Learning Gradient Fields for Molecular Conformation Generation

Chence Shi*[1], Shitong Luo*[2], Minkai Xu[1], Jian Tang[1,3]

[1]Mila-Quebec AI Institute [2]Peking University [3]HEC Montreal, Canada

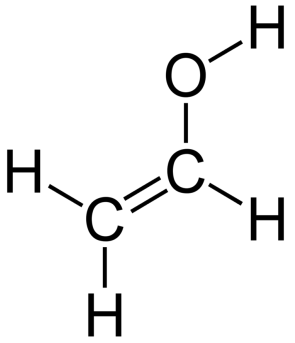
(*: equal contribution)

Molecule Representations

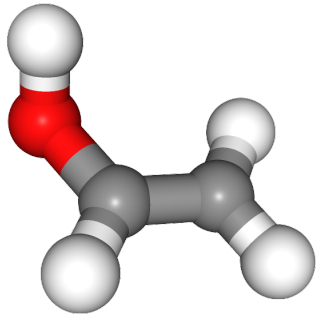
- Understanding properties of molecules is important in a variety of applications
  - Drug discovery, material discovery
- Molecule representations
  - 1D SMILES
  - 2D Molecular graphs
- A more natural and intrinsic representations: **3D conformations**
  - Determines its biological and physical activities
  - e.g., charge distribution, steric constraints, and interaction with other molecules

C=CO

1D SMILES



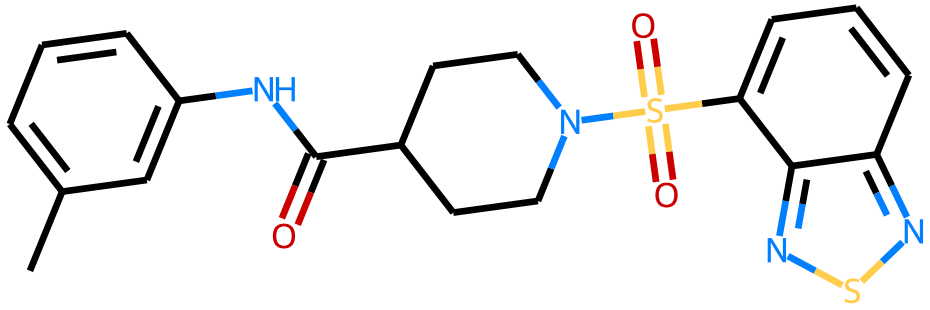
2D Graph



3D Conformation

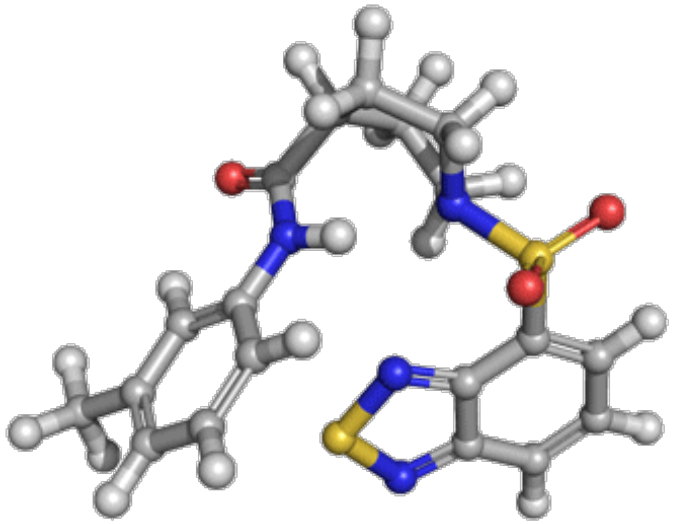
Conformation Prediction

- For real-world molecules, computing 3D structures is expensive
- We study how to predict valid and stable conformations from molecular graph
  - Molecular graph $\mathcal{G}$: 2D atom-bond graph
  - Conformation $\boldsymbol{R}$: atomic 3D coordinates



$\mathcal{G}$

$p(\boldsymbol{R}|\mathcal{G})$



$$\boldsymbol{R} \propto \underbrace{\exp(-E(\boldsymbol{R})/k_B T)}_{\text{Boltzmann distribution}}$$

Limitation of previous works & Motivation

- Likelihood of conformations is not rotation and translation invariant[1]. Distance based methods[2,3] generate outputs (distances) that are proxies of the actual desired object (atomic coordinates)
- This motivates us to pursue an algorithm that (C1) generates conformations within a single stage, and (C2) preserves the roto-translation equivariance of conformations.

