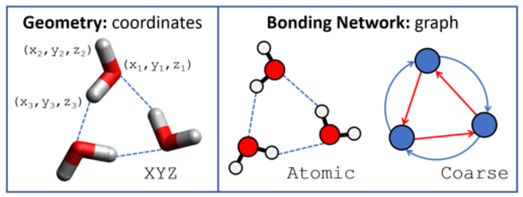
HydroNet: Benchmark Tasks for Preserving Intermolecular Interactions and Structural Motifs in Predictive and Generative Models for Molecular Data

Sutanay Choudhury, Jenna A. Bilbrey, Logan Ward, Sotiris S. Xantheas, Ian Foster, Joseph Heindel, Ben Blaiszik and Marcus Schwarting

I. INTRODUCTION

- Intermolecular and long-range interactions are central to phenomena as diverse as gene regulation, topological states of quantum materials, electrolyte transport in batteries, and the universal solvation properties of water.
- We present a set of challenge problems for preserving intermolecular interactions and structural motifs in machine-learning approaches to chemical problems, through the use of a recently published dataset of 4.95 million water clusters held together by hydrogen bonding interactions and resulting in longer range structural patterns.
- The dataset provides spatial coordinates as well as two types of graph representations, to accommodate a variety of machine-learning practices.



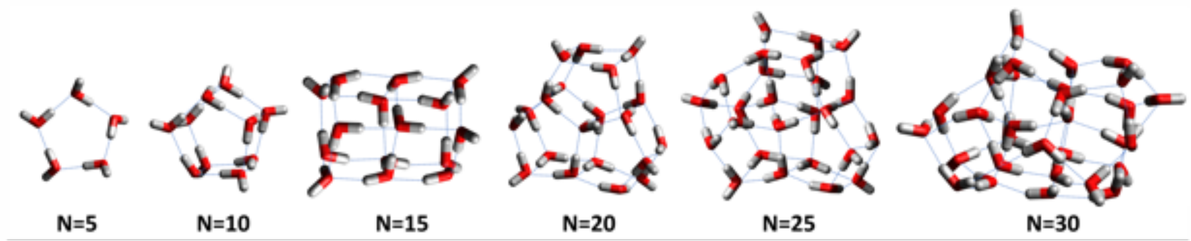
II. DATASET

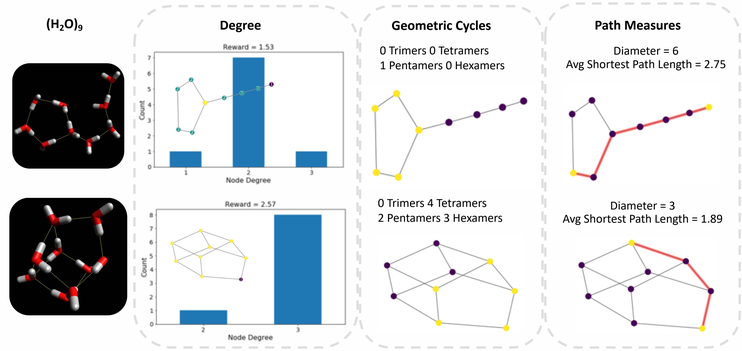
- The dataset of 4.95 million water cluster minima is the largest collection of water cluster minima reported to date

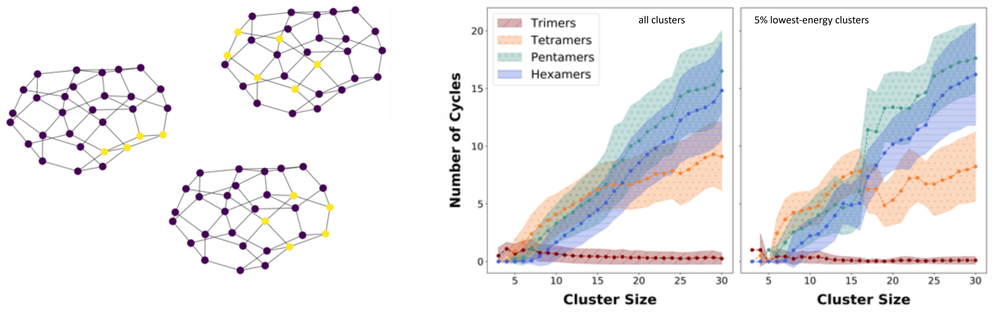
https://exalearn.github.io/hydronet/


```
{
   "z": [8, 1, 1, 8, 1, 1, 8, 1, 1],
   "n_water": 3,
   "n_atom": 9,
   "atom": [0, 1, 1, 0, 1, 1, 0, 1, 1],
   "coords": [[25.3875809, 2.28446364, 8.01933861],
              [24.686451, 2.11461496, 7.36908807],
              [26.1070786, 1.70453322, 7.77935553],
              [22.9643402, 1.68695939, 6.75715494],
              [22.7494984, 1.67431045, 7.70416498],
              [22.2382431, 2.13693213, 6.33168697],
              [23.0780773, 1.86950338, 9.5477314],
              [22.9238548, 2.4637537, 10.2781725],
              [23.9850082, 2.04813766, 9.2500248]],
   "energy": -15.9416428
}
```


