Accelerating the Generation of Molecular Conformations with Progressive Distillation of Equivariant Latent Diffusion Models

Romain Lacombe
Stanford University
`rlacombe@stanford.edu`

Neal Vaidya
NVIDIA
`nealv@nvidia.com`

Abstract

Recent advances in fast sampling methods for diffusion models have demonstrated significant potential to accelerate generation on image modalities. We apply these methods to 3-dimensional molecular conformations by building on the recently introduced GeoLDM equivariant latent diffusion model (Xu et al., 2023). We evaluate trade-offs between speed gains and quality loss, as measured by molecular conformation structural stability. We introduce Equivariant Latent Progressive Distillation, a fast sampling algorithm that preserves geometric equivariance and accelerates generation from latent diffusion models. Our experiments demonstrate **up to 7.5× gains in sampling speed** with limited degradation in molecular stability. These results suggest this accelerated sampling method has strong potential for high-throughput *in silico* molecular conformations screening in computational biochemistry, drug discovery, and life sciences applications.

1 Introduction

Generative approaches in scientific domains have the potential to model complex physical systems and processes without having to solve often intractable physical equations from first principles. Recent work on geometric latent diffusion-based generative models (GeoLDM, Xu et al. (2023)) has demonstrated state-of-the-art results in structural accuracy of 3-dimensional molecular conformations. However, diffusion models require iterative inference, which slows down generation significantly. This limits the tractability of diffusion-based generative models for high-throughput *in silico* screening of large bio-molecules, which are of critical importance to biochemistry and life sciences.

Recent advances on faster sampling of diffusion models (Salimans & Ho, 2022; Song et al., 2023; Luo et al., 2023; Shih et al., 2023; Zheng et al., 2023; Meng et al., 2023) have demonstrated significant generation speed gains for image modalities and robotics tasks, while maintaining high degrees of quality. In this project, we apply recent methods for accelerated sampling of diffusion models to the 3-dimensional molecular conformations generation task, which is of high importance for ligand docking and virtual screening in computational drug discovery. Our objective is to accelerate the generation of physically plausible molecular conformations by diffusion models.

Our core research question: what are the speed vs. quality trade-offs for accelerated sampling of molecular conformations with geometric latent diffusion models (GeoLDM), and which methods provide the most speed gain for the least quality loss?

Given the demonstrated success of diffusion acceleration techniques in speeding up diffusion models for image modalities, both in wall time and number of inference steps, we expect their application to molecular conformations generation to lead to significant speed gains, albeit at the expense of some loss of quality. We implement accelerated diffusion methods for molecular conformation generation, and evaluate these techniques against the GeoLDM baseline. We specifically report on speed/quality trade-offs for each acceleration strategy, as measured by sampling speed improvements vs. structural stability degradation for generated conformations.

2   RELATED WORK

Several recent papers have demonstrated significant potential to accelerate generation from diffusion models by parallelizing inference or learning to sample multiple (or all) diffusion steps at once.

Parallel sampling of diffusion models. Shih et al. (2023) demonstrate in recent work a 2-4x improvement in sampling speed across a range of robotics and image generation tasks, by training a diffusion model to take several steps at once in parallel. Importantly, these speed gains did not result in measurable loss of quality assessed against various quantitative generation evaluation benchmarks.

Fast sampling of diffusion models via operator learning. Zheng et al. (2023) apply operator learning to the fast sampling problem, and solve the induced (deterministic) probability flow ordinary differential equation (ODE) from score-based generative modeling through stochastic differential equations (SDE) using a neural operator. By mapping initial conditions (random sample) to the full trajectory that solves the probability flow ODE, where the last point in the trajectory is the data, this approach parallelizes sampling at each time step and allows for fast sampling from a diffusion process in a single model forward step.

Consistency models. To address the slow, iterative generation process of diffusion-based models, Song et al. (2023) and Luo et al. (2023) introduce consistency models, which are trained to directly map noise to data in a single step, either by distilling pre-trained diffusion models or through direct, end-to-end training. Their recent results surpass existing diffusion model distillation techniques and achieve state-of-the-art performance on image modality benchmarks.