[1]Mansimov, Elman, et al. "Molecular geometry prediction using a deep generative graph neural network." *Scientific reports* 9.1 (2019): 1-13.
[2]Simm, Gregor NC, and José Miguel Hernández-Lobato. "A generative model for molecular distance geometry." *arXiv preprint arXiv:1909.11459* (2019).
[3]Minkai Xu, Shitong Luo, Yoshua Bengio, Jian Peng, and Jian Tang. Learning neural generative dynamics for molecular conformation generation. ICLR 2021

Our Solution: ConfGF (ICML' 21)

- Inspired by molecular dynamics, which use force fields for simulation, we seek to estimate the pseudo forces acting on atoms, i.e., $S_\theta(\boldsymbol{R}) = \nabla_R \log p(\boldsymbol{R} \mid \mathcal{G})$.
- Samples are generated by iteratively applying the pseudo forces to a randomly initialized 3D structure via Langevin dynamics (satisfy C1)
- We develop an algorithm to effectively estimate these gradients and meanwhile preserve their roto-translation equivariance (satisfy C2)

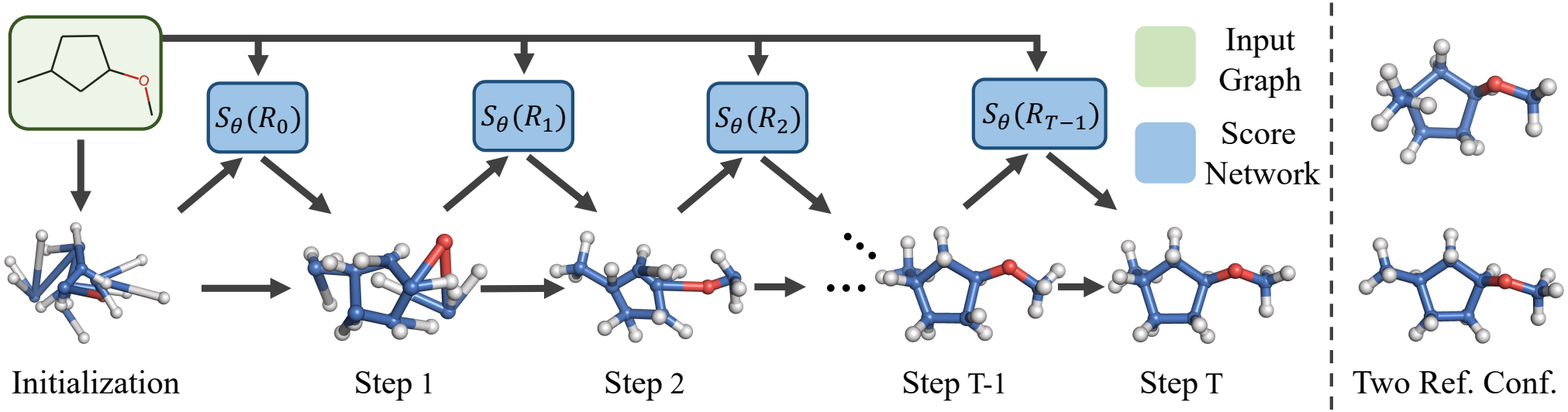
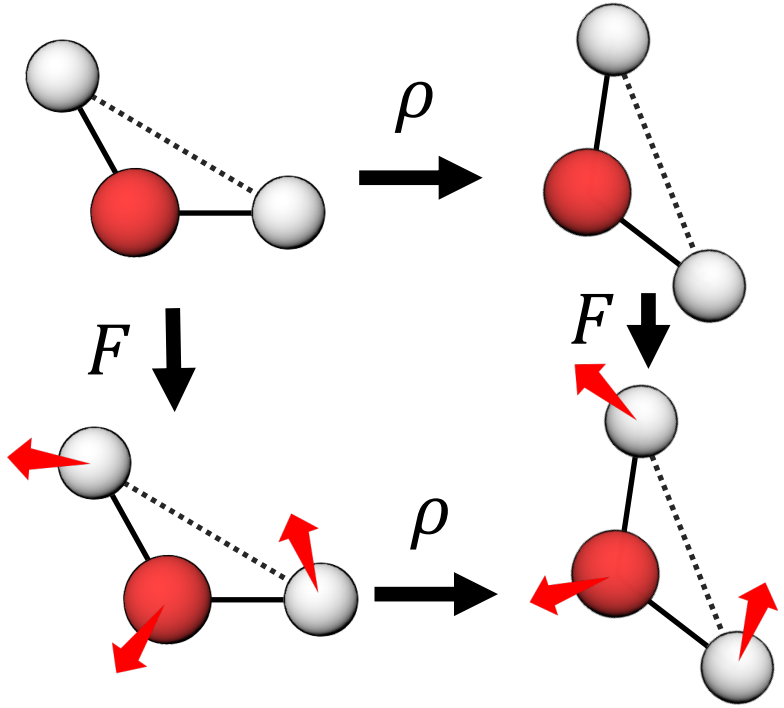


