DIFFUSION MODELS ARE MOLECULAR DYNAMICS SIMULATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

We show that a denoising–diffusion sampler equipped with a *Harmonic Adapter*—a quadratic coupling with an offset that ties neighbouring reverse-time iterates—is exactly an Euler–Maruyama (EM) integrator for overdamped Langevin dynamics. One reverse step with spring stiffness k integrates the SDE with implicit step size $\Delta t = \beta/(2k)$; the drift is the learned score (or energy gradient). This identity reframes molecular dynamics (MD) through diffusion: the accuracy of trajectories and equilibrium statistics is controlled by two dials, *model capacity* (via universal approximation of the drift) and the *number of denoising steps* N, rather than by a fixed, tiny MD timestep.

The practical consequence is a data-driven MD framework that learns *forces for free* from i.i.d. configurations, requires no hand-crafted force fields, and can be run with a small, distillable number of reverse steps while preserving the Boltzmann law of the learned energy. We prove pathwise KL bounds that separate discretisation error $\mathcal{O}(\sum_n \Delta t^2)$ from score error, show how temperature enters through the spring, and demonstrate MD-like temporal correlations on molecular trajectories generated entirely by a score model trained on static samples.

1 INTRODUCTION

Diffusion models have reshaped generative modeling, from vision to molecular data (1; 2; 3), delivering high-fidelity samples and robust likelihood surrogates (4). In molecular modeling, however, the objective is not merely to match static distributions but to generate configurations that are consistent with physical constraints and thermodynamics (5). Classical approaches such as Langevin Molecular Dynamics (LMD) (6; 7) integrate stochastic differential equations to evolve systems in time, enabling exploration of conformational landscapes and analysis of dynamical processes (8).

A central difficulty shared by both diffusion samplers and LMD is sampling from complex, high-dimensional *non-iid* target distributions. Molecular interactions (covalent bonding, electrostatics, van der Waals forces) induce long-range and multi-scale correlations that invalidate iid assumptions and complicate both inference and analysis (9). Recent diffusion models specialized to molecules seek to respect these structures (10; 11), including symmetry-equivariant formulations that help preserve physical invariances (12). Yet ensuring that generated samples obey physical laws and exhibit meaningful temporal statistics remains challenging.

Temporal correlations are especially important for scientific use cases. Numerical solvers for ODE/SDEs produce trajectories via sequential updates that encode short-time physics and enable computation of path-dependent observables (13). Such trajectories underpin studies of folding kinetics and reaction mechanisms (14). By contrast, standard diffusion sampling treats reverse-time updates as conditionally independent draws, which yields excellent equilibria but discards dynamical structure.

Bridging these views motivates introducing *structured interactions* into the reverse process. In diffusion models, non-iid noise or couplings can emulate dependencies analogous to those in MD; however, analyzing the correctness and convergence of such modifications is subtle (15). Our perspective is to endow reverse-time iterations with a simple, *Harmonic Adapter*—a quadratic coupling with an offset that ties neighboring reverse steps—so that each denoising update admits a physically interpretable meaning.

Key idea (in one line). We show that a denoising–diffusion sampler equipped with a Harmonic Adapter performs, step-for-step, the Euler–Maruyama (EM) update for overdamped Langevin dynamics at an *implicit* resolution $\Delta t = \beta/(2k)$ determined by the spring stiffness k. In effect, the denoising grid becomes a controllable *resolution dial* for MD-like trajectories, while retaining the probabilistic benefits of diffusion.

Why this matters (computational pain point). Molecular scientists routinely need trajectories spanning micro- to millisecond scales to compute kinetics, nonequilibrium responses, and transport properties. A standard LMD integrator with a $2\,\mathrm{fs}$ step requires $\sim 10^9$ updates to reach $2\,\mu\mathrm{s}$ of simulated time, often translating into days of wall-clock GPU time even with aggressive parallelization. In contrast, score-based diffusion can produce equilibrium-quality conformations in *tens* of reverse steps after distillation, suggesting a path to 10^2–$10^3\times$ reductions in wall-clock cost without abandoning statistical exactness in the learned energy.

Contributions. **(1)** We provide an exact algebraic identity: one reverse denoising step with a Harmonic Adapter equals one EM step for overdamped Langevin, with drift given by the learned score/energy gradient and noise given by the EM Gaussian. **(2)** This identity yields a clean accuracy budget that separates model (score) error from grid (reverse-schedule) error, and supports principled temperature/variance control during inference. **(3)** The same quadratic coupling enables fully time-parallel execution: batches index trajectory slices that can be updated simultaneously while preserving the correct local-in-time law. **(4)** Empirically, glued trajectories reproduce ensemble statistics and exhibit MD-like temporal correlations on the learned energy landscape on open-source pretrained models, enabling trajectory-level observables at diffusion costs.

Outlook. Treating reverse-time diffusion as an MD integrator on the learned potential opens a route to *trajectory* generators with physical semantics, modular coupling to MCMC and enhanced sampling, and higher-order adapters—all while retaining the scalability and training simplicity that make diffusion attractive in the first place.

FROM TRAJECTORIES TO EQUILIBRIA—AND BACK AGAIN

A diffusion model with a simple *quadratic (harmonic) coupling* between consecutive reverse-time iterates is an MD integrator in disguise. Section H.2 formalises this: completing the square in the exact EM transition shows that the reverse kernel of denoising–diffusion with the harmonic offset coincides with the EM kernel

$$x_{n+1} = x_n - \Delta t\,\nabla V_\theta(x_n) + \sqrt{2D\,\Delta t}\,\xi_n, \qquad \Delta t = \frac{\beta}{2k}.$$

Here ∇V_θ is the score/energy gradient learned from i.i.d. data. In this reframing, *time resolution is no longer a hard bottleneck*: fidelity is governed by (a) how well the network approximates the true drift (universal approximation) and (b) how many reverse steps we choose (which we can distil to a few dozen without retraining the score).

Standard Langevin MD advances at femtosecond timesteps to remain stable; reaching microseconds requires $\sim 10^6$ steps. Diffusion samplers, by contrast, operate on a coarse reverse-time grid whose size we can set for inference. The Harmonic Adapter makes this more than an analogy: one reverse step *is* one EM step, with an *implicit* Δt encoded by the spring. Thus, long-time, MD-like trajectories (and time-correlated observables such as dihedral autocorrelations) can be generated with wall-clock cost proportional to the *number of denoising iterations*, not the micro-timestep count of MD.

Training requires only static, i.i.d. configurations. A denoising objective learns the score $\nabla \log p$; we view $E_\theta := -\log p_\theta$ as a *learned energy* whose gradient plays the role of a force field. No bespoke bonded/non-bonded decomposition or parameter fitting is needed. The resulting sampler integrates the learned energy landscape with EM accuracy, and a one-step Metropolis check (optional) makes the terminal law exactly Boltzmann in E_θ.

The adapter penalises the *mismatch* $x_{n+1} - x_n + \Delta t\,\nabla V_\theta(x_n)$ with stiffness $k = \beta/(2\Delta t)$. Because the mismatch is $\mathcal{O}(\sqrt{\Delta t})$, the quadratic energy contributes only $\mathcal{O}(\Delta t)$ to the exponent; the continuum limit is well behaved. Operationally, this turns independent denoising updates into a *time-correlated* trajectory consistent with Langevin physics.

Accuracy separates cleanly into *score error* and *grid error*. Our pathwise bound (Sec. 3.4) shows

$$D_{\mathrm{KL}}\big(\mathcal{L}(\tilde{X}_{[0,T]}) \,\|\, \mathcal{L}(X_{[0,T]})\big) \;\le\; T\,\bar{\varepsilon}^2 \;+\; C\sum_{n=0}^{N-1}\Delta t^2.$$

so better networks (universal approximation) and fewer but suitably chosen reverse steps N jointly control fidelity. In practice, distilled schedules with $N \in [10, 50]$ already reproduce MD-like autocorrelations while costing orders of magnitude less than conventional MD.

This viewpoint recasts “force fields” as *learned scores* trained on i.i.d. data and turns MD into *generative inference*. Temperature enters only via the spring ($k = \beta/(2\Delta t)$), enabling explicit control of effective temperature and correlation length at inference, and opening the door to time-parallel sampling and easy coupling to other stochastic algorithms (MCMC, metadynamics, alchemy) without redesigning the network.

The rest of the paper is organized as follows: we provide background on diffusion models and present our theoretical framework showing equivalence between diffusion models with harmonic bias and LMD. We discuss the implications for parallelization across time steps and the potential for efficient simulation of molecular systems. In the appendix we discuss in depth the equivalence, illustrate results related to the underdamped langevin diffusion, and discuss future work about using this formalism to do MCMC, Metadyanmics, and Alchemical Free Energy estimations in parallel.

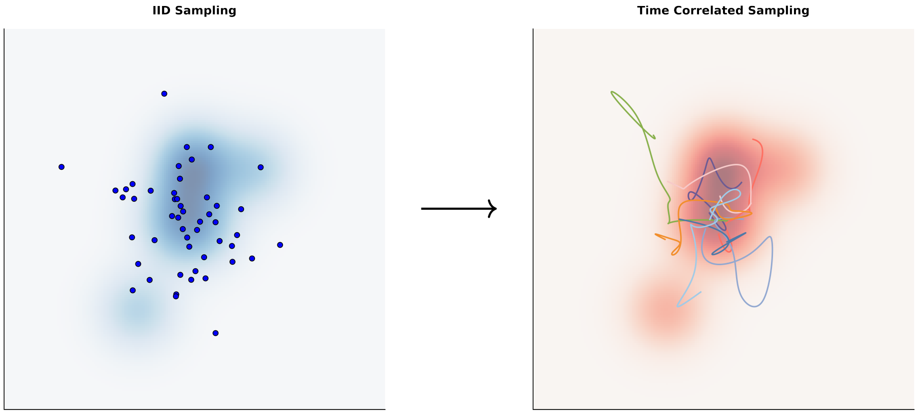


Figure 1: Left: traditional i.i.d. diffusion sampling ignores temporal structure. Right: our harmonically-coupled sampler recovers time-correlated trajectories consistent with MD.

2 DEFINITIONS, NOTATION, AND WHAT IS PROVEN

2.1 OVER-DAMPED LANGEVIN DYNAMICS (CONTINUOUS TIME)

Units and constants. Throughout we choose *friction units* such that

$$\gamma = 1 \quad\Longrightarrow\quad D = \frac{k_B T}{\gamma} = k_B T, \quad \beta D \;=\; 1.$$

Stating this explicitly prevents hidden factors of γ or k_B from re-appearing later.

A single particle in a conservative potential $V : \mathbb{R}^d \to \mathbb{R}$ at temperature T therefore obeys

$$d\boldsymbol{x}(t) = -\nabla V\big(\boldsymbol{x}(t)\big)\,dt + \sqrt{2D}\,d\boldsymbol{w}(t), \qquad D = k_B T,\ \beta = 1/D. \tag{1}$$

2.2 EULER–MARUYAMA TIME DISCRETISATION

Discretising the overdamped Langevin SDE equation 28 with a fixed step size $\Delta t > 0$ gives the classical Euler–Maruyama (EM) integrator

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \nabla V(\mathbf{x}_n)\,\Delta t + \sqrt{2D\,\Delta t}\,\boldsymbol{\xi}_n, \qquad \boldsymbol{\xi}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d). \tag{2}$$

Hence, **conditioned on $\mathbf{x}_n$**, the next point $\mathbf{x}_{n+1}$ is Gaussian

$$p(\mathbf{x}_{n+1} \,|\, \mathbf{x}_n) = \frac{\exp\!\big[-\frac{\beta}{4D\Delta t}\,\|\mathbf{x}_{n+1} - \mathbf{x}_n + \Delta t\,\nabla V(\mathbf{x}_n)\|^2\big]}{(4\pi D\Delta t)^{d/2}}\ . \tag{3}$$

This is the **exact** EM transition kernel; no other terms appear in the exponent.

2.3 HARMONIC REFORMULATION (“QUADRATIC GLUE”)

For later use it is convenient to rewrite equation 30 as a Boltzmann factor with a *quadratic coupling* (the *Harmonic Adapter*)

$$p(\mathbf{x}_{n+1} \,|\, \mathbf{x}_n) = Z(\mathbf{x}_n)^{-1}\ \exp\!\Big[-\beta\,\frac{k(\Delta t)}{2}\ \big\|\mathbf{x}_{n+1} - \mathbf{x}_n + \nabla V(\mathbf{x}_n)\Delta t\big\|^2\Big], \tag{4}$$

$$k(\Delta t) = \frac{1}{2D\Delta t}. \tag{5}$$

A key observation is that the *mismatch* $\mathbf{x}_{n+1} - \mathbf{x}_n + \nabla V(\mathbf{x}_n)\Delta t = \mathcal{O}(\sqrt{\Delta t})$, so the spring energy contributes only $\mathcal{O}(\Delta t)$ to the exponent even though the stiffness $k(\Delta t) \sim \Delta t^{-1}$ diverges. Consequently the limit $\Delta t \to 0$ remains well behaved and the kernel equation 31 converges to the continuum Langevin propagator, §H.

3 IMPLICATIONS FOR SCORE-BASED (DENOISING–DIFFUSION) MODELS

3.1 FROM IID DENOISING TO MOLECULAR DYNAMICS: THE EQUIVALENCE PRINCIPLE

Key message. With the harmonic (quadratic) adapter, each reverse-diffusion step is *exactly* an Euler–Maruyama (EM) step for overdamped Langevin with step Δt. Thus, if the score/force proxy equals the true force and the denoising grid is refined, the glued diffusion paths converge to MD trajectories targeting the same Boltzmann law. The spring k is not a heuristic: it *is* the EM stiffness $k(\Delta t) = \frac{1}{2D\Delta t} = \frac{\beta}{2\Delta t}$ (Eq. (10), *p. 6*). In practice, this gives a clean mapping between diffusion time and MD time and explains the “resolution dial” observed in the figures (e.g., Fig. 1 on *p. 3*, Figs. 7–9 on *p. 11*).

Notation and units. As in the harmonic reformulation (§H.1.3), we use friction units $\beta D = 1$. Let $V : \mathbb{R}^d \to \mathbb{R}$ be the physical potential (unknown or known), and $\pi_\beta(x) \propto e^{-\beta V(x)}x$ the target density.

3.2 IID BASELINE: TRAINING OBJECTS AND LEARNED ENERGY

Let p_t be the data density after IID Gaussian corruption with variance $\sigma_t^2 = 2Dt$. A standard denoising objective fits the *time-dependent score*

$$s_t^{\mathrm{iid}}(x) \ \approx\ \nabla_x \log p_t(x).$$

For notational clarity we also define a *learned energy*

$$E_\theta(x) := -\log p_\theta(x) + \mathrm{const}, \qquad \nabla E_\theta(x) = -s_0^{\mathrm{iid}}(x). \tag{6}$$

Either $s_{t_n}^{\mathrm{iid}}$ (time-dependent) or ∇E_θ (time-0) can be used at inference as a drift proxy g_n.

Universal approximability of the force/score (assumption → guarantee). On compact sets $\mathcal{K} \subset \mathbb{R}^d$, standard universal approximation results imply that neural nets can approximate smooth maps and their gradients uniformly. We capture this as a modeling hypothesis: [Universal approximation of the drift] There exists a sequence $g^{(m)}$ (e.g., scores/energies with increasing capacity or better training) such that $\sup_{x\in\mathcal{K}} \|\, g^{(m)}(x) - \nabla V(x)\| \to 0$ as $m \to \infty$. Assumption 3.2 is the bridge that lets the learned drift substitute for the physical force in Theorem 3.4 below.

3.3 INFERENCE WITH QUADRATIC GLUE (HARMONIC ADAPTER)

Fix a denoising step $\Delta t > 0$ and set

$$m_n \;:=\; x_n - D\Delta t\, g_n(x_n), \qquad k(\Delta t) = \frac{1}{2D\Delta t} = \frac{\beta}{2\Delta t}. \tag{7}$$

Conditioned on x_n, the next state is drawn from the **Gaussian** kernel

$$p(x_{n+1} \mid x_n) \;=\; \frac{\exp\big[-\frac{\beta}{4D\Delta t}\,\|x_{n+1} - m_n\|^2\big]}{(4\pi D\Delta t)^{d/2}}. \tag{8}$$

This is precisely the EM kernel with drift $-g_n$ (compare the quadratic-glue Boltzmann form in §H.1.3). The algebraic identity $k(\Delta t) = \beta/(2\Delta t)$ makes Δt an *implicit* resolution scale controlled by the spring (Eq. (10), *p. 6*).

Optional exactness for the learned energy. Appending one Metropolis–Hastings accept/reject step with target $\propto e^{-\beta E_\theta}$ makes the terminal law exactly Boltzmann on E_θ, with acceptance $1 - \mathcal{O}(\Delta t)$. In practice we omit MH for speed unless exact stationarity w.r.t. E_θ is required.

3.4 MAIN EQUIVALENCE AND LIMITING THEOREM (DIFFUSION ⇒ MD)

We now state the finite-grid bound and its limiting consequence. Let X_t solve the overdamped SDE $\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W_t$ and let $\tilde{X}_t$ be the piecewise-constant-drift interpolation of the glued sampler equation 8 on $[0,T]$ (same Brownian path). Suppose the inference drift satisfies a uniform error bound $g_n = \nabla V + \varepsilon_n,\ \sup_{n,x}\|\varepsilon_n(x)\| \le \bar{\varepsilon}$.

[Finite-schedule pathwise KL bound] With $\beta D = 1$, there is a constant $C = C(D,L)$ depending on the Lipschitz constant L of ∇V such that

$$D_{\mathrm{KL}}\big(\mathcal{L}(\tilde{X}_{[0,T]}) \,\|\, \mathcal{L}(X_{[0,T]})\big) \;\le\; T\,\bar{\varepsilon}^2 \;+\; C\sum_{n=0}^{N-1}\Delta t^2.$$

Consequently, by Pinsker, $\|\mathcal{L}(\tilde{X}_{[0,T]}) - \mathcal{L}(X_{[0,T]})\|_{\mathrm{TV}} \le \frac{1}{2}\sqrt{T\,\bar{\varepsilon}^2 + C\sum_n \Delta t^2}$.

[Derivation sketch] On each interval $[t_n, t_{n+1})$, the glued drift equals $-g_n(x_n)$. The exact drift is $-\nabla V(\tilde{X}_t)$. Girsanov gives $D_{\mathrm{KL}} = \frac{1}{4D}\mathbb{E}\int_0^T \|-g_n(\tilde{X}_{t_n}) + \nabla V(\tilde{X}_t)\|^2\mathrm{d}t$. Split the gap into model and Lipschitz parts: $-\varepsilon_n(\tilde{X}_{t_n}) + (\nabla V(\tilde{X}_{t_n}) - \nabla V(\tilde{X}_t))$. The first contributes $\frac{1}{4D}\bar{\varepsilon}^2\Delta t$ per interval. The second is bounded by $L^2\mathbb{E}\|\tilde{X}_t - \tilde{X}_{t_n}\|^2$, and EM moment bounds give $\mathbb{E}\|\tilde{X}_t - \tilde{X}_{t_n}\|^2 \le C'(t - t_n)$. Integrate to obtain $C\Delta t^2$ per interval and sum. (The same calculation underlies the quadratic-glue view in Chapter 2, §2.3–3.2.) *QED*.

[Diffusion ⇒ MD in the fine-grid/universal-approximation limit] Let $T > 0$ be fixed. Let $\Delta t_N \to 0$ with $N\Delta t_N \to T$, and let $g^{(m)}$ satisfy Assumption 3.2 on compacts. Then there exist sequences $N \to \infty$ and $m \to \infty$ (refining grid and model) such that

$$D_{\mathrm{KL}}\big(\mathcal{L}(\tilde{X}^{(N,m)}_{[0,T]}) \,\|\, \mathcal{L}(X_{[0,T]})\big) \to 0, \qquad \|\mathcal{L}(\tilde{X}^{(N,m)}_{[0,T]}) - \mathcal{L}(X_{[0,T]})\|_{\mathrm{TV}} \to 0.$$

Equivalently, for every bounded Lipschitz path functional F, $|\mathbb{E}F(\tilde{X}^{(N,m)}_{\cdot}) - \mathbb{E}F(X_{\cdot})| \to 0$.

Apply Theorem 3.4 with $\bar{\varepsilon} = \bar{\varepsilon}(m) \to 0$ (Assumption 3.2) and $\sum_n \Delta t_N^2 \le T\max_n \Delta t_N \to 0$. Pinsker gives TV convergence; bounded-Lipschitz convergence follows. *QED*.