3   METHODS

3.1   GEOMETRIC LATENT DIFFUSION MODELS (GEOLDM)

Our paper builds on recent work in molecular conformation generation with equivariant latent diffusion models (Xu et al., 2023; Hoogeboom et al., 2022). In the GeoLDM framework (Figure 1), an encoder learns a latent representation of atom-space molecular conformations as a point cloud in latent space. The denoising model learns to invert a diffusion process in latent space, to generate point cloud representations of plausible conformations through successive inference steps, starting from Gaussian noise. A decoder then transforms the output of this latent diffusion denoising process into molecular conformations in physical space.

Importantly, the original Progressive Distillation paper (Salimans & Ho, 2022) accelerates diffusion models in physical space. Here, we introduce and experiment with progressive distillation of the diffusion denoising process of equivariant latent diffusion models, directly in latent space. We experiment with sampling from the teacher model both deterministically (DDIM) and stochastically (DDPM), as explained below.
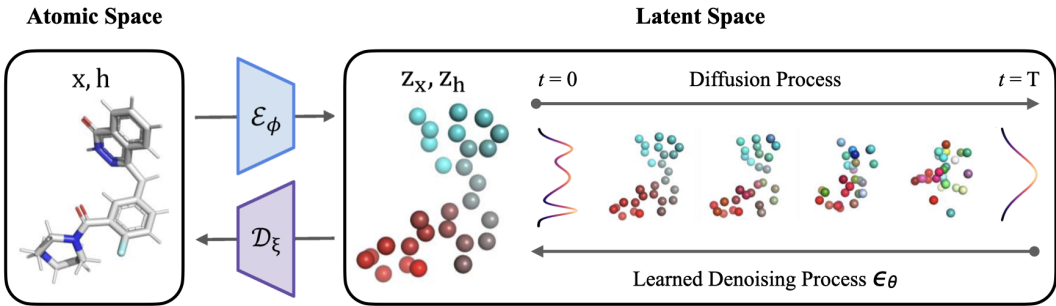

Figure 1: **Geometric latent diffusion model.** In the GeoLDM model (Xu et al., 2023), an equivariant denoising process is trained to generate point clouds in latent space by inverting a diffusion process. A decoder then transforms point clouds into 3-dimensional molecular conformations in atomic space. Image adapted from Xu et al. (2023) and Kreis et al. (2023).

3.2 Denoising Diffusion Implicit Model (DDIM)

Denoising Diffusion Implicit Models (DDIM, Song et al. (2022)) offer a computationally efficient alternative to traditional diffusion models through an implicit denoising process that allows for larger jumps in the state space, significantly reducing the number of required iterative steps.

In a standard Denoising Diffusion Probabilistic Model (DDPM, Ho et al. (2020)), the data sampling procedure is defined as a stochastic Markov process in which some z_T sampled from the prior distribution $p(\mathbf{z}_T)$ is iteratively denoised by sampling from $p(\mathbf{z}_{t-1}|\mathbf{z}_t, t)$ to produce the output $\mathbf{z}_0$. This poses challenges for acceleration and approximation, since each $\mathbf{z}_{t-1}$ relies only on $\mathbf{z}_t$ and cannot be approximated by using previous values.

We can instead define a different, non-Markovian generative sampling process:

$$\mathbf{z}_{t-1} = \sqrt{\alpha_{t-1}}\hat{\mathbf{z}}_0 + \sqrt{1 - \alpha_{t-1} - \sigma_t^2} \cdot \epsilon_\theta^{(t)}(\mathbf{z}_t) + \sigma_t \epsilon_t$$

where $\epsilon_t \sim \mathcal{N}(0, I)$, $\epsilon_\theta^{(t)}(\mathbf{z}_t)$ is a prediction of the noise from the first t steps of the forward noising process, and $\hat{\mathbf{z}}_0$ is an approximation of $\mathbf{z}_0$. Since this process no longer relies only on the value of $\mathbf{z}_t$, we can approximate $\mathbf{z}_{t-1}$ without first explicitly sampling $\mathbf{z}_t$. Song et al. (2022) showed that this alternative sampling process is equivalent to the DDPM process when:

$$\hat{\mathbf{z}}_0 = \frac{\mathbf{z}_t - \sqrt{1 - \alpha_t}\epsilon_\theta^{(t)}(\mathbf{z}_t)}{\sqrt{\alpha_t}} \qquad \sigma_t = \sqrt{\frac{(1 - \alpha_{t-1})}{1 - \alpha_t}}\sqrt{\frac{1 - \alpha_t}{\alpha_{t-1}}}$$

Setting $\sigma_t = 0$ for all t leads to a fully deterministic procedure to sample from a Denoising Diffusion *Implicit* Model (DDIM). Non-stochastic denoising opens up new avenues for accelerating sampling by approximating the deterministic trajectory, either through skipping certain sample steps $\mathbf{z}_t$ or by training a new model to match the denoising trajectory of the original model.