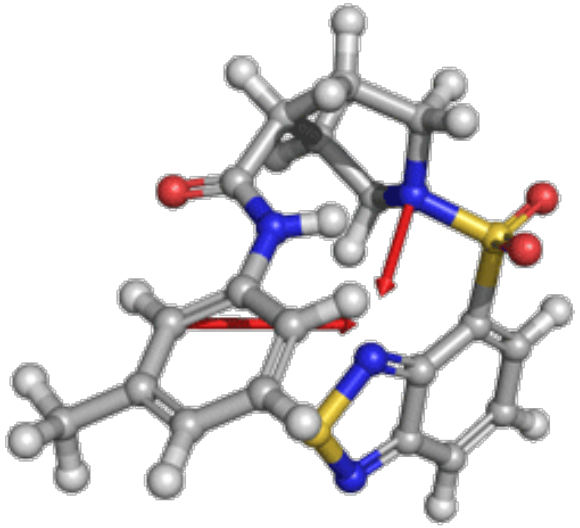
Fig. Generation procedure of GonfGF

Challenge: 3D Rotation Equivariance

$$F \circ \rho(x) = \rho \circ F(x)$$

The equation says that applying the ρ on the input has the same effect as applying it to the output.





A GIF illustrating the rotation equivariance of atomic forces. Two red arrows stand for forces acting on atoms, which rotate together with the molecule.

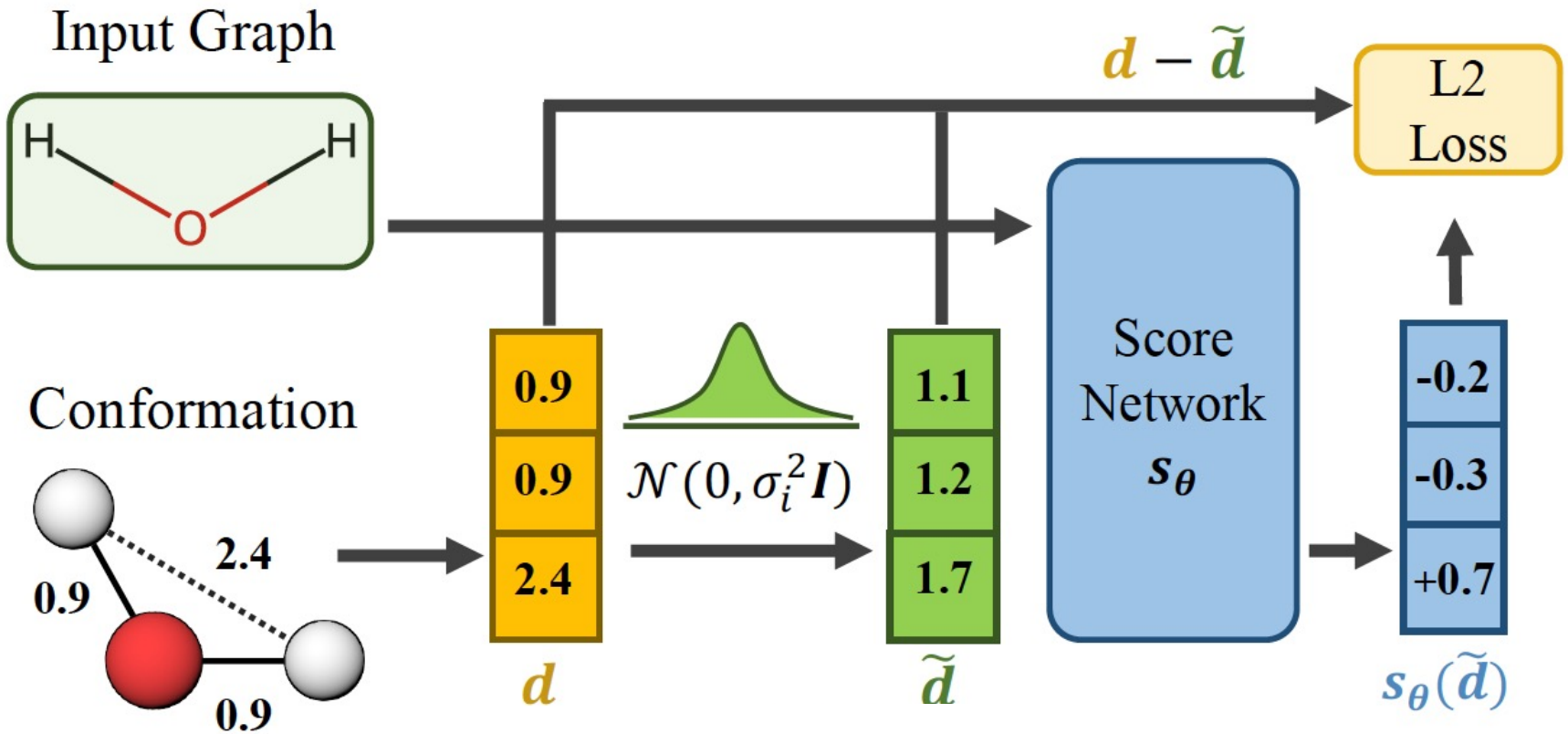
Our Solution: Gradient propagation via chain rule

