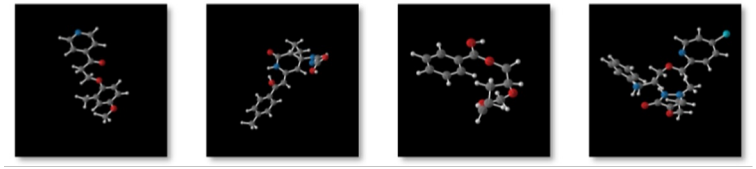
Geometric Latent Diffusion Models for 3D Molecule Generation

Minkai Xu [1] Alexander S. Powers [1 2]
Ron O. Dror [*1] Stefano Ermon [*1] Jure Leskovec [*1]

[1]Department of Computer Science, Stanford University
[2]Department of Chemistry, Stanford University



Outline

- Invariance & Equivariance
- Molecule Generation in 3D
- Motivation
- Model Architecture
- Geometric Autoencoders
- Geometric Latent Diffusion
- Experiment

Invariance & Equivariance

Given a function $f : \mathbf{X} \to \mathbf{Y}$ and transformations S_g, T_g, where g is an element of the group of transformations G.

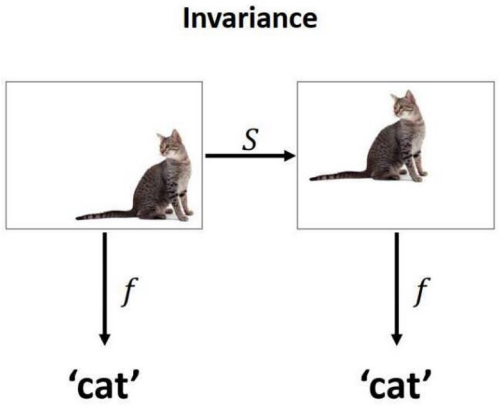
Invariance:

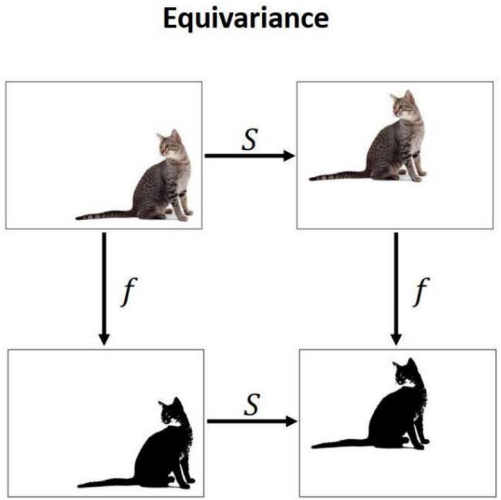
$$f \circ S_g(\mathbf{x}) = f(S_g(\mathbf{x})) = f(\mathbf{x})$$

Equivariance:

$$f \circ S_g(\mathbf{X}) = T_g \circ f(\mathbf{X})$$

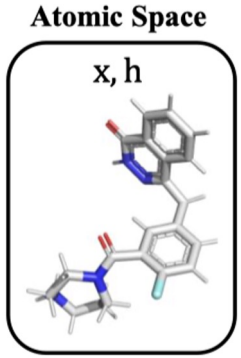
(Note that invariance is a special case of equivariance)





Credit: [1], [3]

Molecule Generation in 3D



Credit: [2]

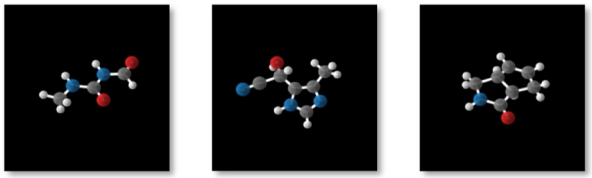
How to represent a **molecule in 3D**?

- Each molecule is represented as **a point cloud:** $\mathcal{G} = \langle \mathbf{x}, \mathbf{h} \rangle$
- **Atom coordinates matrix:** $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \mathbb{R}^{N \times 3}$
- **Node feature matrix:** $\mathbf{h} = (\mathbf{h}_1, \ldots, \mathbf{h}_N) \in \mathbb{R}^{N \times d}$ (e.g., atomic type and charges)

The goal is to learn a probability distribution p over the 3D molecule space with generative models and sample novel 3D molecules from p.

