SAND2021-15366C

Sandia National Laboratories

Exceptional service in the national interest

Low-Dimensional Manifolds Underpin the Structure of Glassy Materials

Thomas J. Hardin (Sandia National Laboratories)

Mark Wilson (Sandia National Laboratories)

Michael Shields (Johns Hopkins University)

with thanks to

Michael Chandross (Sandia National Laboratories)

Tess Smidt (Massachusetts Institute of Technology)

Yannis Kevrekidis (Johns Hopkins University)

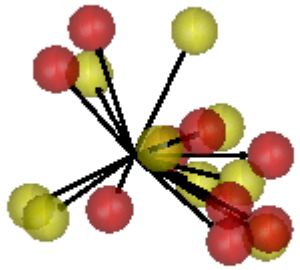
Michael Falk (Johns Hopkins University)

December 6, 2021 ● 2021 MRS Fall Meeting & Exhibit

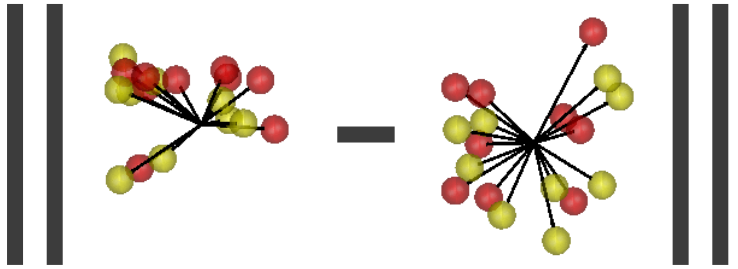
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

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.

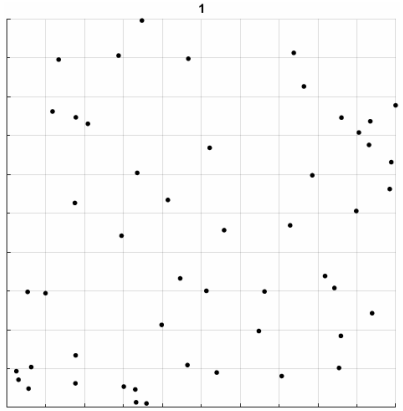
1 Atomic neighborhoods w/cutoff radius



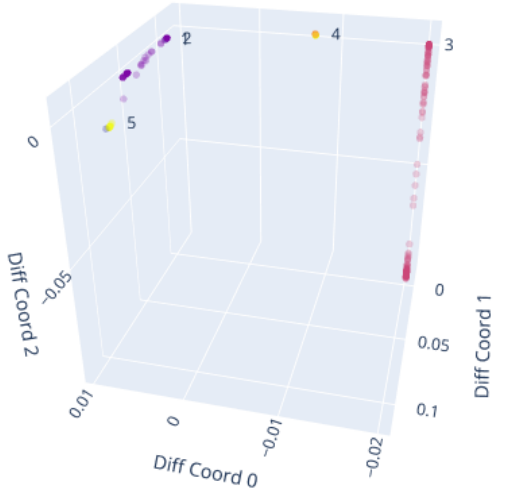
2 Distance measure: Gaussian Integral Inner Product



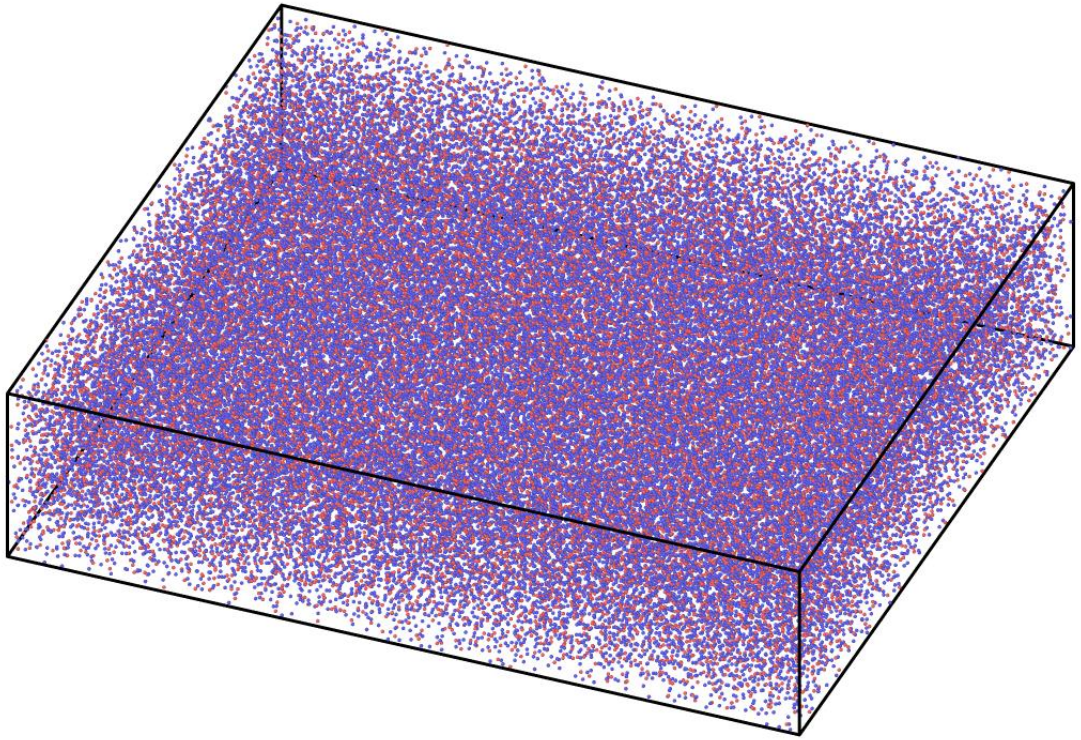
3 Hierarchical (agglomerative) clustering