- We observe that interatomic distances are differentiable *w.r.t.* atomic coordinates.
- We estimate gradient fields of interatomic distances, i.e., $S_\theta(\boldsymbol{d}) = \nabla_d \log p(\boldsymbol{d} \mid \mathcal{G})$ via denoising score matching[1]
- We then backpropagate gradients from $\boldsymbol{d}$ to $\boldsymbol{R}$ via chain rule:

$$\begin{aligned}
\forall i, \mathbf{s}_\theta(\mathbf{R})_i = \frac{\partial f_G(\mathbf{d})}{\partial \mathbf{r}_i} &= \sum_{(i,j), e_{ij} \in E} \underbrace{\frac{\partial f_G(\mathbf{d})}{\partial d_{ij}} \cdot \frac{\partial d_{ij}}{\partial \mathbf{r}_i}}_{\text{chain rule}} \\
&= \sum_{j \in N(i)} \frac{1}{d_{ij}} \cdot \frac{\partial f_G(\mathbf{d})}{\partial d_{ij}} \cdot (\mathbf{r}_i - \mathbf{r}_j) \\
&= \sum_{j \in N(i)} \frac{1}{d_{ij}} \cdot \underbrace{\mathbf{s}_\theta(\mathbf{d})_{ij}}_{\text{score of distance}} \cdot (\mathbf{r}_i - \mathbf{r}_j),
\end{aligned}$$

[1]Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. In Advances in Neural Information Processing Systems,

Training



$$\frac{1}{2L}\sum_{i=1}^{L}\lambda(\sigma_i)\mathbb{E}_{p(\mathbf{d}|G)}\mathbb{E}_{q_{\sigma_i}(\tilde{\mathbf{d}}|\mathbf{d},G)}\left[\left\|\frac{\mathbf{s}_\theta(\tilde{\mathbf{d}})}{\sigma_i}+\frac{\tilde{\mathbf{d}}-\mathbf{d}}{\sigma_i^2}\right\|_2^2\right]$$

Generation

- Given the current conformation ***R***, we calculate atomic gradients by propagating gradients from ***d*** to ***R*** via chain rule:

$$\forall i, \mathbf{s}_\theta(\mathbf{R})_i = \sum_{j \in N(i)} \frac{1}{d_{ij}} \cdot \mathbf{s}_\theta(\mathbf{d})_{ij} \cdot (\mathbf{r}_i - \mathbf{r}_j)$$

- This process can be viewed as calculating resultant force on atoms
- The conformations are sequentially updated based on atomic gradients

