Swallowing the Bitter Pill: Simplified Scalable Conformer Generation

Yuyang Wang [1] **Ahmed A. Elhag** [1 2] **Navdeep Jaitly** [1] **Joshua M. Susskind** [1] **Miguel Ángel Bautista** [1]

Abstract

We present a novel way to predict molecular conformers through a simple formulation that sidesteps many of the heuristics of prior works and achieves state of the art results by using the advantages of scale. By training a diffusion generative model directly on 3D atomic positions without making assumptions about the explicit structure of molecules (*e.g.* modeling torsional angles) we are able to radically simplify structure learning, and make it trivial to scale up the model sizes. This model, called Molecular Conformer Fields (MCF), works by parameterizing conformer structures as functions that map elements from a molecular graph directly to their 3D location in space. This formulation allows us to boil down the essence of structure prediction to learning a distribution over functions. Experimental results show that scaling up the model capacity leads to large gains in generalization performance *without enforcing inductive biases* like rotational equivariance. MCF represents an advance in extending diffusion models to handle complex scientific problems in a conceptually simple, scalable and effective manner.

1. Introduction

In this paper we tackle the problem of molecular conformer generation, *i.e.* predicting the diverse low-energy three-dimensional conformers of molecules. Molecular conformer generation is a fundamental problem in computational drug discovery and chemo-informatics, where understanding the intricate interactions between molecular and protein structures in 3D space is critical, affecting aspects such as charge distribution, potential energy, etc. (Batzner et al., 2022). The core challenge associated with conformer generation is the vast complexity of the 3D structure space, encompassing factors such as bond lengths and torsional angles. Despite the molecular graph dictating potential 3D conformers through specific constraints, such as bond types and spatial arrangements determined by chiral centers, the conformational space experiences exponential growth with the expansion of the graph size and the number of rotatable bonds (Axelrod & Gomez-Bombarelli, 2022). This complicates brute force and exhaustive approaches, making them virtually unfeasible for even moderately small molecules.

Systematic methods, like OMEGA (Hawkins et al., 2010), offer rapid processing through rule-based generators and curated torsion templates. Despite their efficiency, these models typically fail on complex molecules, as they often overlook global interactions and are tricky to extend to inputs like transition states or open-shell molecules. Classic stochastic methods, like molecular dynamics (MD) and Markov chain Monte Carlo (MCMC), rely on extensively exploring the energy landscape to find low-energy conformers. Such techniques suffer from sampling inefficiency for large molecules and struggle to generate diverse representative conformers (Hawkins, 2017; Wilson et al., 1991; Grebner et al., 2011). In the domain of learning-based approaches, several works have looked at conformer generation problems through the lens of probabilistic modeling, using either normalizing flows (Xu et al., 2021a) or diffusion models (Xu et al., 2022; Jing et al., 2022). These approaches tend to use equivariant network architectures to deal with molecular graphs (Xu et al., 2022) or model domain-specific factors like torsional angles (Ganea et al., 2021; Jing et al., 2022). However, explicitly enforcing these domain-specific inductive biases come at a cost. For example, Torsional Diffusion models rely on rule-based methods to find rotatable bonds which may fail especially for complex molecules. Ultimately, the quality of generated conformers is destined to suffer from errors of the non-differentiable cheminformatic methods used to predict local substructures. On the other hand, recent works have proposed domain-agnostic approaches for generative modeling of data in function space (Du et al., 2021; Dupont et al., 2022b;a; Zhuang et al., 2023) obtaining great performance. As an example, Zhuang et al. (2023) use a diffusion model to learn a distribution over functions f, showing great results on different data domains like images (*i.e.* $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$) or 3D geometry (*i.e.*

[1]Apple [2]Work was completed while A.A.E was an intern with Apple. Correspondence to: {yuyang_wang4, aa_elhag, njaitly, jsusskind, mbautistamartin}@apple.com.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.

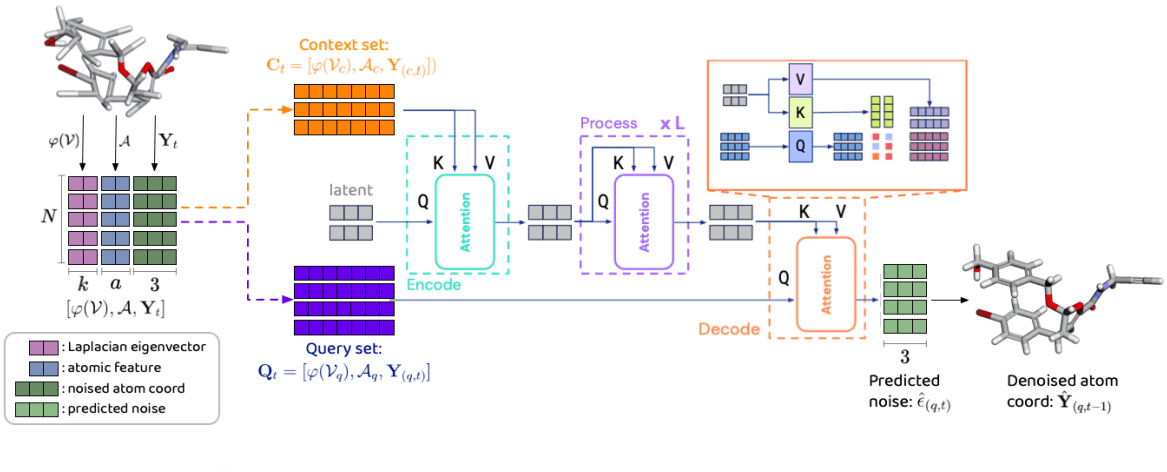

Figure 1. Overview of proposed MCF. The structure of molecular graph is encoded through eigenvectors of Laplacian eigen-decomposition $\varphi(\mathcal{V})$ and atomic features $\mathcal{A}$. *MCF directly operates on atom coordinates in 3D space* and trains the diffusion model to denoise the function in 3D coordinates. The score network is developed with attention-based PerceiverIO architecture. Context pairs $\mathbf{C}_t$ attend to a latent array of learnable parameters via cross attention and the latent array goes through several self attention blocks. Finally, the query pairs $\mathbf{Q}_t$ cross-attend to the latent array to produce the final noise prediction $\hat{\epsilon}_q$ in 3D space.

$f : \mathbb{R}^3 \to \mathbb{R}^1$), where the domain of the function $\mathbb{R}^n$ is fixed across functions. Such frameworks provide a valuable paradigm to investigate whether domain-agnostic methods with little to no inductive biases can be successfully transferred to solve scientific problems (*e.g.* molecular conformer generation).

To this end, we present Molecular Conformer Fields (MCF), a simple and scalable approach to learn generative models of molecular conformers. We leverage a domain-agnostic architecture that makes no assumptions about molecular structures and trivially benefits from scale. We formulate the molecular conformer generation problem as learning a distribution over functions/fields (we use both terms exchangeably), an approach that has been applied widely to various data domains (Zhuang et al., 2023). Specifically, conformers are interpreted as functions that map points on graph $\mathcal{G}_i$ to atom coordinates in $\mathbb{R}^3$, $f_i : \mathcal{G}_i \to \mathbb{R}^3$, which we call a *conformer field*. Unlike many prior efforts that shoe-horn inductive biases of molecular structures into the model (*e.g.* developing equivariant diffusion process, modeling torsional angles, etc.) (Xu et al., 2022; Ganea et al., 2021; Jing et al., 2022), *MCF operates directly on 3D atom coordinates*, without enforcing molecular constraints explicitly, letting the model learn these directly from the data.

Instead of using Graph Neural Networks with intricate equivariance designs, MCF builds a score network using PerceiverIO (Jaegle et al., 2022) (see Fig. 1) which is a scalable and efficient variant of the Transformer architecture. Our model is simple to implement and efficient to scale. Experiments on recent conformer generation benchmarks show MCF surpasses strong baselines by a gap that gets larger as we scale model capacity, potentially revealing a bitter lesson (Sutton, 2019) moment for conformer generation, when large models with fewer domain-specific architectural inductive biases lead to better performance. Superior performance of MCF on molecular conformation generation highlights the potential for building a singe domain-agnostic method that is simple and scalable to work on many different problems.

Our contributions are summarized as follows:

- We introduce a novel approach for molecular conformer generation that has strong scaling properties and surpasses previous methods by a large margin on standard benchmarks.
- Our approach directly predicts the 3D position of atoms as opposed to domain-specific variables, providing a simple and scalable training recipe.
- MCF shows that enforcing inductive biases like rotational equivariance or modeling torsional angles is not required for generalization.

2. Related Work

Recent works have tackled the problem of molecular conformer generation using learning-based generative models. Simm & Hernández-Lobato (2019) and Xu et al. (2021b) develop two-stage methods which first generate interatomic distances following VAE framework and then predict conformers based on the distances. Guan et al. (2021) propose neural energy minimization to optimize low-quality conformers. In Xu et al. (2021a), a normalizing flow approach

is proposed as an alternative to VAEs. To avoid the accumulative errors from two-stage generation, Shi et al. (2021) implement score-based generative model to directly model the gradient of logarithm density of atomic coordinates. In GeoDiff (Xu et al., 2022), a diffusion model is used which focuses on crafting equivariant forward and backward processes with equivariant graph neural networks. In GeoMol (Ganea et al., 2021), the authors first predict 1-hop local structures and then propose a regression objective coupled with an Optimal Transport loss to predict the torsional angles that assemble substructures of a molecule. Following this, Torsional Diffusion (Jing et al., 2022) proposed a diffusion model on the torsional angles of the bonds rather than a regression model used in Ganea et al. (2021).

Our approach extends recent efforts in generative models for functions in Euclidean space (Zhuang et al., 2023; Dupont et al., 2022b;a; Du et al., 2021), to functions defined over graphs (*e.g.* chemical structure of molecules). Different approaches have been proposed to learn distributions over fields in Euclidean space; GASP (Dupont et al., 2022b) leverages a GAN whose generator produces field data whereas a point cloud discriminator operates on discretized data and aims to differentiate real and generated functions. Two-stage approaches (Dupont et al., 2022a; Du et al., 2021) adopt a latent field parameterization (Park et al., 2019) where functions are parameterized via a hypernetwork (Ha et al., 2017) and a generative model is learnt in latent space. MCF presents a generalization over these approaches to deal with training sets where each function f_i is defined on a different graph $\mathcal{G}_i$, as opposed to in Euclidean space. In addition, MCF also related to recent work focusing on fitting a function on a manifold using an intrinsic coordinate system (Koestler et al., 2022; Grattarola & Vandergheynst, 2022), and generalizes it to the problem of learning a probabilistic model over multiple functions defined on different graphs. Intrinsic coordinate systems have also been used in Graph Transformers to tackle supervised learning tasks (Maskey et al., 2022; Sharp et al., 2022; He et al., 2022; Dwivedi et al., 2020).