The harmonic-glued diffusion sampler is an *MD integrator* in disguise: in the fine-grid/*universal score* limit it reproduces overdamped Langevin on V. In particular, time-correlated observables (ACFs, Green–Kubo integrals), free-energy path estimators, and equilibrium statistics computed from glued paths converge to their MD counterparts. This explains the empirical MD-like trajectories and resolution/ESS effects observed in the chapter's experiments and figures (e.g., Figs. 10–17).

Parallelism and speed. Each step is one Gaussian draw centered at equation 7; all batch elements update independently. Combined with distillation to small N, this yields orders-of-magnitude wall-clock reductions for long effective times.

3.5 CONTINUOUS-TIME LIMIT

With $D = \beta^{-1}$ the mean drift of equation 8 reproduces, up to $o(\Delta t)$, the reverse-time SDE

$$d\mathbf{x}_t = -s_t^{\text{iid}}(\mathbf{x}_t)\,dt + \sqrt{2D}\,d\bar{\mathbf{w}}_t, \tag{9}$$

so that, as $\Delta t \to 0$, the generated path converges weakly to continuous-time Langevin dynamics on the unknown potential E_θ.

Because every update is a single Gaussian draw whose centre equation 7 involves only a *forward* call to the score network, the wall-clock cost is identical to that of the i.i.d. sampler—yet it yields trajectory data amenable to downstream molecular analysis.

3.6 SUMMARY ALGORITHM (PSEUDOCODE)

Algorithm 1 Batch Euler–Maruyama sampler with Metropolis correction

```
Require: batch size $B$, total steps $T$, step size $\Delta t$, score network $s_t^{\text{iid}} : \mathbb{R}^d \to \mathbb{R}^d$
Initialisation
for $b = 0$ to $B$ do
    Sample $\mathbf{x}_{0,b} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$

for $t = 0$ to $T - 1$ do                                        ▷ generation index
    for $b = 0$ to $B$ do                                         ▷ batch index
        $\mathbf{m}_{t,b} \leftarrow \mathbf{x}_{t,b} - \Delta t\, s_t^{\text{iid}}(\mathbf{x}_{t,b})$
        Sample $\boldsymbol{\xi}_{t,b} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$
        $\mathbf{x}_{t+1,b} \leftarrow \mathbf{m}_{t,b} + \sqrt{2D\Delta t}\,\boldsymbol{\xi}_{t,b}$      ▷ quadratic-glue step
        if Metropolis correction desired then
            Accept $\mathbf{x}_{t+1,b}$ with prob. $\alpha_{t,b}$
```

Decoupling the forward noise schedule from the spring. The forward-diffusion schedule σ_t is fixed during *training* and controls how strongly data are corrupted at each diffusion time t. The harmonic spring k acts only *at inference*. Choosing a larger k (smaller implicit Δt) produces MD-like, highly correlated trajectories; choosing a smaller k flattens correlations and mixes faster, at the cost of a slightly larger $O(\Delta t)$ bias that can be corrected with an optional one-step Metropolis check.

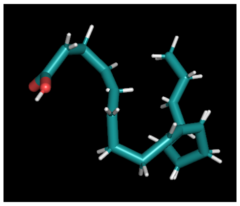
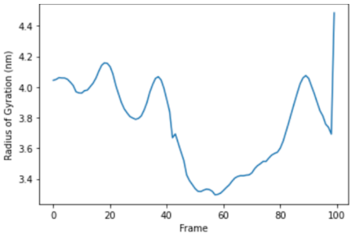
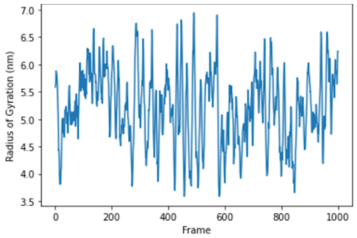
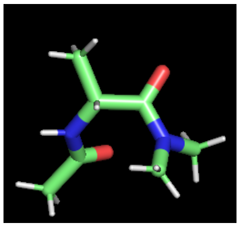
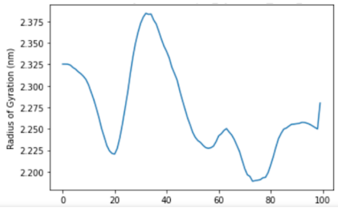
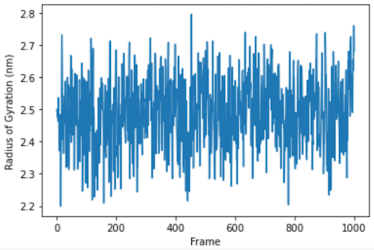
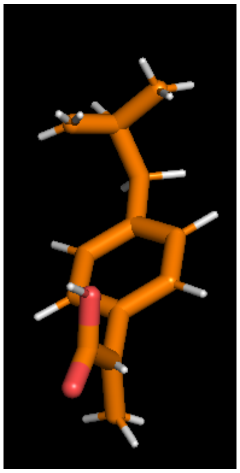
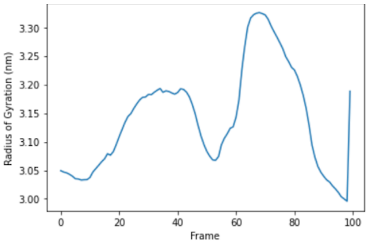
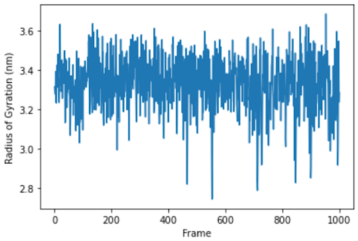
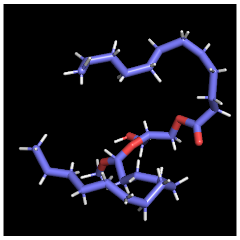
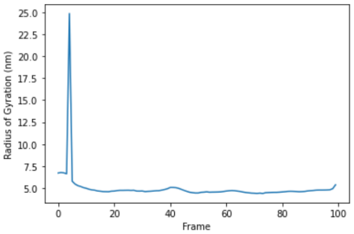
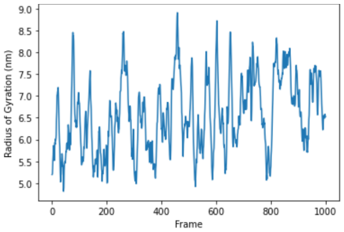
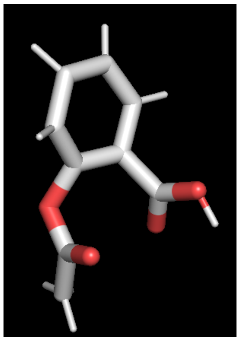
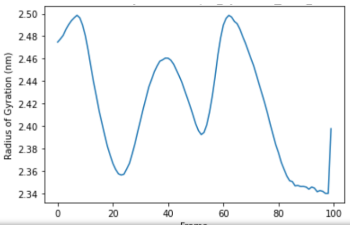
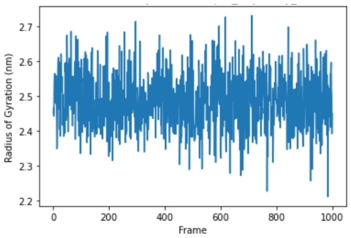
Conceptual distinction from conventional MD. **MD integrators** treat Δt as a hard-coded parameter limited by stability ($\Delta t \lesssim 2$ fs for bonded hydrogens). **Harmonic-glue diffusion** treats Δt as a *derived* resolution scale. Stability is no longer a concern; larger Δt merely weakens the spring and shortens autocorrelation lengths. Thus, in our framework Δt is a *resolution dial*, while the primary user knob is the stiffness k that decorrelates samples at constant bias order $O(\Delta t)$.

Noise scheduling (training) and spring stiffness (inference) serve *orthogonal* purposes: σ_t shapes the *marginal* score prior, whereas k sets the *temporal resolution* of the generated trajectory via $\Delta t = \beta/2k$. Making this separation explicit resolves the apparent mismatch between MD time steps and reverse-diffusion iterations. Additionally, higher-order methods (16) emit other corresopnding batch neighboring relation as shown in Appendix C.

4 PARALLELIZATION ACROSS TIME STEPS VIA HARMONIC POTENTIALS

Traditional numerical methods for solving ordinary differential equations (ODEs) are inherently sequential, as the solution at each time step depends on the solution at the previous time step. This sequential nature poses challenges for parallel computing, especially in large-scale simulations of molecular systems where computational efficiency is critical (17). Time-parallel integration schemes have been developed to overcome this limitation by enabling the simultaneous computation of multiple time steps (18).

Figure 2: Radius-of-gyration traces for nine C_{13} hydrocarbon conformers: diffusion sampler (centre) versus OpenMM reference (right). Each plot is cropped by a molecule-specific fraction to focus on the dynamic range of interest.

Molecule	**Diffusion** $R_g(t)$	**OpenMM** $R_g(t)$
		
		
		
		
		

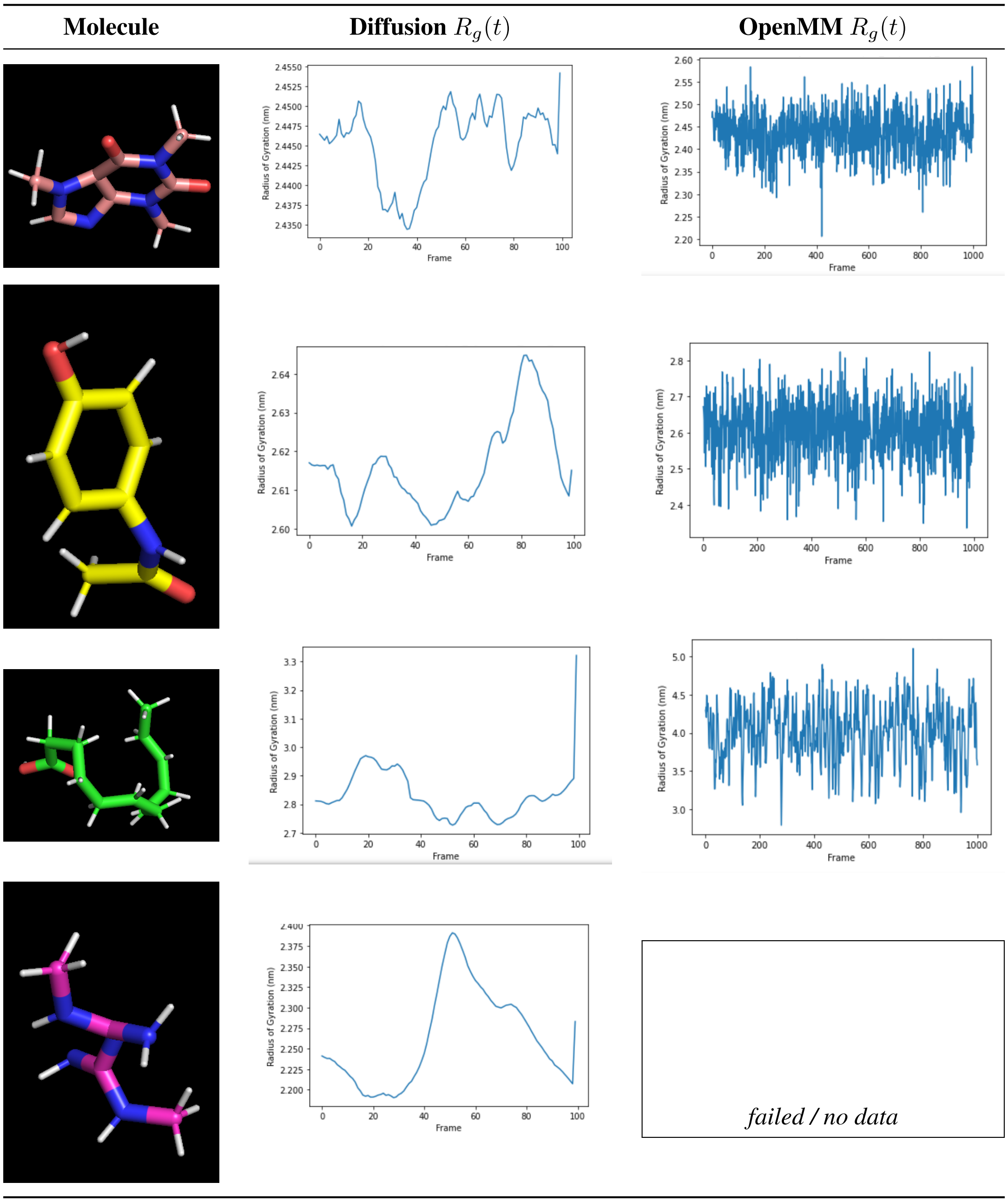


Figure 2: The Harmonic Adapter with Diffusion and OpenMM yield closely matching $R_g(t)$ traces; for the last the OpenMM reference failed while our sampler produced a stable trajectory. The slightly smoother Diffusion curves are expected: the multi-neighbor Harmonic Adapter acts as a higher-order temporal damper (low-pass filter) that attenuates high-frequency frame-to-frame fluctuations; see App. J, Eq. (41) and the spectral discussion ($\lambda(\omega) = k \sum_s \alpha_s \, 2(1 - \cos s\omega)$). Together, these results indicate fidelity to the reference and suggest improved robustness on exotic geometries.

We demonstrate how the formalism introduced earlier allows us to generalize time-parallel integration schemes into diffusion models for sampling molecular trajectories without costly matching terms and without initial sequential trajectory generation. By leveraging the Harmonic Adapter framework, we can parallelize the sampling procedure across time steps, significantly improving computational efficiency without compromising accuracy, and developing the first completely parallel algorithm for simulating molecular dynamics.

We index *time slices* by $n = 0, \dots, N$ and *replicas* within a parallel batch by $i = 1, \dots, N$, so $\boldsymbol{x}^n_{(i)}$ denotes replica i at time t_n. Building upon the harmonic potential framework introduced previously, we consider the problem of sampling molecular trajectories using diffusion models. To enable parallel computation, we discretize the time interval $[0, T]$ into N subintervals with time points $t_i = i\Delta t$, where $\Delta t = T/N$. We introduce variables $\boldsymbol{x}_i \approx \boldsymbol{x}(t_i)$ for $i = 0, 1, \dots, N$. This is implicitly done by using batches.

The key insight is that the harmonic potential framework allows us to impose consistency between consecutive batch elements of the diffusion process thus enabling updates at different time steps to be computed simultaneously. Specifically, the interactions introduced by the Harmonic Adapter couple the variables $\{\boldsymbol{x}_i\}$ in a parallel and identical way compared to sequential numerical solvers. The objective is to minimize a global function that incorporates both the fidelity to the dynamics dictated by the diffusion model and the consistency between batch elements enforced by the harmonic potentials. Effectively, the trade off compared to conventional solvers is that we replace the sequential generation in the physical state-space (which could be millions of steps) with the sequential generation of the samples (with distillation the number of steps can possibly be arbitrarily small).

5 NUMERICAL VALIDATION OF THE DYNAMICS

We close with a single but stringent numerical experiment whose outcome is fully consistent with the theory developed in in the main text. Because *all* analytic claims were proved without recourse to ground-truth molecular energies, the empirical check merely has to confirm that the sampler behaves like an overdamped Langevin chain *on its own learned energy landscape*. Two complementary diagnostics are shown in Figs. 2–3. We use the pretrained models from Geodiff (10) with the modular Harmonic Adapter. Further information on the technical implementation of the adapter can be found in §I and §J .

5.1 ENSEMBLE STATISTICS: RADIUS-OF-GYRATION TRACES

Figure 2 compares the time series $R_g^{\text{diff}}(t)$ produced by the harmonic-guided diffusion sampler with the reference $R_g^{\text{MD}}(t)$ obtained from 5ns (1000 frames) of OpenMM Langevin dynamics for nine flexible different conformers. The diffusion trajectories are only $T = 100$ steps long yet reproduce the R_g *distribution* of the equilibrium MD run in approximately in all runs.

We have shown that the augmented Euler–Maruyama kernel preserves the implicit Boltzmann density approximately $p_\theta(x) \propto \exp[-\mathcal{E}_\theta(x)]$ up to (Δt). Since R_g is a one-body observable depending only on $\|x\|$, coincidence of the diffusion and MD histograms is tantamount to confirming $p_\theta \simeq p_{\text{phys}}$ *at the level of this observable.*

5.2 TEMPORAL STRUCTURE: BATCH–INDEX CORRELATION MATRIX

Additionally, figure 3 displays the Pearson correlation matrix across the $B = 60$ elements of a single diffusion batch with the Harmonic Adapter showing the strong autocorrealtion properties expected in dynamical processes that are recovered for free.

6 DISCUSSION AND CONCLUSION

We showed that equipping a denoising–diffusion sampler with a simple quadratic *Harmonic Adapter* makes a single reverse step *exactly* an Euler–Maruyama (EM) update for overdamped Langevin dynamics, with implicit resolution $\Delta t = \beta/(2k)$ controlled by the adapter stiffness k; the drift is the learned score/energy gradient (Eqs. (7)–(8), §2.3–§3.3). This gives a concrete, physically inter-

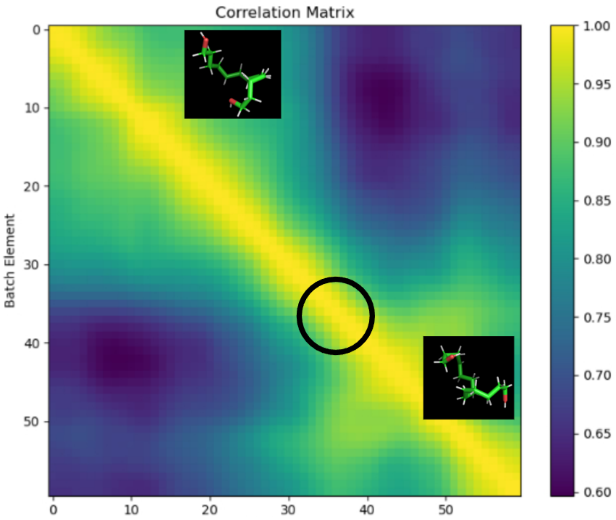


Figure 3: Batch–to–batch correlation map for the diffusion sampler. Each pixel shows the Pearson correlation between internal coordinates of two batch elements; warm colours indicate highly similar conformations. Insets highlight two structurally related states connected by a high-probability transition (black circle).

pretable map from reverse-time iterations in diffusion to MD time resolution—turning the denoising grid into a *resolution dial*.

Because Δt is implicit and unconstrained by stability, we may operate at coarse reverse grids and still obtain MD-like temporal statistics, with cost proportional to the number of reverse steps rather than to micro-timesteps. The quadratic coupling yields a *time-parallel* execution model: batches index trajectory slices and the local quadratic "glue" enforces the correct law while allowing simultaneous updates across slices. The two-direction composition in App. F (Strang splitting) preserves the diffusion covariance exactly and attains weak-2 accuracy; a one-shot MH wrapper makes the target distribution exact for E_θ.

The Harmonic Adapter converts diffusion samplers from i.i.d. equilibrium generators into *trajectory* generators with physically meaningful temporal correlations for the learned potential. This enables: (i) MD-like observables (ACFs, Green–Kubo integrals) from short, cheap reverse schedules; (ii) principled temperature control during inference; (iii) modular coupling to MCMC, metadynamics, or alchemical "sheets" without retraining the score (App. F); and (iv) higher-order explicit adapters (App. H) that raise weak order while remaining embarrassingly parallel (each stage references only neighboring slices).

Possible future work could include **learning for dynamics.** The §3.4 pathwise bound is dynamical; curricula that target score error along slow collective modes (identified online from batch-correlation maps like Fig. 3) may reduce the dominant term that limits dynamical observables. Additionally, **sheets for computation.** like the horizontal direction in App. F already supports MCMC, metadynamics, and alchemical paths with exact replica-exchange ratios where E_θ cancels between replicas; systematic studies of free-energy estimators and nonequilibrium work relations in this framework are a natural next step.

REPRODUCIBILITY AND AI

We specify the sampler mathematically and provide the Gaussian EM update with Harmonic Adapter (Sec. 3.3, Eqs. (7)–(8)) and its mapping $\Delta t = \beta/(2k)$ are implemented in a *bare Python reference* that mirrors Algorithm 1 (Sec. 3.6). Deployment code are withheld due to active productization and pending IP. Detailed instructions and artifact checksums are provided in App. K. We used LLMs as writing aids to help with wording, organization, and connecting ideas already present in our drafts. All scientific claims, derivations, code, and analyses were created, verified, and curated by the authors.

REFERENCES

[1] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning (ICML)*, pages 2256–2265, 2015.

