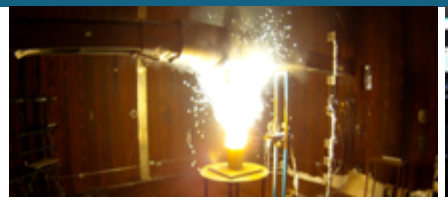
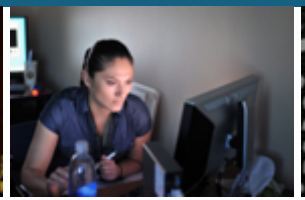
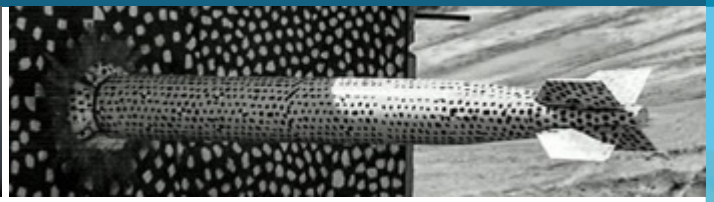
SAND2021-3233C

Sandia National Laboratories

Using Machine Learning Approaches to Predict Atomic-Scale Glass Failure in Environmental Conditions







Presented By

Victoria Lloyd, Harvey Mudd College

March 19, 2021

APS March Meeting 2021

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Co-authors

Claremont Colleges:

- Victoria Lloyd (Harvey Mudd College)
- Sarah Lu (Institute of Mathematical Sciences, Claremont Graduate University)
- Jorge Pena (Institute of Mathematical Sciences, Claremont Graduate University)
- Allon Percus (Institute of Mathematical Sciences, Claremont Graduate University)
- Cora Wang (Institute of Mathematical Sciences, Claremont Graduate University)
- Rory Zhao (Harvey Mudd College)

Sandia National Laboratories:

- Thomas Hardin (Computational Materials and Data Science, Sandia National Laboratories)
- Mark Wilson (Computational Materials and Data Science, Sandia National Laboratories)

Background

A fundamental challenge is to make reliable predictions about where failure occurs in order to gain insight into the physical structures associated with failure.

Prior work has used classical ReaxFF molecular dynamics (MD) simulations to model fracture nucleation of silica-based glasses in aqueous environments, but developing structure-property relationships using this simulation method is computationally expensive.

This project aims to develop a **surrogate model** to make connections between atomic structure and fracture nucleation.

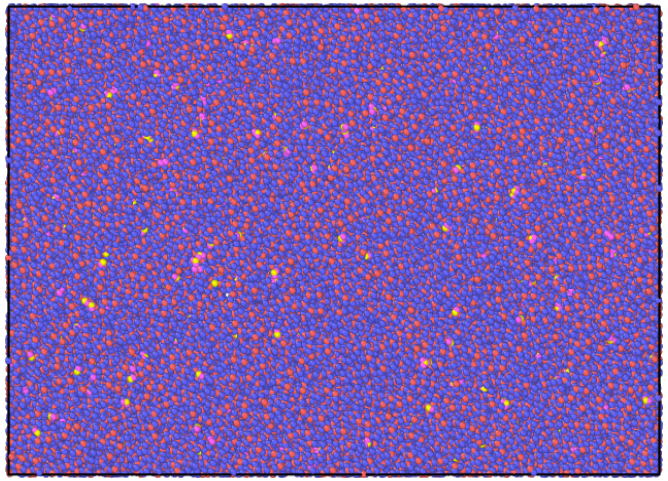
By creating a graph theoretic description of the material we train a supervised **Machine Learning** algorithm on MD simulation data.