Recent strides in the domain of protein folding dynamics have witnessed revolutionary progress, with modern methodologies capable of predicting crystallized 3D structures solely from amino-acid sequences using auto-regressive models like AlphaFold (Jumper et al., 2021). However, transferring these approaches seamlessly to general molecular data is fraught with challenges. Molecules present a unique set of complexities, manifested in their highly branched graphs, varying bond types, and chiral information, aspects that make the direct application of protein folding strategies to molecular data a challenging endeavor.

3. Preliminaries

3.1. Diffusion Probabilistic Fields

Diffusion Probabilistic Fields (DPF) (Zhuang et al., 2023) belongs to the broad family of latent variable models (Everett, 2013) and can be consider a generalization of DDPMs (Ho et al., 2020) to deal with functions $f : \mathbf{M} \to Y$ which are infinite dimensional. Conceptually speaking, DPF (Zhuang et al., 2023) parameterizes functions f with a set of context pairs containing input-outputs to the function. Using these context pairs as input to DPF, the model is trained to denoise any query coordinate (*e.g.* query pairs) in the domain of the function at timestep t (as shown in Fig. 1). In order to learn a parametric distribution over functions $p_\theta(f_0)$ from an empirical distribution of functions s $q(f_0)$, DPF reverses a diffusion Markov Chain that generates function latents $f_{1:T}$ by gradually adding Gaussian noise to (context) input-output pairs randomly drawn from $f \sim q(f_0)$ for T time-steps as follows: $q(f_t|f_{t-1}) := \mathcal{N}\left(f_{t-1}; \sqrt{\bar{\alpha}_t} f_0, (1-\bar{\alpha}_t)\mathbf{I}\right)$. Here, $\bar{\alpha}_t$ is the cumulative product of fixed variances α_t with a handcrafted scheduling up to time-step t. DPF (Zhuang et al., 2023) follows the training recipe in Ho et al. (2020) in which: i) The forward process adopts sampling in closed form. ii) reversing the diffusion process is equivalent to learning a sequence of denoising (or score) networks ϵ_θ, with tied weights. Reparameterizing the forward process as $f_t = \sqrt{\bar{\alpha}_t} f_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$ results in the "simple" DDPM loss: $\mathbb{E}_{t\sim[0,T], f_0\sim q(f_0), \epsilon\sim\mathcal{N}(0,\mathbf{I})}\left[\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} f_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)\|^2\right]$, which makes learning of the data distribution $p_\theta(f_0)$ both efficient and scalable. At inference time, DPF computes $f_0 \sim p_\theta(f_0)$ via ancestral sampling (Zhuang et al., 2023). Concretely, DPF starts by sampling dense query coordinates and assigning a gaussian value to them $f_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then, it iteratively applies the score network ϵ_θ to denoise f_T, thus reversing the diffusion Markov Chain to obtain f_0. In practice, DPFs have obtained amazing results for signals living in an Euclidean geometry.

3.2. Conformers as Functions on Graphs

Following the setting in previous work (Xu et al., 2022; Ganea et al., 2021; Jing et al., 2022) a molecule with n atoms is represented as an undirected graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, where $\mathcal{V} = \{v_i\}_{i=1}^n$ is the set of vertices representing atoms and $\mathcal{E} = \{e_{ij} | (i,j) \subseteq |\mathcal{V}| \times |\mathcal{V}|\}$ is the set of edges representing inter-atomic bonds. We further use $\mathcal{A}$ to denote atomic features which also are leveraged by our generative model. In this paper, we parameterize a molecule's conformer as a function $f : \mathcal{G} \to \mathbb{R}^3$ that takes atoms (*e.g.* vertices) in the molecular graph $\mathcal{G}$ and maps them to 3D space, we call this function a *conformer field*. The training set is composed of conformer fields $f_i : \mathcal{G}_i \to \mathbb{R}^3$, where each field maps atoms of a different molecule $\mathcal{G}_i$ to a 3D point. We then formulate

the task of conformer generation as learning a prior over a training set of conformer fields. We drop the subscript i in the remainder of the text for notation simplicity.

We learn a denoising diffusion generative model (Ho et al., 2020) over conformer fields f. In particular, given conformer field samples $f_0 \sim q(f_0)$ the forward process takes the form of a Markov Chain with progressively increasing Gaussian noise: $q(f_{1:T}|f_0) = \prod_{t=1}^{T} q(f_t|f_{t-1})$, $q(f_t|f_{t-1}) := \mathcal{N}(f_{t-1}; \sqrt{\bar{\alpha}_t} f_0, (1-\bar{\alpha}_t)\mathbf{I})$. We train MCF using the denoising objective function in (Ho et al., 2020): $\mathbb{E}_{t\sim[0,T], f_0\sim q(f_0), \epsilon\sim\mathcal{N}(0,\mathbf{I})}\left[\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} f_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)\|^2\right]$.

3.3. Equivariance in Conformer Generation

Equivariance has become an important topic of study in generative models (Abbott et al., 2023; 2022; Kanwar et al., 2020). In particular, enforcing equivariance as an explicit inductive bias in neural networks can lead to improved generalization (Köhler et al., 2020) by constraining the space of functions that can be represented by a model. On the other hand, recent literature shows that models that can learn these symmetries from data rather than explicitly enforcing them (*e.g.* Transformers vs CNNs) tend to perform better as they are more amenable to optimization (Bai et al., 2021).

Equivariance also plays an interesting role in conformer generation. On one hand, it is important when training likelihood models of conformers, as the likelihood of a conformer is invariant to roto-translations (Köhler et al., 2020). On the other hand, when training models to generate conformers given a molecular graph, explicitly baking roto-translation equivariance might not be as necessary. This is because the intrinsic structure of the conformer encodes far more information about its properties than the extrinsic coordinate system (eg. rotation and translation) in which the conformer is generated (Ruddigkeit et al., 2012). In addition, recent approaches for learning simulations on graphs (Sanchez-Gonzalez et al., 2020) or pre-training models for molecular prediction tasks (Zaidi et al., 2022) have successfully relied on non-equivariant architectures.

In this paper, we ask whether inductive biases like rotational equivariance can be traded for model scale in general purposes architectures like Transformers. Our empirical results show that explicitly enforcing roto-translation equivariance is not a strong requirement for generalization. Furthermore, we show that scalable approaches that do not explicitly enforce roto-translation equivariance (like ours) can outperform approaches that do by a large margin .

4. Method

MCF is a diffusion generative model that captures distributions over conformer fields. We are given observations in the form of an empirical distribution $f_0 \sim q(f_0)$ over fields where a field $f_0 : \mathcal{G} \to \mathbb{R}^3$ maps vertices $v \in \mathcal{G}$ on a molecular graph $\mathcal{G}$ to 3D space $\mathbb{R}^3$.

To tackle the problem of learning a diffusion generative model over conformer fields we extend the recipe in DPF (Zhuang et al., 2023), generalizing from fields defined in ambient Euclidean space to functions on graphs (*e.g.* conformer fields). In order to do this, we compute the k leading eigenvectors of the normalized graph Laplacian $\Delta_\mathcal{G}$ (Maskey et al., 2022; Sharp et al., 2022) as positional encoding for points in the graph. The eigen-decomposition of the normalized graph Laplacian can be computed efficiently using sparse eigen-problem solvers (Hernandez et al., 2009) and only needs to be computed once before training. We use the term $\varphi(v) = \sqrt{n}[\varphi_1(v), \varphi_2(v), \ldots, \varphi_k(v)] \in \mathbb{R}^k$ to denote the normalized Laplacian eigenvector representation of a vertex $v \in \mathcal{G}$.

We adopt an explicit field parametrization where a field is characterized by uniformly sampling a set of vertex-signal pairs $\{(\varphi(v_c), \boldsymbol{y}_{(c,0)})\}$, $v_c \in \mathcal{G}, \boldsymbol{y}_{(c,0)} \in \mathbb{R}^3$, which is denoted as *context set*. We row-wise stack the context set and refer to the resulting matrix via $\mathbf{C}_0 = [\varphi(\mathcal{V}_c),\ \mathbf{Y}_{(c,0)}]$. Here, $\varphi(\mathcal{V}_c)$ denotes the Laplacian eigenvector representation context vertices and $\mathbf{Y}_{(c,0)}$ denotes the 3D position of context vertices at time $t = 0$. We define the forward process for the context set by diffusing the 3D positions and keeping Laplacian eigenvectors fixed:

$$\mathbf{C}_t = [\varphi(\mathcal{V}_c), \mathbf{Y}_{(c,t)} = \sqrt{\bar{\alpha}_t}\mathbf{Y}_{(c,0)} + \sqrt{1-\bar{\alpha}_t}\epsilon_c], \quad (1)$$

where $\epsilon_c \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a noise vector of the appropriate size. We now turn to the task of formulating a score network for fields. The score network needs to take as input the context set (*i.e.* the field parametrization), and needs to accept being evaluated for any point in $\mathcal{G}$. We do this by sampling a *query set* of vertex-signal pairs $\{\varphi(v_q), \boldsymbol{y}_{(q,0)}\}$. Equivalently to the context set, we row-wise stack query pairs and denote the resulting matrix as $\mathbf{Q}_0 = [\varphi(\mathcal{V}_q),\ \mathbf{Y}_{(q,0)}]$. Note that the forward diffusion process is equivalently defined for both context and query sets:

$$\mathbf{Q}_t = [\varphi(\mathcal{V}_q), \mathbf{Y}_{(q,t)} = \sqrt{\bar{\alpha}_t}\mathbf{Y}_{(q,0)} + \sqrt{1-\bar{\alpha}_t}\epsilon_q], \quad (2)$$

where $\epsilon_q \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a noise vector of the appropriate size. The underlying field is solely defined by the context set, and the query set are the function evaluations to be de-noised. The resulting *score field* model is formulated as follows, $\hat{\epsilon_q} = \epsilon_\theta(\mathbf{C}_t, t, \mathbf{Q}_t)$.

