Equivariant Blurring Diffusion for Hierarchical Molecular Conformer Generation

Jiwoong Park, Yang Shen
Department of Electrical and Computer Engineering
Texas A&M University
ptywoong@gmail.com, yshen@tamu.edu

Abstract

How can diffusion models process 3D geometries in a coarse-to-fine manner, akin to our multiscale view of the world? In this paper, we address the question by focusing on a fundamental biochemical problem of generating 3D molecular conformers conditioned on molecular graphs in a multiscale manner. Our approach consists of two hierarchical stages: i) generation of coarse-grained fragment-level 3D structure from the molecular graph, and ii) generation of fine atomic details from the coarse-grained approximated structure while allowing the latter to be adjusted simultaneously. For the challenging second stage, which demands preserving coarse-grained information while ensuring SE(3) equivariance, we introduce a novel generative model termed *Equivariant Blurring Diffusion* (EBD), which defines a forward process that moves towards the fragment-level coarse-grained structure by blurring the fine atomic details of conformers, and a reverse process that performs the opposite operation using equivariant networks. We demonstrate the effectiveness of EBD by geometric and chemical comparison to state-of-the-art denoising diffusion models on a benchmark of drug-like molecules. Ablation studies draw insights on the design of EBD by thoroughly analyzing its architecture, which includes the design of the loss function and the data corruption process. Codes are released at `https://github.com/Shen-Lab/EBD`.

1 Introduction

The advancement of generative models to understand the multiscale properties of objects facilitates their application across a range of granularity levels, transcending individual scales. To enable generative models in processing multiscale structures, there has been a surge in hierarchical design methodologies across multiple domains, spanning from images [35, 40] to speech [20, 17]. These methods initially capture coarse-grained structures and subsequently generate finer details.

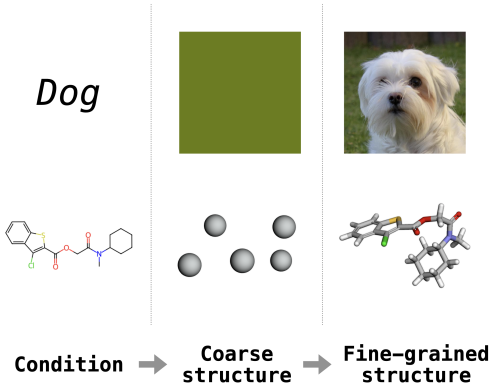


Figure 1: Blurring diffusion generative processes on image [42] and molecular conformer.

In the field of computer vision, recent efforts [16, 42, 2, 7] have yielded successful designs of coarse-to-fine generative frameworks for 2D pixels of images, leveraging denoising diffusion models that corrupt and restore the data by adding and removing noises [49, 50, 52, 15, 27]. Notably, [42] generates images from blurred prior distributions (average of pixel intensities) motivated from the heat equation of partial differential equations as Fig. 1.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

https://doi.org/10.52202/079017-4185

In the field of biochemistry and drug discovery, however, denoising diffusion models for 3D conformers of stable molecular structures have not yet taken advantage of coarse-to-fine multiscale frameworks. Current methods either disregard the scale hierarchy [46, 59, 19, 23, 60] or consider that in very limited ways [38, 41]. For instance, within the recent hierarchical method of unconditional conformer generation [38], a denoising diffusion model [19] is solely applied to generation of coarse-grained structure, without extending to coarse-to-fine generation.

The primary bottleneck in extending denoising diffusion models for molecular conformers to hierarchical designs is that random noise corrupts not only fine atomic details but also structural information of coarse-grained structures indiscriminately. To tackle this challenging problem, we exploit *fragments* that are frequently occurring substructures or functional groups in 2D molecular graphs. These fragments can be a promising candidate for the coarse-grained structural information in 3D geometry. Introducing fragments divides the generation process into two stages: i) generating coarse-grained structures represented by fragments, and ii) restoring fine atomic details from fragment structures. In the first stage of generating fragment coordinates from molecular graphs, we efficiently generate approximations of fragment structures comprising the center of mass and attributes of each fragment from a cheminformatics tool.

For the challenging second step of coarse-to-fine generation, we propose a novel diffusion model, *Equivariant Blurring Diffusion* (EBD) detailed as follows. Motivated from the blurring corruption of the heat equation [42], we design EBD to generate 3D molecular conformers from coarse-grained fragment approximated structures as Fig. 1, rather than from random noise. In our design of EBD, the forward process moves atom positions of conformers towards the center of mass of their respective fragments, while the reverse process restores full-atom details from the prior distribution of the 3D fragment structure. The blurring schedule we designed for EBD allows the diffusion model to focus on restoring fine atomic details while retaining coarse-grained information throughout the entire generative process. We validated our coarse-to-fine EBD model using a benchmark of drug-like molecules. We obtained superior results in conformer generation compared to the denoising diffusion model, even with 100 times fewer diffusion time steps.

The major contribution of this paper can be summarized as follows:

- We design EBD which generates atomic details from coarse-grained estimation of fragment structures using equivariant networks, motivated by the blurring corruption of heat equation.
- We propose a novel blurring scheduler and a revised loss function that significantly impacts performance, instead of directly applying those of existing image blurring diffusion model.
- We conduct a thorough analysis of the effects of fragment granularity, data corruption methods, and loss reformulation. We obtained geometrically and chemically more plausible conformers compared to state-of-the-art denoising diffusion models.

2 Background

2.1 Blurring diffusion

The denoising diffusion models [49, 50, 52, 15, 27], which corrupt data by adding random noise and generate data through denoising, have significantly advanced across diverse domains [55, 54, 44]. Recently, a few works [2, 42, 7, 18] have introduced data corruption into the design space of diffusion models [26], going beyond random noise corruption in the vision domain.

Inverse Heat Dissipation Model (IHDM) [42] proposed a coarse-to-fine generation in the pixel space. Their forward process follows a partial differential equation of heat dissipation on grids:

$$\frac{\partial}{\partial t}\mathbf{x}(i,j,t) = \Delta\mathbf{x}(i,j,t), \tag{1}$$

where $\mathbf{x}$ represents the data on the grid and Δ is the Laplacian operator. IHDM derived the solution of this equation at time step t, $\mathbf{x}_t$, using eigendecomposition of Δ as:

$$\mathbf{x}_t = \mathbf{B}_t\mathbf{x}_0 = \mathbf{V}\exp(-\mathbf{\Lambda}t)\mathbf{V}^T\mathbf{x}_0, \tag{2}$$

where $\mathbf{V}^T$ and $\mathbf{\Lambda}$ are discrete cosine transform and a diagonal matrix whose elements are eigenvalues of Δ, respectively. As $t \to T$, the eigenbasis of eigenvalue 0 only remains and this leads to the

convergence of pixel intensities to their average value. Based upon this blurring process, IHDM defined a forward process as:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t|\mathbf{B}_t\mathbf{x}_0, \sigma^2\mathbf{I}), \tag{3}$$

which means that the state at t is equal to the data blurred until t with small amount of noise. Note that the function of data corruption $\mathbf{B}_t$ was defined at a spectral space of eigenvalues $\mathbf{\Lambda}$. Then, the reverse generative process was defined to deblur each state:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}|\mu_\theta(\mathbf{x}_t, t), \delta^2\mathbf{I}), \tag{4}$$

where the mean at $t-1$ is the result of a deblurring network μ_θ and δ is the small amount of standard deviation for noise. As t approaches 0, μ_θ gradually restores fine details from coarse-grained information about pixel intensities by effectively deblurring state values. The loss was defined to minimize the distance between the result of deblurring network and less blurred state at randomly sampled t as:

$$L_{t-1} = \mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}_t}[\|\mathbf{B}_{t-1}\mathbf{x}_0 - \mu_\theta(\mathbf{x}_t, t)\|^2]. \tag{5}$$

IHDM was evaluated on image generation task using FID score [14], but its performance lagged behind that of denoising diffusion models. For instance, IHDM achieves an FID score of 18.96 while DDPM [15] have 3.17 on CIFAR-10 [30].

2.2 Equivariance

In this work, we consider the SE(3) group to address the roto-translational equivariance of molecular conformers [28, 19, 59]. A function f is equivariant to a group $\mathcal{G}$ if $T_g(f(\mathbf{x})) = f(S_g(\mathbf{x}))$ holds for all $g \in \mathcal{G}$, where T_g, S_g are transformations of the group element g. In our coarse-to-fine generative framework, the invariant prior distribution of coarse-grained structure represents the coordinates of fragments. Therefore, the design of the transition distribution and the loss function in our diffusion model need to ensure that the generated likelihood is invariant, so that the generated conformers are not affected by rotation or translation.

3 Methods

3.1 Problem definition

Suppose that we have a molecular graph G whose nodes $\mathcal{V}$ are n atoms with SE(3)-invariant features $\mathbf{h}^{\text{a}} \in \mathbb{R}^{n\times d}$ and edges $\mathcal{E}$ are inter-atomic bonds. Our objective is to generate an ensemble of 3D molecular conformers $\mathbf{x}^{\text{a}} \in \mathbb{R}^{n\times 3}$ given G. Our hierarchical approach is in two stages. i) $p(\mathbf{x}^{\text{f}}|G)$: generating a coarse-grained 3D structure of fragment coordinates $\mathbf{x}^{\text{f}} \in \mathbb{R}^{m\times 3}$ from G which was decomposed into m fragments, and ii) $p(\mathbf{x}^{\text{a}}|\mathbf{x}^{\text{f}}, G)$: the diffusion model generating fine atomic details $\mathbf{x}^{\text{a}} \in \mathbb{R}^{n\times 3}$ conditioned on the generated fragment structure $\mathbf{x}^{\text{f}}$. To map between atoms and their respective fragments, we defined a mapping matrix $\mathbf{M} \in \mathbb{R}^{n\times m}$ with $\mathbf{M}_{\mathtt{ik}} = 1$ if the i-th atom belongs to the k-th fragment and 0 otherwise. $\mathbf{M}\mathbf{x}^{\text{f}}$ makes each atom located at its respective fragment. On the other hand, $\mathbf{M}^\dagger\mathbf{x}^{\text{a}}$ results in fragment coordinates being the average of the coordinates of its constituent atoms, where $\mathbf{M}^\dagger$ is a pseudoinverse matrix of $\mathbf{M}$ ($\mathbf{M}^\dagger\mathbf{M} = \mathbf{I}$).

3.2 Fragmentation and 3D fragment structures

We decompose a molecule $G = (\mathcal{V}, \mathcal{E})$ into m non-overlapping fragments $\{S_k\}_{k=1}^m$, where $S_k = (\mathcal{V}_k, \mathcal{E}_k)$ and $\mathcal{V} = \bigcup_{k=1}^m \mathcal{V}_k, \mathcal{E} = \bigcup_{k=1}^m \mathcal{E}_k$ using Principal Subgraph (PS) [29]. Starting from all unique atoms in the fragment vocabulary $\mathcal{S}$, PS iteratively merges neighboring fragments. The most frequent fragment among the newly merged fragments was added to the vocabulary at each iteration, which was repeated until the desired size of the vocabulary was reached. The smaller the size of fragment vocabulary, the finer fragments and detailed coarse-grained structures can be obtained. After fragmentation was finished, from the relation between $\{\mathcal{V}_k\}_{k=1}^m$ and $\mathcal{V}$, the mapping matrix $\mathbf{M}$ can be constructed.

To generate the initial coordinates of fragments, we utilize RDKit distance geometry [5, 31], an efficient cheminformatics tool, instead of training an additional deep generative model. After generating initial atom coordinates $\hat{\mathbf{x}}^{\text{a}} \sim p_{\text{RDKit}}(\mathbf{x}^{\text{a}})$, we define the initial fragment coordinates $\mathbf{x}^{\text{f}}$ as averages of their constituent atom coordinates, $\mathbf{M}^\dagger\hat{\mathbf{x}}^{\text{a}}$. Since the atom coordinates generated by

 https://doi.org/10.52202/079017-4185

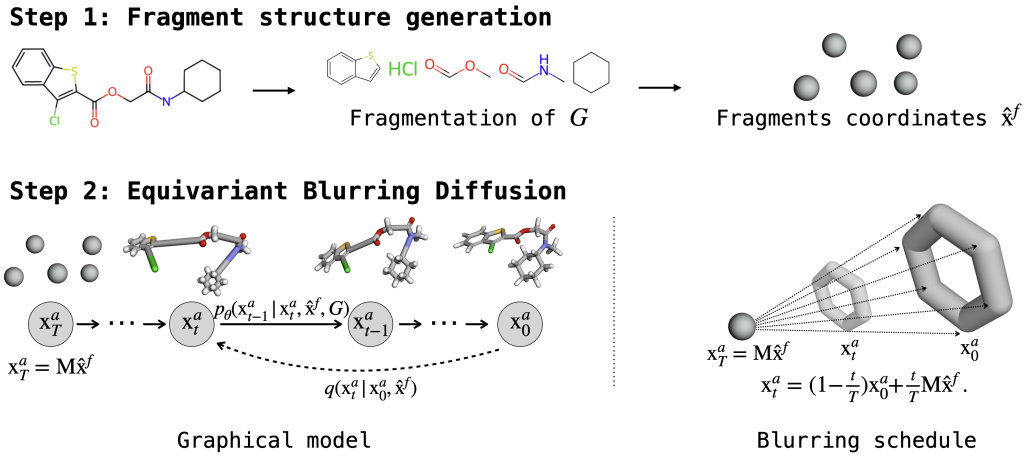


Figure 2: Our hierarchical molecular conformer generation framework. We first decompose a molecular graph G into fragments and generate fragment coordinates $\hat{\mathbf{x}}^f$. Then, conditioned on $\hat{\mathbf{x}}^f$ and G, Equivariant Blurring Diffusion generates atom-level fine details using the blurring schedule.

