FLDWM: World Model for learning approximate SO(3) equivariance for scalable and generalizable molecule generation

Anirban Ray[1,2]

[1]Raydynamics Labs
[2]Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi

1. Introduction

Existing equivariant methods combine diffusion models with equivariant graph neural networks (e.g., EDM and MIDI) to enforce SO(3) rotational equivariance. In contrast, non-equivariant methods (e.g., VoxMol and Nebula) do not model equivariance explicitly.

- **Definition:** A function f is said to be *equivariant* to the action of a group G if:
$$T_g(f(x)) = f(S_g(x)) \quad \forall\, g \in G. \tag{1}$$
where S_g and T_g are linear representations of g.
- **Limitations:** Existing equivariant models are harder to scale, slower to train, and more difficult to implement. Non-equivariant models typically rely on data augmentation, offering limited insight into how equivariance emerges.
- **Our Solution:** I propose a world-modeling approach that explicitly learns approximate equivariance for scalable and generalizable molecule generation. I further provide seed-based novel drug candidates derived from existing drugs that face growing resistance or are unaffordable.

2. Problem Formulation

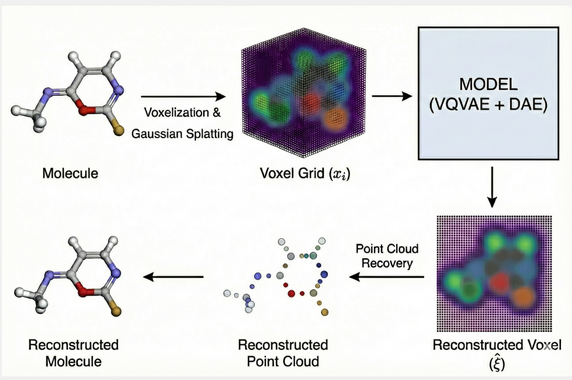

Figure 1: Voxel-based molecular representation

- I train the model using atom-coordinate point clouds from the GEOM-Drugs training set and evaluate it on the GEOM-Drugs test set and PCQM.
- Each point cloud is converted into a voxel grid.
- Number of atom channels: 8. Grid resolution: $64 \times 64 \times 64$.
- Voxel dimension: $\mathbf{V} \in \mathbb{R}^{8\times64\times64\times64}$.

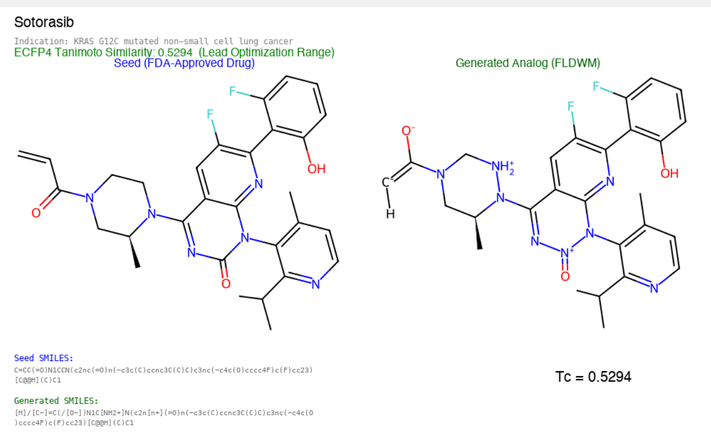

Figure 2: Seeded Sotorasib analogue generated by FLDWM

3. Methodology: FLDWM

Latent construction. The voxelizer, VQ-VAE encoder, and latent normalization define
$$z = \Phi(\mathbf{x}), \qquad \Phi = \psi_\mu \circ \mathcal{E} \circ \mathcal{V}_h.$$
For $R \sim \mu_{\text{Haar}}$ and Gaussian noise ϵ,
$$s_0 = z_{\text{noisy}} = \bar{T}_R(z) + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2 I).$$
World-model formulation. FLDWM treats the latent manifold $\mathcal{Z}$ as its state space and the SO(3) group action as its dynamics. The episode is
$$\begin{aligned} \text{initial state:} \quad & s_0 = \bar{T}_R(z) + \epsilon, \\ \text{observed condition:} \quad & o_R = R_{\text{flat}} \in \mathbb{R}^9, \\ \text{inverse action:} \quad & a_R = \bar{T}_{R^{-1}}, \\ \text{terminal state:} \quad & s_1 = f_\theta(s_0, o_R) \approx z. \end{aligned}$$
Since $R^{-1} = R^\top \in \text{SO}(3)$, the inverse action always exists. The learned transition simultaneously removes noise and moves the rotated latent back to its canonical orientation in one step.