Molecule Generation in 3D

(I) **Unconditional generation:** Give a collection of molecules $\mathcal{G}$, learn parameterized generative models $p_\theta(\mathcal{G})$.



(II) **Controllable generation:** With molecules $\mathcal{G}$ labeled with certain properties s (e.g., protein target), learn conditional generation models $p_\theta(\mathcal{G} \mid s)$.

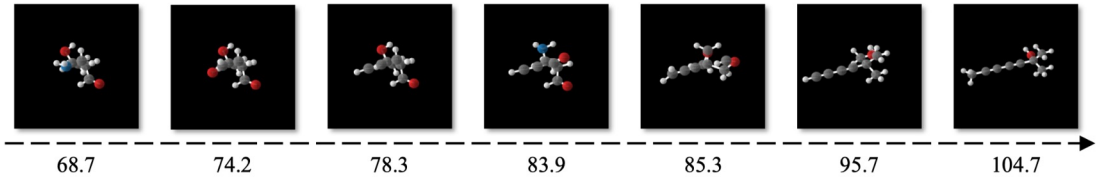


Figure 3. Molecules generated by conditional GEOLDM. We conduct controllable generation with interpolation among different Polarizability α values with the same reparametrization noise ϵ. The given α values are provided at the bottom.

Credit: [2]

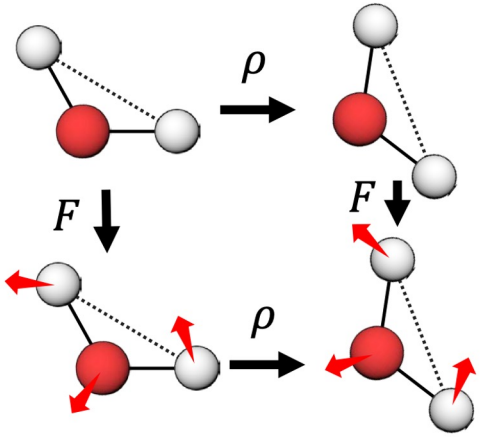
In- & Equivariance - Molecule Case

- **3D Special Euclidean Group SE(3):** The group of rotation and translation in 3D space, where transformations can be represented by a translation t and an orthogonal matrix rotation R.[2]
- When applying transformations $\in$ **SE(3)**:
  - The node features $\mathbf{h}$ are **SE(3)-invariant**
  - The coordinates $\mathbf{x}$ are **affected** as: $\mathbf{R}\mathbf{x} + \boldsymbol{t}$
  - A function f is **SE(3)-equivariant** if

$$\boldsymbol{R}\mathbf{z}_{\mathbf{x}} + \boldsymbol{t} = f(\boldsymbol{R}\mathbf{x} + \boldsymbol{t})$$

, where $\mathbf{z}_{\mathbf{x}} = f(\mathbf{x})$.

$$f \circ S_g(\mathbf{x}) = f(\mathbf{x})$$
$$f \circ S_g(\mathbf{X}) = T_g \circ f(\mathbf{X})$$



Visual explanation of the equivariance

Credit: [3]

Molecule Generation in 3D

- The **central challenge** in **generating 3D molecules from scratch** lies in achieving SE(3)-invariance when generating **3D atom positions**. [6]

- In other words, the generative model should **assign the same probability** to $\mathcal{M}$ and $\mathcal{M}'$ if $\mathcal{M}'$ can be obtained by rotating or translating $\mathcal{M}$ in 3D space.

Motivation

- **Diffusion models** that **satisfy the invariance constraint** have already been applied to 3D molecule generation (e.g., [4]) and show successful applications in downstream tasks like target drug generation, antibody design, and protein design.
- **However**,
  - Existing models **mainly still work on the original atomic space**.
  - Molecules are composed of both categorical (atom types, **discrete**) and **continuous** representations. However, **previous method directly learns Gaussian diffusions on both continuous and discrete features**, which is suboptimal.

