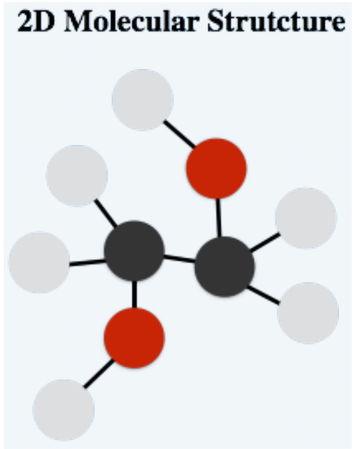
MUDiff: Unified Diffusion for Complete Molecule Generation

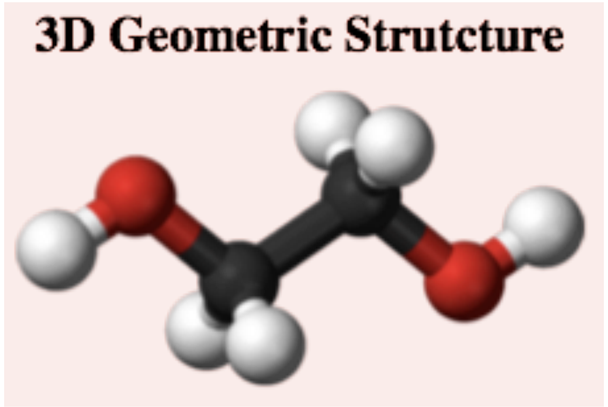
Chenqing Hua, Sitao Luan, Minkai Xu, Rex Ying, Jie Fu, Stefano Ermon, Doina Precup

In Proceedings of the Second Learning on Graphs Conference (LoG 2023)

We Started For A Simple Reason…

- We know that both **molecular structures and conformations** are important for representing a complete molecule





- When we start this project (a year and a half ago), **there is no model** for generating both perspectives for a molecule
- We find it is insufficient, so we aim to design a generative model, **capable of capturing 2D structures and 3D conformations for generation**

We Aim To…

- We realize **2D structures and 3D conformations are essential** for a comprehensive molecular representation

- Therefore, we aim to introduce a diffusion model, which **joinly learns and generates** these two aspects for a molecule

- **The idea is simple**, the joint model should combine the diffusion processes for structures and conformations **in one step**