RDKit are approximations of the ground truth conformer, the resulting fragment coordinates are also an approximation (thus need to be adjusted in the next stage in Sec. 3.3), which we denote as $\hat{\mathbf{x}}^{\mathrm{f}} \sim p_{\mathrm{RDKit}}(\mathbf{x}^{\mathrm{f}})$. For fragment features $\mathbf{h}^{\mathrm{f}} \in \mathbb{R}^{m\times 3}$, we define a 3-dimensional vector as a frequency histogram of its constituent atom types based on their chemical properties, including hydrophobicity, hydrogen bond center, and negative charge center following [38].

The processes of fragment structure generation are illustrated in the step 1 of Fig. 2. Note that every step in this subsection does not harm the efficiency of our framework, as they can be completed before training our diffusion model and the process itself is efficient. To generate 5 distinct fragment coordinates $\hat{\mathbf{x}}^{\mathrm{f}}$ for each of the $45,000$ molecules in the training and validation set of the GEOM-Drug benchmark [1], it took 38 hours, averaging 3.04 seconds per molecule. The details of fragmentation and fragment vocabulary are demonstrated in Appendix C.1.

3.3 Equivariant blurring diffusion

In this subsection, we elaborate on the design of our diffusion model, *Equivariant Blurring Diffusion* (EBD), drawing inspiration from the principles of the heat equation. This model is designed to generate fine details of conformers $\mathbf{x}^{\mathrm{a}}$, starting from a coarse-grained, approximate structure $\hat{\mathbf{x}}^{\mathrm{f}}$ and a molecular graph G. We introduce a forward process and a data corruption function of blurring process in Sec. 3.3.1, a reverse process and a deblurring network to reach an SE(3)-invariant likelihood in Sec. 3.3.2, and a definition and reparameterization of an SE(3)-invariant loss function in Sec. 3.3.3. The overall scheme of EBD is illustrated in the step 2 of Fig. 2.

3.3.1 Forward process and blurring schedule

We define the data corruption of the forward process as a blurring operation that gradually shifts ground truth atom positions $\mathbf{x}_0^{\mathrm{a}} \sim q(\mathbf{x}_0^{\mathrm{a}})$ to their corresponding fragment coordinates:

$$q(\mathbf{x}_t^{\mathrm{a}}|\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}) = \mathcal{N}(\mathbf{x}_t^{\mathrm{a}}|f_{\mathbf{B}}(\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t), \sigma^2\mathbf{I}), \tag{6}$$

where $f_{\mathbf{B}}$ is a deterministic blurring operator. Consequently, every atom will be positioned according to its fragment coordinates $\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}$ in the prior fragment structure distribution.

When defining $f_{\mathbf{B}}$ for the forward process, we cannot directly adopt the spectral blurring operator $\mathbf{B}_t = \mathbf{V}\exp(-\mathbf{\Lambda}t)\mathbf{V}^T$ of IHDM [42] in Eq. (2) for the two reasons: i) For a single molecule, we need to calculate and decompose the fragment graph Laplacian $\{\mathbf{V}_k\mathbf{\Lambda}_k\mathbf{V}_k^T\}_{k=1}^m$ for each fragment $S_k = (\mathcal{V}_k, \mathcal{E}_k)$, unlike a single Laplacian operator per image in IHDM. Given the varying sizes and structures across fragments, it becomes challenging to uniformly adjust the movement of atoms across all fragments using a function of $\{\mathbf{\Lambda}_k\}_{k=1}^m$ in spectral space. ii) As $t \to T$, the ground truth atom coordinates $\mathbf{x}_0^{\mathrm{a}}$ will converge to the ground truth scaffold structure $\mathbf{x}^{\mathrm{f}} = \mathbf{B}_T\mathbf{x}_0^{\mathrm{a}}$ by spectral

graph theory [4]. However, there exists a mismatch between the ground truth coordinates $\mathbf{x}^{\mathrm{f}}$ and the approximation coordinates $\hat{\mathbf{x}}^{\mathrm{f}}$ from RDKit in the generative processes. This distributional shift of the fragment structure can potentially harm the performance during the inference.

To circumvent these issues, we transition the space of the blurring operator from spectral domain to spatial domain while retaining the essence of the blurring process. We define $f_{\mathbf{B}}$ as a linear interpolation between $\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}$ and $\mathbf{x}_0^{\mathrm{a}}$ in Euclidean space:

$$f_{\mathbf{B}}(\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t) = (1 - \frac{t}{T})\mathbf{x}_0^{\mathrm{a}} + \frac{t}{T}\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}. \tag{7}$$

As t progresses from 0 to T, the atom coordinates $\mathbf{x}_t^{\mathrm{a}}$ will gradually converge to the fragment structure $\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}$, allowing for uniform adjustment of atom movement. Additionally, we can mitigate the need for excessive eigendecomposition of the fragment graph Laplacian. The example of our blurring schedule on a single fragment is depicted in the step 2 of Fig. 2.

3.3.2 Reverse process and deblurring networks

The aim of the reverse process is to generate fine details at the atom-level from a prior distribution of 3D fragment structure $p(\mathbf{x}_T^{\mathrm{a}}) = \mathcal{N}(\mathbf{x}_T^{\mathrm{a}} | \mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}, \delta^2\mathbf{I})$ that is roto-translational invariant to the group. Drawing upon proofs regarding the conditions for an invariant likelihood [28, 59], we develop the deblurring process on the zero center-of-mass subspace using equivariant transition distributions:

$$p_\theta(\mathbf{x}_{t-1}^{\mathrm{a}} | \mathbf{x}_t^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, G) = \mathcal{N}(\mathbf{x}_{t-1}^{\mathrm{a}} | \mu_\theta(\mathbf{x}_t^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, G, t), \delta^2\mathbf{I}), \tag{8}$$

where μ_θ is a parameterized mean function consisting of a deblurring network. To ensure equivariance in the transition distribution, we devise μ_θ inspired by equivariant networks [45]. Our equivariant deblurring network updates invariant features of fragments and atoms $\mathbf{h}^{\mathrm{f}}, \mathbf{h}^{\mathrm{a}}$, (Eqs. (9, 10)), and the equivariant coordinates of atoms $\mathbf{x}^{\mathrm{a}}$ (Eq. (11)) by leveraging the hierarchical relationship between atoms and fragments. Let the i-th atom $\mathbf{x}_{\mathtt{i}}^{\mathrm{a}}$ belongs to the k-th fragment $\mathbf{x}_{\mathtt{k}}^{\mathrm{f}}$, then the l-th layer of equivariant deblurring networks for fragment- and atom-level message passing and feature updates is defined as follows:

$$\mathbf{m}_{\mathtt{ij}}^{\mathrm{f}} = \phi_m^{\mathrm{f}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{f},l}, \mathbf{h}_{\mathtt{j}}^{\mathrm{f},l}, \|\mathbf{x}_{\mathtt{i}}^{\mathrm{f}} - \mathbf{x}_{\mathtt{j}}^{\mathrm{f}}\|), \qquad \mathbf{h}_{\mathtt{i}}^{\mathrm{f},l+1} = \phi_h^{\mathrm{f}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{f},l}, \sum\nolimits_{\mathtt{j} \in N(\mathbf{x}_{\mathtt{i}}^{\mathrm{f}})} \mathbf{m}_{\mathtt{ij}}^{\mathrm{f}}, \mathbf{h}^{\mathrm{a},l}), \tag{9}$$

$$\mathbf{m}_{\mathtt{ij}}^{\mathrm{a}} = \phi_m^{\mathrm{a}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{a},l}, \mathbf{h}_{\mathtt{j}}^{\mathrm{a},l}, \|\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{j}}^{\mathrm{a},l}\|, e_{\mathtt{ij}}^{\mathrm{a}}), \qquad \mathbf{h}_{\mathtt{i}}^{\mathrm{a},l+1} = \phi_h^{\mathrm{a}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{a},l}, \sum\nolimits_{\mathtt{j} \in N(\mathbf{x}_{\mathtt{i}}^{\mathrm{a}})} \mathbf{m}_{\mathtt{ij}}^{\mathrm{a}}, \mathbf{h}^{\mathrm{f},l+1}), \tag{10}$$

$$\begin{aligned} \mathbf{x}_{\mathtt{i}}^{\mathrm{a},l+1} = \mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} &+ \sum_{\mathtt{j} \in N(\mathbf{x}_{\mathtt{i}}^{\mathrm{a}})} \frac{\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{j}}^{\mathrm{a},l}}{d_{\mathtt{ij}}^{\mathrm{a},l} + 1} \phi_x^{\mathrm{a}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{a},l+1}, \mathbf{h}_{\mathtt{j}}^{\mathrm{a},l+1}, \mathbf{m}_{\mathtt{ij}}^{\mathrm{a}}, e_{\mathtt{ij}}^{\mathrm{a}}) \\ &+ \frac{\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{k}}^{\mathrm{f}}}{\|\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{k}}^{\mathrm{f}}\| + 1} \phi_x^{\mathrm{f}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{a},l+1}, \mathbf{h}_{\mathtt{k}}^{\mathrm{f},l+1}, \|\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{k}}^{\mathrm{f}}\|), \end{aligned} \tag{11}$$

where $\mathbf{x}_{\mathtt{k}}^{\mathrm{f}}$ is the k-th row of $\mathbf{M}^\dagger \mathbf{x}_t^{\mathrm{a}}$, ϕ are multilayer perceptrons, $e_{\mathtt{ij}}^{\mathrm{a}}$ are inter-atomic bond types, and $d_{\mathtt{ij}}^{\mathrm{a},l} = \|\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{j}}^{\mathrm{a},l}\|$ are inter-atomic distances. We consider a complete graph for fragment-level interactions and expand the edge set by incorporating multi-hop and radius neighbors for atom-level interactions. The details of the deblurring networks are provided in the Appendix A.

3.3.3 Training

Following Eq. (5) of IHDM [42], our loss of previous deblurred state estimation can be defined as:

$$L_{t-1} = \mathbb{E}_{t, \mathbf{x}_0^{\mathrm{a}}, \mathbf{x}_t^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}}[\| f_{\mathbf{B}}(\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t-1) - \rho\big(\mu_\theta(\mathbf{x}_t^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, G, t)\big)\|^2], \tag{12}$$

where ρ is the Kabsch algorithm [25] to obtain the optimal rotation matrix for alignment. Through alignment ρ between the prediction from μ_θ and less blurred state $f_{\mathbf{B}}(\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t-1)$ after translating both terms to the zero center-of-mass subspace, the loss function becomes invariant to the SE(3)-transformation of the prediction.