3.3 Equivariant Latent Progressive Distillation

In the progressive distillation approach, a trained diffusion model is distilled through an iterative student-teacher framework where the student model is trained to output in one denoising step a sample corresponding to two denoising steps of the teacher model. This distilled model can generate samples comparable to those generated by the teacher model, but with only half the required diffusion steps, a $2\times$ speed-up. The process can then be repeated with successive distillation steps.

Here we introduce the Equivariant Latent Progressive Distillation (ELPD) algorithm, which applies progressive distillation to the diffusion denoising process of an equivariant latent diffusion model directly in latent space. The Equivariant Latent Progressive Distillation algorithm is presented in Figure 2, and in algorithm 1 in appendix.
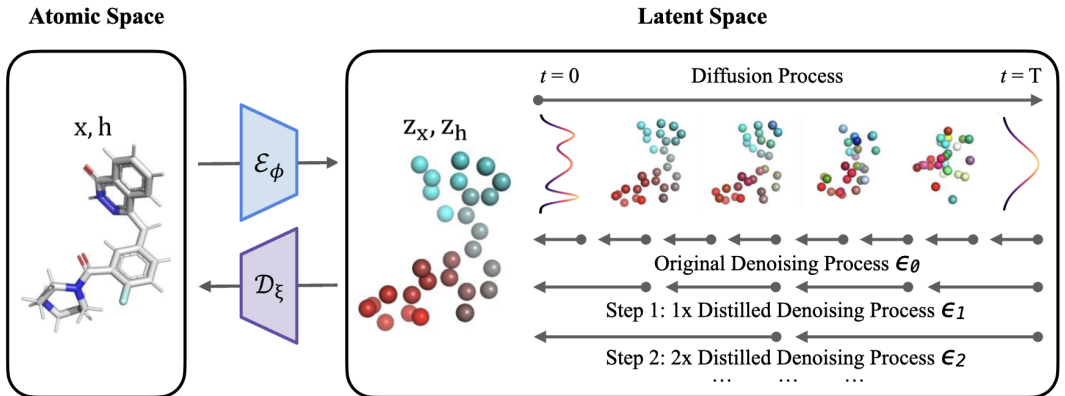

Figure 2: **Equivariant latent progressive distillation:** the denoising process in latent space is iteratively distilled by teaching a student model to sample two steps at a time from the teacher model. The algorithm initializes from the original denoising model, and progresses by successive halvings.

We implement Equivariant Latent Progressive Distillation in the GeoLDM framework, starting from the original GeoLDM model weights pre-trained on *QM9* released by Xu et al. (2023). Following the implementation in Salimans & Ho (2022), we set the weighting function $w(\lambda_t)$ to be the Truncated SNR, defined as $\max(\alpha_t^2/\sigma_t^2, 1)$. We experiment with both deterministic sampling of latent diffusion models using Denoising Diffusion Implicit Models (DDIM), and with progressive distillation of equivariant diffusion models in latent space using both deterministic DDIM sampling (DDIM-PD) and stochastic Denoising Diffusion Probabilistic Models sampling (DDPM-PD).

To our knowledge, this is the first implementation of a progressive distillation setup in latent space for equivariant diffusion generative models of molecular conformations.

4   EVALUATION

4.1   DATASET

Consistent with our baseline model GeoLDM (Xu et al., 2023), we train on the *QM9* small molecules dataset (Ramakrishnan et al., 2014), a comprehensive dataset of 134,000 stable and synthetically accessible small organic molecules, each containing up to 9 non-hydrogen atoms of carbon (C), oxygen (O), nitrogen (N), and fluorine (F) (up to 29 atoms when accounting for hydrogen H). *QM9* provides stable 3-dimensional molecular conformation structures, as well as a set of quantum chemistry features including geometric, energetic, electronic, and thermodynamic properties.

