# Fractional Denoising for 3D Molecular Pre-training

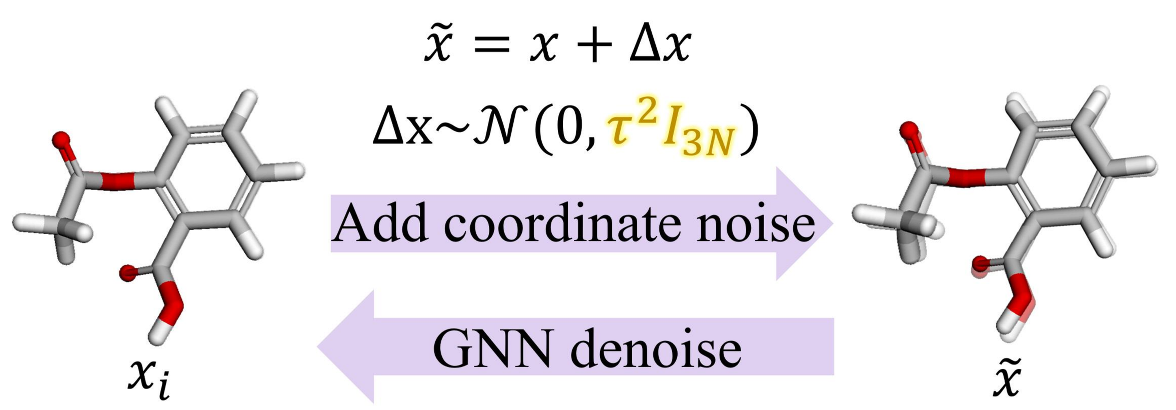
Shikun Feng*[1] , Yuyan Ni *[2] , Yanyan Lan[1], Zhi-Ming Ma[2], Wei-Ying Ma[1]

## Motivation

**Problem Domain**: Self-supervised learning for 3D Molecular representation learning

**Existing SOTA:**

Coordinate Denoising



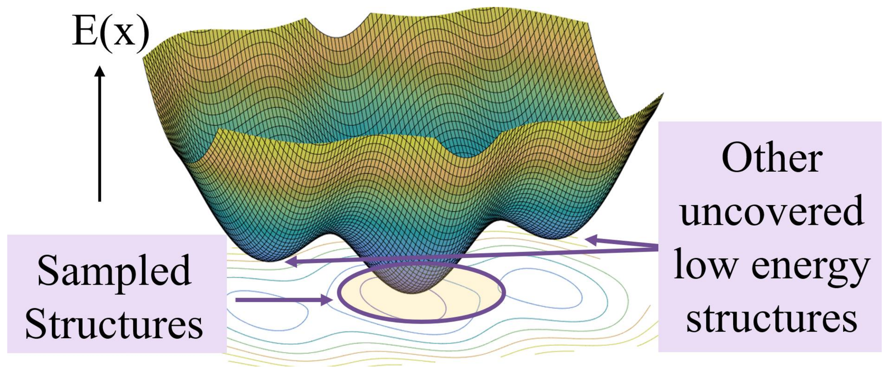

Physical Interpretation:

Equivalent to learning an approximate force field (with Boltzmann distribution and *isotropic* Gaussian assumption)

**Two challenges to Coordinate Denoising:**

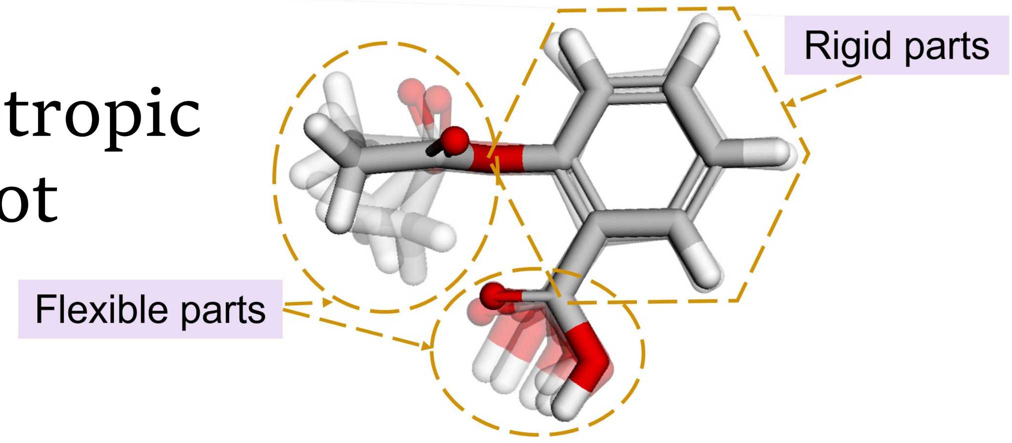
- *Low Sampling Coverage:*

  To avoid irrational structures, the noise level has to be small.




