ICLR

Carnegie Mellon University

Chemistry-Inspired Diffusion with Non-Differentiable Guidance

Yuchen Shen* Chenhao Zhang* Sijie Fu* Chenghui Zhou

Newell Washburn† Barnabás Póczos†

*: equal contribution
†: corresponding author

ChemGuide

- Research Question

  Can quantum chemistry[1, 2] help improve stability in (3D) molecule generation?

- Motivation

  - labels are in general hard to acquire

[1] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. Journal of chemical theory and computation, 15 (3):1652–1671, 2019.
[2] M. J. Frisch et al., Gaussian˜16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.

ChemGuide

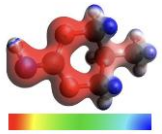
- Research Question

  Can quantum chemistry[1, 2] help improve stability in (3D) molecule generation?

- Motivation

  - labels are in general hard to acquire

    e.g., on QM9[3], to get the label **HOMO** (highest occupied molecular orbital):

DFT ⟶  ⟶ slow, ~10 min

[1] Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. Journal of chemical theory and computation, 15 (3):1652–1671, 2019.
[2] M. J. Frisch et al., Gaussian˜16 Revision C.01, 2016. Gaussian Inc. Wallingford CT.
[3] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. 1(1):140022. ISSN 2052-4463. doi: 10.1038/sdata.2014.22. URL https://www.nature.com/articles/ sdata201422.

ChemGuide

- Motivation
  - labels are in general hard to acquire
  - unconditional diffusion models can be made conditional

ChemGuide

- Motivation
  - labels are in general hard to acquire
  - unconditional diffusion models can be made conditional w. $\nabla_x \log p(y|x)$

    e.g., with a classifier $p(y|x)$, diffusion models can sample conditionally[4]

Algorithm 1 Classifier guided diffusion sampling, given a diffusion model $(\mu_\theta(x_t), \Sigma_\theta(x_t))$, classifier $p_\phi(y|x_t)$, and gradient scale s.

Input: class label y, gradient scale s
$x_T \leftarrow$ sample from $\mathcal{N}(0, \mathbf{I})$
for all t from T to 1 **do**
  $\mu, \Sigma \leftarrow \mu_\theta(x_t), \Sigma_\theta(x_t)$
  $x_{t-1} \leftarrow$ sample from $\mathcal{N}(\mu + s\Sigma \nabla_{x_t} \log p_\phi(y|x_t), \Sigma)$
end for
return x_0

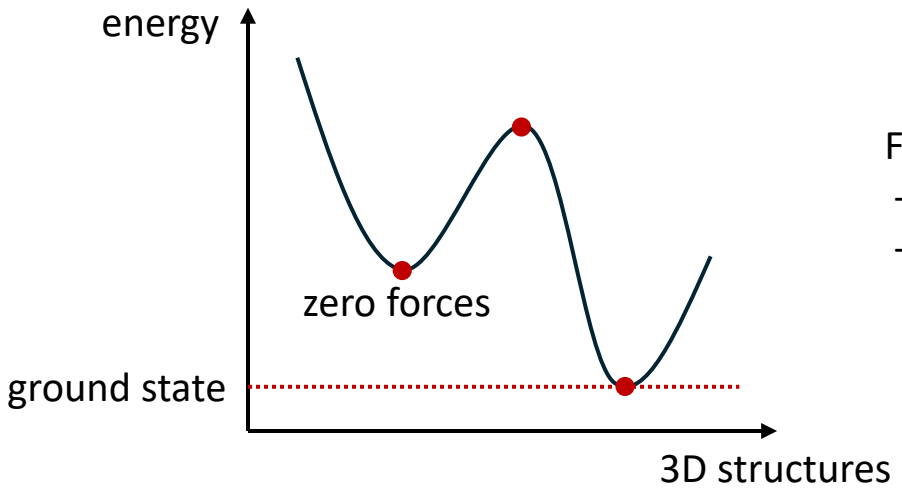
[4] Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis, 2021.

ChemGuide

- Motivation
  - labels are in general hard to acquire
  - unconditional diffusion models can be made conditional w. $\nabla_x \log p(y|x)$
  - stability = ground state energy ($\approx$ forces [- 1st derivative of energy]=0)

ChemGuide

- Motivation
  - labels are in general hard to acquire
  - unconditional diffusion models can be made conditional w. $\nabla_x \log p(y|x)$
  - stability = ground state energy ($\approx$ forces [- 1st derivative of energy]=0)



For example, for the water molecule H_2O:
- stable: V-shaped H-O-H geometry of 104.5 degrees
- unstable: linear H-O-H geometry of 180 degrees

ChemGuide

- Motivation
  - labels are in general hard to acquire
  - unconditional diffusion models can be made conditional w. $\nabla_x \log p(y|x)$
  - stability is **better predicted by quantum chemistry** than surrogate ML models*

  however,

*: the surrogate ML models are usually trained with labels produced by quantum chemistry calculations

ChemGuide

- Motivation
  - labels are in general hard to acquire
  - unconditional diffusion models can be made conditional w. $\nabla_x \log p(y|x)$
  - stability is better predicted by quantum chemistry than surrogate ML models*

  however,

  quantum chemistry as $p(y|x)$: a **non-differentiable** oracle

  surrogate ML models as $p(y|x)$: a **"backpropagate"-able** oracle

  How to introduce guidance from quantum chemistry (accurate but non-differentiable) to diffusion models?