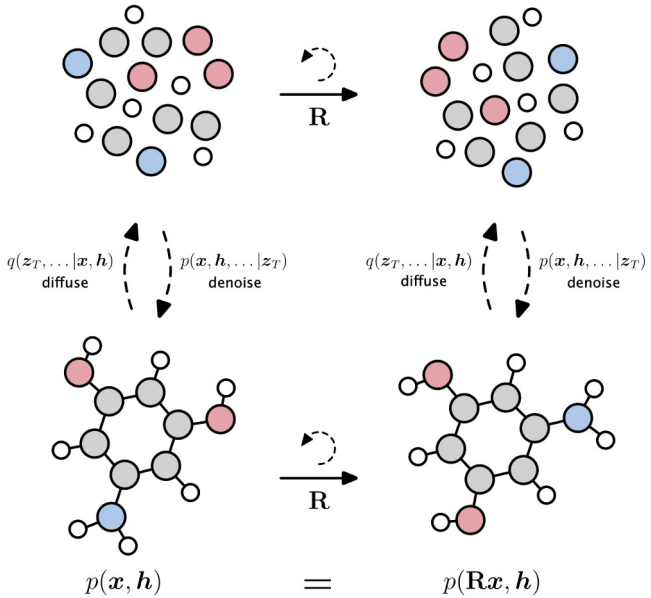


Figure 1. Overview of the EDM. To generate a molecule, a normal distributed set of points is denoised into a molecule consisting of atom coordinates $\boldsymbol{x}$ in 3D and atom types $\boldsymbol{h}$. As the model is rotation equivariant, the likelihood is preserved when a molecule is rotated by $\mathbf{R}$.

Credit: [4]

Motivation

- **Diffusion models** that **satisfy the equivariance constraint** have already been applied to 3D molecule generation (e.g., [4]) and show successful applications in downstream tasks like target drug generation, antibody design, and protein design.
- **However**,
  - Existing models **mainly still work on the original atomic space**.
  - Molecules are composed of both categorical (atom types, **discrete**) and **continuous** representations. However, **previous method directly learns Gaussian diffusions on both continuous and discrete features**, which is suboptimal.

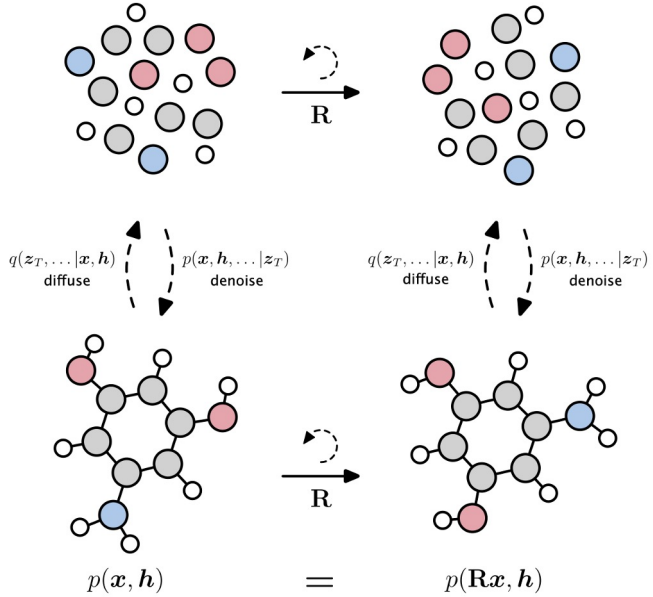


Figure 1. Overview of the EDM. To generate a molecule, a normal distributed set of points is denoised into a molecule consisting of atom coordinates $\boldsymbol{x}$ in 3D and atom types $\boldsymbol{h}$. As the model is rotation equivariant, the likelihood is preserved when a molecule is rotated by $\mathbf{R}$.