Molecular Dynamics Simulations

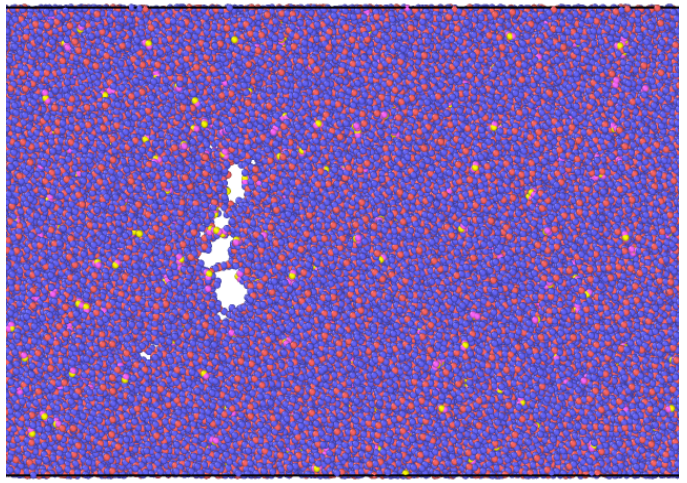
- Previously we have used molecular dynamics (MD) to successfully predict a wide range of detailed properties that cannot be obtained with continuum methods
- Simulations are of size $15 \times 15 \times 4$ nm^3, containing approximately 75,000 atoms
- Uniaxial tension at a constant strain rate of 5×10^8 s^{-1} in the horizontal direction for 1 ns
- Two types of conditions for simulations
  - dry: SiO_2 in vacuum
  - wet: SiO_2 immersed in H_2O
- Simulation output provides physical descriptors, such as:
  - x, y, z coordinates for each atom
  - bond configuration
- 200 simulations, approximately 10GB of data per simulation

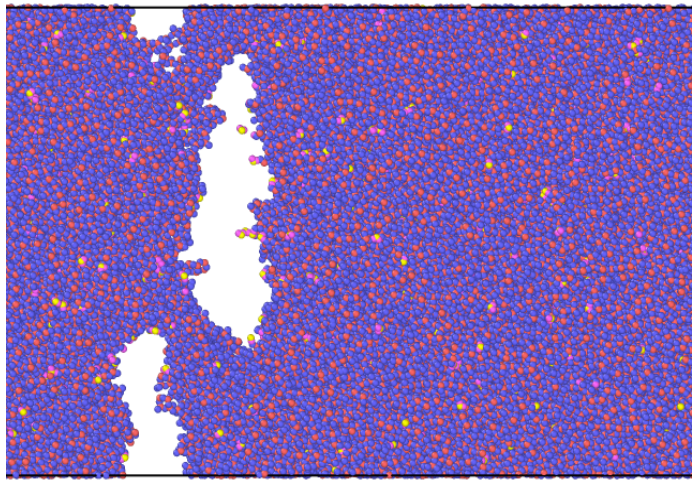
Simulation of Fracture Nucleation

t=580ps

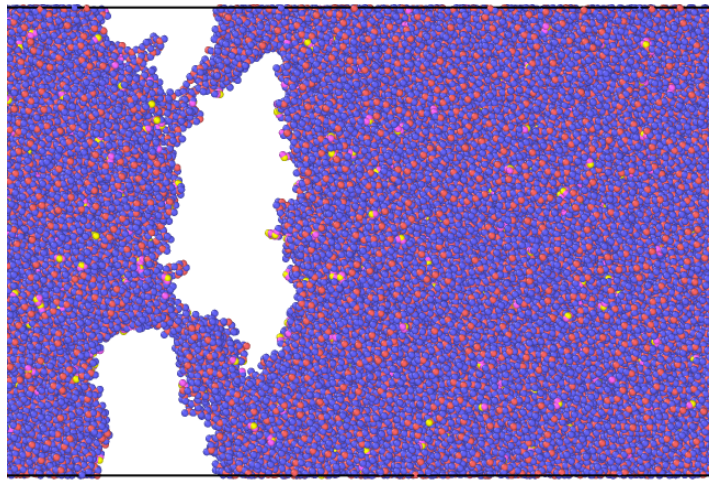


t=640ps





t=690ps



t=750ps

Our Current Objectives

- Computation time for a single run is ≈10,000 CPU hours, making it difficult to study systems directly via simulations
- We aim to develop supervised learning methods, trained on MD simulation data, that generate rapid predictions of where and when atomic-scale fractures occur in samples of **silicate glasses in an aqueous environment** under stress.
  - Generate predictions under multiple environmental conditions.
  - Validate on existing MD simulation results.
- Relate predictions to specific features characterizing local atomic structure.
- Provide new insight into how local structure leads to fracture and failure.
- Examine the impact of aqueous environments on fracture nucleation.