However, we empirically observed that this previous state estimator generates unsatisfactory conformers, similar to the unsatisfactory FID scores observed in image generation of IHDM [42]. We conjectured the reason as the model limited to learn the locally small steps towards the ground truth distribution at each time step [7]. Thus, we reparameterize $\mu_\theta(\mathbf{x}_t^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, G, t)$ as

 https://doi.org/10.52202/079017-4185

Algorithm 1 Training

Sample $\hat{\mathbf{x}}^{\mathrm{f}} \sim p_{\mathrm{RDKit}}(\mathbf{x}^{\mathrm{f}})$
Sample $\mathbf{x}_0^{\mathrm{a}} \sim q(\mathbf{x}_0^{\mathrm{a}})$
Sample $t \sim \mathcal{U}[1, T]$
Sample $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$
$\mathbf{x}_t^{\mathrm{a}} \leftarrow f_{\mathbf{B}}(\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t) + \boldsymbol{\epsilon}$
Minimize $\|\mathbf{x}_0^{\mathrm{a}} - \rho\big(f_\theta(\mathbf{x}_t^{\mathrm{a}}, G, t)\big)\|^2$

Algorithm 2 Generation

Sample $\hat{\mathbf{x}}^{\mathrm{f}} \sim p_{\mathrm{RDKit}}(\mathbf{x}^{\mathrm{f}})$
$\mathbf{x}_T^{\mathrm{a}} \leftarrow \mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}$
for t in $\{T, \ldots, 1\}$ **do**
  Sample $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \delta^2\mathbf{I})$
  $\tilde{\mathbf{x}}_0^{\mathrm{a}} \leftarrow f_\theta(\mathbf{x}_t^{\mathrm{a}} + \boldsymbol{\epsilon}, \hat{\mathbf{x}}^{\mathrm{f}}, G, t)$
  $\mathbf{x}_{t-1}^{\mathrm{a}} \leftarrow f_{\mathbf{B}}(\tilde{\mathbf{x}}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t-1)$
end for

$(1 - \frac{t-1}{T})f_\theta(\mathbf{x}_t^{\mathrm{a}}, G, t) + \frac{t-1}{T}\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}$ to make the deblurring network estimates the ground truth state $\mathbf{x}_0^{\mathrm{a}}$ instead of the previous less blurred state via neural networks f_θ:

$$L_{t-1} = \mathbb{E}_{t,\mathbf{x}_0^{\mathrm{a}},\mathbf{x}_t^{\mathrm{a}},\hat{\mathbf{x}}^{\mathrm{f}}}[\|f_{\mathbf{B}}(\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t-1) - \rho\big((1 - \frac{t-1}{T})f_\theta(\mathbf{x}_t^{\mathrm{a}}, G, t) + \frac{t-1}{T}\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}\big)\|^2] \tag{13}$$

$$\approx \mathbb{E}_{t,\mathbf{x}_0^{\mathrm{a}},\mathbf{x}_t^{\mathrm{a}},\hat{\mathbf{x}}^{\mathrm{f}}}[\|\mathbf{x}_0^{\mathrm{a}} - \rho\big(f_\theta(\mathbf{x}_t^{\mathrm{a}}, G, t)\big)\|^2]. \tag{14}$$

The derivation of the new loss is detailed in Appendix B. By loss reparameterization, ρ aligns the prediction to the ground truth state. At time step t of the sampling process, after estimating ground truth $\tilde{\mathbf{x}}_0^{\mathrm{a}}$ from $\mathbf{x}_t^{\mathrm{a}}$, the next state $\mathbf{x}_{t-1}^{\mathrm{a}}$ is computed from a deterministic blurring function $f_{\mathbf{B}}$ using $\tilde{\mathbf{x}}_0^{\mathrm{a}}$. The training and sampling processes are provided in Algorithms 1, 2.

4 Related work

Hierarchical generation. A hierarchical design of generative models is evident across multiple domains, including image generation [35, 40, 16, 42] and speech synthesis [20, 17], aimed at enhancing the interpretability and quality of samples derived from coarse-grained information. In the field of computational biology, recent studies on molecular graph generation [21, 22, 12], backmapping of protein structure [61] and conformer generation [56] conditioned on the given ground truth coarse-grained information have reported the effectiveness of the hierarchical design. In recent unconditional conformer generation [38], a denoising diffusion model [19] was exclusively used in the fragment structure generation step and not designed for coarse-to-fine generation.

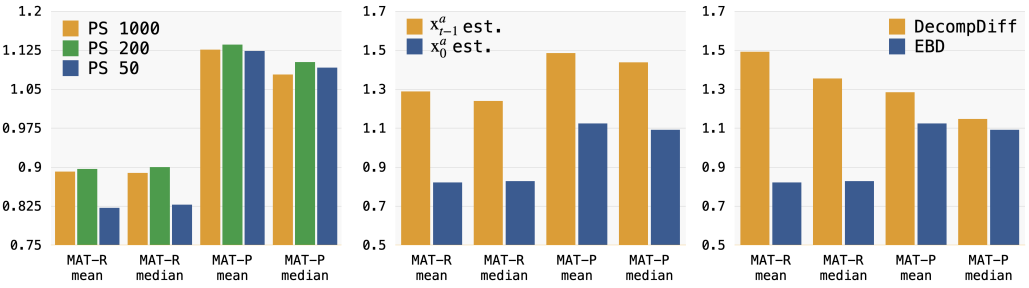
Data corruption in diffusion models. The choice of data corruption can be considered a crucial aspect of the design space of diffusion models [26], depending on the characteristics of the data domain and the specific problem definition. Recently, several studies on diffusion models have revealed that the choice of data corruption can be extended beyond random noise [49, 15, 50, 36, 43] to methods such as masking [2, 7, 8], blurring [42, 2, 7, 18], and varying data dimension [3]. We designed the data corruption process as a blurring operation in Euclidean space, transitioning from atom-level fine details to fragment-level coarse structures. This approach is more effective for multi-scale frameworks compared to random noise, which corrupts both fragment and atom geometries.

5 Experiments

In this section, we evaluate our hierarchical molecular conformer generation framework via Equivariant Blurring Diffusion (EBD) on molecular conformer generation task. We conducted experiments to answer the following questions: i) **Ablation studies** (Sec. 5.2): What are the effects of granularity of the fragment vocabulary, loss reparameterization, and data corruption processes of diffusion models? ii) **Geometric evaluation** (Sec. 5.3): Can EBD generate more diverse and accurate molecular conformers in Euclidean space than previous deep generative approaches? iii) **Property prediction** (Sec. 5.4): Can EBD generate low-energy, stable conformers?

5.1 Experiment setup

Dataset. We use GEOM-QM9 (QM9) [39] and GEOM-Drugs (Drugs) [1] which are small molecules and drug-like molecules, respectively. Each dataset comprises 40,000 molecules for the training set and 5,000 molecules for the validation set, with each molecule containing 5 conformers following data split of [46]. For the test set, we selected 200 molecules for each dataset, resulting in 22,408 and 14,324 conformers existing in QM9 and Drugs, respectively. The details of dataset were demonstrated in Appendix C.1.



(a) What level of granularity? (b) What to estimate? (c) How to corrupt?

Figure 3: Ablation studies on the motivation and design choice of EBD. (a) Fragment vocabulary granularity; (b) Target of state estimator; (c) Choice of data corruption processes.

Metrics. To measure the accuracy and diversity of the generated conformer set $\mathcal{C}$, we adopted metrics proposed by [11]. The metrics are based on root-mean-square deviation (RMSD), which is a normalized Frobenius norm between two atomic coordinate matrices aligned using the Kabsch algorithm [25]. Given the ground truth conformer set $\mathcal{C}^*$ and the generated sample set $\mathcal{C}$, four metrics that follow precision and recall are defined as:

$$\text{COV-R (Recall)} = \frac{1}{|\mathcal{C}^*|}|\{C^* \in \mathcal{C}^* | \text{RMSD}(C^*, C) \leq \delta, C \in \mathcal{C}\}|, \tag{15}$$

$$\text{MAT-R (Recall)} = \frac{1}{|\mathcal{C}^*|} \sum_{C^* \in \mathcal{C}^*} \min_{C \in \mathcal{C}} \text{RMSD}(C^*, C), \tag{16}$$

where COV and MAT are coverage metric and matching metric [57], respectively. COV quantifies the proportion of one set covered by another, with "covered" indicating RMSD values are within a threshold δ. MAT measures the average of RMSD values of one conformer set with its closest conformer in another set. If $\mathcal{C}$ and $\mathcal{C}^*$ are exchanged in Eqs. (15, 16), then metrics become COV-P (Precision) and MAT-P (Precision). The recall metric is focused on the diversity, while the precision metric measures the quality. The threshold δ is set to 0.5Å for QM9 and 1.25Å for Drugs. For each molecule, we generated conformers C that are twice the size of the ground truth conformers C^*.

Baselines. We compare EBD to existing deep generative models for molecular conformer generation. The performance of RDKit [31] that was used to generate the fragment structure of our model was measured as a baseline. Besides RDKit, machine learning models including CVGAE [34], GraphDG [47], CGCF [57], ConfVAE [58], GeoMol [11], ConfGF [46], and GeoDiff [59] were compared to our model. Among these, GeoDiff is the denoising diffusion model restoring the target distribution from random noise in atom coordinates. We adhered to their settings by configuring the maximum time step T to 5,000. For EBD, we use the $T = 50$, a noise scale of 0.01 for the forward process (σ in Eq. 6) and 0.0125 for the reverse process (δ in Eq. 8) in every experiments. The implementation details were reported in Appendix C.2.

5.2 Ablation studies

We conducted ablation studies to validate our model design, encompassing the size of the fragment vocabulary, the reparameterization of loss, and the blurring data corruption. For each ablation study, we calculated the mean and median of matching scores MAT-R and MAT-P on Drugs test set. Note that lower values of MAT-R and MAT-P indicate better results.

Effects of fragment granularity. We assessed the performance variation as fragment structure became more detailed and informative by measuring the generation performances across different fragment vocabulary sizes $|\mathcal{S}| \in \{50, 200, 1000\}$. Since PS [29], the fragmentation method we used, initializes the vocabulary from unique single atoms, reducing the size $|\mathcal{S}|$ results in obtaining finer fragments $\hat{\mathbf{x}}^f$. We reported the statistics of the average number of fragments per graph (#frags/G) and atoms per fragment (#atoms/frag) in Table 1 and the generation results in Fig. 3 (a).

Table 1: Statistics of fragment vocabulary $\mathcal{S}$ in Drugs.

$\lvert\mathcal{S}\rvert$	#frags/G	#atoms/frag
50	11.77	4.02
200	7.60	6.34
1000	5.26	9.25

https://doi.org/10.52202/079017-4185

Thanks to the increased level of detail in fragments, $|\mathcal{S}| = 50$ can obtain better performance compared to other vocabulary sizes. This is because more specific fragment structures decrease the amount of atomic-level detail that needs to be generated. From this observation, we use $|\mathcal{S}| = 50$ in all subsequent experiments. We also conducted fine-to-fine generation to observe the ability of EBD. We measured the performance when the prior is $\hat{\mathbf{x}}^{\text{a}} \sim p_{\text{RDKit}}(\mathbf{x}^{\text{a}})$. Table 2 presents coarse-to-fine (C2F) result when $|\mathcal{S}| = 50$ and fine-to-fine (F2F) result. As expected, F2F shows more accurate results compared to C2F as it includes more detail in the prior distribution.

Table 2: Fine-to-fine generation on Drugs.

	COV-R (%) ↑		MAT-R(Å) ↓		COV-P (%) ↑		MAT-P (Å) ↓	
	Mean	Med	Mean	Med	Mean	Med	Mean	Med
C2F	**92.60**	**98.73**	0.8216	0.8279	66.24	68.39	1.1237	1.0916
F2F	89.44	**98.73**	**0.7986**	**0.7710**	**76.62**	**88.64**	**1.0090**	**0.9397**

Effects of loss reparameterization. We presented the performance comparison between the less blurred previous state estimator in Eq. (12) and ground truth estimator in Eq. (14) after loss reprarameterization in Fig. 3 (b). From the previous state estimator, we acquired degenerated conformers with relatively high matching scores, which align with low FID score of IHDM [42] in image generation. On the other hand, we observed distinct advantages in introducing the ground truth estimator across all metrics. We speculate that the ground truth estimator facilitates the diffusion model in learning beyond locally blurring distributions towards the target distribution.

Effects of data corruptions. We provide the same initial fragment structures $\hat{\mathbf{x}}^{\text{f}}$ to both EBD and DecompDiff [13] so that the data corruption method becomes the primary distinction to examine. DecompDiff is a better candidate than GeoDiff [59] to compare for this purpose because, unlike GeoDiff that generates conformers from a single prior distribution, DecompDiff denoises multiple prior distributions, where each mean corresponds to the coordinates of each fragment $\hat{\mathbf{x}}^{\text{f}}$. The generation results and sampling trajectories are compared between the two models ($T = 50$ for both) in Fig. 3 (c) and Fig. 4. At first, we observed that the conformers generated from DecompDiff exhibit lower diversity scores compared to EBD. This is because the results of DecompDiff tend to adhere closely to the approximate fragment structure $\hat{\mathbf{x}}^{\text{f}}$, whereas EBD attempts to transition towards the ground truth fragment structure $\mathbf{x}^{\text{f}}$. We speculate that our blurring schedule, which entails a linear interpolation between $\mathbf{M}\hat{\mathbf{x}}^{\text{f}}$ and $\mathbf{x}_0^{\text{a}}$, facilitates the learning process for the diffusion model compared to a stochastic trajectory between prior and target distributions. As empirical evidence, we observed that DecompDiff primarily focuses on denoising the fragment structure throughout most of the sampling process in Fig. 4. On the other hand, EBD focuses on the entirety of the sampling process to generate fine details, resulting in better quality.