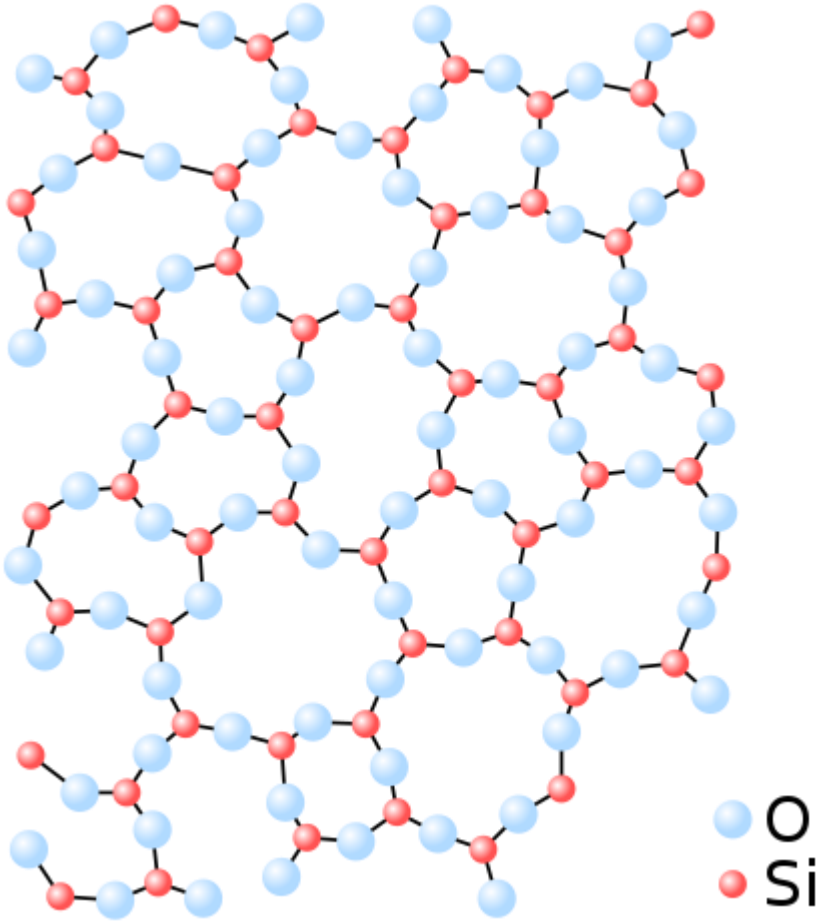


4 Diffusion Maps for low-dimensional latent space



1. Extraction of neighborhoods from silicate glass





- ReaxFF SiO potential (van Duin et al, 2003)
- 71k atoms: Quartz -> Melt -> Quench @ 3.7K/ps
- Averaged over first 500 steps (not relaxed)
- 5000 random Si or O atoms
- First nearest neighborhoods extracted

2. A Meaningful Distance Metric Between Atomic Neighborhoods

Integral inner product of two functions

(1) $$\langle a, b \rangle = \int_{\mathbb{R}^3} a(\mathbf{x}') \cdot b(\mathbf{x}') \, d\mathbf{x}'$$

Norm of a function induced by the inner product above

(2) $$||a|| = \sqrt{\langle a, a \rangle}$$

Gaussian function with standard deviation sigma, normalized to 1

(3) $$G_\sigma(\mathbf{x}) = \exp\left[-|\mathbf{x}|^2/(2\sigma^2)\right] / (\pi^{3/4}\sigma^{3/2})$$

Atomic density function consisting of weighted Gaussians centered on atomic positions

(4) $$\rho_{\mathcal{X}}(\mathbf{x}) = \sum_{\mathbf{x}' \in \mathcal{X}} w_{\mathbf{x}'} G_{\sigma_{\mathbf{x}'}}(\mathbf{x} - \mathbf{x}')$$

Gaussian Integral Inner Product (GIIP) between two configurations

(5) $$\left\langle \mathcal{X}^\alpha, \mathcal{X}^\beta \right\rangle = \langle \rho_{\mathcal{X}^\alpha}, \rho_{\mathcal{X}^\beta} \rangle$$

Distance between two configurations

(6) $$\left|\mathcal{X}^\alpha - \mathcal{X}^\beta\right|^2 = \left\langle \mathcal{X}^\alpha - \mathcal{X}^\beta, \mathcal{X}^\alpha - \mathcal{X}^\beta \right\rangle = \left\langle \mathcal{X}^\alpha, \mathcal{X}^\alpha \right\rangle + \left\langle \mathcal{X}^\beta, \mathcal{X}^\beta \right\rangle - 2 \cdot \left\langle \mathcal{X}^\alpha, \mathcal{X}^\beta \right\rangle$$

Orientation invariance by minimizing over all possible rotations/rotoinversions

(7) $$\min_{\mathbf{R} \in \mathrm{O}(3)} \left|\mathcal{X}^\alpha - \mathbf{R}\mathcal{X}^\beta\right|^2$$

