Analytic Gaussian Convolution for Faster Molecular Optimization and Sampling

Soumya Ram [1] **Akhila Ram** [2]

Abstract

For macromolecular structure optimization over internal torsion angles, the rugged, highly non-convex Lennard–Jones (LJ) potential poses a major obstacle. Our novel contribution lies in applying analytic Gaussian convolution to the LJ potential as a linear, shift-invariant smoothing operator that uniquely guarantees monotonic reduction of non-convexity without introducing new extrema. By deriving closed-form radial integrals and exact gradients in $\mathbb{R}^{3N}$, our method enables efficient, rotation- and translation-invariant smoothing. Across systems of 3 to 500 particles, we observe up to a 56% reduction in the optimization gap to the global minimum and over 75% closure of the sampling gap in toy benchmarks, translating into orders-of-magnitude higher probabilities of visiting near-optimal configurations. This framework opens a new avenue for scaling advanced neural samplers in biomolecular modeling.

1. Introduction

Accurately identifying the ground-truth structures of biomolecules and atomic or molecular clusters is a foundational task across domains such as drug development, materials design, and catalysis. While recent advances like AlphaFold3 have significantly improved protein structure prediction, their generalization to other biomolecule types remains limited. This is largely due to the fact that approximately 75 percent of AlphaFold3's training set consists solely of proteins, reducing its reliability for non-protein biomolecules. For example, it fails to correctly predict the structures of RNAs from orphan families (Zonta & Pantano, 2024).

Recent efforts aim to circumvent this issue by utilizing neural networks to sample from the Boltzmann distribution of macromolecules using only the closed-form energy function, without samples (Tan et al., 2025; Arbel et al., 2021; Phillips et al., 2024; Vargas et al., 2023).

However, these approaches have not scaled to large macromolecules due to the challenging nature of the energy function. The current SOTA (Akhound-Sadegh et al., 2024) is for a 55-dimensional Lennard-Jones particle system. This is because the number of local minima increases exponentially with dimension, and there are high energy barriers between local minima, preventing exploration.

In the typical coarse-grained approximation, the energy function's non-convexity is primarily driven by the Lennard-Jones potential. This is because bond lengths and bond angles are relatively stiff degrees of freedom, fluctuating minimally around their equilibrium values. As a result, many systems—including Rosetta—fix these degrees of freedom to their default values and optimize only over torsion angles (Das & Baker, 2008; Schwieters & Clore, 2001; Chen et al., 2005). In this setting, the energy function consists of only Lennard-Jones, Coloumb electrostatics, torsional contributions, with the torsion energy being far more benign. Torsion terms in modern biomolecular force fields are periodic and bounded, with Fourier amplitudes typically in the range 0.2–4 kcal mol^{-1}, giving total barrier heights of at most a few kcal mol^{-1} (Ponder & Case, 2003; MacKerell et al., 2000). By contrast, Lennard–Jones (LJ) non-bonded interactions have an attractive well depth ε of only $\lesssim 1$ kcal mol^{-1} for common biomolecular atom types, yet the repulsive part rises steeply—exceeding $+100$ kcal mol^{-1} when two atoms approach 0.8σ—making the LJ surface effectively unbounded and far more rugged (Allen & Tildesley, 1987). Coloumb electrostatics can also be analytically smoothed in the same manner, as it is simply a function of the distance between atoms. However, we focus this work the Lennard-Jones potential as it is a greater contributor to the non-convexity.

We introduce analytical Gaussian convolution of the Lennard–Jones potential as a principled, scalable smoothing operator that accelerates both optimization and sampling. Convolution with an isotropic Gaussian admits a closed-form—factorizing into one-dimensional integrals over pairwise distances—that directly reduces non-convexity by dampening sharp wells and rugged barriers. Crucially, it is

[*]Equal contribution [1]Independent Researcher [2]Department of Electrical Engineering and Computer Science, MIT. Correspondence to: Soumya Ram <soumya1910@gmail.com>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

the only linear, shift-invariant operator that obeys the causality property from scale-space theory (Babaud et al., 1986): as the smoothing scale σ increases, it removes all existing local minima and maxima without ever introducing new ones, gradually simplifying the landscape until it collapses to a single global minimum. Moreover, Gaussian smoothing can be viewed as the optimal affine approximation to the convex-envelope PDE (Mobahi & III, 2012). This guarantees a controlled and artifact-free way to smooth highly non-convex energy surfaces.

Once σ exceeds the system's diameter, the smoothed energy has a single critical point which is a global minimum. This structure is valuable for both major use cases: in optimization, the unique minimum provides a stable anchor point for homotopy continuation, allowing the original landscape to be recovered via gradual de-smoothing; in sampling, the convex energy defines a simple prior distribution for annealing.

Utilizing this analytical Gaussian convolution, we observe consistent improvements across a broad range of settings: small and large systems (3 to 500 particles), both optimization and sampling tasks, and both neural and non-neural samplers. For optimization, Gaussian homotopy continuation shrinks the gap to the global minimum by up to 56%, indicating better convergence and solution quality. For sampling, using the smoothed energy as an annealing prior accelerates mixing and dramatically improves recovery of low-energy states—achieving a $\approx 3.3\times$ lower minimum energy in the 13-particle system (from -4.65 to -15.28, reducing the gap to the -44.33 optimum by $\sim 27\%$), a $\approx 3.6\times$ lower minimum energy in the 55-particle system (from -14.26 to -51.16, reducing the gap to the -279.25 optimum by $\sim 14\%$), and recovering over 75% of the gap to the global minimum in the 3-particle system. These results confirm that analytic Gaussian smoothing enables SMC to penetrate deep energy basins that mixture-based paths systematically miss.

2. Related Work

Neural-network-accelerated Boltzmann Sampling There has been a flurry of recent work on utilizing neural networks to sample from Boltzmann distributions in the challenging scenario where one is given the closed-form unnormalized density but no samples from the distribution. Major methods include sequential Monte Carlo based approaches (Tan et al., 2025; Arbel et al., 2021) and diffusion samplers (Phillips et al., 2024; Vargas et al., 2023). However, scalability has been a challenge as the energy barriers and number of local minima increase exponentially with dimension.