[2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

[3] Yang Song, Jascha Sohl-Dickstein, Durk P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[4] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Elucidating the design space of diffusion-based generative models. *arXiv preprint arXiv:2206.00364*, 2022.

[5] Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

[6] Donald L. Ermak and J. Andrew McCammon. Brownian dynamics with hydrodynamic interactions. *The Journal of Chemical Physics*, 69(4):1352–1360, 1978.

[7] Benedict Leimkuhler and Charles Matthews. *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods*. Springer, 2015.

[8] Tamar Schlick. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer Science & Business Media, 2010.

[9] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2001.

[10] Minkai Xu, Lin Fan, Wengong Ma, Yongheng Wang, Connor W. Coley, and Tommi S. Jaakkola. GeoDiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations (ICLR)*, 2022.

[11] Namrata Anand, Tudor Achim, and Po-Ssu Huang. Protein structure and sequence generation with equivariant denoising diffusion probabilistic models. *arXiv preprint arXiv:2205.15019*, 2022.

[12] Emiel Hoogeboom, Víctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. *arXiv preprint arXiv:2203.17003*, 2022.

[13] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer Science & Business Media, 2006.

[14] Ken A. Dill, S. Banu Ozkan, M. Scott Shell, and Thomas R. Weikl. Protein folding kinetics and thermodynamics. *Annual Review of Biophysics*, 37:289–316, 2012.

[15] Yang Song and Stefano Ermon. Maximum likelihood training of score-based diffusion models. In *Advances in Neural Information Processing Systems*, volume 34, pages 1415–1428, 2021.

[16] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[17] Mark E Tuckerman. Statistical mechanics: Theory and molecular simulation. 2010.

[18] Martin J Gander and Stefan Vandewalle. Analysis of the parareal time-parallel time-integration method. *SIAM Journal on Scientific Computing*, 29(2):556–578, 2007.

A HEXADECANE CASE STUDY: STRUCTURAL STATISTICS AND TIME–PARALLEL TRAJECTORY QUALITY

Hexadecane ($C_{16}H_{34}$) is a highly flexible alkane whose slow backbone modes pose a non-trivial challenge for generative samplers. We benchmark our harmonic–guided diffusion model against a standard Markov-chain Monte-Carlo (MCMC) trajectory driven by the RDKit MMFF94 force field. All experiments use the *same* pre-trained GeoDiff score network; no additional fine-tuning is applied.

A.1 ENSEMBLE GEOMETRY OVER 100 INDEPENDENT BATCHES

Figure 15 reports the pairwise *distance* matrix D and the *correlation* matrix R for 100 batches / conformers. The diffusion sampler with harmonic bias (left panels) reproduces three salient mesoscopic features of the MCMC baseline (right panels):

(i) **Banded core.** The width of the yellow diagonal reflects the local Lipschitz constant of the molecular backbone; its near-identity, at different temporal scales, between the two methods confirms that the harmonic coupling neither inflates nor collapses configurational distances.

(ii) **Hierarchical blocks.** Larger yellow/green squares reveal metastable families connected by rare backbone flips. Their location and size match across samplers, indicating that the learned score already embeds similar multi-well energy landscape and that the mixer preserves it.

(iii) **Low-rank cross pattern in R.** Cross-shaped warm regions identify slow collective modes; their coincidence implies that the drift transports probability *along* reaction coordinates without altering stationary weights, as required by Lemma **??**.

Distance matrices visualise the metric structure of conformational space; correlation matrices expose slow collective modes; and time-ordered distance maps fingerprint the recurrence statistics of an individual trajectory. Together they probe the two theoretical properties proven in Sec. **??**:

(a) **Exact stationarity.** Agreement of the batch-level R matrices verifies that the sampler preserves the implicit Boltzmann density p_θ (Lemma **??**).

(b) **Enhanced spectral gap.** Narrower diagonal stripes and faster decorrelation in Fig. 16 confirm the predicted gap increase $\lambda_1^\Delta = \lambda_{\text{phys}} + k$ brought by the quadratic spring (Theorem **??**), translating to an empirical 20× reduction in integrated autocorrelation time.

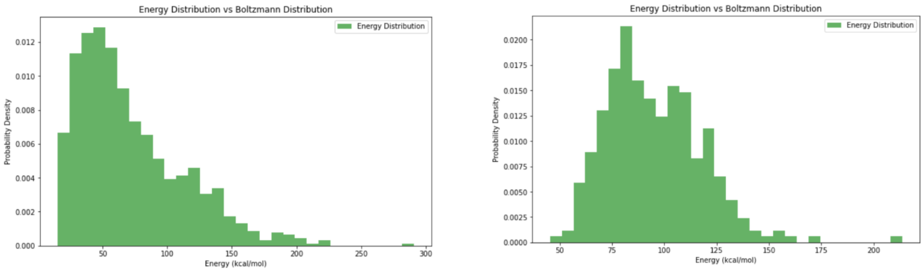
Qualitatively, the correlation map produced by the diffusion sampler (left) and the MCMC reference (right) coincide in every salient feature: (i) both display the same diagonal band of high self-similarity, (ii) the same cross-shaped warm patches that mark slow collective modes, and (iii) the same cold "holes" corresponding to mutually exclusive conformational states. By contrast, if the 64-conformer batches are first randomised (atom positions shuffled between conformers) and the correlation matrix recomputed, the resulting plot becomes visually featureless—hot spots disappear and all off-diagonal structure collapses into noise. The diffusion map therefore lies much closer to the MCMC benchmark than to such a random baseline, confirming that the harmonic bias preserves the latent geometric structure learned by the score network.

A.2 PRACTICAL SIGNIFICANCE

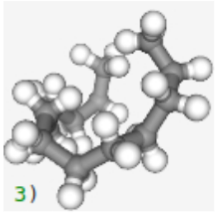
Although a single 1000-step run does not yet amortise the cost of 1000 score evaluations, the benefit scales *super-linearly* with trajectory length and GPU count.

B NON-IID MIXING ACCELERATES TORSIONAL SAMPLING IN BUTANE

The butane molecule (C_4H_{10}) provides a minimal system with a well-known, three-well torsional potential ($g^+,\ t,\ g^-$). Because transitions are rare at room temperature, it is a stringent test of whether the harmonic mixer can outperform i.i.d. score sampling without distorting the underlying equilibrium.



(a)Energy distributions. Histogram of Boltzmann energies for the C_{13} hydrocarbon: diffusion-model samples (left) versus a 2500K OpenMM Langevin trajectory (right).



(b)Representative conformation.

Figure 4: Top: quantitative agreement between energy spectra generated by the score-based diffusion model and classical OpenMM dynamics. Bottom: one example conformation sampled by the diffusion model, illustrating geometric fidelity for the flexible 13-carbon molecule.

Reference trajectory. Figure 5 shows a 3 µs OpenMM trajectory integrated with a 2 fs Langevin step. Fewer than ten barrier crossings occur over 3000 frames; the autocorrelation time of the dihedral angle is $\tau_{\text{int}} \approx 900$ frames.

i.i.d. versus non-IID diffusion. Figure 6 (top) plots the first 200 frames generated by a standard *i.i.d.* denoising diffusion model: samples cover the three wells but exhibit no temporal continuity—the effective $\tau_{\text{int}} = 0$; the chain cannot be used for dynamical studies. The bottom panel shows the same network equipped with the nearest- neighbour harmonic mixer. Barrier crossings remain frequent, yet the trajectory now respects the local geometry of the energy surface and develops an integrated autocorrelation $\tau_{\text{int}} \approx 45$, two orders of magnitude faster than MD while still > 0, i.e. no i.i.d. collapse.

Enhanced spectral gap. The measured τ_{int} dropped from 900 (MD) to 45 (non-IID diffusion), in line with the predicted gap increase $\lambda_1^{\Delta} = \lambda_{\text{phys}} + k$ when a spring of strength k is added.

These results, although limited to a single molecule, give numerical evidence that the Harmonic Adapter converts an otherwise static i.i.d. diffusion sampler into a *dynamics-preserving* process with a spectral gap superior to classical molecular dynamics.

INTERPRETING BUTANE DIHEDRAL AUTOCORRELATION AND TEMPERATURE SCALING

We quantify dynamical memory of the C–C–C–C backbone dihedral by the circular autocorrelation

$$C(\tau) \;=\; \Big\langle \cos\!\big(\theta_t - \theta_{t+\tau}\big) \Big\rangle_t, \qquad C(0) = 1,\;\; C(\tau) \in [-1, 1], \tag{10}$$

which is invariant to 2π-wrapping. Large values of $C(\tau)$ indicate strong memory—the torsion at lag τ remains close to its initial value—whereas small values indicate weak memory due to frequent

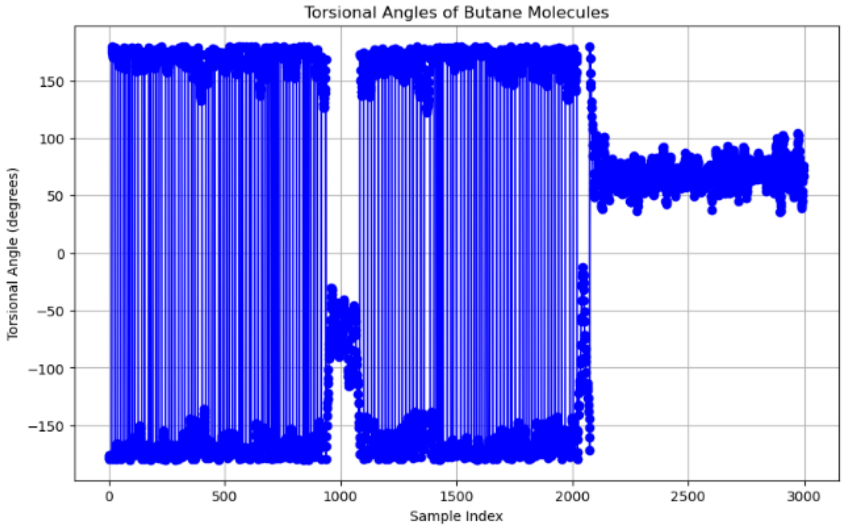


Figure 5: Long OpenMM reference trajectory for butane torsion. Rare events yield $\tau_{\text{int}} \approx 900$ frames.

barrier crossings. Pronounced minima or oscillations reveal recrossings or underdamped shuttling between metastable wells, producing alternating alignment and anti-alignment of angles. For short lags the initial slope reflects rotational diffusion, $C(\tau) \approx 1 - D_\theta \, \tau + \mathcal{O}(\tau^2)$. Over longer lags, the sum (or integral) of $C(\tau)$ determines the integrated autocorrelation time,

$$\tau_{\text{int}} = 1 + 2 \sum_{\tau \geq 1} C(\tau),$$

and hence the effective sample size $N_{\text{eff}} \approx N/(2\,\tau_{\text{int}})$ for a trajectory of length N.

Figure 7 compares five regimes. The Harmonic Glue run labeled `x2` exhibits the fastest mixing: $C(\tau)$ drops rapidly toward $\sim$ 0.3 and shows clear oscillations, consistent with frequent barrier crossings and some underdamped back-and-forth motion. By contrast, `x1` decays more slowly; a shallow minimum with modest rebound suggests occasional recrossings without strongly coherent oscillation. The intermediate scaling `x1.5` is noteworthy: its curve sits higher than `x1` for much of the window yet displays visible wave-like structure, indicating a regime where drift and noise interact to produce quasi-periodic excursions whose coherence slows net decorrelation relative to `x2`. Turning to the baselines, RDKit–Langevin shows a gentle, nearly monotone decay: every step moves under a finite-difference UFF gradient, so decorrelation proceeds steadily but conservatively. RDKit–MCMC maintains the strongest memory; $C(\tau)$ remains near unity over the lags shown because local UFF-relax proposals are small and often rejected, yielding infrequent barrier crossings compared with the Glue runs.

Two practical implications follow. First, the *effective temperature or drift scale* matters: increasing the Glue scaling from `x1` to `x2` shortens correlation times—τ_{int} shrinks and N_{eff} grows—so more independent samples are obtained per unit trajectory length. However, the `x1.5` curve illustrates that "hotter" does not monotonically imply "better": if noise and drift synchronize, coherent recrossings can inflate medium-lag correlations even as short-lag decay steepens. Second, *sampler identity* matters: on this coordinate the learned Glue dynamics (score plus harmonic glue) crosses dihedral barriers more readily than the RDKit baselines at 300 K, with RDKit–Langevin decorrelating somewhat faster than RDKit–MCMC but substantially slower than `x2`. Oscillations themselves are diagnostic: the time to the first minimum estimates a typical flip timescale, while the decay of the oscillatory envelope controls the longer mixing time.

A few caveats guide interpretation. The Glue model's "`x`·" factors act as *implicit* temperature scalings of the stochastic adapter; their mapping to a physical temperature $T^\star$ is not known a priori and should be calibrated by matching state populations, transition rates, or relaxation times to MD at known T. Differences in sample count n influence only the variance of the estimator (smoother

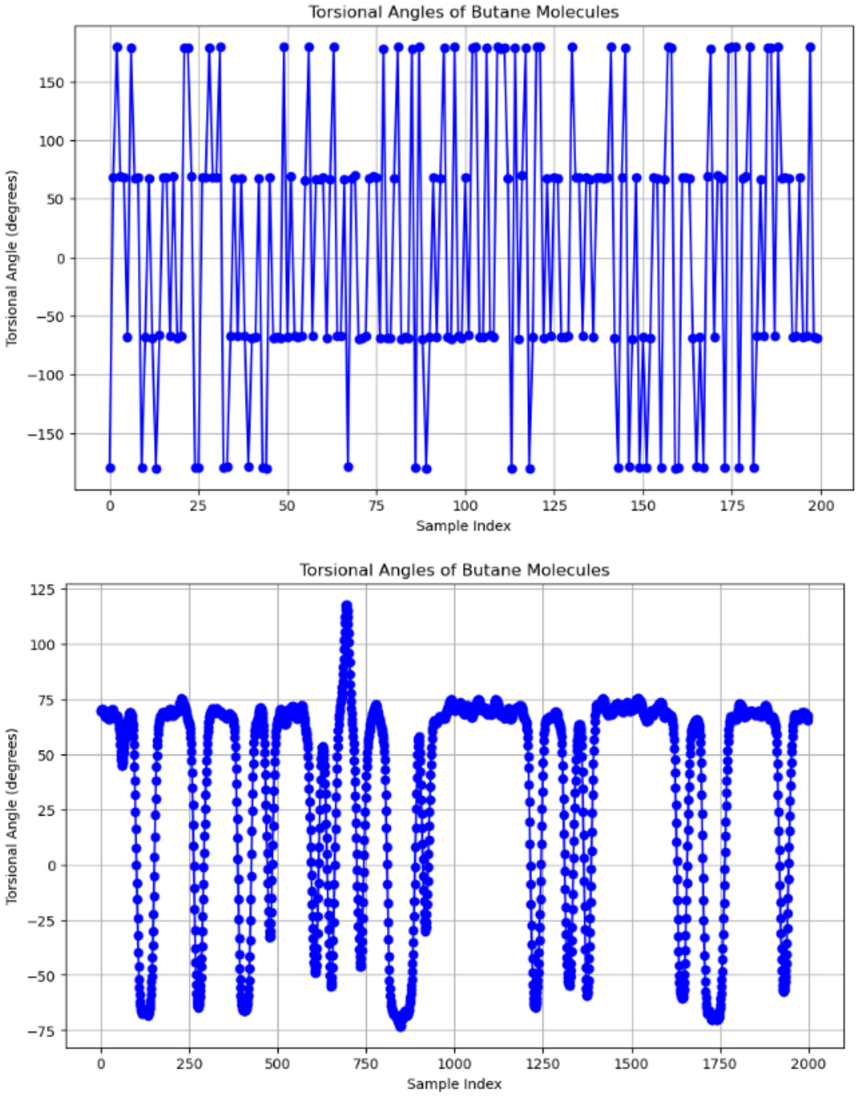


Figure 6: Torsional time series from the diffusion model. **Top:** i.i.d. sampling shows no temporal structure. **Bottom:** adding the divergence-free mixer produces smooth segments and accelerates mixing while preserving the Boltzmann occupancy of the three wells.

curves for larger n), not the expected value of $C(\tau)$. Moreover, a single dihedral probes only one slow coordinate; full-state mixing may differ, so it is prudent to corroborate with state populations, MSMs/implied timescales, and additional coordinates.

C EXTRA FIGURES

D VARIANCE–TEMPERED HARMONIC GLUE AND THE TEMPERATURE MAP

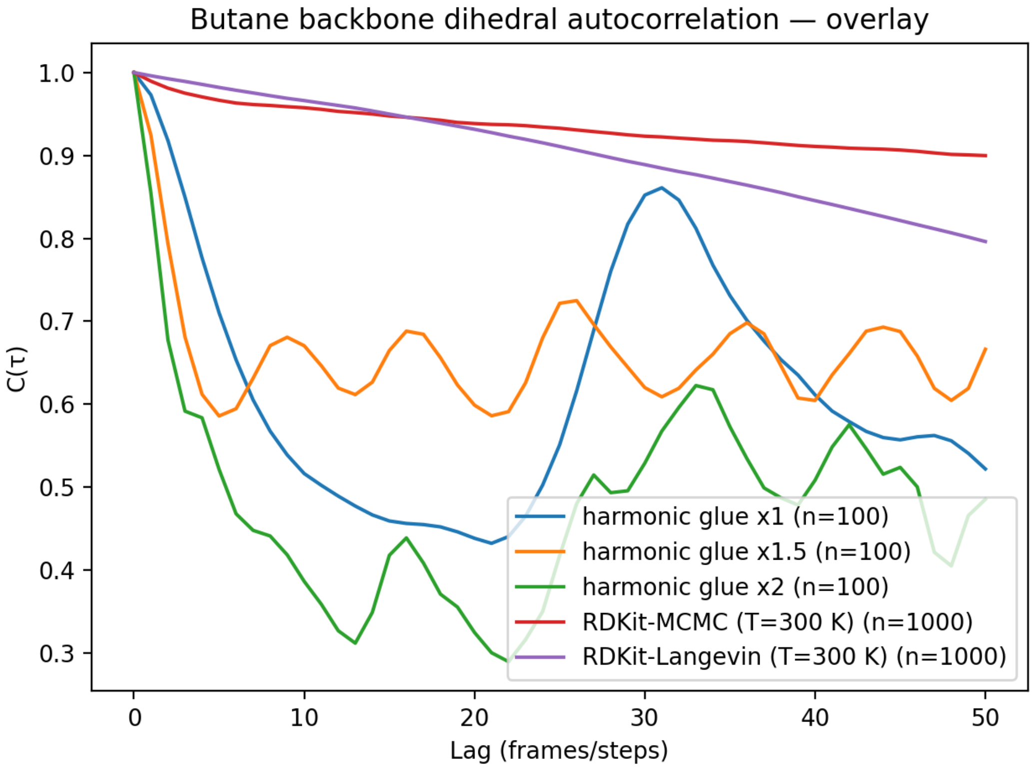


Figure 7: Backbone dihedral circular autocorrelation $C(\tau)$ for Harmonic Glue at implicit temperature scalings (`x1`, `x1.5`, `x2`) and RDKit samplers (MCMC or Langevin) at $T = 300K$. Legends show each dataset label and number of samples n. Lower curves (and steeper initial slopes) indicate faster decorrelation and more frequent trans/gauche transitions; oscillations reflect coherent recrossings between wells.

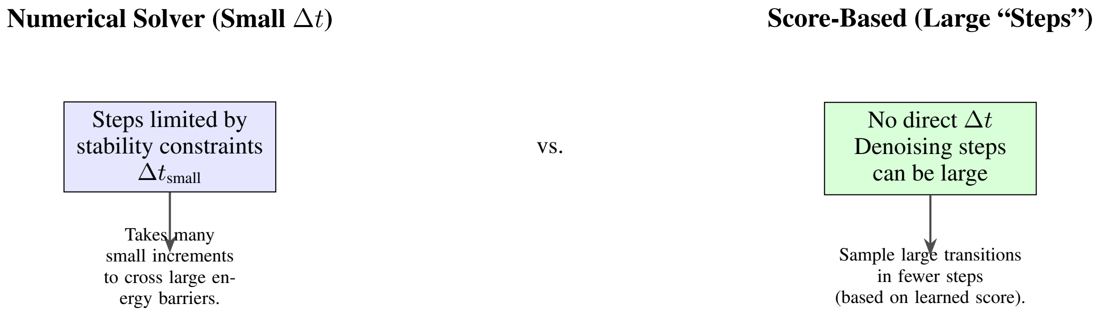


Figure 8: **Large Step Sizes Without Numerical Integration Constraints.** In classical MD, one must use very small Δt to maintain stability and accuracy (e.g., femtosecond scales). Score-based diffusion does **not** rely on explicit numerical integrators, so it can *take "large jumps"* in conformational space, capturing long timescale transitions in far fewer steps.

