

拡散点雲法による原子クラスターの結合

Diffusion point cloud method applied to connecting atomic clusters

大阪大学¹, 鳥取大学² ○高野翔大¹, 小谷岳生^{1,2}, 佐藤和則¹, 大信田丈志²

takano.shota@mat.eng.osaka-u.ac.jp



「羽の生えた2匹の猫が空中で戦う」図

Introduction

➤ 材料設計と生成AI

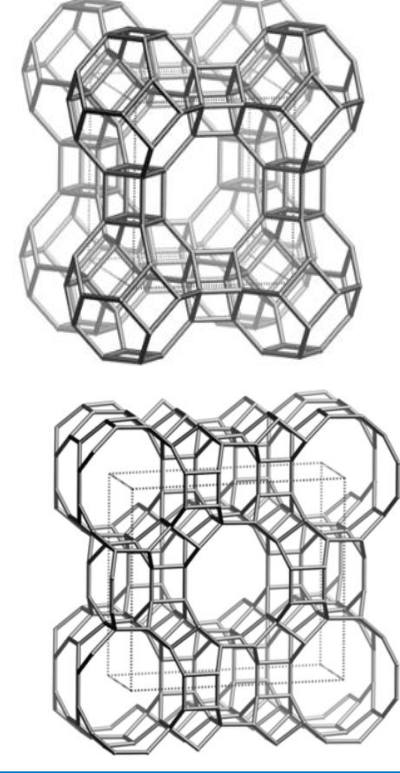
- ✓ 生成AIを用いた材料設計は近年進歩しており、特にMatterGen^[1]では、**逆拡散法**を応用した手法が試みられている。
- ✓ 「イオン伝導度の大きい結晶」など特性制約に応じた結晶の生成も可能であるが、手法的には改善の余地が大きい。

[1] C. Zeni et al., [arXiv:2312.03687](https://arxiv.org/abs/2312.03687) (2023)

➤ ゼオライトの結晶構造

本研究では簡単のためゼオライトを対象とした。

- ✓ SiO₂の組成を持ち、Si-O-Siの結合をもつ。Oを除けばSiのみであり、**Siのみで多形を生成する**課題となる。
- ✓ 250程度の構造が合成/観測されているが、10個程度のみが触媒、吸着、分子ふるいなどとして活用されている。



➤ 本研究の目的

結晶データベース^[2]の多数の結晶構造から局所的な構造(第n隣接までのクラスター)を取り出し、**クラスターを結合して新たな結晶構造を生成する。**

[2] <http://www.iza-structure.org/databases/>

➤ 画像生成AIの著しい進歩

- ✓ 画像生成では、逆拡散法, GANs, AEを組み合わせる方法が主流であり、精緻な絵を生成することが可能。
- ✓ 右図はbingで「羽の生えた2匹の猫が空中で戦っているところ」を指示して生成された図。
- ✓ **結晶でも同様の手法を適用できるはずである**
- ✓ さまざまな応用、拡張が考えられる基本的かつ重要な課題

➤ 逆拡散法とは？

- ✓ 拡散過程:
ある原子の配置に対し温度 T を上げていくと、原子は拡散され高温ではランダムな分布になる。(温度 T に比例するガウスノイズによるランダム化)
機械学習においては、温度 T に対応する仮想的な時刻 t を導入し、時刻 t を進めるにつれてデータがランダム化される過程について考える。
- ✓ 逆拡散法:
拡散過程を逆向きにたどる手法。(逆拡散法=アニーリングに対応)
各拡散ステップにおいて、拡散過程のランダムな動きの傾向を機械学習させ、その逆方向の動きを再現することで逆拡散を行う。

