Adjoint Schrödinger Bridge Sampler

NeurIPS 2025 Oral

Guan-Horng Liu*, Jaemoo Choi*, Yongxin Chen, Benjamin K. Miller, Ricky T. Q. Chen*





Meta
Georgia Tech

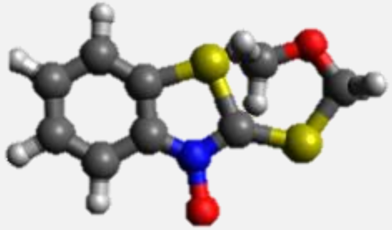
Problem Setup

Boltzmann Distribution

$$\rho_{\text{target}}(x) := \frac{1}{Z}\exp(-E(x)), \quad Z := \int_{\mathcal{X}} \exp(-E(x))\mathrm{d}x$$

unnormalized energy function
(differentiable)

normalization constant
(usually intractable)

$E($  $) = 0.1$

conformer
(3D point clouds)

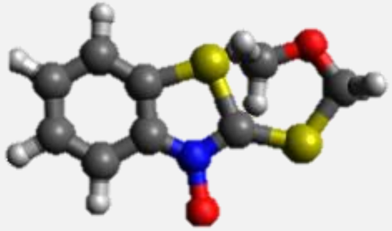
Problem Setup

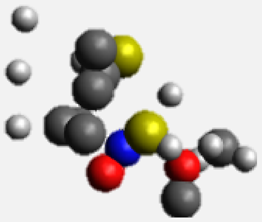
Boltzmann Distribution

$$\rho_{\text{target}}(x) := \frac{1}{Z}\exp(-E(x)), \quad Z := \int_{\mathcal{X}} \exp(-E(x))\mathrm{d}x$$

unnormalized energy function
(differentiable)

normalization constant
(usually intractable)

$E($  $)$ = 0.1

$E($  $)$ = 100

low energy → stable structure → likely to appear → high probability
high energy → unstable structure → unlikely to appear → low probability

Problem Setup

Boltzmann Distribution

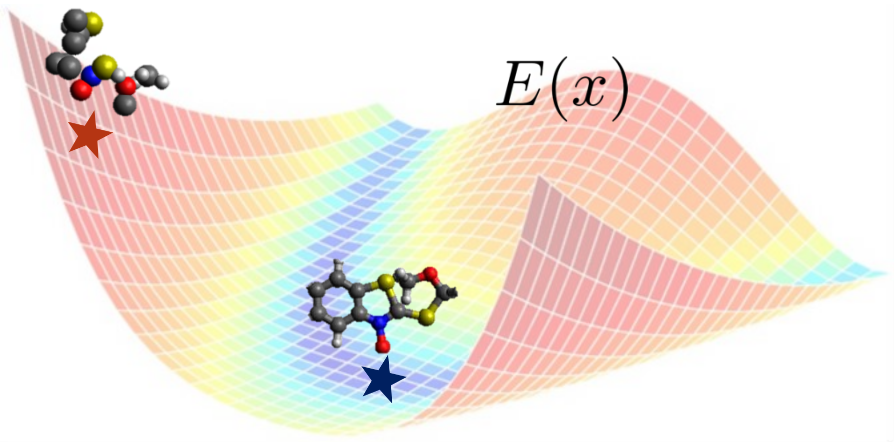
$$\rho_{\text{target}}(x) := \frac{1}{Z}\exp(-E(x)), \quad Z := \int_{\mathcal{X}} \exp(-E(x))\mathrm{d}x$$

unnormalized energy function
(differentiable)

normalization constant
(usually intractable)

Problem Statement

Given $E(x)$, generate $X \sim \rho_{\text{target}}$.



Computational Method

Given $E(x)$, generate $X \sim \rho_{\text{target}}$, where $\rho_{\text{target}}(x) := \frac{1}{Z}\exp(-E(x))$

MCMC methods

Metropolis-Hastings
Annealed Importance Sampling
Hamiltonian Monte Carlo
...