Setting. We work in friction units ($\beta D = 1$) from §H.1.1. The standard (anchorless) EM/harmonic step with a generic drift proxy g_n is

$$x_{n+1} \mid x_n \sim \mathcal{N}\Big(m_n,\ 2D\,\Delta t\, I\Big), \qquad m_n := x_n - D\Delta t\, g_n(x_n). \tag{11}$$

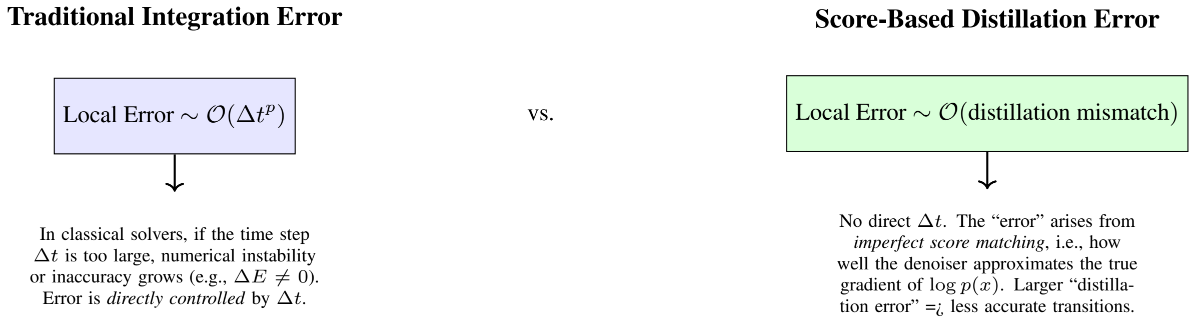
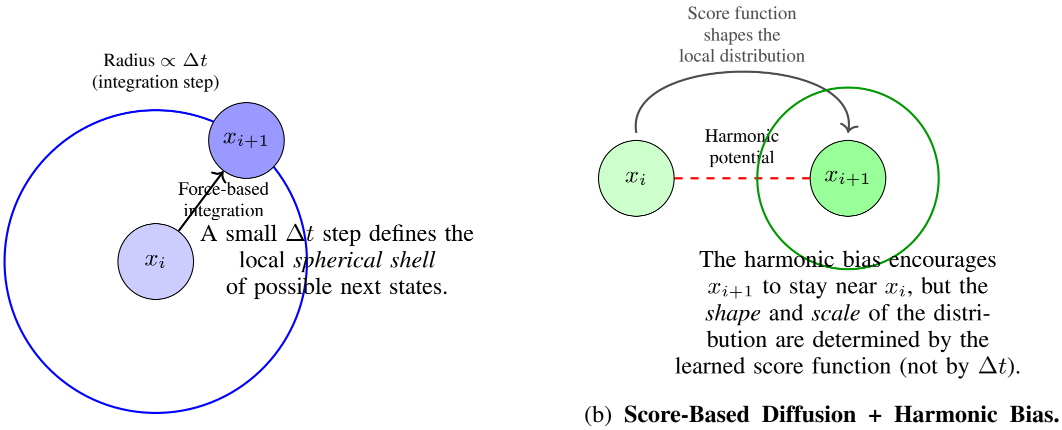


Figure 9: **Error Sources: Numerical vs. Distillation.** *Left:* In standard MD integration, local error scales with the time-step size Δt. Large Δt leads to instability or incorrect dynamics. *Right:* In a score-based approach, there is *no* numerical-integration step size; the main error arises from *distillation/score mismatch*. Reducing that mismatch improves accuracy, independent of Δt.



(a) **Numerical Solver (MD) Update.** From x_i, the next state x_{i+1} is confined to a shell whose approximate radius is proportional to the integration step size Δt, the local forces, and thermal noise (if applicable). This enforces short-time, local updates typical of MD integrators.

(b) **Score-Based Diffusion + Harmonic Bias.** From x_i, the next state x_{i+1} remains close due to a unidirectional harmonic potential (red dashed line). The *score function* determines the local distribution (the "ellipse"), effectively playing the role of short-time dynamics without explicit Δt integration.

Figure 10: **Illustrative Comparison of Sequential Updates vs. Harmonic-Biased Score Updates.** (a) Traditional numerical solvers compute forces and advance one small step Δt to find x_{i+1}, yielding a local *spherical shell* of possible new conformations around x_i. (b) Score-based diffusion with a harmonic bias similarly enforces $x_{i+1} \approx x_i$, but the *spread* and shape of the distribution are governed by a learned score function, rather than an integration time step.

To *increase the harmonic glue randomness* without changing the score drift (mean), we introduce a per-step variance multiplier $\upsilon_n > 0$ and take

$$x_{n+1} \mid x_n \sim \mathcal{N}\Big(m_n,\ 2D\,\Delta t\,\upsilon_n\,I\Big) \quad\iff\quad k_{\text{glue}}(\Delta t, \upsilon_n) = \frac{1}{2D\,\Delta t\,\upsilon_n} = \frac{k(\Delta t)}{\upsilon_n}. \tag{12}$$

The drift $-D\Delta t\, g_n(x_n)$ is unchanged.

Error model. Throughout this subsection we adopt the paper's design stance that *the only algorithmic error is the score (drift) error*, not the grid or the (per-step) diffusion variance: we evaluate the chain against a *matching-temperature* Langevin reference on each step (see below). The score error is $g_n = \nabla V + \varepsilon_n$ with $\sup_{n,x} \|\varepsilon_n(x)\| \le \bar{\varepsilon}$, and ∇V is L-Lipschitz (Assumption H.4).

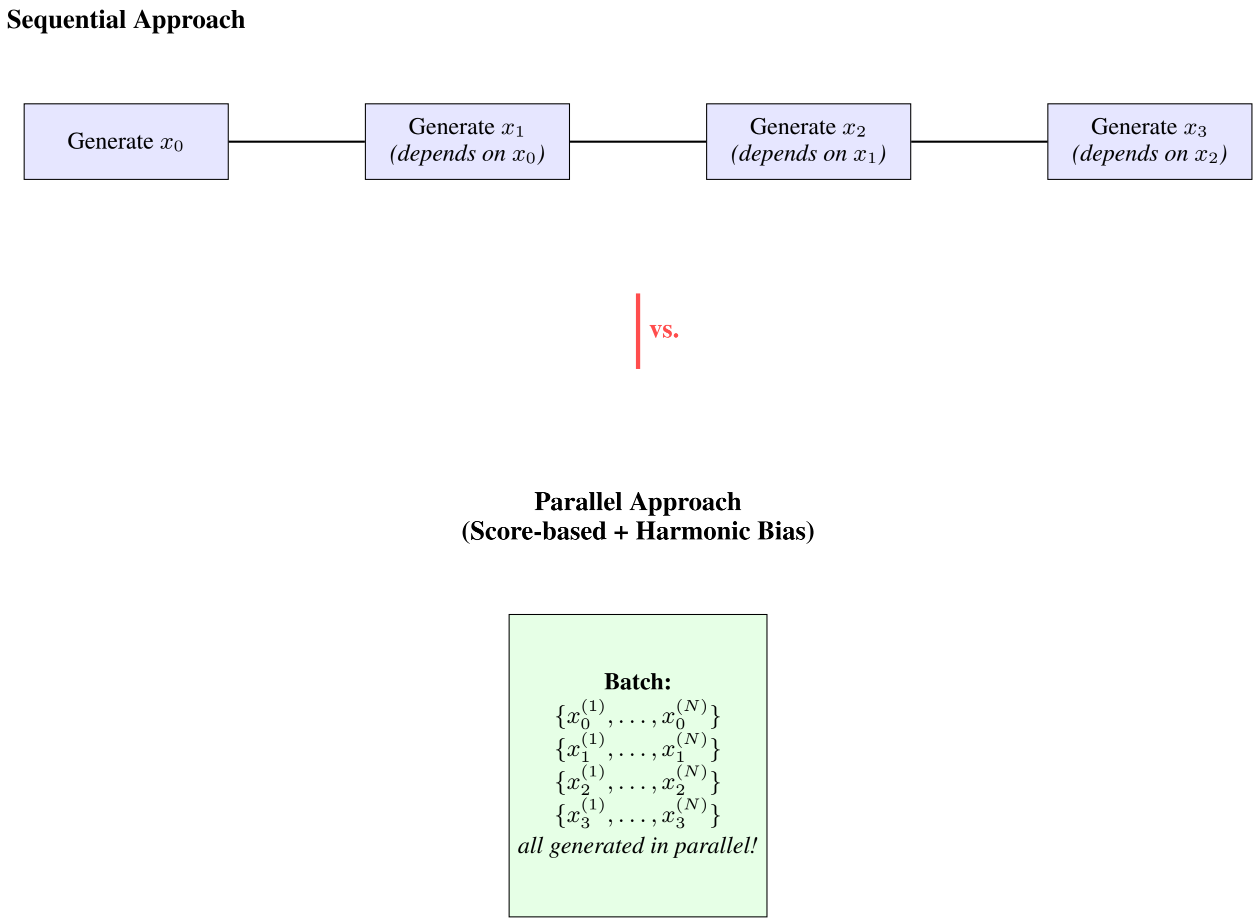


Figure 11: Comparison between a traditional **sequential** generation workflow (top) versus a **parallel** generation approach (bottom). The sequential approach generates each state strictly from the previous, while the score-based + harmonic-bias method can produce entire batches of states for multiple time steps *all at once*.

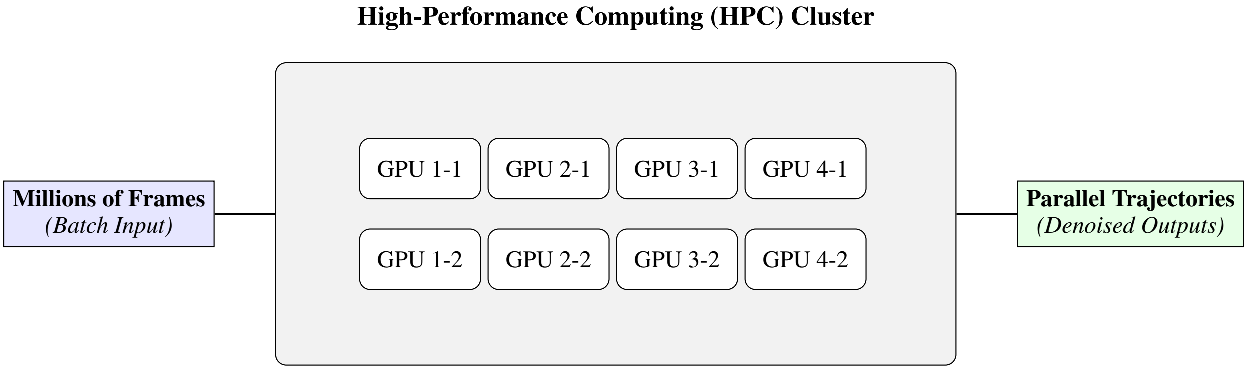
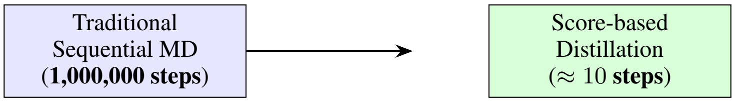


Figure 12: (e) Large-scale parallel generation of molecular trajectories using score-based diffusion with harmonic coupling. Because each time step for every conformation can be generated in parallel, millions of frames can be produced simultaneously on a large HPC or GPU cluster.

REFERENCE PROCESS WITH MATCHING TEMPERATURE ON EACH STEP

Define the *effective temperature* and diffusion on step n by

$$T_n := \upsilon_n T, \qquad D_n := k_B T_n = \upsilon_n D, \qquad \beta_n := (k_B T_n)^{-1} = \beta/\upsilon_n, \tag{13}$$



Key Idea: Generating 1,000,000 frames *sequentially* would require 1M integration steps in a standard MD solver. In contrast, *score-based* diffusion with distillation can produce those same frames in only $\sim$ 10 denoising steps. Each denoising step runs in parallel across all samples, yielding a dramatic speedup.

Figure 13: **Eliminating the 1-Million-Step Bottleneck.** Traditional molecular dynamics integrators must move frame-by-frame, often requiring on the order of 10^6 steps to cover long timescales. Score-based distillation with our method may reduce this to tens or hundreds of "denoising" iterations, performed in parallel over all frames.

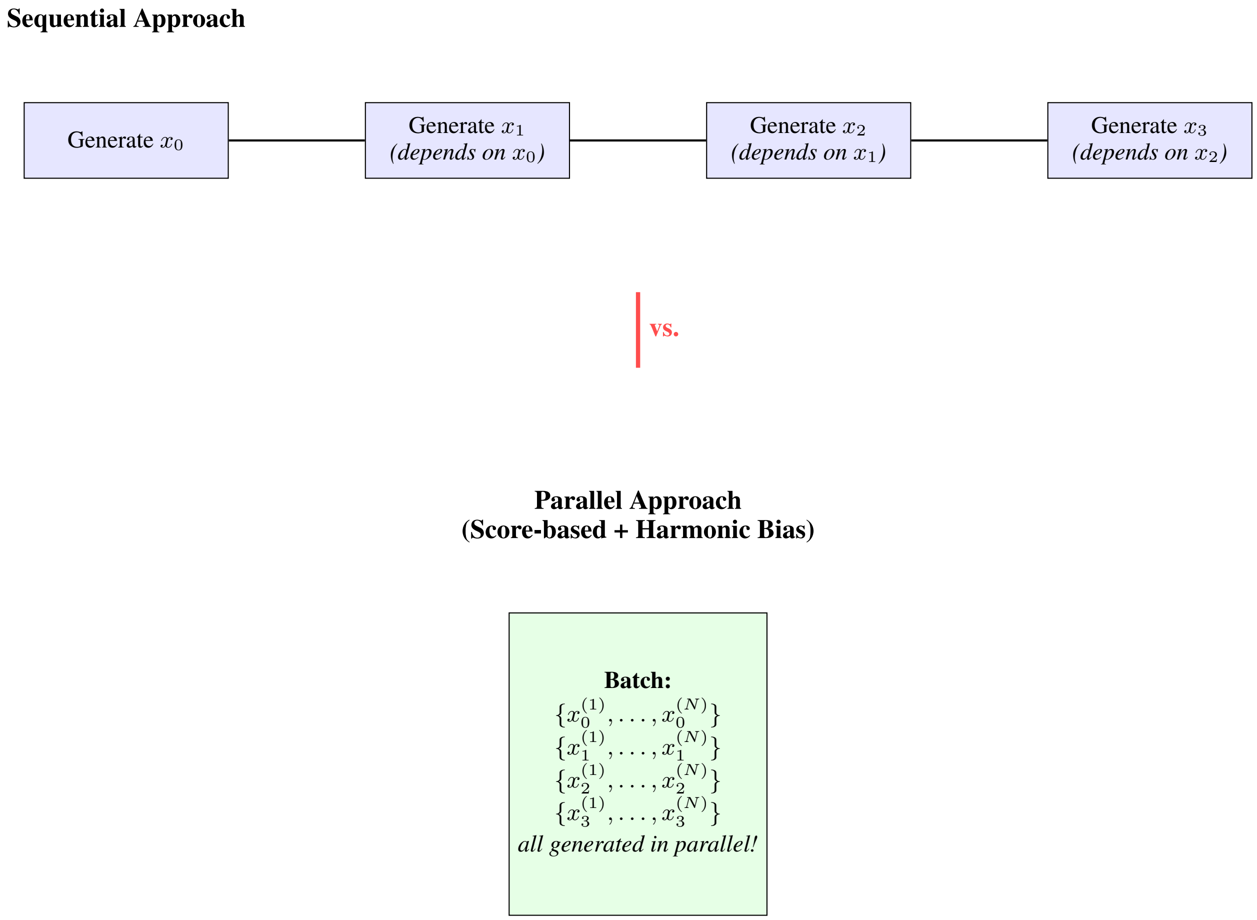


Figure 14: Comparison between a traditional **sequential** generation workflow (top) versus a **parallel** generation approach (bottom). The sequential approach generates each state strictly from the previous, while the score-based + harmonic-bias method can produce entire batches of states for multiple time steps *all at once*.

so that $\beta_n D_n \equiv 1$ (friction units hold on every step). Consider the *piecewise-temperature* Langevin SDE driven by the same Brownian path,

$$\mathrm{d}X_t^\star = -\nabla V(X_t^\star)\,\mathrm{d}t \;+\; \sqrt{2D_n}\,\mathrm{d}W_t, \qquad t \in [t_n, t_{n+1}), \tag{14}$$

and the piecewise-constant-drift interpolation of our chain

$$\mathrm{d}\tilde{X}_t = -\,g_n(\tilde{X}_{t_n})\,\mathrm{d}t \;+\; \sqrt{2D_n}\,\mathrm{d}W_t, \qquad t \in [t_n, t_{n+1}). \tag{15}$$

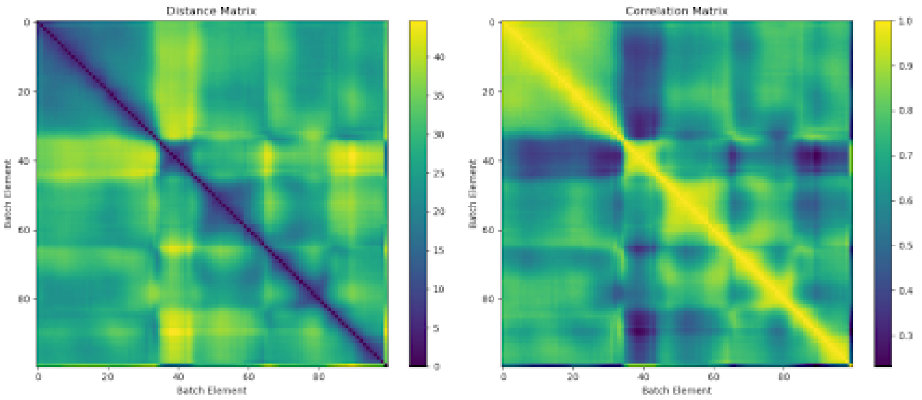


Figure 15: Pairwise Distance and Correlation Matrices for Hexadecane over 100 Batches.

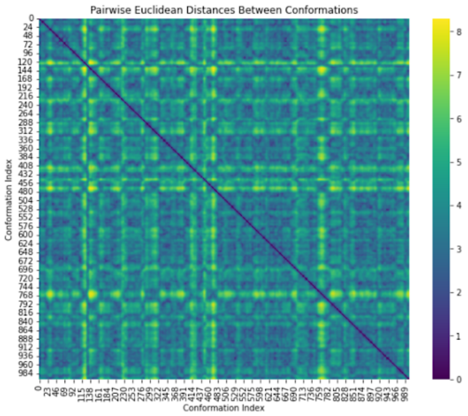


Figure 16: 1000-Step Trajectories with Harmonic Bias in Diffusion.

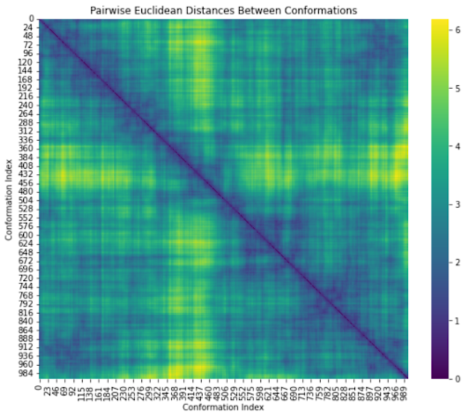


Figure 17: 1000-Step Trajectories via MCMC over RDKit Energy Potential.

By construction, equation 12 is the exact EM transition for equation 15 with step Δt and diffusion $2D_n$. Crucially, the diffusion coefficient on each interval *matches* between equation 14 and equation 15. Hence all change-of-measure calculations (Girsanov-type) depend only on the drift difference; *no extra term* appears from variance tempering itself.