Method

➤ 入力データ

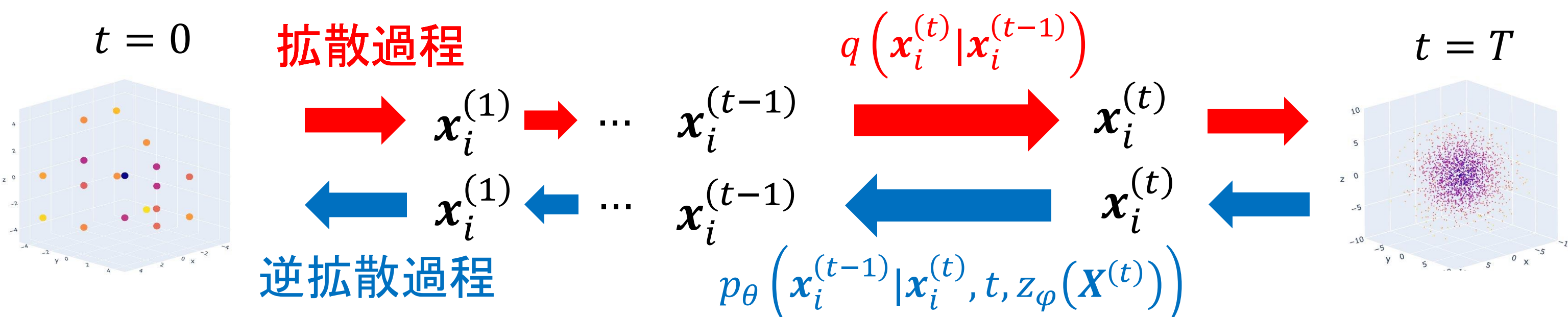
$$X = \{x_i | i = 1, \dots, N\}$$

- ゼオライトの結晶構造データベース^[2]から得たSiO₂ゼオライトのcifファイル227種。
- Si原子を中心として7ÅのSi位置のみのクラスターを取り出す(6200個)。
- クラスターに含まれる原子は30個程度。これをランダムに回転させる。
- クラスターは点雲であらわす。 $N = 36$ としているので一つの原子点に複数の点が重なることもある。

➤ 拡散点雲法^[3]

[3] S. Luo and W. Hu (2021) [arXiv:2103.01458](https://arxiv.org/abs/2103.01458)

各時刻 t ごとに X を考えるので、 $X^{(t)} = \{x_i^{(t)} | i = 1, \dots, N\}$ を考える。



- ✓ θ と φ を機械学習により決定する。
- ✓ z をラテントベクトルと呼び、 $X^{(t)}$ の情報を押し込んでいる。

✓ 拡散カーネル $q(x_i^{(t)} | x_i^{(t-1)})$

単純なガウスノイズによるランダム化。通常は時刻 t に比例して大きくする。

✓ 逆拡散カーネル $p_\theta(x_i^{(t-1)} | x_i^{(t)}, t, z_\varphi(X^{(t)}))$

$t, X^{(t)}$ を与えたときの点 x_i に働く力に対応。(シフトバックすべき量)
 q とは異なり $X^{(t)}$ が必要。(実際的には x_i の周辺の点の位置が重要)

逆拡散法の簡単な説明

- 時刻 t におけるデータ X の分布
- 確率分布 $P(X, t)$
- 初期分布 $P(X, t=0)$ は再現されるべきデータの分布
- 拡散の指定により $P(X, t)$ は決定されている
- 拡散をドリフト(流れ)として捉えると、ドリフト量を逆にすることにより逆拡散させることができる
- ドリフト量は多次元ベクトルと時刻 (X, t) の関数であり、これを学習させるのが逆拡散法