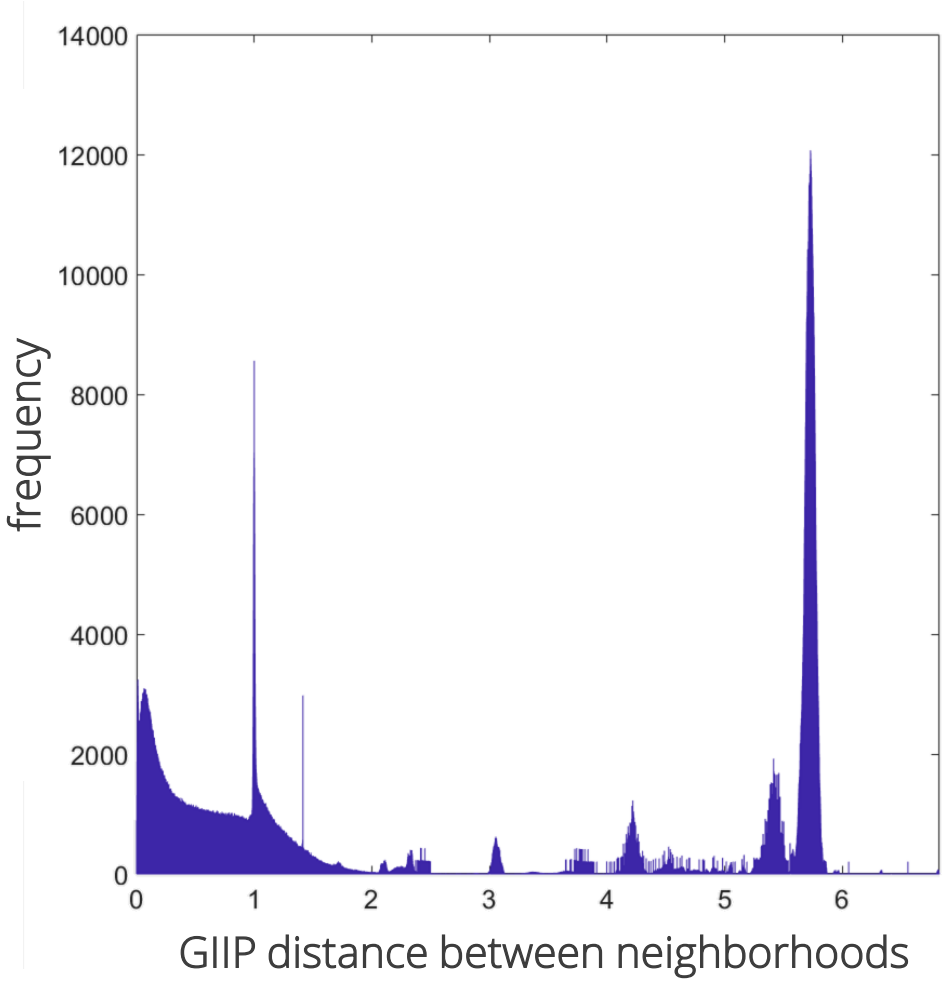
GIIP is analytically tractable in a computationally convenient form

(8) $$\left\langle \mathcal{X}^\alpha, \mathcal{X}^\beta \right\rangle = 2\sqrt{2} \sum_{\mathbf{x}^\alpha \in \mathcal{X}^\alpha} \sum_{\mathbf{x}^\beta \in \mathcal{X}^\beta} w_{\mathbf{x}^\alpha} w_{\mathbf{x}^\beta} \left(\frac{\sigma_{\mathbf{x}^\alpha} \sigma_{\mathbf{x}^\beta}}{\sigma_{\mathbf{x}^\alpha}^2 + \sigma_{\mathbf{x}^\beta}^2} \right)^{3/2} \exp\left[-|\mathbf{x}^\alpha - \mathbf{x}^\beta|^2 / (2\sigma_{\mathbf{x}^\alpha}^2 + 2\sigma_{\mathbf{x}^\beta}^2)\right]$$

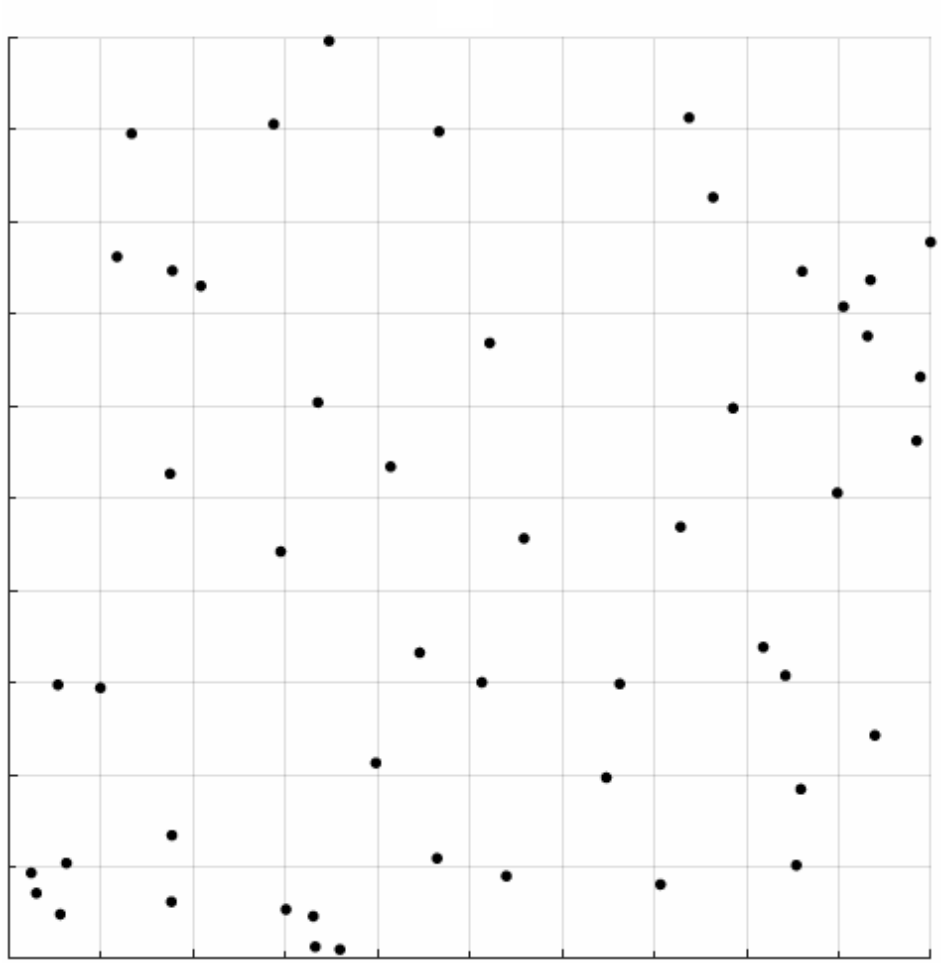
2. A Meaningful Distance Metric Between Atomic Neighborhoods