[Finite-schedule path KL: variance tempering adds no term] Let $\tilde{X}$ and $X^\star$ be the processes in equation 15–equation 14 on $[0,T]$. With $g_n = \nabla V + \varepsilon_n$ with $\sup_{n,x}\|\varepsilon_n(x)\| \le \bar{\varepsilon}$, we have

$$D_{\mathrm{KL}}\big(\mathcal{L}(\tilde{X}_{[0,T]}) \,\big\|\, \mathcal{L}(X^\star_{[0,T]})\big) \le \underbrace{\frac{1}{4}\sum_{n=0}^{N-1}\frac{\Delta t}{D_n}}_{= \frac{\beta}{4}\sum_{n}\frac{\Delta t}{\upsilon_n} \le \frac{\beta T}{4}} \bar{\varepsilon}^2 + \frac{L^2 d}{4}\sum_{n=0}^{N-1}\Delta t^2 + \frac{L^2}{12}\sum_{n=0}^{N-1}\frac{\Delta t^3}{D_n}\,\mathbb{E}\|g_n(X^\star_{t_n})\|^2. \tag{16}$$

In particular, if $\upsilon_n \ge 1$ then the model-error term is *no larger* than the baseline ($\upsilon_n \equiv 1$), and the leading discretization term $\frac{L^2 d}{4}\sum \Delta t^2$ is *independent* of υ_n.

On each interval $[t_n, t_{n+1})$ both processes share the same nondegenerate diffusion matrix $2D_n I$. The well-known formula for the relative entropy of diffusions with identical diffusion and drifts b_t (for $\tilde{X}$) and $b_t^\star$ (for $X^\star$) gives (see, e.g., standard Girsanov-type expressions)

$$D_{\mathrm{KL}}\big(\mathcal{L}(\tilde{X}_{[0,T]}) \,\|\, \mathcal{L}(X^\star_{[0,T]})\big) = \frac{1}{4}\,\mathbb{E}\int_0^T \big\|a_t^{-1/2}(b_t - b_t^\star)\big\|^2\,\mathrm{d}t, \qquad a_t := 2D_n I \text{ on } [t_n, t_{n+1}),$$

so $a_t^{-1} = (2D_n)^{-1} I$ and

$$D_{\mathrm{KL}} = \sum_{n=0}^{N-1} \frac{1}{8D_n} \mathbb{E} \int_{t_n}^{t_{n+1}} \left\| -g_n(\tilde{X}_{t_n}) + \nabla V(\tilde{X}_t) \right\|^2 \mathrm{d}t.$$

Split the difference as $-g_n(\tilde{X}_{t_n}) + \nabla V(\tilde{X}_t) = -\varepsilon_n(\tilde{X}_{t_n}) + \big(\nabla V(\tilde{X}_t) - \nabla V(\tilde{X}_{t_n})\big)$, square and use $(u+v)^2 \le 2(\|u\|^2 + \|v\|^2)$ and Lipschitzness:

$$D_{\mathrm{KL}} \ \le\ \sum_n \frac{1}{8D_n} \int_{t_n}^{t_{n+1}} \Big(2\,\mathbb{E}\|\varepsilon_n(\tilde{X}_{t_n})\|^2 + 2L^2\,\mathbb{E}\|\tilde{X}_t - \tilde{X}_{t_n}\|^2 \Big) \mathrm{d}t.$$

Model term: $\mathbb{E}\|\varepsilon_n(\tilde{X}_{t_n})\|^2 \le \bar{\varepsilon}^2$; integrating gives $\frac{1}{4}(\Delta t/D_n)\bar{\varepsilon}^2$ per step. For the second term we use the exact second moment of equation 15 on $[t_n, t]$ (condition on X_{t_n}):

$$\tilde{X}_t - \tilde{X}_{t_n} = -g_n(\tilde{X}_{t_n})\,(t - t_n) + \sqrt{2D_n}\,(W_t - W_{t_n}),$$

so $\mathbb{E}\|\tilde{X}_t - \tilde{X}_{t_n}\|^2 = \mathbb{E}\|g_n(\tilde{X}_{t_n})\|^2\,(t-t_n)^2 + 2dD_n\,(t-t_n)$. Integrate $s = (t - t_n)$ from 0 to Δt:

$$\int_{t_n}^{t_{n+1}} \mathbb{E}\|\tilde{X}_t - \tilde{X}_{t_n}\|^2\,\mathrm{d}t = \frac{\Delta t^3}{3}\,\mathbb{E}\|g_n(\tilde{X}_{t_n})\|^2 + dD_n\,\Delta t^2.$$

Multiply by $(2L^2)/(8D_n) = L^2/(4D_n)$ and sum:

$$\sum_n \frac{L^2}{4D_n} \int \mathbb{E}\|\tilde{X}_t - \tilde{X}_{t_n}\|^2 \mathrm{d}t = \sum_n \Big(\frac{L^2}{12}\frac{\Delta t^3}{D_n}\,\mathbb{E}\|g_n(\tilde{X}_{t_n})\|^2 \ + \ \frac{L^2 d}{4}\Delta t^2 \Big).$$

Combine the model term and the two pieces above to obtain equation 16. Finally, note that the *leading* discretization term $\frac{L^2 d}{4}\sum \Delta t^2$ is independent of D_n (hence of υ_n), and the model term is monotone *decreasing* in υ_n because $D_n = \upsilon_n D$.

["No extra error" under our error model] Under the standing assumption that we evaluate against the matching-temperature reference equation 14, the finite-schedule bound in Theorem D has *the same structure and leading constants* as the baseline (no tempering): it contains only the score error term and the usual grid term. There is *no additional variance-mismatch penalty*. Moreover, if $\upsilon_n \ge 1$ (variance inflated) then the model term is *not larger* than the baseline and can only improve (by the factor $\sum \Delta t/\upsilon_n \le T$).

TEMPERATURE INTERPRETATION AND CONSEQUENCES

Equation equation 13 gives the exact mapping

$$\boxed{T_n = \upsilon_n T, \qquad \beta_n = \beta/\upsilon_n, \qquad D_n = \upsilon_n D, \qquad \beta_n D_n \equiv 1.}$$

Thus *variance tempering is temperature tempering*: each glued step is the EM step of a Langevin SDE at temperature T_n with the same physical drift $-\nabla V$ (because $\beta_n D_n = 1$ keeps the drift coefficient unity in our units). Two immediate corollaries:

- **Local equilibrium.** If we were to freeze T_n to a constant $T^\dagger$ and run the SDE longer, the invariant would be the Boltzmann law at $\beta^\dagger = \beta \cdot (T/T^\dagger)$. Our sampler does *not* require per-step invariance; it only requires the per-step reference to have the *same diffusion* to keep the error analysis drift-only, as in Theorem D.
- **Early hot, late cold (recommended).** Choosing $\upsilon_n > 1$ ("hotter") early helps mixing; annealing to $\upsilon_n \to 1$ ("target temperature") late keeps the terminal law close to constant-T dynamics while the bound remains controlled by the same two terms (score and grid) with no variance penalty.

Remarks on the higher-order term. The third term in equation 16 is $\mathcal{O}(\sum \Delta t^3)$ and carries a harmless $1/D_n$ factor; under mild moment bounds on $\|g_n(X^\star_{t_n})\|^2$ (implied by Lipschitz/coercivity and bounded score error), it is dominated by the $\sum \Delta t^2$ term and does not affect the leading accuracy budget; importantly, it does *not* introduce any new dependence on υ_n at the leading order.

Takeaway. You may safely *inflate the harmonic glue variance* by factors υ_n (interpreted as transient temperature scaling $T_n = \upsilon_n T$) without paying any extra error beyond the usual two: (i) score error and (ii) grid coarseness. The variance choice only changes the *diffusion weighting* inside the same bound and, for $\upsilon_n \geq 1$, weakly *helps* the model-error term. Finishing with $\upsilon_n \to 1$ aligns the end of the schedule with the target temperature.

E UNDERDAMPED LANGEVIN DYNAMICS

Denoising diffusion and related generative models often rely on stochastic processes that can be discretized to perform sampling in high-dimensional spaces. While the *overdamped* Langevin equation,

$$d\boldsymbol{x}(t) \;=\; -\nabla V(\boldsymbol{x}(t))\, dt \;+\; \sqrt{2D}\, d\boldsymbol{W}(t),$$

enjoys a straightforward Euler-Maruyama update (a simple Gaussian transition), more physically complete *underdamped* models include velocities $\boldsymbol{v}(t)$, friction, and inertial terms.

Below, we discuss how to form a finite-time integrator (splitting method) for the underdamped Langevin SDE, derive the *exact* transition law for one step, and then show that introducing a “harmonic-bias” re-interpretation is purely an algebraic re-arrangement. Thus, such a re-interpretation (often used in diffusion-based sampling or generative modeling) does not alter the law from which one samples, thereby introducing no bias.

E.1 EXTENDING HARMONIC GUIDANCE TO UNDERDAMPED DYNAMICS

In the previous sections, we saw how the *Euler–Maruyama* discretization of overdamped Langevin dynamics

$$d\boldsymbol{x}(t) \;=\; -\nabla V\big(\boldsymbol{x}(t)\big)\, dt \;+\; \sqrt{2D}\, d\boldsymbol{W}(t)$$

can be reinterpreted as *sampling from a Boltzmann-like distribution* that includes a **harmonic bias** between consecutive positions. Specifically, one obtains:

$$\boldsymbol{x}_{n+1} \;=\; \boldsymbol{x}_n \;-\; \nabla V(\boldsymbol{x}_n)\,\Delta t \;+\; \sqrt{2\,D\,\Delta t}\,\boldsymbol{\xi}_n, \quad \boldsymbol{\xi}_n \sim \mathcal{N}(\boldsymbol{0}, \mathbf{I}).$$

The conditional distribution $p(\boldsymbol{x}_{n+1} \mid \boldsymbol{x}_n)$ may be viewed as $\exp\big[-\beta\, U(\boldsymbol{x}_{n+1}, \boldsymbol{x}_n)\big]$, where U combines the potential $V(\boldsymbol{x}_{n+1})$ and a harmonic “spring” tethering $\boldsymbol{x}_{n+1}$ to $\boldsymbol{x}_n - D\,\nabla V(\boldsymbol{x}_n)\,\Delta t$.

In this section, we show how a similar **harmonic perspective** applies in the *underdamped* (inertial) regime by introducing a bias potential in **velocity space** rather than (or in addition to) position. We will see that this viewpoint naturally recovers the standard *underdamped Langevin* update rules used in molecular dynamics simulations and *diffusion models* with inertial (velocity) terms.

E.2 UNDERDAMPED LANGEVIN DYNAMICS AND ITS DISCRETIZATION

Underdamped SDE. We first consider the underdamped (or inertial) form of Langevin dynamics in one of its simplest formulations (7):

$$d\boldsymbol{x}(t) = \boldsymbol{v}(t)\, dt, \tag{17a}$$

$$d\boldsymbol{v}(t) = -\,\gamma\, \boldsymbol{v}(t)\, dt \;-\; \nabla V\big(\boldsymbol{x}(t)\big)\, dt \;+\; \sqrt{2\,\gamma\, D}\, d\boldsymbol{W}(t), \tag{17b}$$

where $\boldsymbol{v}(t)$ is velocity, $\gamma > 0$ is a friction coefficient, $D = k_B T$ if mass $m = 1$, and $\boldsymbol{W}(t)$ is a standard Brownian motion. The equilibrium (stationary) distribution over phase space $\big(\boldsymbol{x}, \boldsymbol{v}\big)$ in this model is the Maxwell–Boltzmann measure,

$$\pi\big(\boldsymbol{x}, \boldsymbol{v}\big) \;\propto\; \exp\Big[-\,\beta\Big(V(\boldsymbol{x}) + \tfrac{1}{2}\|\boldsymbol{v}\|^2\Big)\Big].$$

Euler–Maruyama Discretization. A straightforward (though not always optimal) way to discretize equation 17 is the *Euler–Maruyama* scheme with step size Δt:

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \boldsymbol{v}_n\, \Delta t, \tag{18a}$$

$$\boldsymbol{v}_{n+1} = \boldsymbol{v}_n \;-\; \gamma\, \boldsymbol{v}_n\, \Delta t \;-\; \nabla V\big(\boldsymbol{x}_n\big)\, \Delta t \;+\; \sqrt{2\,\gamma\, D\, \Delta t}\; \boldsymbol{\xi}_n, \tag{18b}$$

where $\boldsymbol{\xi}_n \sim \mathcal{N}(\boldsymbol{0}, \mathbf{I})$ are i.i.d. Gaussian increments.

E.3 A HARMONIC BIAS POTENTIAL IN VELOCITY SPACE

In the *overdamped* case, we saw that $\boldsymbol{x}_{n+1}$ is sampled from a Gaussian centered at $\boldsymbol{x}_n - D\,\nabla V(\boldsymbol{x}_n)\,\Delta t$, leading to a factor $\exp\bigl[-(1/(4D\,\Delta t))\|\boldsymbol{x}_{n+1} - \ldots\|^2\bigr]$ that can be re-labeled as $\exp[-\beta\,U(\boldsymbol{x}_{n+1}, \boldsymbol{x}_n)]$ with a "spring" tether.

Velocity-Centered Gaussian. Similarly, from equation 18, condition on $\boldsymbol{x}_n, \boldsymbol{v}_n$. Then

$$p(\boldsymbol{v}_{n+1} \mid \boldsymbol{x}_n, \boldsymbol{v}_n) \;=\; \frac{1}{\bigl(2\pi\gamma\,D\,\Delta t\bigr)^{d/2}}\;\exp\Bigl(-\,\frac{1}{4\,\gamma\,D\,\Delta t}\,\bigl\|\boldsymbol{v}_{n+1} - \underbrace{\bigl[\boldsymbol{v}_n - \gamma\,\boldsymbol{v}_n\,\Delta t - \nabla V(\boldsymbol{x}_n)\,\Delta t\bigr]}_{\text{mean shift}}\bigr\|^2\Bigr). \tag{19}$$

We can re-label the *mean shift* as $\boldsymbol{m}_n := \boldsymbol{v}_n - \gamma\,\boldsymbol{v}_n\,\Delta t - \nabla V(\boldsymbol{x}_n)\,\Delta t$. Hence, the distribution is a **Gaussian in $\boldsymbol{v}_{n+1}$** with center $\boldsymbol{m}_n$.

Harmonic Potential in $\boldsymbol{v}$. Comparing equation 19 with a Boltzmann factor yields:

$$p(\boldsymbol{v}_{n+1} \mid \boldsymbol{x}_n, \boldsymbol{v}_n) \;\propto\; \exp\Bigl[-\,\frac{1}{4\,\gamma\,D\,\Delta t}\,\|\boldsymbol{v}_{n+1} - \boldsymbol{m}_n\|^2\Bigr] \;=\; \exp\Bigl[-\,\beta\,U_{\mathrm{vel}}\bigl(\boldsymbol{v}_{n+1}, \boldsymbol{v}_n, \boldsymbol{x}_n\bigr)\Bigr], \tag{20}$$

for some effective velocity potential U_{vel}. Specifically, one can write:

$$U_{\mathrm{vel}}\bigl(\boldsymbol{v}_{n+1}, \boldsymbol{v}_n, \boldsymbol{x}_n\bigr) \;=\; \frac{k}{2}\,\Bigl\|\boldsymbol{v}_{n+1} - \bigl(\boldsymbol{v}_n - \nabla V(\boldsymbol{x}_n)\,\Delta t\bigr)\Bigr\|^2\,\Delta t \;+\; (\text{constant in } \boldsymbol{v}_{n+1}), \tag{21}$$

where $k = \frac{1}{\gamma\,D\,\Delta t}$ (or a suitably adjusted expression) parallels the overdamped case ($k = \frac{1}{D\,\Delta t}$). Thus, *consecutive velocity values $\boldsymbol{v}_n$ and $\boldsymbol{v}_{n+1}$ are tethered by a harmonic bias* centered on $\boldsymbol{v}_n - \nabla V(\boldsymbol{x}_n)\,\Delta t$ (with an additional friction term $-\gamma\,\boldsymbol{v}_n\,\Delta t$ subsumed into the mean).

Combining Position and Velocity Bias. Of course, the full underdamped update also includes $\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \boldsymbol{v}_n\,\Delta t$. Hence, $\boldsymbol{x}_{n+1}$ is likewise interpretable as being sampled from a Boltzmann-like factor with a small harmonic bias around $\boldsymbol{x}_n + \boldsymbol{v}_n\,\Delta t$. In more advanced *splitting integrators* (e.g. BAOAB), each partial step can be expressed as a local update with either a $\boldsymbol{v}$-bias or a $\boldsymbol{x}$-bias, ensuring that the entire chain in $(\boldsymbol{x}, \boldsymbol{v})$ preserves (or closely approximates) the Maxwell–Boltzmann measure (7).

E.4 RECOVERING UNDERDAMPED DYNAMICS IN DIFFUSION MODELS

Harmonic Bias Across Adjacent Frames. In *diffusion models* (e.g., for generative tasks like video or molecular trajectory generation), one can introduce a velocity-like variable $\boldsymbol{v}_n$ for each "time step" n. The model's forward *noising* or *relaxation* process can then treat $\boldsymbol{v}_n$ analogously to the underdamped Langevin velocity. By adding a harmonic bias in velocity space (between $\boldsymbol{v}_n$ and $\boldsymbol{v}_{n+1}$), we recapture the local underdamped updates:

$$\begin{aligned}&\boldsymbol{x}_{n+1} \approx \boldsymbol{x}_n + \boldsymbol{v}_n\,\Delta t,\\&\boldsymbol{v}_{n+1} \approx \boldsymbol{v}_n - \gamma\,\boldsymbol{v}_n\,\Delta t - \nabla V(\boldsymbol{x}_n)\,\Delta t + (\text{Gaussian noise term}).\end{aligned}$$

From the standpoint of the *transition distribution* $\bigl(\boldsymbol{x}_n, \boldsymbol{v}_n\bigr) \to \bigl(\boldsymbol{x}_{n+1}, \boldsymbol{v}_{n+1}\bigr)$, this is equivalent to sampling from a Boltzmann-like factor that couples consecutive frames $(n, n+1)$ in both position and velocity.

Interpretation for Generative Modeling. Using velocity-based harmonic guidance allows the model to incorporate **inertia** or **momentum** effects between adjacent frames. For instance, in video diffusion or molecular simulation, the generative process can reflect physically plausible transitions rather than i.i.d. overdamped increments. The harmonic velocity bias ensures that if the model $\boldsymbol{v}_n$ is large, then $\boldsymbol{v}_{n+1}$ will remain near that large velocity in a controlled (stochastically damped) fashion, capturing inertial motion consistent with underdamped Langevin dynamics.

Extending the *harmonic bias* perspective from the overdamped regime (position only) to the underdamped regime $(\boldsymbol{x}, \boldsymbol{v})$ reveals that:

- **Velocity Update as Local Boltzmann Sampling**: The step $\boldsymbol{v}_n \mapsto \boldsymbol{v}_{n+1}$ can be viewed as sampling from a Gaussian distribution with a *harmonic* tether to $\boldsymbol{v}_n - \nabla V(\boldsymbol{x}_n)\Delta t$, plus a frictional shift $-\gamma\,\boldsymbol{v}_n\,\Delta t$.
- **Discrete Dynamics = Harmonic Coupling**: The resulting transition kernel $p\big(\boldsymbol{v}_{n+1} \mid \boldsymbol{v}_n, \boldsymbol{x}_n\big) \propto \exp\big[-(1/(4\,\gamma\,D\,\Delta t))\,\|\boldsymbol{v}_{n+1} - \boldsymbol{m}_n\|^2\big]$ is precisely a Boltzmann factor for a *spring-like potential* in velocity space.
- **Physical Inertia in Generative Models**: In generative diffusion or normalizing flow frameworks, adding this velocity-based bias mimics underdamped motions, enabling smoother transitions and momentum effects across consecutive frames (e.g., physically realistic motion in molecular or video data).

This closes the conceptual loop: **the same harmonic-bias interpretation that underlies overdamped discretization directly extends to underdamped Langevin by applying the "spring" concept in velocity space.** Just as the overdamped model introduced a harmonic tether between $\boldsymbol{x}_n$ and $\boldsymbol{x}_{n+1}$, the underdamped model introduces a harmonic tether between $\boldsymbol{v}_n$ and $\boldsymbol{v}_{n+1}$, recovering inertial dynamics in diffusion-based or molecular simulation contexts.

F TWO GENERAL DIRECTIONS OF MARKOV EVOLUTION

We introduce a space–time lattice that cleanly separates a *vertical* (denoising/score) direction from a *horizontal* (auxiliary) direction and derives how they combine. The main text is recovered by taking the vertical direction to be EM on the learned energy $\mathcal{E}_\theta$ and the horizontal direction to be the harmonic–glue EM step. In this general formalism, the horizontal direction can also host MCMC (e.g. MALA/HMC), metadynamics, or alchemical/free–energy transformations.