We Introduce MUDiff…

Algorithm 1 Training MUDiff

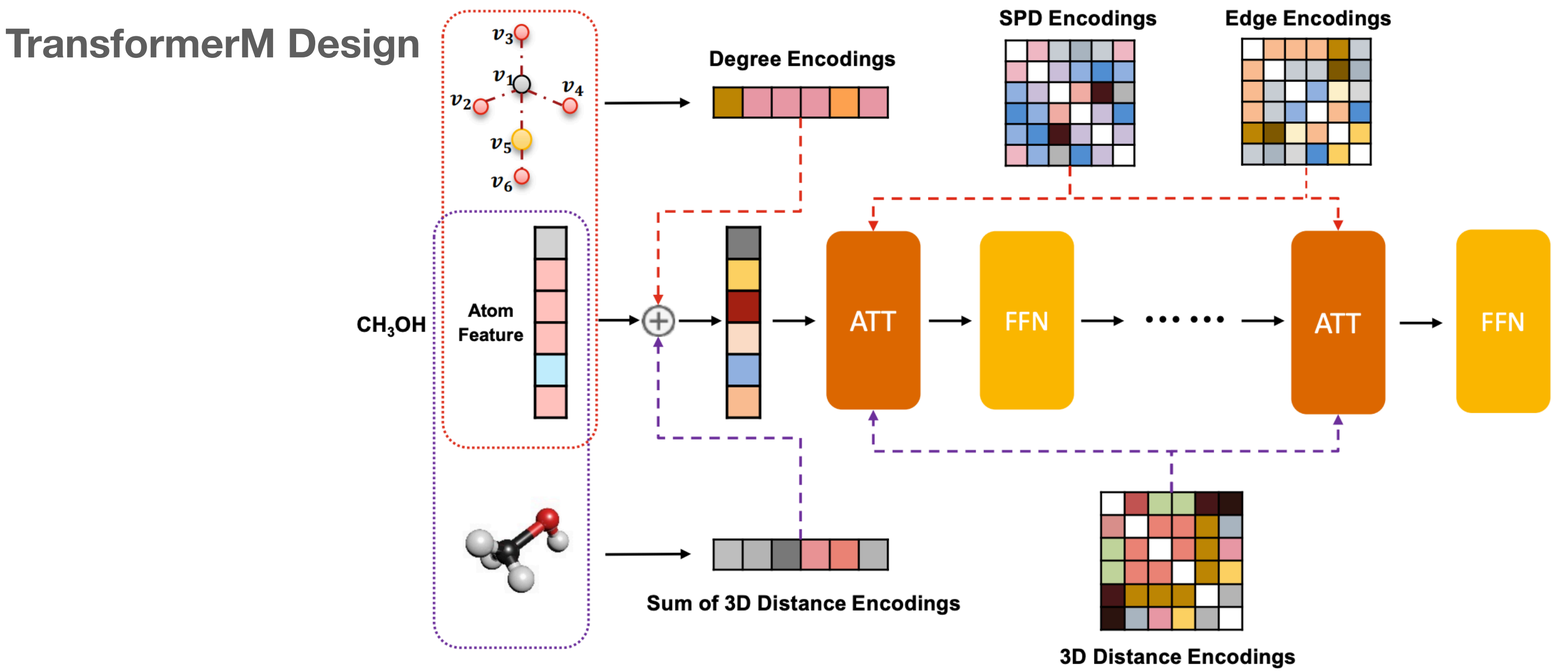
Input: A complete molecule $\mathbf{M} = (\mathbf{H}, \mathbf{E}, \mathbf{X})$
Sample $t \sim \mathcal{U}(1, \cdots, T)$
Sample $\boldsymbol{\epsilon}_{\mathbf{H}}, \boldsymbol{\epsilon}_{\mathbf{X}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
Subtract center of gravity from $\boldsymbol{\epsilon}_{\mathbf{X}}$
Compute $\tilde{\mathbf{H}}_t = \alpha_t \mathbf{H} + \sigma_t \boldsymbol{\epsilon}^t_{\mathbf{H}},\ \tilde{\mathbf{X}}_t = \alpha_t \mathbf{X} + \sigma_t \boldsymbol{\epsilon}^t_{\mathbf{X}}$
Sample $\tilde{\mathbf{E}}_t \sim \mathbf{E}\tilde{Q}_t$
Compute $\hat{\boldsymbol{\epsilon}}^t_{\mathbf{H}}, \hat{\boldsymbol{\epsilon}}^t_{\mathbf{X}}, p(\hat{\mathbf{E}}) = \psi_\theta([\tilde{\mathbf{H}}_t, \frac{t}{T}], \tilde{\mathbf{X}}_t, \tilde{\mathbf{E}}_t) - (\mathbf{0}, \tilde{\mathbf{X}}_t, \mathbf{0})$
Minimize $\|\boldsymbol{\epsilon}^t_{\mathbf{H}} - \hat{\boldsymbol{\epsilon}}^t_{\mathbf{H}}\|^2 + \|\boldsymbol{\epsilon}^t_{\mathbf{X}} - \hat{\boldsymbol{\epsilon}}^t_{\mathbf{X}}\|^2 + \mathrm{CE}(\mathbf{E}, p(\hat{\mathbf{E}}))$