Using the explicit field characterization and the score field network, we obtain the training and inference procedures

Algorithm 1 Training

$\Delta_{\mathcal{G}}\varphi_i = \varphi_i\lambda_i$ // Compute Laplacian eigenvectors
repeat
  $(\mathbf{C}_0, \mathbf{Q}_0) \sim \text{Uniform}(q(f_0))$
  $t \sim \text{Uniform}(\{1, \dots, T\})$
  $\epsilon_c \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \epsilon_q \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\mathbf{C}_t = [\varphi(\mathcal{V}_c), \sqrt{\bar{\alpha}_t}\mathbf{Y}_{(c,0)} + \sqrt{1-\bar{\alpha}_t}\epsilon_c]$
  $\mathbf{Q}_t = [\varphi(\mathcal{V}_q), \sqrt{\bar{\alpha}_t}\mathbf{Y}_{(q,0)} + \sqrt{1-\bar{\alpha}_t}\epsilon_q]$
  Take gradient descent step on
  $\nabla_\theta \left\|\epsilon_q - \epsilon_\theta(\mathbf{C}_t, t, \mathbf{Q}_t)\right\|^2$
until converged

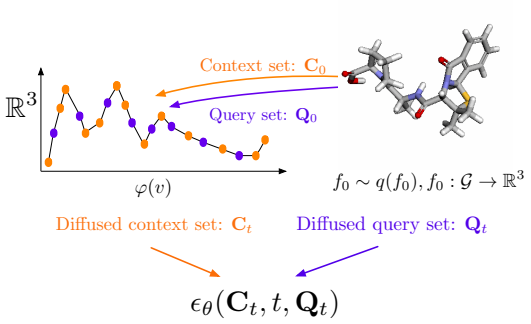

Figure 2. **Left:** MCF training algorithm. **Right**: Visual depiction of a training iteration for a conformer field. See Sect. 4 for definitions (.

Algorithm 2 Sampling

$\Delta_{\mathcal{G}}\varphi_i = \varphi_i\lambda_i$ // LBO eigen-decomposition
$\mathbf{Q}_T = [\varphi(\mathcal{V}_q), \mathbf{Y}_{(q,t)} \sim \mathcal{N}(\mathbf{0}_q, \mathbf{I}_q)]$
$\mathbf{C}_T \subseteq \mathbf{Q}_T$ {Random subset}
for $t = T, \dots, 1$ **do**
  $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\boldsymbol{z} = \mathbf{0}$
  $\mathbf{Y}_{(q,t-1)} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{Y}_{(q,t)} - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{C}_t, t, \mathbf{Q}_t)\right) + \sigma_t \boldsymbol{z}$
  $\mathbf{Q}_{t-1} = [\varphi(\mathcal{V}_q), \mathbf{Y}_{(q,t-1)}]$
  $\mathbf{C}_{t-1} \subseteq \mathbf{Q}_{t-1}$ {Same subset as in step 2}
end for
return f_0 evaluated at coordinates $\varphi(\mathcal{V}_q)$

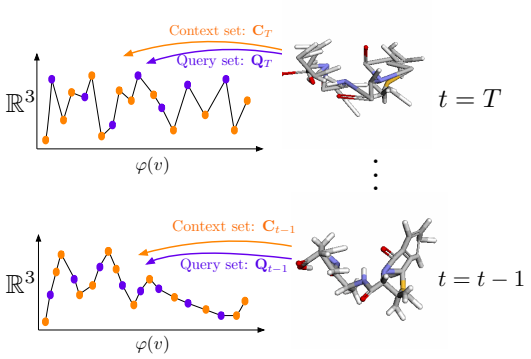

Figure 3. **Left:** MCF sampling algorithm. **Right**: Visual depiction of the sampling process of a conformer field.

in Alg. 1 and Alg. 2, respectively, which are accompanied by illustrative examples of sampling a conformer field. For training, we uniformly sample context and query sets from $f_0 \sim \text{Uniform}(q(f_0))$ and only corrupt their signal using the forward process in Eq. equation 1 and Eq. equation 2. We train the score field network ϵ_θ to denoise the signal portion of the query set, given the context set. During sampling, to generate a conformer fields $f_0 \sim p_\theta(f_0)$ we first define a query set $\mathbf{Q}_T = [\varphi(\mathcal{V}_q), \mathbf{Y}_{(q,T)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})]$ of random atom positions to be de-noised. We set the context set to be a random subset of the query set. We use the context set to denoise the query set and follow ancestral sampling as in the vanilla DDPM (Ho et al., 2020). Note that during inference the eigen-function representation $\varphi(v)$ of the context and query sets does not change, only their corresponding signal value (*e.g.* their 3D position).

4.1. Score Field Network ϵ_θ

In MCF, the score field's design space covers all architectures that can process irregularly sampled data, such as Transformers (Vaswani et al., 2017) and their corresponding Graph counterparts (Maskey et al., 2022; Sharp et al., 2022; He et al., 2022; Dwivedi et al., 2020) which have recently gained popularity in the supervised learning setting. The score field network ϵ_θ is primarily implemented using PerceiverIO (Jaegle et al., 2022), an effective Transformer encoder-decoder architecture. A PerceiverIO is chosen due to its nature of a general-purposed architecture that can handle data of a wide variety domains. It provides a suitable test bed for evaluating how well models without domain-specific inductive bias (*e.g.* equivariance) perform in solving scientific problems (*e.g.* molecular conformer generation as investigated in this work). PerceiverIO encodes interactions between elements in sets using attention, which has been demonstrated to be scalable in many previous works (Brown et al., 2020). Fig. 1 demonstrates how these sets are used within the PerceiverIO architecture. To elaborate, the encoder maps the context set into latent arrays (*i.e.* a group of learnable vectors) through a cross-attention layer, while the decoder does the same for query set. For a more detailed analysis of the PerceiverIO architecture refer to (Jaegle et al., 2022).The time-step t is incorporated into the score computation by concatenating a positional embedding representation of t to both context and query sets.

	Recall				Precision			
	COV ↑		AMR ↓		COV ↑		AMR ↓	
	mean	median	mean	median	mean	median	mean	median
CGCF	69.5	96.2	0.425	0.374	38.2	33.3	0.711	0.695
GeoDiff	76.5	**100.0**	0.297	0.229	50.0	33.5	0.524	0.510
GeoMol	91.5	**100.0**	0.225	0.193	87.6	**100.0**	0.270	0.241
Tor. Diff.	92.8	**100.0**	0.178	0.147	92.7	**100.0**	0.221	0.195
MCF	**95.0**	100.0	**0.103**	**0.044**	**93.7**	100.0	**0.119**	**0.055**

Table 1. Molecule conformer generation results on GEOM-QM9. MCF obtains better results than the state-of-the-art baselines.

	Recall				Precision			
	COV ↑		AMR ↓		COV ↑		AMR ↓	
	mean	median	mean	median	mean	median	mean	median
GeoDiff	42.1	37.8	0.835	0.809	24.9	14.5	1.136	1.090
GeoMol	44.6	41.4	0.875	0.834	43.0	36.4	0.928	0.841
Tor. Diff.	72.7	80.0	0.582	0.565	55.2	56.9	0.778	0.729
MCF-S	79.4	87.5	0.512	0.492	57.4	57.6	0.761	0.715
MCF-B	84.0	91.5	0.427	0.402	64.0	66.2	0.667	0.605
MCF-L	**84.7**	**92.2**	**0.390**	**0.247**	**66.8**	**71.3**	**0.618**	**0.530**

Table 2. Molecule conformer generation results on GEOM-DRUGS. MCF surpasses state-of-the-art baselines by large margin.

5. Experiments

We use two popular datasets: GEOM-QM9 and GEOM-DRUGS (Axelrod & Gomez-Bombarelli, 2022). Datasets are preprocessed and split as described in Ganea et al. (2021). We deploy PerceiverIO with small (S), base (B) and large (L) sizes, which contain 13M, 64M and 242M parameters respectively. More implementation details can be found in Appendix A.2. We provide additional experiments that validate the design choices for the score network architecture, as well as empirically validating the chemical properties of generated conformers in the Appendix A.3.

5.1. GEOM-QM9

Following the standard setting for molecule conformer prediction we use the GEOM-QM9 dataset which contains $\sim$ 130K molecules ranging from 3 to 29 atoms. We report our results with base size model (*i.e.* MCF-B) in Tab. 1 and compare with CGCF (Xu et al., 2021a), GeoDiff (Xu et al., 2022), GeoMol (Ganea et al., 2021) and Torsional Diff. (Jing et al., 2022). Note that all baselines make strong assumptions about the geometric structure of molecules. They either develop equivariant diffusion process (Xu et al., 2022) or model domain-specific characteristics like interatomic distances (Xu et al., 2021a) and torsional angles of rotatable bonds (Ganea et al., 2021; Jing et al., 2022). In contrast, MCF simply models the distribution of 3D coordinates of atoms without making any assumptions about the underlying structure. Finally we report the same metrics as Torsional Diff. (Jing et al., 2022) to compare the generated and ground truth conformer ensembles: average minimum RMSD (AMR) and coverage (COV). These metrics are reported both for precision, measuring the accuracy of the generated conformers, and recall, measuring how well the generated ensemble covers the ground-truth ensemble (details about metrics can be found in Appendix A.2.4). We generate $2K$ conformers for a molecule with K ground truth conformers.

Tab. 1 shows that MCF outperforms previous approaches by a substantial margin. In addition, it is important to note that MCF is a general approach for learning functions on graphs that does not make any assumptions about the intrinsic geometric factors important in conformers like torsional angles. This makes MCF simpler to implement and applicable to other settings in which intrinsic geometric factors are not known or expensive to compute.

5.2. GEOM-DRUGS

To test the capacity of MCF to deal with larger molecules we also report experiments on GEOM-DRUGS, the largest and most pharmaceutically relevant part of the GEOM dataset (Axelrod & Gomez-Bombarelli, 2022) — consisting of 304k drug-like molecules (average 44 atoms). We report our results in Tab. 2 and compare with GeoDiff (Xu et al., 2022), GeoMol (Ganea et al., 2021) and Torsional Diff. (Jing et al., 2022). Note again that all baseline approaches make strong assumptions about the geometric structure of molecules and model domain-specific characteristics like torsional angles of bonds. MCF simply models the distribution of 3D coordinates of atoms without making any assumptions about the underlying structure.