*: the surrogate ML models are usually trained with labels produced by quantum chemistry calculations

ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization

$$\hat{\nabla}_{\boldsymbol{z}_t} \log p_\phi(y|\boldsymbol{z}_t) \propto -\nabla_{\mathcal{F}(\boldsymbol{z}_t)} \mathcal{L}(y, \mathcal{F}(\boldsymbol{z}_t)) \frac{\mathcal{F}([\boldsymbol{z}_{\mathbf{x},t} + \zeta \mathbf{U}, \boldsymbol{z}_{\mathbf{h},t}]) - \mathcal{F}([\mathbf{z}_{\mathbf{x},t} - \zeta \mathbf{U}, \boldsymbol{z}_{\mathbf{h},t}])}{2\zeta} \mathbf{U}$$

ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization

$$\hat{\nabla}_{\boldsymbol{z}_t} \log p_\phi(y|\boldsymbol{z}_t) \propto -\nabla_{\mathcal{F}(\boldsymbol{z}_t)} \mathcal{L}(y, \mathcal{F}(\boldsymbol{z}_t)) \frac{\mathcal{F}([\boldsymbol{z}_{\mathbf{x},t} + \zeta \mathbf{U}, \boldsymbol{z}_{\mathbf{h},t}]) - \mathcal{F}([\mathbf{z}_{\mathbf{x},t} - \zeta \mathbf{U}, \boldsymbol{z}_{\mathbf{h},t}])}{2\zeta} \mathbf{U}$$

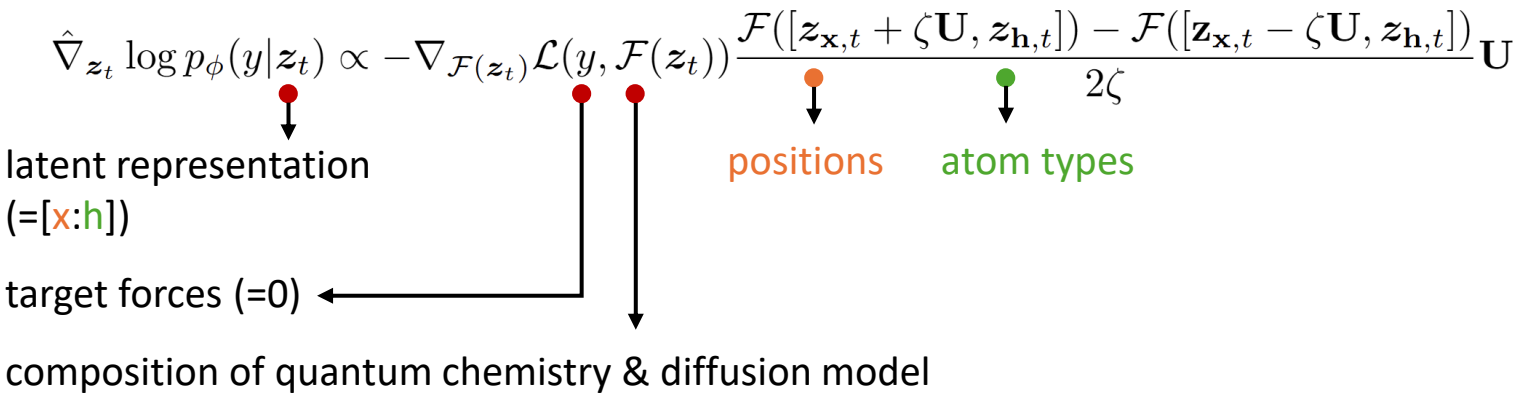
latent representation (=[x:h])