Metropolis-Adjusted Langevin Algorithm

$$\mathrm{d}X_t = -\nabla E(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t$$

Learning-based methods

Boltzmann generators
...
Diffusion samplers

$$\mathrm{d}X_t = [\underbrace{f_t(X_t)}_{\text{prescribed base drift}} + \underbrace{u_t^\theta(X_t)}_{\text{learned control drift}}]\mathrm{d}t + \underbrace{\sigma_t}_{\text{prescribed noise schedule}}\mathrm{d}W_t,$$

$$X_0 \sim \rho_{\text{source}}$$

- ⊗ slow mixing rate
- ⊗ expansive energy evaluation

Diffusion Sampler

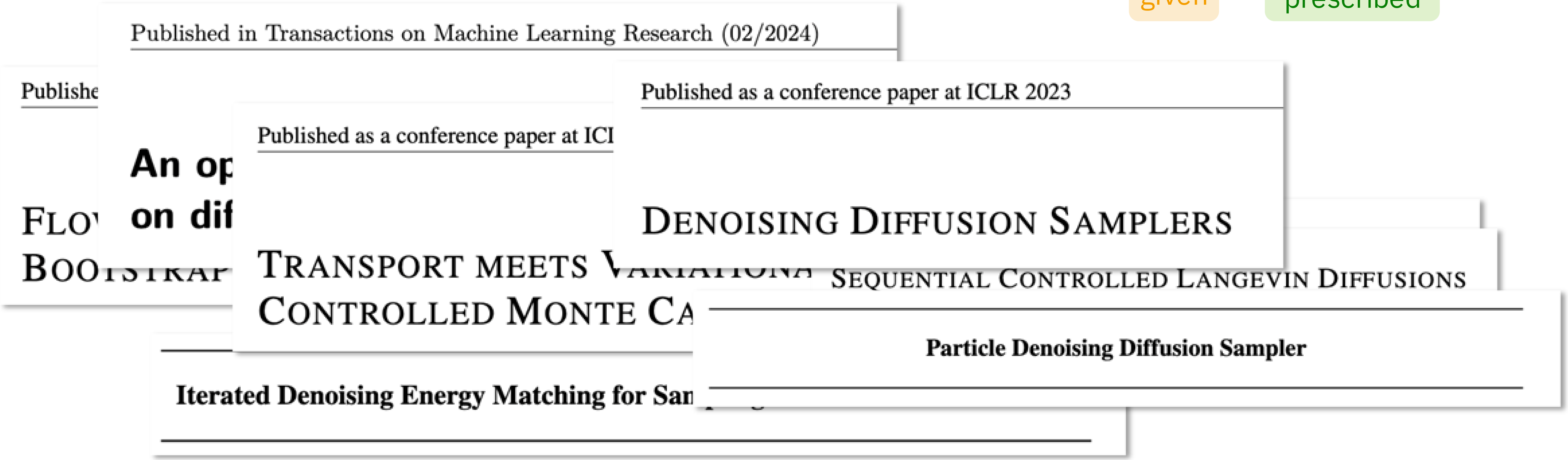
Find $u_t^\theta(\cdot)$ such that $X_1 \sim \rho_{\text{target}}$, where

$$X_0 \sim \rho_{\text{source}}$$

$$\mathrm{d}X_t = [f_t(X_t) + u_t^\theta(X_t)]\mathrm{d}t + \sigma_t \mathrm{d}W_t,$$

$$\rho_{\text{target}}(x) := \frac{1}{Z}\exp(-E(x))$$

given prescribed


Published in Transactions on Machine Learning Research (02/2024)
Publishe
An op
on dif
FLO
BOOTSTRAP
Published as a conference paper at ICI
TRANSPORT MEETS V
CONTROLLED MONTE CA
Published as a conference paper at ICLR 2023
DENOISING DIFFUSION SAMPLERS
SEQUENTIAL CONTROLLED LANGEVIN DIFFUSIONS
Particle Denoising Diffusion Sampler
Iterated Denoising Energy Matching for San