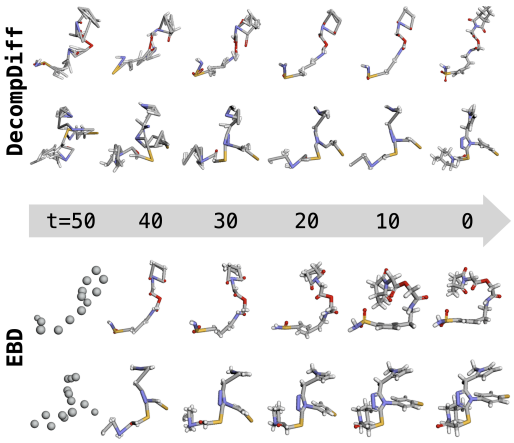


Figure 4: Sampling processes of two conformers depending on data corruptions.

We conducted further analysis to observe how accurately the fragment coordinates generated by RDKit were corrected toward the ground truth by EBD. For 200 molecules in Drugs test set, we measured RMSD($\mathbf{x}^{\text{f}}_{\text{RDKit}}, \mathbf{x}^{\text{f}}_{\text{gt}}$) and RMSD($\mathbf{x}^{\text{f}}_{\text{EBD}}, \mathbf{x}^{\text{f}}_{\text{gt}}$). If RMSD($\mathbf{x}^{\text{f}}_{\text{EBD}}, \mathbf{x}^{\text{f}}_{\text{gt}}$) is lower than RMSD($\mathbf{x}^{\text{f}}_{\text{RDKit}}, \mathbf{x}^{\text{f}}_{\text{gt}}$) for a molecule, then it indicates that the model has corrected the fragment coordinates. In Fig. 5, the points below the red line represent cases where the model corrected the coordinates accurately. We observed that the greater the RMSD($\mathbf{x}^{\text{f}}_{\text{RDKit}}, \mathbf{x}^{\text{f}}_{\text{gt}}$) (points further to the right on the x-axis), the larger the reduction in RMSD($\mathbf{x}^{\text{f}}_{\text{EBD}}, \mathbf{x}^{\text{f}}_{\text{gt}}$). In other words, the lower the quality of the coarse-grained prior, the more accurately the model tends to make corrections.

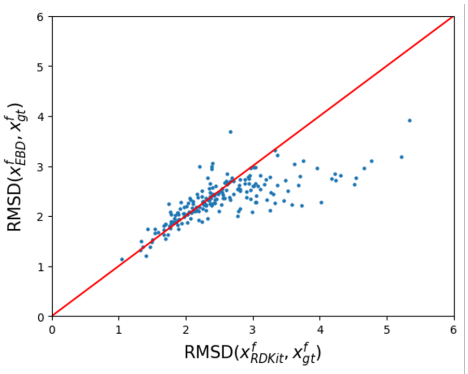


Figure 5: Correction on fragment coordinates

https://doi.org/10.52202/079017-4185

Table 3: Geometric evaluation on Drugs ($\delta = 1.25$Å).

Models	COV-R (%) ↑		MAT-R(Å) ↓		COV-P (%) ↑		MAT-P (Å) ↓	
	Mean	Med	Mean	Med	Mean	Med	Mean	Med
RDKit	45.74	31.75	1.5376	1.4004	54.78	59.48	1.3341	1.1996
CVGAE	0.00	0.00	3.0702	2.9937	-	-	-	-
GraphDG	8.27	0.00	1.9722	1.9845	2.08	0.00	2.4340	2.4100
CGCF	53.96	57.06	1.2487	1.2247	21.68	13.72	1.8571	1.8066
ConfVAE	55.20	59.43	1.2380	1.1417	22.96	14.05	1.8287	1.8159
GeoMol	67.16	71.71	1.0875	1.0586	-	-	-	-
ConfGF	62.15	70.93	1.1629	1.1596	23.42	15.52	1.7219	1.6863
GeoDiff ($T = 5000$)	89.40	96.86	0.8571	0.8495	61.28	65.00	1.1642	1.1272
EBD ($T = 50$)	**92.60**	**98.73**	**0.8216**	**0.8279**	**66.24**	**68.39**	**1.1237**	**1.0916**

Additional visualizations of sampling results on Drugs and QM9, as well as the sampling processes, are illustrated in the Appendix E.

5.3 Geometric evaluation

We compared our hierarchical framework to the baseline RDKit and machine learning models for molecular conformer generation on Drugs, and the results are reported in Table 3. EBD achieves superior performance across all metrics by generating diverse and accurate molecular conformers. In comparison to RDKit, which was used to generate fragment structures $\hat{\mathbf{x}}^{\mathrm{f}}$, EBD achieved a significant improvement in the generation of diverse fine atomic details, as evidenced by higher COV-R and MAT-R scores. We also observed that, due to the informative fragment structure prior distribution and the proposed blurring schedule, EBD produces more diverse and higher-quality conformers with statistical significance (see Appendix D), even with 100 times fewer T compared to GeoDiff. We also reported EBD's better performance on QM9 with statistical significance in Appendix D.

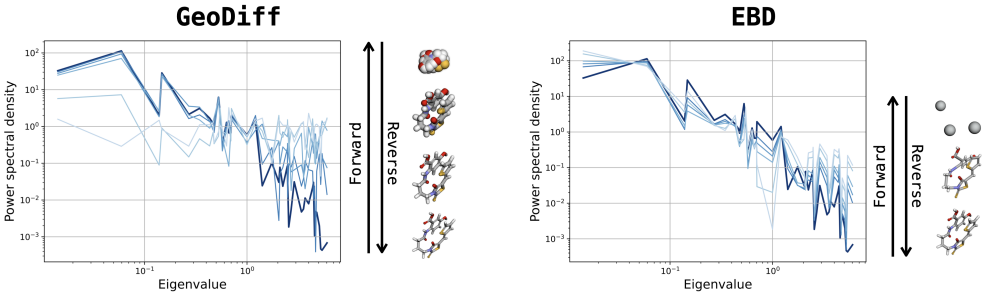


Figure 6: Power spectral density analysis on forward processes of GeoDiff and EBD. The darkest line is the PSD coefficient of $\mathbf{x}_0^{\mathrm{a}}$ and the lines become lighter as $t \to T$.

We take an additional step to delve into the rationale behind the ability of EBD to achieve better performance with smaller T by analyzing the spectral domain in Fig 6. Motivated from the analysis of [42], we calculated the power spectral density, $\mathrm{PSD}(\mathbf{x}_t^{\mathrm{a}})_{\mathtt{i}} = \frac{1}{3}\sum_{c\in\{x,y,z\}} |\mathbf{V}_{\mathtt{i}}^T \mathbf{x}_t^{\mathrm{a},c}|^2$, where $\mathbf{V}_{\mathtt{i}}$ is the i-th smallest eigenvector of graph Laplacian Δ of the molecular graph G, and $\mathbf{x}_t^{\mathrm{a},c}$ is one of the $\{x, y, z\}$ coordinate vectors of atomic coordinate matrix $\mathbf{x}_t^{\mathrm{a}}$. The smaller eigenvalues correspond to coarser-grained structures, while higher eigenvalues correspond to finer details [4]. Therefore, by measuring $\mathrm{PSD}(\mathbf{x}_t^{\mathrm{a}})_{\mathtt{i}}$ during the forward processes $\{\mathbf{x}_t^{\mathrm{a}}\}_{t=0}^T$ of EBD and GeoDiff, we can ascertain which structural components are corrupted during the forward process and will be restored during the reverse process, respectively. For EBD, there is less significant perturbation across lower frequency parts of the PSD that corresponds to the coarser-grained discrepancy between ground truth $\mathbf{x}^{\mathrm{f}}$ and approximate $\hat{\mathbf{x}}^{\mathrm{f}}$ fragment positions. Thus, EBD primarily focuses on restoring perturbed fine-grained structures in the higher frequency parts throughout the entire generative process. This explains why EBD does not require excessive T. In contrast, GeoDiff requires relatively more T because random noise corrupts the overall structural information, and the amount of perturbations is also significant.

https://doi.org/10.52202/079017-4185

5.4 Chemical evaluation

In addition to geometric evaluation, we assessed the quality of generated conformers by their chemical properties. After training EBD on QM9, following [46], we generated 50 samples for each of the 30 molecules, which constitute a subset of QM9. Using PSI4 [48], we calculated properties of each conformer including the energy $\overline{E}$, the lowest energy E_{min}, HOMO-LUMO gap ϵ, the average gap $\overline{\Delta\epsilon}$, the minimum gap $\Delta\epsilon_{\text{min}}$ and the maximum gap $\Delta\epsilon_{\text{max}}$. Then, we measured the mean absolute errors between the properties of generated and ground truth (Table. 4). We observed that EBD can generate the most stable conformers compared to other methods, as evidenced by lower energy and HOMO-LUMO gap.

Table 4: Mean absolute errors between generated and ground truth ensemble properties in eV.

Models	$\overline{E}$	E_{min}	$\overline{\Delta\epsilon}$	$\Delta\epsilon_{\text{min}}$	$\Delta\epsilon_{\text{max}}$
RDKit	0.9233	0.6585	0.3698	0.8021	0.2359
GraphDG	9.1027	0.8882	1.7973	4.1743	0.4776
CGCF	28.9661	2.8410	2.8356	10.6361	0.5954
ConfVAE	8.2080	0.6100	1.6080	3.9111	0.2429
ConfGF	2.7886	0.1765	0.4688	2.1843	**0.1433**
GeoDiff	0.2597	0.1551	0.3091	0.7033	0.1909
EBD	**0.1812**	**0.1214**	**0.1253**	**0.5306**	0.2153

6 Conclusion

We introduced a novel hierarchical generative model for molecular conformers via Equivariant Blurring Diffusion (EBD), a diffusion model designed for coarse-to-fine generative scheme. After generating the initial distribution of fragment coordinates from a cheminformatics tool, EBD generated fine atomic details from coarse-grained structures through equivariant networks. We also proposed a simple and effective linear blurring scheduler and ground truth state estimator to enhance the model's ability to produce diverse and accurate conformers. Through extensive analysis of the proposed model and comparison between competitive denoising diffusion models, we substantiated the validity of the model design.

As a coarse-to-fine generative scheme, EBD can be extended for larger and more complex molecular structures, because hierarchical systems similar to those utilized in EBD exists widely across molecular systems (ranging from proteins as linear polymers of amino acids to materials as lattices of molecules). We discussed a few limitations of our model in the Appendix F.

Acknowledgments and Disclosure of Funding

We thank Yuning You and Hoseok Do for valuable feedback on the manuscript. This work was partially funded by the U.S. National Science Foundation under grant CCF-1943008. Portions of this research were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing.

References

[1] Axelrod, Simon and Gomez-Bombarelli, Rafael. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):185, 2022.

[2] Bansal, Arpit, Borgnia, Eitan, Chu, Hong-Min, Li, Jie, Kazemi, Hamid, Huang, Furong, Goldblum, Micah, Geiping, Jonas, and Goldstein, Tom. Cold diffusion: Inverting arbitrary image transforms without noise. *Advances in Neural Information Processing Systems*, 36, 2023.

[3] Campbell, Andrew, Harvey, William, Weilbach, Christian, De Bortoli, Valentin, Rainforth, Thomas, and Doucet, Arnaud. Trans-dimensional generative modeling via jump diffusion models. *Advances in Neural Information Processing Systems*, 36, 2023.

[4] Chung, Fan RK. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[5] Crippen, Gordon M, Havel, Timothy F, et al. *Distance geometry and molecular conformation*, volume 74. Research Studies Press Taunton, 1988.

[6] Crouse, David F. On implementing 2d rectangular assignment algorithms. *IEEE Transactions on Aerospace and Electronic Systems*, 52(4):1679–1696, 2016.

[7] Daras, Giannis, Delbracio, Mauricio, Talebi, Hossein, Dimakis, Alexandros, and Milanfar, Peyman. Soft diffusion: Score matching with general corruptions. *Transactions on Machine Learning Research*, 2023.

[8] Daras, Giannis, Shah, Kulin, Dagan, Yuval, Gollakota, Aravind, Dimakis, Alex, and Klivans, Adam. Ambient diffusion: Learning clean distributions from corrupted data. *Advances in Neural Information Processing Systems*, 36, 2023.