To address this challenge, two major techniques have been proposed in the literature: temperature annealing and smoothing.

Temperature annealing can lower energy barriers (Luo et al., 2018), facilitating exploration. Indeed, this is behind the success of the neural sequential Monte Carlo approaches mentioned above. Tan et al. (2025) use sequential Monte Carlo along with a new architecture to achieve SOTA performance.

Another helpful technique has been smoothing the energy landscape. Akhound-Sadegh et al. (2024), the first to model a 55-particle Lennard-Jones system, convolved $\exp(-E(x))$ with the Gaussian distribution where $E(x)$ is the Lennard-Jones energy function. Because this convolution cannot be calculated analytically, it was estimated via MCMC sampling. However, MCMC sampling does not scale well to higher dimensions.

Gaussian Homotopy Continuation Gaussian homotopy continuation constructs a family of progressively less smoothed surrogates of a non-convex energy landscape by convolving the target function with Gaussian kernels of decreasing variance. Empirically, following this homotopy path has consistently outperformed vanilla gradient descent on highly multimodal benchmarks, enabling escape from narrow local minima and discovery of deeper basins (Mobahi, 2015; Hazan et al., 2016). It also enjoys theoretical grounding - it is the best affine approximation to the Vese's nonlinear PDE, a PDE that evolves a function to its convex envelope (Mobahi & III, 2012). In most practical applications, the closed-form expressions for the smoothed energy and its gradients are unavailable, so each homotopy step is estimated via MCMC sampling.

Analytically smoothing Lennard-Jones Convolving the Lennard-Jones potential with a Gaussian kernel reduces to a sum of 1D integrals, allowing efficient and precise evaluation. To our knowledge, this property has not been exploited in any prior work to accelerate sampling—e.g., through annealing or Langevin dynamics. Earlier work before 2000 did apply this idea for optimization via Gaussian homotopy continuation (Wu, 1996; Pappu et al., 1998), but this direction has seen little to no follow-up since 2000. Given modern developments—such as the ability to compute high-precision integrals and adaptive optimizers like Adam—a fresh analysis of its use for optimization is warranted.

3. Problem Formulation

Mixing time and spectral gap introduction According to Section 12.2 (Levin et al., 2017), let P be a discrete-time ergodic Markov kernel on state space $\mathcal{X}$ with stationary distribution π, and denote its spectral gap by $\gamma > 0$. Then

for all $x \in \mathcal{X}$ and $t \geq 0$,

$$\|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \;\leq\; e^{-\gamma\, t},$$

and the mixing time to within total-variation error ε satisfies

$$\tau_{\mathrm{mix}}(\varepsilon) = \inf\{t : \sup_x \|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq \varepsilon\} = \mathcal{O}\big(\gamma^{-1} \log \tfrac{1}{\varepsilon}\big).$$

For functions with small spectral gaps, it can be helpful to sample via temperature annealing as more time is spent in the high-temperature regime where the spectral gap is higher.

For a family of kernels $\{P_\beta\}_{\beta \geq 0}$ indexed by inverse temperature β, the mixing time at fixed β is $\tau_{\mathrm{mix}}(\beta) = \mathcal{O}\big(\gamma(\beta)^{-1}\big)$. By (Woodard et al., 2009), under a discrete annealing schedule $\beta_0 < \beta_1 < \cdots < \beta_K$, one may write the *effective* mixing time of tempered MCMC as

$$\tau_{\mathrm{temp}} \;=\; \sum_{k=0}^{K-1} \frac{\Delta\beta_k}{\gamma(\beta_k)} \;\xrightarrow{\text{cont.}}\; \int_{\beta_{\min}}^{\beta_{\max}} \frac{d\beta}{\gamma(\beta)}.$$

To improve the mixing time of the annealing schedule further, one can learn a schedule. However, by the perturbation bound, small perturbations in the learned transition kernel can induce large errors in its stationary distribution if the target function has a low spectral gap, by Section 4.4 (Levin et al., 2017).

$$\|\pi - \pi'\|_{\mathrm{TV}} \;\leq\; \tau_{\mathrm{mix}} \sup_x \big\|P(x,\cdot) - P'(x,\cdot)\big\|_{\mathrm{TV}},$$

so that approximation errors in the kernel are amplified by the mixing time.

Mixing time increases exponentially with particle size In the small-noise ($\beta \to \infty$) regime of overdamped Langevin dynamics on a potential $E : \mathbb{R}^d \to \mathbb{R}$ with multiple local minima separated by barrier height ΔE, the spectral gap obeys

$$\gamma(\beta) \;\approx\; A(\beta)\, \exp\big(-\beta\, \Delta E\big),$$

where $A(\beta)$ is a prefactor determined by the local curvatures around minima and saddle points, according to the Eyring–Kramers law (Bouchet & Reygner, 2016).

For Lennard-Jones, empirical evidence suggests ΔE grows (roughly) with the number of particles N (Wales & Doye, 1997), causing the spectral gap to decay exponentially in N.

The above indicates that biological energy functions have an inherent hardness that increases exponentially with dimension. This hardness hinders temperature annealing and causes learned samplers to learn an incorrect stationary distribution.

Learning a fast-mixing schedule via samples from the target distribution, as done in standard diffusion model training, avoids the curse of dimensionality (Gupta et al., 2024; Li & Yan, 2024). Neural sampling, which aims to sample from the target distribution without access to any ground-truth samples, is a far more challenging problem.

To address this, we propose smoothing the Lennard-Jones function with the Gaussian distribution. Biological energy functions consist of smooth, low-frequency, deep funnel(s) composed with shallow, high-frequency variation (Bryngelson et al., 1995). At an appropriate level of smoothing, one can remove the variation to recover the funnel structure. We demonstrate empirically that this leads to significant improvements in mixing time.