Results on Tab. 2 are where we see MCF outperforms strong baseline approaches by substantial margins. All MCF models achieve better performance than previous state-of-the-art Torsional Diff. model. On both recall and precision, MCF of small size (MCF-S) surpasses Torsional Diff. by approximately 15%. This indicates that our proposed MCF not only generates high-quality conformers that are close with ground truth but also covers a wide variety of conformers in the distribution. In addition, it is important to note that MCF does not make any assumptions about the intrinsic geometric factors in conformers like torsional angles and thus provides a simple recipe to scale up the model. With growing number of parameters, larger MCF constantly achieves better performance than smaller counterpart in all metrics. In particular, when compared with MCF-S, MCF-B shows approximately 15% improvement on precision and even larger MCF-L improves it by approximately 20%. The experimental results demonstrate the power of scaling up proposed MCF in better solving conformer generation problem. Since our proposed method simply operates on 3D atomic positions, it provides a straightforward recipe for

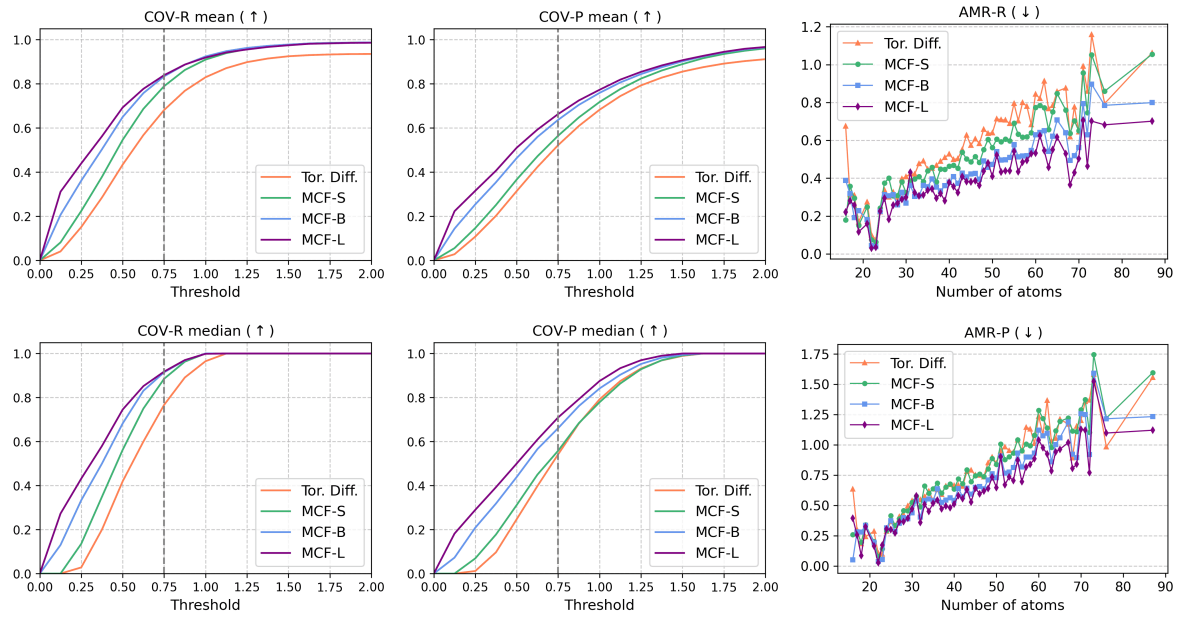

Figure 4. (a) Recall coverage and (b) precision coverage as a function of the threshold distance. MCF outperforms Torsional Diff. across the full spectrum of thresholds. (c) Averaged AMR of recall and precision as a function of the number of atoms in molecules.

scaling up the model. This sheds light on how scaling law could potentially benefit applications of deep generative models to scientific domains.

In Fig. 4, we further show a breakdown of the performance on GEOM-DRUGS of MCF with different sizes vs. Torsional diffusion (Jing et al., 2022) as a function of the threshold distance, as well as a function of the number of atoms in molecules. MCF outperforms Torsional Diff. across the full spectrum of thresholds in both recall and precision. When looking at the break-down AMR on different number of atoms in Fig. 4(c), MCF also demonstrates its superior performance for molecules of different sizes. It is indicated that MCF better captures the fine intrinsic geometric structure of conformers and scaling up the model helps improve the performance of proposed model. Also, as the number of parameters increases, MCF demonstrates better performance across all threshold levels in terms of both recall and precision. This provides further evidence on the performance gain from increasing the models size of MCF which is designed to be scalable in a straightforward way. We further investigate the ensemble properties of generated conformers in Appendix A.3.3. Fig. 10 in the Appendix shows examples of MCF generated conformers in GEOM-DRUGS.

5.3. Generalization to GEOM-XL

We now turn to the task of evaluating how well a model trained on GEOM-DRUGS transfers to unseen molecules with large numbers of atoms. Following Jing et al. (2022), we use the GEOM-XL dataset, a subset of GEOM-MoleculeNet that contains 102 molecules with more than 100 atoms. Note that this evaluation not only tests the capacity of models to generalize to larger and more complex molecules but also serves as an out-of-distribution generalization experiment.

	AMR-P ↓		AMR-R ↓		# mols
	mean	median	mean	median	
GeoDiff	2.92	2.62	3.35	3.15	-
GeoMol	2.47	2.39	3.30	3.14	-
Tor. Diff.	2.05	1.86	**2.94**	2.78	-
MCF-S	2.22	1.97	3.17	2.81	102
MCF-B	2.01	1.70	3.03	2.64	102
MCF-L	**1.97**	**1.60**	**2.94**	**2.43**	102
Tor. Diff. (our eval)	1.93	1.86	2.84	2.71	77
MCF-S	2.02	1.87	2.9	2.69	77
MCF-B	1.71	1.61	2.69	2.44	77
MCF-L	**1.64**	**1.51**	**2.57**	**2.26**	77

Table 3. Generalization results on GEOM-XL.

In Tab. 3 we report AMR for both precision and recall and compare with GeoDiff (Xu et al., 2022), GeoMol (Ganea et al., 2021) and Torsional Diff. (Jing et al., 2022). In particular, when taking the numbers directly from Jing et al. (2022), MCF-B achieves better or comparable performance than Torsional Diff. Further, in running the checkpoint provided by Torsional Diff. and following their validation process we found that 25 molecules failed to be generated,

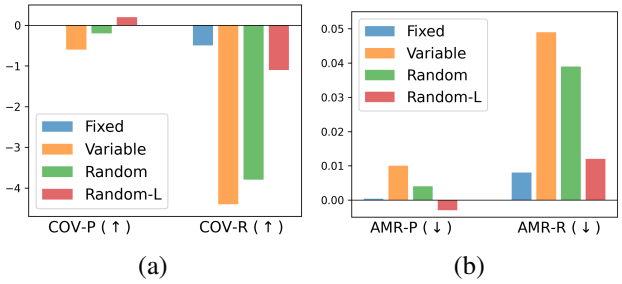

Figure 5. (a) Mean Coverage and (b) mean AMR of different rotation augmentation strategies on GEOM-QM9 when compared with training on original dataset.

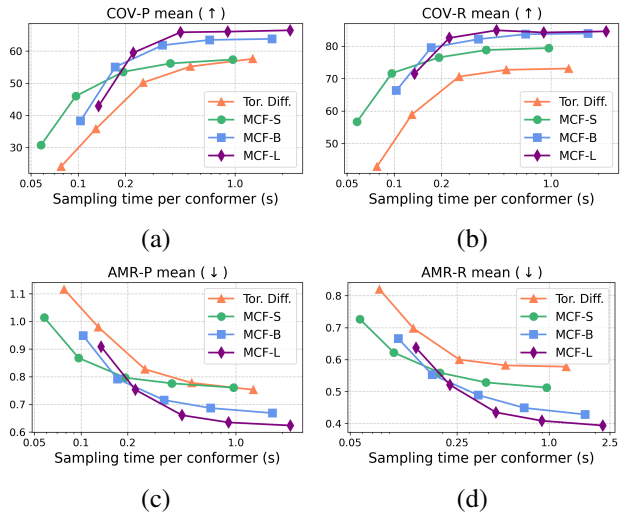

Figure 6. Inference wall clock time v.s. (a) precision coverage, (b) recall coverage, (c) precision AMR, and (d) recall AMR with Torsional Diff. and our MCF.

this is due to the fact that Torsional Diff. generates torsional angles conditioned on the molecular graph $\mathcal{G}$ and the local structures obtained from RDKit. And RDKit can fail to find local structures and Torsional Diff. cannot generate conformers in these cases. In our experiments with the same 77 molecules in GEOM-XL from our replica, MCF surpasses Torsional Diff. by a large margin. Leveraging little or no inductive bias in modeling molecular conformers, our proposed method is adaptable to wider variety of data.

Results also show that scaling up the model size to MCF-L further improves the generalizability to large and unseen molecules in GEOM-XL. In our replica, MCF-L demonstrates better performance than smaller model counterparts (*i.e.* MCF-S and MCF-B) and surpasses Torsional Diff. by a large margin. The results highlight the generalizability of MCF to large and complex molecules, which may shed a light on pre-training molecular conformer generation model.

5.4. Why does MCF generalize?

A natural question to ask is why does MCF generalize given its non-equivariant design. To answer this question we devised an experiment to understand if conformers in training and validation sets share a canonical coordinate system. In our experiment we apply different rotation transformations to GEOM-QM9 and train MCF on this transformed training set, while keeping the validation set unchanged. Three rotation transformations are investigated: 1) "Fixed" applies a single random rotation to all conformers, 2) "Variable" applies a different rotation to each conformer and keeps it through training, 3) "Random" applies a different random rotation to each conformer in each training epoch. Fig. 5 shows the results in these different settings.

Not surprisingly, applying a fixed rotation to the training set minimally affects performance. This is because a fixed rotation does not break relative SO(3) relations between conformers in the training set. However, rotating each conformer independently once during training (*e.g.* "Variable") negatively impacts performance. This finding points to the fact that the DFT simulations used to generate the data might be implicitly encoding a canonical coordinate system, which affects generalization if broken. Finally, applying a random rotation to each conformer on each training epoch forces the model to be invariant to any coordinate system, which is a more challenging task. Notably, though randomly rotating each conformer leads to worse results in recall, the performance drop is still marginal. Finally, by training a bigger model on this randomly rotated training set (*i.e.* Random-L) we can recover most of the performance gap in comparison with training on the original dataset.

These experiments show that inductive biases like roto-equivariance can be traded for scale in general-purpose models. Our results highlight that a domain-agnostic model at scale can achieve better performance than intricate models with strong domain-specific inductive biases in molecular conformer generation. We hope our findings will inspire the community to develop simple models that are prone to benefit from "scaling laws" especially when taking into account the fast growth of available scientific data.

5.5. Sampling

In this section, we investigate the performance of our MCF under limited computation budget in inference. To this end, we report COV and AMR of MCF with respect to different wall-clock sampling times. DDIM (Song et al., 2021), an efficient sampler, is applied, which uses a significantly smaller number of sampling steps than vanilla DDPM (*i.e.* 1000 sample steps). Specifically, we sample conformers with 3, 5, 10, 20, and 50 sampling steps with DDIM and compare the performance as well as inference time with

Torsional Diff. (Jing et al., 2022). All models are benchmarked on a single A100 GPU for comparison. It is shown that MCF is more efficient than Torsional Diffusion, which means that for the same inference time in seconds, MCF always outperforms Torsional Diffusion across all metrics and all model sizes. Notably, due to application of equivariant operations, Torsional Diff. can be time consuming in inference. Even when using a very limited sampling steps (*e.g.* 5 steps) MCF achieves comparable or better COV and AMR than Torsional Diff. using 50 sampling steps, while being more efficient in wall-clock inference time. MCF with different sizes achieves Pareto frontier of performance (*i.e.* COV and AMR) and sampling efficiency compared with Torsional Diff. This further indicates rotational or translational equivariance may not be a strong requirement while simple and scalable framework like MCF can own the merits in efficiency. Examples of MCF sampled conformers with different sampling steps can be found in Fig. 9 in the Appendix.