positions

atom types

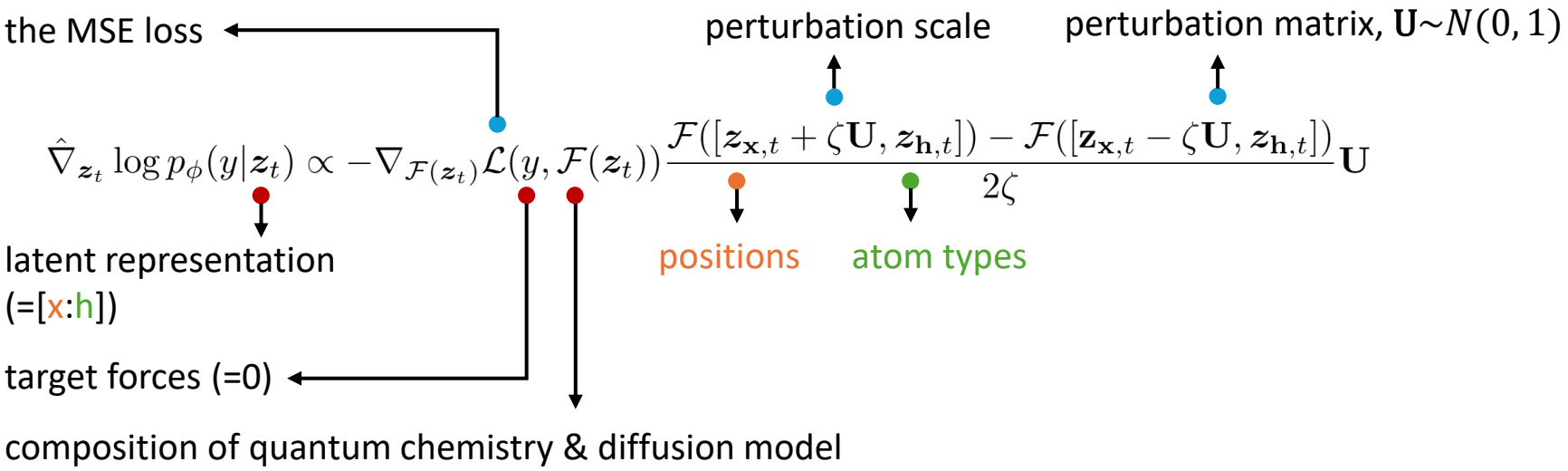
ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization



ChemGuide

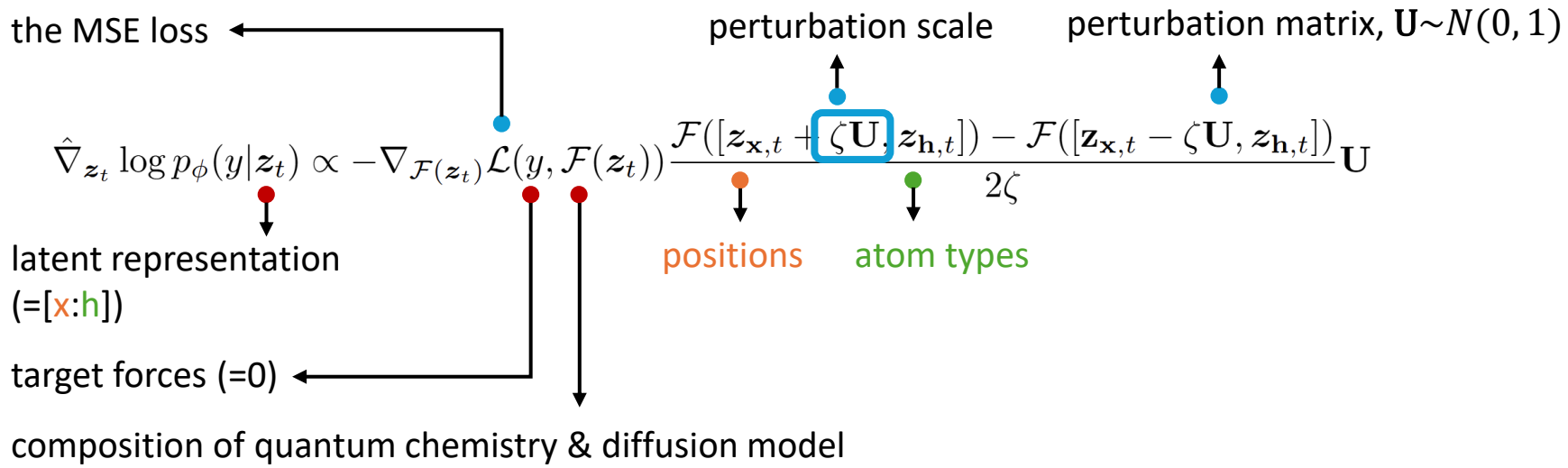
- Method
  - estimating guidance w. zeroth-order optimization



ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization

use **SPSA**[5, 6] to estimate gradient

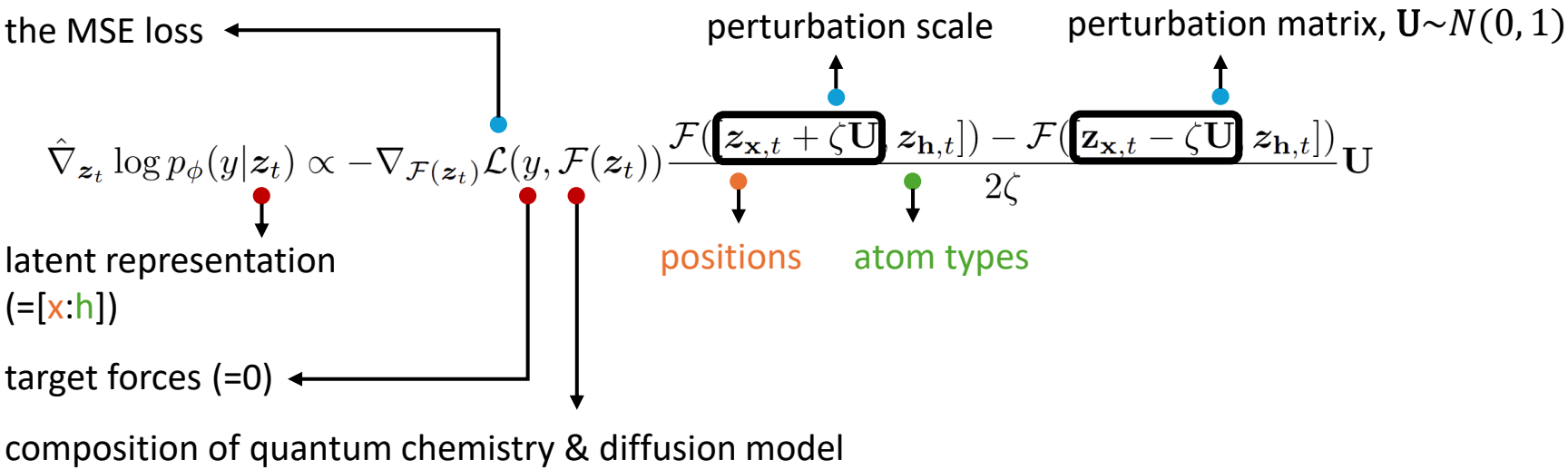


[5] James C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Transactions on Automatic Control, 37:332–341, 1992. URL https://api.semanticscholar.org/CorpusID:122365276.
[6] Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions.Foundations of Computational Mathematics, 17(2):527–566, 2017.

ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization

use **SPSA**[5, 6] to estimate gradient, while respecting the **zero-mean** requirement[7] (i.e., $\mathbb{E}[\mathbf{U}] = 0$)



[5] James C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Transactions on Automatic Control, 37:332–341, 1992. URL https://api.semanticscholar.org/CorpusID:122365276.
[6] Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions.Foundations of Computational Mathematics, 17(2):527–566, 2017.
[7] Minkai Xu, Alexander S Powers, Ron O Dror, Stefano Ermon, and Jure Leskovec. Geometric latentdiffusion models for 3d molecule generation. In International Conference on Machine Learning,pp. 38592–38610. PMLR, 2023.

ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization
  - stable & property-guided generation w. bilevel optimization

$$\begin{aligned} \min_{\boldsymbol{z}_{x,t}} \quad & \mathcal{F}([\boldsymbol{z}_{x,t}, \boldsymbol{z}^*_{h,t}]) \\ \text{s.t.} \quad & [\boldsymbol{z}^*_{x,t}, \mathbf{z}^*_{h,t}] = \underset{[\boldsymbol{z}_{x,t}, \boldsymbol{z}_{h,t}]}{\arg\min} \| f_\eta([\boldsymbol{z}_{x,t}, \boldsymbol{z}_{h,t}]) - y \|^2 \end{aligned}$$

ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization
  - stable & property-guided generation w. bilevel optimization

$$\min_{\boldsymbol{z}_{x,t}} \quad \mathcal{F}([\boldsymbol{z}_{x,t}, \boldsymbol{z}^*_{h,t}])$$

$$\text{s.t.} \quad [\boldsymbol{z}^*_{x,t}, \mathbf{z}^*_{h,t}] = \arg\min_{[\boldsymbol{z}_{x,t}, \boldsymbol{z}_{h,t}]} \| f_\eta([\boldsymbol{z}_{x,t}, \boldsymbol{z}_{h,t}]) - y \|^2$$

property predictor

target property

ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization
  - stable & property-guided generation w. bilevel optimization

$$\min_{\boldsymbol{z}_{x,t}} \quad \mathcal{F}([\boldsymbol{z}_{x,t}, \boldsymbol{z}^*_{h,t}])$$

$$\text{s.t.} \quad [\boldsymbol{z}^*_{x,t}, \mathbf{z}^*_{h,t}] = \arg\min_{[\boldsymbol{z}_{x,t}, \boldsymbol{z}_{h,t}]} \|f_\eta([\boldsymbol{z}_{x,t}, \boldsymbol{z}_{h,t}]) - y\|^2$$