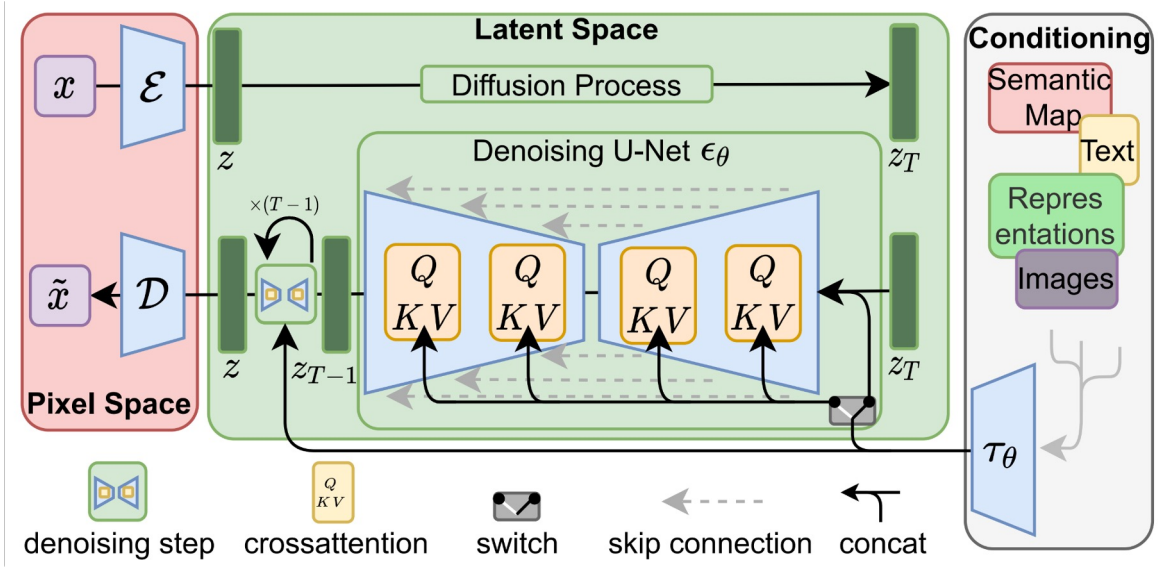
Credit: [4]

Motivation

Stable (Latent) Diffusion

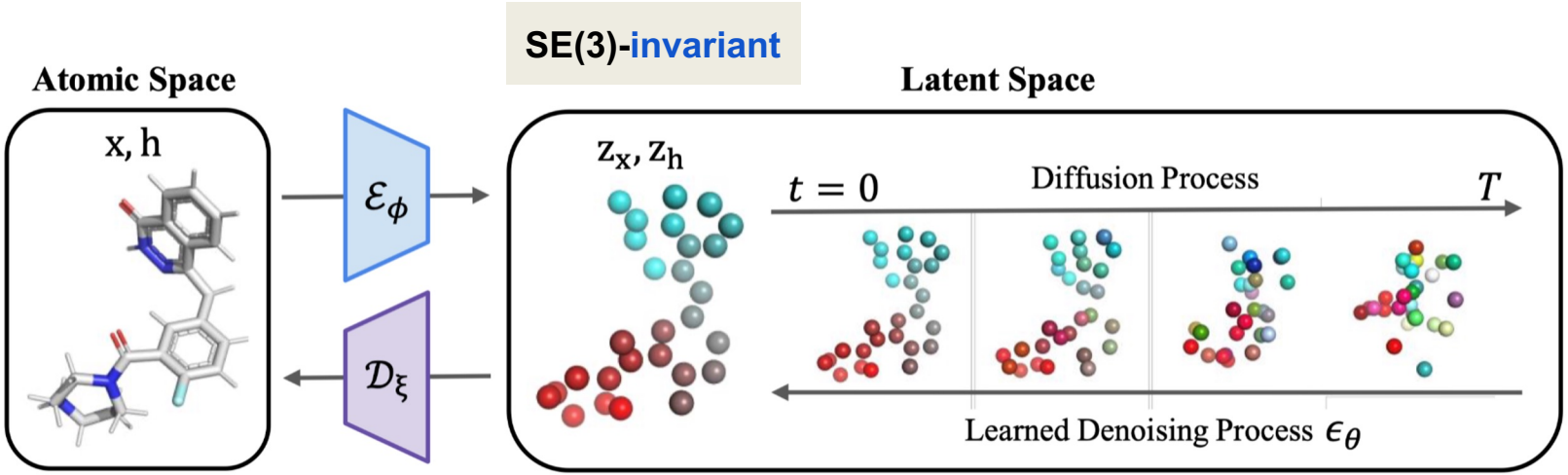
Compared with DDPM (we learned last week):

- Perform denoising diffusion in **the latent space** rather than the data space.
- **Better fidelity** and **controllable generation**.