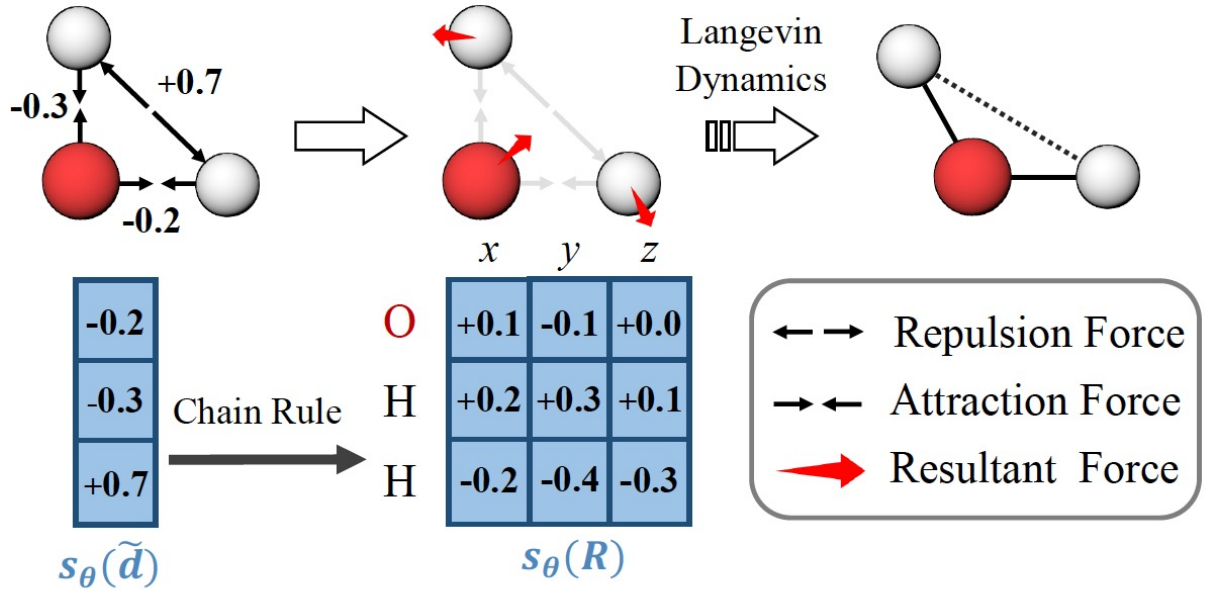


Fig. Score estimation via chain rule

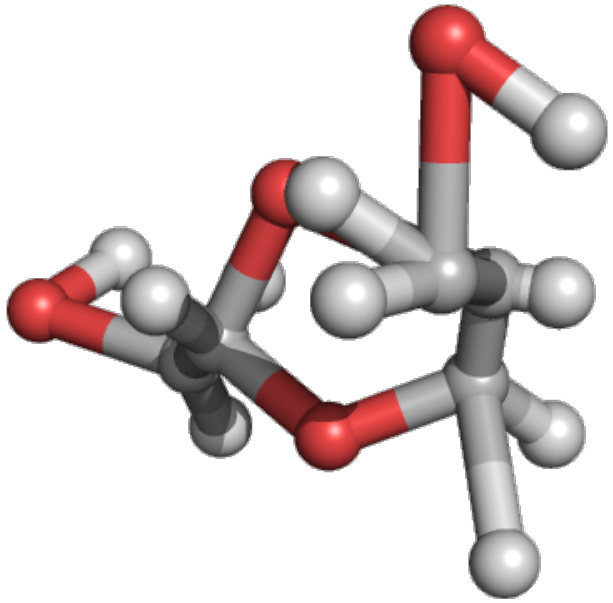
Algorithm 1 Annealed Langevin dynamics sampling

input molecular graph G, noise levels $\{\sigma_i\}_{i=1}^{L}$, the smallest step size ϵ, and the number of sampling steps per noise level T.