6. Conclusions

In this paper we introduced MCF, where we formulate the problem of molecular conformer generation as learning a diffusion model over functions on molecular graphs. MCF achieves state-of-the-art performance across different molecular generation benchmarks, surpassing models with hard-coded inductive biases by a large margin. Notably, MCF uses general-purpose Transformer-based score network rather than a model designed with specific inductive biases for molecules. MCF achieves superior results without explicitly modeling geometric properties of molecules like torsional angles, which makes it simpler to understand and scale. We believe MCF represents an exciting first step for future research on scaling conformer generation to proteins and other macro molecular structures. We hope our work serves as a reminder to the community to carefully consider the interplay between baking inductive biases in architectures while also considering the benefits of efficient and scalable approaches.

Impact Statement

This paper introduces a novel diffusion model based approach to generate molecular conformers, contributing to the advancement of computational chemistry and molecular modeling. The impact of this work extends to various scientific domains, enabling more accurate predictions of molecular structures and behaviors. The potential societal implications lie in the acceleration of drug discovery, materials science, and environmental studies, offering efficient solutions to complex molecular design challenges.

References

Abbott, R., Albergo, M. S., Boyda, D., Cranmer, K., Hackett, D. C., Kanwar, G., Racanière, S., Rezende, D. J., Romero-López, F., Shanahan, P. E., et al. Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions. *Physical Review D*, 106(7): 074506, 2022.

Abbott, R., Albergo, M. S., Botev, A., Boyda, D., Cranmer, K., Hackett, D. C., Kanwar, G., Matthews, A. G., Racanière, S., Razavi, A., et al. Normalizing flows for lattice gauge theory in arbitrary space-time dimension. *arXiv preprint arXiv:2305.02402*, 2023.

Arts, M., Garcia Satorras, V., Huang, C.-W., Zügner, D., Federici, M., Clementi, C., Noé, F., Pinsler, R., and van den Berg, R. Two for one: Diffusion models and force fields for coarse-grained molecular dynamics. *Journal of Chemical Theory and Computation*, 19(18):6151–6159, 2023.

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

Bai, Y., Mei, J., Yuille, A. L., and Xie, C. Are transformers more robust than cnns? *Advances in neural information processing systems*, 34:26831–26843, 2021.

Bannwarth, C., Ehlert, S., and Grimme, S. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of chemical theory and computation*, 15(3):1652–1671, 2019.

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):2453, 2022.

Berthelot, D., Autef, A., Lin, J., Yap, D. A., Zhai, S., Hu, S., Zheng, D., Talbot, W., and Gu, E. Tract: Denoising diffusion models with transitive closure time-distillation. *arXiv preprint arXiv:2303.04248*, 2023.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901, 2020.

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

Dao, T., Fu, D., Ermon, S., Rudra, A., and Ré, C. Flashattention: Fast and memory-efficient exact attention with io-awareness. *Advances in Neural Information Processing Systems*, 35:16344–16359, 2022.

Du, Y., Collins, K., Tenenbaum, J., and Sitzmann, V. Learning signal-agnostic manifolds of neural fields. In *NeurIPS*, 2021.

Dupont, E., Kim, H., Eslami, S., Rezende, D., and Rosenbaum, D. From data to functa: Your data point is a function and you should treat it like one. In *ICML*, 2022a.

Dupont, E., Teh, Y., and Doucet, A. Generative models as distributions of functions. In *AISTATS*, 2022b.

Dwivedi, V. P., Joshi, C. K., Laurent, T., Bengio, Y., and Bresson, X. Benchmarking graph neural networks. 2020.

Everett, B. *An introduction to latent variable models*. Springer, 2013.

Flam-Shepherd, D. and Aspuru-Guzik, A. Language models can generate molecules, materials, and protein binding sites directly in three dimensions as xyz, cif, and pdb files. *arXiv preprint arXiv:2305.05708*, 2023.

Ganea, O., Pattanaik, L., Coley, C., Barzilay, R., Jensen, K., Green, W., and Jaakkola, T. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. *Advances in Neural Information Processing Systems*, 34: 13757–13769, 2021.

Grattarola, D. and Vandergheynst, P. Generalised implicit neural representations. *arXiv preprint arXiv:2205.15674*, 2022.

Grebner, C., Becker, J., Stepanenko, S., and Engels, B. Efficiency of tabu-search-based conformational search algorithms. *Journal of Computational Chemistry*, 32(10): 2245–2253, 2011.

Gruver, N., Finzi, M., Goldblum, M., and Wilson, A. G. The lie derivative for measuring learned equivariance. *arXiv preprint arXiv:2210.02984*, 2022.

Guan, J., Qian, W. W., Ma, W.-Y., Ma, J., Peng, J., et al. Energy-inspired molecular conformation optimization. In *international conference on learning representations*, 2021.

Ha, D., Dai, A., and Le, Q. Hypernetworks. In *ICLR*, 2017.

Hawkins, P. C. Conformation generation: the state of the art. *Journal of chemical information and modeling*, 57 (8):1747–1756, 2017.

Hawkins, P. C., Skillman, A. G., Warren, G. L., Ellingson, B. A., and Stahl, M. T. Conformer generation with omega: algorithm and validation using high quality structures from the protein databank and cambridge structural database. *Journal of chemical information and modeling*, 50(4):572–584, 2010.

He, X., Hooi, B., Laurent, T., Perold, A., LeCun, Y., and Bresson, X. A generalization of vit/mlp-mixer to graphs. *arXiv preprint arXiv:2212.13350*, 2022.

Hernandez, V., Roman, J., Tomas, A., and Vidal, V. A survey of software for sparse eigenvalue problems. *Universitat Politecnica De Valencia, SLEPs technical report STR-6*, 2009.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.

Jabri, A., Fleet, D., and Chen, T. Scalable adaptive computation for iterative generation. *arXiv preprint arXiv:2212.11972*, 2022.

Jaegle, A., Borgeaud, S., Alayrac, J., et al. Perceiver io: A general architecture for structured inputs & outputs. In *ICLR*, 2022.

Jing, B., Corso, G., Chang, J., Barzilay, R., and Jaakkola, T. Torsional diffusion for molecular conformer generation. *Advances in Neural Information Processing Systems*, 35: 24240–24253, 2022.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

Kanwar, G., Albergo, M. S., Boyda, D., Cranmer, K., Hackett, D. C., Racaniere, S., Rezende, D. J., and Shanahan, P. E. Equivariant flow-based sampling for lattice gauge theory. *Physical Review Letters*, 125(12):121601, 2020.

Koestler, L., Grittner, D., Moeller, M., Cremers, D., and Lähner, Z. Intrinsic neural fields: Learning functions on manifolds. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, pp. 622–639. Springer, 2022.

Köhler, J., Klein, L., and Noé, F. Equivariant flows: exact likelihood generative learning for symmetric densities. In *International conference on machine learning*, pp. 5361–5370. PMLR, 2020.

Lee, J., Lee, Y., Kim, J., Kosiorek, A., Choi, S., and Teh, Y. W. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International conference on machine learning*, pp. 3744–3753. PMLR, 2019.

Lim, D., Robinson, J., Zhao, L., Smidt, T., Sra, S., Maron, H., and Jegelka, S. Sign and basis invariant networks for spectral graph representation learning. *arXiv preprint arXiv:2202.13013*, 2022.

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Maskey, S., Parviz, A., Thiessen, M., Stärk, H., Sadikaj, Y., and Maron, H. Generalized laplacian positional encoding for graph representation learning. *arXiv preprint arXiv:2210.15956*, 2022.

Nightingale, M. P. and Umrigar, C. J. *Quantum Monte Carlo methods in physics and chemistry*. Number 525. Springer Science & Business Media, 1998.

O Pinheiro, P. O., Rackers, J., Kleinhenz, J., Maser, M., Mahmood, O., Watkins, A., Ra, S., Sresht, V., and Saremi, S. 3d molecule generation by denoising voxel grids. *Advances in Neural Information Processing Systems*, 36, 2024.

Park, J., Florence, P., Straub, J., Newcombe, R., and Lovegrove, S. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019.

Ruddigkeit, L., Van Deursen, R., Blum, L. C., and Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17. *Journal of chemical information and modeling*, 52(11):2864–2875, 2012.

Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J., and Battaglia, P. Learning to simulate complex physics with graph networks. In *International conference on machine learning*, pp. 8459–8468. PMLR, 2020.

Sharp, N., Attaiki, S., Crane, K., and Ovsjanikov, M. Diffusionnet: Discretization agnostic learning on surfaces. *ACM Transactions on Graphics (TOG)*, 41(3):1–16, 2022.

Shi, C., Luo, S., Xu, M., and Tang, J. Learning gradient fields for molecular conformation generation. In *International conference on machine learning*, pp. 9558–9568. PMLR, 2021.

Simm, G. N. and Hernández-Lobato, J. M. A generative model for molecular distance geometry. *arXiv preprint arXiv:1909.11459*, 2019.

Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. In *ICLR*, 2021.

Song, Y., Dhariwal, P., Chen, M., and Sutskever, I. Consistency models. 2023.

Sutton, R. The bitter lesson. *Incomplete Ideas (blog)*, 13(1), 2019.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., et al. De novo design of protein structure and function with rfdiffusion. *Nature*, pp. 1–3, 2023.

Wilson, S. R., Cui, W., Moskowitz, J. W., and Schmidt, K. E. Applications of simulated annealing to the conformational analysis of flexible molecules. *Journal of computational chemistry*, 12(3):342–349, 1991.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Xu, M., Luo, S., Bengio, Y., Peng, J., and Tang, J. Learning neural generative dynamics for molecular conformation generation. *arXiv preprint arXiv:2102.10240*, 2021a.

Xu, M., Wang, W., Luo, S., Shi, C., Bengio, Y., Gomez-Bombarelli, R., and Tang, J. An end-to-end framework for molecular conformation generation via bilevel programming. In *International Conference on Machine Learning*, pp. 11537–11547. PMLR, 2021b.

Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S., and Tang, J. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.

Zaidi, S., Schaarschmidt, M., Martens, J., Kim, H., Teh, Y. W., Sanchez-Gonzalez, A., Battaglia, P., Pascanu, R., and Godwin, J. Pre-training via denoising for molecular property prediction. *arXiv preprint arXiv:2206.00133*, 2022.

Zhuang, P., Abnar, S., Gu, J., Schwing, A., Susskind, J., and Bautista, M. A. Diffusion probabilistic fields. In *ICLR*, 2023.