(property-)optimized positions

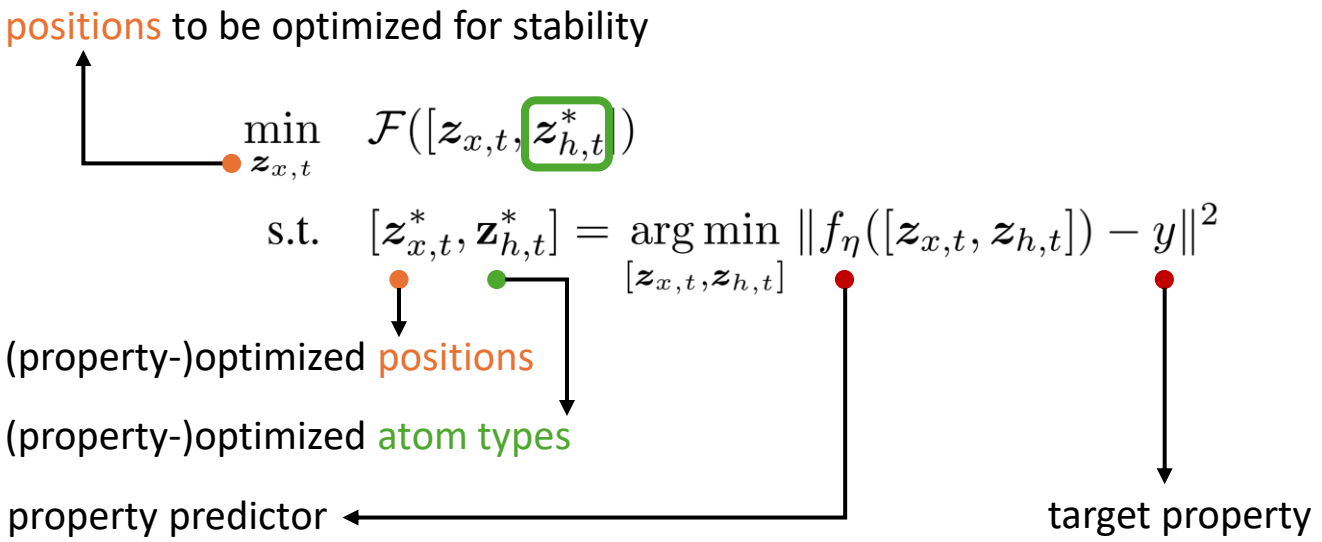
(property-)optimized atom types

property predictor

target property

ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization
  - stable & property-guided generation w. bilevel optimization



ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization
  - stable & property-guided (**inner**) generation w. bilevel optimization

$$\min_{\boldsymbol{z}_{x,t}} \quad \mathcal{F}([\boldsymbol{z}_{x,t}, \boldsymbol{z}^*_{h,t}])$$

inner objective $$\text{s.t.} \quad [\boldsymbol{z}^*_{x,t}, \mathbf{z}^*_{h,t}] = \underset{[\boldsymbol{z}_{x,t}, \boldsymbol{z}_{h,t}]}{\arg\min} \|f_\eta([\boldsymbol{z}_{x,t}, \boldsymbol{z}_{h,t}]) - y\|^2$$ **existing** guidance methods

ChemGuide

- Method
  - estimating guidance w. zeroth-order optimization
  - stable (**outer**) & property-guided generation w. bilevel optimization

outer objective $\min_{\boldsymbol{z}_{x,t}} \quad \mathcal{F}([\boldsymbol{z}_{x,t}, \boldsymbol{z}^*_{h,t}])$ **ChemGuide**

inner objective $\text{s.t.} \quad [\boldsymbol{z}^*_{x,t}, \mathbf{z}^*_{h,t}] = \underset{[\boldsymbol{z}_{x,t}, \boldsymbol{z}_{h,t}]}{\arg\min} \|f_\eta([\boldsymbol{z}_{x,t}, \boldsymbol{z}_{h,t}]) - y\|^2$ existing guidance methods

ChemGuide

- Result
  - dataset, property & model*

*: for the detailed introduction & citations of the baselines, please refer to our paper

ChemGuide

- Result
  - **dataset**, property & model*

    QM9 (~134K molecules) & GEOM (~450K molecules, 37M conformations)

*: for the detailed introduction & citations of the baselines, please refer to our paper

ChemGuide

- Result

  - dataset, **property** & model*

    QM9 (~134K molecules) & GEOM (~450K molecules, 37M conformations)

    - the norm of static polarizability (**α**, $Bohr^3$),
    - the norm of dipole moment (**μ**, Debye),
    - heat capacity at room temperature (**Cv** , cal/(mol·K)),
    - the energy of the electron in the highest occupied molecular orbital(**ϵHOMO**, eV),
    - the energy of the electron in the lowest unoccupied molecular orbital (**ϵLUMO**, eV),
    - the HOMO-LUMO energy gap (**Δϵ**, eV).

*: for the detailed introduction & citations of the baselines, please refer to our paper

ChemGuide

- Result
  - dataset, property & **model***

    QM9 (~134K molecules) & GEOM (~450K molecules, 37M conformations)

    α, μ, Cv , ϵHOMO, ϵLUMO, Δϵ

    EDM & GeoLDM (trained unconditionally & conditionally [C-xxx])
  - guidance oracle

    xTB (faster compared with DFT, but less accurate)

*: for the detailed introduction & citations of the baselines, please refer to our paper

ChemGuide

- Result

  - metric*

    • **force RMS** (Eh/Bohr): the root mean square (RMS) of the forces on the atoms of a molecule; it is the lower the better.

    • **energy above ground state** (Eh): it is the difference between the energy before and after using xTB or DFT to optimize the geometry of the generated molecules, and a smaller energy above ground state indicates the geometry is closer to the stable ground state of the molecule.

    • **validity**: it measures how many percent of the generated molecules are valid, and is calculated by either xTB or DFT.

    • **uniqueness**: it is the percentage of unique molecules among the generated molecule

*: for the detailed introduction & citations of the baselines, please refer to our paper

ChemGuide

- Result
  - ChemGuide improves stability

    on QM9 (measured with GFN2-xTB level of theory)

Table 1: Metrics of 500 generated molecules from *QM9* using GeoLDM with non-differentiable oracle (i.e. xTB) guidance. * and **bold** denote the overall best result and our best result, respectively. Percentage changes between our results and GeoLDM are shown in parentheses.