4.2   PERFORMANCE METRICS: SPEED VS. QUALITY

Speed: Sampling speed is reported both as the number of iterations at sample time (diffusion steps) and as the number of samples generated per second, averaged over the generation of 10,000 samples. Except for models with very small numbers of diffusion steps, where model overhead makes each step relatively slower, we find that wall-time generation speed and number of diffusion steps are in almost perfect correlation with one another.

Quality: Following GeoLDM (Xu et al., 2023), we evaluate the quality of our models by assessing chemical plausibility of the molecular conformations generated by sampling on the *QM9* dataset. We report the following performance metric for each evaluation run:

- **Atom-level stability:** proportion of atoms that are stable, i.e. generated with correct valence.
- **Molecule-level stability:** proportion of generated molecules where all atoms are stable.
- **Validity:** proportion of molecules measured as valid using RDKIT (Landrum, 2006).
- **Uniqueness:** the percentage of unique molecules in all generated conformations.

4.3   EXPERIMENTS

We report sample quality and generation speed for the following baseline models:

- **GeoLDM-1000** equivariant latent diffusion model (in latent space) Xu et al. (2023) pretrained with the original $T = 1000$ diffusion steps;
- **GeoLDM-100** trained with a reduced number of diffusion steps $T = 100$.

We then compare performance metrics for these baselines with the following experiments:

- **Denoising diffusion implicit model with reduced diffusion steps:** we sample from GeoLDM as a DDIM model with $T = \{1000, 500, 250, 125, 63, 32, 16, 8, 4, 2, 1\}$ steps.
- **Latent equivariant progressive distillation with deterministic sampling:** we train successive student models through progressive distillation, with $T = \{1000, 500, 250, 125, 63, 32, 16\}$ diffusion steps, and use deterministic sampling (DDIM) at train time.
- **Latent equivariant progressive distillation with stochastic sampling:** we train successive student models through progressive distillation, with $T = \{1000, 500, 250, 125, 63, 32, 16\}$ diffusion steps, and use stochastic sampling (DDPM) at train time.

We report results for baselines and all experiments in Table 1 and Figure 3. We also report examples of visualizations of conformations generated by our models in appendix (Figure 4).

Model	Sampling Method	Steps	Speed (sec^{-1})	Mol Sta %	Valid %	Valid & Unique %
GeoLDM	DDPM	1000	3.70	89.4	93.8	92.7
		100	33.30	55.8	70.6	79.7
GeoLDM	DDIM	1000	3.59	76.3	87	86.1
		125	28.30	74.6	85.3	84.1
		16	196.69	31.4	53.0	52.7
GeoLDM-PD	DDPM	125	28.28	88.4	93.3	91.6
		16	196.51	51.0	73.2	72.3
GeoLDM-PD	DDIM	125	28.28	81.6	91.7	83.6
		16	196.51	50.4	73.4	72.6

Table 1: **Comparison of base model quality and timing based on number of diffusion steps.** Speed is reported as the average number of molecules generated per second. Equivariant latent progressive distillation enables $\sim 7.5\times$ speed-up in sample time with minimal loss in sample quality.
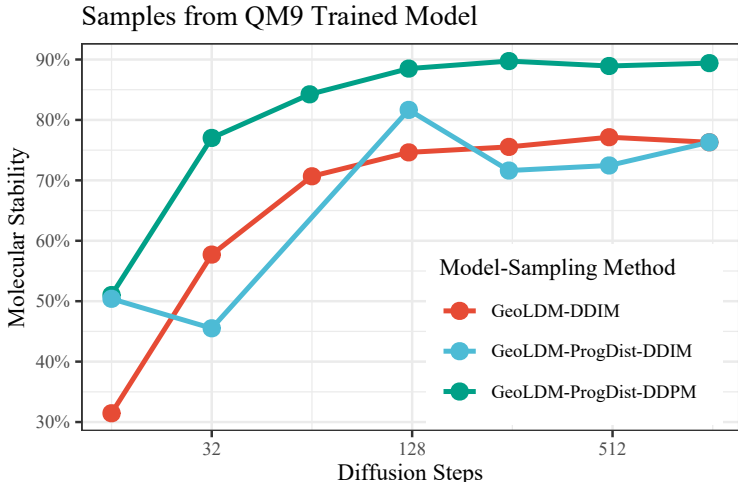