A. Appendix

A.1. Limitations and Future Work

While MCF shows competitive performance in molecular conformer generation, it does encounter limitations and potential improvements for future explorations. One limitation is that our proposed method is computationally expensive. Extensive computations first stem from the Transformer-based (Vaswani et al., 2017) score network. In MCF, we use a PerceiverIO (Jaegle et al., 2022) as score network, an efficient Transformer that allows for sub-quadratic compute, as well as FlashAttention (Dao et al., 2022) in implementation. Other efficient Transformer architectures and tricks like Jabri et al. (2022) can be used to improve training efficiency. The other factor is computational cost during inference. In MCF, we iterate 1000 timesteps to sample a conformer following DDPM (Ho et al., 2020). Experiments in Section 5.5 show that efficient sampling strategies, *i.e.* DDIM (Song et al., 2021), can help significantly increase inference efficiency while maintain high-quality in sampled conformers. Other efficient variants of diffusion models like consistency model (Song et al., 2023) as well as distillation approaches (Berthelot et al., 2023) may be adapted to further decrease the sampling to single step. Also, recent works have demonstrated that diffusion generative model can generate samples following Boltzmann distributions when provided with Boltzmann-distributed training data (Arts et al., 2023). Driven by this, our proposed MCF can be adapted to generate molecular conformers that follow Boltzmann distributions when trained with corresponding data. Besides, recent flow matching generative model (Lipman et al., 2022) provides the flexibility of mapping between arbitrary distributions and access to exact log-likelihood estimation. Integrating flow matching framework could help sample molecular conformer from Boltzmann distribution instead of standard Gaussian. Some recent works (Flam-Shepherd & Aspuru-Guzik, 2023; O Pinheiro et al., 2024; Gruver et al., 2022) also show that expressive models can learn equivariance from data, but they have not thoroughly investigated molecular conformer generation.

Another limitation could be how well MCF performs in low data regime. The propsed method may not perform as well as conformer generation when applied to problems with limited data or related to sequential problems like molecular dynamics (MD) simulations. In future work, we plan to extend MCF to conditional inference. For example, molecular docking can be formulated as conformer generation problem conditioned on proteins (Corso et al., 2022). Also, current framework can be expanded to *de novo* drug designs where no molecule information is provided (Hoogeboom et al., 2022). Besides, scaling up our model to large molecules, like proteins, can be of great interest. MCF by nature provides the flexibility to generate from partially observed sample, which can be suitable for designing proteins with known functional motifs (Watson et al., 2023).

A.2. Implementation details

In this section we describe implementation details for all our experiments. We also provide hyper-parameters and settings for the implementation of the score field network ϵ_θ and compute used for each experiment in the paper. In our experiments, we split GEOM-QM9 and GEOM-DRUGS randomly based on molecules into train/validation/test (80%/10%/10%). At the end, for each dataset, we report the performance on 1000 test molecules. Thus, the splits contain 106586/13323/1000 and 243473/30433/1000 molecules for GEOM-QM9 and GEOM-DRUGS, respectively. We follow the exact same training splits for all baselines (Ganea et al., 2021; Jing et al., 2022).

A.2.1. Score Field Network implementation details

The time-step t is incorporated into the score computation by concatenating a positional embedding representation of t to the context and query sets. The specific PerceiverIO settings used in all quantitatively evaluated experiments are presented in Tab. 4. An AdamW (Loshchilov & Hutter, 2017) optimizer is employed during training with a learning rate of $1e-4$. Cosine learning rate decay is deployed with 30K warmup steps. We use EMA with a decay of 0.999. Models are trained for 300K steps on GEOM-QM9 and 750K steps on GEOM-DRUSG. All models use an effective batch size of 512. A modified version of the publicly available repository is used for PerceiverIO [1]. Since molecules have different number of atoms, we set the number of context and query sets as the number of atoms during training and inference.

A.2.2. Atomic Features

We include atomic features alongside the graph Laplacians to model the key descriptions of molecules following previous works (Ganea et al., 2021; Jing et al., 2022). Detailed features are listed in Tab. 5. The atomic features are concatenated with graph Laplacian eigenvectors in both context and query inputs.

[1] https://huggingface.co/docs/transformers/model_doc/perceiver

Hyper-parameter	Small	Base	Large
num_freq_pos_embed	128	128	128
num_latent	128	512	1024
d_latennt	256	512	1024
d_model	512	1024	1024
num_enc_block	6	8	12
num_dec_block	2	2	2
num_self_attn_per_block	2	2	2
num_self_attn_head	4	4	8
num_cross_attn_head	4	4	8
# param	13M	64M	242M

Table 4. Hyperparameters and settings for MCF on different datasets.

Name	Description	Range
atomic	Atom type	one-hot of 35 elements in dataset
degree	Number of bonded neighbors	$\{x : 0 \leq x \leq 6, x \in \mathbb{Z}\}$
charge	Formal charge of atom	$\{x : -1 \leq x \leq 1, x \in \mathbb{Z}\}$
valence	Implicit valence of atom	$\{x : 0 \leq x \leq 6, x \in \mathbb{Z}\}$
hybrization	Hybrization type	{sp, sp^2, sp^3, sp^3d, sp^3d^2, other}
aromatic	Whether on a aromatic ring	{True, False}
num_rings	number of rings atom is in	$\{x : 0 \leq x \leq 3, x \in \mathbb{Z}\}$

Table 5. Atomic features included in MCF.

A.2.3. Compute

For GEOM-QM9, we train models using a machine with 4 Nvidia A100 GPUs using precision BF16. For GEOM-DRUGS, we train models using precision FP32, where MCF-B is trained with 8 Nvidia A100 GPUs and MCF-L is trained with 16 Nvidia A100 GPUs.

A.2.4. Evaluation Metrics

Following previous works (Xu et al., 2022; Ganea et al., 2021; Jing et al., 2022), we apply Average Minimum RMSD (AMR) and Coverage (COV) to measure the performance of molecular conformer generation. Let C_g denote the sets of generated conformations and C_r denote the one with reference conformations. For AMR and COV, we report both the Recall (R) and Precision (P). Recall evaluates how well the model locates ground-truth conformers within the generated samples, while precision reflects how many generated conformers are of good quality. The expressions of the metrics are given in the following equations:

$$\text{AMR-R}(C_g, C_r) = \frac{1}{|C_r|} \sum_{\mathbf{R} \in C_r} \min_{\hat{\mathbf{R}} \in C_g} \text{RMSD}(\mathbf{R}, \hat{\mathbf{R}}), \tag{3}$$

$$\text{COV-R}(C_g, C_r) = \frac{1}{|C_r|} |\{\mathbf{R} \in C_r | \text{RMSD}(\mathbf{R}, \hat{\mathbf{R}}) < \delta, \hat{\mathbf{R}} \in C_g\}|, \tag{4}$$

$$\text{AMR-P}(C_r, C_g) = \frac{1}{|C_g|} \sum_{\hat{\mathbf{R}} \in C_g} \min_{\mathbf{R} \in C_r} \text{RMSD}(\hat{\mathbf{R}}, \mathbf{R}), \tag{5}$$

$$\text{COV-P}(C_r, C_g) = \frac{1}{|C_g|} |\{\hat{\mathbf{R}} \in C_g | \text{RMSD}(\hat{\mathbf{R}}, \mathbf{R}) < \delta, \mathbf{R} \in C_r\}|, \tag{6}$$

where δ is a threshold. In general, a lower AMR scores indicate better accuracy and a higher COV score indicates a better diversity for the generative model. Following (Jing et al., 2022), δ is set as 0.5Å for GEOM-QM9 and 0.75Å for GEOM-DRUGS.

A.3. Additional experiments

In this section we include additional experiments ablating architecture choices, as well as prediction the ensemble properties of generated conformers.

A.3.1. Ablation experiments

In this section we provide an ablation study over the key design choices of MCF. We run all our ablation experiments on the GEOM-QM9 dataset following the settings in GeoMol (Ganea et al., 2021) and Torsional Diff. (Jing et al., 2022). In particular we study: (i) how does performance behave as a function of the number of Laplacian eigenvectors used in $\varphi(v)$. (ii) How does the model perform without atom features (*e.g.* how predictable conformers are given only the graph topology, without using atom features). Results in Tab. 6 show that the graph topology $\mathcal{G}$ encodes a surprising amount of information for sampling reasonable conformers in GEOM-QM9, as shown in row 2. In addition, we show how performance of MCF changes as a function of the number of eigen-functions k. Interestingly, with as few as $k = 2$ eigen-functions MCF is able to generate consistent accurate conformer.

A.3.2. Architectural choices

To further investigate the design choices of architecture in proposed MCF, we include additional experiments on GEOM-QM9 as shown in Tab. 6. To investigate the effectiveness of using Laplacian eigenvectors from LBO eigen-decomposition as positional embedding, we leverage SignNet (Lim et al., 2022) as the positional embedding, which explicitly models symmetries in eigenvectors. Using SignNet does not benefit the performance when compared with the standard MCF. Though adding edge attributes in SignNet achieves better performance than SignNet alone, the performance is still not rival. Also, it's worth mentioning that SignNet includes graph neural networks (Xu et al., 2018) and Set Transformer (Lee et al., 2019) which makes training less efficient.

In addition, we also report results using a vanilla Transformer encoder-decoder (TF) (Vaswani et al., 2017) as the backbone instead of PerceiverIO (PIO) (Jaegle et al., 2022). TF-base model contains 6 encoder layers and 6 decoder layers with 4 attention heads while TF-large contains 12 encoder layers and 12 decoder layers. The number of parameters match approximately with base and large sized PerceiverIO investigated in this work. Tab. 6 shows that TF-base is performing significantly worse than PIO-base with similar number of parameters. When increasing the model size, TF-large achieves on par performance as PIO-base, which validates the design choice of architecture in MCF.

				Precision				Recall			
				COV ↑		AMR ↓		COV ↑		AMR ↓	
k	atom feat.	PE	backbone	mean	median	mean	median	mean	median	mean	median
28	YES	LBO	PIO-base	95.00	100.00	0.103	0.044	93.67	100.00	0.119	0.055
28	NO	LBO	PIO-base	90.70	100.00	0.187	0.124	79.82	93.86	0.295	0.213
16	YES	LBO	PIO-base	94.87	100.00	0.139	0.093	87.54	100.00	0.220	0.151
8	YES	LBO	PIO-base	94.28	100.00	0.162	0.109	84.27	100.00	0.261	0.208
4	YES	LBO	PIO-base	94.57	100.00	0.145	0.093	86.83	100.00	0.225	0.151
2	YES	LBO	PIO-base	93.15	100.00	0.152	0.088	86.97	100.00	0.211	0.138
28	YES	SignNet	PIO-base	94.10	100.00	0.153	0.098	87.50	100.0	0.222	0.152
28	YES	$\text{SignNet}_{\text{attr}}$	PIO-base	95.30	100.00	0.143	0.091	90.20	100.00	0.197	0.135
28	YES	LBO	TF-base	94.92	100.00	0.131	0.083	89.33	100.00	0.194	0.132
28	YES	LBO	TF-large	95.49	100.00	0.110	0.061	93.48	100.00	0.135	0.073

Table 6. Ablation study with different network architectures on GEOM-QM9.