Algorithm 2 Sampling from MUDiff

Sample $\tilde{\mathbf{M}}_T$: $\tilde{\mathbf{H}}_T, \tilde{\mathbf{X}}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),\ \tilde{\mathbf{E}}_T \sim q_\infty$
for $t = T, T-1, \ldots, 1$ **do**
  Compute $\hat{\boldsymbol{\epsilon}}^t_{\mathbf{H}}, \hat{\boldsymbol{\epsilon}}^t_{\mathbf{X}}, \hat{\mathbf{E}} = \psi_\theta([\tilde{\mathbf{H}}_t, \frac{t}{T}], \tilde{\mathbf{X}}_t, \tilde{\mathbf{E}}_t) - (\mathbf{0}, \tilde{\mathbf{X}}_t, \mathbf{0})$
  Sample $\tilde{\mathbf{E}}_{t-1} \sim p(\tilde{\mathbf{E}}_{t-1} | \tilde{\mathbf{E}}_t)$
  Sample $\boldsymbol{\epsilon}_{\mathbf{H}}, \boldsymbol{\epsilon}_{\mathbf{X}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  Compute $\tilde{\mathbf{H}}_{t-1} = \frac{\tilde{\mathbf{H}}_t}{\alpha_{t|t-1}} - \frac{\sigma^2_{t|t-1} \hat{\boldsymbol{\epsilon}}^t_{\mathbf{H}}}{\alpha_{t|t-1}\sigma_t} + \sigma_{t \to t-1} \boldsymbol{\epsilon}_{\mathbf{H}}$
  Subtract center of gravity from $\boldsymbol{\epsilon}_{\mathbf{X}}$
  Compute $\tilde{\mathbf{X}}_{t-1} = \frac{\tilde{\mathbf{X}}_t}{\alpha_{t|t-1}} - \frac{\sigma^2_{t|t-1} \hat{\boldsymbol{\epsilon}}^t_{\mathbf{X}}}{\alpha_{t|t-1}\sigma_t} + \sigma_{t \to t-1} \boldsymbol{\epsilon}_{\mathbf{X}}$
)**end for**
Sample $\mathbf{M} \sim p(\mathbf{M} | \tilde{\mathbf{M}}_0)$

- We introduce MUDiff, a model that **jointly learns and generates molecular structures and conformations in one step**
- The design is simple, consisting of discrete and continuous diffusion processes, **adding noises to atom features, edge features, and atom positions**
- To achieve this, we introduce MUformer as the backbone autoencoder model

