Task Units	α Bohr^3	$\Delta\varepsilon$ meV	$\varepsilon_{\text{HOMO}}$ meV	$\varepsilon_{\text{LUMO}}$ meV	μ D	C_v $\frac{\text{cal}}{\text{mol}}$K
Naive (Upper-bound)	9.01	1470	645	1457	1.616	6.857
# Atom	3.86	866	426	813	1.053	1.971
EDM	2.76	655	356	584	1.111	1.101
GCDM	<u>1.97</u>	602	344	<u>479</u>	<u>0.844</u>	<u>0.689</u>
GeoLDM	2.37	<u>587</u>	<u>340</u>	522	1.108	1.025
GFMDiff	**1.74**	**558**	**321**	**430**	**0.728**	**0.593**
QM9 (Lower-bound)	0.10	64	39	36	0.043	0.040

Table 2: Performance comparison for conditioned molecule generation on QM9. With conditioned samples, Results are in the form of mean absolute error (MAE) for property prediction of 10000 conditional samples by an EGNN classifier.

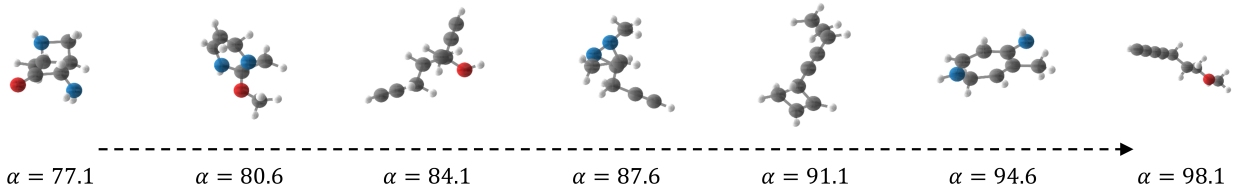

Figure 5: Generated samples of GFMDiff on QM9 conditioned with increasing values of α

Setup

In order to make comprehensive comparisons, we conduct experiments on two benchmark datasets in molecule generation: GEOM-QM9 (**?**) and GEOM-Drugs (**?**). GEOM-QM9 dataset consists of over 130K molecules and their corresponding conformations, where molecules have 18 atoms with hydrogen included on average. GEOM-Drugs includes over 450K molecules and 37M conformations, where size of molecule is 44 on average.

To assess the performance of GFMDiff in a fair and comprehensive manner, we compare it against six representative baselines in this fiels, which are E-NF (**?**), G-SchNet (**?**), EDM (**?**), models of Wu et al. (**?**), GCDM (**?**), and GeoLDM (**?**). We refer to the performances of the first three models stated in EDM, as well as results of the remaining baselines reported in GCDM and GeoLDM.

In terms of evaluation metrics, we adopt the same ones used in previous research, which are stability, validity, and uniqueness. Stability measures the proportion of atoms with correct valencies and molecules whose atoms are all stable. Validity is defined as the percentage of molecules that are theoretically correct and uniqueness shows the probability of non-repetitive samples. Arrows in Table **??** and Table **??** signify the preferred direction of each criteria. The best results are highlightened in bold and the second best results are underlined.

De Novo Molecule Generation on QM9

In order to analyze results fairly, we use the same dataset settings as previous methods. To evaluate the effectiveness of GFLoss and triplet geometric information learning, we includes GFMDiff w/o GFLoss and GFM w/o tri for comparison. In GFM w/o tri, we replace the pair-triplet track multi-head attention module with a self-attention module of pair-wise features. The weight for GFLoss λ is set 0.01. On QM9, GFMDiff is trained for around 1000 epochs, with a five layer DTN and the embedding size of 256.

As it is shown in Table **??**, GFMDiff outperforms all baselines and achieves the best performance in stability, validity, and uniqueness times validity. GFMDiff and recent SOTA methods show no major difference in stability of atoms, but the performance lead of GFMDiff over the second-best method using the same generative methods in terms of stability of molecules is 2.1%. This indicates that our model is capable of genrating stable molecules. We believe that the molecule stabilty could be further improved using latent diffusion in GeoLDM. The performance lead of GFMDiff over the SOTA method in validity and validity times uniqueness is 1.1% and 1.3%, respectively. The superior performance in validity means that GFMDiff generates molecules not only with accurate conformations, but also with correct valid and unique structres. It is intriguing to find out that GFMDiff exhibits lower performance in terms of the negative log-likelihood of data (NLL) compared to GCDM, but still surpasses other baselines. A possible explanation could be the different ways of applying geomteric information between GFMDiff and GCDM.