4. Gaussian Convolution Properties

Our goal is to accelerate optimization and sampling on the N-particle Lennard–Jones landscape by constructing a smoothed energy E_σ that

1. *monotonically reduces non-convexity*—as σ increases it can only eliminate local extrema and never introduce new ones, smoothing away narrow, deep wells that cause metastable trapping;
2. is *linear*, so that the full $\mathbb{R}^{3N}$ convolution reduces to an explicit one-dimensional integral over each pairwise distance, enabling efficient evaluation (see Eq. (1));
3. is *shift-invariant and radially symmetric*, preserving the translational and rotational invariance of the original energy landscape;
4. guarantees a *unimodal landscape* once σ exceeds the system's diameter, allowing us to control the complexity of the prior distribution by modulating σ.

In the sections that follow, we show that convolution with an isotropic Gaussian kernel meets these requirements: it reduces to the one-dimensional radial form in Eq. (1), yields existence and uniqueness of the smoothed minimum (Corollary 1), and—by the classical uniqueness theorem for linear, shift-invariant convolutions—is the only such operator that never introduces new local extrema as σ grows (Babaud et al., 1986).

Moreover, Gaussian convolution serves as the optimal linear proxy for PDEs that drive the function to its convex envelope, as further discussed Section 4.3.

4.1. Gaussian smoothing of the LJ energy

Let

$$X = (x_1, \ldots, x_N) \in \mathbb{R}^{3N},$$

where each $x_i \in \mathbb{R}^3$ is the position of particle i. Define the pairwise Lennard–Jones potential

$$V_{\mathrm{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma_{\mathrm{LJ}}}{r}\right)^{12} - \left(\frac{\sigma_{\mathrm{LJ}}}{r}\right)^{6}\right],$$

so that the total energy is

$$E(X) = \sum_{1\le i<j\le N} V_{\mathrm{LJ}}\big(\|x_i - x_j\|\big).$$

We smooth E by convolution with the isotropic Gaussian in $\mathbb{R}^{3N}$,

$$G_\sigma^{(3N)}(\Delta X) = \prod_{k=1}^{N} G_\sigma^{(3)}(\Delta x_k),$$
$$G_\sigma^{(3)}(u) = \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\!\left(-\frac{\|u\|^2}{2\sigma^2}\right).$$

where

$$\Delta X = (\Delta x_1, \ldots, \Delta x_N) \in \mathbb{R}^{3N},$$

and each $\Delta x_k \in \mathbb{R}^3$ denotes the isotropic Gaussian displacement applied to particle k.

We define the smoothed energy as

$$E_\sigma(X) = (E * G_\sigma^{(3N)})(X)$$
$$= \int_{\mathbb{R}^{3N}} E(X - \Delta X)\, G_\sigma^{(3N)}(\Delta X)\, d^{3N}\Delta X.$$

By linearity and the product form of $G_\sigma^{(3N)}$, each 3-dimensional "spectator" convolution integrates to one, and one is left with an integral only over the relative distance u for each particle-pair. In Appendix A we show that

$$E_\sigma(X) = \sum_{1\le i<j\le N} V_\sigma(\|x_i - x_j\|),$$

where the *radialized* potential is

$$V_\sigma(r) = \int_0^\infty \frac{u\, V_{\mathrm{LJ}}(u)}{2\, r\sqrt{\pi}\,\sigma}\left[e^{-\frac{(r-u)^2}{4\sigma^2}} - e^{-\frac{(r+u)^2}{4\sigma^2}}\right] du. \quad (1)$$

4.2. Existence and uniqueness of the smoothed minimum

We show that Gaussian smoothing of a regularized Lennard-Jones energy admits a unique critical point – a global minimum – once the smoothing scale exceeds the radius of the configuration domain.

Regularized LJ and compact support. We take the standard LJ pair potential

$$V_{\mathrm{LJ}}(r) = 4\varepsilon\big[(\sigma/r)^{12} - (\sigma/r)^{6}\big], \quad r > r_0 > 0,$$

and replace its singular core for $r \le r_0$ by the unique affine extension matching value and derivative at r_0. The resulting bounded-above $V_{\mathrm{LJ}}(r)$ yields a total energy

$$E(X) = \sum_{1\le i<j\le N} V_{\mathrm{LJ}}(\|x_i - x_j\|).$$

Fix a compact set $D \subset (\mathbb{R}^3)^N$ containing all configurations of interest, and let

$$C = \max_{X\in D} E(X).$$

Define

$$f(X) = \begin{cases} C - E(X), & X \in D, \\ 0, & X \notin D, \end{cases}$$

so that $f \ge 0$ is compactly supported. Denote by r the minimal radius so that $\mathrm{supp}(f) \subset B_r(0)$.

Uniqueness via Gaussian smoothing. Let $g_\sigma(X) = (2\pi\sigma^2)^{-3N/2} e^{-\|X\|^2/(2\sigma^2)}$, and write

$$f_\sigma = f * g_\sigma, \qquad E_\sigma = E * g_\sigma.$$

By a classical result of Loog & Welling (Loog et al., 2001),

Theorem 1. *If $f \ge 0$ is compactly supported in $B_r(0)$, then for every $\sigma > r$, f_σ has exactly one critical point, which is a global maximum.*

Since $f_\sigma = C - E_\sigma$, it follows immediately that:

Corollary 1. *For $\sigma > r$, the smoothed energy E_σ has exactly one critical point (where $\nabla E_\sigma = 0$), and this point is a global minimum. This point is at X = 0.*

4.3. Optimality of the Gaussian kernel

Finally, we note that among all linear, shift-invariant convolution operators, only the Gaussian kernel is guaranteed never to introduce new local extrema as the smoothing scale increases—a result established by Babaud et al. (Babaud et al., 1986).

This extremum-preserving property mirrors the behavior of Vese's geometric flow for convex-envelope generation, which evolves an initial function $v(x,t)$ according to

$$\frac{\partial v}{\partial t} = \sqrt{1 + \|\nabla v\|^2}\, \min\{0,\ \lambda_{\min}(\nabla^2 v)\},\ v(x,0) = f(x),$$

so that smoothing occurs only at points of local non-convexity (where the smallest Hessian eigenvalue is negative), and as $t \to \infty$, $v(\cdot,t)$ converges to the convex

envelope of f (Vese, 1999). Moreover, among all linear, shift-invariant operators, Gaussian convolution arises as the best affine approximation to this inherently nonlinear PDE (Mobahi & III, 2012).