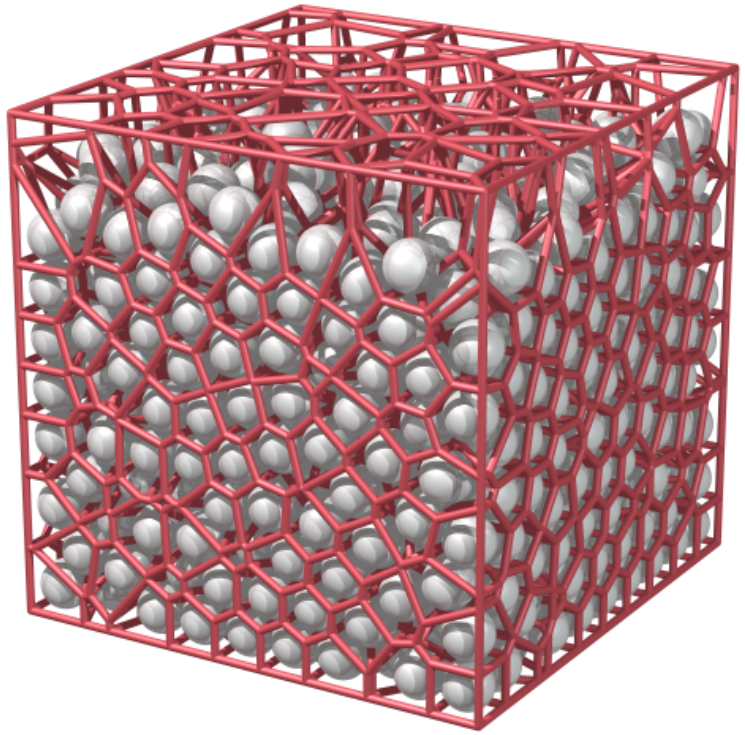
Machine Learning Approach

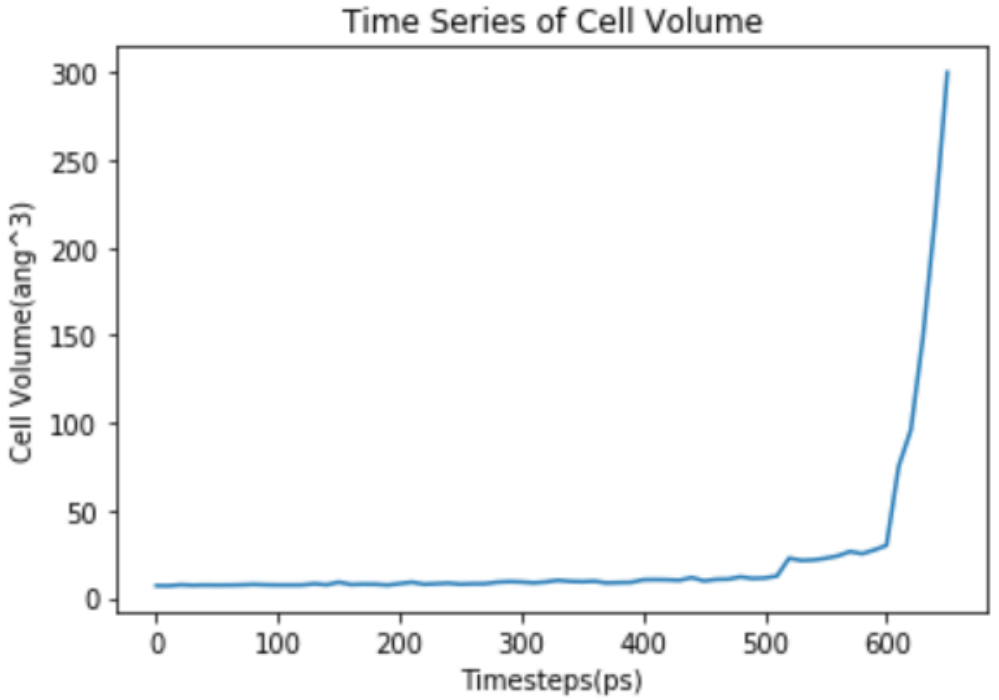
In this project we are training machine learning models using the following features which are obtained using MD simulations:

- Physical Features
  - Cell volume
  - Atom displacement
  - Stress tensor
  - Local elasticity
  - Kinetic and potential energy
- Topological Features
  - Coordination number
  - Number of bridging oxygens
  - k-neighborhood
  - Bond activity
  - Bond angle