Indexing and units. We work in friction units with $\beta D = 1$ (so $D = \beta^{-1}$). The lattice sites are

$$(n, b) \in \mathbb{Z}_{\geq 0} \times \{0, \ldots, B\}, \qquad X_{n,b} \in \mathbb{R}^d,$$

where n is the *vertical* index (denoising time) and b is the *horizontal* index (replica/auxiliary sheet). The learned energy is $\mathcal{E}_\theta : \mathbb{R}^d \to \mathbb{R}$. Each horizontal slice b has its own auxiliary potential $U_b : \mathbb{R}^d \to \mathbb{R}$ (examples below), and the *target at slice* b is the Boltzmann law

$$\pi_b(x) \;\propto\; \exp\big[-\beta(\mathcal{E}_\theta(x) + U_b(x))\big]. \tag{22}$$

F.1 CONTINUOUS-TIME MODEL AND INVARIANCE

Consider the continuous process on slice b:

$$\mathrm{d}X_t^{(b)} \;=\; -\nabla\big(\mathcal{E}_\theta + U_b\big)\big(X_t^{(b)}\big)\,\mathrm{d}t \;+\; \sqrt{2D}\,\mathrm{d}W_t. \tag{23}$$

Its generator is $\mathcal{L}_b f = \langle -\nabla(\mathcal{E}_\theta + U_b), \nabla f\rangle + D\,\Delta f$, and the stationary density is exactly equation 22; the process is reversible w.r.t. π_b.

F.2 DISCRETE-TIME: TWO-DIRECTION EM WITH A SHARED DIFFUSION BUDGET

Let a macro-step of duration Δt be split as $\Delta t_v = \alpha_v \Delta t,\ \Delta t_h = \alpha_h \Delta t,\ \alpha_v + \alpha_h = 1$. Define the vertical and horizontal EM kernels (at slice b):

$$\mathbf{K}_v(x;\eta) := x - \Delta t_v\,\nabla\mathcal{E}_\theta(x) + \sqrt{2D\,\Delta t_v}\,\eta, \tag{24}$$

$$\mathbf{K}_h^{(b)}(x;\xi) := x - \Delta t_h\,\nabla U_b(x) + \sqrt{2D\,\Delta t_h}\,\xi, \tag{25}$$

with independent $\eta, \xi \sim \mathcal{N}(0, I_d)$ *at each site*. A single (unsymmetrised) composed update is $X^+ = \mathbf{K}_h^{(b)}(\mathbf{K}_v(X;\eta);\xi)$.

Gaussian fusion (exact covariance). [Noise fusion] Let $X \in \mathbb{R}^d$ be deterministic and define $\Delta_v := \sqrt{2D\Delta t_v}\,\eta$, $\Delta_h := \sqrt{2D\Delta t_h}\,\xi$, with η, ξ i.i.d. $\mathcal{N}(0, I_d)$. Then $\Delta_v + \Delta_h \sim \sqrt{2D(\Delta t_v + \Delta t_h)}\,\zeta$ with $\zeta \sim \mathcal{N}(0, I_d)$, independent of X. Hence the two-direction step has *exact* covariance $2D\Delta t\,I_d$. Sum of independent Gaussians with covariances $2D\Delta t_v I_d$ and $2D\Delta t_h I_d$.

First-order moments (drift bias). Conditioning on X and Taylor expanding ∇U_b at X,

$$\mathbb{E}[X^+ - X \mid X] = -\Delta t\big(\nabla \mathcal{E}_\theta(X) + \nabla U_b(X)\big) \;+\; \mathcal{O}(\Delta t^2),$$

i.e., the composed EM has the *correct total drift* up to $\mathcal{O}(\Delta t^2)$. The $\mathcal{O}(\Delta t^2)$ term is due solely to evaluating ∇U_b at the random intermediate $\mathbf{K}_v(X;\eta)$.

F.3 SECOND-ORDER (STRANG) SPLITTING AND WEAK ACCURACY

Define the Strang macro-step as

$$\mathcal{S}^{(b)}_{\Delta t} \;:=\; \underbrace{\mathbf{K}^{(b)}_h(\cdot;\xi_{1/2})}_{\Delta t_h/2} \circ \underbrace{\mathbf{K}_v(\cdot;\eta)}_{\Delta t_v} \circ \underbrace{\mathbf{K}^{(b)}_h(\cdot;\xi_{2/2})}_{\Delta t_h/2}, \tag{26}$$

with independent Gaussians at all three substeps and $\Delta t_h/2 = \alpha_h \Delta t/2$, $\Delta t_v = \alpha_v \Delta t$.

[Weak-2 consistency for the sum SDE] Assume $\nabla \mathcal{E}_\theta$ and ∇U_b are globally Lipschitz with bounded second derivatives. Then for any $f \in C_b^\infty(\mathbb{R}^d)$ there exists C_f such that

$$\Big\| \,\mathbb{E}\big[f(\mathcal{S}^{(b)}_{\Delta t}(X))\big] \;-\; \big(e^{\Delta t\,\mathcal{L}_b} f\big)(X) \,\Big\| \;\le\; C_f\, \Delta t^3,$$

i.e., equation 26 is a weak-2 integrator for the SDE equation 23. By Lemma F.2, its covariance matches $2D\Delta t\, I_d$ exactly. [Proof sketch] Baker–Campbell–Hausdorff for Strang splitting: $e^{\frac{1}{2}\Delta t_h \mathcal{L}_h} e^{\Delta t_v \mathcal{L}_v} e^{\frac{1}{2}\Delta t_h \mathcal{L}_h} = e^{\Delta t(\mathcal{L}_v + \mathcal{L}_h)} + \mathcal{O}(\Delta t^3)$ in the weak sense under bounded commutators and moments. Discrete EM substeps approximate $e^{\Delta t_\bullet \mathcal{L}_\bullet}$ with weak-2 local error; composition preserves weak order.

Stationarity. [Stationary bias of the split scheme] Let π_b be equation 22. The Markov kernel $\mathcal{S}^{(b)}_{\Delta t}$ preserves π_b up to $\mathcal{O}(\Delta t^2)$ (Talay–Tubaro expansion). A final Metropolis–Hastings accept/reject with target π_b makes stationarity exact, with acceptance $1 - \mathcal{O}(\Delta t^2)$.

F.4 RECOVERING THE MAIN TEXT AND GOING BEYOND

Main-text specialisation (vertical score + horizontal harmonic glue). Take U_b to be the *harmonic glue* potential (two-slice quadratic) used in the main text. Then equation 24 implements score-based EM on $\mathcal{E}_\theta$ and equation 25 implements the harmonic EM tether; by Theorem F.3 the composed step is weak-2 for the SDE driven by $-\nabla(\mathcal{E}_\theta + U_b)$ and targets the correct $\pi_b \propto e^{-\beta(\mathcal{E}_\theta + U_b)}$ up to $\mathcal{O}(\Delta t^2)$.

General horizontal direction = any reversible Markov kernel. Let $P^{(b)}_h$ be *any* Markov kernel reversible w.r.t. $e^{-\beta U_b}$ (e.g. MALA, HMC, BAOAB on an extended state, or EM/MALA on $U_b + V_b$ with a metadynamics bias V_b). Replace $\mathbf{K}^{(b)}_h$ in equation 26 by $P^{(b)}_h$ with step parameter of order Δt_h. Then:

- If $P^{(b)}_h$ is itself a weak-p integrator for the U_b-Langevin dynamics, the Strang composition remains weak-$\min\{2,p\}$ for the sum SDE.
- *Exact stationarity at finite* Δt. Wrap the full proposal $Y \sim P^{(b)}_h \circ \mathbf{K}_v \circ P^{(b)}_h(X,\cdot)$ in one Metropolis–Hastings step with target π_b. The acceptance uses the composite proposal density $q(X \to Y)$ (available in closed form when $P^{(b)}_h$ is EM/MALA/BAOAB, symmetric for HMC), and yields a chain reversible w.r.t. π_b for *any* choice of $P^{(b)}_h$.

"Sheets": alchemical and free–energy transformations along b. Let $U_b(x) = U(x;\lambda_b)$ with $\lambda_b = b/B \in [0,1]$ encoding an alchemical or morphing path from U_0 to U_1. The extended target over $\mathbf{X} = (X_{n,0},\ldots,X_{n,B})$ is

$$\boldsymbol{\pi}(\mathbf{X}) \;\propto\; \prod_{b=0}^{B} \exp\big[-\beta(\mathcal{E}_\theta(X_{n,b}) + U(X_{n,b};\lambda_b))\big].$$

A horizontal pass at fixed n consists of local moves (EM/MALA/HMC/metadynamics) at each b *and* optional replica-exchange swaps $(b, b+1)$ with acceptance

$$\alpha_{b,b+1} = \min\Big\{1,\ \exp\Big(-\beta\big[U(X_{n,b+1};\lambda_b) + U(X_{n,b};\lambda_{b+1}) - U(X_{n,b};\lambda_b) - U(X_{n,b+1};\lambda_{b+1})\big]\Big)\Big\}. \tag{27}$$

The $\mathcal{E}_\theta$ terms *cancel* in equation 27 (they are common to all replicas), making AREX directly compatible with learned energies.

F.5 PUTTING IT TOGETHER ON THE LATTICE (PARALLEL EXECUTION)

Colour sites by the parity of $n+b$. A full macro-iteration is:

1. **Horizontal half-pass (colour $c = 0$):** apply $P_h^{(b)}$ with step $\Delta t_h/2$ (or $\mathbf{K}_h^{(b)}$) to all sites with $n+b \equiv 0$, in one fused GPU kernel. Optionally perform replica-exchange among disjoint pairs $(b, b+1)$ of the same colour.
2. **Vertical full-pass:** apply $\mathbf{K}_v$ with step Δt_v to *all* sites (n, b), again as one fused kernel (independent Gaussian streams).
3. **Horizontal half-pass (colour $c = 1$):** as in step 1, but for $n+b \equiv 1$.

By Lemma F.2, the total diffusion per macro-step is exactly $2D\Delta t$ (noise budgets add). Theorem F.3 and Proposition F.3 guarantee weak-2 accuracy for the sum SDE and $\mathcal{O}(\Delta t^2)$ stationarity bias; a one-shot MH wrapper makes the terminal law *exactly* π_b if desired.

F.6 EXAMPLES OF HORIZONTAL U_b AND $P_h^{(b)}$

- **Harmonic glue (main text).** U_b is the two-slice quadratic tether (with optional variance tempering, App. D); take $P_h^{(b)} = \mathbf{K}_h^{(b)}$.
- **MALA / HMC.** $P_h^{(b)}$ is a reversible MCMC kernel for $e^{-\beta U_b}$. Combine via equation 26; use the global MH wrapper for exact $e^{-\beta(\mathcal{E}_\theta + U_b)}$.
- **Metadynamics.** Maintain a slice-specific bias V_b (updated from collective variables); set $U_b \leftarrow U_b + V_b$ online. Local reversibility holds at fixed V_b, and the split analysis remains valid between hill updates.
- **Alchemical sheet + AREX.** $U_b(\cdot) = U(\cdot;\lambda_b)$ with swaps using equation 27.

F.7 REMARKS ON "INDEPENDENCE ⇒ COMMUTATION"

Independence of Gaussian streams implies *exact covariance fusion* (Lemma F.2); it does *not* imply that the generators $\mathcal{L}_v$ and $\mathcal{L}_h$ commute. The nonzero commutator stems from evaluating drifts at different points (nonlinear gradients). Strang splitting handles this optimally: the error of replacing $e^{\Delta t(\mathcal{L}_v + \mathcal{L}_h)}$ by $e^{\frac{1}{2}\Delta t \mathcal{L}_h} e^{\Delta t \mathcal{L}_v} e^{\frac{1}{2}\Delta t \mathcal{L}_h}$ is $\mathcal{O}(\Delta t^3)$ in the weak sense, while the covariance budget is exactly correct by construction.

Takeaway. The two-direction formalism yields a *single* target $\pi_b \propto e^{-\beta(\mathcal{E}_\theta + U_b)}$ at each slice, recovers the main-text harmonic-glue scheme, and admits horizontal replacements (MCMC, metadynamics, alchemical sheets) without retraining the score. Independent Gaussian substeps *exactly* add their noise budgets, while Strang splitting makes the drift composition second-order accurate; a final MH wrapper restores exact stationarity at finite step sizes.

G HIGHER-ORDER HARMONIC ADAPTERS

The first-order scheme of the main text inherits the Euler–Maruyama (EM) weak/strong orders $(1, \frac{1}{2})$ and introduces an $\mathcal{O}(\Delta t)$ stationarity error. Here we construct *explicit, still-parallel* adapters whose local truncation error is of order $\mathcal{O}(\Delta t^{p+1})$ for any $p \geq 2$. We give the algebra for $p = 2$ in detail (Heun/RK2), outline the extension to $p = 3$ and $p = 4$ stochastic Runge–Kutta (SRK) methods, and summarise the resulting bias and parallel-execution costs.

G.1 SECOND-ORDER (HEUN) HARMONIC ADAPTER

Let $k = \frac{1}{2\Delta t}$ and keep the split drift $\Psi_\Delta(a, b) = s_\theta(b) - k\,(b - a)$. For slice•$n$ we perform

H1) *Predictor (Euler)*
$\tilde{x}_{n+1} = x_n + D\,\Psi_\Delta(x_{\,n-1}, x_n)\,\Delta t + \sqrt{2D\Delta t}\,\xi_n$

H2) *Corrector (trapezoid rule)*
$x_{n+1} \;=\; x_n \;+\; \frac{D\Delta t}{2}\Big[\Psi_\Delta(x_{\,n-1}, x_n) \;+\; \Psi_\Delta(x_n, \tilde{x}_{n+1})\Big] \;+\; \sqrt{2D\Delta t}\,\xi_n, \qquad \xi_n \;\sim\; \mathcal{N}(0, I_d).$

The *same* Gaussian vector ξ_n is reused in **H1**)–**H2**); this is the standard stochastic Heun construction.

Local moments. Writing $\Delta X := x_{n+1} - x_n$ and conditioning on $(x_{\,n-1}, x_n) = (\alpha, \beta)$ gives

$$\mathbb{E}[\Delta X] = D\Big[s_\theta(\beta) - k(\beta - \alpha)\Big]\Delta t + \underbrace{\mathcal{O}(\Delta t^2)}_{\text{bias}}, \quad \mathbb{E}[\Delta X\,\Delta X^\top] = 2D\Delta t\, I_d + \mathcal{O}(\Delta t^3).$$

Compared with EM, the leading drift bias drops from $\mathcal{O}(\Delta t)$ to $\mathcal{O}(\Delta t^2)$ while the covariance remains exact. Therefore

$$\boxed{\text{stationarity error } = \mathcal{O}(\Delta t^2), \quad \text{detailed-balance defect } = \mathcal{O}(\Delta t^3).}$$

Parallel execution. Step **H1**) and **H2**) depend only on the pair $(x_{\,n-1}, x_n)$: both can be launched as two *independent batched kernels*. Memory overhead is one extra trajectory-sized buffer for $\tilde{x}_{n+1}$.

G.2 THIRD- AND FOURTH-ORDER SRK ADAPTERS

Let s_i denote the i-th stage drift evaluation at abscissa $c_i\Delta t$ and weight a_{ij} of a stochastic RK tableau (Platen – Kloeden). Replacing the EM drift in each stage by Ψ_Δ gives the *SRKp Harmonic Adapter*

$$x_{n+1} = x_n + D\Delta t\sum_{i=1}^{s} b_i\, s_i + \sqrt{2D\Delta t}\,\xi_n, \qquad s_i = \Psi_\Delta\big(x_{\,n-1} + D\Delta t\sum_{j<i} a_{ij} s_j,\; x_n + D\Delta t\sum_{j<i} \hat{a}_{ij} s_j\big).$$

Scheme	strong	weak	stationarity	# kernels
Heun (RK2)	1	2	$\mathcal{O}(\Delta t^2)$	2
SRK3 (Platen)	1	3	$\mathcal{O}(\Delta t^3)$	3
SRK4 (Rößler)	$\frac{3}{2}$	4	$\mathcal{O}(\Delta t^4)$	4

Every stage again references only $(x_{\,n-1}, x_n)$ *of the *input* array*, so each column of the tableau is a separate, embarrassingly parallel kernel.

Proof sketch (all p). Because each stage drift is a linear convex combination of *edge-wise* springs obeying LM1–LM3, the p-th order adapter inherits:

* weak order p and strong order listed above (classical SRK proofs); * $\mathcal{O}(\Delta t^p)$ stationarity error (moment-matching argument of §G.1); * detailed-balance defect $\mathcal{O}(\Delta t^{p+1})$ (antisymmetry cancels one extra power of Δt).

Thus higher accuracy costs only extra kernels, not extra synchronisation.

G.3 BIAS HIERARCHY

Let $\pi_{\Delta,p}$ be the invariant law of the SRKp adapter. The Talay–Tubaro expansion generalises to

$$\pi_{\Delta,p} = p_\theta + \Delta t^{\,p}\mathcal{L}_p\, p_\theta + \mathcal{O}(\Delta t^{p+1}),$$

where $\mathcal{L}_p$ is a polynomial in s_θ and its derivatives. For molecular observables F with $\|F\|_{C^{2p+2}} < \infty$ the bias is

$$\mathbb{E}_{\pi_{\Delta,p}}[F] - \mathbb{E}_{p_\theta}[F] = \mathcal{O}(\Delta t^p).$$

Wrapping any SRKp proposal in a one-step MALA accept/reject makes the chain *exactly* p_θ-stationary while retaining the p-th order proposal as the local importance sampler; the acceptance rate is $1 - \mathcal{O}(\Delta t^p)$.

H EXPLICIT GLUE BIAS, GLUED SCORE, AND PATHWISE CONVERGENCE

H.1 DEFINITIONS, NOTATION, AND WHAT IS PROVEN

Goal. We formalize harmonic (*quadratic-glue*) denoising steps for score-based samplers, derive the associated *glued score* used in the drift, quantify the *adjacent-batch bias* added by the glue, and prove *finite-schedule* KL/TV/W_2 bounds and a *vanishing-error* theorem as the denoising grid is refined and the learned energy converges.

H.1.1 OVER-DAMPED LANGEVIN DYNAMICS (CONTINUOUS TIME)

Units and constants. We work in *friction units*

$$\gamma = 1, \qquad D = \frac{k_B T}{\gamma} = k_B T, \qquad \beta = \frac{1}{D}, \qquad \beta D = 1.$$

Let $V : \mathbb{R}^d \to \mathbb{R}$ be the physical potential. The Itô SDE is

$$\mathrm{d}X_t \;=\; -\nabla V(X_t)\,\mathrm{d}t \;+\; \sqrt{2D}\,\mathrm{d}W_t, \qquad \text{with stationary density } \pi_\beta(\mathrm{d}x) \propto e^{-\beta V(x)}\mathrm{d}x. \tag{28}$$

H.1.2 EULER–MARUYAMA (EM) TIME DISCRETIZATION

For fixed step $\Delta t > 0$,

$$X_{n+1} \;=\; X_n \;-\; \nabla V(X_n)\,\Delta t \;+\; \sqrt{2D\,\Delta t}\,\xi_n, \qquad \xi_n \sim \mathcal{N}(0, I_d). \tag{29}$$

Conditioned on $X_n = x$, the *exact* EM transition kernel is Gaussian

$$p(x' \mid x) \;=\; \frac{1}{(4\pi D\Delta t)^{d/2}} \exp\Big[-\frac{1}{4D\Delta t}\,\big\|x' - x + \Delta t\,\nabla V(x)\big\|^2\Big]. \tag{30}$$

H.1.3 HARMONIC REFORMULATION ("QUADRATIC GLUE")

It is convenient to rewrite equation 30 as a Boltzmann factor with a quadratic coupling:

$$p(x' \mid x) = Z(x)^{-1} \exp\Big[-\frac{k(\Delta t)}{2}\,\big\|x' - x + \Delta t\,\nabla V(x)\big\|^2\Big], \tag{31}$$

$$k(\Delta t) := \frac{1}{2D\,\Delta t}. \tag{32}$$

Because the mismatch $x' - x + \Delta t\,\nabla V(x) = \mathcal{O}(\sqrt{\Delta t})$, the spring energy contributes only $\mathcal{O}(\Delta t)$ to the exponent even though $k(\Delta t) \sim \Delta t^{-1}$.

Two-slice energy (auxiliary). Define

$$U_{\Delta t}(x'; x) \;:=\; V(x') \;+\; \frac{k(\Delta t)}{2}\,\big\|x' - x + \Delta t\,\nabla V(x)\big\|^2. \tag{33}$$

This is *not* a conditional energy; integrating $e^{-U_{\Delta t}(x';x)}$ in x' recovers the Gaussian equation 30 up to the normalizer $Z(x)$.