5. Experiments

In this section, we validate our analytic Gaussian smoothing in two complementary settings that together cover the core tasks of molecular energy landscapes. First, we use Gaussian homotopy continuation as a drop-in replacement for standard gradient-based optimizers, testing whether our closed-form convolution helps escape local minima and converge more reliably toward known global optima on systems ranging from 100-500 particles. Second, we embed the same smoothing into sequential Monte Carlo samplers to measure mixing efficiency and minimum energies under a fixed MCMC budget on moderate-dimensional Lennard–Jones problems. By evaluating both optimization and sampling, we demonstrate that Gaussian convolution annealing delivers practical speedups and deeper basin exploration compared to traditional mixture-based schedules.

We find that Gaussian homotopy continuation reduces the energy gap to the global minimum by around 40% compared to Adam on high-dimensional optimization tasks. Furthermore, these benefits also extend to neural samplers - where Gaussian-convolution annealing was able to close over 75% of the gap to the global optimum compared to standard methods on a toy problem. Notably, since sampling probabilities scale as $e^{-E(x)}$, even a modest reduction in the energy gap yields an exponential boost in the likelihood of visiting near–global-minimum configurations.

5.1. Gaussian Homotopy Continuation

Gaussian homotopy continuation constructs a family of smoothed objectives

$$E_\sigma(x) \;=\; (E * \mathcal{G}_\sigma)(x),$$

where $\mathcal{G}_\sigma$ is an isotropic Gaussian of variance σ^2. Starting from a large σ, we solve

$$x^{(\sigma)} \;=\; \arg\min_x E_\sigma(x)$$

using a standard gradient-based optimizer. We then decrease σ according to a predefined schedule and re-initialize the optimizer at $x^{(\sigma)}$, thereby obtaining a sequence of approximate minimizers $\{x^{(\sigma)}\}$ that migrate toward the original landscape $E(x)$.

As shown in Section 4, Gaussian convolution is preferred over ad-hoc mixtures (e.g. $\pi_0(x)^{\lambda}\, e^{-(1-\lambda)E(x)}$, where π_0 is a simple base distribution) or over other non-Gaussian convolutions due to its unique theoretical properties. This ensures that critical basin structures and low-energy manifolds of E are retained more faithfully during smoothing.

While most objective functions lack a closed-form Gaussian convolution—forcing prior Gaussian homotopy continuation methods to approximate E_σ via Monte Carlo sampling— the Lennard–Jones potential admits an exact analytic convolution with an isotropic Gaussian. Leveraging this analytic reduction to a single one-dimensional radial integral, we compute gradients for 500-particle systems on the order of minutes.

Experimental setup. In this section, we compare homotopy continuation (with our analytic smoothing) against standard Adam optimization on the original, unsmoothed LJ potential. Both use learning rate decay on plateau (by a factor of 10) and early stopping. Each setting is repeated across 5 independent runs. See Appendix D for more details.

Results. Gaussian homotopy continuation consistently achieves lower-energy minima than Adam. For $N = 100$, Adam's best run ends at -530.4, leaving a gap of 26.6 energy units, whereas Gaussian homotopy reaches -541.8, reducing the gap to 15.2 units—an absolute gain of 11.4 units and a relative reduction of 42.9%. At $N = 250$, the minimum gap shrinks from 81.7 to 35.6 units (an absolute improvement of 46.1 and a 56.4% reduction), and at $N = 500$, from 192.1 to 120.1 units (a 72.0-unit gain, or 37.5% reduction). Median energies likewise improve. These gaps in $E(x)$ become even more significant in the context of particle likelihood, which scales as $\exp(-E(x))$, indicating that the solutions found by Gaussian homotopy are many orders of magnitude more likely.

Method	**Metric**	$N = 100$	$N = 250$	$N = 500$
Adam	Median	-526.9	-1482.2	-3170.2
	Min	-530.4	-1498.1	-3190.6
Homotopy	Median	$\mathbf{-534.7}$	$\mathbf{-1501.4}$	$\mathbf{-3205.0}$
	Min	$\mathbf{-541.8}$	$\mathbf{-1544.2}$	$\mathbf{-3262.6}$
Global Minimum		-557.0	-1579.8	-3382.7
% Gap Reduction		42.9%	56.4%	37.5%

Table 1. Final Lennard-Jones (LJ) energies across 5 independent runs per method and particle count. We report the median and minimum energy achieved, along with the known global minimum from the Cambridge Cluster Database (Wales et al., 2001). Gaussian homotopy continuation consistently outperforms Adam, converging to lower energies and reducing the gap to the global minimum more effectively. The final row quantifies the percent reduction in this energy gap (Adam's minimum to global min) achieved by switching from Adam to Gaussian Homotopy.

5.2. Sequential Monte Carlo Methods

In this section, we employ two related population-based sampling frameworks—standard Sequential Monte Carlo (SMC) and Neural SMC via Annealed Flow Transport (Arbel et al., 2021)—to explore the Lennard–Jones energy landscape.

Standard SMC. SMC approximates a target distribution $\pi(x) \propto e^{-E(x)}$ by evolving a population of M weighted particles $\{(x_t^{(i)}, w_t^{(i)})\}_{i=1}^M$ through a sequence of intermediate densities $\{\pi_t\}_{t=0}^T$. At each stage t, the algorithm proceeds as follows:

1. **Mutation (ULA step).** We apply one Unadjusted Langevin Algorithm (ULA) update, which discretizes the continuous overdamped Langevin diffusion

$$dX = \nabla \log \pi_t(X)\, dt + \sqrt{2}\, dW_t$$

with step size ε:

$$x_t^{(i)} = x_{t-1}^{(i)} + \tfrac{\varepsilon}{2} \nabla \log \pi_t(x_{t-1}^{(i)}) + \sqrt{\varepsilon}\, \xi_t^{(i)}, \quad \xi_t^{(i)} \sim \mathcal{N}(0, I).$$

This gradient-informed proposal mixes more efficiently than a random-walk Metropolis kernel.

