Equivariant Flow Matching for Molecular Conformer Generation

Majdi Hassan*, Nikhil Shenoy*, Jungyoon Lee*, Hannes Stark, Stephan Thaler, Dominique Beaini

Mila · Valence Labs · Université de Montréal · The University of British Columbia · Massachusetts Institute of Technology

Motivation: Efficiently and accurately generate equilibrium conformations responsible for several biological, chemical and physical properties.

Contributions

1. **ET-Flow obtains state-of-the-art precision results for conformer prediction and ensemble property prediction.**
2. **Counter to the prevailing status quo in the literature, we show that equivariance proves beneficial.**
3. **ET-Flow demonstrates competitive performance while using orders of magnitude fewer sampling steps than GeoDiff and significantly fewer parameters than MCF**

Background

Static, linear interpolant to bridge the particles sampled from the base (ρ_0) and target (ρ_1) densities, $x_0 \sim \rho_0$ and $x_1 \sim \rho_1$ respectively.

$$I_t(x_0, x_1) = \alpha x_1 + \beta x_0$$

Add stochasticity by sampling from a gaussian centered at the interpolant.

$$\rho_t(x) = \mathcal{N}(x \mid I_t(\mathbf{x}_0, \mathbf{x}_1), \sigma_t^2 \mathbf{I})$$

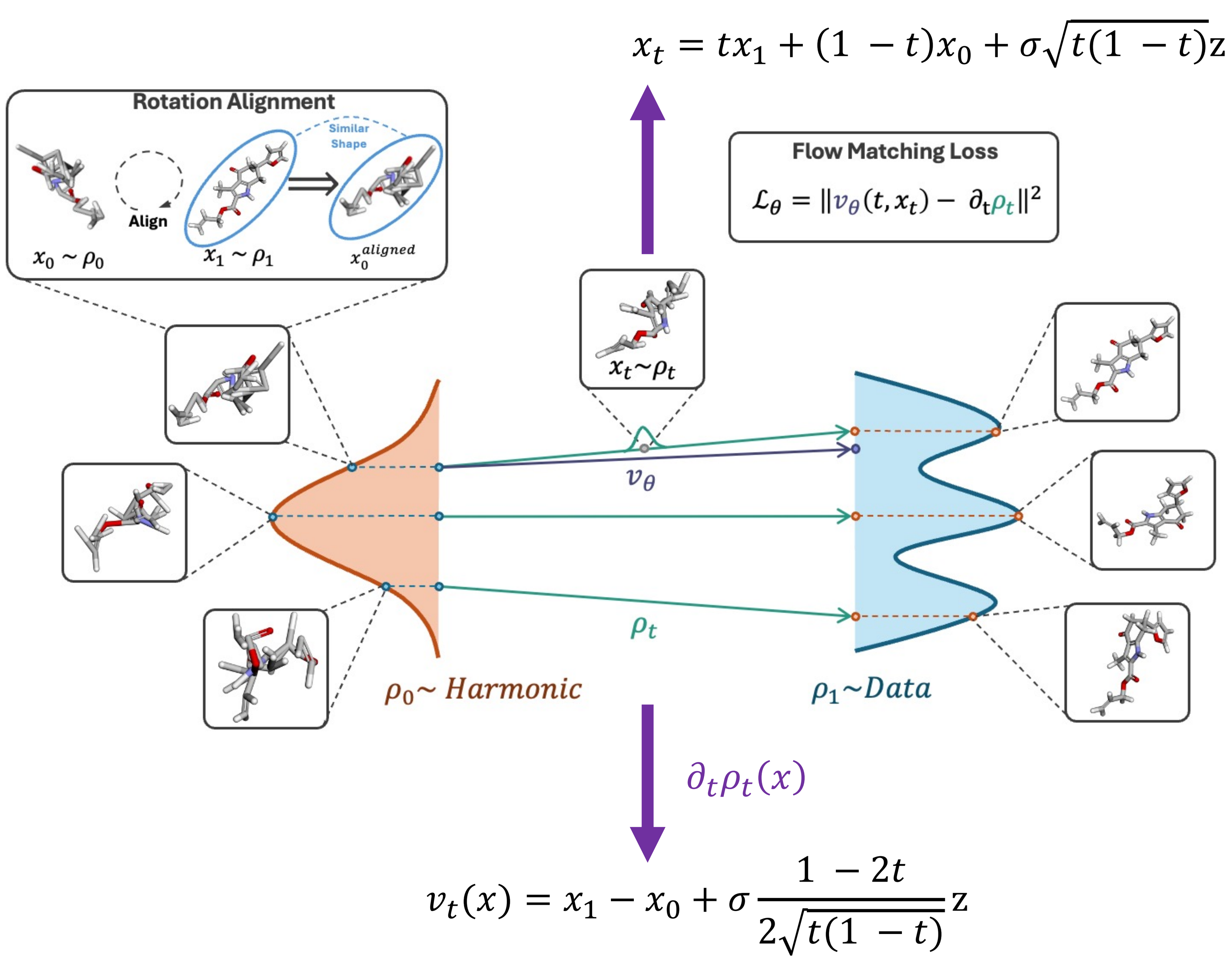
Flow matching (FM) learns the vector field $v_{t(x)} = \partial_t \rho_t(x)$.