Figure 3: **Molecular stability of generated samples for various diffusion steps.** GeoLDM-DDIM is the original model sampled for a varying number of DDIM steps. GeoLDM-ProgDist-DDIM and GeoLDM-ProgDist-DDPM are sequences of progressively distilled models, using DDIM and DDPM samplers at train time. Note: x axis is logarithmic; distillation progresses from right to left.

5 ANALYSIS

5.1 ACCELERATED VS BASELINE MODELS

Compared to the baseline, all distilled models and reduced-step models achieve significant speed-ups in terms of wall-clock time to generate a sample molecule. There are no major differences in speed between the different methods after accounting for the number of diffusion steps. The GeoLDM model that was trained for 100 diffusion steps improves speed but suffers a large decrease in sample quality, with a 33.6% drop in molecular stability. The models that were trained with 1000 steps and subsequently accelerated all exhibit significantly better quality at similar speed.

5.2 COMPARING ACCELERATION METHODS

The deterministic DDIM sampler performs significantly worse than the DDPM sampler even at an equivalent number of denoising steps. There is a small drop in quality when sampling with 125 steps, and a much sharper drop with fewer steps. Because of this immediate drop in quality, DDIM sampling is likely not a viable option for accelerating GeoLDM models. Progressive distillation with deterministic sampling, as proposed in the initial formulation, performs similarly to direct DDIM sampling on the teacher model except at very low numbers of steps.

Despite the fact that the progressive distillation process trains the student model to match DDIM trajectories with the teacher model, we see that performing stochastic DDPM sampling to train the student models offers better sample quality than DDIM sampling until reaching 16 steps (6 halvings). This is consistent with the results in Salimans & Ho (2022), who show that for larger numbers of steps, stochastic sampling works better than DDIM sampling. Specifically, equivariant latent progressive distillation (ELPD) with DDPM sampling leads to molecular stability scores higher than regular DDIM, and than ELPD with DDIM, by and average of **+24.0%** and **+26.4%**, respectively.

As a conclusion, we find that **progressive distillation to 125 steps (3 halvings) with stochastic sampling (DDPM) offers the best overall trade-off between speed gains and quality loss, with a roughly 7.5× speed-up and only a 1 point drop in molecular stability**. Further speed gains can be achieved by performing further distillations, at a trade-off with decreased molecular stability.

5.3 KNOWN LIMITATIONS

Compute constraints and the need to sequentially train models for each step of distillation lead us to report results for a single sequence successive distillation training run for each setup, and we report evaluation results for a single distilled model at each step (hence, no error bars). Of note: we encountered numerical instabilities when evaluating models at further stages of training, leading to unstable generations such as with the 63-step DDIM GeoLDM-ProgDist model (see Table 2).

Lastly, stability as measured by the valence of atoms is an imperfect measurement of generation quality, as it doesn't take into account conformation energy, likelihood, steric effects, or any other consideration than valence. As a result, some generations show implausible molecular conformations, especially at low numbers of sampling steps (e.g. images (e) and (f) sampled from the 16-step DDPM model after 6 halvings). Further work could incorporate conformation energy estimations to realistically sample the Boltzmann distribution with higher-likelihood generations.

6 CONCLUSION

In conclusion, our work applies recent advances in fast diffusion model sampling to accelerate 3-dimensional molecular conformations generation by building on the recently introduced GeoLDM equivariant geometric diffusion model (Xu et al., 2023).

Our primary contribution lies in the introduction and evaluation of Equivariant Latent Progressive Distillation, a diffusion model acceleration technique in latent space which enforces equivariance by performing progressive distillation of the equivariant latent diffusion process of an original GeoLDM pre-trained teacher model via successive halving of diffusion steps. We implement and evaluate speed and quality of latent diffusion models trained through ELPD with both stochastic (DDPM) and deterministic (DDIM) diffusion steps sampling at train and inference time.