Credit: [5]

Model Architecture



Geometric Latent Diffusion

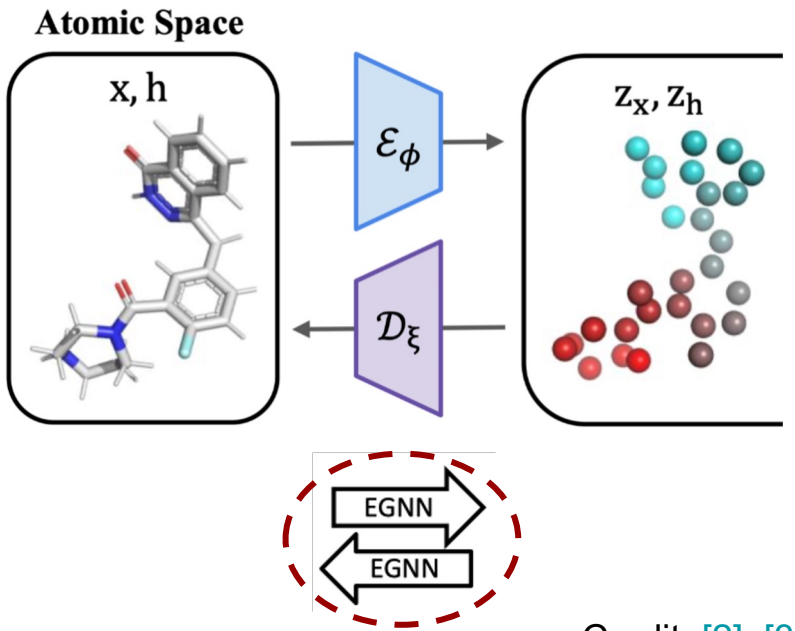
Credit: [2]

Geometric Autoencoders

- **Geometry consistency**: the latent features should faithfully represent the geometry (can be decoded back to the original structures). This requires the latent variable should also hold both invariant and equivariant features.

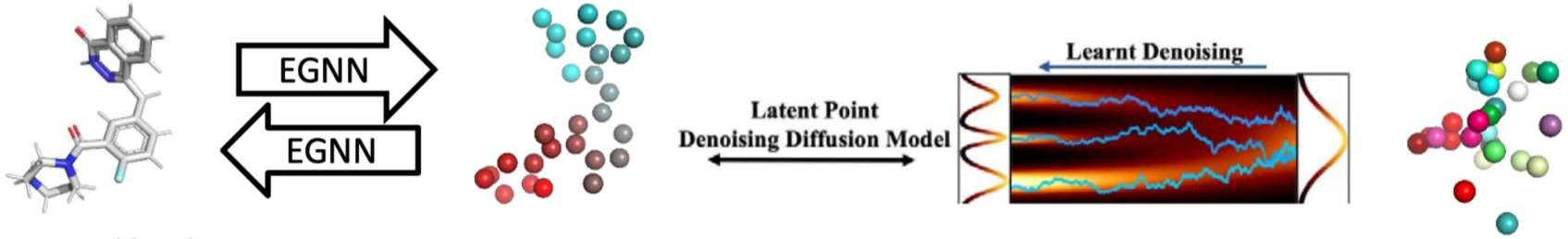
$$\mathbf{z} = \langle \mathbf{z}_{\mathrm{x}}, \mathbf{z}_{\mathrm{h}} \rangle \in \mathbb{R}^{N \times (3+k)}$$

- **SE(3)-equivariance**: given the fact that AE itself is also generative models, it should also be SE(3)-invariant, which requires the encoding and decoding transformations should be SE(3)-equivariant. [3]



