

Learning Energy-Based Generative Models via Potential Flow: A Variational Principle Approach to Probability Density Homotopy Matching

Junn Yong Loo

Monash University Malaysia

loo.junnyong@monash.edu

Michelle Adeline

Monash University Malaysia

made0008@student.monash.edu

Julia Kaiwen Lau

Monash University Malaysia

julia.lau@monash.edu

Fang Yu Leong

Monash University Malaysia

leong.fangyu@monash.edu

Hwa Hui Tew

Monash University Malaysia

hwa.tew@monash.edu

Arghya Pal

Monash University Malaysia

arghya.pal@monash.edu

Vishnu Monn Baskaran

Monash University Malaysia

vishnu.monn@monash.edu

Chee-Ming Ting

Monash University Malaysia

ting.cheeming@monash.edu

Raphaël C.-W. Phan

Monash University Malaysia

raphael.phan@monash.edu

Abstract

Energy-based models (EBMs) are a powerful class of probabilistic generative models due to their flexibility and interpretability. However, relationships between potential flows and explicit EBMs remain underexplored, while contrastive divergence training via implicit Markov chain Monte Carlo (MCMC) sampling is often unstable and expensive in high-dimensional settings. In this paper, we propose Variational Potential Flow Bayes (VPFB), a new energy-based generative framework that eliminates the need for implicit MCMC sampling and does not rely on auxiliary networks or cooperative training. VPFB learns an energy-parameterized potential flow by constructing a flow-driven density homotopy that is matched to the data distribution through a variational loss minimizing the Kullback-Leibler divergence between the flow-driven and marginal homotopies. This principled formulation enables robust and efficient generative modeling while preserving the interpretability of EBMs. Experimental results on image generation, interpolation, out-of-distribution detection, and compositional generation confirm the effectiveness of VPFB, showing that our method performs competitively with existing approaches in terms of sample quality and versatility across diverse generative modeling tasks.

1 Introduction

Energy-based models (EBMs) have emerged as a flexible and expressive class of probabilistic generative models (Nijkamp et al., 2019; Du & Mordatch, 2019; Grathwohl et al., 2020b; Gao et al., 2020; Du et al., 2021; Gao et al., 2021; Grathwohl et al., 2020a; Yang et al., 2023; Zhu et al., 2024). By assigning a potential energy that correlates with the unnormalized data likelihood (Song & Kingma, 2021), EBMs offer a structured energy landscape for probability density estimation, providing several notable advantages. First, EBMs are interpretable, as the underlying energy function can be visualized in terms of energy surfaces. Second, they are highly expressive and do not impose strong architectural constraints (Bond-Taylor et al., 2022), enabling them to capture complex data distributions. Third, EBMs exhibit inherent robustness to Out-of-Distribution (OOD) inputs, given that regions with low likelihood are naturally penalized (Du & Mordatch, 2019; Grathwohl et al., 2020a). Building on their origins in Boltzmann machines (Hinton, 2002), EBMs also share conceptual ties with statistical physics, allowing practitioners to adapt physical insights and tools for model design and analysis (Feinauer & Lucibello, 2021). They have demonstrated promising performance in various applications beyond image modeling, including text generation (Deng et al., 2020), robot learning (Du et al., 2020), point cloud synthesis (Xie et al., 2021a), trajectory prediction (Pang et al., 2021; Wang et al., 2023), molecular design (Liu et al., 2021), and anomaly detection (Yoon et al., 2023).

Despite these advantages, training deep EBMs often relies on implicit Markov Chain Monte Carlo (MCMC) sampling for contrastive divergence. In high-dimensional settings, MCMC suffers from poor mode mixing and slow mixing (Du & Mordatch, 2019; Nijkamp et al., 2019; Gao et al., 2020; Grathwohl et al., 2020a; Nijkamp et al., 2022; Bond-Taylor et al., 2022), yielding biased estimates that may optimize unintended objectives (Grathwohl et al., 2020b). Truncated chains, in particular, can lead models to learn an implicit sampler rather than a true density, which prevents valid steady-state convergence and inflates computational overhead. As a result, the generated samples can deviate significantly from the target distribution (Grathwohl et al., 2020b). To mitigate these issues, some works propose auxiliary or cooperative strategies that learn complementary models to either avoid MCMC via variational inference (Xiao et al., 2021a) or combine short-run MCMC refinements with learned generator distributions (Xie et al., 2020; Grathwohl et al., 2021; Hill et al., 2022). Nevertheless, these approaches could complicate model architectures and training procedures.