Metric	Guidance Scale					Baseline	
	0.0001	0.001	0.01	0.1	1.0	EDM	GeoLDM
Force RMS (Eh/Bohr)	**0.0104*** (-6.76%↓)	0.0104	0.0107	0.0108	0.0125	0.0114	0.0111
Validity	**91.40%*** (1.60%↑)	91.20%	91.20%	90.00%	89.40%	86.60%	89.80%
Uniqueness	**100.00%***	**100.00%***	**100.00%***	**100.00%***	**100.00%***	100.00%*	100.00%*
Atom Stability	99.02%	98.97%	98.98%	**99.03%*** (0.10%↑)	98.67%	98.53%	98.93%
Molecule Stability	90.60%	90.40%	90.20%	**91.00%*** (2.20%↑)	87.80%	83.60%	88.80%
Energy above Ground State (Eh)	**0.0042*** (-15.78%↓)	0.0042	0.0045	0.0050	0.0061	0.0072	0.0050

ChemGuide

- Result
  - ChemGuide improves stability

    on QM9 (measured with DFT level of theory)

Table 3: Metrics of 500 generated molecules from *QM9* using GeoLDM with CHEMGUIDE using scale=0.0001, calculated at DFT/B3LYP/6-31G(2df,p) level of theory, more precise but computationally expensive than GFN2-xTB. *Strict validity* is defined as generated geometries within 50 kcal/mol (0.07968 Eh) of the optimized geometries. * and **bold** denote the overall best result and our best result. Percentage changes between our results and GeoLDM are shown in parentheses.

Metric	Ours	EDM	GeoLDM
Validity	**96.2%*** (2.6%↑)	89.0%	93.6%
Strict Validity	**95.6%*** (2.4%↑)	88.4%	93.2%
Energy above Ground State (Eh)	**0.004051*** (-9.6%↓)	0.006314	0.004480
Energy above GS (kcal/mol)	**2.542*** (-9.6%↓)	3.962	2.811

with a low-precision oracle (xTB), ChemGuide achieves improvements measured with theory (DFT) of higher precision

ChemGuide

- Result
  - ChemGuide improves stability

    on QM9 & GEOM (distribution plots of force RMS / energy change)

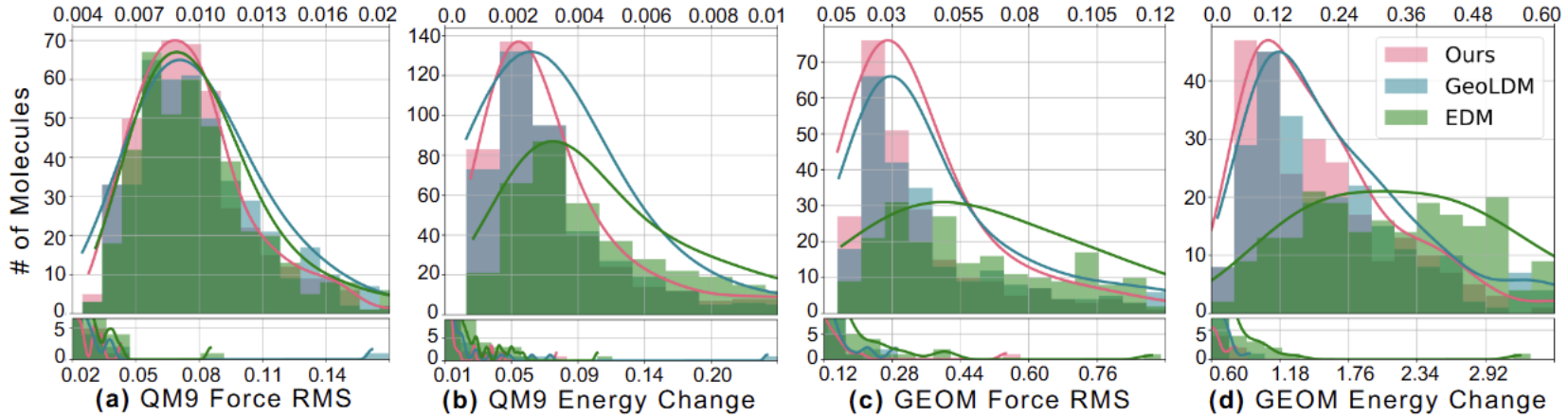


Figure 2: Histograms and distributions of force RMS and energy change of 500 generated molecules from *QM9* and *GEOM* using GeoLDM with CHEMGUIDE under scale=0.0001. Energy change refers to the energy of our generated molecule above its ground state energy.