- *Isotropic Force Field:*

  Isotropic Gaussian leads to isotropic energy function, which does not match the actual molecular distribution.




## Main Contribution

- We introduce a new hybrid noise strategy and design a novel fractional denoising method (Frad) for 3D molecule pre-training.
- We use Frad to improve Noisy Nodes that fixes the non-converging issue and improve performance in fine-tuning.
- Theoretically, we prove Frad is equivalent to learning the force field with an anisotropic covariance, which captures the important characteristic of molecules. Frad also has the ability to sample more low-energy conformations.
- Empirically, we achieve a new SOTA on QM9 and MD17. Ablation study shows the effectiveness of our Frad as a pre-training strategy and our improved Noisy Nodes as a fine-tuning technique.

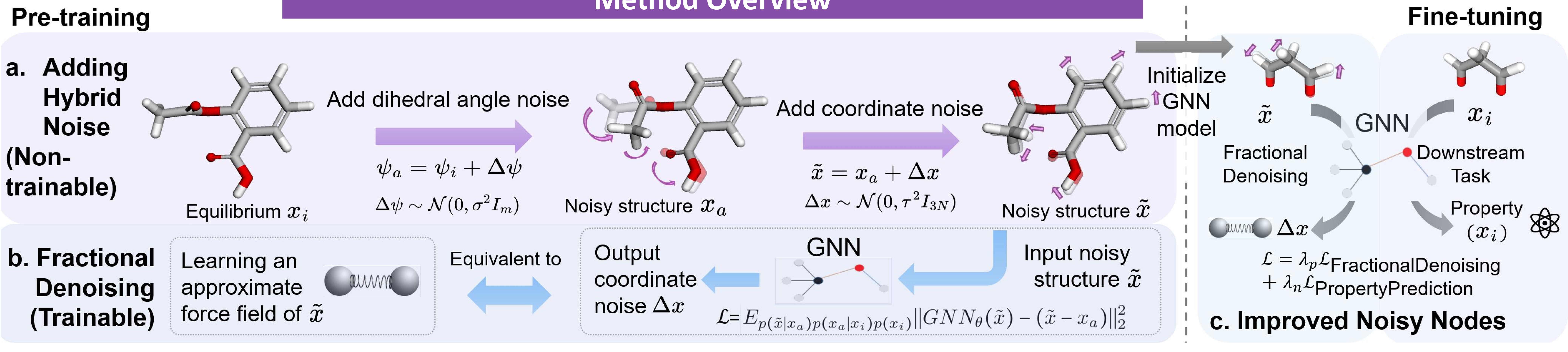
## Results

| Training set size | Models | Aspirin | Benzene | Ethanol | Malonaldehyde | Naphthalene | Salicylic Acid | Toluene | Uracil |
|---|---|---|---|---|---|---|---|---|---|
| 9500 | TorchMD-NET | 0.1216 | 0.1479 | 0.0492 | 0.0695 | 0.0390 | 0.0655 | 0.0393 | 0.0484 |
| | 3D-EMGP | 0.1560 | 0.1648 | 0.0389 | 0.0737 | 0.0829 | 0.1187 | 0.0619 | 0.0773 |
| | 3D-EMGP (TorchMD-NET) | 0.1124 | 0.1417 | 0.0445 | 0.0618 | 0.0352 | 0.0586 | 0.0385 | 0.0477 |
| | DP-TorchMD -NET($\tau = 0.04$) | 0.0920 | 0.1397 | 0.0402 | 0.0661 | 0.0544 | 0.0790 | 0.0495 | 0.0507 |
| | Frad ($\sigma = 2, \tau = 0.04$) | **0.0680** | 0.1606 | **0.0332** | **0.0427** | **0.0277** | **0.0410** | **0.0305** | **0.0323** |
| 1000 | SphereNet | 0.430 | 0.178 | 0.208 | 0.340 | 0.178 | 0.360 | 0.155 | 0.267 |
| | SchNet | 1.35 | 0.31 | 0.39 | 0.66 | 0.58 | 0.85 | 0.57 | 0.56 |
| | DimeNet | 0.499 | 0.187 | 0.230 | 0.383 | 0.215 | 0.374 | 0.216 | 0.301 |
| | SE(3)-DDM | 0.453 | **0.051** | 0.166 | 0.288 | 0.129 | 0.266 | 0.122 | 0.183 |
| 950 | PaiNN | 0.338 | 0.052* | 0.224 | 0.319 | 0.077 | 0.195 | 0.094 | 0.139 |
| | TorchMD-NET | 0.2450 | 0.2187 | 0.1067 | 0.1667 | 0.0593 | 0.1284 | 0.0644 | 0.0887 |
| | Frad ($\sigma = 2, \tau = 0.04$) | **0.2087** | 0.1994 | **0.0910** | **0.1415** | **0.0530** | **0.1081** | **0.0540** | **0.0760** |

