Navigating the Design Space of Equivariant Diffusion-Based Generative Models for De Novo 3D Molecule Generation

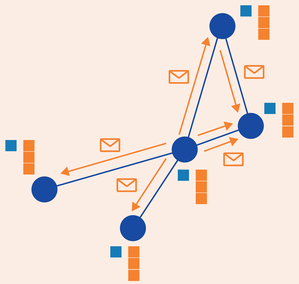
TUAN LE*[1,2], JULIAN CREMER*[1,3], FRANK NOÉ[2,4], DJORK-ARNÉ CLEVERT[1] AND KRISTOF T. SCHÜTT[1]

[1] Pfizer Research & Development
[2] Freie Universität Berlin
[3] Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB)
[4] Microsoft Research

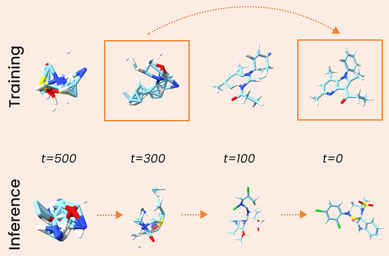
*Equal contribution

Overview

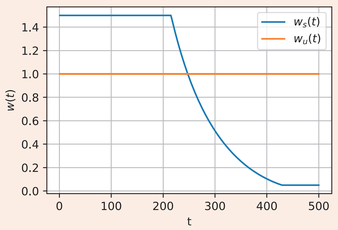
EQGAT-DIFF



TRAINING AND INFERENCE



TIMESTEP-DEPENDENT LOSS WEIGHTING



$$L_{t,\epsilon} = w(t)||\epsilon_t - \hat{\epsilon}_\theta(x_t, t)||^2$$

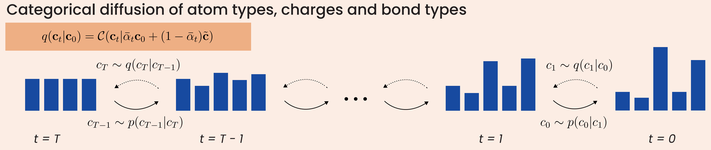
$$L_{t,x_0} = w(t) \cdot l_d(x_0, \hat{x}_\theta(x_t, t); \lambda_m)$$

$$w_s(t) = \max(0.05,\ \min(1.5,\ \mathrm{SNR}(t)))$$

CATEGORICAL AND CONTINUOUS DIFFUSION

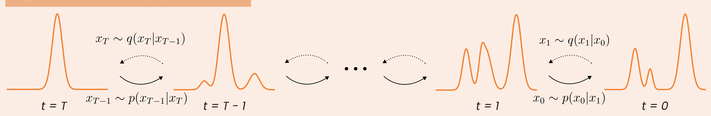
Categorical diffusion of atom types, charges and bond types

$q(\mathbf{c}_t|\mathbf{c}_0) = \mathcal{C}(\mathbf{c}_t|\bar{\alpha}_t\mathbf{c}_0 + (1-\bar{\alpha}_t)\tilde{\mathbf{c}})$

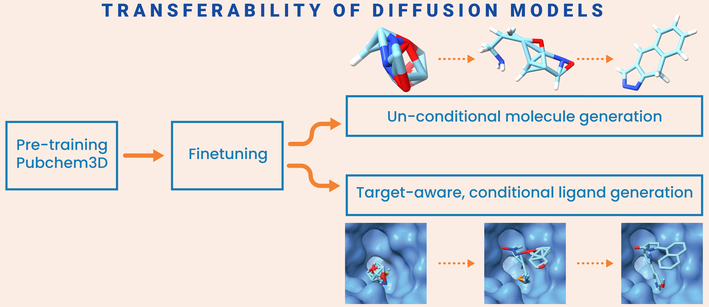


Continuous diffusion of atom positions

$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t|\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$



TRANSFERABILITY OF DIFFUSION MODELS



Experiments

MOTIVATION AND BACKGROUND

We propose EQGAT-diff, a generative model p_θ for 3D molecules $x = (\mathbf{H}, \mathbf{X}, \mathbf{E})$ with $\mathbf{H} \in \{0,1\}^{N\times K_a}$, $\mathbf{E} \in \{0,1\}^{N\times N\times K_b}$, $\mathbf{X} \in \mathbb{R}^{N\times 3}$, operating on continuous and discrete variables invariant to the permutation of atoms and roto-translations, i.e., $p_\theta(x) = p_\theta(g.x)$