Example of Physical Features

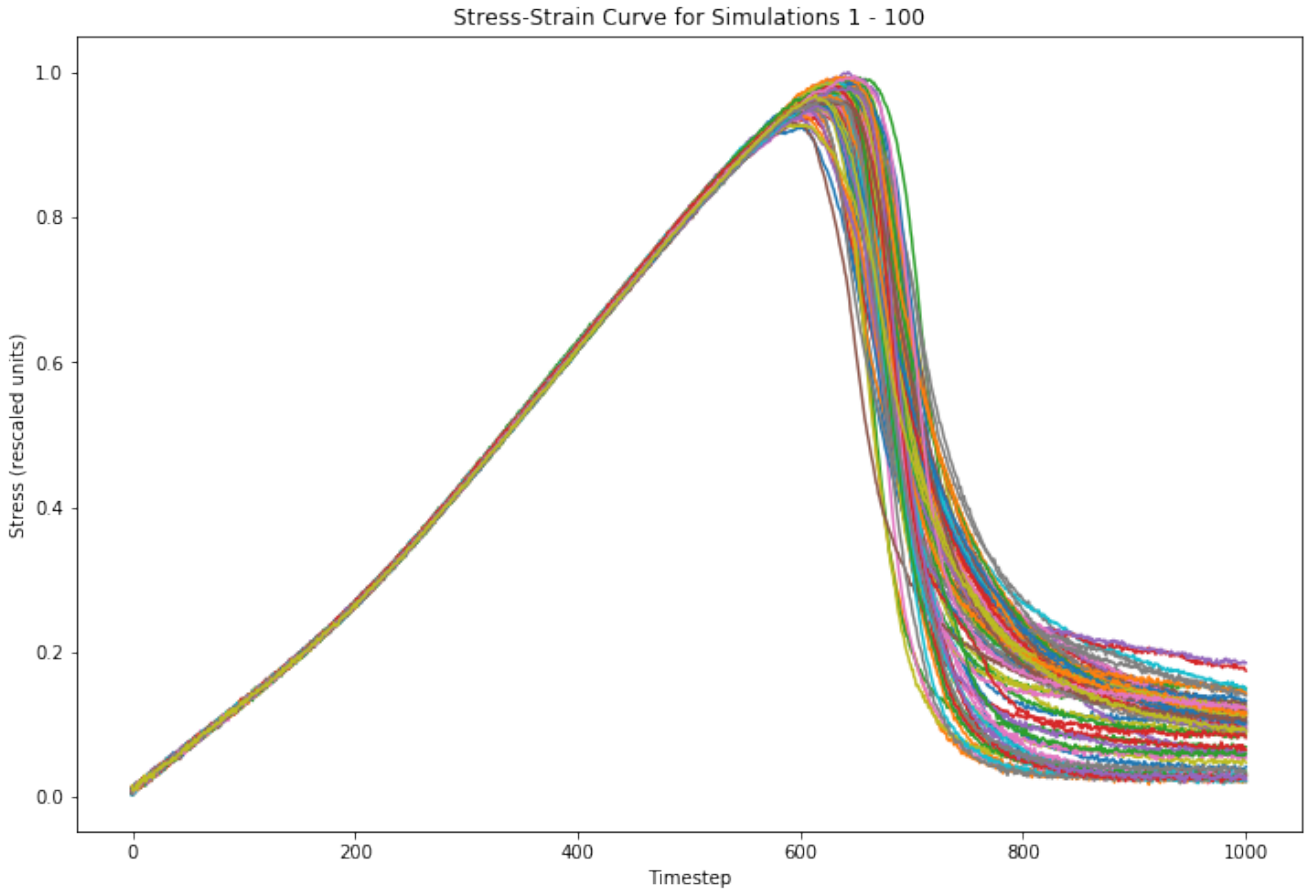
The **Voronoi cell volume**, v_i, is a local density measure. It measures how large the empty space is surrounding an atom *i*.





Example of Physical Features

- Elasticity is given by the slope of the stress-strain curve.
- Stress may not increase uniformly over all regions when placed under load.
- **Local elasticity**, as defined through an appropriate level of local spatial averaging, may serve as an indicator of physical changes leading to nucleation.



Example of Topological Features

Number of Bridging Oxygens: In a Q_n unit, an Si atom is surrounded by *n* bridging O atoms, each forming an Si–O–Si group. The value of *n* supplies crucial network connectivity information.

- With the reduced graph representation, the number of bridging oxygens is equivalent to the degree of a vertex
- ML algorithms may benefit from having node degree as an explicit node-based feature

***k*-neighborhood:** A crucial assumption in our modeling approach is that structure beyond nearest-neighbor information can help in predicting fracture nucleation.