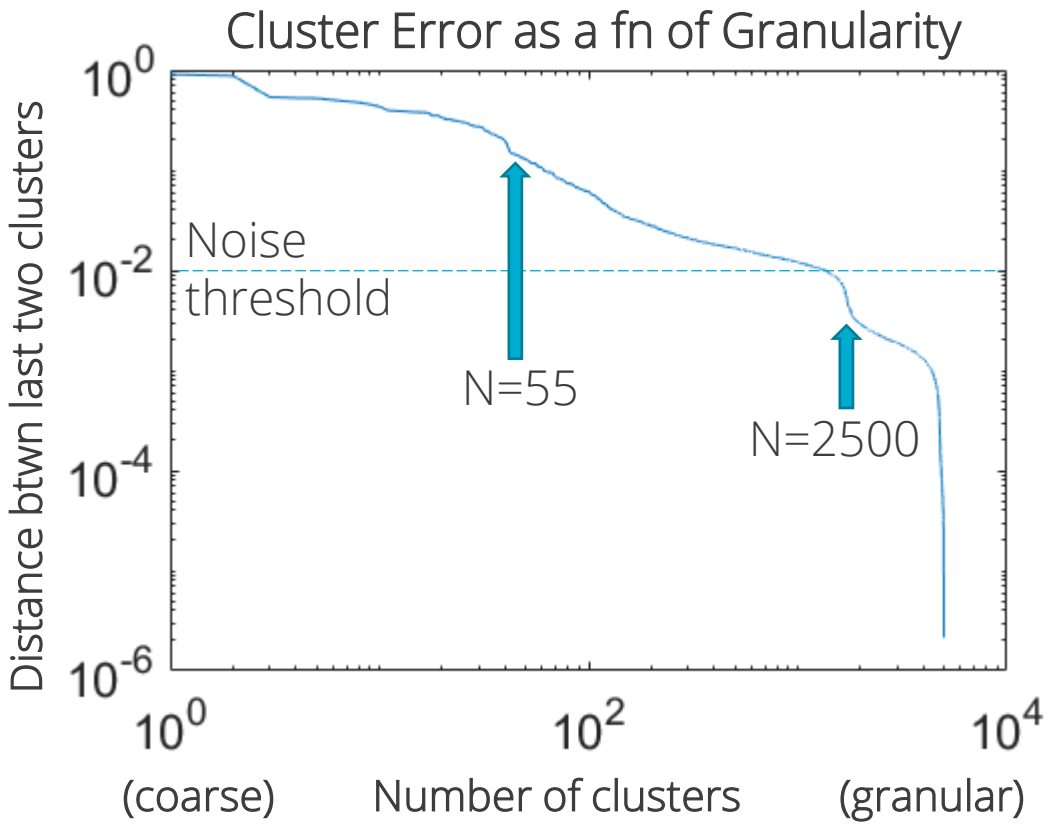
- PyTorch GPU-enabled code for GIIP-distance calculations
- Pairwise GIIP-distance between all 5000 atomic neighborhoods @ angular resolution of 5 degrees

- Histogram of 2.5 million GIIP distances at right immediately suggests the presence of classes of “defects” in the glass



3. Hierarchical (Agglomerative) Clustering of Atomic Neighborhoods

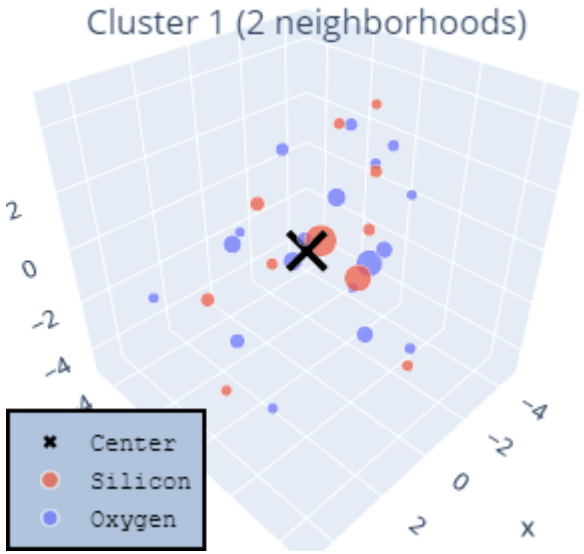


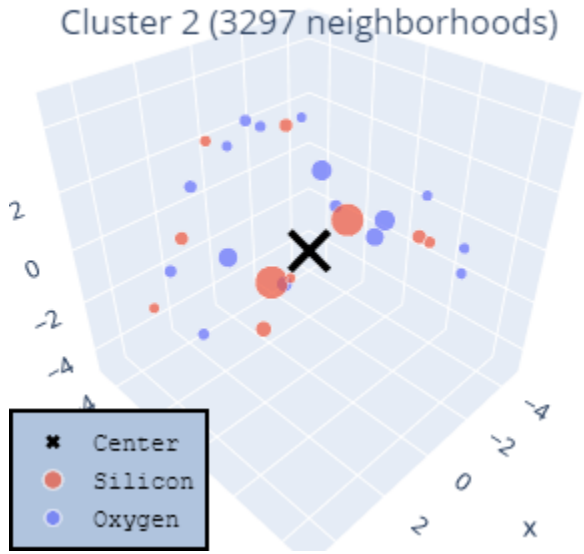


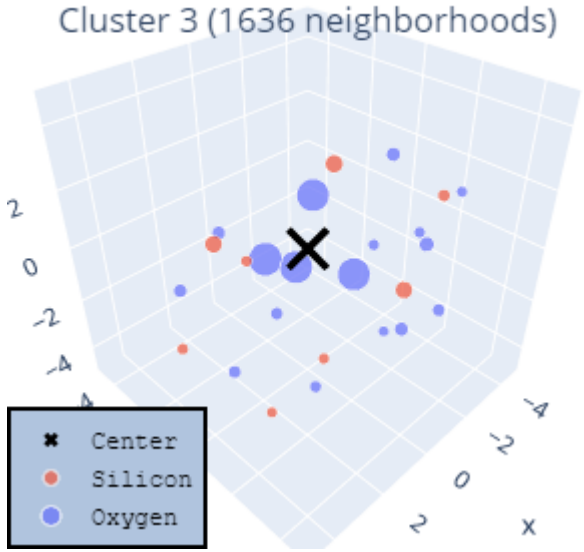
Plotting (number of clusters) against (cluster error), we see two natural inflections. These inflections tell us something about the structure of the glass: the 5000 neighborhoods sampled can be naturally divided into 55 coarse classes (motifs) or 2500 fine classes.