Research Questions:

1. Prior works using equivariant autoregressive or diffusion models for molecule design showed suboptimal performance on more complex data distributions, like GEOM-Drugs. Are those fundamental architectural limitations or can we improve the performance with better design choices and which are those?
2. Should we respect the data modalities and apply data-dependent noising and de-noising steps? That is, for continuous data, Gaussian diffusion and for discrete data, categorical diffusion.
3. For continuous variables, should the noise or the data parameterization be used, i.e., $\hat{\epsilon}$ or $\hat{x}_0$?
4. Can we increase performance and generality by pre-training the diffusion model on large corpora of molecular data derived at a low level of theory? Can this be transferred to structure-based drug design?

TIMESTEP-DEPENDENT LOSS WEIGHTING

We hypothesize that denoising requires high accuracy close to the data distribution for generating valid molecules, while errors close to the noise distribution are negligible. We propose using the time-dependent weighting:

$$w_s(t) = \max(0.05,\ \min(1.5,\ \mathrm{SNR}(t))). \tag{1}$$

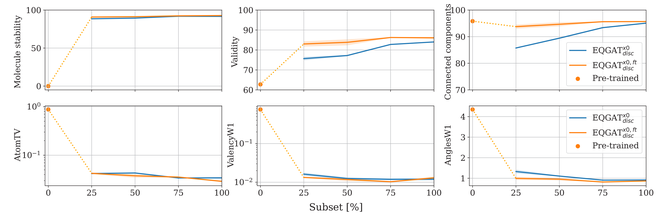
	QM9			GEOM-Drugs		
Weighting	Mol. Stability ↑	Validity ↑	Connect. Comp. ↑	Mol. Stability ↑	Validity ↑	Connect. Comp. ↑
w_u	$97.39_{\pm 0.23}$	$97.99_{\pm 0.20}$	$99.70_{\pm 0.03}$	$87.59_{\pm 0.19}$	$71.44_{\pm 0.22}$	$86.57_{\pm 0.33}$
$w_s(t)$	$98.68_{\pm 0.11}$	$98.96_{\pm 0.07}$	$99.94_{\pm 0.03}$	$91.60_{\pm 0.14}$	$84.02_{\pm 0.19}$	$95.08_{\pm 0.12}$