Diffusion Sampler

Find $u_t^\theta(\cdot)$ such that $X_1 \sim \rho_{\text{target}}$, where $\underbrace{\mathrm{d}X_t = [f_t(X_t) + u_t^\theta(X_t)]\mathrm{d}t + \sigma_t \mathrm{d}W_t,}_{}$ $X_0 \sim \rho_{\text{source}}$

$$\rho_{\text{target}}(x) := \frac{1}{Z}\exp(-E(x))$$

given

prescribed

Matching methods

$$\arg\min_\theta \mathbb{E}\left[\|u_t^\theta(X_t) - \square\|^2\right]$$

✓ scalable

Non-Matching methods

✗ backpropagation through SDE, higher-order derivatives

Diffusion Sampler

Find $u_t^\theta(\cdot)$ such that $X_1 \sim \rho_{\text{target}}$, where

$$X_0 \sim \rho_{\text{source}}$$

$$\mathrm{d}X_t = [f_t(X_t) + u_t^\theta(X_t)]\mathrm{d}t + \sigma_t \mathrm{d}W_t,$$

$$\rho_{\text{target}}(x) := \frac{1}{Z}\exp(-E(x))$$

given   prescribed

Matching methods

$$\arg\min_\theta \mathbb{E}\left[\|u_t^\theta(X_t) - \square\|^2\right]$$

score matching

❌ requires $X_1 \sim \rho_{\text{target}}$

adjoint matching

✅ does NOT require $X_1 \sim \rho_{\text{target}}$

Adjoint Matching

Method for Stochastic Optimal Control

control cost / kinetic energy

terminal cost / reward

$$\min_u \mathbb{E}\left[\int_0^1 \frac{1}{2\sigma_t^2}\|u_t(X_t)\|^2 \mathrm{d}t + g(X_1)\right]$$

$$\text{s.t.}\ \ \mathrm{d}X_t = [f_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_0 \sim \rho_{\text{source}}$$

$$\Updownarrow$$

$$u^\star = \arg\min_u \mathbb{E}_{p^{\bar{u}}}\left[\|u_t(X_t) - \Phi_t(\boldsymbol{X}) \cdot \nabla g(X_1)\|^2\right]$$

$$\bar{u} = \texttt{stopgrad}(u)$$

- Matching objective
- Does *not* require $X_1 \sim \rho_{\text{target}}$
- Detached model samples

Domingo-Enrich et al., *Adjoint Matching: Fine-tuning Flow and Diffusion Generative Models with Memoryless Stochastic Optimal Control,* ICLR 2025

Adjoint Matching

Application to sampling

$$\min_u \mathbb{E}\left[\int_0^1 \frac{1}{2\sigma_t^2}\|u_t(X_t)\|^2 \mathrm{d}t + g(X_1)\right]$$

$$\text{s.t. } \mathrm{d}X_t = [f_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_0 \sim \rho_{\text{source}}$$

$\Updownarrow$ $\Downarrow$

$$u^\star = \arg\min_u \mathbb{E}_{p^{\bar{u}}}\left[\|u_t(X_t) - \Phi_t(\boldsymbol{X}) \cdot \nabla g(X_1)\|^2\right]$$

$$\bar{u} = \texttt{stopgrad}(u)$$

$\Rightarrow$

optimal distribution induced by u^*

$$\rho_{\text{target}}(X_1) = \int p^\star(X_0, X_1)\mathrm{d}X_0$$

$$\rho_{\text{target}}(x) := \frac{1}{Z}\exp(-E(x))$$

Adjoint-based Diffusion Sampler

Variational Inference
(Diffusion models → Stochastic Optimal Control)

Scalable Computational Method
(SOC → adjoint matching)

Probabilistic Characteristics

Adjoint-based Diffusion Sampler

connection to Schrödinger Bridge

$$\min_u \mathbb{E}\left[\int_0^1 \frac{1}{2\sigma_t^2}\|u_t(X_t)\|^2 \mathrm{d}t + g(X_1)\right]$$

$$\text{s.t.}\ \ \mathrm{d}X_t = [f_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_0 \sim \rho_{\text{source}}$$

$\Updownarrow$ $\quad\quad$ $\Downarrow$

Scalable Computational Method
(SOC → adjoint matching)

$\Rightarrow$

$$\rho_{\text{target}}(X_1) = \int p^\star(X_0, X_1)\mathrm{d}X_0$$

$\rho_{\text{target}}(x) := \frac{1}{Z}\exp(-E(x))$

For ***any* base process**, there's an **SOC problem** satisfying the equality constraint.

largest model class

This is **Schrödinger Bridge**!