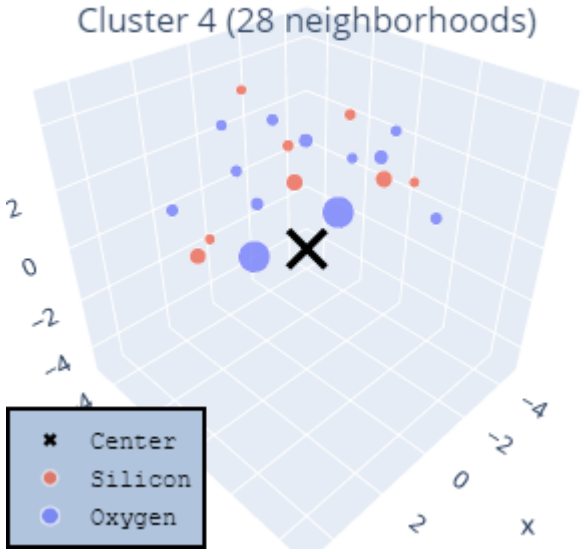
3. Hierarchical (Agglomerative) Clustering of Atomic Neighborhoods

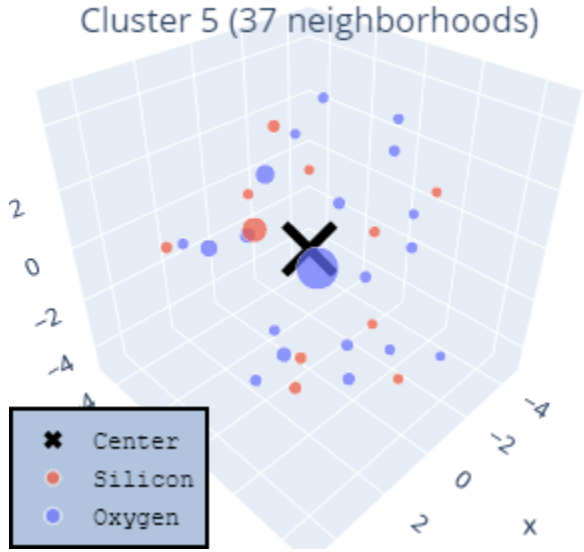
- Agglomerative clustering, n=5 classes
- Exemplar atomic environments from each class are plotted here
- Classes 2 and 3 are typical of SiO_2 glass
- Classes 1, 4, and 5 are broad classes of defects