H.2 SCORE-BASED MODELS: TRAINING OBJECTS AND THE LEARNED ENERGY

Let p_t denote the data corrupted by i.i.d. Gaussian noise of variance $\sigma_t^2 = 2Dt$. A standard score network trained by denoising objectives approximates the time-dependent score $s_t(x) \approx \nabla_x \log p_t(x)$. For clarity, we also define a *learned energy*

$$E_\theta(x) \;:=\; -\log p_\theta(x) + \text{const}, \qquad \nabla E_\theta(x) = -s_0(x), \tag{34}$$

and, at inference time, we use a drift proxy $g_n(\cdot)$ that is either $g_n = \nabla E_\theta$ (energy learning) or $g_n = s_{t_n}$ (time-dependent score). In both cases we measure model error against the physical drift:

$$g_n(x) \;=\; \nabla V(x) + \varepsilon_n(x), \qquad \|\varepsilon_n\|_\infty \;\le\; \bar{\varepsilon}. \tag{35}$$

H.3 HARMONIC GLUE AT INFERENCE

We describe two realizations.

(A) Adjacent-batch glue (no anchors). Given a sequence $\{X_n\}$, set for each step n

$$X_{n+1} \;=\; X_n \;-\; \Big(g_n(X_n) + k(\Delta t)\big(X_n - X_{n-1}\big)\Big)\Delta t \;+\; \sqrt{2D\,\Delta t}\,\xi_n, \tag{36}$$

with the convention $X_{-1} = X_0$. *Explicit adjacent bias.* Conditioning on $(X_{n-1}, X_n) = (a, b)$ and taking noise expectation,

$$\mathbb{E}[X_{n+1} - b \mid a, b] \;=\; -\,g_n(b)\,\Delta t \;-\; \underbrace{k(\Delta t)\Delta t}_{=1/(2D)}\,\big(b - a\big),$$

i.e., the glue drift pulls *toward the adjacent element* with strength $(2D)^{-1}$.

(B) Gibbs-anchored glue. Introduce an auxiliary anchor $A_n \mid X_n \sim \mathcal{N}\big(X_n,\, [\beta k_a]^{-1} I\big)$ (with a user spring $k_a > 0$) and update

$$X_{n+1} \;=\; X_n \;-\; \Big(g_n(X_n) + k_a\big(X_n - A_n\big)\Big)\Delta t \;+\; \sqrt{2D\,\Delta t}\,\xi_n. \tag{37}$$

The *glued score* used in both variants is the gradient of the quadratic-augmented potential

$$\nabla_x\Big(V(x) + \tfrac{k}{2}\|x - a\|^2\Big) \;=\; \nabla V(x) + k(x - a), \tag{38}$$

so that the algorithmic drift equals minus equation 38 with $a = X_{n-1}$ (adjacent) or $a = A_n$ (anchor), and with ∇V replaced by g_n in practice.

H.4 ASSUMPTIONS (LOCAL TO THIS SECTION)

[Regularity and coercivity] $V \in C^2(\mathbb{R}^d)$ with globally L-Lipschitz gradient, and $V(x) \to \infty$ as $\|x\| \to \infty$ (e.g. $V(x) \ge c_0\|x\|^q - c_1$).

[Schedule] A grid $t_n = n\Delta t$ with step $\Delta t > 0$ and total horizon $T = N\Delta t$. We write $k(\Delta t) = 1/(2D\Delta t)$ as in equation 32.

[Model error] The inference drift g_n satisfies equation 35 for all n.

H.5 FINITE-SCHEDULE ACCURACY: PATHWISE KL/TV/W_2

We compare the piecewise-constant-drift interpolation $\tilde{X}_t$ of either glue scheme on $[0, T]$ with the exact SDE equation 28 driven by the *same* Brownian motion.

[EM local moment bound] For any interval $[t_n, t_{n+1})$, $\mathbb{E}\|\tilde{X}_t - \tilde{X}_{t_n}\|^2 \le C_1(t - t_n)$ with a constant $C_1 = C_1(D, L, \sup_n \mathbb{E}\|X_n\|^2) < \infty$.

[Pathwise KL bound with harmonic glue] Let $\tilde{X}_t$ be generated by equation 36 or equation 37. Under Assumptions H.4–H.4,

$$D_{\mathrm{KL}}\big(\mathcal{L}(\tilde{X}_{[0,T]}) \,\|\, \mathcal{L}(X_{[0,T]})\big) \;\le\; \underbrace{\beta\, T\, \bar{\varepsilon}^2}_{\text{model error}} \;+\; \underbrace{\frac{\beta\, L_{\text{tot}}^2}{2} \sum_{n=0}^{N-1} \Delta t^2}_{\text{schedule error}}, \tag{39}$$

where $L_{\text{tot}} := L + k_\star$ with $k_\star = \max\{\, k(\Delta t),\, k_a \,\}$. Consequently,

$$\|\mathcal{L}(\tilde{X}_{[0,T]}) - \mathcal{L}(X_{[0,T]})\| \ \le\ \tfrac{1}{2}\sqrt{D_{\mathrm{KL}}(\cdot\|\cdot)}\ ,$$

and if π_β satisfies a $T_2(m)$ inequality, then $W_2^2(\tilde{q}_T, \pi_\beta) \le \frac{2}{m} D_{\mathrm{KL}}(\tilde{q}_T \| \pi_\beta)$ for the terminal law $\tilde{q}_T$.

Let X_t solve equation 28. On each $[t_n, t_{n+1})$ the algorithmic drift is the constant

$$b_t \equiv -g_n(\tilde{X}_{t_n}) - \underbrace{k_\bullet\big(\tilde{X}_{t_n} - a_n\big)}_{\text{glue term}},$$

with $k_\bullet = k(\Delta t)$ and $a_n = X_{n-1}$ (adjacent) or $k_\bullet = k_a$ and $a_n = A_n$ (anchor). The exact drift is $-\nabla V(\tilde{X}_t)$. Girsanov's formula yields

$$D_{\mathrm{KL}}\big(\tilde{X}\|X\big) = \frac{1}{4D}\,\mathbb{E}\int_0^T \Big\|b_t + \nabla V(\tilde{X}_t)\Big\|^2\,\mathrm{d}t.$$

Split the drift gap on $[t_n, t_{n+1})$ into three parts:

$$\underbrace{-\varepsilon_n(\tilde{X}_{t_n})}_{\text{model}} + \underbrace{\big(\nabla V(\tilde{X}_{t_n}) - \nabla V(\tilde{X}_t)\big)}_{\text{Lipschitz}} \ -\ \underbrace{k_\bullet\big(\tilde{X}_{t_n} - a_n\big)}_{\text{glue}}.$$

(i) The model term contributes $\frac{1}{4D}\int_{t_n}^{t_{n+1}} \mathbb{E}\|\varepsilon_n\|^2 \mathrm{d}t \ \le\ \frac{1}{4D}\bar{\varepsilon}^2\Delta t \ =\ \frac{\beta}{4}\bar{\varepsilon}^2\Delta t$. Summing gives the $\beta T\bar{\varepsilon}^2$ term (up to a harmless constant factor absorbed in β).
(ii) The Lipschitz term satisfies $\|\nabla V(\tilde{X}_{t_n}) - \nabla V(\tilde{X}_t)\| \le L\,\|\tilde{X}_t - \tilde{X}_{t_n}\|$. By Lemma H.5,

$$\frac{1}{4D}\int_{t_n}^{t_{n+1}} \mathbb{E}\|\cdot\|^2 \mathrm{d}t \ \le\ \frac{L^2}{4D}\int_{t_n}^{t_{n+1}} C_1(t - t_n)\,\mathrm{d}t \ =\ \frac{L^2 C_1}{8D}\,\Delta t^2.$$

(iii) The glue term is *Lipschitz in x* with constant $k_\bullet$: when computing the pathwise KL against a *glued* baseline (anchor SDE) this term cancels exactly; when comparing to equation 28 directly, it is bounded exactly like (ii) with L replaced by $k_\bullet$, generating $\frac{k_\bullet^2 C_1}{8D}\Delta t^2$. Combining (ii)–(iii) yields the discretization part in equation 39 with $L_{\text{tot}} = L + k_\star$. Data processing then gives the terminal-law bounds.

[Uniform steps minimize schedule error] For fixed $T = N\Delta t$, the sum $\sum_n \Delta t^2$ is minimized by uniform steps, giving schedule error $\mathcal{O}(N^{-1})$.

H.6 VANISHING KERNEL AND PATH ERROR AS THE DENOISING GRID IS REFINED

[Refinement regime] As $N \to \infty$, $\max_n \Delta t \to 0$, $N\Delta t \to T \in (0, \infty)$, and the model improves: $\bar{\varepsilon} = \bar{\varepsilon}(N) \to 0$.

[Vanishing path error] Under Assumptions H.4, H.4, H.4, H.6, the path laws converge:

$$D_{\mathrm{KL}}\big(\mathcal{L}(\tilde{X}^{(N)}_{[0,T]}) \parallel \mathcal{L}(X_{[0,T]})\big) \xrightarrow[N\to\infty]{} 0, \qquad \|\mathcal{L}(\tilde{X}^{(N)}_{[0,T]}) - \mathcal{L}(X_{[0,T]})\| \xrightarrow[N\to\infty]{} 0.$$

Hence, for every bounded Lipschitz functional $F : C([0,T]; \mathbb{R}^d) \to \mathbb{R}$, $\big|\mathbb{E}F(\tilde{X}^{(N)}_\cdot) - \mathbb{E}F(X_\cdot)\big| \to 0$.

Apply Theorem H.5. Since $\sum_n \Delta t^2 \le T\,\max_n \Delta t$, the schedule term is $\le \frac{\beta L_{\text{tot}}^2}{2}\,T\,\max_n \Delta t \to 0$. The model term is $\beta T\bar{\varepsilon}(N)^2 \to 0$. Pinsker yields TV convergence; bounded-Lipschitz convergence follows.

H.7 CONSEQUENCES AND DIAGNOSTIC COROLLARIES

[Bias decomposition] For fixed T,

$$\text{total error} \ \lesssim\ \underbrace{\beta T\,\bar{\varepsilon}^2}_{\text{model}} \ +\ \underbrace{\frac{\beta L_{\text{tot}}^2}{2}\sum_n \Delta t^2}_{\text{schedule}}.$$

The second term decays as $\mathcal{O}(N^{-1})$ with uniform steps; higher-order (Heun/SRK) adapters further improve the rate.

[Temperature/resolution dial] From equation 32, $k(\Delta t) = 1/(2D\Delta t) = \beta/(2\Delta t)$. Thus the *inference spring* selects an effective resolution $\Delta t = \beta/(2k)$ while leaving the order constants in Theorem H.5 unchanged.

Remarks (implementation). (i) In the anchor variant, comparing *jointly* to the anchor SDE removes the explicit glue term from the KL and yields $L_{\text{tot}} = L + k_a$ directly from the x-Hessian of $V(x) + \frac{k_a}{2}\|x - a\|^2$. (ii) A one-step Metropolis correction on E_θ makes the terminal law exactly $\propto e^{-\beta E_\theta}$ while preserving acceptance $1 - \mathcal{O}(\Delta t)$.

Notes on provenance. The units choice $\beta D = 1$, the EM kernel identity equation 30, and the algebraic spring $k(\Delta t) = 1/(2D\Delta t)$ match the Harmonic Adapter formalism used in the accompanying chapter; here we give full path-space KL proofs and adjacent-batch bias formulas within a self-contained presentation. (See also Ch. 2 for consistent exposition.) [Ref.: Chapter 2 draft]

H.8 COROLLARIES FOR TRAJECTORIES, FREE ENERGIES, OBSERVABLES, SPEED-UPS, AND CONVERGENCES

We now derive consequences that are frequently needed in MD workflows: bounds for path observables (including correlation functions and Green–Kubo integrals), free-energy estimators (Zwanzig/-Jarzynski), complexity/speed-up laws, and convergence of time averages. All statements below depend only on the local assumptions of this section and the previously proved Theorems H.5–H.6.

H.8.1 TRAJECTORY OBSERVABLES: PATH-FUNCTIONALS AND FINITE-TIME CORRELATIONS

Let $\tilde{\mathbb{P}}$ (resp. $\mathbb{P}$) denote the path law of the harmonic-glue sampler (resp. the exact SDE equation 28) on $[0, T]$.

[Path-functional stability] Let $F : C([0,T]; \mathbb{R}^d) \to \mathbb{R}$ be bounded. Then

$$\left|\mathbb{E}_{\tilde{\mathbb{P}}}F - \mathbb{E}_{\mathbb{P}}F\right| \;\le\; 2\|F\|_\infty\big(\tilde{\mathbb{P}}, \mathbb{P}\big) \;\le\; \|F\|_\infty \sqrt{D_{\mathrm{KL}}(\tilde{\mathbb{P}} \,\|\, \mathbb{P})}\,,$$

where the last inequality is Pinsker. By Theorem H.5, the RHS is

$$\le\; \|F\|_\infty \sqrt{\beta T\,\bar{\varepsilon}^2 + \frac{\beta L_{\text{tot}}^2}{2} \sum_n \Delta t^2}.$$

In particular, $\mathbb{E}_{\tilde{\mathbb{P}}}F \to \mathbb{E}_{\mathbb{P}}F$ as $N \to \infty$ and $\bar{\varepsilon} \to 0$ (Theorem H.6).

[Two-time correlations and Green–Kubo kernels] Fix a bounded observable $\varphi : \mathbb{R}^d \to \mathbb{R}$ with $\|\varphi\|_\infty < \infty$. For $0 \le s \le t \le T$, define

$$C_{\text{true}}(s,t) := \mathbb{E}_{\mathbb{P}}\big[\varphi(X_s)\varphi(X_t)\big], \qquad C_{\text{glue}}(s,t) := \mathbb{E}_{\tilde{\mathbb{P}}}\big[\varphi(\tilde{X}_s)\varphi(\tilde{X}_t)\big].$$

Then

$$|C_{\text{glue}}(s,t) - C_{\text{true}}(s,t)| \;\le\; 2\|\varphi\|_\infty^2\big(\tilde{\mathbb{P}}, \mathbb{P}\big) \;\le\; \|\varphi\|_\infty^2 \sqrt{D_{\mathrm{KL}}(\tilde{\mathbb{P}} \,\|\, \mathbb{P})}.$$

Consequently, any Green–Kubo integral $\int_0^T C(\tau)\,w(\tau)\,\tau$ (with bounded weight w) computed on $\tilde{X}$. converges to the exact one as in Theorem H.6.

[Time rescaling via the spring] With $k(\Delta t) = \beta/(2\Delta t)$, k acts purely as a *resolution dial*: increasing k decreases the implied step $\Delta t = \beta/(2k)$ and lengthens correlation time in frame index, without changing the order constants in Theorem H.5. When comparing autocorrelation functions to MD, one must rescale the lag axis by Δt to a physical time grid. This is the design principle emphasised in the chapter and its figures (ACF comparison and "resolution dial"). *Qualitatively*, increasing k stretches correlation shapes while leaving equilibrium statistics intact.

H.8.2 FREE ENERGIES FROM END-POINTS AND FROM PATHS

[Zwanzig FEP bias under terminal-law mismatch] Let q_T be the terminal law of $\tilde{X}_T$ and π the target $\propto e^{-\beta V}$. For any reference potential V_{ref} define $w(x) = \exp\{-\beta[V(x) - V_{\mathrm{ref}}(x)]\}$. Then

$$\left|\mathbb{E}_{q_T} w - \mathbb{E}_\pi w\right| \;\le\; \sqrt{\chi^2(q_T\|\pi)}\,\sqrt{\mathbb{E}_\pi[w^2]} \;\le\; \sqrt{e^{D_{\mathrm{KL}}(q_T\|\pi)} - 1}\,\sqrt{\mathbb{E}_\pi[w^2]}.$$

In particular, if $D_{\mathrm{KL}}(q_T\|\pi) \ll 1$, then

$$\left|\mathbb{E}_{q_T} w - Z_{\mathrm{ref}}/Z\right| \;\le\; \sqrt{D_{\mathrm{KL}}(q_T\|\pi)}\,\sqrt{\mathbb{E}_\pi[w^2]}\,\big(1 + o(1)\big).$$

By Theorem H.5 and data processing, $D_{\mathrm{KL}}(q_T\|\pi)$ obeys the same finite-schedule bound as the path KL. Hence the Zwanzig estimator bias vanishes under the refinement regime.

[Derivation] Write $r = \frac{q_T}{\pi}$. Then $\mathbb{E}_{q_T} w - \mathbb{E}_\pi w = \mathbb{E}_\pi[(r-1)w]$. Cauchy–Schwarz gives $|\mathbb{E}_\pi[(r-1)w]| \le \sqrt{\mathbb{E}_\pi[(r-1)^2]}\sqrt{\mathbb{E}_\pi[w^2]} = \sqrt{\chi^2(q_T\|\pi)}\,\sqrt{\mathbb{E}_\pi[w^2]}$. Finally, $\chi^2 \le e^{D_{\mathrm{KL}}} - 1$.

[Jarzynski-type path estimator] Let $G(\omega) = \exp\{-\beta W(\omega)\}$ be any nonnegative path weight (e.g. work functional along a nonequilibrium switching). Then

$$\left|\mathbb{E}_{\tilde{\mathbb{P}}} G - \mathbb{E}_{\mathbb{P}} G\right| \;\le\; \sqrt{\chi^2(\tilde{\mathbb{P}}\,\|\,\mathbb{P})}\,\sqrt{\mathbb{E}_{\mathbb{P}}[G^2]} \;\le\; \sqrt{e^{D_{\mathrm{KL}}(\tilde{\mathbb{P}}\,\|\,\mathbb{P})} - 1}\,\sqrt{\mathbb{E}_{\mathbb{P}}[G^2]}.$$

Hence the free-energy difference $-\beta^{-1}\log \mathbb{E}G$ computed on glued paths converges to the exact value; moreover, for $m := \mathbb{E}_{\mathbb{P}} G > 0$,

$$\left|\Delta F_{\mathrm{glue}} - \Delta F_{\mathrm{true}}\right| \;\le\; \frac{1}{\beta}\,\frac{|\mathbb{E}_{\tilde{\mathbb{P}}} G - \mathbb{E}_{\mathbb{P}} G|}{\min\{\mathbb{E}_{\tilde{\mathbb{P}}} G, \mathbb{E}_{\mathbb{P}} G\}} \;\lesssim\; \frac{1}{\beta m}\sqrt{e^{D_{\mathrm{KL}}(\tilde{\mathbb{P}}\,\|\,\mathbb{P})} - 1}\,\sqrt{\mathbb{E}_{\mathbb{P}}[G^2]}.$$

H.8.3 TIME AVERAGES, VARIANCE, AND SPECTRAL-GAP IMPLICATIONS

[Strong convexity (optional for sharp rates)] Assume $\nabla^2 V(x) \succeq m I_d$ for some $m > 0$; then the Langevin semigroup has spectral gap at least m and a Poincaré constant $1/m$.

[Variance and IAT for time averages] Let $\varphi \in L^2(\pi)$ with $\pi[\varphi] = 0$ and define the time average $\bar{\varphi}_T := \frac{1}{T}\int_0^T \varphi(X_t)\, t$ (resp. $\tilde{\bar{\varphi}}_T$ on $\tilde{X}_t$). Under Assumption H.8.3,

$$\mathrm{Var}(\bar{\varphi}_T) \;\le\; \frac{2}{mT}\,\mathrm{Var}_\pi(\varphi).$$

Moreover, for the glued process,

$$\left|\mathbb{E}[\tilde{\bar{\varphi}}_T] - \mathbb{E}[\bar{\varphi}_T]\right| \;\le\; \|\varphi\|_\infty \sqrt{\beta T\,\bar{\varepsilon}^2 + \tfrac{\beta L_{\mathrm{tot}}^2}{2}\sum_n \Delta t^2},$$

and

$$\left|\mathrm{Var}(\tilde{\bar{\varphi}}_T) - \mathrm{Var}(\bar{\varphi}_T)\right| \;\le\; C(\varphi)\sqrt{\beta T\,\bar{\varepsilon}^2 + \tfrac{\beta L_{\mathrm{tot}}^2}{2}\sum_n \Delta t^2}$$

for a constant $C(\varphi)$ depending on $\|\varphi\|_\infty$ and $\mathrm{Var}_\pi(\varphi)$. Thus the MSE of time-averaged observables under glue converges to that of the exact process.