Adjoint Schrödinger Bridge Sampler

$$\min_u \mathbb{E}\left[\int_0^1 \frac{1}{2\sigma_t^2}\|u_t(X_t)\|^2 \mathrm{d}t + E(X_1) + \log \hat{\varphi}_1(X_1)\right]$$

$$\text{s.t. } \mathrm{d}X_t = u_t(X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_0 \sim \rho_{\text{source}}$$

$\Updownarrow$ $\Downarrow$

Scalable Computational Method (SOC → adjoint matching)

$\Rightarrow$

$$\rho_{\text{target}}(X_1) = \int p^\star(X_0, X_1)\mathrm{d}X_0$$

$\rho_{\text{target}}(x) := \frac{1}{Z}\exp(-E(x))$

$$g(X_1) := E(X_1) + \log \hat{\varphi}_1(X_1)$$

$$\varphi_0(x) = \int p_{1|0}^{\text{base}}(y|x)\varphi_1(y)\mathrm{d}y, \quad \varphi_0(x)\hat{\varphi}_0(x) = \rho_{\text{prior}}(x),$$

$$\hat{\varphi}_1(y) = \int p_{1|0}^{\text{base}}(y|x)\hat{\varphi}_0(x)\mathrm{d}x, \quad \varphi_1(y)\hat{\varphi}_1(y) = \rho_{\text{target}}(y),$$

Adjoint Schrödinger Bridge Sampler

intractable

$$\min_u \mathbb{E}\left[\int_0^1 \frac{1}{2\sigma_t^2}\|u_t(X_t)\|^2 \mathrm{d}t + E(X_1) + \log\hat{\varphi}_1(X_1)\right]$$

$$\text{s.t. } \mathrm{d}X_t = u_t(X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_0 \sim \rho_{\text{source}}$$

$\Updownarrow$ $\quad\quad$ $\Downarrow$

$$u^\star = \arg\min_u \mathbb{E}_{p^{\bar{u}}}\left[\|u_t(X_t) + \sigma_t^2 (\nabla E + \nabla \log\hat{\varphi}_1)(X_1)\|^2\right]$$

$$\nabla \log\hat{\varphi}_1 = \arg\min_h \mathbb{E}_{p^{\bar{u}}}\left[\|h(X_1) - \nabla_{x_1}\log p^{\text{base}}(X_1|X_0)\|^2\right]$$

tractable!

$\bar{u} = \texttt{stopgrad}(u)$

$\Rightarrow$

$$\rho_{\text{target}}(X_1) = \int p^\star(X_0, X_1)\mathrm{d}X_0$$

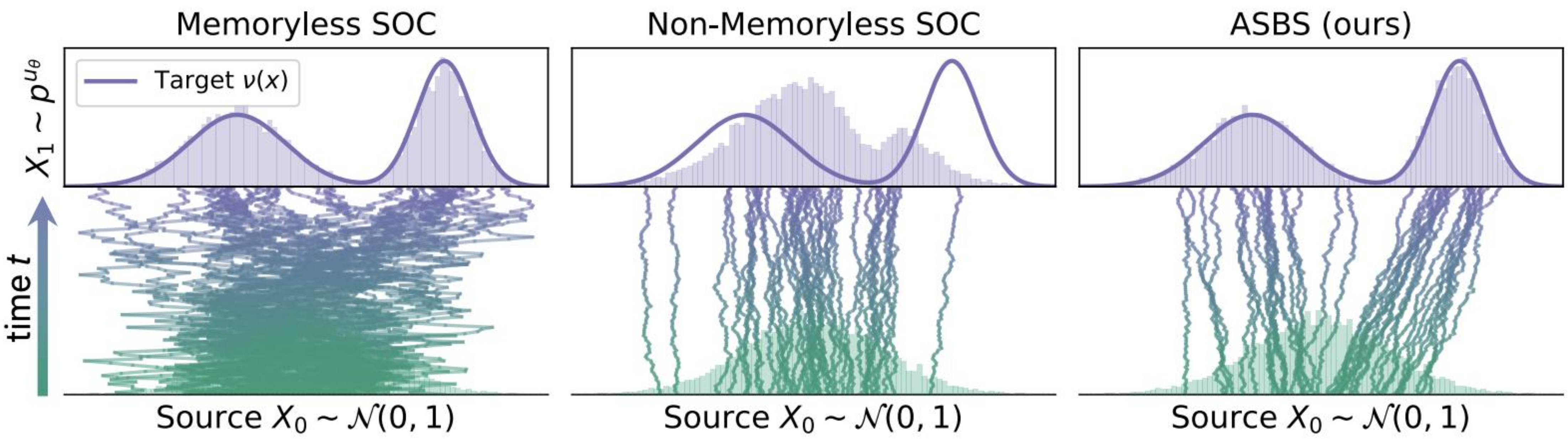
$$g(X_1) := E(X_1) + \log\hat{\varphi}_1(X_1)$$