In parallel, flow-based models have advanced generative modeling by leveraging continuous normalizing flows and optimal transport techniques to surpass diffusion models in sample quality and efficiency (Kim et al., 2021; Song et al., 2021). Notable examples include Flow Matching (Lipman et al., 2023), which models diffeomorphic mappings between noise and data; Rectified Flow (Liu et al., 2023b), which optimizes sampling paths; Stochastic Interpolants (Albergo & Vanden-Eijnden, 2023; Rezende & Mohamed, 2015; Chen et al., 2018), which incorporate stochastic processes into flows for complex data geometries; Schrödinger Bridge Matching (Shi et al., 2023), which integrates entropy-regularized optimal transport with diffusion; and Poisson Flow Generative Model (PFGM) (Xu et al., 2022), which introduces an augmented space governed by the Poisson equation. However, these methods do not directly parameterize probability density and lack the theoretical advantages of EBMs, such as generating conservative vector fields aligned with log-likelihood gradients (Salimans & Ho, 2021).

Recent approaches, such as Action Matching (Neklyudov et al., 2023), explicitly model the energy (action) to generate data-recovery vector fields, thus providing a structured approach to learning conservative dynamics. Meanwhile, Diffusion Recovery Likelihood (DRL) (Gao et al., 2021) and Denoising Diffusion Adversarial EBMs (DDAEBM) (Geng et al., 2024) refine conditional EBMs by improving sampling efficiency and training stability through diffusion-based probability paths. However, a direct connection between energy-parameterized flow models and explicit (marginal) EBMs remains unexplored, limiting the application of flow-based techniques for learning EBMs. Furthermore, existing generative models have yet to adopt variational formulations, such as the Deep Ritz method, to align the evolution of density paths.

To address the computational challenges of existing energy-based methods, we propose Variational Potential Flow Bayes (VPFB), a novel generative framework grounded in variational principles that eliminates the need for auxiliary models and implicit MCMC sampling. VPFB employs the Deep Ritz method to learn an energy-parameterized potential flow, ensuring alignment between the flow-driven density homotopy and the data-recovery likelihood homotopy. To address the intractability of homotopy matching, we formulate

a variational loss function that minimizes the Kullback-Leibler (KL) divergence between these density homotopies. Additionally, we validate the learned potential energy as an effective parameterization of the stationary Boltzmann energy. Through empirical validations, we benchmark VPFB against state-of-the-art generative models, showcasing its competitive performance in Fréchet Inception Distance (FID) for image generation and excellent OOD detection with high Area Under the Receiver Operating Characteristic Curve (AUROC) scores across multiple datasets.

2 Background and Related Works

In this section, we provide an overview of EBMs, particle flow, and the Deep Ritz method, collectively forming the cornerstone of the proposed VPFB framework.

2.1 Energy-based Models (EBMs)

Denote $\bar{x} \in \Omega \subseteq \mathbb{R}^n$ as the training data, EBMs approximate the data likelihood $p_{\text{data}}(\bar{x})$ via defining a Boltzmann distribution