Credit: [2], [3]

Geometric Latent Diffusion



$X \sim N \times (d+3)$ $z \sim N \times (k+3)$

$$\mathcal{L}_{AE} = \mathcal{L}_{recon} + \mathcal{L}_{reg},$$
$$\mathcal{L}_{recon} = -\mathbb{E}_{q_\phi(\mathbf{z}_\mathrm{x},\mathbf{z}_\mathrm{h}|\mathbf{x},\mathbf{h})} p_\xi(\mathbf{x},\mathbf{h}|\mathbf{z}_\mathrm{x},\mathbf{z}_\mathrm{h}),$$

$$\mathcal{L}_{LDM} = \mathbb{E}_{\mathcal{E}(\mathcal{G}),\epsilon\sim\mathcal{N}(0,I),t}\left[w(t)||\epsilon - \epsilon_\theta(\mathbf{z}_{\mathrm{x},t},\mathbf{z}_{\mathrm{h},t},t)||^2\right]$$
$$\mathcal{L} := \mathcal{L}_{recon} + \mathcal{L}_{LDM}$$

- **Model architecture**: **equivariant** autoencoders (utilizing EGNNs) **encodes** geometric data $\mathbf{x}$ into latent space $\mathbf{z}$ **with scalar and tensor variables**; latent **equivariant** diffusion (parameterized as time-conditional EGNNs) learns to **generate** latent variables $\mathbf{z}$.
- **Training**:
  - **First stage**: learn autoencoders by reconstruction objective
  - **Second stage**: learn denoising diffusion models in latent space
- **Sampling**:
  - Denoising from prior Gaussian noise $\mathbf{z}_\mathrm{T}$ to generate latent variables $\mathbf{z}$
  - Decoder decodes $\mathbf{z}$ back to $\mathbf{x}$

Credit: [3]

Geometric Latent Diffusion

GEOLDM is parameterized and optimized in an SE(3)-invariant fashion. Detailed proof in [2].

- Learn **equivariant** diffusions in the regularized and **lower-dimensional** latent space, enabling **better generation quality and efficiency**;
- Building upon the concept of stable diffusion, we can achieve improved **controllable generation** in the latent space by conditioning it on desired properties.

$$\boldsymbol{\epsilon}_\theta(\mathbf{z}, t) \rightarrow \boldsymbol{\epsilon}_\theta(\mathbf{z}, t, s)$$

$$\varepsilon_\phi(\mathbf{x}, \mathbf{h}), \mathcal{D}_\xi(\mathbf{z}_\mathbf{x}, \mathbf{z}_\mathbf{h}) \rightarrow \varepsilon_\phi(\mathbf{x}, \mathbf{h}, s), \mathcal{D}_\xi(\mathbf{z}_\mathbf{x}, \mathbf{z}_\mathbf{h}, s)$$

Credit: [3]

Experiment

Datasets

- ***QM9***: 130k small molecules, limited to 9 heavy atoms
- ***GEOM-DRUG***: 37 million molecular conformations for around 450k molecules, up to 181 atoms and 44.2 atoms on average

Evaluation Metrics

- **Atom Stability**: Percentage of atoms with correct valency
- **Molecule Stability**: Percentage of molecules in which all atoms are stable.
- **Validity and Uniqueness**: Percentage of valid and unique molecules as determined by RDKit

Experiment

Table 1. Results of atom stability, molecule stability, validity, and validity×uniqueness. A higher number indicates a better generation quality. Metrics are calculated with 10000 samples generated from each model. On QM9, we run the evaluation for 3 times and report the derivation. Note that, for DRUG dataset, molecule stability and uniqueness metric are omitted since they are nearly 0% and 100% respectively for all the methods. Compared with previous methods, the latent space with both invariant and equivariant variables enables GEOLDM to achieve up to 7% improvement for the validity of large molecule generation.