$$\varphi_0(x) = \int p^{\text{base}}_{1|0}(y|x)\varphi_1(y)\mathrm{d}y, \quad \varphi_0(x)\hat{\varphi}_0(x) = \rho_{\text{prior}}(x),$$

$$\hat{\varphi}_1(y) = \int p^{\text{base}}_{1|0}(y|x)\hat{\varphi}_0(x)\mathrm{d}x, \quad \varphi_1(y)\hat{\varphi}_1(y) = \rho_{\text{target}}(y),$$

Adjoint Matching $\quad$ **Corrector Matching**

Demo on 1D



Memoryless SOC
Non-Memoryless SOC
ASBS (ours)
Target $\nu(x)$
$X_1 \sim p^{u_\theta}$
time t
Source $X_0 \sim \mathcal{N}(0, 1)$
Source $X_0 \sim \mathcal{N}(0, 1)$
Source $X_0 \sim \mathcal{N}(0, 1)$

Experiment Results

Synthetic multi-particle systems

	MW-5 (d=5)	DW-4 ($d=8$)		LJ-13 ($d=39$)		LJ-55 ($d=165$)	
Method	Sinkhorn ↓	$\mathcal{W}_2$ ↓	$E(\cdot)$ $\mathcal{W}_2$ ↓	$\mathcal{W}_2$ ↓	$E(\cdot)$ $\mathcal{W}_2$ ↓	$\mathcal{W}_2$ ↓	$E(\cdot)$ $\mathcal{W}_2$ ↓
PDDS (Phillips et al., 2024)	—	$0.92_{\pm 0.08}$	$0.58_{\pm 0.25}$	$4.66_{\pm 0.87}$	$56.01_{\pm 10.80}$	—	—
SCLD (Chen et al., 2025)	$0.44_{\pm 0.06}$	$1.30_{\pm 0.64}$	$0.40_{\pm 0.19}$	$2.93_{\pm 0.19}$	$27.98_{\pm 1.26}$	—	—
PIS (Zhang and Chen, 2022)	$0.65_{\pm 0.25}$	$0.68_{\pm 0.28}$	$0.65_{\pm 0.25}$	$1.93_{\pm 0.07}$	$18.02_{\pm 1.12}$	$4.79_{\pm 0.45}$	$228.70_{\pm 131.27}$
DDS (Vargas et al., 2023)	$0.63_{\pm 0.24}$	$0.92_{\pm 0.11}$	$0.90_{\pm 0.37}$	$1.99_{\pm 0.13}$	$24.61_{\pm 8.99}$	$4.60_{\pm 0.09}$	$173.09_{\pm 18.01}$
LV-PIS (Richter and Berner, 2024)	—	$1.04_{\pm 0.29}$	$1.89_{\pm 0.89}$	—	—	—	—
iDEM (Akhound-Sadegh et al., 2024)	—	$0.70_{\pm 0.06}$	$0.55_{\pm 0.14}$	$1.61_{\pm 0.01}$	$30.78_{\pm 24.46}$	$4.69_{\pm 1.52}$	$93.53_{\pm 16.31}$
AS (Havens et al., 2025)	$0.32_{\pm 0.06}$	$0.62_{\pm 0.06}$	$0.55_{\pm 0.12}$	$1.67_{\pm 0.01}$	$2.40_{\pm 1.25}$	$4.04_{\pm 0.05}$	$30.83_{\pm 8.19}$
ASBS (**Ours**)	$0.15_{\pm 0.02}$	$0.43_{\pm 0.05}$	$0.20_{\pm 0.11}$	$1.59_{\pm 0.03}$	$1.99_{\pm 1.01}$	$4.00_{\pm 0.03}$	$28.10_{\pm 8.15}$