III. GRAPH-THEORETIC DESCRIPTORS FOR THE DATASET







IV. TASK-I: POTENTIAL ENERGY PREDICTION

This task is to predict the potential energy of a water cluster without the use of expensive ab initio methods, in two settings. In the geometry-to-energy setting, we assume the geometry of the structure is known accurately. In the graph-to-energy setting, we require predictions to be made from the connectivity of the water molecules only.

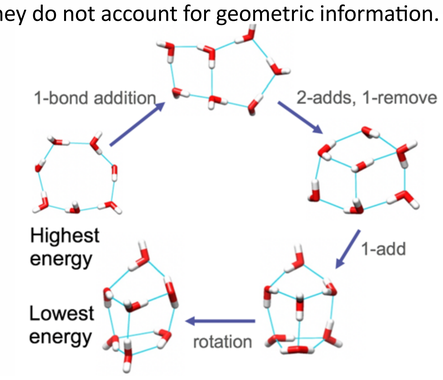
$$m_v^{t+1} = \sum_{w \in Nbors(v)} M_t(h_v^t, h_w^t, \alpha_{vw}^t) \qquad h_v^{t+1} = h_v^t + m_v^{t+1} \qquad \alpha_{vw}^{t+1} = \alpha_{vw}^t + M_t(h_v^t, h_w^t, \alpha_{vw}^t)$$

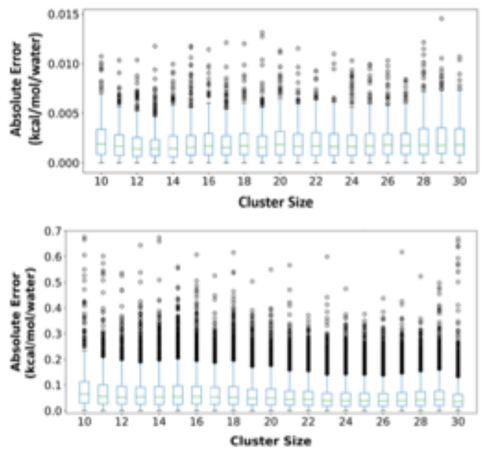
TASK-II: STRUCTURAL MEASURE-PRESERVING MOLECULAR GENERATION

We define the task as follows: given *N* water molecules, generate a geometric/atomic/coarse representation that a) satisfies certain graph-theoretic structural measures derived from its hydrogen bonding network and b) minimizes the cluster energy by optimizing the relative spatial arrangements of atoms and molecules.

V. INTEGRATING GRAPH AND SPATIAL INFORMATION

Both tasks require awareness of geometric and structural information, especially in the context of minimum energy water clusters. The performance of graph neural networks is limited when they do not account for geometric information.





REFERENCES

1. Jenna A. Bilbrey, Joseph P. Heindel, Malachi Schram, Pradipta Bandyopadhyay, Sotiris S. Xantheas, and Sutanay Choudhury. A look inside the black box: Using graph-theoretical descriptors to interpret a Continuous-Filter Convolutional Neural Network (CF-CNN) trained on the global and local minimum energy structures of neutral water clusters. The Journal of Chemical Physics, 153(2):024302, 2020.
2. Frank N Keutsch and Richard J Saykally. Water clusters: Untangling the mysteries of the liquid, one molecule at a time. Proceedings of the National Academy of Sciences, 98(19):10533–10540, 2001.
3. Avijit Rakshit, Pradipta Bandyopadhyay, Joseph P. Heindel, and Sotiris S. Xantheas. Atlas of putative minima and low-lying energy networks of water clusters n = 3–25. The Journal of Chemical Physics, 151(21):214307, 2019.
4. K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller. Schnet- pack: A deep learning toolbox for atomistic systems. Journal of Chemical Theory and Compu- tation, 15(1):448–455, 2019.
5. Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. arXiv preprint arXiv:1704.01212, 2017.

Pacific Northwest NATIONAL LABORATORY · Argonne NATIONAL LABORATORY · University of Washington · The University of Chicago