$$p_B(x) = \frac{e^{\Phi_B(x)}}{Z} \quad (1)$$

where Φ_B is the Boltzmann energy parameterized via neural networks and $Z = \int_{\Omega} e^{\Phi_B(x)} dx$ is the normalizing constant. Given that this partition function is analytically intractable for high-dimensional data, EBMs perform the Maximum Likelihood Estimation (MLE) by minimizing the negative log-likelihood loss $\mathcal{L}_{\text{MLE}}(\theta) = -\mathbb{E}_{p_{\text{data}}(\bar{x})}[\log p_B(\bar{x})] = \mathbb{E}_{p_{\text{data}}(\bar{x})}[\Phi_B(\bar{x})] - \mathbb{E}_{p_{\text{data}}(\bar{x})}[\log Z]$. The gradient of this MLE loss with respect to model parameters θ is approximated via the contrastive divergence (Hinton, 2002) loss $\nabla_{\theta} \mathcal{L}_{\text{MLE}} = \mathbb{E}_{p_{\text{data}}(\bar{x})}[\nabla_{\theta} \Phi_B(\bar{x})] - \mathbb{E}_{p_B(x)}[\nabla_{\theta} \Phi_B(x)]$. Nonetheless, EBMs are computationally intensive due to the implicit MCMC generating procedure required for generating negative samples $x \sim p_B(x)$ implicitly during training.

2.2 Particle Flow

Particle flow, introduced by Daum & Huang (2007), is a class of nonlinear Bayesian filtering (sequential inference) methods designed to approximate the posterior distribution $p(x_t | \bar{x}_{\leq t})$ of the sampling process given observations. While closely related to normalizing flows (Rezende & Mohamed, 2015) and neural ordinary differential equations (ODEs) (Chen et al., 2018), these frameworks do not explicitly accommodate a Bayes update. Instead, particle flow achieves Bayes update $p(x_t | \bar{x}_{\leq t}) \propto p(x_t | \bar{x}_{< t}) p(\bar{x}_t | x_t, \bar{x}_{< t})$ by transporting the prior samples $x_t \sim p(x_t | \bar{x}_{< t})$ through an ODE $\frac{dx}{dt} = v(x, t)$ parameterized by a velocity field $v(x, t)$, over pseudo-time $t \in [0, 1]$. The velocity field is designed such that the sample density follows a log-homotopy that induces the Bayes update. Despite its effectiveness in time-series inference (Pal et al., 2021b; Chen et al., 2019b; Yang et al., 2014) and its robustness against the curse of dimensionality (Surace et al., 2019), particle flow, particularly potential flow where the velocity field $v(x, t) = \Phi(x, t)$ is the gradient of potential energy, remains largely unexplored in energy-based generative modeling.

2.3 Deep Ritz Method

The Deep Ritz method is a deep learning-based variational numerical approach, originally proposed by E & Yu (2018), for solving scalar elliptic partial differential equations (PDEs) in high dimensions. Consider the following Poisson’s equation, fundamental to many physical models:

$$\begin{aligned} \Delta_x u(x) &= \Gamma(x), \quad x \in \Omega \\ u(x) &= 0, \quad x \in \partial\Omega \end{aligned} \quad (2)$$

where Δ_x is the Laplace operator, and $\partial\Omega$ denotes the boundary of Ω . For a Sobolev function $u \in \mathcal{H}_0^1(\Omega)$ (definition in Proposition 2) and square-integrable $\Gamma \in L^2(\Omega)$, the variational principle ensures that a weak

solution of the Euler-Lagrange boundary value equation (2) is equivalent to the variational problem of minimizing the Dirichlet energy (Müller & Zeinhofer, 2019), as follows:

$$u = \arg \min_v \int_{\Omega} \left(\frac{1}{2} \|\nabla_x v(x)\|^2 - \Gamma(x) v(x) \right) dx + \eta \int_{\partial\Omega} v(x)^2 dx \quad (3)$$

where ∇_x denotes the Del operator (gradient). In particular, the Deep Ritz method parameterizes the trial function v using neural networks and performs the optimization (3) via stochastic gradient descent. To enforce the Dirichlet boundary condition, the second component of the Dirichlet energy (3), weighted by a positive constant η , must be evaluated on the boundary $\partial\Omega$. This necessitates acquiring additional boundary samples $x \in \partial\Omega$ during neural network training, thereby introducing extra computational overhead. The Deep Ritz method is predominantly applied for finite element analysis (Liu et al., 2023a) due to its versatility and effectiveness in handling high-dimensional PDE systems. In (Olmez et al., 2020), the Deep Ritz method is employed to solve the density-weighted Poisson equation arising from the feedback particle filter (Yang et al., 2013). However, its application in generative modeling remains unexplored.