[9] Degen, Jorg, Wegscheid-Gerlach, Christof, Zaliani, Andrea, and Rarey, Matthias. On the art of compiling and using'drug-like'chemical fragment spaces. *ChemMedChem*, 3(10):1503, 2008.

[10] Du, Weitao, Zhang, He, Du, Yuanqi, Meng, Qi, Chen, Wei, Zheng, Nanning, Shao, Bin, and Liu, Tie-Yan. Se (3) equivariant graph neural networks with complete local frames. In *International Conference on Machine Learning*, pp. 5583–5608. PMLR, 2022.

[11] Ganea, Octavian, Pattanaik, Lagnajit, Coley, Connor, Barzilay, Regina, Jensen, Klavs, Green, William, and Jaakkola, Tommi. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. *Advances in Neural Information Processing Systems*, 34, 2021.

[12] Geng, Zijie, Xie, Shufang, Xia, Yingce, Wu, Lijun, Qin, Tao, Wang, Jie, Zhang, Yongdong, Wu, Feng, and Liu, Tie-Yan. De novo molecular generation via connection-aware motif mining. In *The Eleventh International Conference on Learning Representations*, 2023.

[13] Guan, Jiaqi, Zhou, Xiangxin, Yang, Yuwei, Bao, Yu, Peng, Jian, Ma, Jianzhu, Liu, Qiang, Wang, Liang, and Gu, Quanquan. Decompdiff: Diffusion models with decomposed priors for structure-based drug design. In *International Conference on Machine Learning*, pp. 11827–11846. PMLR, 2023.

[14] Heusel, Martin, Ramsauer, Hubert, Unterthiner, Thomas, Nessler, Bernhard, and Hochreiter, Sepp. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, 30, 2017.

[15] Ho, Jonathan, Jain, Ajay, and Abbeel, Pieter. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 2020.

[16] Ho, Jonathan, Saharia, Chitwan, Chan, William, Fleet, David J, Norouzi, Mohammad, and Salimans, Tim. Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research*, 23(47):1–33, 2022.

[17] Hono, Yukiya, Tsuboi, Kazuna, Sawada, Kei, Hashimoto, Kei, Oura, Keiichiro, Nankaku, Yoshihiko, and Tokuda, Keiichi. Hierarchical multi-grained generative model for expressive speech synthesis. *Interspeech 2020*, 2020.

[18] Hoogeboom, Emiel and Salimans, Tim. Blurring diffusion models. In *The Eleventh International Conference on Learning Representations*, 2023.

[19] Hoogeboom, Emiel, Satorras, Vıctor Garcia, Vignac, Clément, and Welling, Max. Equivariant diffusion for molecule generation in 3d. In *International Conference on Machine Learning*, pp. 8867–8887. PMLR, 2022.

[20] Hsu, Wei-Ning, Zhang, Yu, Weiss, Ron, Zen, Heiga, Wu, Yonghui, Cao, Yuan, and Wang, Yuxuan. Hierarchical generative modeling for controllable speech synthesis. In *International Conference on Learning Representations*, 2019.

[21] Jin, Wengong, Barzilay, Regina, and Jaakkola, Tommi. Junction tree variational autoencoder for molecular graph generation. In *International Conference on Machine Learning*, pp. 2323–2332. PMLR, 2018.

[22] Jin, Wengong, Barzilay, Regina, and Jaakkola, Tommi. Hierarchical generation of molecular graphs using structural motifs. In *International Conference on Machine Learning*, pp. 4839–4848. PMLR, 2020.

[23] Jing, Bowen, Corso, Gabriele, Chang, Jeffrey, Barzilay, Regina, and Jaakkola, Tommi. Torsional diffusion for molecular conformer generation. *Advances in Neural Information Processing Systems*, 35, 2022.

 https://doi.org/10.52202/079017-4185

[24] Joshi, Chaitanya K, Bodnar, Cristian, Mathis, Simon V, Cohen, Taco, and Lio, Pietro. On the expressive power of geometric graph neural networks. In *International Conference on Machine Learning*, pp. 15330–15355. PMLR, 2023.

[25] Kabsch, Wolfgang. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32 (5):922–923, 1976.

[26] Karras, Tero, Aittala, Miika, Aila, Timo, and Laine, Samuli. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, 35, 2022.

[27] Kingma, Diederik, Salimans, Tim, Poole, Ben, and Ho, Jonathan. Variational diffusion models. *Advances in Neural Information Processing Systems*, 34, 2021.

[28] Köhler, Jonas, Klein, Leon, and Noé, Frank. Equivariant flows: exact likelihood generative learning for symmetric densities. In *International Conference on Machine Learning*, pp. 5361–5370. PMLR, 2020.

[29] Kong, Xiangzhe, Huang, Wenbing, Tan, Zhixing, and Liu, Yang. Molecule generation by principal subgraph mining and assembling. *Advances in Neural Information Processing Systems*, 35, 2022.

[30] Krizhevsky, Alex, Hinton, Geoffrey, et al. Learning multiple layers of features from tiny images. 2009.

[31] Landrum, Greg et al. Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*, 8(31.10):5281, 2013.

[32] Lewell, Xiao Qing, Judd, Duncan B, Watson, Stephen P, and Hann, Michael M. Recap retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *Journal of chemical information and computer sciences*, 38(3):511–522, 1998.

[33] Loshchilov, Ilya and Hutter, Frank. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

[34] Mansimov, Elman, Mahmood, Omar, Kang, Seokho, and Cho, Kyunghyun. Molecular geometry prediction using a deep generative graph neural network. *Scientific reports*, 9(1):20381, 2019.

[35] Menick, Jacob and Kalchbrenner, Nal. Generating high fidelity images with subscale pixel networks and multidimensional upscaling. In *International Conference on Learning Representations*, 2019.

[36] Nichol, Alexander Quinn and Dhariwal, Prafulla. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pp. 8162–8171. PMLR, 2021.

[37] Paszke, Adam, Gross, Sam, Chintala, Soumith, Chanan, Gregory, Yang, Edward, DeVito, Zachary, Lin, Zeming, Desmaison, Alban, Antiga, Luca, and Lerer, Adam. Automatic differentiation in pytorch. 2017.

[38] Qiang, Bo, Song, Yuxuan, Xu, Minkai, Gong, Jingjing, Gao, Bowen, Zhou, Hao, Ma, Wei-Ying, and Lan, Yanyan. Coarse-to-fine: a hierarchical diffusion model for molecule generation in 3d. In *International Conference on Machine Learning*, pp. 28277–28299. PMLR, 2023.

[39] Ramakrishnan, Raghunathan, Dral, Pavlo O, Rupp, Matthias, and Von Lilienfeld, O Anatole. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

[40] Razavi, Ali, Van den Oord, Aaron, and Vinyals, Oriol. Generating diverse high-fidelity images with vq-vae-2. *Advances in Neural Information Processing Systems*, 32, 2019.

[41] Reidenbach, Danny and Krishnapriyan, Aditi. Coarsenconf: Equivariant coarsening with aggregated attention for molecular conformer generation. In *NeurIPS 2023 Generative AI and Biology (GenBio) Workshop*, 2023.

[42] Rissanen, Severi, Heinonen, Markus, and Solin, Arno. Generative modelling with inverse heat dissipation. In *The Eleventh International Conference on Learning Representations*, 2023.

[43] Rombach, Robin, Blattmann, Andreas, Lorenz, Dominik, Esser, Patrick, and Ommer, Björn. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

[44] Saharia, Chitwan, Chan, William, Saxena, Saurabh, Li, Lala, Whang, Jay, Denton, Emily L, Ghasemipour, Kamyar, Gontijo Lopes, Raphael, Karagol Ayan, Burcu, Salimans, Tim, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35, 2022.

[45] Satorras, Vıctor Garcia, Hoogeboom, Emiel, and Welling, Max. E (n) equivariant graph neural networks. In *International Conference on Machine Learning*, pp. 9323–9332. PMLR, 2021.

[46] Shi, Chence, Luo, Shitong, Xu, Minkai, and Tang, Jian. Learning gradient fields for molecular conformation generation. In *International Conference on Machine Learning*, pp. 9558–9568. PMLR, 2021.

[47] Simm, Gregor and Hernandez-Lobato, Jose Miguel. A generative model for molecular distance geometry. In *International Conference on Machine Learning*, pp. 8949–8958. PMLR, 2020.

[48] Smith, Daniel GA, Burns, Lori A, Simmonett, Andrew C, Parrish, Robert M, Schieber, Matthew C, Galvelis, Raimondas, Kraus, Peter, Kruse, Holger, Di Remigio, Roberto, Alenaizan, Asem, et al. Psi4 1.4: Open-source software for high-throughput quantum chemistry. *The Journal of chemical physics*, 152(18), 2020.

[49] Sohl-Dickstein, Jascha, Weiss, Eric, Maheswaranathan, Niru, and Ganguli, Surya. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pp. 2256–2265. PMLR, 2015.

[50] Song, Jiaming, Meng, Chenlin, and Ermon, Stefano. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2020.

[51] Song, Yang and Ermon, Stefano. Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 32, 2019.

[52] Song, Yang, Sohl-Dickstein, Jascha, Kingma, Diederik P, Kumar, Abhishek, Ermon, Stefano, and Poole, Ben. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2020.

[53] Thomas, Nathaniel, Smidt, Tess, Kearnes, Steven, Yang, Lusann, Li, Li, Kohlhoff, Kai, and Riley, Patrick. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

[54] Vahdat, Arash, Kreis, Karsten, and Kautz, Jan. Score-based generative modeling in latent space. *Advances in Neural Information Processing Systems*, 34, 2021.

[55] Vahdat, Arash, Williams, Francis, Gojcic, Zan, Litany, Or, Fidler, Sanja, Kreis, Karsten, et al. Lion: Latent point diffusion models for 3d shape generation. *Advances in Neural Information Processing Systems*, 35:10021–10039, 2022.

[56] Wang, Wujie, Xu, Minkai, Cai, Chen, Miller, Benjamin K, Smidt, Tess, Wang, Yusu, Tang, Jian, and Gomez-Bombarelli, Rafael. Generative coarse-graining of molecular conformations. In *International Conference on Machine Learning*, pp. 23213–23236. PMLR, 2022.

[57] Xu, Minkai, Luo, Shitong, Bengio, Yoshua, Peng, Jian, and Tang, Jian. Learning neural generative dynamics for molecular conformation generation. In *International Conference on Learning Representations*, 2021.

[58] Xu, Minkai, Wang, Wujie, Luo, Shitong, Shi, Chence, Bengio, Yoshua, Gomez-Bombarelli, Rafael, and Tang, Jian. An end-to-end framework for molecular conformation generation via bilevel programming. In *International Conference on Machine Learning*, pp. 11537–11547. PMLR, 2021.

 https://doi.org/10.52202/079017-4185

[59] Xu, Minkai, Yu, Lantao, Song, Yang, Shi, Chence, Ermon, Stefano, and Tang, Jian. Geodiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2022.

[60] Xu, Minkai, Powers, Alexander S, Dror, Ron O, Ermon, Stefano, and Leskovec, Jure. Geometric latent diffusion models for 3d molecule generation. In *International Conference on Machine Learning*, pp. 38592–38610. PMLR, 2023.

[61] Yang, Soojung and Gómez-Bombarelli, Rafael. Chemically transferable generative backmapping of coarse-grained proteins. In *International Conference on Machine Learning*, pp. 39277–39298, 2023.

[62] Zhu, Jinhua, Xia, Yingce, Liu, Chang, Wu, Lijun, Xie, Shufang, Wang, Yusong, Wang, Tong, Qin, Tao, Zhou, Wengang, Li, Houqiang, et al. Direct molecular conformation generation. *Transactions on Machine Learning Research*, 2022.

A Deblurring network architectures

In SE(3)-equivariant deblurring networks, there are update functions of SE(3)-invariant fragment and atom features $\mathbf{h}^{\mathrm{f}}$, $\mathbf{h}^{\mathrm{a}}$, as well as an update function of SE(3)-equivariant atom coordinates $\mathbf{x}^{\mathrm{a}}$ motivated from equivariant graph neural networks [45]. For the fragments, we constructed a complete graph to account for dense interactions among them. In the case of atoms, we expanded the neighbor set of each atom by including multi-hop neighbors derived from the powers of the adjacency matrix and a radius graph, which includes atoms within a specified cutoff distance. The benefits of dense interactions for accurate conformers estimation have been confirmed in several studies [47, 59, 19].

