MetalAI: Atomic structure calculations guided by machine learning

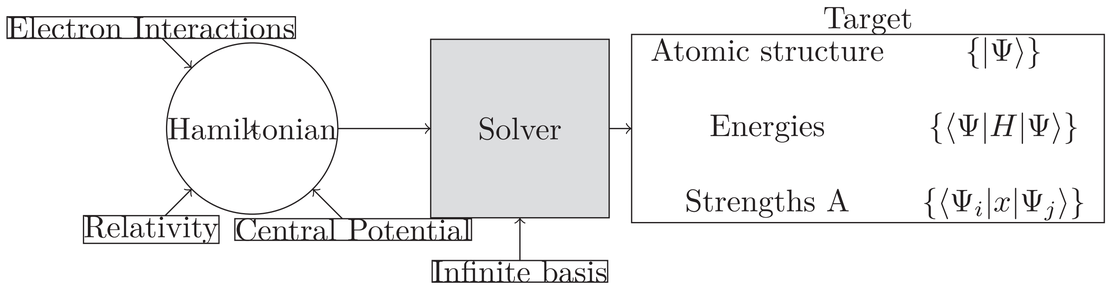
Moust Holmes[1], Rasmus Damgaard Nielsen[2]
[1]Department of Biology, University of Copenhagen
[2]Niels Bohr Institute, University of Copenhagen

1.Motivation

The advent of gravitational wave astronomy has enabled the direct study of the synthesis of heavy elements such as rare earths, lanthanides, and actinides. This was exemplified by the kilonova AT2017gfo/GW170817, where several heavy element emission lines were confirmed. The primary challenge in advancing this research is the lack of accurate and complete atomic data for these elements in ionization stages I-IV.

2. GRASP and Atomic structure calculations

- Atomic structure:
  In theory a solved problem
- In practice horrible scaling $\sim O(n^3)$ with vector space size
- "The magic" is choosing the right subspace = set of CSFs



- CSF: The used family of basis vectors, spanning all bound states Defined by number of electrons in each orbital

3. The Learning Problem

Goal:

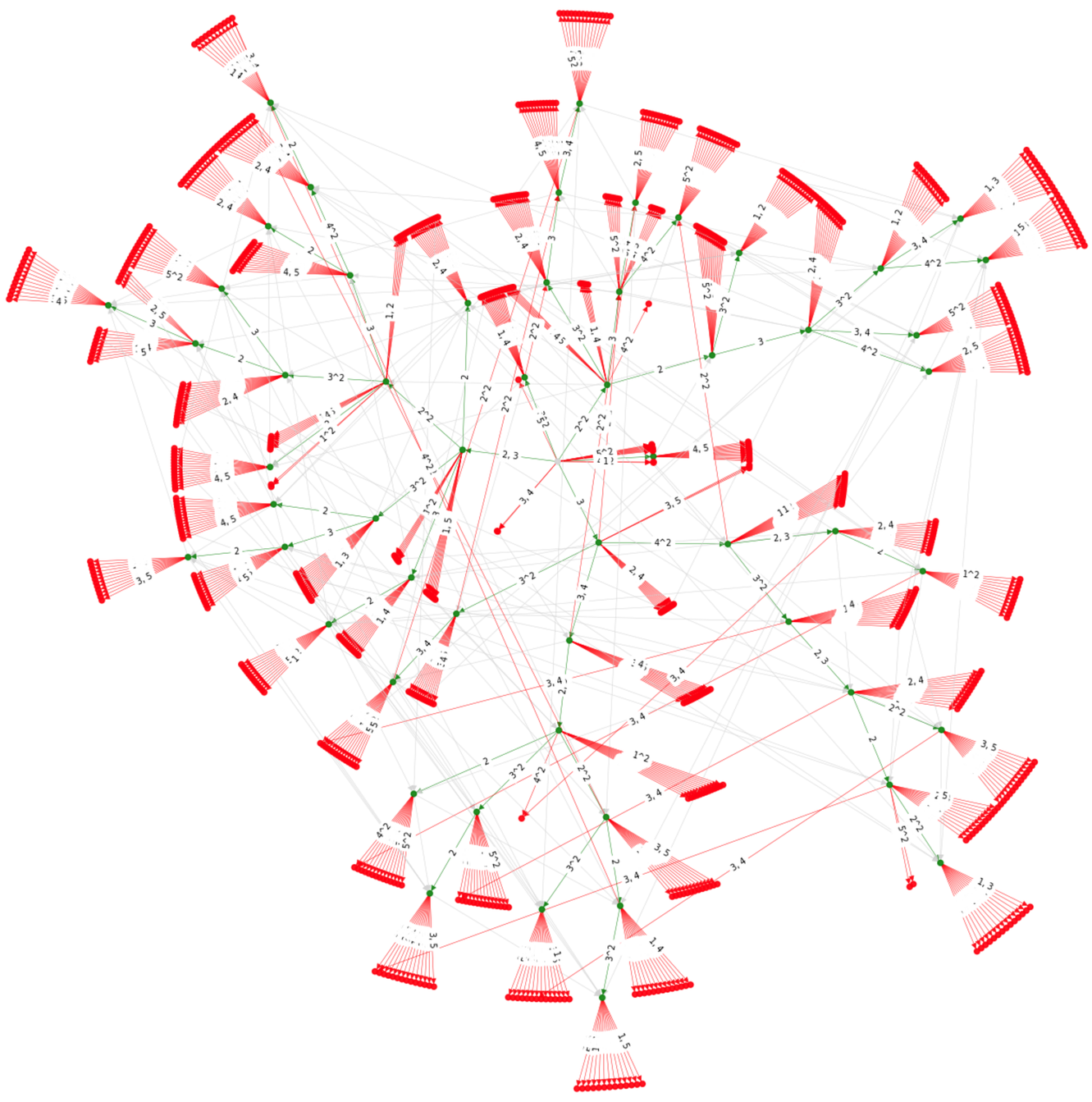
- Predict if adding a new CSF to a set of CSFs will lead to convergence
- Use convergence predictions to navigate the graph of possible CSFs.
- Use trained model to predict which CSF to add to create more training data