A.3.3. Ensemble properties

To fully assess the quality of generated conformers we also compute chemical property resemblance between the synthesized and the authentic ground truth ensembles. We select a random group of 100 molecules from the GEOM-DRUGS and produce a minimum of $2K$ and a maximum of 32 conformers for each molecule following (Jing et al., 2022). Subsequently, we undertake a comparison of the Boltzmann-weighted attributes of the created and the true ensembles. To elaborate, we calculate the following characteristics using xTB (as documented by (Bannwarth et al., 2019)): energy (E), dipole moment (μ), the gap between HOMO and LUMO ($\Delta\epsilon$), and the lowest possible energy, denoted as E_{min}. Since we don't

	E	μ	$\Delta\epsilon$	E_{min}
OMEGA	0.68	0.66	0.68	0.69
GeoDiff	0.31	0.35	0.89	0.39
GeoMol	0.42	0.34	0.59	0.40
Tor. Diff.	**0.22**	0.35	**0.54**	0.13
MCF	0.68±0.06	**0.28±0.05**	0.63±0.05	**0.04±0.00**
Tor. Diff. (our eval)	3.07±2.32	0.61±0.38	1.71±1.69	4.11±7.91
MCF	1.00±0.70	0.44±0.36	1.32±1.40	1.16±2.02

Table 7. Median averaged errors of ensemble properties between sampled and generated conformers (E, $\Delta\epsilon$, E_{min} in kcal/mol, and μ in debye).

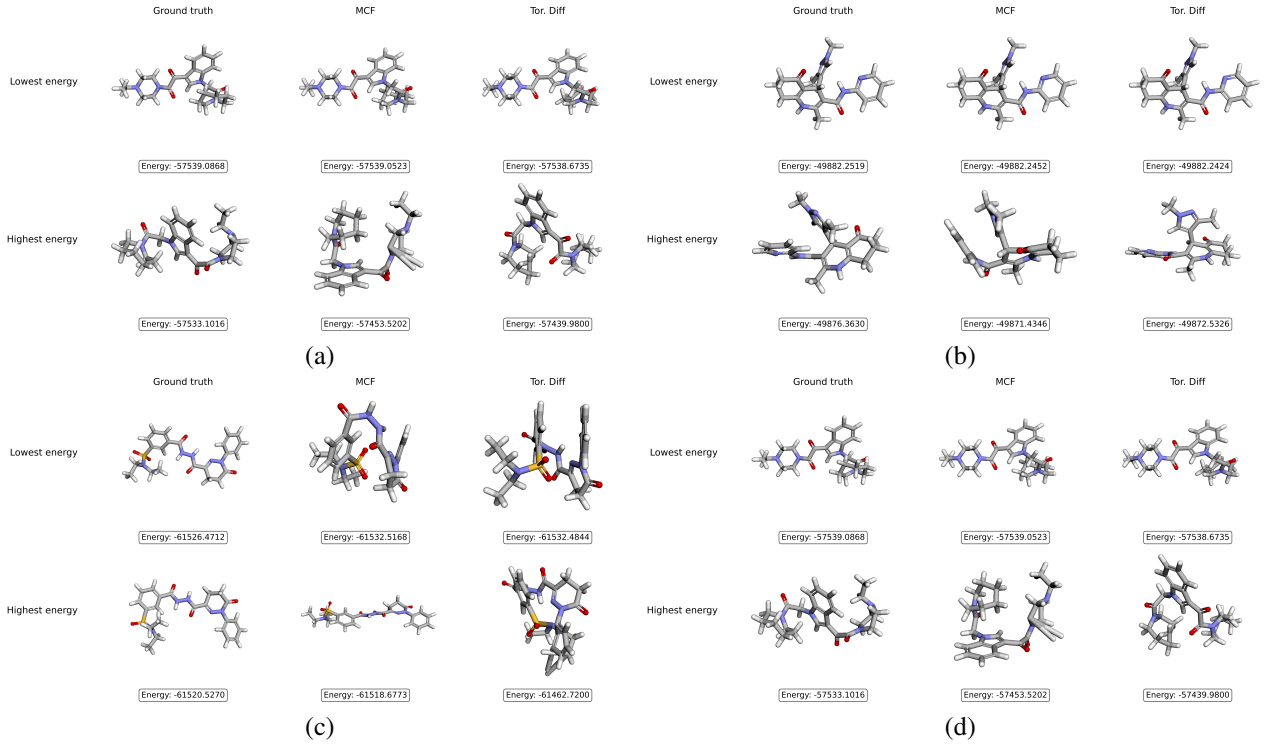

Figure 7. Examples of conformers with lowest and highest energies in ground truth, MCF samples, and Torsional Diff. samples for different molecules.

have the access to the exact subset of DRUGS used in (Jing et al., 2022), we randomly pick three subsets and report the averaged and standard deviation over three individual runs with different random seeds. The results are listed in Tab. 7. Our model achieves the lowest error on E_{min} when compared with other baselines, which demonstrates that MCF succeeds at generating stable conformers that are very close to the ground states. This could root from the fact that MCF doesn't rely on rule-based cheminfomatics methods and the model learns to better model stable conformers from data. Besides, MCF achieves competitive performance on μ and $\Delta\epsilon$. However, the error of E is high compared to the rest of approaches, meaning that though MCF performs well in generating samples close to ground states, it may also generate conformers with high energy that are not plausible in the dataset.

To further evaluate the performance on ensemble properties, we randomly pick 10 molecules from test set of GEOM-DRUGS and compare MCF with our replica Torsional Diff. on the subset as shown in the last two rows of Tab. 7. We use the checkpoints from the public GitHub repository[2] of Torsional Diff. to sample conformers. Unlike previous setting which only sample 32 conformers, we sample $2K$ conformers for a molecule with K ground truth conformers. We report the average and standard deviation of errors over the 10 molecules. It is indicated that MCF generates samples with ensemble

[2] https://github.com/gcorso/torsional-diffusion

properties that are closer to the ground truth. Fig. 7 shows the conformers with lowest and highest energy in ground truth, MCF samples, and Torsional Diff. samples.

A.4. Continuous conformers

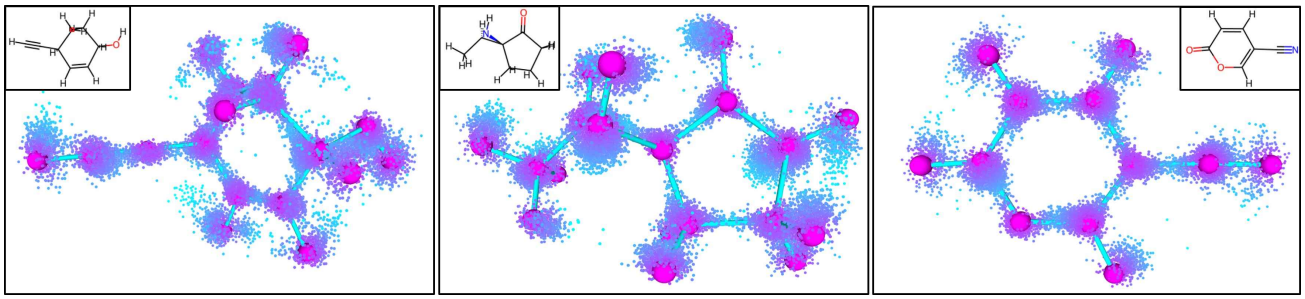

Figure 8. Continuously evaluating generated conformer fields for different molecules in GEOM-QM9.

Molecular conformers are defined as discrete atomic positions in 3D Euclidean space. Since MCF encodes continuous conformer field, it can be *continuously evaluated* in $\mathcal{G}$, which maps arbitrary points in $\mathcal{G}$ to 3D positional space $\mathbb{R}^3$. In order to do this, for a point p in a bond connecting atoms (v_i, v_j) we linearly interpolate the Laplacian eigenvector representation of it's endpoints $\varphi(p) = \alpha\varphi(v_i) + (1-\alpha)\varphi(v_j)$, we then feed this interpolated Laplacian eigenvector into the model to sample its 3D position in the conformer field. We visualize results in Fig. 1 and 8. We generated this visualizations an MCF model trained on GEOM-QM9 without atom features. Note that while MCF is never trained on points along molecular bonds, it manages to generate plausible 3D positions for such points.

Here, the experiment conceptually investigates the flexibility of defining conformer generation problem as a field. It is shown that MCF can generate feasible conformers even when the input interpolated eigenfunctions have never been seen during training. Such that MCF is not over-fitted to certain eigenfunctions and learns to generate distributional aspects of atomic positions purely from correlations in training data. Also, when provided molecular conformer data with distribution of electron density from Quantum Monte Carlo methods (Nightingale & Umrigar, 1998), MCF may be extended to predict electron density beyond atomic positions in future works as well. We recognize this is highly speculative and needs further empirical investigation to substantiate in future works.

A.5. Additional visualization

Fig. 9 show some examples of sampled conformers from MCF with different sampling steps. It is illustrated that even with very limited sample steps, MCF can still generate plausible conformers especially for the heavy atoms. We also show examples of conformers from ground truth, Torsional Diff., and our MCF. Fig. 10 and 11 depict samples from GEOM-DRUGS and GEOM-XL, respectively. We found the samples that are most aligned with ground truth and plot them side by side.