The architecture of the SE(3)-invariant message passing and feature update functions at the fragment- and atom-level is as follows:

$$\mathbf{m}_{\mathtt{ij}}^{\mathrm{f}} = \phi_m^{\mathrm{f}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{f},l}, \mathbf{h}_{\mathtt{j}}^{\mathrm{f},l}, \|\mathbf{x}_{\mathtt{i}}^{\mathrm{f}} - \mathbf{x}_{\mathtt{j}}^{\mathrm{f}}\|), \qquad \mathbf{h}_{\mathtt{i}}^{\mathrm{f},l+1} = \phi_h^{\mathrm{f}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{f},l}, \sum\nolimits_{\mathtt{j}\in N(\mathbf{x}_{\mathtt{i}}^{\mathrm{f}})} \mathbf{m}_{\mathtt{ij}}^{\mathrm{f}}, \mathbf{h}^{\mathrm{a},l}), \tag{A.1}$$

$$\mathbf{m}_{\mathtt{ij}}^{\mathrm{a}} = \phi_m^{\mathrm{a}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{a},l}, \mathbf{h}_{\mathtt{j}}^{\mathrm{a},l}, \|\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{j}}^{\mathrm{a},l}\|, e_{\mathtt{ij}}^{\mathrm{a}}), \quad \mathbf{h}_{\mathtt{i}}^{\mathrm{a},l+1} = \phi_h^{\mathrm{a}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{a},l}, \sum\nolimits_{\mathtt{j}\in N(\mathbf{x}_{\mathtt{i}}^{\mathrm{a}})} \mathbf{m}_{\mathtt{ij}}^{\mathrm{a}}, \mathbf{h}^{\mathrm{f},l+1}), \tag{A.2}$$

where $\mathbf{m}_{\mathtt{ij}} \in \mathbb{R}^d$ is the message for each interactions, and $\mathbf{h} \in \mathbb{R}^d$ is the feature vector from the aggregated messages and features from different hierarchy level. For every invariant update functions $\phi_m^{\mathrm{f}}, \phi_h^{\mathrm{f}}, \phi_m^{\mathrm{a}}, \phi_h^{\mathrm{a}}$, we used multilayer perceptrons. For initial features $\mathbf{h}_{\mathtt{i}}^{\mathrm{f},0}$ of fragments, we defined a 3-dimensional vector as a frequency histogram of its constituent atom types based on their chemical properties, including hydrophobicity, hydrogen bond center, and negative charge center following [38]. The detailed definition of the initial fragment features is in Table 5. For initial atom features $\mathbf{h}_{\mathtt{i}}^{\mathrm{a},0} \in \mathbb{R}^d$ and bond features $e_{\mathtt{ij}}^{\mathrm{a}} \in \mathbb{R}^d$, we used embeddings from atom types and bond types, respectively.

Table 5: Initial fragment feature based on chemical properties

Properties	Details	Types
Hydrophobicity	Frequency of C element	Integer
Hydrogen bond center	Frequency of O, N, S, P elements	Integer
Negative charge center	Frequency of F, Cl, Br, I elements	Integer

For the i-th atom $\mathbf{x}_{\mathtt{i}}^{\mathrm{a}}$ belongs to the k-th fragment $\mathbf{x}_{\mathtt{k}}^{\mathrm{f}}$, the architecture of the equivariant atom coordinate update function is as follows:

$$\begin{aligned}\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l+1} = \mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} &+ \sum_{\mathtt{j}\in N(\mathbf{x}_{\mathtt{i}}^{\mathrm{a}})} \frac{\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{j}}^{\mathrm{a},l}}{d_{\mathtt{ij}}^{\mathrm{a},l} + 1} \phi_x^{\mathrm{a}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{a},l+1}, \mathbf{h}_{\mathtt{j}}^{\mathrm{a},l+1}, \mathbf{m}_{\mathtt{ij}}^{\mathrm{a}}, e_{\mathtt{ij}}^{\mathrm{a}}) \\ &+ \frac{\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{k}}^{\mathrm{f}}}{\|\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{k}}^{\mathrm{f}}\| + 1} \phi_x^{\mathrm{f}}(\mathbf{h}_{\mathtt{i}}^{\mathrm{a},l+1}, \mathbf{h}_{\mathtt{k}}^{\mathrm{f},l+1}, \|\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{k}}^{\mathrm{f}}\|),\end{aligned} \tag{A.3}$$

where $\mathbf{x}_{\mathtt{k}}^{\mathrm{f}}$ is the k-th row of $\mathbf{M}^{\dagger}\mathbf{x}_t^{\mathrm{a}}$, and $d_{\mathtt{ij}}^{\mathrm{a},l} = \|\mathbf{x}_{\mathtt{i}}^{\mathrm{a},l} - \mathbf{x}_{\mathtt{j}}^{\mathrm{a},l}\|$ are inter-atomic distances. For every equivariant update functions $\phi_x^{\mathrm{a}}, \phi_x^{\mathrm{f}}$, we used multilayer perceptrons. For three terms in right-hand side of Eq. (A.3), the first term is the coordinate from the previous layer, the second term is an equivariant update function that accounts for atom-level interactions, and the third term is an equivariant update function that considers the deviation of the current atom coordinate from its respective fragment's coordinate.

B Derivation of loss function

In this section, we explain the derivation of the loss function for the ground truth state estimator from the previous state estimator. The loss function of previous state estimation is defined as:

$$L_{t-1} = \mathbb{E}_{t,\mathbf{x}_0^{\mathrm{a}},\mathbf{x}_t^{\mathrm{a}},\hat{\mathbf{x}}^{\mathrm{f}}}[\|f_{\mathbf{B}}(\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t-1) - \rho\big(\mu_\theta(\mathbf{x}_t^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, G, t)\big)\|^2], \tag{A.4}$$

where ρ is the Kabsch algorithm [25] to obtain the optimal rotation matrix for alignment. Through alignment ρ of the prediction from μ_θ to the less blurred state $f_{\mathbf{B}}(\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t-1)$ after translating both terms to the zero center-of-mass subspace, the loss function becomes invariant to the translation and rotation of the prediction.

https://doi.org/10.52202/079017-4185

However, this previous state estimator generates unsatisfactory conformers as empirically observed in Sec. 5.2. We conjectured the reason as the model limited to learn the locally small steps towards the ground truth distribution at each time step [7]. Thus, we reparameterize $\mu_\theta(\mathbf{x}_t^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, G, t)$ as $(1-\frac{t-1}{T})f_\theta(\mathbf{x}_t^{\mathrm{a}}, G, t) + \frac{t-1}{T}\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}$ to make the deblurring network estimates the ground truth state $\mathbf{x}_0^{\mathrm{a}}$ instead of the previous less blurred state via neural networks f_θ. We first start with the non-invariant previous state estimation, which is without the alignment ρ:

$$L_{t-1} = \mathbb{E}_{t,\mathbf{x}_0^{\mathrm{a}},\mathbf{x}_t^{\mathrm{a}},\hat{\mathbf{x}}^{\mathrm{f}}}[\|f_{\mathbf{B}}(\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t-1) - \mu_\theta(\mathbf{x}_t^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, G, t)\|^2], \tag{A.5}$$

$$= \mathbb{E}_{t,\mathbf{x}_0^{\mathrm{a}},\mathbf{x}_t^{\mathrm{a}},\hat{\mathbf{x}}^{\mathrm{f}}}[\|f_{\mathbf{B}}(\mathbf{x}_0^{\mathrm{a}}, \hat{\mathbf{x}}^{\mathrm{f}}, t-1) - (1-\tfrac{t-1}{T})f_\theta(\mathbf{x}_t^{\mathrm{a}}, G, t) - \tfrac{t-1}{T}\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}\|^2] \tag{A.6}$$

$$= \mathbb{E}_{t,\mathbf{x}_0^{\mathrm{a}},\mathbf{x}_t^{\mathrm{a}},\hat{\mathbf{x}}^{\mathrm{f}}}[\|(1-\tfrac{t-1}{T})\mathbf{x}_0^{\mathrm{a}} + \tfrac{t-1}{T}\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}} - (1-\tfrac{t-1}{T})f_\theta(\mathbf{x}_t^{\mathrm{a}}, G, t) - \tfrac{t-1}{T}\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}}\|^2] \tag{A.7}$$

$$= \mathbb{E}_{t,\mathbf{x}_0^{\mathrm{a}},\mathbf{x}_t^{\mathrm{a}},\hat{\mathbf{x}}^{\mathrm{f}}}[\|(1-\tfrac{t-1}{T})(\mathbf{x}_0^{\mathrm{a}} - f_\theta(\mathbf{x}_t^{\mathrm{a}}, G, t)) + \tfrac{t-1}{T}(\mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}} - \mathbf{M}\hat{\mathbf{x}}^{\mathrm{f}})\|^2] \tag{A.8}$$

$$= \mathbb{E}_{t,\mathbf{x}_0^{\mathrm{a}},\mathbf{x}_t^{\mathrm{a}},\hat{\mathbf{x}}^{\mathrm{f}}}(1-\tfrac{t-1}{T})^2[\|\mathbf{x}_0^{\mathrm{a}} - f_\theta(\mathbf{x}_t^{\mathrm{a}}, G, t)\|^2] \tag{A.9}$$

$$\approx \mathbb{E}_{t,\mathbf{x}_0^{\mathrm{a}},\mathbf{x}_t^{\mathrm{a}},\hat{\mathbf{x}}^{\mathrm{f}}}[\|\mathbf{x}_0^{\mathrm{a}} - \rho\big(f_\theta(\mathbf{x}_t^{\mathrm{a}}, G, t)\big)\|^2]. \tag{A.10}$$

In the last stage from Eq. (A.9) to Eq. (A.10), we simplified the loss function by discarding the time-dependent weight as [15]. Finally, we make the loss function for ground truth estimation invariant by aligning the prediction from f_θ to the ground truth state using Kabsch alignment ρ [25].

C Implementation details

C.1 Datasets

We used GEOM-QM9 (QM9) [39] and GEOM-Drugs (Drugs) [1] for analysis and comparison between molecular conformer generation models. Each dataset comprises 40,000 molecules for the training set and 5,000 molecules for the validation set, with each molecule containing 5 conformers following data split of [46]. We obtained the raw data, the pre-processed data and the data split at `https://github.com/DeepGraphLearning/ConfGF`. For the test set, we selected 200 molecules for each dataset, resulting in 22,408 and 14,324 conformers existing in QM9 and Drugs, respectively.

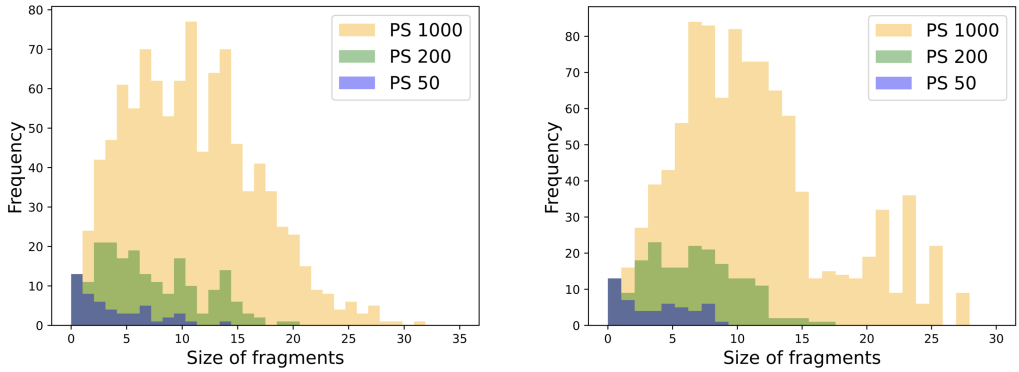
For fragmentation of the molecular graphs $G = (\mathcal{V}, \mathcal{E})$ in Drugs and QM9, we used Principal Subgraph (PS) [29] (`https://github.com/THUNLP-MT/PS-VAE`) which can construct a fragment vocabulary $\mathcal{S}$ whose elements are the largest and frequent repetitive subgraphs of molecules. Starting from all unique atoms in $\mathcal{S}$ at initial stage, PS iteratively merges neighboring fragments. The most frequent fragment among the newly merged fragments was added to the vocabulary at each iteration, and this operation was repeated until the desired size of the vocabulary was reached. Thus, the smaller the fragment vocabulary, the finer fragments can be obtained. One of the advantages of PS compared to existing fragmentation methods such as RECAP [32], BRICS [9], junction tree decomposition [21] is the ability to control the vocabulary size, allowing us to observe how performance varies with fragment granularity. We constructed $\mathcal{S}$ for each dataset with three fragment vocabulary sizes $|\mathcal{S}| \in \{50, 200, 1000\}$. The average numbers of fragments per graph ($\#\mathrm{frags}/G$) and atoms per fragment ($\#\mathrm{atoms}/\mathrm{frag}$) of Drugs and QM9 were reported in Table 6.