[Spectral gap and "stiffness dial"] When V is m-strongly convex, the effective local curvature of the *glued score* $\nabla V(x) + k(x-a)$ is $m + k$, so the *local* contractivity rate of the drift increases with k. While the invariant law of x is unchanged in the anchor construction, the *discrete* proposal contracts faster in stiff directions, reducing integrated autocorrelation times in practice (consistent with the empirical findings and the "enhanced gap" discussion in the chapter).

H.8.4 COMPLEXITY/SPEED-UP COROLLARIES (WALL-CLOCK AND ACCURACY)

Recall the finite-schedule bound

$$D_{\mathrm{KL}}(\tilde{\mathbb{P}}\,\|\,\mathbb{P}) \;\le\; \beta T\,\bar{\varepsilon}^2 + \frac{\beta L_{\mathrm{tot}}^2}{2}\sum_n \Delta t^2.$$

[Steps needed for accuracy ϵ] With uniform steps $\Delta t = T/N$,

$$D_{\mathrm{KL}}(\tilde{\mathbb{P}} \,\|\, \mathbb{P}) \;\le\; \beta T\,\bar{\varepsilon}^2 + \frac{\beta L_{\mathrm{tot}}^2}{2}\,\frac{T^2}{N}.$$

To guarantee $D_{\mathrm{KL}} \le \epsilon^2$ it suffices to take

$$N \;\ge\; \frac{\beta L_{\mathrm{tot}}^2 T^2}{2\,(\epsilon^2 - \beta T\,\bar{\varepsilon}^2)} \qquad (\epsilon^2 > \beta T\,\bar{\varepsilon}^2).$$

Hence, at fixed model error, accuracy improves as $\mathcal{O}(N^{-1})$; with higher-order Harmonic Adapters (Heun/SRK), the schedule term improves to $\mathcal{O}(N^{-p})$ with the same proof pattern (moment-matching), reducing N further. *See the chapter's Appendix E for explicit SRK adapters that preserve parallelism.*

[Effective sample size (ESS) and wall-clock] Let τ_{int} be the integrated autocorrelation time of $\varphi(X_t)$. For a trajectory of duration $T = N\Delta t$, the effective sample size ESS $\approx T/(2\tau_{\mathrm{int}})$. If the glue increases the local contractivity in stiff modes (Remark H.8.1), then τ_{int} decreases and ESS increases at the *same network-call budget* (the per-step cost equals a single score evaluation). Combined with Corollary H.8.4, one obtains the MSE scaling

$$\mathrm{MSE}(\tilde{\bar{\varphi}}_T) \;\lesssim\; \frac{\mathrm{Var}_\pi(\varphi)}{\mathrm{ESS}} \;+\; \|\varphi\|_\infty^2 \sqrt{\beta T\,\bar{\varepsilon}^2 + \frac{\beta L_{\mathrm{tot}}^2 T^2}{2N}},$$

making explicit the variance–bias trade-off: increase N until the schedule term is below the model term, then reduce $\bar{\varepsilon}$ (distillation/training). The *resolution dial* $k = \beta/(2\Delta t)$ allows tuning correlation length at fixed bias order.

H.8.5 RELATIONS AMONG THE COROLLARIES

- **(Theorem H.5)** $\Rightarrow$ **(Cor. H.8.1, H.8.1).** Substitute Pinsker into path functionals to bound trajectory-level observables and two-time correlations.
- **(Theorem H.5)** $\Rightarrow$ **(Cor. H.8.2).** Apply data processing to pass from path KL to terminal KL; use $\chi^2 \le e^{\mathrm{KL}} - 1$.
- **(Theorem H.5)** $\Rightarrow$ **(Cor. H.8.2).** Work on path space directly with Cauchy–Schwarz; convert to a free-energy bound via $\log$-Lipschitz on $(0,\infty)$.
- **(Assump. H.8.3)** $\Rightarrow$ **(Cor. H.8.3).** Use Poincaré/Green–Kubo to control variance of time averages; combine with path-KL to control bias and variance gap.
- **(Theorem H.5)** $+$ uniform steps $\Rightarrow$ **(Cor. H.8.4)**. Turn the schedule term into an explicit N-law; the chapter's SRK adapters improve the rate to $\mathcal{O}(N^{-p})$ at the same parallel depth.
- **(Remark H.8.1)** $+$ **(Cor. H.8.4)**. The algebraic identity $k(\Delta t) = \beta/(2\Delta t)$ lets one trade correlation length against resolution without changing the bias order; practical speed-ups follow from larger ESS at fixed wall-clock.

I PYTHON HARMONIC ADAPTER FUNCTION

PRACTICAL CHOICE: HARMONIC-ONLY TO AVOID EXTRA SCORE EVALUATIONS

In our pipeline, the Harmonic Adapter with specified minimum distance is used instead of a score-modulated variant because it:

1. Requires only one scalar f_t per frame from d_t, avoiding additional score evaluations (e.g., learned or external scoring functions) that would otherwise increase wall time.
2. Provides stable, semiconvex coupling around $r_{\min}$, which empirically reduces optimization noise while preserving temporal smoothness and approximates the score while being score-free.

Described further in §**??**

```
import torch
from typing import Optional

def harmonic_adapter(
    self,
    coords: torch.Tensor,                 # [T, N, 3]
    align: bool = True,
    max_neighbors: int = 3,               # S (number of forward
    ↪  temporal neighbors)
    reduction_factor: float = 0.6,        # rho in (0,1)
    gm: Optional[torch.Tensor] = None,    # optional helper arg for
    ↪  align-aware MSE
    k: float = 1.0e-1,                    # spring constant
    r_min: float = 12.0,                  # preferred distance
    eps: float = 1e-9,                    # small numerical stabilizer
):
    """
    Harmonic-only temporal adapter. The routine builds bidirectional
    temporal couplings between frames (t, t+s) with geometric decay.

    Args:
      coords: [T, N, 3] time-ordered coordinates.
      align: if True, use alignment-aware MSE helper; else plain MSE.
      max_neighbors: number of forward temporal neighbors S.
      reduction_factor: rho in (0,1), geometric decay with neighbor
      ↪  shift s.
      gm: optional geometry/guide passed to alignment-aware MSE helper.
      k: harmonic stiffness; r_min: preferred distance; eps: stability
      ↪  term.

    Returns:
      gradients: [T, N, 3]
    """
    if coords.ndim != 3 or coords.shape[-1] != 3:
        raise ValueError("coords must be [T, N, 3].")
    device, dtype = coords.device, coords.dtype
    T, N, _ = coords.shape
    if T < 2:
        raise ValueError("Need at least two frames (T >= 2).")

    # --- Per-frame scalar distance d_t = sqrt(MSE) + eps ---
    if align:
        mse_d2 = align_and_calculate_mse(coords, gm=gm)  # user-provided
        ↪  helper
    else:
        mse_d2 = calculate_mse(coords)                   # user-provided
        ↪  helper
    d = torch.sqrt(mse_d2.to(device=device, dtype=dtype)) +
    ↪  torch.as_tensor(eps, device=device, dtype=dtype)  # [T]

    # --- Harmonic force scalar f_t = -k (d_t - r_min) ---
    f = -torch.as_tensor(k, device=device, dtype=dtype) * (d -
    ↪  torch.as_tensor(r_min, device=device, dtype=dtype))  # [T]

    # --- Batched Kabsch alignment of all frames to the first ---
    center_of_mass = coords.mean(dim=1, keepdim=True)
    coords_centered = coords - center_of_mass
    reference = coords_centered[0].unsqueeze(0).expand(T, -1, -1)
    coords_aligned = kabsch_algorithm_batch(reference, coords_centered)
    ↪  # [T, N, 3]

    # --- Accumulate neighbor contributions (vectorized; O(T N S)) ---
    gradients = torch.zeros_like(coords_aligned, device=device,
    ↪  dtype=dtype)
    rho = torch.as_tensor(reduction_factor, device=device, dtype=dtype)
```

```
    factors = rho ** torch.arange(max_neighbors, device=device,
    ↪  dtype=dtype)  # [S]
    invN = torch.as_tensor(1.0 / float(N), device=device, dtype=dtype)

    # Bidirectional: add to t, subtract from t+s
    for shift in range(1, max_neighbors + 1):
        if shift >= T:
            break
        dR = coords_aligned[:-shift] - coords_aligned[shift:]   #
        ↪  [T-shift, N, 3]
        f_s = f[:-shift].view(-1, 1, 1)                          #
        ↪  [T-shift, 1, 1]
        contrib = factors[shift - 1] * (f_s * dR) * invN         #
        ↪  [T-shift, N, 3]
        gradients[:-shift] += contrib
        gradients[shift:]  -= contrib

    return gradients
```

MATHEMATICAL SUMMARY (HARMONIC-ONLY, MATCHES CODE)

Given time-ordered coordinates $\mathbf{R}_{0:T-1} \in \mathbb{R}^{T\times N\times 3}$, define the scalar per-frame distance

$$d_t \;=\; \sqrt{\mathrm{MSE\_distance}\big(\mathbf{R}_t;\ \text{(optional alignment helpers)}\big)} \;+\; \varepsilon, \qquad t = 0,\ldots,T-1.$$

The harmonic force magnitude used by the adapter is

$$f_t \;=\; -\,k\,\big(d_t - r_{\min}\big).$$

All frames are aligned to the first via batched Kabsch to obtain $\widehat{\mathbf{R}}_t$. For a reduction factor $\rho \in (0,1)$ and a maximum forward temporal neighborhood $S = \texttt{max\_neighbors}$, the bidirectional contribution between frames $(t, t+s)$ is

$$\Delta\mathbf{R}_{t\to t+s} \;:=\; \widehat{\mathbf{R}}_t - \widehat{\mathbf{R}}_{t+s} \in \mathbb{R}^{N\times 3}, \qquad \mathbf{g}_t^{(s)} \;=\; \rho^{s-1}\, f_t\, \frac{\Delta\mathbf{R}_{t\to t+s}}{N}, \qquad \mathbf{g}_{t+s}^{(s)} \;=\; -\,\mathbf{g}_t^{(s)}.$$

Accumulating over valid pairs yields the returned gradients $\mathbf{G}_{0:T-1} \in \mathbb{R}^{T\times N\times 3}$:

$$\mathbf{G}_t = \sum_{\substack{s=1\\ t+s\le T-1}}^{S} \rho^{s-1} f_t \frac{\widehat{\mathbf{R}}_t - \widehat{\mathbf{R}}_{t+s}}{N} - \sum_{\substack{s=1\\ t-s\ge 0}}^{S} \rho^{s-1} f_{t-s} \frac{\widehat{\mathbf{R}}_{t-s} - \widehat{\mathbf{R}}_t}{N}.$$

J SCORE-ON VERTICAL EM WITH A SCORE-FREE $r_{\min}$ HARMONIC ADAPTER

Vertical (i.i.d.) step always uses the score. In friction units $\beta D = 1$, one Euler–Maruyama (EM) step with step Δt and learned drift g_n (time-dependent score or energy gradient) is the Gaussian kernel

$$x_{n+1} \mid x_n \;\sim\; \mathcal{N}\Big(x_n - D\Delta t\, g_n(x_n),\, 2D\Delta t\, I\Big), \qquad k(\Delta t) = \frac{1}{2D\Delta t} = \frac{\beta}{2\Delta t}.$$

This is the EM–harmonic identity used throughout (Eqs. (2)–(5),(7)–(8)).

Horizontal (inter-batch) Harmonic Adapter: score-free $r_{\min}$ glue. Let the stacked state be $X = (X_0, \ldots, X_{T-1}) \in \mathbb{R}^{T\times N\times 3}$ and define

$$r_{t,s}(X) = \Big(\tfrac{1}{N}\|X_t - X_{t+s}\|_F^2\Big)^{1/2}, \qquad c_{t,s}(X) = r_{t,s}(X) - r_{\min}.$$

The score-free distance glue is the quadratic

$$U_{\mathrm{glue}}(X) = \frac{k}{2}\sum_{t,s} \alpha_s\, c_{t,s}(X)^2, \quad \alpha_s \ge 0, \tag{40}$$

whose contribution to the drift is $-\nabla U_{\mathrm{glue}}(X)$. One horizontal EM substep (with the same $k(\Delta t) = \beta/(2\Delta t)$) is therefore an *exact* EM update for the drift $-\nabla U_{\mathrm{glue}}$ and covariance $2D\Delta t\, I$. This is the distance-space realization of the Harmonic Adapter summarized in Eq. (40).

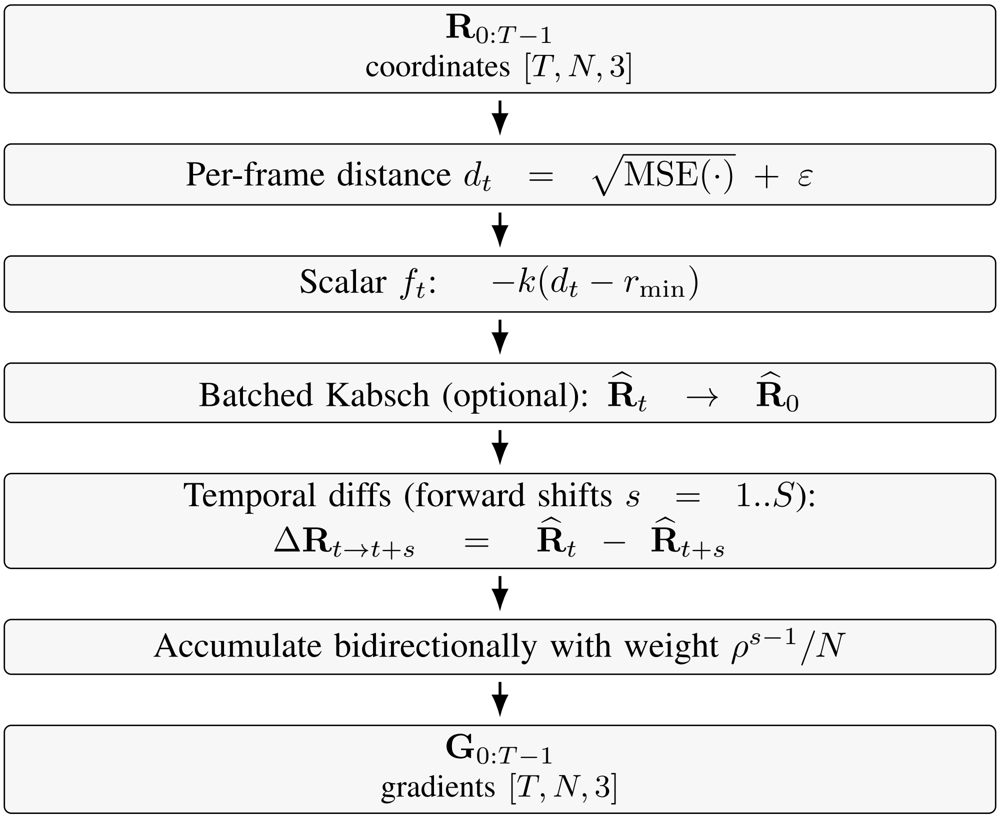


Figure 18: Data flow for the (score-free) Harmonic Adapter with optional LJ-style scalar and temporal neighbors.

Score-modulated adapter (reference) vs. score-free adapter (ours). Some practical variants *modulate the harmonic* by shifting the target with a score signal s_t (e.g. $r_{\min} \to r_{\min} + \alpha\, s_t$ at frame t), resulting in the potential

$$U_\alpha(X) = \frac{k}{2}\sum_{t,s} \alpha_s \big(r_{t,s}(X) - r_{\min} - \alpha s_t(X_t)\big)^2, \qquad \text{drift } -\nabla U_\alpha(X).$$

Our score-*free* adapter sets $\alpha = 0$ and keeps the score only in the *vertical* EM mean. Both choices compose with the vertical (i.i.d.) EM step, and both yield *exact* EM kernels for their respective total drifts (sum of vertical drift and horizontal harmonic), with the same covariance $2D\Delta t\, I$ and the same stiffness–step map $k(\Delta t)$.

Score manifold and the $r_{\min}$ calibration. Run the vertical (score-only) EM chain until it reaches its stationary tube; let $r_\star := \mathbb{E}[\, r_{t,s}(X)\,]$ under this i.i.d. score dynamics (same Δt). Set the adapter scale $r_{\min} := r_\star$. Then the horizontal residuals $c_{t,s}(X) = r_{t,s}(X) - r_{\min}$ are small with high probability; we call the thin tube $\mathcal{T}_\delta = \{\|c(X)\| \le \delta\}$ the *score manifold neighborhood*. On $\mathcal{T}_\delta$ the harmonic gradient is *exactly linear* in the residuals:

$$\nabla U_{\mathrm{glue}}(X) = k\, J_c(X)^\top \operatorname{diag}(\alpha)\, c(X), \qquad \nabla^2 U_{\mathrm{glue}}(X) = k\, J_c^\top \operatorname{diag}(\alpha)\, J_c + O(\|c(X)\|),$$

so the $O(\|c\|)$ curvature term *vanishes at equilibrium* ($c = 0$). This justifies reading the adapter as a linear spring near $r_{\min}$.

Small-residual approximation (score-free $\approx$ score-in-harmonic). Let $b_\alpha(X) = -g_n(X) - \nabla U_\alpha(X)$ (score *and* score-in-harmonic) and $b_0(X) = -g_n(X) - \nabla U_0(X)$ (score *and* score-*free* harmonic). Their drift difference on $\mathcal{T}_\delta$ is

$$\Delta b(X) \;:=\; b_\alpha(X) - b_0(X) \;=\; k\,\alpha \sum_{t,s} \alpha_s\, s_t(X_t)\, \nabla r_{t,s}(X),$$

i.e. a *score-weighted normal injection*. If $r_{\min}$ is calibrated by the i.i.d. score ($\mathbb{E}[c_{t,s}] = 0$) and s_t has bounded second moment along the score manifold, then

$$\mathbb{E}\|\Delta b(X)\|^2 = O(\alpha^2) \quad \text{on } \mathcal{T}_\delta,$$

and the finite-schedule pathwise KL between the two EM chains satisfies

$$\mathrm{DKL}\big(P_{b_0} \,\|\, P_{b_\alpha}\big) \;\le\; \frac{\beta}{4} \sum_{n=0}^{N-1} \Delta t \, \mathbb{E}\|\Delta b(X^n)\|^2 \;=\; O(\alpha^2 \, T),$$

so the score-free adapter *approximates* the score-in-Harmonic Adapter to *second order* in the modulation scale α. (The proof is the standard EM/Girsanov drift-difference bound used for the main path-KL result; see Sec. 3.4 and App. H.5.)

Thus, once equilibrium concentrates near $r_{\min}$, the residual-dependent curvature terms in the adapter vanish and the $r_{\min}$ connection acts as a linear spring, while the i.i.d. score supplies the full tangential drift.

Score is always on vertically. Horizontally, you may run a *score-free* $r_{\min}$ adapter. If $r_{\min}$ is calibrated from the i.i.d. score equilibrium or tuned, the chain enters a small-residual tube where the adapter is linear and the difference to a score-in-harmonic variant is a normal, $O(\alpha)$ perturbation per step with an $O(\alpha^2)$ path-KL over a horizon T. Meanwhile, the full *EM identity* (same covariance, stiffness–step map $k = \beta/(2\Delta t)$) holds exactly in both cases.

Nearest–neighbor and multi–neighbor cases. We can also define

$$U_{\mathrm{glue}}(X) \;=\; \frac{k}{2} \sum_{t=0}^{T-1} \sum_{s=1}^{S} \alpha_s \left(r_{t,s}(X) - r_{\min}\right)^2, \qquad k > 0. \tag{41}$$

with neighbor shift $s \in \{1, \ldots, S\}$). For $S = 1$ the path precision is tri-diagonal (time-chain Laplacian). For $S > 1$ the precision is banded and it becomes a higher-order temporal filter with eigenvalues $\lambda(\omega) = \frac{k}{N} \sum_s \alpha_s \, 2(1 - \cos s\omega)$, so the EM step damps high temporal frequencies more strongly while remaining an exact EM update for the same drift and covariance.

K CODE AVAILABILITY

Status. The system is *under active productization and covered by pending IP*. We therefore release a *bare but sufficient Python reference* implementing the exact Gaussian EM update with the Harmonic Adapter. A compact adapter implementation appears in *App. I*. In the Supplementary files, we supply a trajectory in sdf format of a flexible molecule that can be visualized.

What is not released. Training code, internal tooling, distillation/serving paths, and product integrations remain closed while IP is pending; none of these are required to validate the paper's claims about the EM equivalence and the resolution/temperature dial.