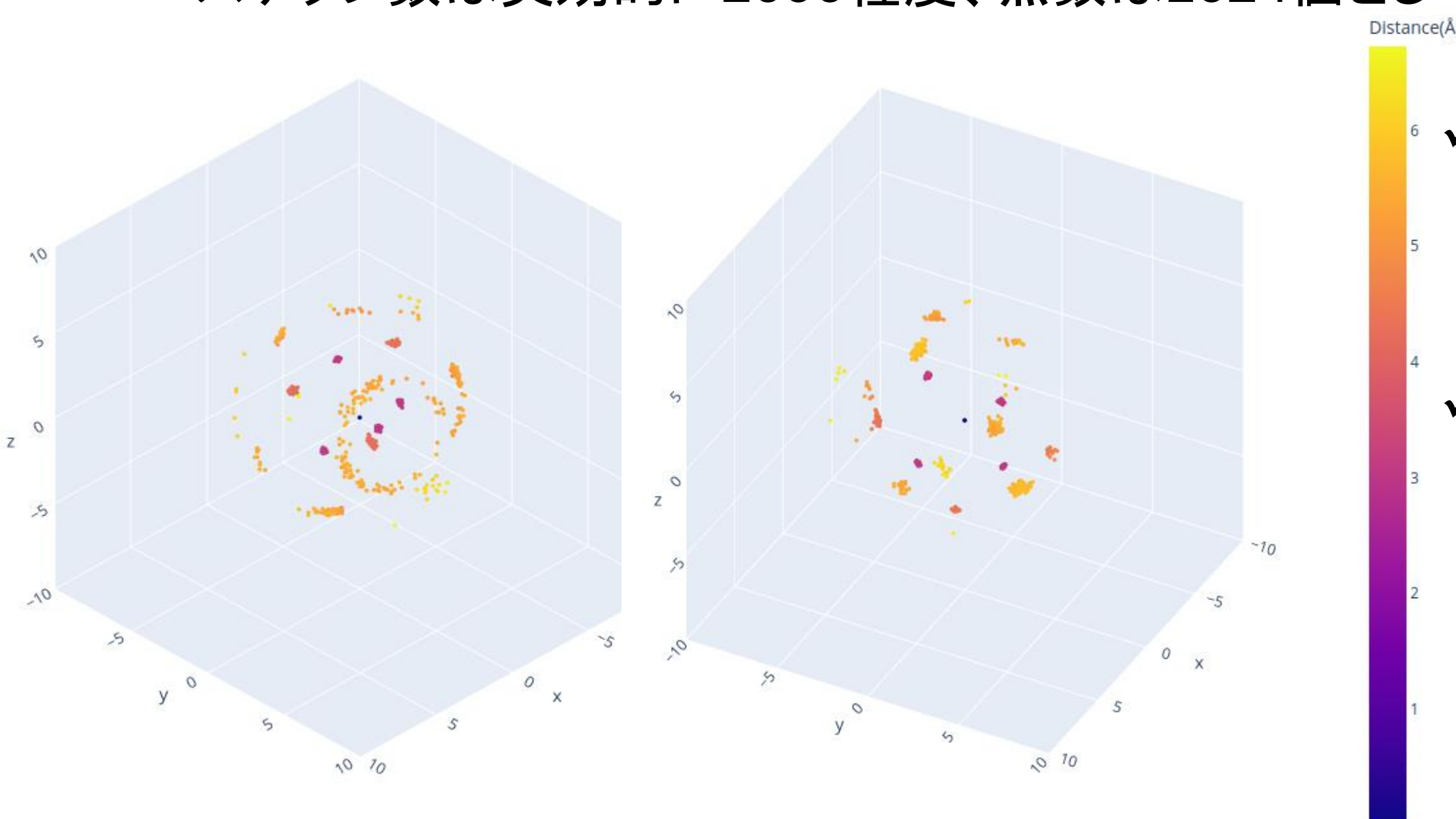
Result

ソースコード※

マシンの情報

➤ クラスターの再生テスト

- ✓ 原子位置が確率分布として点雲で表現される
- ✓ 学習させたもののうちのどれか(もしくは類似したもの)が生成されることが期待される。
- ✓ 学習コンディション
 - 学習にはRTX4090で数時間程度以上必要。生成は数秒程度(バッチで大量一括生成)
 - ステップ数は実効的に2000程度、点数は1024個として生成した。