ChemGuide

- Result
  - ChemGuide improves stability

    on QM9 & GEOM (distribution plots, closer to **0 [left]** the better)

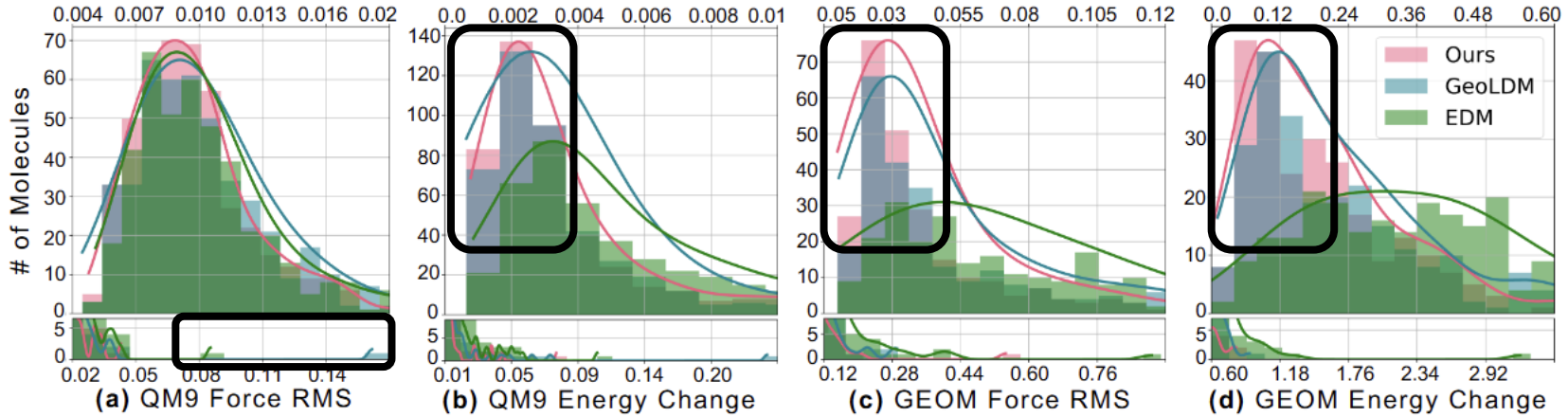


Figure 2: Histograms and distributions of force RMS and energy change of 500 generated molecules from *QM9* and *GEOM* using GeoLDM with CHEMGUIDE under scale=0.0001. Energy change refers to the energy of our generated molecule above its ground state energy.

ChemGuide

- Result
  - ChemGuide improves both stability & molecule property

ChemGuide

- Result

  - ChemGuide improves both **stability** & molecule property

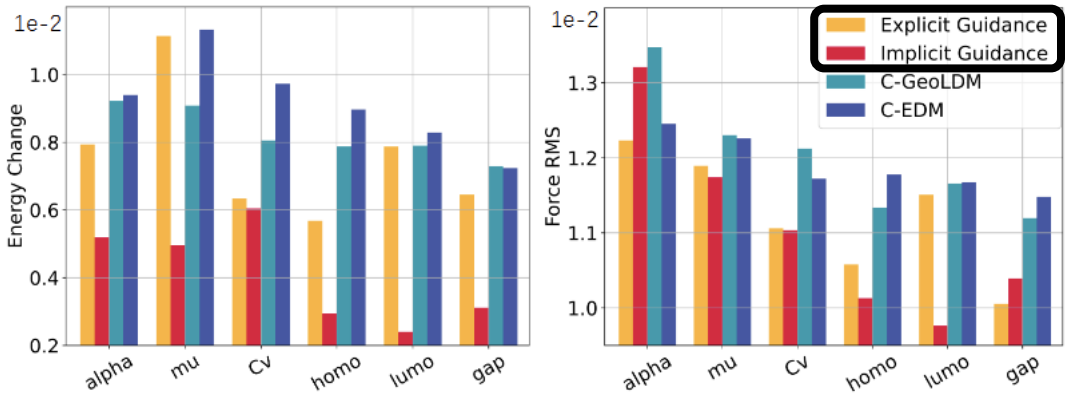


Figure 4: Force RMS and energy change of 200 generated molecules using explicit bilevel optimization, implicit bilevel optimization with noisy neural guidance, C-EDM, and C-GeoLDM. Energy change refers to energy above ground state.