3 Variational Potential Flow Bayes (VPFB)

In this section, we introduce VPFB, a novel generative modeling framework inspired by particle flow and the Deep Ritz method. VPFB encompasses four key elements: constructing a Bayesian marginal homotopy between the Gaussian prior and data likelihood (Section 3.1), designing a potential flow that aligns the flow-driven homotopy with the marginal homotopy (Section 3.2), formulating a variational loss function using the Deep Ritz method (Section 3.4), and establishing connections between homotopy matching, diffusion, and EBMs (Section 3.3).

3.1 Interpolating Between Prior and Data Likelihood: Log-Homotopy Bayesian Transport

Let $\bar{x} \in \Omega$ denote the training data, with likelihood $p_{\text{data}}(\bar{x})$, and let $x \in \Omega$ represent the generative samples. First, we define a Gaussian prior $q(x) = \mathcal{N}(0, \omega^2 I)$ and a Gaussian conditional data likelihood $p(\bar{x} | x) = \mathcal{N}(\bar{x}; x, \nu^2 I)$, both with isotropic covariances. This data likelihood satisfies the state space model $x = \bar{x} + \nu \epsilon$, where ϵ is the standard Gaussian noise. The standard deviation ν is usually set to be small so that x closely resembles \bar{x} . The aim of flow-based generative modeling is to learn a density homotopy (path) interpolating between the prior and the data likelihood for generative modeling. On that account, consider the following conditional (data-conditioned) probability density log-homotopy $\rho : \Omega^2 \times [0, 1] \rightarrow \mathbb{R}$:

$$\rho(x | \bar{x}, t) = \frac{e^{h(x|\bar{x},t)}}{\int_{\Omega} e^{h(x|\bar{x},t)} dx} \quad (4)$$

where $h : \Omega^2 \times [0, 1] \rightarrow \mathbb{R}$ is a log-linear function:

$$h(x | \bar{x}, t) = \alpha(t) \log q(x) + \beta(t) \log p(\bar{x} | x) \quad (5)$$

where $\alpha : [0, 1] \rightarrow [0, 1]$ and $\beta : [0, 1] \rightarrow [0, 1]$ are both monotonically increasing functions parameterized by time t . The following proposition shows that this log-homotopy transformation results in a Gaussian perturbation kernel.

Proposition 1. *Consider a Gaussian prior $q(x) = \mathcal{N}(x; 0, \omega^2 I)$ and a conditional data likelihood $p(\bar{x} | x) = \mathcal{N}(\bar{x}; x, \nu^2 I)$. The log-homotopy transport (4) corresponds to a Gaussian perturbation kernel $\rho(x | \bar{x}, t) = \mathcal{N}(x; \mu(t) \bar{x}, \sigma(t)^2 I)$, characterized by the time-varying mean and standard deviation:*

$$\mu(t) = \text{sigmoid} \left(\log \left(\frac{\beta(t)}{\alpha(t)} \frac{\omega^2}{\nu^2} \right) \right), \quad \sigma(t) = \sqrt{\frac{\nu^2}{\beta(t)}} \mu(t) \quad (6)$$

where $\text{sigmoid}(z) = \frac{1}{1+e^{-z}}$ denotes the logistic (sigmoid) function.