Initialize conformation $\mathbf{R}_0$ from a prior distribution
for $i \leftarrow 1$ to L **do**
  $\alpha_i \leftarrow \epsilon \cdot \sigma_i^2 / \sigma_L^2$ ▷ α_i is the step size.
  for $t \leftarrow 1$ to T **do**
    $\mathbf{d}_{t-1} = g_G(\mathbf{R}_{t-1})$ ▷ get distances from $\mathbf{R}_{t-1}$
    $\mathbf{s}_\theta(\mathbf{R}_{t-1}, \sigma_i) \leftarrow \text{convert}(\mathbf{s}_\theta(\mathbf{d}_{t-1}, \sigma_i))$ ▷ Eq. 3.
    Draw $\mathbf{z}_t \sim \mathcal{N}(0, \mathbf{I})$
    $\mathbf{R}_t \leftarrow \mathbf{R}_{t-1} + \alpha_i \mathbf{s}_\theta(\mathbf{R}_{t-1}, \sigma_i) + \sqrt{2\alpha_i}\mathbf{z}_t$
  end for
  $\mathbf{R}_0 \leftarrow \mathbf{R}_T$
end for

output Generated conformation $\mathbf{R}_T$.

Demo



Starting from a random initialization, the conformation is sequentially updated with the gradient information of atomic coordinates calculated from the score network

Experiments

- Data Sets
  - GEOM: > 33 million molecular conformers by MIT group, including both small molecules in QM9 and medium-sized drug-like molecules
- Baselines
  - CVGAE (Mansimov et al. 2019): learning atom representations with GNNs and then predicting the coordinates of atoms
  - GraphDG (Simm&Hernandez-Lobato, 2020) and CGCF (Xu et al., 2021): generating the pairwise distances between atoms and then generating conformers based on distances
  - RDKit: a classical Euclidean Distance Geometry-based approach

Evaluation Metrics

- Discrepancy between two conformations: Root-Mean-Square Deviation (RMSD)

$$\mathrm{RMSD}(\boldsymbol{R}, \hat{\boldsymbol{R}}) = \left(\frac{1}{n} \sum_{i=1}^{n} \| \boldsymbol{R}_i - \hat{\boldsymbol{R}}_i \|^2 \right)^{\frac{1}{2}}$$

- Coverage (COV): the fraction of conformations in the reference set that are matched by at least one conformation in the generated conformations

$$\mathrm{COV}(\mathbb{S}_g(\mathcal{G}), \mathbb{S}_r(\mathcal{G})) = \frac{1}{|\mathbb{S}_r|} \left| \left\{ \boldsymbol{R} \in \mathbb{S}_r \middle| \, \mathrm{RMSD}(\boldsymbol{R}, \boldsymbol{R}') < \delta, \boldsymbol{R}' \in \mathbb{S}_g \right\} \right|$$

- Matching (MAT): measure the average distance of the reference conformations with their nearest neighbors in the generated conformations

$$\mathrm{MAT}(\mathbb{S}_g(\mathcal{G}), \mathbb{S}_r(\mathcal{G})) = \frac{1}{|\mathbb{S}_r|} \sum_{\boldsymbol{R}' \in \mathbb{S}_r} \min_{\boldsymbol{R} \in \mathbb{S}_g} \mathrm{RMSD}(\boldsymbol{R}, \boldsymbol{R}').$$