- For example, total number of Si atoms within *k* bridging oxygens of a given Si atom without any hydrogen bonds.
- For increasing *k*, these features consider local neighborhoods of increasing size.

Machine Learning Methods

Objective is to produce surrogate model of MD simulations using supervised learning.

- **Static approach:** Use **linear regression** to determine feature importance.
- **Dynamic approach:** Use **recurrent neural networks** (RNN) to learn dynamics leading to nucleation.

Logistic Regression

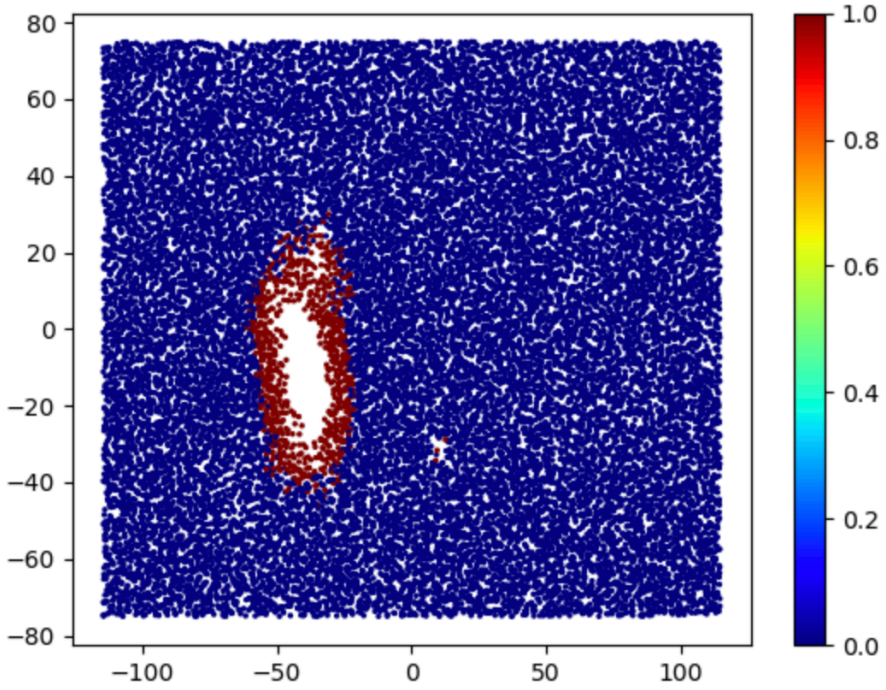
We aim to estimate relative feature importance quickly and easily before using more computationally intensive modeling techniques. To do so, we use:

- Target identification: uses the volume of the entire Q_n unit thresholded to identify only atoms within the region of the crack at nucleation time T
- Features: uses physical and topological at a fixed time $t < T$ as described above

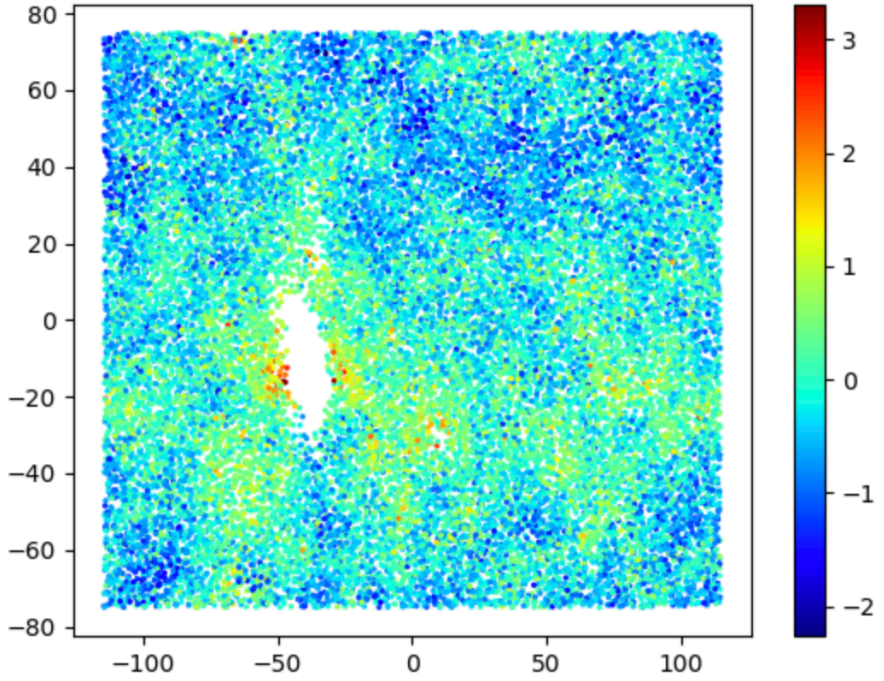
This allows us to quickly determine the best methods to increase the predictive power of the RNN.