Table 6: Statistics of fragment vocabulary $\mathcal{S}$.

	Drugs		QM9	
$\lvert\mathcal{S}\rvert$	$\#\mathrm{frags}/G$	$\#\mathrm{atoms}/\mathrm{frag}$	$\#\mathrm{frags}/G$	$\#\mathrm{atoms}/\mathrm{frag}$
50	11.77	4.02	5.17	3.91
200	7.60	6.34	3.70	5.45
1000	5.26	9.25	2.91	6.98

Additionally, the frequency depends on the size of fragments (number of constituent atoms) in Drugs and QM9 was reported in Fig. 7. For each $|\mathcal{S}|$, the frequency distribution across fragment sizes is smooth and not biased toward certain sizes.

In the training and validation sets of the Drugs and QM9 datasets, there are 5 different ground truth conformers for each molecule. Thus, we generated 5 different conformers from RDKit to compute the fragment coordinates $\hat{\mathbf{x}}^{\mathrm{f}}$ for each molecule in train and validation sets. Following [23], we computed the optimal matching between 5 RDKit generated conformers and 5 different ground truth conformers



(a) Drugs (b) QM9

Figure 7: Frequency depends on the size of fragments in the fragment vocabulary of Drugs and QM9.

for a single molecule. After computing the cost matrix whose the (i, j)-th element means RMSD between the i-th RDKit generated conformers and the j-th ground truth conformer, we assigned optimal RDKit conformer to each ground truth conformer using linear sum assignment problem [6]. After finding the optimal matching, we aligned each ground truth conformer to its assigned RDKit conformer using the Kabsch algorithm [25]. The aligned ground truth conformers were then used in the blurring schedule (Eq. (7)) and loss function (Eq. (14)) of the training process.

C.2 Training and time

We used a single NVIDIA A100 GPU for every training and generation tasks. For training, we used a learning rate 10^{-4} with the AdamW optimizer [33]. The training time for both Drugs and QM9 was required around 3.8 days. For sampling, Drugs required 145 minutes for 14,324 conformers in 200 molecules, and QM9 required 71 minutes for 22,408 conformers in 200 molecules. We reported hyperparameters of EBD training including the maximum time step T, number of layers ($\# \, l$) and number of features ($\# \, d$) in the deblurring networks, number of multi-hops ($\#$ of hops) and cutoff value for the expansion of atom interactions, batch size, and number of iteration in Table 7.

Table 7: Hyperparameters of EBD.

Dataset	T	$\# \, l$	$\# \, d$	$\#$ of hops	cutoff	batch size	training iter.
Drugs	50	6	128	3	10 Å	32	650k
QM9	50	6	128	3	10 Å	64	650k

C.3 Performance of compared methods

For the results of compared methods in geometric evaluation of Drugs (Table 3) and QM9 (Table 8), COV-R and MAT-R scores of CVGAE [34], GraphDG [47], CGCF [57], and ConfGF [46] were borrowed from [46]. The performance of GeoMol and ConfVAE were borrowed from [62] and [59], respectively. In a case of RDKit [31], we reported the performance from the generated conformers from RDKit that we utilized to compute the approximate fragment coordinates $\hat{\mathbf{x}}^{\mathrm{f}} \sim p_{\mathrm{RDKit}}(\mathbf{x}^{\mathrm{f}})$. For GeoDiff [59], we downloaded their implementation code from `https://github.com/MinkaiXu/GeoDiff/tree/main` and trained GeoDiff model for our experiments. We reported the performance of GeoDiff after sampling conformers using Langevin dynamics [51], as they did in their implementation.

For the method in the ablation study, DecompDiff [13] is a denoising diffusion model conditioned on coarse-grained structures, where the number of prior distributions corresponds to the number of fragments, and the mean of each prior is the respective fragment coordinates. By comparing the proposed method with DecompDiff in a controlled manner, we aimed to isolate the effect of the proposed blurring scheduler and random noise injection on learning in the coarse-to-fine molecular conformer generation task. Our use of DecompDiff was not to demonstrate its suitability for the molecular conformer generation task but rather to show that the stochastic trajectory from random

https://doi.org/10.52202/079017-4185

noise corruption is more challenging for the coarse-to-fine generation task than the proposed blurring schedule, even when the prior distributions are conditioned on the coarse-grained structures.

C.4 Pseudo-code

In this subsection, we provide the Pytorch-style [37] pseudo-codes. The RDKit conformer generator to obtain the approximate fragment structure, linear interpolation blurring schedule, training process, and sampling process were given in Pseudo-codes 1, 2, 3, and 4, respectively.

```python
import torch
import copy
from rdkit.Chem import AllChem

def get_multiple_rdkit_coords(molecule, num_conf):
    mol = copy.deepcopy(molecule)
    mol.RemoveAllConformers()
    ps = AllChem.ETDG()
    ps.maxIterations = 5000
    ps.randomSeed = 2023
    ps.useBasicKnowledge = False
    ps.useExpTorsionAnglePrefs = False
    ps.useRandomCoords = False
    ids = AllChem.EmbedMultipleConfs(mol, num_conf, ps)
    if -1 in ids or mol.GetNumConformers() != num_conf:
        print("Use DG random coords.")
        ps.useRandomCoords = True
        ids = AllChem.EmbedMultipleConfs(mol, num_conf, ps)
    confs = []
    for cid in range(num_conf):
        confs.append(torch.tensor(mol.GetConformer(cid).GetPositions()
    )

    return confs
```

Pseudo-code 1: Initial atom coordinate generation from RDKit.

```python
import torch

def blurring(t, x_a_gt, x_f_rdkit, mapping_matrix):
    # prior distribution
    x_f_rdkit_extend = mapping_matrix @ x_f_rdkit

    # move positions to zero center-of-mass subspace
    x_a_gt = remove_mean(x_a_gt)
    x_f_rdkit_extend = remove_mean(x_f_rdkit_extend)

    # linear interpolation
    blurred_pos = torch.lerp(x_a_gt, x_f_ref_ext_split, t)

    return blurred_pos
```

Pseudo-code 2: Blurring schedule in Eq. 7.

```python
import torch

def loss(x_a_gt, x_f_rdkit, mapping_matrix, sigma, T):
    # sample time
    t = torch.randint(1, T, (1,)) / T

    # blurred atom position from blurring schedule
    blurred_pos = blurring(t, x_a_gt, x_f_rdkit, mapping_matrix)

    # add noise
    noise = torch.randn_like(blurred_pos)
    noise = remove_mean(noise)
```

```
    blurred_pos = blurred_pos + noise * sigma

    # estimate ground truth state from blurred atom position
    x_a_gt_estimated = deblur_network(blurred_pos, mapping_matrix, t)

    # translate to the zero center-of-mass subspace
    x_a_gt = remove_mean(x_a_gt)
    x_a_gt_est = remove_mean(x_a_gt_estimated)

    # optimal rotation matrix from Kabsch algorithm
    rot_matrix = Kabsch_alignment(x_a_gt_est, x_a_gt)

    # mean squared error
    loss = mean((x_a_gt - rot_matrix @ x_a_gt_est) ** 2)

    return loss
```

Pseudo-code 3: Training process in Algorithm 1.

```
import torch
import copy

def sample(x_f_rdkit, mapping_matrix, delta, T):
    # initial atom position located at fragment position
    x_a_init = mapping_matrix @ x_f_rdkit
    x_a_init = remove_mean(x_a_init)
    x_a = copy.deepcopy(x_a_init)

    for i in range(T-1, 0, -1):
        t = i/T

        # add noise
        noise = torch.randn_like(x_a)
        noise = remove_mean(noise)
        x_a = x_a + noise * delta

        # estimate ground truth state from blurred atom position
        x_a_gt_est = deblur_network(x_a, mapping_matrix, t)

        # translate to the zero center-of-mass subspace
        x_a_gt_est = remove_mean(x_a_gt_est)

        # optimal rotation matrix from Kabsch algorithm
        rot_matrix = Kabsch_alignment(x_a_gt_est, x_a_init)

        # next step from estimated ground truth and initial positions
        x_a_gt_est = rot_matrix @ x_a_gt_est
        x_a = blurring((i-1)/T, x_a_gt_est, x_a_init, mapping_matrx)

    return x_a
```

Pseudo-code 4: Sampling process in Algorithm 2.

D Further results on geometric evaluation

GEOM-QM9. We compared our EBD to the baseline RDKit and machine learning models on small molecules GEOM-QM9, and the results are reported in Table 8. Compared to the most of machine learning models, EBD achieved superior performances especially on the precision score. We observed that RDKit, the distance geometry-based conformer generator, outperformed in coverage metrics for small molecules. However, as the size of molecules increases and the tasks become more challenging, RDKit suffers a significant performance drop, as shown in Table 3.

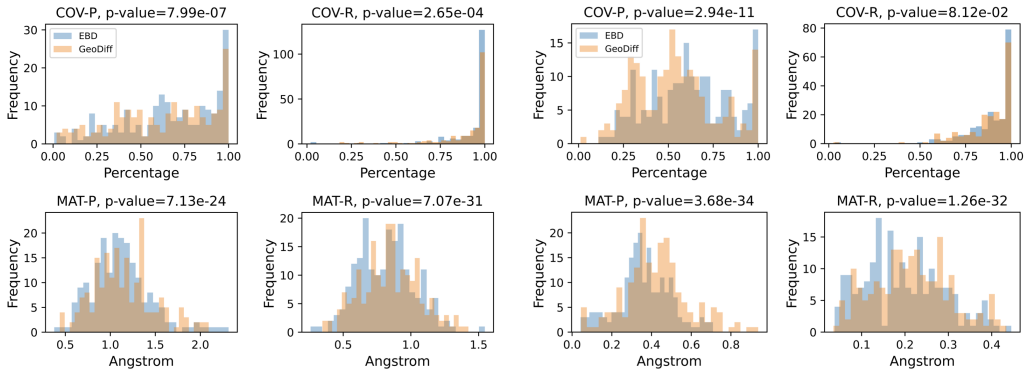
Statistical significance. We report the statistical significance of our model's improvements in geometric evaluation (COV-P, COV-R, MAT-P, and MAT-R scores). We measured p-value from

 https://doi.org/10.52202/079017-4185

Table 8: Geometric evaluation on GEOM-QM9 benchmark ($\delta = 0.5$Å).

Models	COV-R (%) ↑ Mean	COV-R (%) ↑ Med	MAT-R(Å) ↓ Mean	MAT-R(Å) ↓ Med	COV-P (%) ↑ Mean	COV-P (%) ↑ Med	MAT-P (Å) ↓ Mean	MAT-P (Å) ↓ Med
RDKit	88.34	**95.08**	0.3544	0.2974	**83.42**	**88.17**	0.3747	0.3692
CVGAE	0.09	0.00	1.6713	1.6088	-	-	-	-
GraphDG	73.33	84.21	0.4245	0.3973	43.90	35.33	0.5809	0.5823
CGCF	78.05	82.48	0.4219	0.3900	36.49	33.57	0.6615	0.6427
ConfVAE	77.84	88.20	0.4154	0.3739	38.02	34.67	0.6215	0.6091
GeoMol	71.26	72.00	0.3731	0.3731	-	-	-	-
ConfGF	88.49	94.31	0.2673	0.2685	46.43	43.41	0.5224	0.5124
GeoDiff ($T = 5000$)	88.02	92.33	**0.2199**	0.2116	53.72	52.36	0.4362	0.4259
EBD ($T = 50$)	**89.37**	93.21	0.2374	**0.1903**	61.31	60.46	**0.3622**	**0.3517**

one-sided Wilcoxon signed-rank test (a non-parametric version of paired t-test) over those scores of EBD and GeoDiff [59] on Drugs and QM9, and the results are reported in Fig. 8. Except for the COV-R score on QM9, our EBD achieved statistically significant improvement in generating more diverse and more accurate conformers for every score on either dataset, as evidenced by the p-value.



(a) Drugs (b) QM9

Figure 8: p-value of COV-P, COV-R, MAT-P, and MAT-R on Drugs and QM9.

E Visualizations

We provide additional samples and sampling processes of EBD for the test set of Drugs in Figs. 9, 11 and the test set of QM9 in Figs. 10, 12.