Proof. Refer to Appendix C.1. □

Hence, the density homotopy equation 4 represents a tempered Bayesian transport mapping from the Gaussian prior $q(x)$ to the posterior kernel

$$\rho(x \mid \bar{x}, 1) = \frac{e^{h(x|\bar{x},1)}}{\int_{\Omega} e^{h(x|\bar{x},1)} dx} = \frac{p(\bar{x} \mid x) q(x)}{\int_{\Omega} p(\bar{x} \mid x) q(x) dx} = p(x \mid \bar{x}) \quad (7)$$

which is the maximum a posteriori estimation centered on discrete data samples. To approximate the intractable data likelihood, we can then consider the following marginal probability density homotopy:

$$\bar{\rho}(x, t) = \int_{\Omega} p_{\text{data}}(\bar{x}) \rho(x \mid \bar{x}, t) d\bar{x}, \quad (8)$$

where it remains that $p(x, 0) = q(x)$, and we have $\bar{\rho}(x, 1) = \int_{\Omega} p_{\text{data}}(\bar{x}) p(x \mid \bar{x}) d\bar{x} = p(x)$. Therefore, this marginal homotopy defines a data-recovery path interpolation between the Gaussian prior $q(x)$ and the approximate data likelihood $p(x)$. In particular, $p(x)$ represents a Bayesian approximation of the true data likelihood, by convolving the discrete data likelihood $p_{\text{data}}(\bar{x})$ with the posterior distribution $p(x \mid \bar{x})$. Nevertheless, the marginalization in (8) is intractable, thereby precluding a closed-form solution for the marginal homotopy. To overcome this challenge, we propose a potential flow-driven density homotopy, whose time evolution is aligned with this data-recovery marginal homotopy.

3.2 Modeling Potential Flow in a Data-Recovery Homotopy Landscape

Our goal is to model a potential flow whose density evolution aligns with the marginal homotopy, thereby directing samples toward the data likelihood. We begin by deriving the time evolution of the marginal homotopy in the following proposition.

Proposition 2. *Consider the conditional homotopy $\rho(x \mid \bar{x}, t)$ in (4) with Gaussian conditional data likelihood $p(\bar{x} \mid x) = \mathcal{N}(\bar{x}; x, \nu^2 I)$. Then, the time evolution (derivative) of the marginal homotopy $\bar{\rho}(x, t)$ is given by the following partial differential equation (PDE):*

$$\frac{\partial \bar{\rho}(x, t)}{\partial t} = -\frac{1}{2} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x \mid \bar{x}, t) \left(\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t) \right) \right] \quad (9)$$

where γ denotes the innovation term

$$\gamma(x, \bar{x}, t) = \frac{\dot{\alpha}(t)}{\omega^2} \|x\|^2 + \frac{\dot{\beta}(t)}{\nu^2} \|x - \bar{x}\|^2 \quad (10)$$

Here, $\dot{\alpha}(t)$ and $\dot{\beta}(t)$ denote the time-derivatives, and $\bar{\gamma}(x, \bar{x}, t) = \mathbb{E}_{\rho(x|\bar{x},t)}[\gamma(x, \bar{x}, t)]$ denotes the expectation.

Proof. Refer to Appendix C.2. □

A potential flow involves subjecting the prior samples to an energy-generated velocity field, where their trajectories $(x(t))$ satisfy the following ODE:

$$\frac{dx(t)}{dt} = \nabla_x \Phi(x, t) \quad (11)$$

where $\Phi : \Omega \times [0, 1] \rightarrow \mathbb{R}$ is a scalar potential energy, and ∇_x denotes the Del operator (gradient) with respect to the data samples $x(t)$. The vector field $\nabla_x \Phi \in \Omega$ represents the divergence (irrotational) component in the Helmholtz decomposition. By incorporating this potential flow, the flow-driven density homotopy $\rho_{\Phi}(x, t)$ evolves via the continuity equation (Gardiner, 2009):

$$\frac{\partial \rho_{\Phi}(x, t)}{\partial t} = -\nabla_x \cdot \left(\rho_{\Phi}(x, t) \nabla_x \Phi(x, t) \right) \quad (12)$$

which corresponds to the transport equation for modeling fluid advection. Our aim is to model the potential energy such that the evolution of the prior density under the potential flow emulates the evolution of the

marginal homotopy. In other words, we seek to achieve homotopy matching, $\rho_\Phi \equiv \bar{\rho}$, by aligning their respective time evolutions as described in (9) and (12). This leads to the following PDE, which takes the form of a density-weighted Poisson equation:

$$\nabla_x \cdot \left(\rho_\Phi(x, t) \nabla_x \Phi(x, t) \right) = \frac{1}{2} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) \left(\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t) \right) \right] \quad (13)$$

However, this Poisson equation remains intractable due to the lack of a closed-form expression for ρ_Φ . To overcome this limitation, we substitute the intractable ρ_Φ with the target marginal homotopy $\bar{\rho}$, enabling direct sampling and a variational principle approach. In the following proposition, we demonstrate that the revised Poisson's equation minimizes the KL divergence between the flow-driven and conditional homotopies, yielding statistically optimal homotopy matching.