| Models | $\mu$ (D) | $\alpha$ ($a_0^3$) | $\epsilon_{HOMO}$ (meV) | $\epsilon_{LUMO}$ (meV) | $\Delta\epsilon$ (meV) | $< R^2 >$ ($a_0^2$) | ZPVE (meV) | $U_0$ (meV) | $U$ (meV) | $H$ (meV) | $G$ (meV) | $C_v$ ($\frac{cal}{mol K}$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SchNet | 0.033 | 0.235 | 41.0 | 34.0 | 63.0 | 0.07 | 1.70 | 14.00 | 19.00 | 14.00 | 14.00 | 0.033 |
| E(n)-GNN | 0.029 | 0.071 | 29.0 | 25.0 | 48.0 | 0.11 | 1.55 | 11.00 | 12.00 | 12.00 | 12.00 | 0.031 |
| DimeNet++ | 0.030 | 0.043 | 24.6 | 19.5 | 32.6 | 0.33 | 1.21 | 6.32 | 6.28 | 6.53 | 7.56 | 0.023 |
| PaiNN | 0.012 | 0.045 | 27.6 | 20.4 | 45.7 | 0.07 | 1.28 | 5.85 | 5.83 | 5.98 | 7.35 | 0.024 |
| SphereNet | 0.027 | 0.047 | 23.6 | 18.9 | 32.3 | 0.29 | **1.120** | 6.26 | 7.33 | 6.40 | 8.00 | 0.022 |
| TorchMD-NET | 0.011 | 0.059 | 20.3 | 18.6 | 36.1 | **0.033** | 1.840 | 6.15 | 6.38 | 6.16 | 7.62 | 0.026 |
| Transformer-M | 0.037 | 0.041 | 17.5 | 16.2 | **27.4** | 0.075 | 1.18 | 9.37 | 9.41 | 9.39 | 9.63 | 0.022 |
| SE(3)-DDM | 0.015 | 0.046 | 23.5 | 19.5 | 40.2 | 0.122 | 1.31 | 6.92 | 6.99 | 7.09 | 7.65 | 0.024 |
| 3D-EMGP | 0.020 | 0.057 | 21.3 | 18.2 | 37.1 | 0.092 | 1.38 | 8.60 | 8.60 | 8.70 | 9.30 | 0.026 |
| DP-TorchMD -NET($\tau = 0.04$) | 0.012 | 0.0517 | 17.7 | 14.3 | 31.8 | 0.4496 | 1.71 | 6.57 | 6.11 | 6.45 | 6.91 | **0.020** |
| Frad ($\sigma = 2, \tau = 0.04$) | **0.010** | **0.0374** | **15.3** | **13.7** | 27.8 | 0.3419 | 1.418 | **5.33** | **5.62** | **5.55** | **6.19** | **0.020** |

New SOTAs on 9/12 tasks of QM9 and 7/8 targets of MD17 !

## Method Overview




*Equal contribution. [1] Institute for AI Industry Research (AIR), Tsinghua University, [2]Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences.

Contact us: lanyanyan@air.tsinghua.edu.cn; fsk21@mails.tsinghua.edu.cn; niyuyan17@mails.ucas.ac.cn.

References:
Zaidi, S et al. Pre-training via denoising for molecular property prediction. 2022.
Zhou, G et al. Uni-mol: A universal 3D molecular representation learning framework. 2022.
Luo, S et al. One transformer can understand both 2D & 3D molecular data. 2022.
Godwin et al. Simple GNN Regularisation for 3D Molecular Property Prediction and Beyond. 2021