Logistic Regression

Because of the significant class imbalance – only ~1% of atoms lie on the crack face at t=650 – we spatially averaged the features and target definition and lowered the threshold to include a larger neighborhood around the crack.



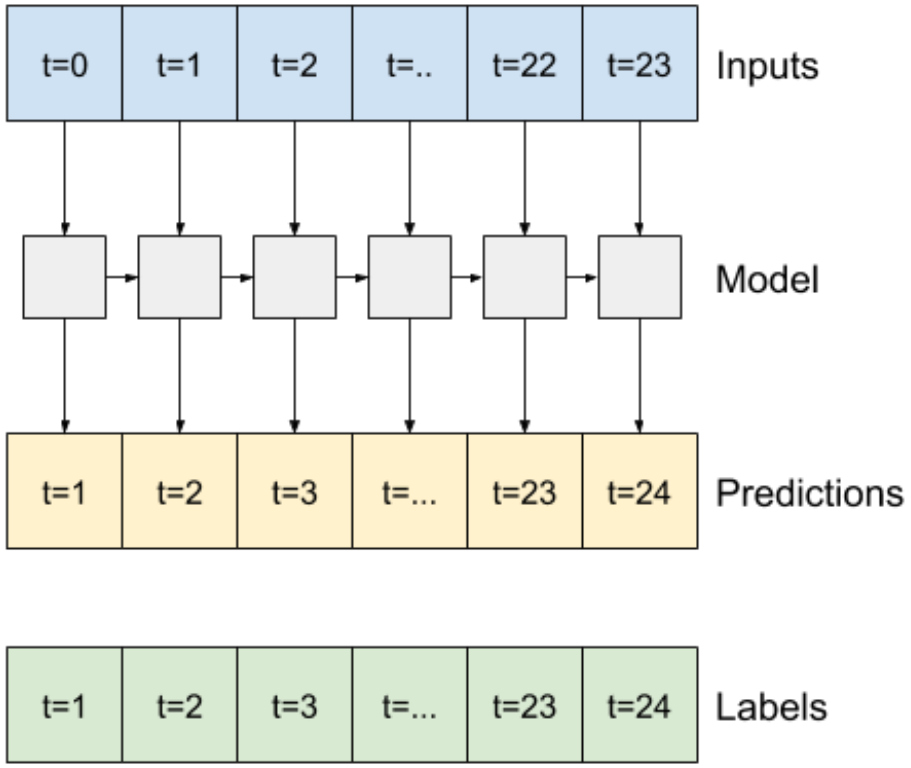
Thresholded Atoms



Predicting 650ps from 600ps

Dynamic Approach: Recurrent Neural Network

RNN trains on an entire time series, learning dynamics of a process which it stores using a "memory" of internal states.



Training input is time series of features from MD simulation data.