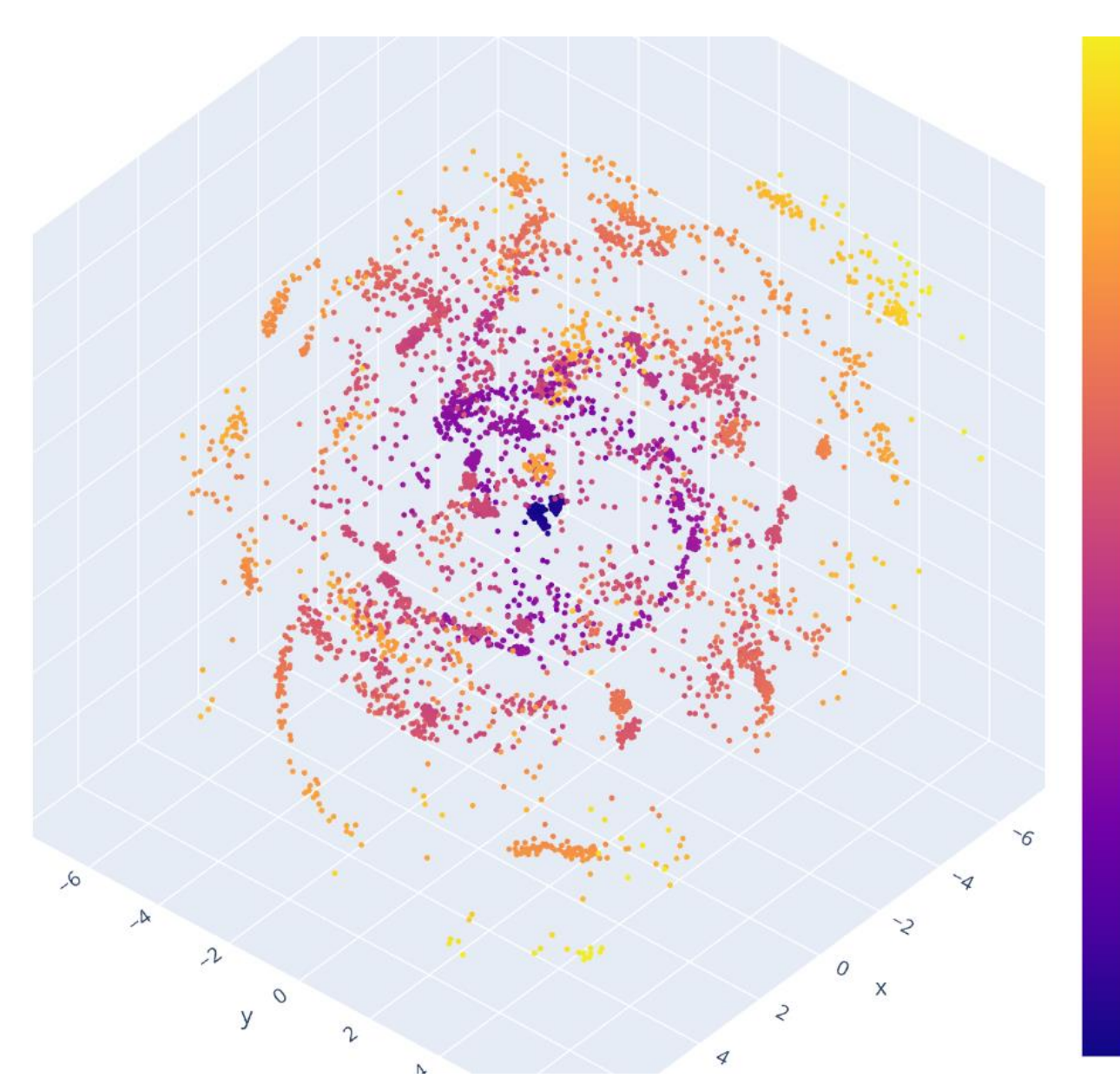
- ✓ 第一隣接($\sim 3\text{\AA}$)くっきりとした4配位を示す。
- ✓ 第二隣接それなりに学習できているが、中央から遠いと少しランダムになる。

➤ クラスターの結合生成テスト

1. まずひとつの中央クラスターを生成する。
2. その周辺の4つのSi位置を中心して、4つのクラスターを中央クラスターの原子分布を尊重するようにして生成する。(本研究では中央クラスター原子位置を混ぜ込んでラテントを生成)
3. 5つのクラスターが結合したものが生成される。