$$v_t(\mathbf{x}) = \dot{\alpha}_t \mathbf{x}_1 + \dot{\beta}_t \mathbf{x}_0 + \dot{\sigma}_t \mathbf{z} \qquad \mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$$

Method

Overview

Architecture: TorchMD-Equivariant Transformer (8.3M)



Chirality Correction

- Post-hoc orientation correction when Oriented Volume (OV) does not match RDKit chiral tags.

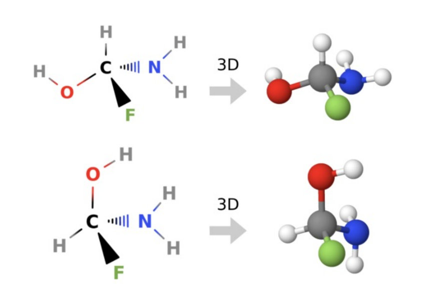
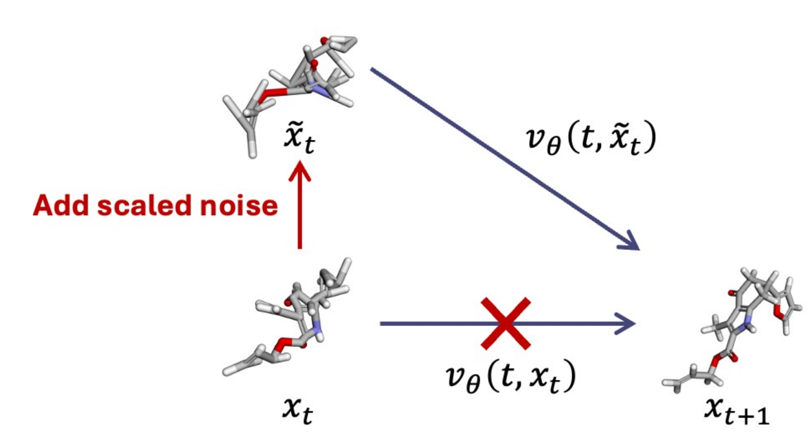


Figure from Ganea et al 2021

$$OV(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4) = sign\left(\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{vmatrix}\right)$$

Stochastic Sampling

- Introduce stochasticity into inference by adding noise at each time step between (t=0.8~1.0) and evaluating the vector field from the state



Results

RMSD Metrics

DRUGS

	Recall				Precision			
	Coverage ↑		AMR ↓		Coverage ↑		AMR ↓	
	mean	median	mean	median	mean	median	mean	median
GeoDiff	42.10	37.80	0.835	0.809	24.90	14.50	1.136	1.090
GeoMol	44.60	41.40	0.875	0.834	43.00	36.40	0.928	0.841
Torsional Diff.	72.70	80.00	0.582	0.565	55.20	56.90	0.778	0.729
MCF - S (13M)	79.4	87.5	0.512	0.492	57.4	57.6	0.761	0.715
MCF - B (62M)	84.0	91.5	0.427	0.402	64.0	66.2	0.667	0.605
MCF - L (242M)	**84.7**	**92.2**	**0.390**	**0.247**	66.8	71.3	0.618	0.530
ET-Flow (8.3M)	79.53	84.57	0.452	0.419	74.38	81.04	0.541	0.470
ET-Flow - SS (8.3M)	79.62	84.63	0.439	0.406	**75.19**	**81.66**	**0.517**	**0.442**