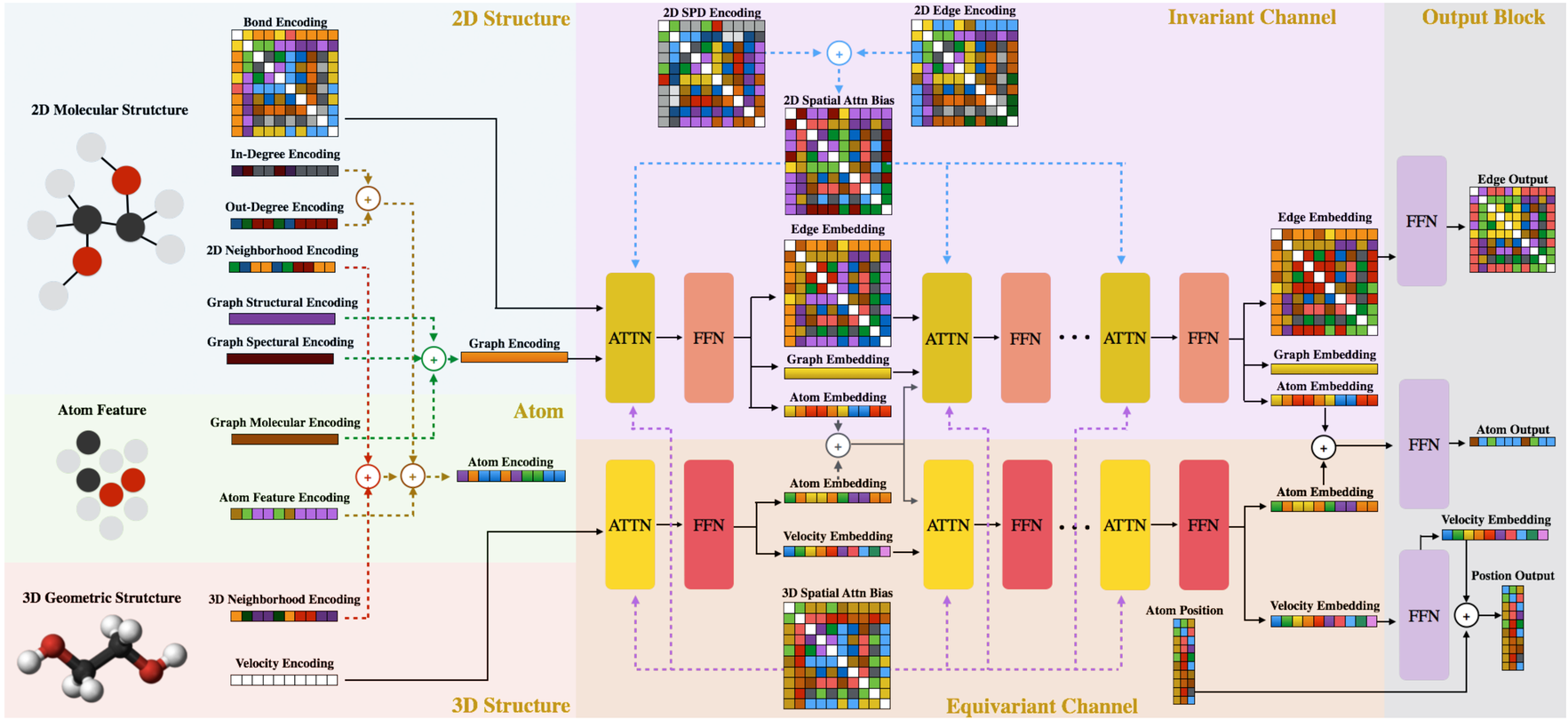
We Draw Inspiration From…



- We draw inspiration from TransformerM [1] to design MUformer, taking its advantage to process different data types of a molecule
- TransformerM take multiple views of a molecule, having **two seperate channels to process edge encodings (structures) and distance encodings (conformations)** for property predictions
- We are aware the two-channel design is well-aligned with our task

[1] ONE TRANSFORMER CAN UNDERSTAND BOTH 2D & 3D MOLECULAR DATA

We Design Backbone MUformer Autoencoder…



- MUformer process 2D molecular graphs and 3D molecular conformations, but with invariant (purple) and equivariant (brown) channels, similar to TransformerM
- A distinct advantage is, the model still works when either 2D or 3D data is missing
- **With missing 2D structures**, MUformer activate equivariant (brown) channel, **predicting atom features and positions**
- **With missing 3D conformations**, MUformer activate invariant (purple) channel, **predicting atom features and structures**
- **With 2D structures and 3D conformations provided**, MUformer activate both invariant and equivariant channels, **predicting atom features, positions, and molecular structures**

Experiment

Molecule Generation

- We evaluate the atom and molecule stability of the generated compounds by measuring the proportion of atoms that have the correct valency for atom stability, and the proportion of generated molecules in which all atoms are stable for molecule stability.

- The table shows that MUDiff can generate molecules that are significantly more stable than the baseline models in terms of negative log-likelihood and molecule stability

.

Table 2: Negative log-likelihood, atom stability, and molecule stability are evaluated with standard deviation across 3 runs on QM9, using 10K samples (with hydrogen) from the model. The results surpass those of previous models, as reported in [4, 6].

Method	NLL	Atom Stable(%)	Mol Stable(%)
Data	-	99.0	95.2
ENF	-59.7	85.0	4.9
GSchnet	-	95.7	68.1
GDM	-92.5	97.6	71.6
EDM	-110.7 $\pm$ 1.5	98.7 $\pm$ 0.1	82.0 $\pm$ 0.4
DiGress	-	98.1 $\pm$ 0.3	79.8 $\pm$ 5.6
MDM	-	98.6	**91.9**
GeoLDM	-	**98.9** $\pm$ 0.1	89.4 $\pm$ 0.5
MUDiff (ours)	**-135.5** $\pm$ 2.1	**98.8** $\pm$ 0.2	**89.9** $\pm$ 1.1

Special Experiment

Molecule Generation with Limited 3D Data

- We introduce a new molecule generation task that incorporates **limited 3D data**, as many real-world datasets lack complete 3D structures.
- We randomly split the 100K training molecules into two sets: **30K with both 2D and 3D structures and 70K with only 2D structures**. We train the model on the 30K samples using both the invariant and equivariant channels and validate on 18K samples until NLL converges.
- We fine-tune the trained model on the remaining 70K molecules with only 2D structures and validate/test on 18K/13K samples.
- MUDiff achieved **competitive results in generating stable molecules**, even with limited 3D information in the training set, compared to the baselines.

Table 1: Negative log-likelihood, atom stability, and molecule stability are evaluated with standard deviation across 3 runs on QM9, using 10K samples from the model. 30K+70K means model trained with limited 3D data.

Method	NLL	Atom Stable(%)	Mol Stable(%)
EDM	-110.7 ± 1.5	98.7 ± 0.1	82.0 ± 0.4
DiGress	-	98.1 ± 0.3	79.8 ± 5.6
MUDiff	−135.5 ± 2.1	98.8 ± 0.2	89.9 ± 1.1
MUDiff (30K+70K)	−120.6 ± 3.4	98.2 ± 0.7	84.5 ± 2.5