2. **Weighting:** Particles are reweighted to account for the change in target density:

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{\pi_t(x_t^{(i)})}{\pi_{t-1}(x_t^{(i)})}.$$

3. **Resampling:** When the effective sample size falls below a threshold, particles are resampled to prevent weight degeneracy.

Neural SMC. Annealed Flow Transport (AFT) (Arbel et al., 2021) enhances standard SMC by inserting learned, invertible transport maps between successive intermediate targets. At each stage k, a lightweight normalizing flow is trained to push the particle approximation of π_{k-1} toward π_k; the transported particles are then reweighted (and resampled if necessary) before a brief ULA mutation. This learned alignment sharply reduces weight variance and accelerates mixing compared to pure importance-resampling.

The choice of annealing path $\{\pi_t\}$ critically controls mixing: poor paths trap particles in narrow wells or collapse the effective sample size.

For both methods, we consider two annealing paths:

- **Mixture-based annealing**, which interpolates between a simple Gaussian prior with variance σ_{fixed}^2 and the Boltzmann target by gradually decreasing the mixing parameter $\lambda \in [0, 1]$:

$$\pi_k(x) \propto \pi_0(x)^{\lambda}\, \pi_K(x)^{1-\lambda},$$

where

$$\pi_0(x) = \mathcal{N}\big(x; 0, \sigma_{fixed}^2 I\big),\ \pi_K(x) \propto \exp\big(-E(x)\big).$$

- **Gaussian-convolution annealing**, which replaces the original energy E with its Gaussian-smoothed version E_σ and then linearly decreases the smoothing width σ:

$$E_\sigma(x) = (E * G_\sigma^{(3N)})(x),\ \pi_\sigma(x) \propto \exp\big(-E_\sigma(x)\big).$$

We find that in both SMC and Neural SMC, Gaussian-convolution annealing accelerates mixing and drives particles into deeper minima—improving minimum energies by roughly an order of magnitude compared to mixture-based annealing. Even in low-dimensional settings (e.g. a 3-particle system), mixture annealing suffers severe sampling pathologies and struggles to locate deep wells.

5.2.1. Standard Sequential Monte Carlo

We benchmark the Gaussian convolution and mixture-based annealing schedules on 13- and 55-particle Lennard-Jones systems. Each schedule drives $T = 10^4$ ULA steps per chain across five independent runs, using a population of 1,000 replicas.

To isolate mixing rather than importance–weight correction, we omit the weighting/resampling stages: every replica evolves with identical step sizes and remains equally weighted. This choice makes the experiment *more demanding*—without resampling, the chains themselves must mix rapidly—so only genuinely efficient annealing paths will perform well. Because the ULA proposal is Markov–chain–exact for the intermediate targets in the limit of small ε, any residual bias is identical for both schedules; differences in performance therefore isolate the mixing efficiency of each path.

Gaussian convolution annealing. We begin with a burn-in of ten ULA steps targeting the distribution defined by the LJ energy smoothed with a Gaussian kernel of initial width $\sigma_{initial}$ = 1.5. We then linearly decrease σ from $\sigma_{initial}$ to 0.025 over the remaining 10^4 ULA steps, performing one ULA update per σ and carrying forward the chain state.

Mixture annealing. Particles are initialized from an isotropic Gaussian with variance σ_{fixed}^2 (where $\sigma_{fixed} = 1.5$ to match the above setting). Over 10^4 ULA steps, we decrease the mixing parameter λ from 1 to 0 under two spacing strategies:

- *Linear schedule*: λ decreases by a fixed increment each step.
- *Geometric schedule*: λ values are updated by multiplying by a fixed ratio of 0.95 at each step, starting from $\lambda = 1$ and decaying toward 0.

At each step, we perform one ULA update targeting the current mixture distribution and carry forward the chain state.

Results. As shown in Table 2 (13-particle LJ) and Table 3 (55-particle LJ), Gaussian-convolution annealing consistently achieves substantially lower mean minimum energies than either mixture-based schedule. In the 13-particle system, it attains -15.28 ± 0.78 compared to 4.65 ± 1.06 for the geometric λ schedule and -0.02 ± 0.002 for the linear λ schedule. In the 55-particle system, it reaches -51.16 ± 1.60 versus -14.26 ± 5.93 for the geometric schedule and -0.12 ± 0.005 for the linear schedule, demonstrating a clear and consistent advantage of Gaussian-convolution annealing across both problem sizes.

The top panel of Figure 1 reveals that under Gaussian convolution for the LJ-13 system, the sample distribution (orange) spans roughly -15 to 0 and is tightly concentrated between -11 and -3, almost entirely below the minimum energies reached by mixture-based schedule (blue). This demonstrates that analytic Gaussian smoothing drives SMC into deep energy basins that mixture annealing systematically fails to explore.

Similarly, the bottom panel of Figure 1 shows that for LJ-55 the Gaussian-smoothed distribution extends from about -55 to 0, whereas the mixture-based schedule yields virtually no low-energy samples. These observations confirm that Gaussian convolution annealing consistently penetrates deep minima that mixture-based paths miss across both system sizes.

Method	Mean Min. Energy	Std. Dev.
Gauss. ann. (lin. σ)	-15.28	0.78
Mix. ann. (geom. λ)	-4.65	1.06
Mix. ann. (lin. λ)	-0.02	0.002
Global Minimum	-44.33	—

Table 2. 13 dimensional LJ. Mean minimum energy over 5 chains after 10^4 ULA steps. Gaussian convolution annealing achieves substantially lower energies than either mixture variant.

5.2.2. Neural Sequential Monte Carlo

We evaluate the performance of the two annealing distributions for Annealed Flow Transport (AFT) Monte Carlo (Arbel et al., 2021).