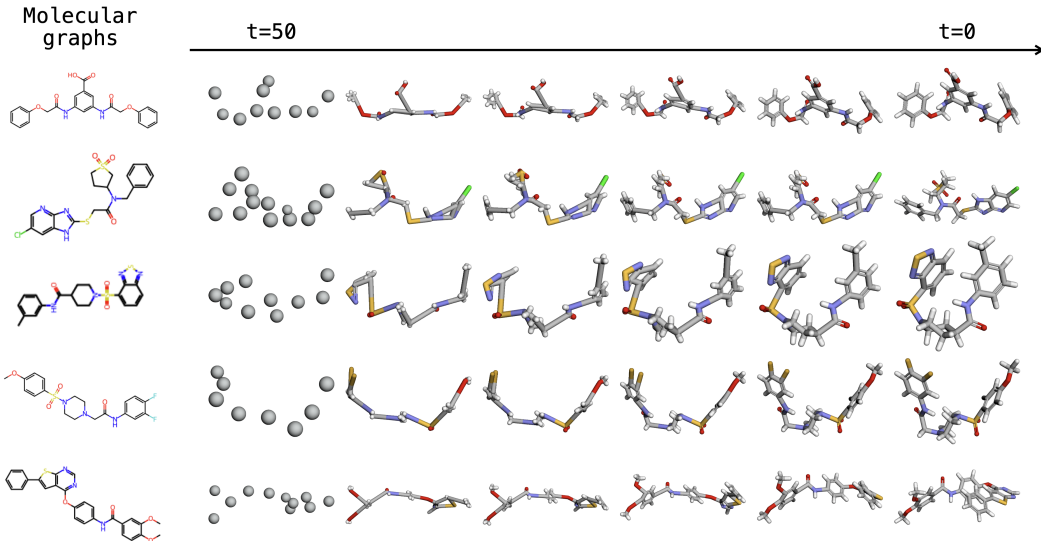


Figure 9: Sampling processes of EBD on Drugs.

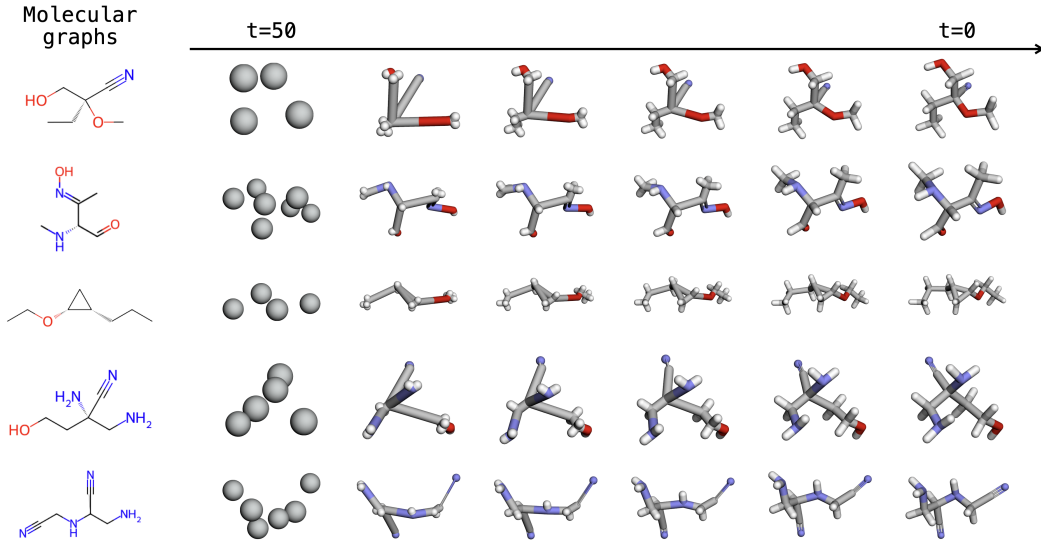


Figure 10: Sampling processes of EBD on QM9.

https://doi.org/10.52202/079017-4185

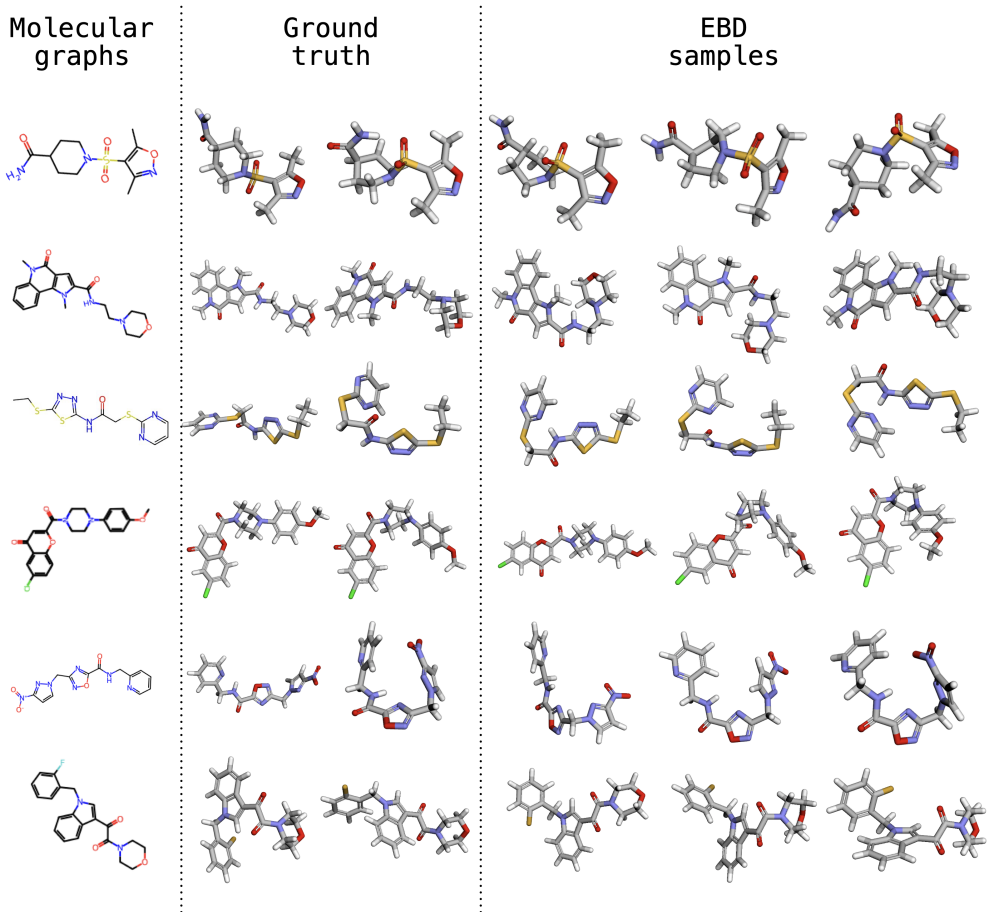


Figure 11: Visualization of molecular graphs, ground truth conformers, and samples of EBD on Drugs.

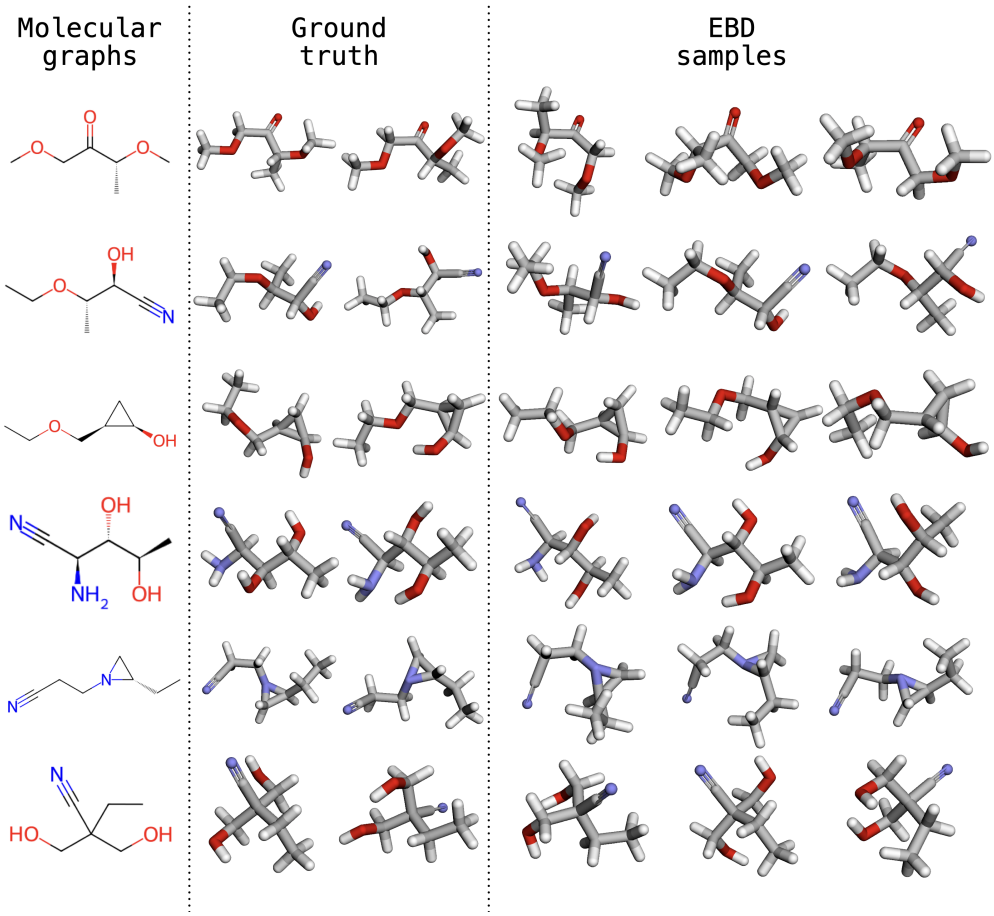


Figure 12: Visualization of molecular graphs, ground truth conformers, and samples of EBD on QM9.

https://doi.org/10.52202/079017-4185

F Limitations

Although our Equivariant Blurring Diffusion achieves significant performance on coarse-to-fine generative problems in a hierarchical molecular conformer generation scheme, there are still several limitations.

Due to the change of the estimation target from the previous state (Eq. (12)) to the ground truth state (Eq. (14)) during sampling (Algorithm 2), the next step $\mathbf{x}^{\mathrm{a}}_{t-1}$ cannot be directly computed from the current state $\mathbf{x}^{\mathrm{a}}_{t}$ and requires an additional step of the deterministic blurring function $f_{\mathbf{B}}$. This additional step in the sampling process can make the entire process slower compared to the previous state estimator.

As the size of molecule increases, the discrepancy between the ground truth $\mathbf{x}^{\mathrm{f}}$ and the approximate $\hat{\mathbf{x}}^{\mathrm{f}} \sim p_{\mathrm{RDKit}}(\mathbf{x}^{\mathrm{f}})$ of fragment structures becomes more severe. This increased discrepancy can make it more challenging for the model to learn the trajectory from the coarse fragment structures to the fine atomic details. To circumvent this issue, increasing the time step T to more than 50 can be applied. Also, for a more accurate deblurring network than equivariant graph neural networks [45] we used, more powerful geometric graph neural networks [24] can be applied such as local complete frames [10] and higher-order tensors from spherical harmonics [53].

G Broader impacts

We presented a deep generative model for the coarse-to-fine generation of molecular conformers. Our proposed model can be applied to problems in fragment-based drug discovery, such as scaffold hopping and linker generation, to achieve improved performance. In drug discovery applications, potential negative societal impacts may arise if the training set is contaminated and includes toxins. In such cases, the generated samples could potentially be harmful to humans. From a more general perspective, the generative models to which our model belongs can be misused to create false information that appears authentic. Therefore, users must be aware of the potential risks associated with generative models before using them.

https://doi.org/10.52202/079017-4185

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: See Sections 3 and 5.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: See Appendix F.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

https://doi.org/10.52202/079017-4185

Justification: No theoretical assumption was made.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: See Section 5.1 and Appendix C.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: See Appendix C.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: See Appendix C.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: See Appendix D.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

https://doi.org/10.52202/079017-4185

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: See Appendix C.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in this paper conform with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: See Appendix G.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We used the pre-processed data of GEOM [1] from [46] (`https://github.com/DeepGraphLearning/ConfGF`) which was released under MIT license. For experiments, we used the codes of EDM [19] (`https://github.com/ehoogeboom/e3_diffusion_for_molecules`), GeoDiff [59] (`https://github.com/MinkaiXu/GeoDiff`), GeoMol [11] (`https://github.com/PattanaikL/GeoMol`), Torsional Diffusion [23] (`https://github.com/gcorso/torsional-diffusion`), and RDKit [31] (`http://www.rdkit.org/docs/index.html`). EDM, GeoDiff, GeoMol, and Torsional Diffusion were released under MIT license. RDKit is licensed under Creative Commons Attribution-ShareAlike 4.0 License.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

https://doi.org/10.52202/079017-4185

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

https://doi.org/10.52202/079017-4185