	QM9				DRUG	
# Metrics	Atom Sta (%)	Mol Sta (%)	Valid (%)	Valid & Unique (%)	Atom Sta (%)	Valid (%)
Data	99.0	95.2	97.7	97.7	86.5	99.9
ENF	85.0	4.9	40.2	39.4	-	-
G-Schnet	95.7	68.1	85.5	80.3	-	-
GDM	97.0	63.2	-	-	75.0	90.8
GDM-AUG	97.6	71.6	90.4	89.5	77.7	91.8
EDM	98.7	82.0	91.9	90.7	81.3	92.6
EDM-Bridge	98.8	84.6	92.0*	90.7	82.4	92.8*
GRAPHLDM	97.2	70.5	83.6	82.7	76.2	97.2
GRAPHLDM-AUG	97.9	78.7	90.5	89.5	79.6	98.0
GEOLDM	**98.9** $\pm$ 0.1	**89.4** $\pm$ 0.5	**93.8** $\pm$ 0.4	**92.7** $\pm$ 0.5	**84.4**	**99.3**

*Results obtained by our own experiments. Other results are borrowed from recent studies (Hoogeboom et al., 2022; Wu et al., 2022).

Experiment

Table 2. Mean Absolute Error for molecular property prediction. A lower number indicates a better controllable generation result. Results are predicted by a pretrained EGNN classifier ω on molecular samples extracted from individual methods.

Property Units	α Bohr^3	$\Delta\varepsilon$ meV	$\varepsilon_{\text{HOMO}}$ meV	$\varepsilon_{\text{LUMO}}$ meV	μ D	C_v $\frac{\text{cal}}{\text{mol}}$K
QM9*	0.10	64	39	36	0.043	0.040
Random*	9.01	1470	645	1457	1.616	6.857
N_{atoms}	3.86	866	426	813	1.053	1.971
EDM	2.76	655	356	584	1.111	1.101
GeoLDM	2.37	587	340	522	1.108	1.025

*The results of *QM9* and *Random* can be viewed as lower and upper bounds of MAE on all properties.

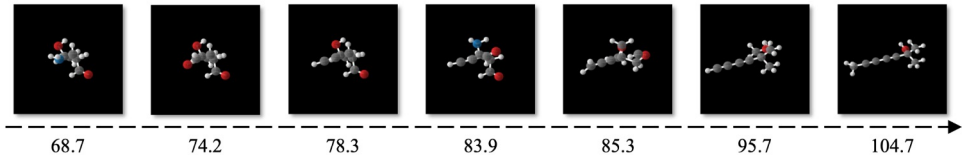


Figure 3. Molecules generated by conditional GeoLDM. We conduct controllable generation with interpolation among different Polarizability α values with the same reparametrization noise ϵ. The given α values are provided at the bottom.