Method	Mean Min. Energy	Std. Dev.
Gauss. ann. (lin. σ)	-51.16	1.60
Mix. ann. (geom. λ)	-14.26	5.93
Mix. ann. (lin. λ)	-0.12	0.005
Global Minimum	-279.25	—

Table 3. 55 dimensional LJ. Mean minimum energy over 5 chains after 10^4 ULA steps. Gaussian convolution annealing achieves substantially lower energies than either mixture variant.

To highlight sampling pathologies in even the simplest case, we evaluate on a three-particle Lennard–Jones system using the equivariant flow from (Köhler et al., 2020). The Gaussian convolution and mixture annealing distributions are the same as those described in the previous section. In addition, we set σ to be 1.1.

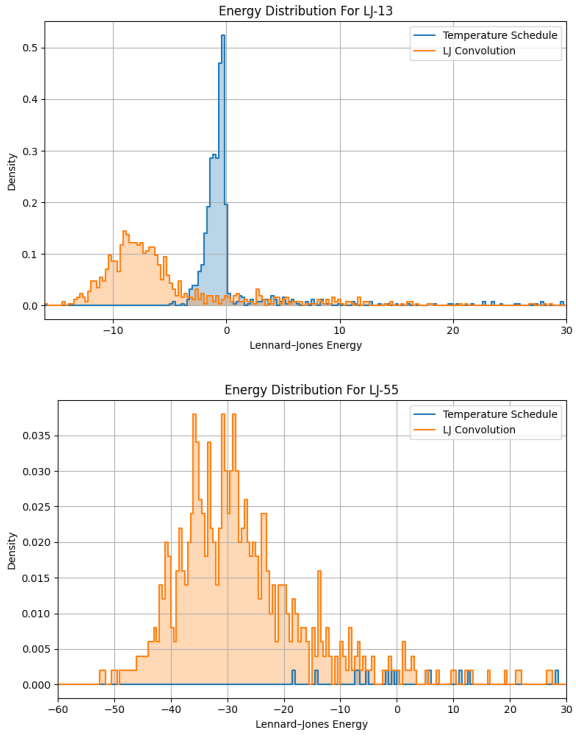

Figure 1. Energy histograms for LJ_{13} (top) and LJ_{55} (bottom) after 10^4 sampling steps from Gaussian convolution annealing (orange) and mixture annealing (blue). Gaussian annealing exhibits a heavier low-energy tail, indicating faster mixing into deep minima compared to mixture annealing. The experiment was run with 1000 particles. For clarity's sake, the x-axis of the graphs are clipped at 30. However, the sample fractions are the true fractions prior to clipping.

We train each flow for two stages.

Annealing schedules.

- **Mixture-based:** In stage one, the mixing parameter λ decreases from 1.0 to 0.6, and in stage two from 0.6 to 0, interpolating between the target $\exp(-E)$ and the Gaussian prior $\mathcal{N}(0,\, 1.1^2 I)$.

- **Gaussian-smoothed:** In stage one, the smoothing width σ decreases from 1.1 to 0.6, and in stage two from 0.6 to 0, targeting $\exp(-E_\sigma)$.

Training details. We train two flow models sequentially (one per annealing stage) for 150 epochs each using Adam ($\text{lr} = 2 \times 10^{-4}$) on 2,000 training, 500 validation, and 1,000 test samples. A log-determinant penalty $\mathbb{E}[(\log\det J_f)^2]$ and energy rescaling of $\frac{1}{40}$ are applied to prevent effective sample size collapse upon reweighting.

Results. Gaussian-smoothed AFT drives the mean test-set minimum from -1.13 ± 0.07 (mixture-based) down to -2.58 ± 0.06, a net improvement of 1.45 LJ units (roughly a 130% deeper well). Moreover, the gap to the known global minimum -3.00 shrinks by over 75%, from 1.87 to 0.42 units. These gains confirm that even in the simplest 3-particle setting, analytic Gaussian smoothing substantially improves sampler quality and basin exploration compared to conventional mixture schedules.

Method	Mean	Std
Mixture-based	-1.1329	0.0715
Gaussian-smoothed	-2.5822	0.0629
Global minimum	-3.0000	–

Table 4. Test-set minimum LJ energy (mean $\pm$ std) over 10 seeds.

6. Extension to Torsional Energy Terms

In this work we restrict our smoothing to Lennard-Jones interactions (our method generalizes to Coloumb electrostatics as well) and leave torsional terms unsmoothed, but three natural extensions could be pursued. First, one might simply omit torsion smoothing and rely on the dominant LJ convolution, as LJ is the main contributor to non-convexity. Second, one could apply Gaussian convolution directly in dihedral-angle space—performing a one-dimensional integral on each S^1 torsion angle—to eliminate local extrema; this is computationally trivial and preserves periodicity, yet it operates in a different domain than the coordinate-space LJ smoothing. Third, one could convolve each four-atom torsion term with an isotropic Gaussian in $\mathbb{R}^{12}$ while restricting the integrand to the torsion manifold of fixed angles and bonds via an indicator or delta function. This leads to a double integral with an outer 1-D convolution over the dihedral angle ϕ, and for each ϕ, an inner integral over the residual rigid-body rotations (3-D). We leave a thorough evaluation of these torsion-smoothing strategies to future work.

7. Conclusion

Energy landscapes of molecular systems are notoriously difficult to navigate: even when optimizing just the intrinsic coordinates (torsion angles), the Lennard-Jones potential drives extreme non-convexity, with steep repulsive walls exceeding $+100\ \text{kcal}\,\text{mol}^{-1}$ and a proliferation of local minima. We offer a principled remedy—analytical Gaussian convolution of the LJ potential—as the unique linear, shift-invariant smoothing operator that monotonically reduces non-convexity without introducing artifacts. This novel observation yields closed-form, one-dimensional radial integrals and exact gradients, enabling efficient, scalable evaluation in $\mathbb{R}^{3N}$.