Training objective.
$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x},R,\epsilon}\big[\| f_\theta(\bar{T}_R(\Phi(\mathbf{x})) + \epsilon, R_{\text{flat}}) - \Phi(\mathbf{x})\|^2\big]. \tag{2}$$

Theorem 1 (Approximate $SO(3)$ equivariance)**.** *Suppose f_{θ^*} achieves $\mathcal{L}(\theta^*) \le \delta^2$. Decompose the conditioned model as $f_{\theta^*}(\,\cdot\,, R_{\text{flat}}) = \bar{T}_{R^{-1}} \circ g_{\theta^*}(\,\cdot\,)$, where g_{θ^*} is the denoiser backbone and $\bar{T}_{R^{-1}}$ is the inverse latent rotation learned via FiLM conditioning on R. Then g_{θ^*} satisfies $SO(3)$-equivariance on the image of Φ:*
$$g_{\theta^*}\big(\bar{T}_R(z)\big) \;=\; \bar{T}_R\big(g_{\theta^*}(z)\big), \qquad \textit{with error } \delta + \mathcal{O}(h), \tag{3}$$
where $z := \Phi(\mathbf{x})$.

Proof. (i) The world model learns $\bar{T}_{R^{-1}}$. Write $z = \Phi(\mathbf{x})$. The loss (2) requires the full model, given the rotated latent $\bar{T}_R(z)$ and the conditioning R_{flat}, to recover the canonical code z. Since $\bar{T}_R$ is a bijection on $\mathcal{Z}$ (with inverse $\bar{T}_{R^{-1}}$, guaranteed by $R^{-1} = R^\top \in SO(3)$), the *unique* transformation that maps $\bar{T}_R(z)$ back to z for every z is $\bar{T}_{R^{-1}}$. Without learning this transformation, the loss cannot decrease. The convergence $\mathcal{L}(\theta^*) \le \delta^2$ therefore certifies that the FiLM layers, conditioned on R, have learned to apply $\bar{T}_{R^{-1}}$ to the backbone output $g_{\theta^*}(\bar{T}_R(z))$.

(ii) From convergence. Substituting the decomposition $f_{\theta^*}(\,\cdot\,, R_{\text{flat}}) = \bar{T}_{R^{-1}} \circ g_{\theta^*}(\,\cdot\,)$ into the loss optimality condition:
$$\bar{T}_{R^{-1}}\big(g_{\theta^*}(\bar{T}_R(z))\big) \;=\; z, \qquad \text{residual} \le \delta. \tag{4}$$
(iii) Identity case. Setting $R = I$ in Eq. (4): $\bar{T}_I(g_{\theta^*}(z)) = z$. Since $\bar{T}_I = \text{Id}$, this gives
$$g_{\theta^*}(z) = z. \tag{5}$$
(iv) Equivariance. Since $\bar{T}_{R^{-1}} = \bar{T}_R^{-1}$ (group inverse property), left-multiply both sides of Eq. (4) by $\bar{T}_R$:
$$g_{\theta^*}\big(\bar{T}_R(z)\big) \;=\; \bar{T}_R(z). \tag{6}$$
Substituting Eq. (5) into the right-hand side:
$$g_{\theta^*}\big(\bar{T}_R(z)\big) \;=\; \bar{T}_R\big(g_{\theta^*}(z)\big), \tag{7}$$
which is the $SO(3)$-equivariance condition (1): the denoiser backbone commutes with the latent rotation. Since μ_{Haar} has full support on $SO(3)$, this holds for *all* rotations. □

4. World Model Conditioning

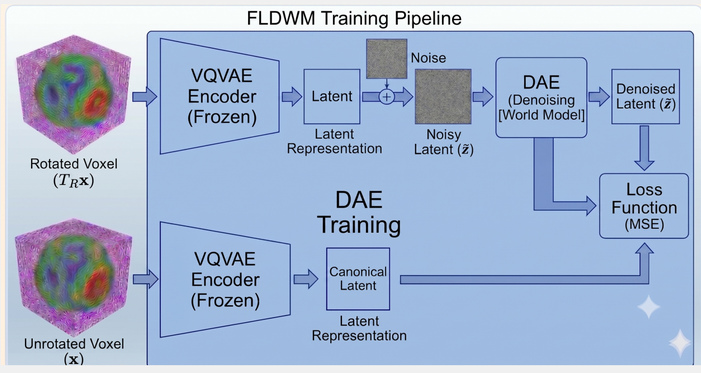

Figure 3: FLDWM training pipeline

The flattened rotation matrix is embedded by a five-layer MLP:
$$c = \text{MLP}(R_{\text{flat}}) \in \mathbb{R}^{3072}.$$
At each residual block, unbounded linear FiLM modulation is applied after group normalization:
$$\begin{aligned} [\gamma, \beta] &= \text{Linear}(c), \\ \bar{F} &= \text{GroupNorm}(F) \odot (1+\gamma) + \beta. \end{aligned}$$
Conditioning on the physical rotation enables the decomposition
$$f_\theta(\cdot, R_{\text{flat}}) = \bar{T}_{R^{-1}} \circ g_\theta(\cdot),$$
where g_θ denoises the latent and $\bar{T}_{R^{-1}}$ maps it back to its canonical orientation.