Data:

- 132000 sets of CSF calculations
- 160 ions from lithium to silver
- only 5% of our allocated compute time (EuroHPC development∼500.000 CPU-hr).
- No ML guided CSF choice.
- 50/50 converged and non-converged

Extrapolation over Interpolation:

- Iteratively adding new CSFs creates highly correlated data.
- Validation will be on sets of CSFs larger than what the model was trained on.
- Train first on small sets and gradually increase the size to bias towards extrapolation.



Converged and non-converged CSF sets for Yttrium I. Nodes are green if at least one edge to it converged and red if none. Grey node in center is starting ground state

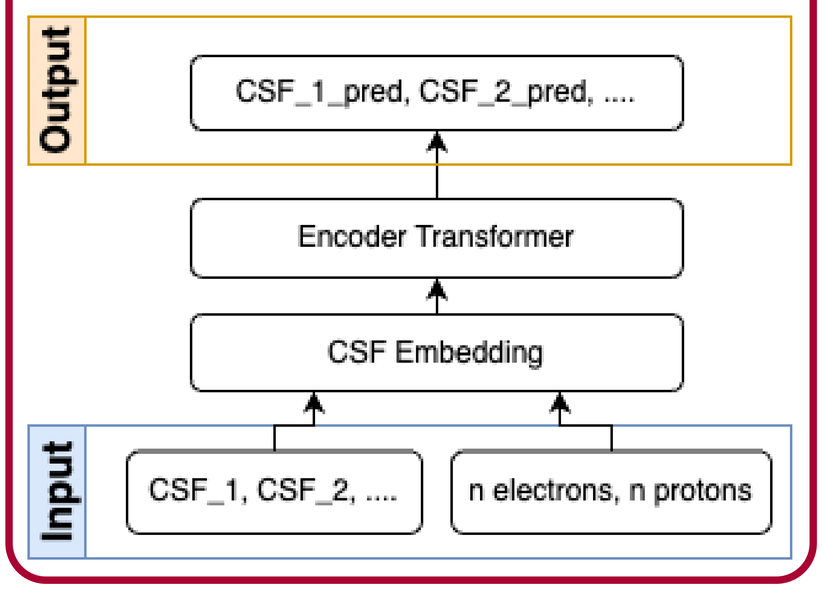
4. Machine Learning

Model architecture:

- Predict if adding a new CSF to a set of CSFs will lead to convergence.
- Reformulate the problem for training: set of CSFs as input and predict for each CSF if adding it to the subset without it would converge.
- Use a Transformer model due to the set-based nature of the input.

CSF embedding:

- CSFs needs to be converted to a fixed size vector before being passed to the transformer model.
- Excitation representation: limited number of perturbation of the filled ground state. e.g. [2,1] would mean the outermost electron have been moved 2 spaces out and the next outermost electron have been moved 1 space out.
- Incorporate ion information in embedding



4.Convergence Predictions

Best validation accuracy 73%. Validated on 15% of largest sets.