Our empirical results on systems from 3 to 500 particles confirm substantial quantitative gains. In optimization, Gaussian homotopy continuation reduces the gap to the global minimum by 56.4% compared to Adam for $N = 250$. In sampling, Gaussian-convolution annealing achieves a mean minimum energy of -15.28 ± 0.78 versus -4.65 ± 1.06 under geometric mixture for the 13-particle LJ system—reducing the energy gap to the global optimum by approximately 26.8%. For the 55-particle system, it reaches -51.16 ± 1.60 compared to -14.26 ± 5.93, corresponding to a gap reduction of about 13.9%. In Neural SMC, it reduces the residual gap from 1.87 to 0.42 LJ units (a 77.5% improvement). These improvements consistently outperform both Adam and mixture-based schedules.

By bridging scale-space theory with molecular modeling, we provide a scalable, artifact-free framework to smooth highly rugged energy surfaces. We hope this technique can be adopted to scale advanced neural samplers to larger and more complex bio-molecular systems.

Software and Data

Code to reproduce all experiments can be found here.

Impact Statement

We present a technique that improves optimization and sampling from biological energy functions. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Akhound-Sadegh, T., Rector-Brooks, J., Bose, A. J., Mittal, S., Lemos, P., Liu, C.-H., Sendera, M., Ravanbakhsh, S., Gidel, G., Bengio, Y., Malkin, N., and Tong, A. Iterated

denoising energy matching for sampling from boltzmann densities. *arXiv preprint arXiv:2402.06121*, 2024.

Allen, M. P. and Tildesley, D. J. *Computer Simulation of Liquids*. Oxford University Press, 1987.

Arbel, M., Matthews, A. G. D. G., and Doucet, A. Annealed flow transport monte carlo. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 318–330. PMLR, 2021.

Babaud, J., Witkin, A. P., Baudin, M., and Duda, R. O. Uniqueness of the gaussian kernel for scale-space filtering. *IEEE transactions on pattern analysis and machine intelligence*, (1):26–33, 1986.

Bouchet, F. and Reygner, J. Generalisation of the eyring–kramers transition rate formula to irreversible diffusion processes. *Annales Henri Poincaré*, 17(12):3499–3532, 2016. doi: 10.1007/s00023-016-0507-4.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Genetics*, 21(3):167–195, March 1995. doi: 10.1002/prot.340210302.

Chen, J., Im, W., and Brooks, C. L. Application of torsion angle molecular dynamics for efficient sampling of protein conformations. *Journal of Computational Chemistry*, 26(15):1565–1578, 2005. doi: 10.1002/jcc.20269.

Das, R. and Baker, D. Macromolecular modeling with rosetta. *Annual Review of Biochemistry*, 77:363–382, 2008. doi: 10.1146/annurev.biochem.77.062906.171838.

Gupta, S., Parulekar, A., Price, E., and Xun, Z. Improved sample complexity bounds for diffusion model training. In Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J. M., and Zhang, C. (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10–15, 2024*, 2024.

Hazan, E., Levy, K. Y., and Shalev-Shwartz, S. On graduated optimization for stochastic nonconvex problems. In *Proceedings of the 33rd International Conference on Machine Learning*, pp. 1833–1841, 2016.

Köhler, J., Klein, L., and Noé, F. Equivariant flows: Exact likelihood generative learning for symmetric densities. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 5361–5370. PMLR, 2020.

Levin, D. A., Peres, Y., and Wilmer, E. L. *Markov Chains and Mixing Times*. AMS Textbooks. American Mathematical Society, Providence, RI, 2 edition, 2017. ISBN 978-1-4704-2962-1. doi: 10.1090/mbk/107. URL https://doi.org/10.1090/mbk/107.

Li, G. and Yan, Y. O(d/t) convergence theory for diffusion probabilistic models under minimal assumptions. *arXiv preprint arXiv:2409.18959*, 2024. doi: 10.48550/arXiv.2409.18959. URL https://arxiv.org/abs/2409.18959.

Loog, M., JisseDuistermaat, J., and Florack, L. M. On the behavior of spatial critical points under gaussian blurring a folklore theorem and scale-space constraints. In *Scale-Space and Morphology in Computer Vision: Third International Conference, Scale-Space 2001 Vancouver, Canada, July 7–8, 2001 Proceedings 3*, pp. 183–192. Springer, 2001.

Luo, R., Wang, J., Yang, Y., Zhu, Z., and Wang, J. Thermostat-assisted continuously-tempered hamiltonian monte carlo for bayesian learning. In *Advances in Neural Information Processing Systems*, volume 31, pp. 1–11. Curran Associates, Inc., 2018. URL https://arxiv.org/abs/1711.11511.

MacKerell, A. D., Banavali, N. G., and Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers: Original Research on Biomolecules*, 56(4):257–265, 2000.

Mobahi, H. A theoretical analysis of optimization by gaussian continuation. *arXiv preprint arXiv:1510.05632*, 2015.

Mobahi, H. and III, J. W. F. On the link between gaussian homotopy continuation and convex envelopes. In *NIPS*, 2012.

Pappu, R. V., Hart, R. K., and Ponder, J. W. Analysis and application of potential energy smoothing and search methods for global optimization. *The Journal of Physical Chemistry B*, 102(48):9725–9742, 1998. doi: 10.1021/jp982255t. URL https://pubs.acs.org/doi/10.1021/jp982255t.

Phillips, A., Dau, H.-D., Hutchinson, M. J., De Bortoli, V., Deligiannidis, G., and Doucet, A. Particle denoising diffusion sampler. *arXiv preprint arXiv:2402.06320*, 2024. URL https://arxiv.org/abs/2402.06320.

Ponder, J. W. and Case, D. A. Force fields for protein simulations. *Advances in Protein Chemistry*, 66:27–85, 2003.

Schwieters, C. D. and Clore, G. M. Internal coordinates for molecular dynamics and minimization in structure determination and refinement. *Journal of Magnetic Resonance*, 152(2):288–302, 2001. doi: 10.1006/jmre.2001.2421.

URL https://www.gmclore.org/clore/Pub/pdf/345.pdf.

Tan, C. B., Bose, A. J., Lin, C., Klein, L., Bronstein, M. M., and Tong, A. Scalable equilibrium sampling with sequential boltzmann generators. *arXiv preprint arXiv:2502.18462*, 2025.