STATE-OF-THE-ART MOLECULE GENERATION

Dataset	GEOM-Drugs					
Model	EQGAT$^{x0}_{disc}$	EQGAT$^{x0,ft}_{disc}$	EQGAT$^{x0,af}_{disc}$	EQGAT$^{x0,af,ft}_{disc}$	EDM	MiDi
Mol. Stab. ↑	$93.11_{\pm 0.31}$	$93.92_{\pm 0.13}$	$94.51_{\pm 0.18}$	$95.01_{\pm 0.37}$	40.3	$89.7_{\pm 0.60}$
Atom. Stab ↑	$99.79_{\pm 0.01}$	$99.81_{\pm 0.01}$	$99.83_{\pm 0.01}$	$99.84_{\pm 0.00}$	97.8	$99.7_{\pm 0.01}$
Validity ↑	$85.86_{\pm 0.33}$	$88.04_{\pm 0.17}$	$87.89_{\pm 0.31}$	$88.42_{\pm 0.26}$	87.8	$70.5_{\pm 0.41}$
Connect. Comp. ↑	$96.32_{\pm 0.25}$	$96.57_{\pm 0.18}$	$96.36_{\pm 0.25}$	$96.71_{\pm 0.20}$	41.4	$88.76_{\pm 0.55}$
Novelty ↑	$99.82_{\pm 0.05}$	$99.84_{\pm 0.02}$	$99.82_{\pm 0.05}$	$99.82_{\pm 0.03}$	100.00	100.00$_{\pm 0.00}$
Diversity ↑	$89.03_{\pm 0.03}$	$89.05_{\pm 0.05}$	$88.98_{\pm 0.02}$	$88.96_{\pm 0.01}$	-	-
KL Divergence ↑	$87.66_{\pm 0.31}$	$87.58_{\pm 0.56}$	$88.38_{\pm 0.25}$	$87.62_{\pm 0.19}$	-	-
Train Similarity ↓	$0.114_{\pm 0.0}$	$0.113_{\pm 0.0}$	$0.114_{\pm 0.0}$	$0.114_{\pm 0.0}$	-	-
AtomsTV $[10^{-2}]$ ↓	$3.02_{\pm 0.08}$	$3.02_{\pm 0.10}$	$2.88_{\pm 0.10}$	$2.91_{\pm 0.10}$	21.2	$5.11_{\pm 0.19}$
BondsTV $[10^{-2}]$ ↓	$2.44_{\pm 0.01}$	$2.40_{\pm 0.00}$	$2.42_{\pm 0.00}$	$2.40_{\pm 0.00}$	4.8	$2.44_{\pm 0.00}$
ValencyW$_1$ $[10^{-2}]$ ↓	$1.18_{\pm 0.09}$	$1.20_{\pm 0.00}$	$0.85_{\pm 0.12}$	$0.90_{\pm 0.10}$	28.5	$2.48_{\pm 0.52}$
BondLenghtsW$_1$ $[10^{-2}]$ ↓	$0.56_{\pm 0.38}$	$0.10_{\pm 0.00}$	$0.50_{\pm 0.51}$	$0.20_{\pm 0.10}$	0.2	$0.2_{\pm 0.10}$
BondAnglesW$_1$ ↓	$0.83_{\pm 0.03}$	$0.79_{\pm 0.02}$	$0.65_{\pm 0.01}$	$0.62_{\pm 0.01}$	6.23	$1.73_{\pm 0.32}$

TRANSFERABILITY OF DIFFUSION MODELS

Diffusion models are effective in learning vast data distributions but require large datasets to be trained on. We explore the effect of pre-training on PubChem3d for generalization on the Geom-Drugs dataeset. We observe that the fine-tuned model performs better than models trained from scratch, even if the fine-tuning was done on a subset of data.



STRUCTURE-BASED LIGAND DESIGN

Here, we train a conditional diffusion model $p_\theta(x|P)$ where P is a protein pocket. Particularly, a pre-trained model that learned on a vast, yet unconditional chemical space, generalizes better when fine-tuned on CrossDocked2020 than a model trained from scratch.

Model	Validity ↑	Connect. Comp. ↑	BondLengths W1 $[10^{-2}]$ ↓	BondAngles W1 ↓
EQGAT$^{x0}_{disc}(w_u)$	$85.51_{\pm 0.09}$	$95.15_{\pm 0.14}$	$0.20_{\pm 0.0}$	$4.37_{\pm 0.20}$
EQGAT$^{x0}_{disc}(w_s(t))$	$89.62_{\pm 0.08}$	$97.65_{\pm 0.11}$	$0.12_{\pm 0.0}$	$2.12_{\pm 0.26}$
EQGAT$^{x0,ft}_{disc}(w_s(t))$	**95.65**$_{\pm 0.12}$	**99.66**$_{\pm 0.10}$	**0.11**$_{\pm 0.0}$	**1.55**$_{\pm 0.21}$

Model	Vina (All) ↓	Vina (Top-10%) ↓	QED ↑	SA ↑	Lipinski ↑	Diversity ↑
EQGAT$^{x0,ft}_{disc}(w_s(t))$	**-7.423**$_{\pm 2.33}$	-9.571$_{\pm 2.14}$	**0.522**$_{\pm 0.18}$	**0.697**$_{\pm 0.20}$	4.66$_{\pm 0.72}$	**0.742**$_{\pm 0.07}$
TargetDiff	-7.318$_{\pm 2.47}$	**-9.669**$_{\pm 2.55}$	0.483$_{\pm 0.20}$	0.584$_{\pm 0.13}$	4.594$_{\pm 0.83}$	0.718$_{\pm 0.09}$
DiffSBDD-cond	-6.950$_{\pm 2.06}$	-9.120$_{\pm 2.16}$	0.469$_{\pm 0.21}$	0.578$_{\pm 0.13}$	4.562$_{\pm 0.89}$	0.728$_{\pm 0.07}$