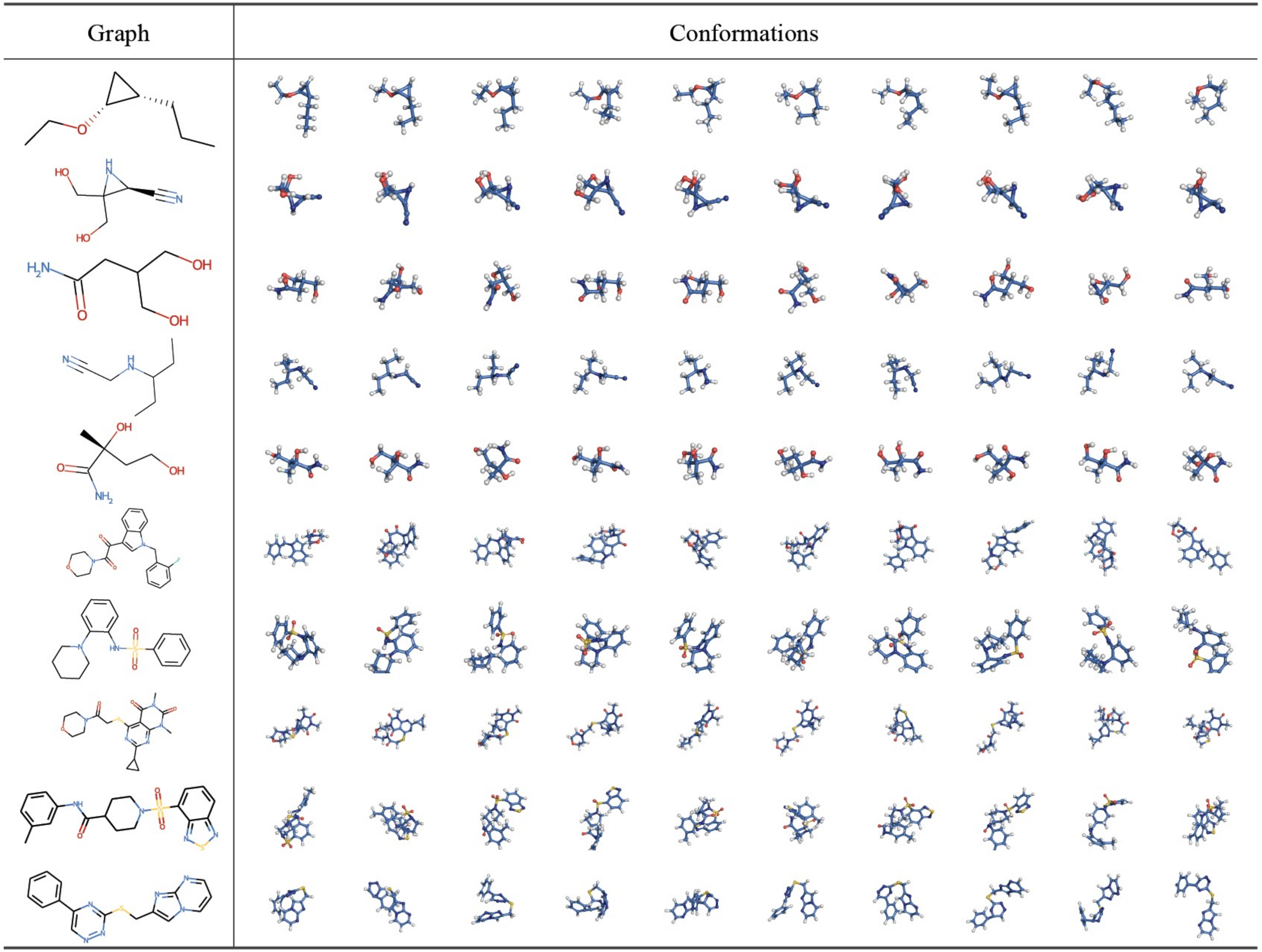
Results

Table 1: COV and MAT scores on GEOM-QM9 and GEOM-Drugs datasets. The threshold δ of COV score is 0.5Å for GEOM-QM9 and 1.25Å for GEOM-Drugs following Xu et al. [41]. (↑): the higher the better. (↓): the lower the better.

	GEOM-QM9				GEOM-Drugs			
	COV (%, ↑)		MAT (Å, ↓)		COV (%, ↑)		MAT (Å, ↓)	
Method	Mean	Median	Mean	Median	Mean	Median	Mean	Median
RDKit	83.26	90.78	0.3447	0.2935	60.91	65.70	1.2026	1.1252
CVGAE	0.09	0.00	1.6713	1.6088	0.00	0.00	3.0702	2.9937
GraphDG	73.33	84.21	0.4245	0.3973	8.27	0.00	1.9722	1.9845
CGCF	77.52	80.40	0.4206	0.3903	54.19	56.35	1.2575	1.2356
ConfGF	**88.49**	**94.13**	**0.2673**	**0.2685**	**62.15**	**70.93**	**1.1629**	**1.1596**

ConfGF achieves the state-of-the-art performance on all four metrics

Examples



Code will be released at:

https://github.com/DeepGraphLearning

Thanks for listening!

Also feel free to contact me later at chence.shi@umontreal.ca

Take-away Messages

- The one-stage framework, which avoid generating proxies of the atomic coordinates (distances), greatly enhance the performance of conformation generation
- Interatomic distances are continuously differentiable *w.r.t.* atomic coordinates. Gradients can be propagated from distances to Cartesian Coordinates, and the roto-translation equivariance is preserved.
- The framework is very general and can be applied to other systems, e.g., proteins