Dynamic Approach

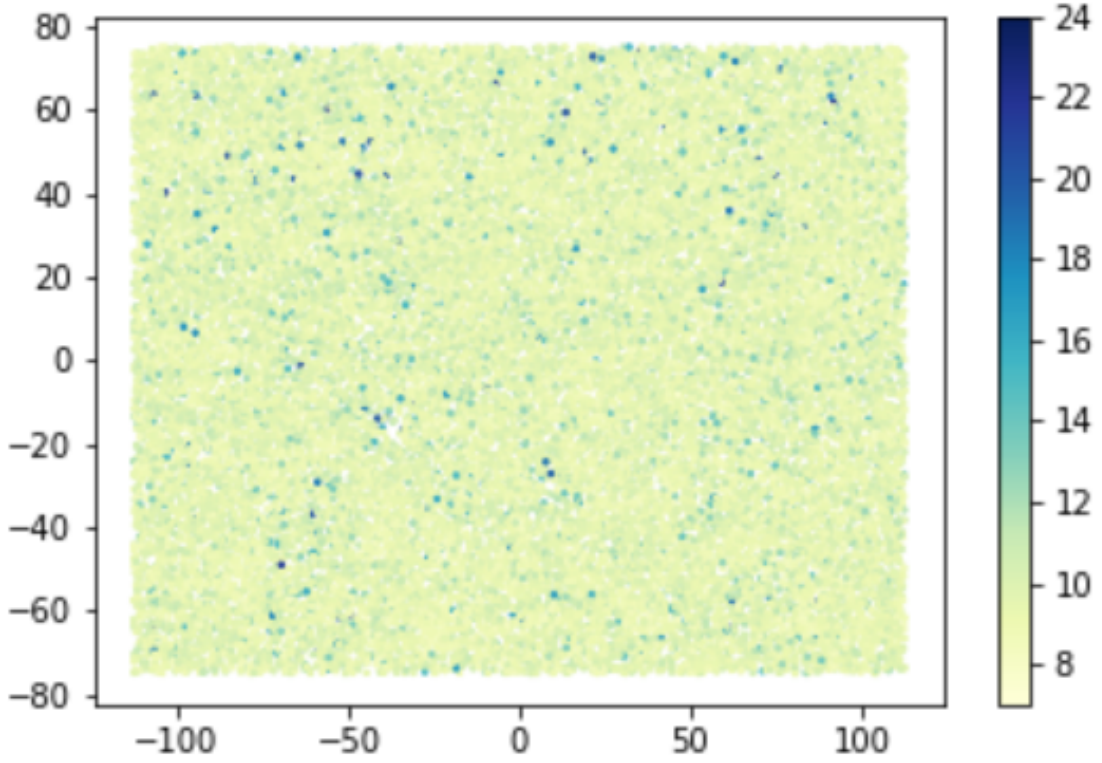
The ultimate goal of this approach is to generate a model which learns the time evolution of the system and, given some early time series for a different system, can generate accurate predictions about its time evolution during nucleation.

Our current model trains the model on each of the atoms individually with the feature(s) for each atom as input & output

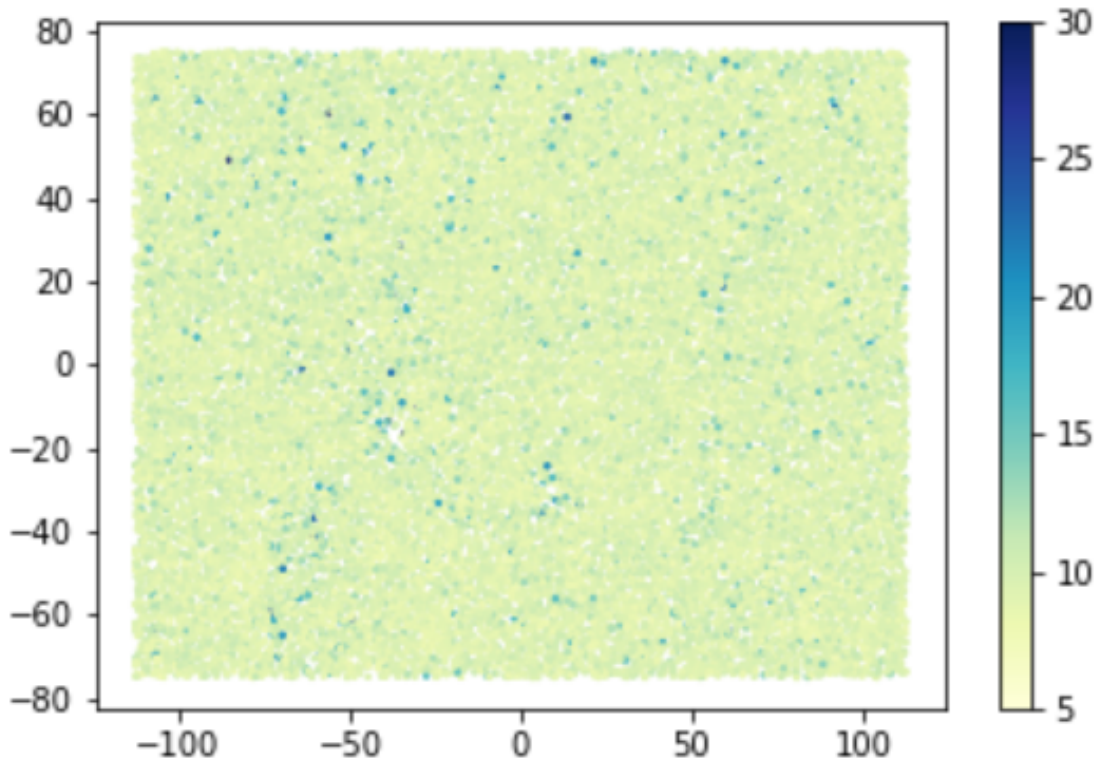
- Train with time steps 0-590ps as warmup, predict 600ps
- Test with time steps 0-600ps as warmup, predict 610ps
- Batch size is all atoms in a single simulation, model has a single layer with hidden size 64, and trains for 1500 epochs