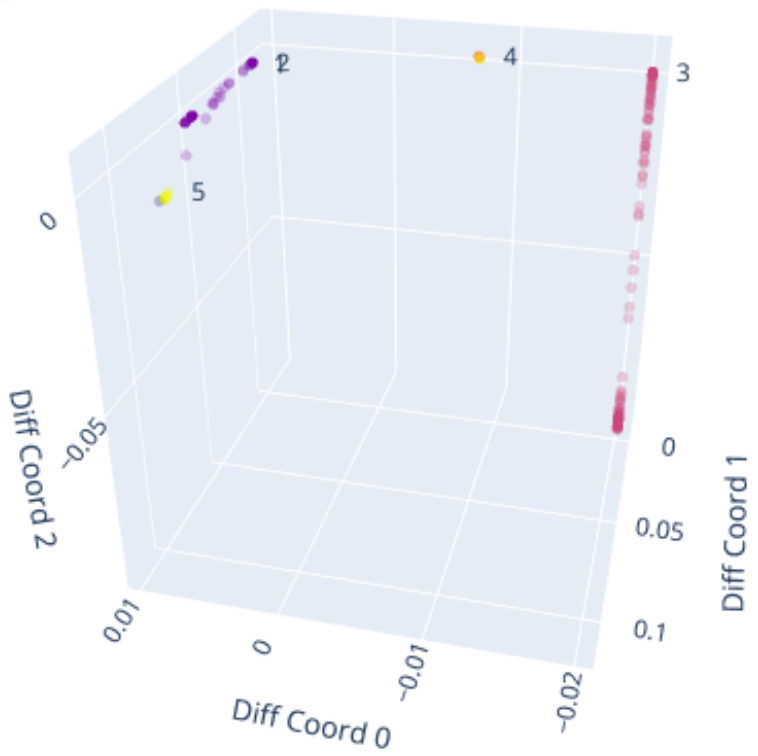


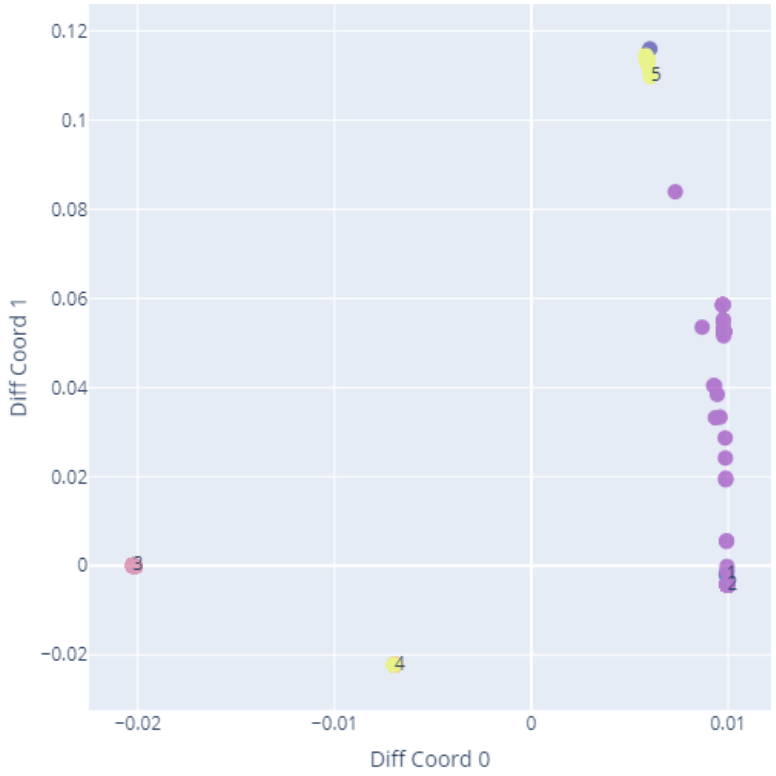


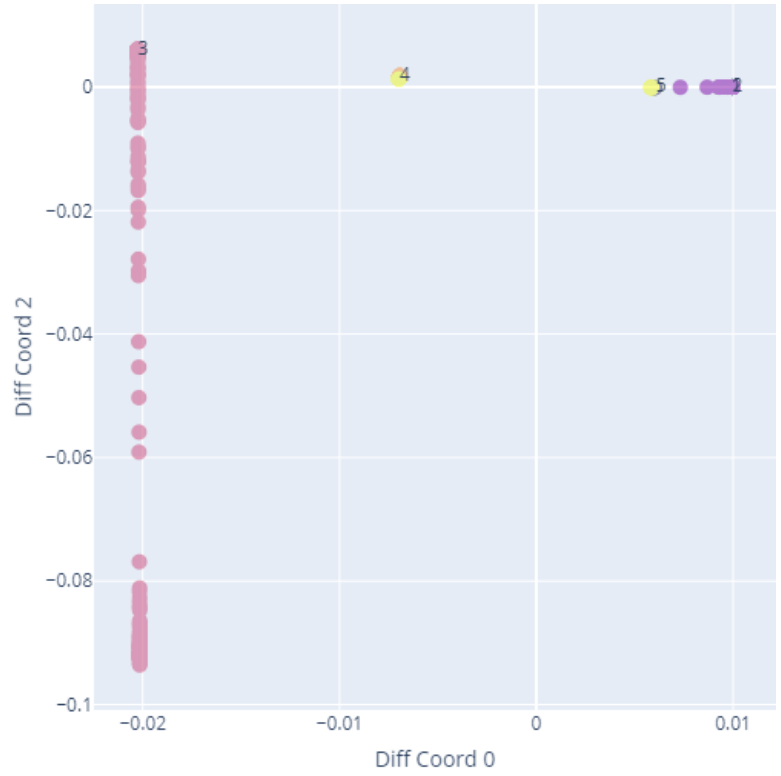




4. Diffusion Maps

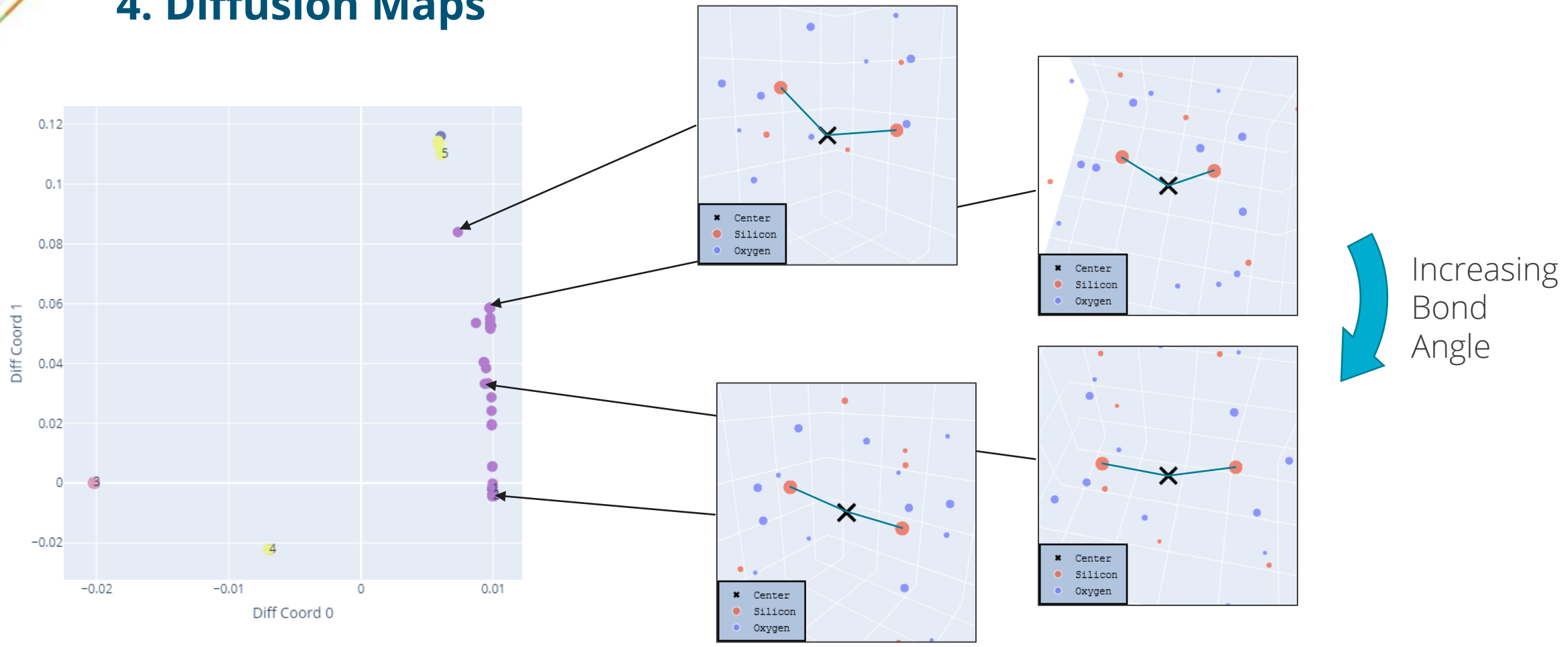






Diffusion maps are a well-established way of extracting a latent space from a pairwise distance matrix. Here, I've plotted all 5000 neighborhoods on the first three parsimonious latent space coordinates, and colored them according to the five clusters we extracted previously.
Notice that the classes (which we extracted directly from the pairwise distances) are, for the most part, also clustered on our latent space coordinates.