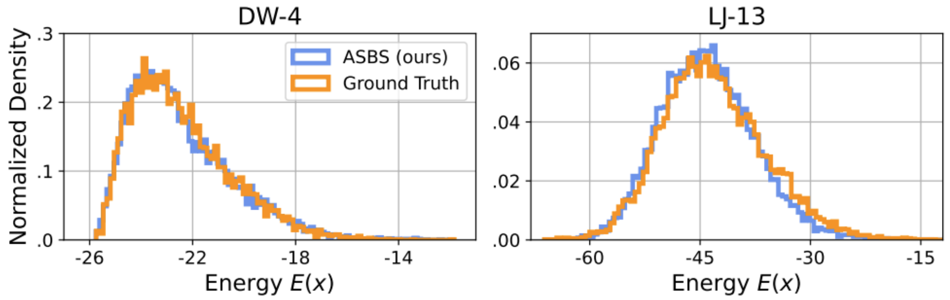


Figure 3 The energy histograms of DW-4 and LJ-13 from Table 2. ASBS generates samples whose energy profiles closely match those of the ground-truth samples.

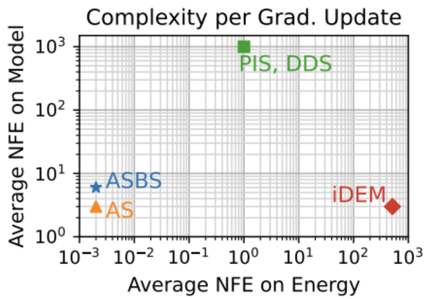


Figure 4 Complexity w.r.t. the number of function evaluation (NFE) on LJ-13 potential.

Experiment Results

Large-Scale Amortized Conformer Generation

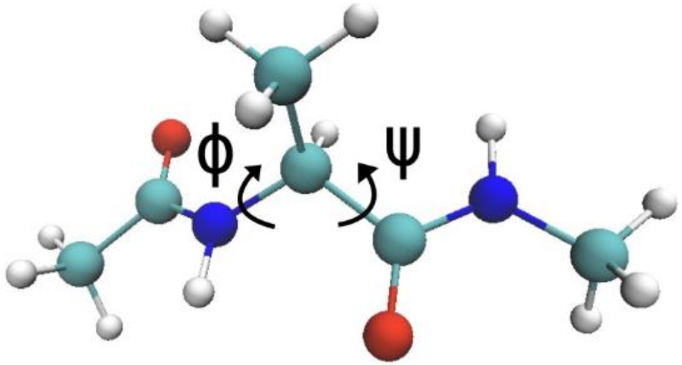
$$\rho_{\text{target}}(x|g) = e^{-\frac{1}{\tau}E(x|g)},\quad g \in \mathcal{G},\quad \tau \ll 1$$