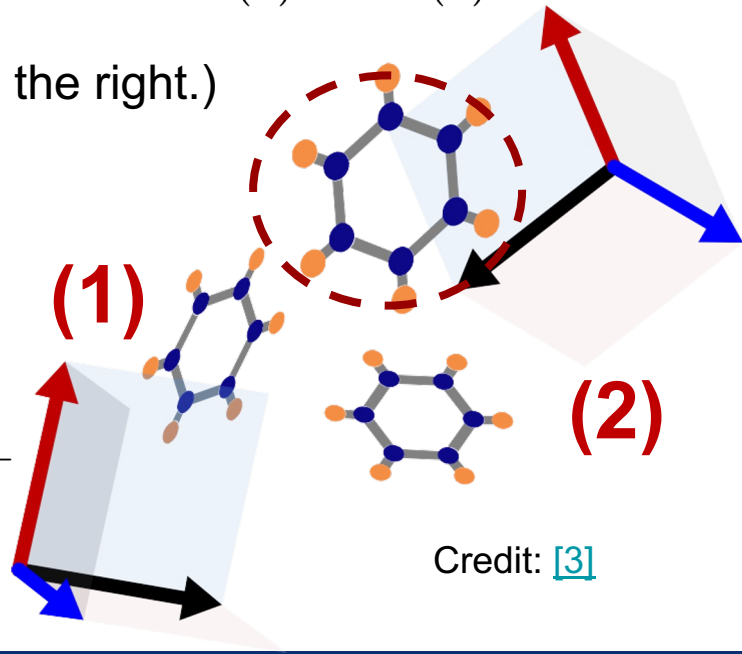
References

[1] https://www.doc.ic.ac.uk/~bkainz/teaching/DL/notes/equivariance.pdf
[2] Xu, Minkai, et al. "Geometric latent diffusion models for 3d molecule generation." International Conference on Machine Learning. PMLR, 2023.
[3] Xu, Minkai, et al. "Graph and Geometry Generative Modeling for Drug Discovery." Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023.
[4] Hoogeboom, Emiel, et al. "Equivariant diffusion for molecule generation in 3d." International conference on machine learning. PMLR, 2022.
[5] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
[6] Zhang, Xuan, et al. "Artificial intelligence for science in quantum, atomistic, and continuum systems." arXiv preprint arXiv:2307.08423 (2023).

Quiz

Q1: Why is it important to ensure invariance when representing molecules in 3D?
(Please answer the question based on the image on the right.)

$$\mathbf{x}_{(2)} = [-1, -2, -3]^\top$$
$$\mathbf{h}_{(2)} = \mathbf{h}_{(1)}$$



$$\mathbf{x}_{(1)} = [1, 2, 3]^\top$$
$$\mathbf{h}_{(1)} = \mathbf{h}_{(2)}$$

Credit: [3]

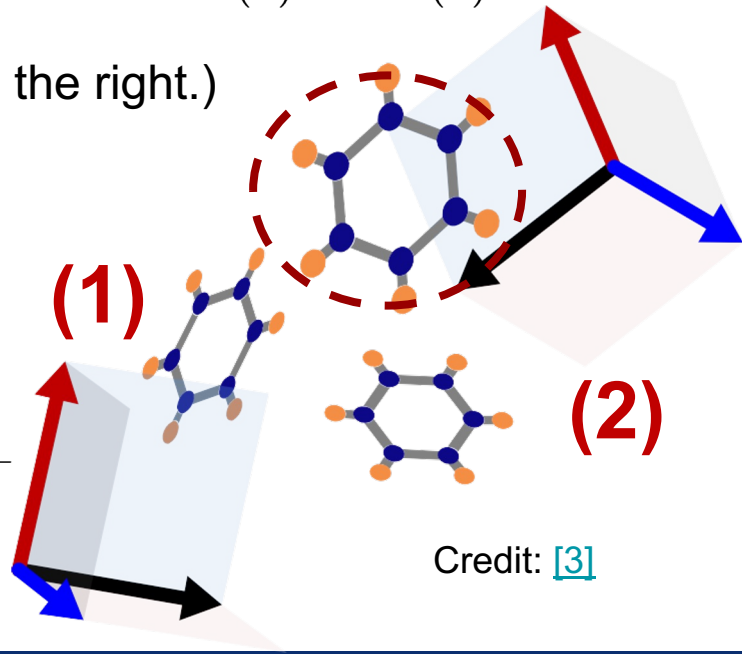
Quiz

$$\mathbf{x}_{(2)} = [-1, -2, -3]^\top$$
$$\mathbf{h}_{(2)} = \mathbf{h}_{(1)}$$

Q1: Why is it important to ensure invariance when representing molecules in 3D?
(Please answer the question based on the image on the right.)

- (1) & (2) use different coordinate systems to describe the **same** molecular geometry.
  But output of some models like traditional GNNs given (1) and (2) as completely different!

$$\mathbf{x}_{(1)} = [1, 2, 3]^\top$$
$$\mathbf{h}_{(1)} = \mathbf{h}_{(2)}$$



Credit: [3]

Quiz

Q2: What is **not** the benefit of modeling molecular geometries in a regularized latent space compared to directly in the atomic feature space?

a) Latent space is lower dimensional

b) Allow for better control over the generation process

c) Removes the need to enforce geometric constraints

d) Learn to model a much smoother distribution

Quiz

Q2: What is **not** the benefit of modeling molecular geometries in a regularized latent space compared to directly in the atomic feature space?

a) Latent space is lower dimensional

b) Allow for better control over the generation process

c) Removes the need to enforce geometric constraints

d) Learn to model a much smoother distribution

Thank you!