Proposition 3. *Consider a potential flow of the form (11) and given that $\Phi \in \mathcal{H}_0^1(\Omega, p)$, where \mathcal{H}_0^n denotes the (Sobolev) space of n -times differentiable functions that are compactly supported, and square-integrable with respect to marginal homotopy $\bar{\rho}(x, t)$. Solving for the potential energy $\Phi(x)$ that satisfies the following density-weighted Poisson's equation:*

$$\nabla_x \cdot \left(\bar{\rho}(x, t) \nabla_x \Phi(x, t) \right) = \frac{1}{2} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) \left(\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t) \right) \right] \quad (14)$$

is then equivalent to minimizing the KL divergence $\mathcal{D}_{\text{KL}}[\rho_\Phi(x, t) \| \bar{\rho}(x, t)]$ between the flow-driven homotopy and the conditional homotopy.

Proof. Refer to Appendix C.3. □

Therefore, solving this density-weighted Poisson's equation corresponds to performing a homotopy matching $\rho_\Phi \equiv \bar{\rho}$. In the following section, we demonstrate that this homotopy matching gives rise to a Boltzmann energy expressed in terms of the potential energy Φ when the marginal homotopy $\bar{\rho}$ reaches its stationary equilibrium, thereby establishing a connection between our proposed potential flow framework and EBMs.

3.3 Connections to Diffusion Process and Energy-Based Modeling

In this section, we clarify the relationship between diffusion models and flow matching within the homotopy matching framework. Building on this insight, we establish a link between our proposed potential flow framework and energy-based modeling.

First, we present results from diffusion models. It has been outlined in Song et al. (2021) that the conditional density homotopy, represented by the Gaussian perturbation kernel $\rho(x | \bar{x}, t) = \mathcal{N}(x; \mu(t) \bar{x}, \sigma(t)^2 I)$, characterizes a diffusion process governed by the following stochastic differential equation (SDE):

$$dx(t) = -f(t) x(t) dt + g(t) dW(t) \quad (15)$$

where $W(t) \in \mathbb{R}^n$ denote the standard Wiener process. Note that the time parameterization with respect to t here is the reverse of the conventional parameterization used in diffusion models, where the diffusion process transitions from $x(1) \sim p_{\text{data}}(\bar{x})$ to $x(0) \sim q(x) = \mathcal{N}(0, \omega^2 I)$ as defined in Section 3.1. In addition, the time-varying drift $f : [0, 1] \rightarrow \mathbb{R}$ and diffusion $g : [0, 1] \rightarrow \mathbb{R}$ coefficients are shown by Karras et al. (2022) to be given by

$$f(t) = -\frac{\dot{\mu}(t)}{\mu(t)}, \quad g(t) = -\sqrt{2\sigma(t) \left(\dot{\sigma}(t) + f(t) \sigma(t) \right)} \quad (16)$$

where $\dot{\mu}(t)$ and $\dot{\sigma}(t)$ denote the time-derivatives. It has also been shown in Song et al. (2021) that the following deterministic probability flow ODE:

$$\frac{dx(t)}{dt} = -f(t) x(t) + \frac{1}{2} g(t)^2 \nabla_x \log \bar{\rho}(x, t) \quad (17)$$

results in the same marginal probability homotopy $\bar{\rho}(x, t)$ as the forward-time diffusion SDE (16). Subsequently, we highlight the link between the diffusion process and the vector field modeled in flow matching.