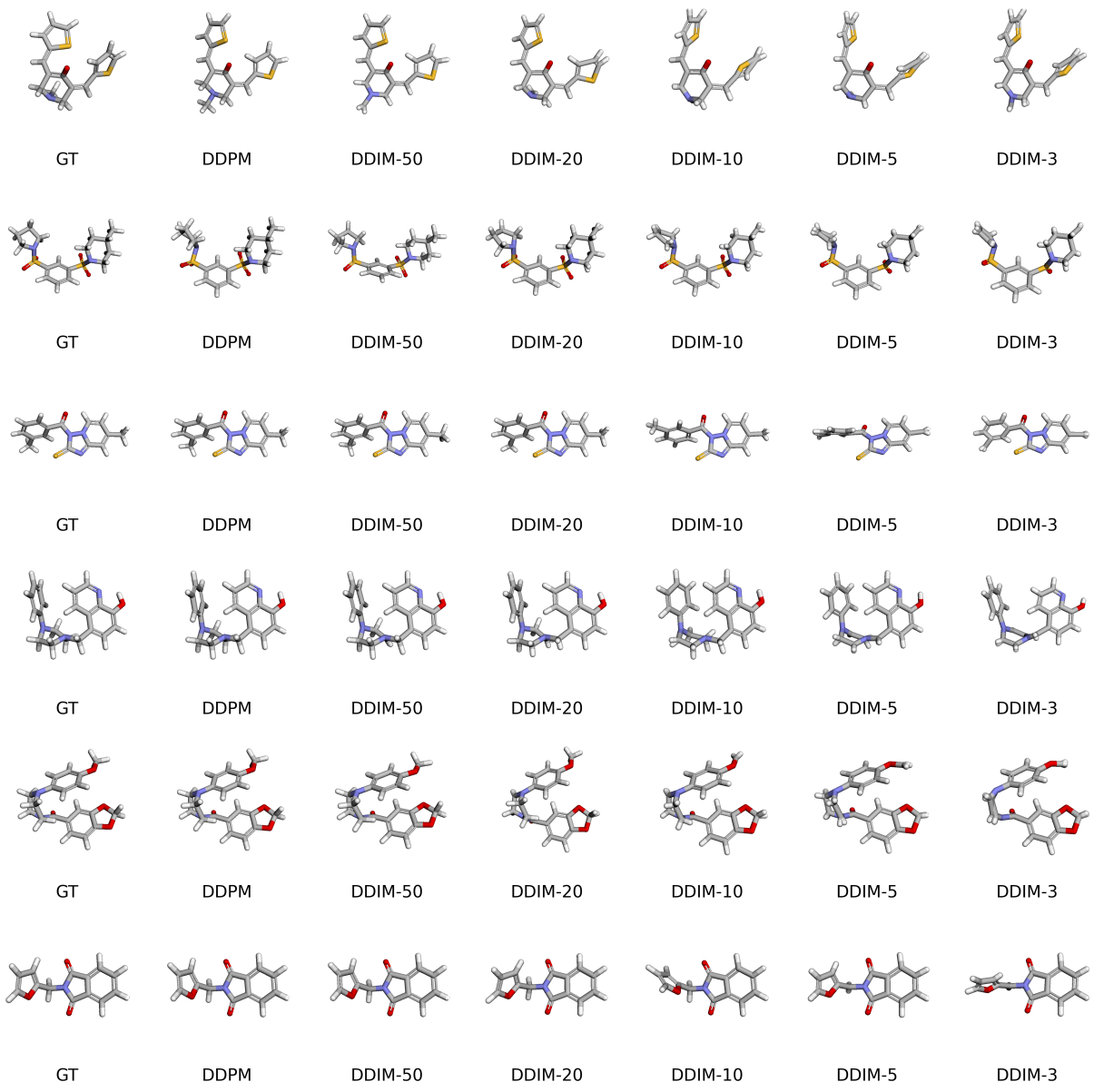

Figure 9. Examples of conformers with different sampling steps.

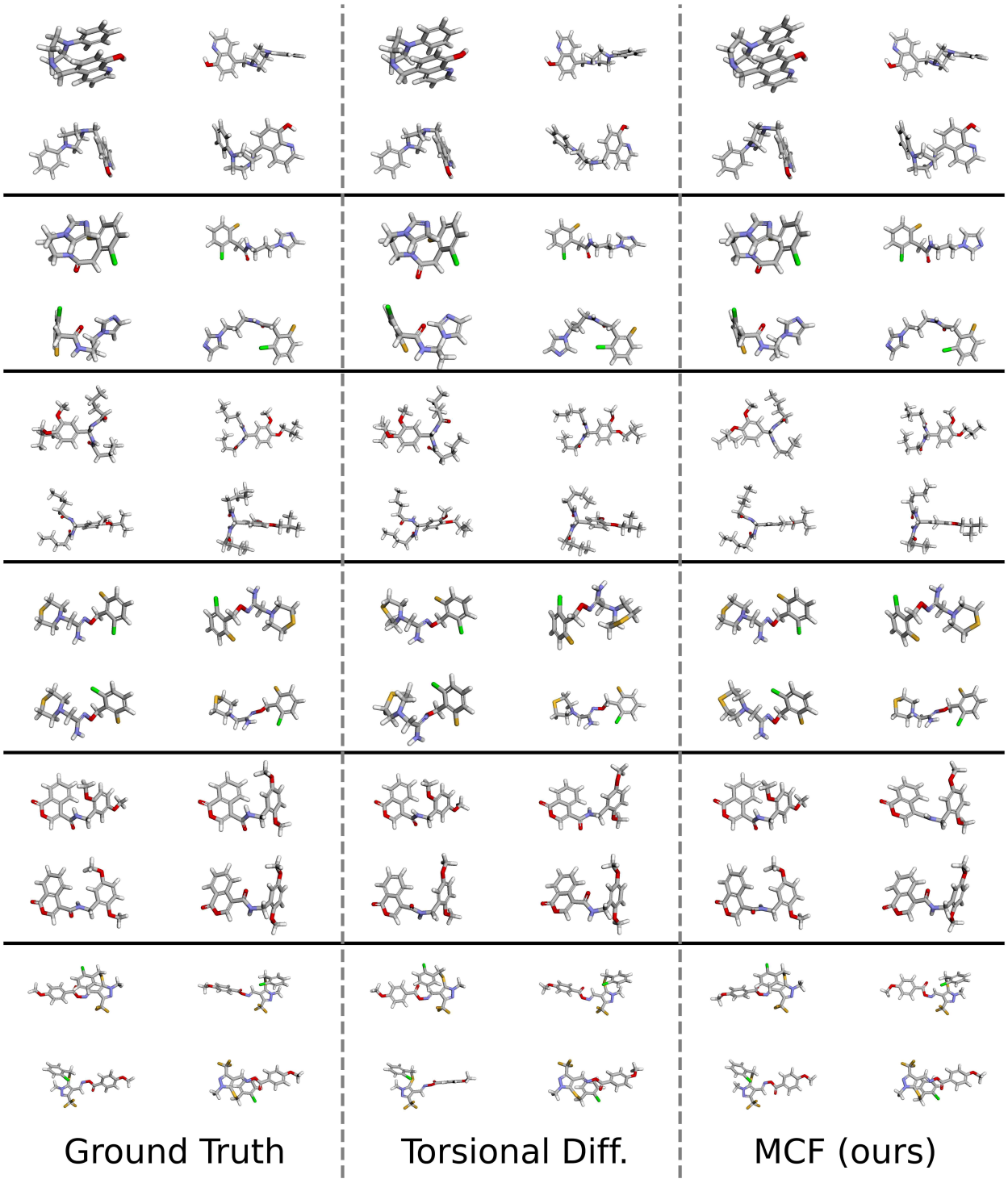

Figure 10. Examples of conformers of ground truth, Torsional Diff. samples, and MCF samples from GEOM-DRUGS.

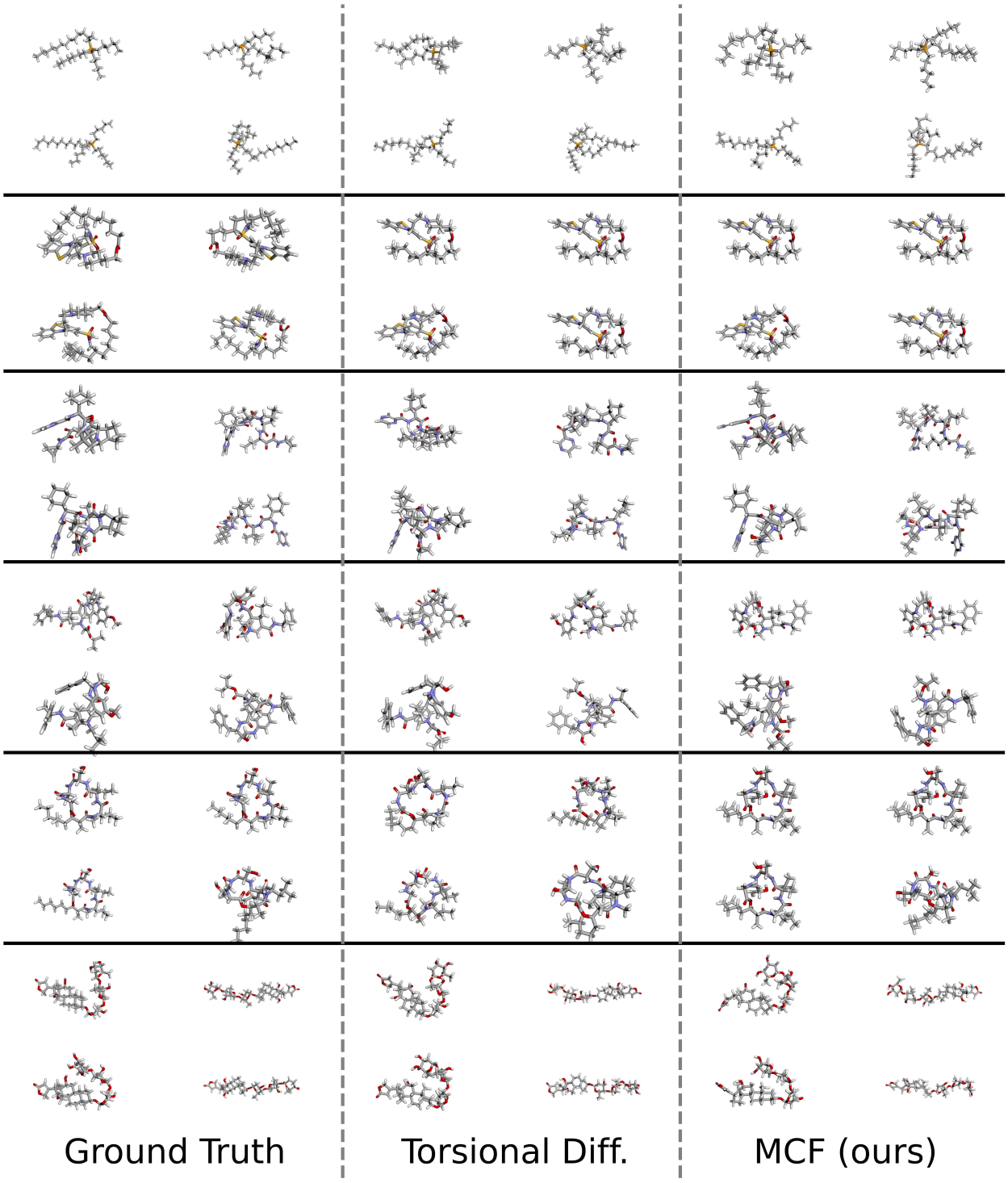

Figure 11. Examples of conformers of ground truth, Torsional Diff. samples, and MCF samples from GEOM-XL.