Our findings indicate that stochastic sampling (DDPM) of equivariant latent diffusion models trained through progressive distillation maintain molecular structure quality significantly better than other fast sampling methods at comparable sampling speed-ups. **These results suggest that the Equivariant Latent Progressive Distillation method has strong potential for high-throughput *in silico* molecular conformations screening.**

Overall, our research provides a deeper understanding of the trade-offs between sampling speed gains and quality losses in fast sampling of generative models for molecular conformations. The insights and methods presented in this paper pave the way for faster molecular structures generation of large bio-molecules, which are essential to computational biochemistry, drug discovery, and life sciences.

CODE AND DATA

Code and data for our experiments are available on the following GitHub repository:

`https://github.com/rlacombe/AccGeoLDM`.

FUTURE WORK

In future work, these results could be extended to larger molecules by experimenting with similar models and training setups on the *GEOM-DRUG* (Geometric Ensemble Of Molecules) dataset introduced by Axelrod & Gómez-Bombarelli (2022), which features larger organic compounds.

Also, the GeoLDM implementation uses a 'light' VAE where only the decoder is trained in practice. Further experiments could explore whether richer variational representations of the input molecules in physical space could lead to higher structural stability for molecules sampled from latent diffusion.

Additionally, a GeoLDM model could be trained with a continuous-time denoising diffusion model (Karras et al., 2022) in latent space, or other continuous-time accelerated sampling methods such as consistency models (Song et al., 2023).

Lastly, stability of molecules as measured by the valence of generated atoms is an imperfect measurement of the quality of conformations. Further work could incorporate conformation energy estimations to realistically sample the Boltzmann distribution with higher-likelihood generations.

REFERENCES

Simon Axelrod and Rafael Gómez-Bombarelli. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9:185, 2022. doi: 10.1038/s41597-022-01288-4. URL `https://doi.org/10.1038/s41597-022-01288-4`.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.

Emiel Hoogeboom, Victor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d, 2022.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models, 2022.

Karsten Kreis, Ruiqi Gao, and Arash Vahdat. Latent Diffusion Models: Is the Generative AI Revolution Happening in Latent Space? Workshop at the Conference on Neural Information Processing Systems, 2023. URL `https://neurips2023-ldm-tutorial.github.io/`.

Greg Landrum. *RDKit: Open-Source Cheminformatics*, 2006. URL `https://www.rdkit.org/`.

Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent consistency models: Synthesizing high-resolution images with few-step inference, 2023.

Chenlin Meng, Robin Rombach, Ruiqi Gao, Diederik P. Kingma, Stefano Ermon, Jonathan Ho, and Tim Salimans. On distillation of guided diffusion models, 2023.

Raghunathan Ramakrishnan, Pavlo Dral, Matthias Rupp, and Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1:140022, 2014. doi: 10.1038/sdata.2014.22. URL `https://doi.org/10.1038/sdata.2014.22`.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models, 2022.

Andy Shih, Suneel Belkhale, Stefano Ermon, Dorsa Sadigh, and Nima Anari. Parallel sampling of diffusion models, 2023.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2022.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models, 2023.

Minkai Xu, Alexander Powers, Ron Dror, Stefano Ermon, and Jure Leskovec. Geometric Latent Diffusion Models for 3D Molecule Generation, 2023.

Hongkai Zheng, Weili Nie, Arash Vahdat, Kamyar Azizzadenesheli, and Anima Anandkumar. Fast sampling of diffusion models via operator learning, 2023.

A Appendix: Detailed Results

We report the detailed results of our experiences with fast sampling methods in Table 2.

We also report a selection of example visualizations of conformations generated by various models in Figure 4 (next page).