Proposition 4. *The conditional vector field in flow matching (Lipman et al., 2023), given by*

$$\frac{dx(t)}{dt} = v(x | \bar{x}, t) = \dot{\mu}(t) \bar{x} + \dot{\sigma}(t) \epsilon \quad (18)$$

with standard Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$, satisfies the conditional probability flow ODE governing the diffusion process conditioned on boundary condition $x(1) \sim p_{\text{data}}(\bar{x})$. It follows that the marginal vector field, given by the law of iterated expectation (tower property) $\mathbb{E}[U|X=x] = \mathbb{E}[\mathbb{E}[U | X=x, Y] | X=x]$:

$$\frac{dx(t)}{dt} = v(x, t) = \mathbb{E}_{p_{\text{data}}(\bar{x}|x)}[v(x | \bar{x}, t) | x] = \int_{\Omega} v(x | \bar{x}, t) \frac{\rho(x | \bar{x}, t) p_{\text{data}}(\bar{x})}{\bar{\rho}(x, t)} d\bar{x} \quad (19)$$

also satisfies the marginal probability flow ODE (17).

Proof. Refer to Appendix C.4. □

Building on this result, we establish a connection between the proposed potential flow framework and EBMs. The following proposition demonstrates that homotopy matching, e.g., $\rho_{\Phi} \equiv \bar{\rho}$ leads to an energy-parameterized Boltzmann equilibrium.

Proposition 5. *Given that the flow-driven homotopy $\rho_{\Phi}(x, t)$ matches the data-recovery marginal homotopy $\bar{\rho}(x, t)$, they exhibit the same Fokker–Planck dynamics. As the time-varying marginal density $\bar{\rho}(x, t)$ converges to its stationary equilibrium $\bar{\rho}_{\infty}(x)$, i.e., when $\frac{\partial \bar{\rho}(x, t)}{\partial t} \rightarrow 0$, the Fokker–Planck dynamics reach the Boltzmann distribution (1), where the Boltzmann energy $\Phi_B(x)$ is defined as follows:*

$$\Phi_B(x) = \frac{4\Phi_{\infty}(x) + f_{\infty} \|x\|^2}{g_{\infty}^2} \quad (20)$$

where $\Phi_{\infty}(x)$, f_{∞} , and g_{∞} denote the steady-state potential energy, drift, and diffusion coefficients, respectively, associated with the stationary equilibrium.

Proof. Refer to Appendix C.5. □

On that note, we uncover the connection between the proposed VPFB framework and EBMs, demonstrating the validity of the potential energy as a parameterization of a Boltzmann energy. This holds provided that $\bar{\rho}$ converges to its stationary equilibrium and ρ_{Φ} learns to match these convergent dynamics. In the following section, we introduce a variational principle approach to solving the density-weighted Poisson equation (14), thereby addressing the intractable homotopy matching problem.

3.4 Variational Potential Energy Loss Formulation: Deep Ritz Method

Solving the density-weighted Poisson’s equation (14) is particularly challenging in high-dimensional settings. Numerical approximation struggles to scale with higher dimensionality, as selecting suitable basis functions, such as in the Galerkin approximation, becomes increasingly complex (Yang et al., 2016). Similarly, a diffusion map-based algorithm demands an exponentially growing number of particles to ensure error convergence (Taghvaei et al., 2020). To address these challenges, we propose a variational loss function using the Deep Ritz method. This approach casts Poisson’s equation as a variational problem compatible with stochastic gradient descent. Consequently, the proposed approach solves Eq. (14), effectively aligning the flow-driven homotopy with the marginal homotopy. Directly solving Poisson’s equation (14) is challenging. Therefore, we first consider the following weak formulation:

$$\int_{\Omega} \frac{1}{2} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] \Psi dx = \int_{\Omega} \nabla_x \cdot \left(\bar{\rho}(x, t) \nabla_x \Phi(x, t) \right) \Psi dx \quad (21)$$

This PDE must hold for all differentiable trial functions Ψ . In the following proposition, we introduce a variational loss function that is equivalent to solving this weak formulation of the density-weighted Poisson’s equation.

