML-rendered Jupyter Notebook with all code, explanation, and results

t.ly/-80v