Model	Sample Method	Steps	Speed (sec^{-1})	Mol Sta (%)	Atom Sta (%)	Valid (%)	Valid & Unique (%)
EDM	DDPM	1000	3.70	82.00	98.70	91.90	90.70
GeoLDM	DDPM	1000	3.70	89.40	98.90	93.80	92.70
		100	33.33	55.80	95.40	70.60	79.70
GeoLDM	DDIM	1000	3.59	76.30	97.00	87.00	86.06
		500	7.14	77.14	97.13	87.10	85.92
		250	14.16	75.54	96.90	85.91	84.96
		125	28.40	74.63	96.66	85.37	84.16
		63	54.3	70.67	95.95	82.47	81.48
		32	105.06	57.71	93.43	74.91	74.09
		16	196.69	31.46	85.83	53.05	52.75
		8	349.04	3.67	66.33	20.36	19.96
		4	569.16	0.00	31.87	13.31	2.96
		2	833.26	0.00	1.29	75.38	0.14
		1	1097.26	0.00	0.14	97.93	0.06
GeoLDM-ProgDist	DDPM	1000	3.70	89.40	98.90	93.80	92.70
		500	7.17	88.93	98.85	93.23	91.56
		250	14.31	89.74	98.91	93.53	92.01
		125	28.28	88.50	98.78	93.38	91.70
		63	55.18	84.24	98.20	90.89	89.15
		32	104.77	77.02	97.24	87.23	86.19
		16	196.51	51.02	92.08	73.22	72.39
GeoLDM-ProgDist	DDIM	1000	3.59	76.30	97.00	87.00	86.06
		500	7.17	72.46	96.27	85.74	85.54
		250	14.31	71.62	96.45	82.22	81.23
		125	28.28	81.66	97.62	91.71	83.67
		63	55.18	0.00	0.00	0.00	0.00
		32	104.77	45.52	91.51	58.52	57.60
		16	196.51	50.40	92.07	73.44	72.65

Table 2: Full comparison of model quality and timing based on different numbers of diffusion steps. All metrics are reported on the QM9 dataset.
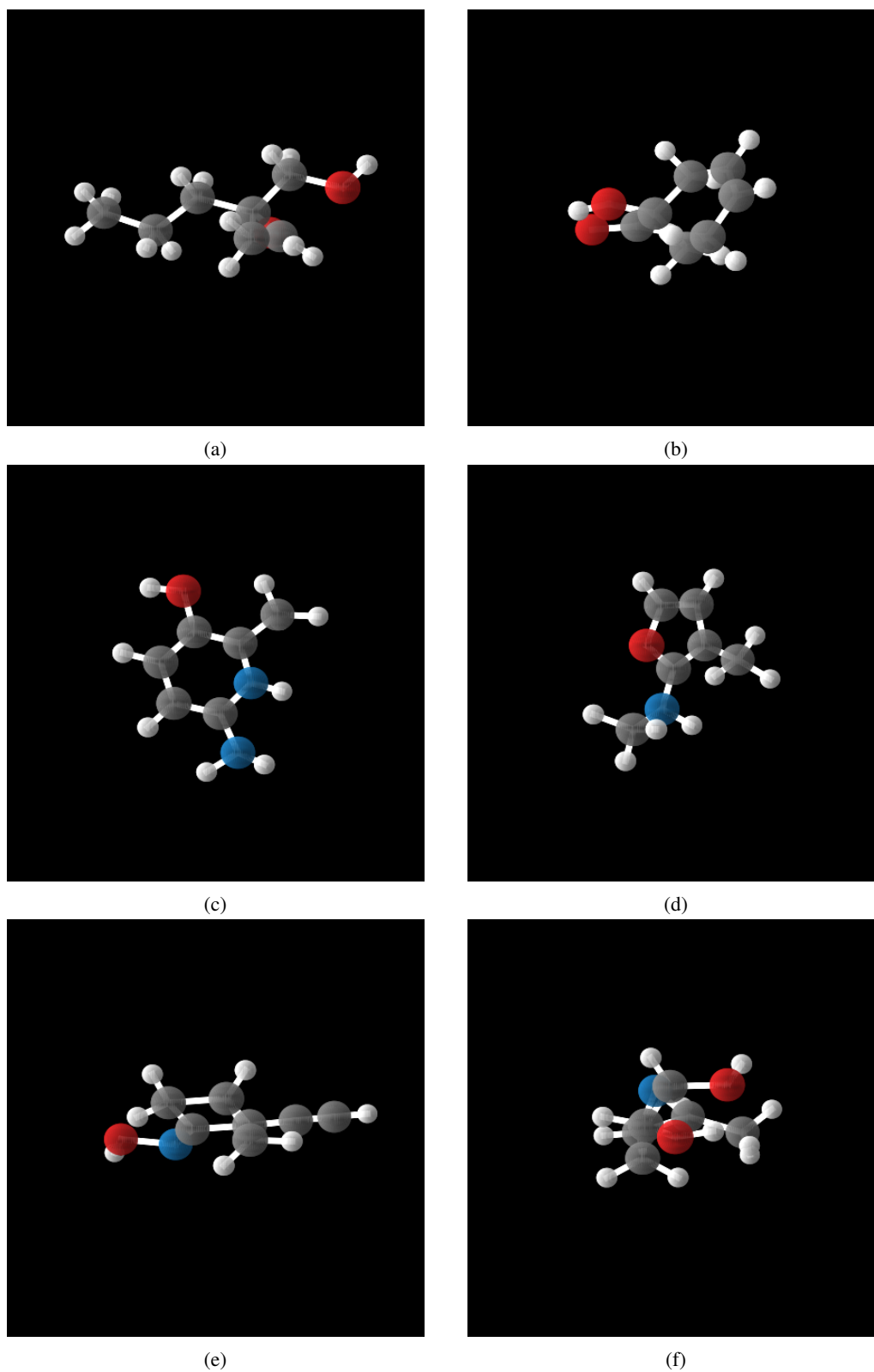
Figure 4: **Examples of conformation generations**, sampled from: (a, b) original 1000-step GeoLDM model; (c, d) 125-step model distilled with DDPM sampler after 3 halvings; (e, f) 16-step model distilled with DDPM sampler after 6 halvings.