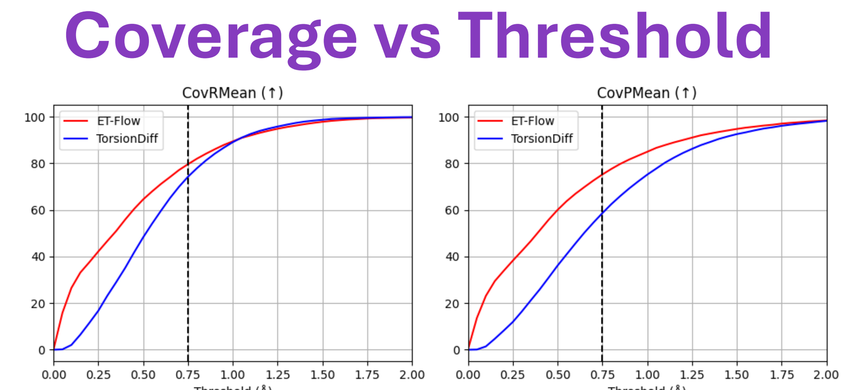
QM9

	Recall				Precision			
	Coverage ↑		AMR ↓		Coverage ↑		AMR ↓	
	mean	median	mean	median	mean	median	mean	median
CGCF	69.47	96.15	0.425	0.374	38.20	33.33	0.711	0.695
GeoDiff	76.50	**100.00**	0.297	0.229	50.00	33.50	1.524	0.510
GeoMol	91.50	**100.00**	0.225	0.193	87.60	**100.00**	0.270	0.241
Torsional Diff.	92.80	**100.00**	0.178	0.147	92.70	**100.00**	0.221	0.195
MCF	**95.0**	**100.00**	0.103	0.044	**93.7**	**100.00**	0.119	0.055
ET-Flow (ours)	**94.99**	**100.00**	**0.083**	**0.035**	91.00	**100.00**	**0.116**	**0.047**

Energy Metrics

	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
Torsional 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
ET-Flow	**0.18±0.01**	**0.18±0.01**	**0.35±0.06**	**0.02±0.00**

Coverage vs Threshold



Ablations

	Recall				Precision			
	Coverage ↑		AMR ↓		Coverage ↑		AMR ↓	
	mean	median	mean	median	mean	median	mean	median
ET-Flow	75.37	82.35	0.557	0.529	58.90	60.87	0.742	0.690
ET-Flow ($O(3)$)	72.74	79.21	0.576	0.556	54.84	54.11	0.794	0.739
ET-Flow (w/o Alignment)	68.67	74.71	0.622	0.611	47.09	44.25	0.870	0.832
ET-Flow (Gaussian Prior)	66.53	73.01	0.640	0.625	44.41	40.88	0.903	0.864

	Recall				Precision			
	Coverage ↑		AMR ↓		Coverage ↑		AMR ↓	
	mean	median	mean	median	mean	median	mean	median
ET-Flow (5 Steps)	77.84	82.21	0.476	0.443	74.03	80.8	0.55	0.474
ET-Flow (10 Steps)	79.05	84.00	0.451	0.415	74.64	**81.38**	0.533	0.457
ET-Flow (20 Steps)	79.29	84.04	**0.449**	**0.413**	**74.89**	81.32	**0.531**	**0.454**
ET-Flow (50 Steps)	**79.53**	**84.57**	0.452	0.419	74.38	81.04	0.541	0.470

Conclusion

- Incorporating equivariance still proves useful in generative modeling.
- ET-Flow has great flexibility in incorporating more useful priors.
- FM improves inference efficiency generating accurate samples with as few as 5 steps.

Future Work

- Incorporate SO(3) equivariance to avoid additional chirality correction step.
- Scaling parameters to boost recall metrics.