4. Diffusion Maps



We can identify the physical interpretation of some diffusion coordinates by inspection. For example, Diffusion Coordinate 1 captures the bond angle across bridging oxygens.

But, since the whole difficulty with glass is that the “by inspection” method is inadequate, is there a more data-driven approach available to us?

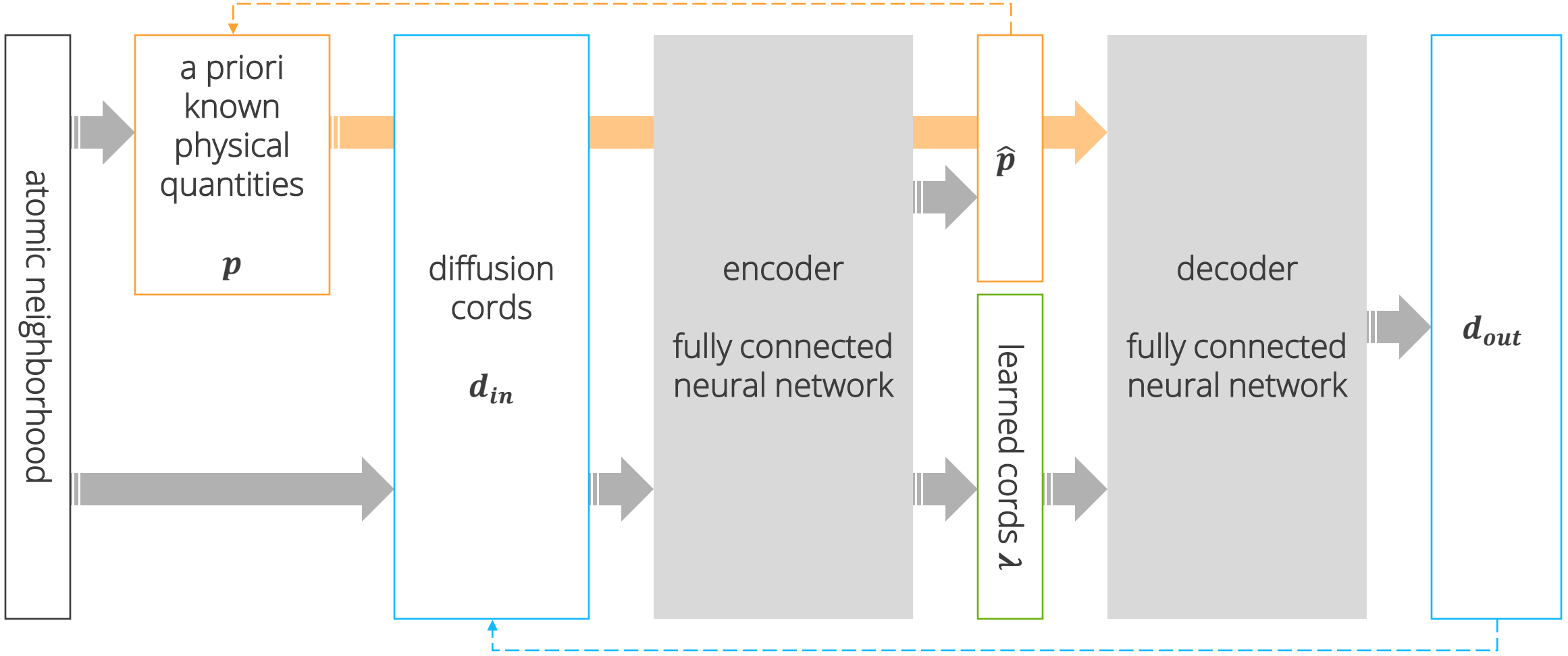
Conclusions

1. Gaussian Integral Inner Product (GIIP) provides an intuitive and computationally convenient distance metric between atomic configurations. This formulation addresses noise sensitivity, continuity, smoothness, radial cutoff, and permutation and (optionally) orientation invariance.
2. Using hierarchical clustering on GIIP distance, we can classify the atomic neighborhoods in a material to varying degrees of coarseness. This provides a detailed discrete descriptor for local structure.
3. Diffusion maps on GIIP distance provide a continuous latent space, parameterizing the implicit low-enthalpy manifold of the material. That parameterization is occasionally, but often not, friendly to interpretation by inspection.