Moreover, the ablations of GFLoss and triplet-wise geomtery illustrate the effectiveness of them. Among GFMDiff and its abalation models, GFMDiff w/o tri achieves the lowest results. This means the incorporation of complete local geometry information contributes more to the performance lift than GFLoss. In summary, GFMDiff exhibits the ability to generate stable molecules while addressing validities of samplessimultaneously.

Type	Method	Atom Stable (%) ↑	Mol Stable (%) ↑
Normalizing flow	E-NF	75.0	0
DDPM	EDM	81.3	0.0
	Bridge+Force	82.4±0.8	0.0
	GCDM	86.4±0.2	3.7±0.3
	GeoLDM	84.4	3.2
Ours	GFMDiff	**86.5**±0.2	**3.9**±0.2
Data		86.5	2.8

Table 3: Performance comparison on GEOM-Drugs. Results of 10000 generated samples are reported with standard deviations across 3 runs using different seeds.

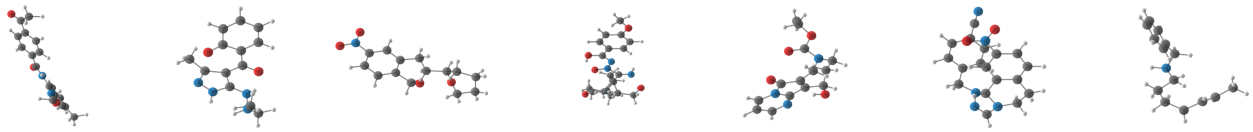

Figure 6: Molecule samples generated by GFMDiff for GEOM-Drugs

Conditional Molecule Generation on GEOM-QM9

For conditional molecule generation on QM9, we compare our GFMDiff with existing methods along with naive baselines. In Table **??**, we show the comparison of MAE on property prediction task. The "Naive (Upper-bound)" is a baseline where samples and labels are shuffled and the "#Atoms" is the property prdiction method which simply relies of the number of atoms. Lower mean absolute errors of a model than these two baselines indicate the model is capable to incorporate properties and molecule conformation information into generated samples.

As it is demonstrated in Table **??**, our methods outperforms the state-of-the-art method in this task. Samples of samples with various values of α is shown in Figure **??** as well. The performence lead of GFMDiff over the second-best method on for 6 properies are 11.7%, 4.9%, 6.2%, 10.2%, 13.7%, and 13.9%, respectively. Results indicates the superiority of our GFMDiff in generate molecules with desirable properties.

De Novo Molecule Generation on GEOM-Drugs

It is a challenging task to generate molecules for GEOM-Drugs dataset, since it is a large scale dataset of big molecules with up to 181 atoms. The relatively large scale of molecules and low stabilities of ground truth data bring huge challenges to 3D molecule generation. In experiments on GEOM-Drugs, we compares GFMDiff with E-NF, EDM, Bridge + Force, GCDM, and GeoLDM. Since current methods performs poorly in the novelty of molecules, we only list the stability of generated samples for comparison.

Due to the size of molecules in GEOM-Drugs, the stability of ground truth data are much lower than that in QM9. The proposed GFMDiff outperforms GCDM in terms of atom stability by a small margin, while GFMDiff outperforms the second-best result on molecule stability by 5.4%. It's worth noting that GeoLDM, which generates samples with high stabilities on QM9, encounters a bottleneck in generating large molecules. Some samples generated by GFMDiff are shown in Figure **??**. Results on Drugs also demonstrate the capability of our proposed GFMDiff to generate stable molecule geometries.

Conclusion

In this paper, we propose GFMDiff, a novel molecule generation methods that fully excevates geometric information to help expressive representation learning and accurate bonds formation in molecule graphs. Unlike earlier methods that did not comprehensively model molecular geometries and heavily rely on predefined rules to generate bonds, GFMDiff makes full use of spatial information to assist on representation learning and facilitate accurate edge generation. We adopt DTN as the denoising kernel to update atom features and coordinates based on interatomic forces and multi-body interactions. The GFLoss is also implemented to actively intervene the formation of bonds during each time step at the training stage. We conduct comprehensive experiments to evaluate the effectiveness and performance edge of the proposed techniques over SOTA methods. It is shown that GFMDiff is capable to generate valid molecules with accurate conformations and correct atom valencies.

Acknowledgments

This project is supported by National Key Research and Development Program of China (2022YFB4500300), the National Natural Science Foundation of China (72273132), in part by Key Research Project of Zhejiang Lab (No. 2022PI0AC01). We also gratefully acknowledge the valuable comments from anonymous reviewers.

excepturi non exercitationem sint tenetur quibusdam repellat, magni nemo dicta, sunt voluptate officia doloremque ipsa temporibus ea a accusantium impedit sit quod, modi debitis ut ad autem laudantium, alias quaerat ad sunt consectetur sed repellendus labore.