Vargas, F., Grathwohl, W., and Doucet, A. Denoising diffusion samplers. In *Proceedings of the Eleventh International Conference on Learning Representations (ICLR)*, 2023. URL https://arxiv.org/abs/2302.13834.

Vese, L. A method to convexify functions via curve evolution. *Communications in partial differential equations*, 24(9-10):1573–1591, 1999.

Wales, D. J. and Doye, J. P. K. Evolution of the potential energy surface with size for lennard-jones clusters. *The Journal of Chemical Physics*, 107(16):6160–6170, 1997. doi: 10.1063/1.474547.

Wales, D. J., Doye, J. P. K., Dullweber, A., Hodges, M. P., Naumkin, F. Y., Calvo, F., Hernández-Rojas, J., and Middleton, T. F. The cambridge cluster database. http://www-wales.ch.cam.ac.uk/CCD.html, 2001.

Woodard, D. B., Schmidler, S. C., and Huber, M. L. Conditions for rapid mixing of parallel and simulated tempering on multimodal distributions. *Annals of Applied Probability*, 19(2):617–640, 2009. doi: 10.1214/08-AAP552.

Wu, Z. The effective energy transformation scheme as a special continuation approach to global optimization with application to molecular conformation. *SIAM Journal on Optimization*, 6(3):748–768, 1996. doi: 10.1137/S1052623493254698. URL https://epubs.siam.org/doi/10.1137/S1052623493254698.

Zonta, F. and Pantano, S. From sequence to mechanobiology? promises and challenges for alphafold 3. *Heliyon*, 10(10):e2949907024000469, 2024. doi: 10.1016/j.heliyon.2024.e2949907024000469. URL https://www.sciencedirect.com/science/article/pii/S2949907024000469.

A. Derivation of the one-dimensional radial integral

We show how the 3-dimensional convolution for one particle pair

$$I(r) \;=\; \int_{\mathbb{R}^3} V_{\mathrm{LJ}}(\|r-u\|)\; G_{\mathrm{rel}}(u)\; d^3u,$$

with

$$G_{\mathrm{rel}}(u) = \frac{1}{(4\pi\sigma^2)^{3/2}}\exp\Big(-\frac{\|u\|^2}{4\sigma^2}\Big),$$

reduces to a 1-dimensional radial integral. Here $r = \|x_i - x_j\|$.

1. Switch to spherical coordinates

Align r along the polar axis. In spherical (u, θ, ϕ):

$$d^3u = u^2 \sin\theta\, du\, d\theta\, d\phi, \quad \|r-u\|^2 = r^2 + u^2 - 2\,r\,u\,\cos\theta.$$

Thus

$$I(r) = \frac{2\pi}{(4\pi\sigma^2)^{3/2}} \int_0^\infty u^2\, e^{-u^2/(4\sigma^2)}\, du \int_{-1}^{1} V_{\mathrm{LJ}}\big(\sqrt{r^2+u^2-2ru\mu}\big)\, d\mu,$$

with $\mu = \cos\theta$.

2. Exact angular integral

Use

$$\int_{-1}^{1} e^{-a(r^2+u^2-2ru\mu)}\, d\mu = \frac{e^{-a(r-u)^2} - e^{-a(r+u)^2}}{a\,r\,u}, \quad a = \frac{1}{4\sigma^2}.$$

Since V_{LJ} depends only on $\sqrt{r^2+u^2-2ru\mu}$, inserting this identity yields the "difference-of-Gaussians" form inside the radial integral.

3. Final 1-D form

Collecting constants and simplifying gives

$$V_\sigma(r) = \int_0^\infty u\, V_{\mathrm{LJ}}(u)\; \frac{1}{2\,r\,\sqrt{\pi}\,\sigma}\Big[e^{-(r-u)^2/(4\sigma^2)} \;-\; e^{-(r+u)^2/(4\sigma^2)}\Big]\, du.$$

Hence the globally smoothed energy is

$$E_\sigma(X) = \sum_{1\le i<j\le N} V_\sigma\big(\|x_i - x_j\|\big).$$

B. Computational Efficiency of Convolution

For each fixed σ we compute one integral up to the largest pair distance; every other r simply re-uses this result, so the overhead minimally increases with dimension.

C. Lennard-Jones formulation

To deal with the Lennard-Jones singularity at r = 0, for r less than 0.8, the function was modified to be linear. Furthermore, the minimum sigma value for gaussian smoothing was 0.025, as the integral become unstable for lower values of sigma.

D. Gaussian Homotopy Continuation Details

All experiments were run for 10,000 steps and the configuration with the best energy was taken. The initial positions were sampled from a standard normal distribution. The learning rate started at 1e-1 and decreased by a factor of 10 upon 1000 steps with no improvement. For the gaussian homotopy experiment, the initial sigma value for smoothing was 1.0. When the gradient norm dipped below 1e-5, the sigma value was multiplied by 0.95. We take the configuration with the best energy for a given sigma value as the starting point for the minimization on the subsequent sigma value. After the sigma value of 0.025 was reached, we additionally minimize with the un-smoothed Lennard-Jones function.

E. Sequential Monte Carlo

Prior Selection. We find that a σ value between 1.0 to 1.5 works well across tasks and particle sizes. This is because a sigma value in this range smooths the Lennard-Jones function to be almost flat. To sample from the initial sigma value, we first sample from an iid Gaussian with the same sigma value and then apply 10 burn-in Metropolis-Hastings steps. For the temperature annealing, we simply sample from the iid Gaussian with the same sigma value.

Ground-Truth MCMC Processing We retrieve the raw data for the 100 k experiment from (Köhler et al., 2020) here. We extracted near-independent energy samples from the 100 k–step baseline as follows:

1. Discard the first 20 k steps (burn-in).
2. Estimate the integrated autocorrelation time τ_{int} of the post–burn-in energy series.
3. Thin the trajectory by a factor of $\lceil 2\tau_{\mathrm{int}} \rceil$.
4. Evaluate LJ energies on the thinned configurations for all downstream analyses.