Given that each atom has graph topological features associated with it, training on individual atoms should be sufficient for capturing information about the graph structure.

Prediction Generated from Time Series



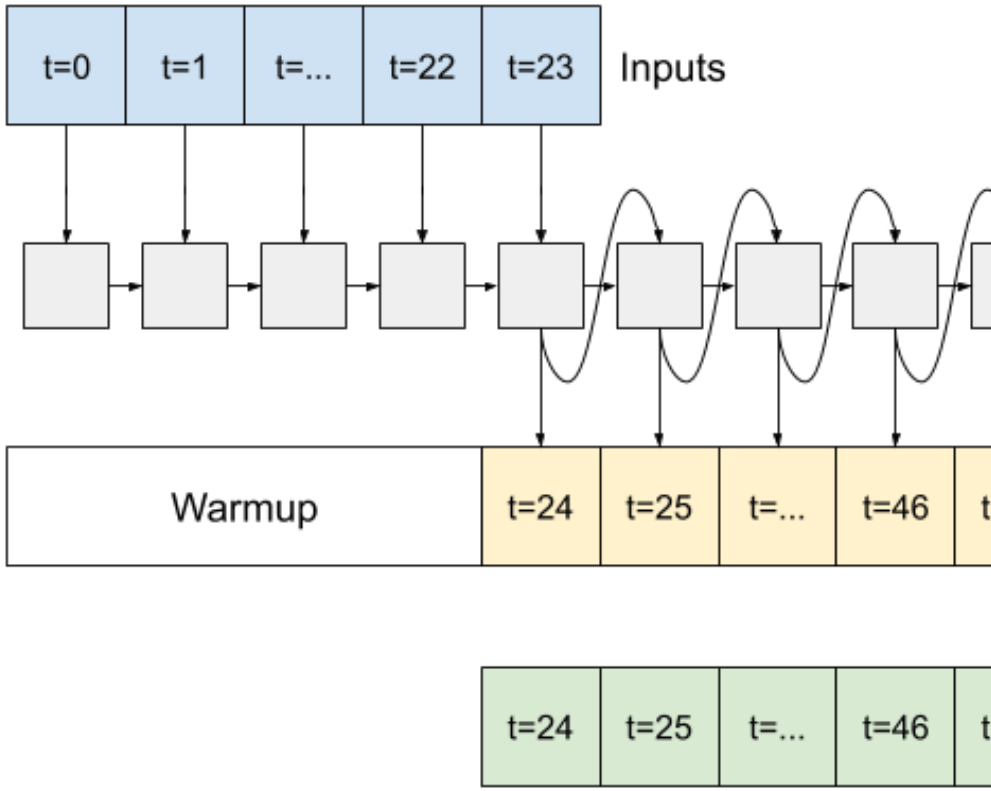
Predicted Cell Volume at t=610



Actual Cell Volume at t=610

Autoregressive Recurrent Neural Network

An autoregressive RNN predicts individual time steps which are used recursively to make predictions about subsequent time steps.



Training input is time series of features from MD simulation data.

Summary & Next Steps

- Logistic regression: machine learning methods are capable of generating meaningful predictions about where fracture nucleation will occur using features at earlier time steps
- Recurrent Neural Network: including the entire time series in our model should enhance the predictions made by logistic regression, and has demonstrated the ability to accurately predict the system at subsequent time steps
- Autoregressive Model: moving forward, we should be able to recursively generate new predictions spanning the ~50ps where nucleation generally occurs in order to determine where and when fractures nucleate when stress is applied to a system
- Using the autoregressive model to study fracture nucleation, we may examine how local atomic structure influences fracture and failure much more quickly than using MD simulations directly

Thank You!