Diffusion Generative Flow Samplers: Improving Learning Signals Through Partial Trajectory Optimization

Dinghuai Zhang*, Ricky T. Q. Chen, Cheng-Hao Liu, Aaron Courville & Yoshua Bengio

Thomas Mousseau

October 18, 2025

Overview

Sampling from Unnormalized Densities

Goal. Sample from a D-dimensional target with unnormalized density $\mu(x)$:

$$\pi(x) = \frac{\mu(x)}{Z}, \qquad Z = \int_{\mathbb{R}^D} \mu(x)\, dx \text{ (unknown)}.$$

We assume we can evaluate $\mu(x)$ (and sometimes $\nabla \log \mu(x)$), but we have no samples from π and do not know Z.

Context. We seek a *sampler* (peer to MCMC/VI) that produces calibrated samples and, ideally, estimates of $\log Z$, *without* any dataset from π.

Chemistry (small-molecule conformers). Use *conformer strain/steric clash energy* as $E(x)$ for a 3D shape x (lower when bonds/angles/rotations are comfortable). Sampling $\pi(x) \propto e^{-\beta E(x)}$ gives the *population of stable rotamers* and how often each shape occurs at temperature T.

The integral for Z over all 3D configurations is intractable, so we use the unnormalized score $\mu(x) = e^{-\beta E(x)}$, where $E(x)$ comes from chemistry/force-field physics (bonds, angles, nonbonded terms) and encodes how likely a conformation is, a lower E means more likely.

SOC formulation (part 1/2)

Forward (controlled) process Q. A Markov chain with Gaussian transitions:

$$Q(x_{0:N}): \quad x_0 \sim p_0^{\text{ref}}, \quad x_{n+1} \sim P_F(\cdot \mid x_n) = \mathcal{N}\big(x_n + h\, f(x_n, n),\ h\sigma^2 I\big).$$

Reference process Q^{ref}. Same covariance, zero drift:

$$x_{n+1} \sim P_F^{\text{ref}}(\cdot \mid x_n) = \mathcal{N}\big(x_n,\ h\sigma^2 I\big), \qquad x_0 \sim p_0^{\text{ref}},\ \ p_n^{\text{ref}} \text{ known}.$$

Target process P. Tie the terminal marginal to π via the reference:

$$P(x_{0:N}) \ := \ Q^{\text{ref}}(x_{0:N})\ \frac{\pi(x_N)}{p_N^{\text{ref}}(x_N)}.$$

Then $P(x_N) \propto \mu(x_N)$, making P a valid path-space target.

SOC formulation (part 2/2)

Learning objective (discrete-time SOC). Learn f by minimizing the path KL:

$$\min_f \; D_{\mathrm{KL}}(Q \parallel P) \iff \min_f \; \mathbb{E}_Q\Big[\sum_{n=0}^{N-1} \frac{h}{2\sigma^2} \|f(x_n, n)\|^2 + \log \Psi(x_N)\Big],$$

with $\Psi(x_N) = \dfrac{p_N^{\mathrm{ref}}(x_N)}{\mu(x_N)}$. (Continuous-time limit recovers the classic VE-SDE SOC formulation.)

Why SOC, and what still hurts (motivation for DGFS)

Why SOC/control-as-inference?

- Principled *path-space* objective; Z cancels, so only μ (and optionally $\nabla \log \mu$) is needed.
- Calibrated sampling by *steering* a simple reference process toward the target.

Pain point in prior SOC samplers (PIS/DDS).

- Training signal sits *only at terminal time* N and losses use *full trajectories* $\Rightarrow$ poor credit assignment, high variance, weaker mode coverage.

DGFS in one line.

- Keep the same SOC/path-KL setup, but introduce a learned *flow function* $F_n(x_n)$ and enforce *subtrajectory balance* to inject *intermediate* learning signals and enable *partial-trajectory* training.

Prior Work & Limitations

SOC framing via SDEs. Representing diffusion models as controlled SDEs lets us cast sampling as a stochastic optimal control (SOC) or control-as-inference problem (minimize a KL over *path measures*).

Path Integral Sampler (PIS).

- Uses Schrödinger bridge / path-integral control to learn u_θ that transports a reference diffusion to match π at time N.
- Principled *trajectory-space* objective that works with *unnormalized* μ; supports off-policy training.
- *Limitation:* Objective requires *full trajectories* and rewards at terminal time only $\Rightarrow$ sluggish credit assignment, higher gradient variance.

Denoising Diffusion Samplers (DDS).

- Also recast sampling as control over the forward path; improves practical sampling and stability vs. vanilla diffusion heuristics.

DGFS Core

Trajectories with Learned Flow F_n

DGFS uses trajectories guided by a learned flow F_n for sampling.

- Partial-trajectory updates via SubTB (Subtrajectory Balance).
- Intermediate training signals for better credit assignment.
- Controlled forward process and subtrajectory balance constraint.
- Partition-function estimator for evaluation.
- Two nets: Drift f and scalar flow F_n.
- Training setup, step embeddings, and cost vs. inference speed trade-offs.

Why It Works

Evidence for Effectiveness

DGFS reduces gradient variance and improves stability.

- Lower gradient variance through partial trajectories.
- Better mode coverage and stability in sampling.
- Compatibility with off-policy exploration.

Results

Key Findings

Evaluation of log Z bias across various targets.

- Visual grids for distributions like MoG (Mixture of Gaussians) and Manywell.
- Key ablations on training components and hyperparameters.

Limitations & Open Questions

Challenges and Future Work

Designing intermediate signals remains an open problem.

- Scaling to high-dimensional and costly μ (target distributions).
- Fixed backward policy P_B limits flexibility.

Conclusion