B APPENDIX: EQUIVARIANT LATENT PROGRESSIVE DISTILLATION ALGORITHM

Here we present the detailed Equivariant Latent Progressive Distillation algorithm introduced in the Methods section and represented in Figure 2.

Algorithm 1 Equivariant Latent Progressive Distillation:

Require: Trained teacher model $\hat{z}_\eta(z_t)$
Require: Data set D
Require: Loss weight function $w()$
Require: Student sampling steps N
Require: Noise schedule $\sigma_t, \alpha_t \, \forall t \in [1, N]$

1: **for** K iterations **do**
2:      $\theta \leftarrow \eta$ $\triangleright$ Init student from teacher
3:      **while** not converged **do**
4:          $x, h \sim D$
5:          **Without gradient computation:** $\triangleright$ VAE and teacher are fixed
6:              $t = i/N, i \sim \text{Cat}[1, 2, \ldots, N]$
7:              $z_{xh} = \mathcal{E}_\phi(x, h)$ $\triangleright$ Encode x, h in latent space
8:              $\epsilon \sim \mathcal{N}(0, I)$
9:              $z_t = \alpha_t z_{xh} + \sigma_t \epsilon$
10:             $t' = t - 0.5/N, t'' = t - 1/N$ $\triangleright$ Take 2 sampling steps with Teacher
11:             $z_{t'} = \alpha_{t'} \hat{z}_\eta(z_t) + \sigma_{t'} \frac{\sigma_{t'}}{\sigma_t}(z_t - \alpha_t \hat{z}_\eta(z_t))$
12:             $z_{t''} = \alpha_{t''} \hat{z}_\eta(z_{t'}) + \sigma_{t''} \frac{\sigma_{t''}}{\sigma_{t'}}(z_{t'} - \alpha_{t'} \hat{z}_\eta(z_{t'}))$
13:             $\tilde{z} = \frac{z_{t''} - (\sigma_{t''}/\sigma_t) z_t}{\alpha_{t''} - (\sigma_{t''}/\sigma_t)\alpha_t}$ $\triangleright$ Teacher $\tilde{z}$ latent target
14:             $\lambda_t = \log[\alpha_t^2/\sigma_t^2]$
15:         Compute $\hat{z}_\theta(z_t)$ $\triangleright$ Student $\hat{z}_\theta$ latent prediction
16:         $L_\theta = w(\lambda_t) \|\tilde{z} - \hat{z}_\theta(z_t)\|^2$
17:         $\theta \leftarrow \theta - \gamma \nabla_\theta L_\theta$ $\triangleright$ Gradient descent step to update student
18:     **end while**
19:     $\eta \leftarrow \theta$ $\triangleright$ Student becomes next teacher
20:     $N \leftarrow \lceil N/2 \rceil$ $\triangleright$ Halve number of sampling steps
21: **end for**