Backup Slides

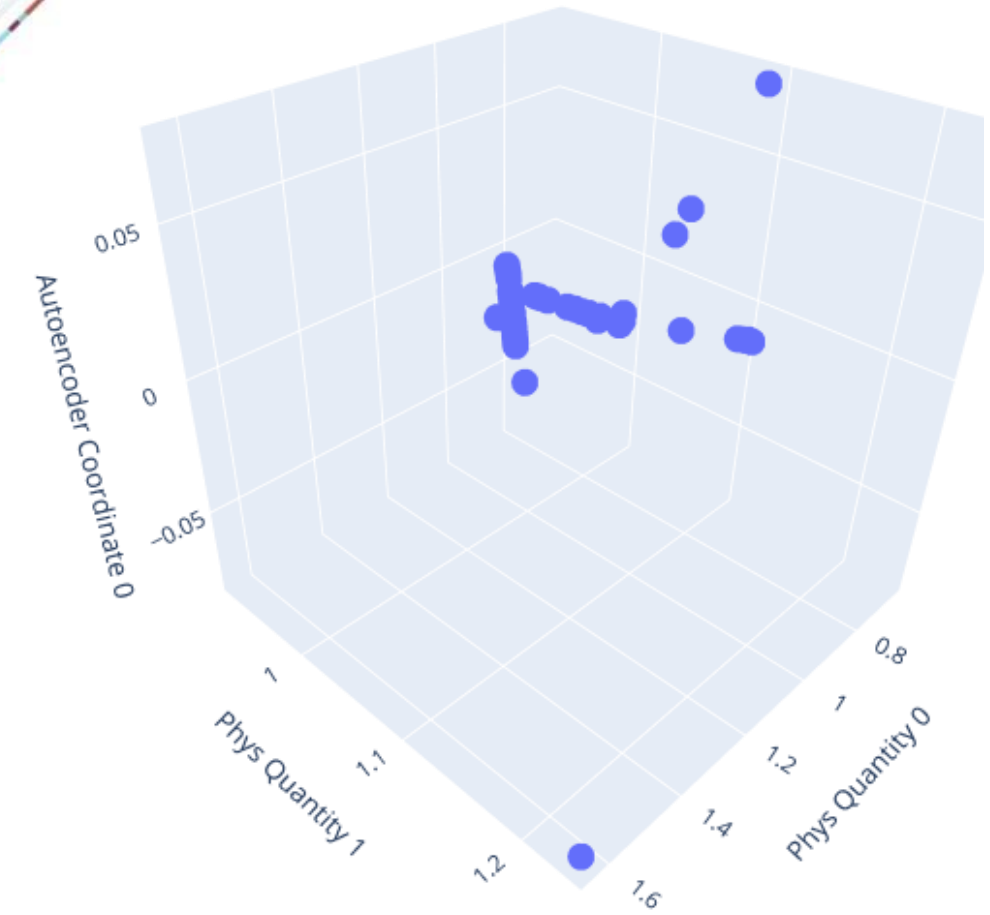
	Smooth Overlap of Atomic Position (SOAP)	Gaussian Integral Inner Product (GIIP)
Goal	Differentiable representation for potential energy functions; MD potentials.	Distance between configurations.
Atomic neighbor density function	Gaussian attached to each atom	Weighted Gaussian attached to each atom
Similarity Measure	Integral inner product	Integral inner product GIIP distance induced by inner product.
Rotational invariance	Integrate over orientations	Minimize GIIP distance over orientations
Implementation	Expand in spherical harmonic & radial basis fns. Kernel is dot product of expansion coefficients. Pretty fast!	Inner prod. analytically tractable from positions. Brute-force search over orientations. Pretty slow! Use GPUs.
Continuity/smoothness at boundaries	Enforced by selection of radial basis functions.	Enforced by weights going smoothly to zero as radial distance increases.
Permutation invariance Tolerance to variable atom quantities	Enforced by expansion into atomic neighbor density function.	Enforced by expansion into atomic neighbor density function.

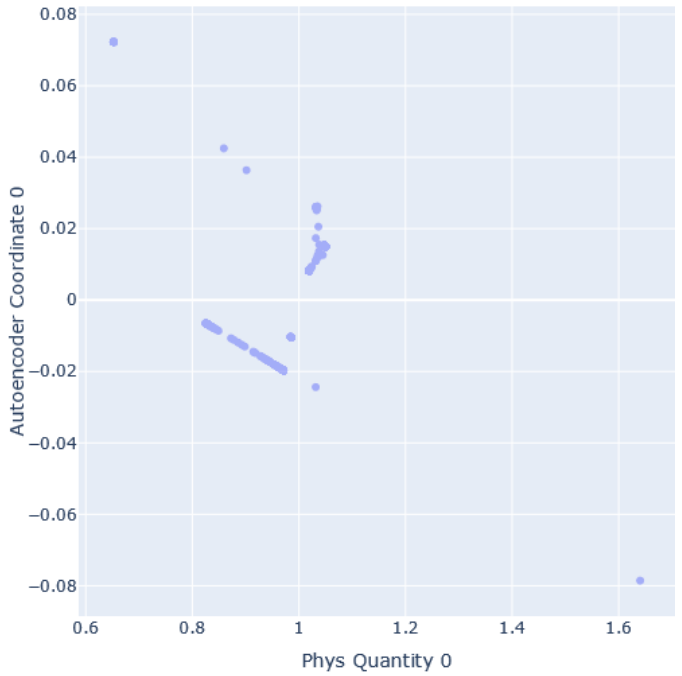
5. Conformal autoencoder for interpretability

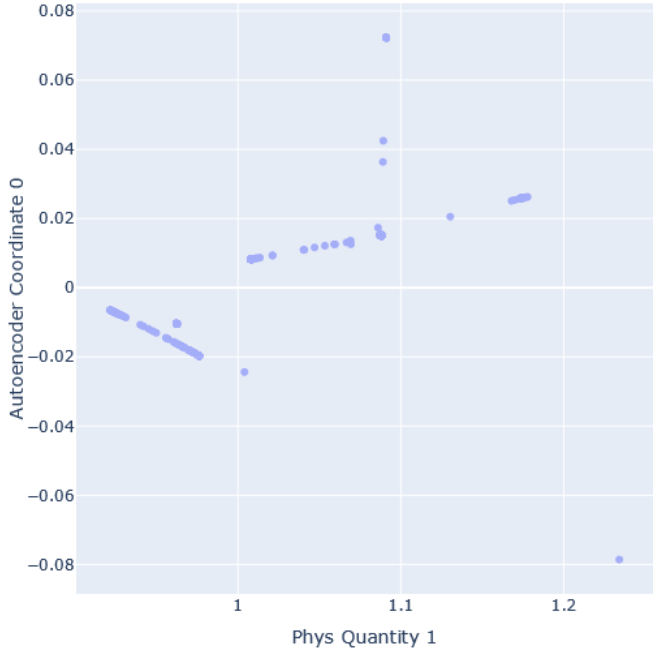


$$\text{loss} = \|\boldsymbol{d_{out}} - \boldsymbol{d_{in}}\|^2 + \|\hat{\boldsymbol{p}} - \boldsymbol{p}\|^2 + \alpha \left\| \frac{\partial \boldsymbol{\lambda}}{\partial \hat{\boldsymbol{p}}} \right\|^2 + \beta \left\| \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\lambda}} - \boldsymbol{I} \right\|^2$$

5. Conformal autoencoder for interpretability







Using the conformal autoencoder, we extract the portion of the diffusion coordinates that is orthogonal to (that shares no information with) physical quantities 0 and 1– that is, with central species and local volume. In other words, we have interpreted our diffusion coordinates in terms of two intuitive physical quantities, and one yet-to-be-determined parameter.