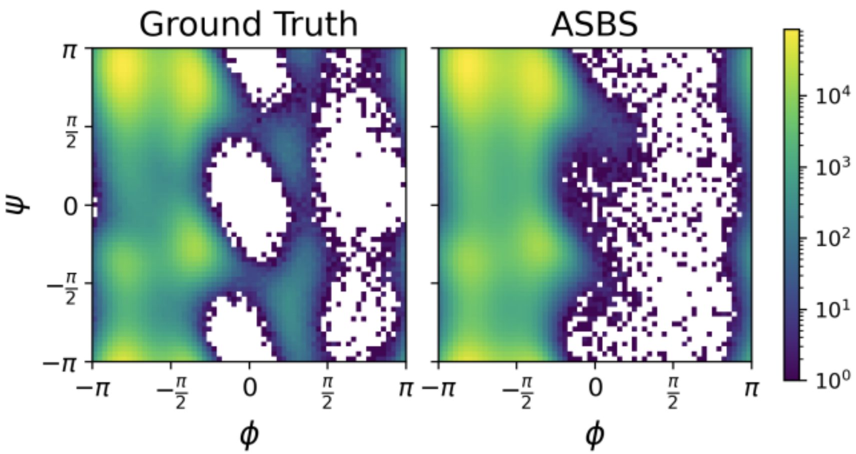
graph structure — conformer

	without relaxation				with relaxation			
	SPICE		GEOM-DRUGS		SPICE		GEOM-DRUGS	
Method	Coverage ↑	AMR ↓	Coverage ↑	AMR ↓	Coverage ↑	AMR ↓	Coverage ↑	AMR ↓
RDKit ETKDG (Riniker and Landrum, 2015)	$56.94_{\pm 35.82}$	$1.04_{\pm 0.52}$	$50.81_{\pm 34.69}$	$1.15_{\pm 0.61}$	$70.21_{\pm 31.70}$	$0.79_{\pm 0.44}$	$62.55_{\pm 31.67}$	$0.93_{\pm 0.53}$
AS (Havens et al., 2025)	$56.75_{\pm 38.15}$	$0.96_{\pm 0.26}$	$36.23_{\pm 33.42}$	$1.20_{\pm 0.43}$	$82.41_{\pm 25.85}$	$0.68_{\pm 0.28}$	$64.26_{\pm 34.57}$	$0.89_{\pm 0.45}$
ASBS w/ Gaussian prior (**Ours**)	$73.04_{\pm 31.95}$	$0.83_{\pm 0.24}$	$50.23_{\pm 35.98}$	$1.05_{\pm 0.43}$	$88.26_{\pm 20.57}$	$0.60_{\pm 0.24}$	$72.32_{\pm 29.68}$	$0.77_{\pm 0.35}$
ASBS w/ harmonic prior (**Ours**)	$74.05_{\pm 31.61}$	$0.82_{\pm 0.23}$	$53.14_{\pm 35.69}$	$1.03_{\pm 0.42}$	$88.71_{\pm 18.63}$	$0.59_{\pm 0.24}$	$72.77_{\pm 29.94}$	$0.78_{\pm 0.35}$

Experiment Results

Boltzmann distribution of Alanine dipeptide





Method	D_{KL} on each torsion's marginal ↓				
	ϕ	ψ	γ_1	γ_2	γ_3
PIS (Zhang and Chen, 2022)	$0.05_{\pm 0.03}$	$0.38_{\pm 0.49}$	$5.61_{\pm 1.24}$	$4.49_{\pm 0.03}$	$4.60_{\pm 0.03}$
DDS (Vargas et al., 2023)	$0.03_{\pm 0.01}$	$0.16_{\pm 0.07}$	$2.44_{\pm 0.96}$	$0.03_{\pm 0.00}$	$0.03_{\pm 0.00}$
AS (Havens et al., 2025)	$0.09_{\pm 0.09}$	$0.04_{\pm 0.04}$	$0.17_{\pm 0.17}$	$0.56_{\pm 0.09}$	$0.51_{\pm 0.06}$
ASBS (**Ours**)	$0.02_{\pm 0.00}$	$0.01_{\pm 0.00}$	$0.03_{\pm 0.01}$	$0.02_{\pm 0.00}$	$0.02_{\pm 0.00}$

*image credit to FAB (Midgley et al 2022)

Poster Session
11am-2pm #612
(today)

Adjoint-based Diffusion Sampler

Variational Inference / Stochastic Optimal Control

$$\min_u \mathbb{E}\left[\int_0^1 \frac{1}{2\sigma_t^2}\|u_t(X_t)\|^2 \mathrm{d}t + g(X_1)\right]$$

$$\text{s.t.}\ \ \mathrm{d}X_t = [f_t(X_t) + u_t(X_t)]\mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad X_0 \sim \rho_{\text{source}}$$

$\Updownarrow$ $\qquad\qquad$ $\Downarrow$

Scalable Adjoint-based Method

$$u^\star = \arg\min_u \mathbb{E}_{p^{\bar{u}}}\left[\|u_t(X_t) - \Phi_t(\boldsymbol{X}) \cdot \nabla g(X_1)\|^2\right]$$

$$\bar{u} = \mathtt{stopgrad}(u)$$

$\Rightarrow$

Probabilistic Characteristics

$$\rho_{\text{target}}(X_1) = \int p^\star(X_0, X_1)\mathrm{d}X_0$$

$$\rho_{\text{target}}(x) := \tfrac{1}{Z}\exp(-E(x))$$

[1] *Adjoint Sampling*, ICML 2025.

[2] *Adjoint Schrödinger Bridge Sampler*, NeurIPS 2025 (*Oral*).

[3] *Non-equilibrium Annealed Adjoint Sampler*, NeurIPS 2025.

[4] *Well-Tempered Adjoint Schrödinger Bridge Sampler*, preprint 2025.

[5] *Functional Adjoint Sampler*, preprint 2025.

arXiv *code* *slides*