explicit guidance:
ChemGuide+C-GeoLDM

implicit guidance:
ChemGuide+property-guidance+GeoLDM

ChemGuide

- Result

  - ChemGuide improves both **stability** & molecule property

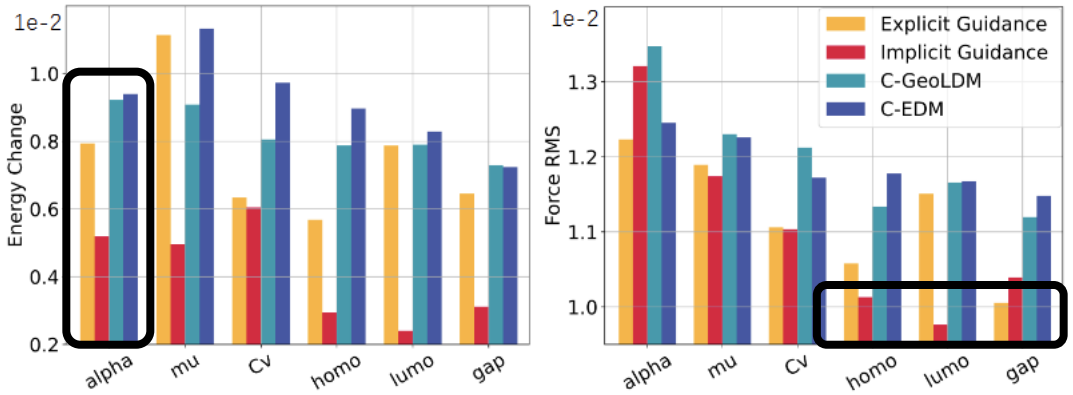


Figure 4: Force RMS and energy change of 200 generated molecules using explicit bilevel optimization, implicit bilevel optimization with noisy neural guidance, C-EDM, and C-GeoLDM. Energy change refers to energy above ground state.

ChemGuide achieves **lowest** energy change & force RMS across different properties, with both **conditional models** & **bilevel optimization** framework

ChemGuide

- Result
  - ChemGuide improves both stability & **molecule property**

Table 5: MAE of 200 generated molecules using C-GeoLDM with CHEMGUIDE and GeoLDM with noisy neural guidance, where the result from the best configuration is reported. We use C-EDM and C-GeoLDM as baselines. $*$ and **bold** denote the overall best and the best within different scales, respectively. Percentage changes between our results and C-GeoLDM are shown in parentheses.

Property Units	α Bohr^3	μ D	C_v $\frac{\text{cal}}{\text{mol}}$K	ϵ_{HOMO} meV	ϵ_{LUMO} meV	$\Delta\epsilon$ meV
C-EDM	2.5089	1.0571	1.0624	0.3304	0.5046	0.6191
C-GeoLDM	2.5593	1.1582	0.9646	0.3407	0.5927	0.4982*
Bilevel Optimization w/ Explicit Guidance	**2.4430*** (-4.54%↓)	1.0483	0.9393	**0.3404*** (-0.09%↓)	**0.5077*** (-14.34%↓)	**0.5225** (4.88%↑)
Bilevel Optimization w/ Noisy Guidance	2.4860	**0.7246*** (-37.43%↓)	**0.9369*** (-2.87%↓)	0.4445	0.8099	0.7041

ChemGuide

- Result
  - ChemGuide improves both stability & **molecule property**

Table 5: MAE of 200 generated molecules using C-GeoLDM with CHEMGUIDE and GeoLDM with noisy neural guidance, where the result from the best configuration is reported. We use C-EDM and C-GeoLDM as baselines. $*$ and **bold** denote the overall best and the best within different scales, respectively. Percentage changes between our results and C-GeoLDM are shown in parentheses.

Property Units	α Bohr^3	μ D	C_v $\frac{\text{cal}}{\text{mol}}$K	ϵ_{HOMO} meV	ϵ_{LUMO} meV	$\Delta\epsilon$ meV
C-EDM	2.5089	1.0571	1.0624	0.3304	0.5046	0.6191
C-GeoLDM	2.5593	1.1582	0.9646	0.3407	0.5927	0.4982*
Bilevel Optimization w/ Explicit Guidance	**2.4430*** (-4.54%↓)	1.0483	0.9393	**0.3404*** (-0.09%↓)	**0.5077*** (-14.34%↓)	**0.5225** (4.88%↑)
Bilevel Optimization w/ Noisy Guidance	2.4860	**0.7246*** (-37.43%↓)	**0.9369*** (-2.87%↓)	0.4445	0.8099	0.7041

ChemGuide is **compatible** with other property guidance methods (with reduced property MAE) in the bilevel optimization framework

ChemGuide

- Result
  - ChemGuide beyond stability

Table 6: CHEMGUIDE for properties (α, μ, ϵ_{HOMO}, ϵ_{LUMO}, $\Delta\epsilon$) with 200 generated molecules on *QM9* and *GEOM*, where the result from the best configuration is reported. We use GeoLDM as the baseline. $*$ and **bold** denote the overall best and the best within different scales, respectively. Percentage changes between our results and GeoLDM are shown in parentheses.

	Property	α	μ	ϵ_{HOMO}	ϵ_{LUMO}	$\Delta\epsilon$
	Units	Bohr^3	D	meV	meV	meV
QM9	GeoLDM	4.7555*	1.5490	0.6264*	2.1790	2.1568
	GeoLDM w/ CHEMGUIDE	**4.8781** (2.58%↑)	**1.5295*** (-1.26%↓)	**0.6392** (2.04%↑)	**2.0091*** (-7.80%↓)	**1.9318*** (-10.43%↓)
GEOM	GeoLDM	24.3424	3.3676	1.1922*	5.0903	1.9956
	GeoLDM w/ CHEMGUIDE	**24.2168*** (-0.52%↓)	**3.3412*** (-0.78%↓)	**1.2939** (8.53%↑)	**5.0802*** (-0.20%↓)	**1.9264*** (-3.47%↓)

ChemGuide **generalizes** to other molecule properties besides stability

Thank you for your interest!