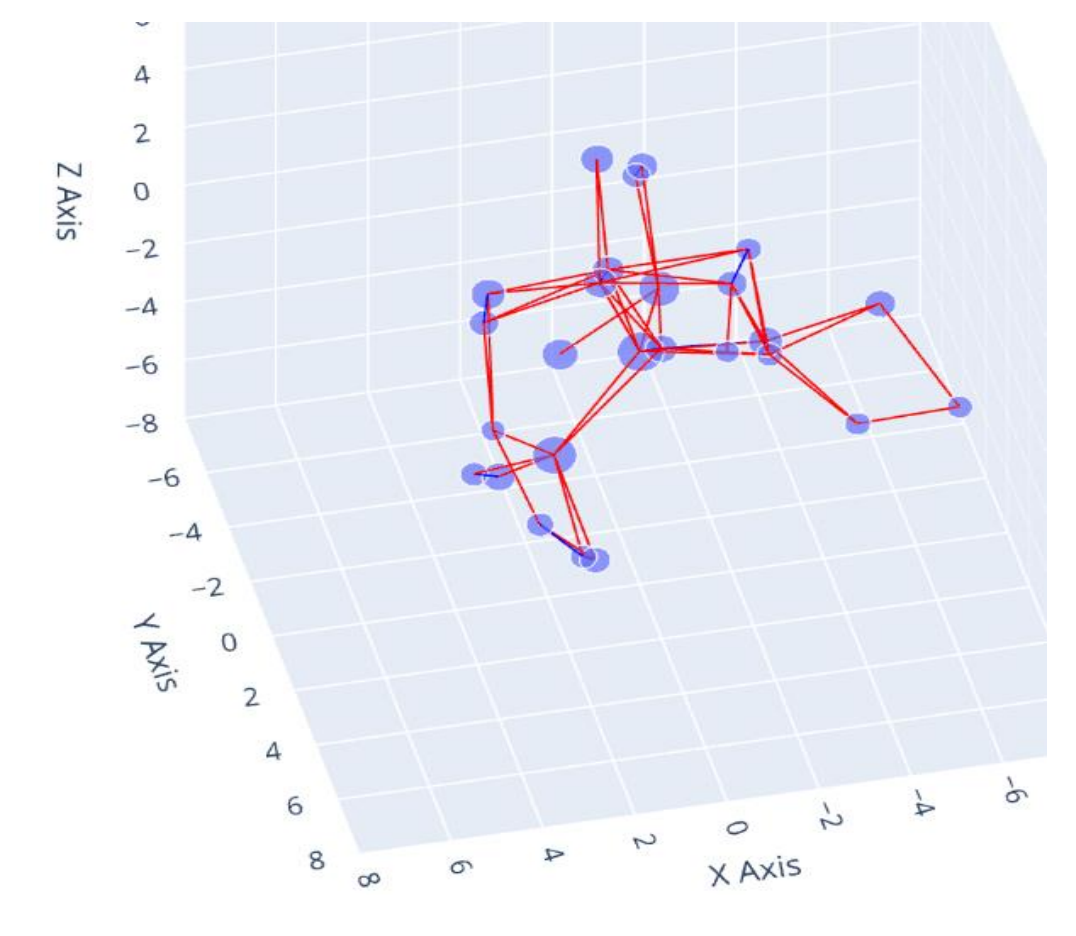
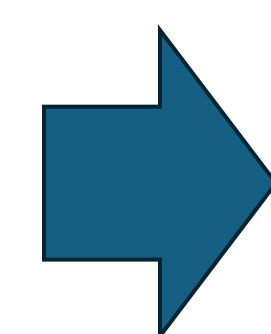
5つのクラスターの合成結果

くっきりとは合成できていない。
これだけではよくわからない→右図



点雲をクラスタリングしてボンドを描いた

- 丸の体積が点数に比例。
- 第一隣接はそれなりにくっきり見える(少しずれがあり、クリーンな4配位ではない)
- 合成によってできた5員環なども見えている
- 青線は2.7Å以下で短すぎるボンド



Conclusion

結果

- ✓ **点雲逆拡散法によってゼオライトの原子クラスターを再現した。**改善の余地はあるが、良好な学習結果を示した。
- ✓ **逆拡散の際に5つの原子クラスターの結合を試みた。**改善の余地はあるが、逆拡散によるクラスター結合は可能であると言える。
- ✓ 原子数の増減が自由な点で、MatterGen^[1]の方法(原子座標のみを用いる)に比べて優位性を持ち得る。

課題

- ✓ クラスターの合成: 中央のクラスターを固定せず同時的に生成したい。また、結晶周期性を取り込みたい。(この点は本質的には問題はない)
- ✓ 学習法: バッチ正規化などで改善の余地が見られる。試行錯誤を行い、より論理的なモデルアーキテクチャを探索する。

※現在のコードはリクエストに応じて配布可(RX3090, 16GB程度以上を推奨)