5. Seeded Generation Results

Table 1: Seeded generation on GEOM. Results are averaged over five repeats.

Model (WJS)	Tan. sim. (%)	Stable sanit. (%)	Stable atom (%)	Valid (%)
0 (data)	100.00	99.90	99.90	99.80
VoxMol (10)	$71.44_{\pm1.82}$	$85.71_{\pm0.36}$	$99.36_{\pm0.04}$	$89.86_{\pm0.28}$
NEBULA (10)	$\mathbf{88.47}_{\pm0.18}$	$85.73_{\pm0.12}$	$\mathbf{99.46}_{\pm0.01}$	$90.29_{\pm0.11}$
FLDWM (10)	$88.39_{\pm0.00}$	$\mathbf{87.29}_{\pm0.09}$	$98.20_{\pm0.03}$	$\mathbf{96.62}_{\pm0.07}$
NEBULA (20)	$84.30_{\pm0.46}$	$\mathbf{85.46}_{\pm0.31}$	$\mathbf{99.42}_{\pm0.01}$	$89.82_{\pm0.31}$
FLDWM (20)	$\mathbf{86.04}_{\pm0.10}$	$84.85_{\pm0.11}$	$97.86_{\pm0.01}$	$\mathbf{96.43}_{\pm0.05}$

Table 2: Cross-dataset seeded generation on PCQM. Results are averaged over five repeats.

Model (WJS)	Tan. sim. (%)	Stable sanit. (%)	Stable atom (%)	Valid (%)
0 (data)	100.00	95.51	99.62	99.40
VoxMol (5)	$10.95_{\pm1.85}$	$92.45_{\pm0.29}$	$95.60_{\pm0.07}$	$97.11_{\pm0.17}$
VoxMol (10)	$9.74_{\pm0.62}$	$92.08_{\pm0.83}$	$96.89_{\pm0.11}$	$96.49_{\pm0.63}$
VoxMol (50)	$9.75_{\pm0.38}$	$90.73_{\pm0.54}$	$98.44_{\pm0.07}$	$95.17_{\pm0.30}$
VoxMol (100)	$9.81_{\pm0.21}$	$89.93_{\pm0.47}$	$98.62_{\pm0.04}$	$94.81_{\pm0.43}$
VoxMol (200)	$9.86_{\pm0.22}$	$90.44_{\pm0.85}$	$98.77_{\pm0.07}$	$94.68_{\pm0.56}$
NEBULA (10)	$35.10_{\pm0.57}$	$95.18_{\pm0.26}$	$75.64_{\pm0.13}$	$98.86_{\pm0.23}$
NEBULA (20)	$32.19_{\pm0.38}$	$95.44_{\pm0.27}$	$76.43_{\pm0.18}$	$98.80_{\pm0.13}$
NEBULA (50)	$25.62_{\pm0.26}$	$95.64_{\pm0.26}$	$78.35_{\pm0.18}$	$98.36_{\pm0.12}$
NEBULA (100)	$18.38_{\pm0.27}$	$\mathbf{96.16}_{\pm0.19}$	$80.04_{\pm0.20}$	$97.88_{\pm0.35}$
NEBULA (200)	$12.58_{\pm0.09}$	$95.72_{\pm0.49}$	$79.77_{\pm0.51}$	$96.72_{\pm0.26}$
FLDWM (10)	$\mathbf{88.62}_{\pm0.07}$	$95.94_{\pm0.08}$	$\mathbf{99.63}_{\pm0.01}$	$\mathbf{99.95}_{\pm0.03}$
FLDWM (20)	$86.69_{\pm0.12}$	$93.96_{\pm0.13}$	$99.44_{\pm0.02}$	$99.84_{\pm0.02}$
FLDWM (50)	$70.90_{\pm0.31}$	$79.26_{\pm0.25}$	$97.96_{\pm0.05}$	$98.56_{\pm0.18}$
FLDWM (100)	$41.04_{\pm0.16}$	$57.52_{\pm0.30}$	$94.96_{\pm0.07}$	$96.58_{\pm0.07}$
FLDWM (200)	$18.41_{\pm0.08}$	$43.41_{\pm0.11}$	$90.47_{\pm0.04}$	$94.24_{\pm0.21}$

6. Conclusion

FLDWM learns approximate equivariance through latent rotation inversion, avoiding equivariant architectures. It achieves state-of-the-art performance on GEOM-Drugs, zero-shot generalization to PCQM, and seed-conditioned generation of alternative drug candidates. Future work will extend the framework to macromolecules and conditional drug design.

Acknowledgements

Anirban Ray acknowledges the financial support received from the Ministry of Education (MoE) Government of India(GoI), in the form of the Prime Minister's Research Fellowship (PMRF).

LaTeX TikZposter