Proposition 6. *The variational problem of minimizing the following loss functional:*

$$\mathcal{L}(\Phi, t) = \text{Cov}_{\rho(x|\bar{x},t) p_{\text{data}}(\bar{x})} [\Phi(x, t), \gamma(x, \bar{x}, t)] + \mathbb{E}_{\bar{\rho}(x,t)} [\|\nabla_x \Phi(x, t)\|^2] \quad (22)$$

with respect to the potential energy Φ , is equivalent to solving the weak formulation (21) of the density-weighted Poisson’s equation (14). Here, $\|\cdot\|$ denotes the Euclidean norm, and Cov denotes the covariance. Furthermore, the variational problem (22) admits a unique solution $\Phi \in \mathcal{H}_0^1(\Omega; \rho)$ if the marginal homotopy p satisfy the Poincaré inequality:

$$\mathbb{E}_{\bar{\rho}(x,t)} [\|\nabla_x \Phi(x, t)\|^2] \geq \eta \mathbb{E}_{\bar{\rho}(x,t)} [\|\Phi(x, t)\|^2] \quad (23)$$

for some positive scalar constant $\eta > 0$ (spectral gap).

Proof. Refer to Appendix C.6. □

Remark 1. The integration by parts in (66) and (87) require that the marginal density $\bar{\rho}(x)$ vanishes on the boundary $\partial\Omega$ of some open, bounded domain $\Omega \subset \mathbb{R}^n$, so that the boundary integral $\int_{\partial\Omega} \bar{\rho}(x) (\nabla_x \Phi \cdot \hat{n}) dx = 0$ holds. In standard implementations, although the training data \bar{x} are typically normalized to lie within $[-1, 1]^n$, we may define the perturbed samples as $x \in \Omega \subset \mathbb{R}^n$, where Ω is chosen to contain the support of the data distribution. Accordingly, the open bounded domain Ω can be defined sufficiently large so that the conditional homotopy $\rho(x | \bar{x}, t)$ approaches zero at the boundary $\partial\Omega$. Since $\rho(x | \bar{x}, t)$ is a Gaussian perturbation kernel, it decays exponentially and is effectively negligible near the boundary, thereby satisfying the required condition. As a result, the marginal distribution $\bar{\rho}(x, t)$ also vanishes at $\partial\Omega$, ensuring the validity of the integration by parts required to formulate both Proposition 3 Proposition 6.

Overall, Propositions 3 and 6 recast the intractable problem of minimizing the KL divergence between the flow-driven homotopy and the marginal homotopy as an equivalent variational problem of solving the loss function (22). By optimizing the potential energy with respect to this loss and transporting the prior samples through the ODE (11), the prior particles evolve along a trajectory that aligns with the marginal homotopy. In particular, the covariance loss here plays an important role by ensuring that the normalized innovation (residual sum of squares) is inversely proportional to the potential energy. As a result, the energy-generated velocity field $\nabla_x \Phi$ consistently points in the direction of greatest potential ascent, thereby driving the flow of prior particles towards high likelihood regions. Given that homotopy matching is performed over the entire time horizon, we apply stochastic integration to the loss function over time, where $t \sim \mathcal{U}(0, t_{\text{end}})$ is drawn from the uniform distribution.

3.5 Training Implementation

In our implementation, we adopt the Optimal Transport Flow Matching (OT-FM) framework (Lipman et al., 2023) for training, where it corresponds to the SDE parameterization $f(t) = -\frac{1}{t}$, $g(t) = \sqrt{\frac{2(1-t)}{t}}$ as outline in (Kingma & Gao, 2023), or equivalently $\alpha(t) = \frac{\omega^2}{1-t}$, $\beta(t) = \frac{\nu^2 t}{(1-t)^2}$ for the log-homotopy transformation derived in (5). To establish a Boltzmann equilibrium, we further require $\frac{\partial \bar{\rho}(x,t)}{\partial t} \rightarrow 0$ so that the time-varying marginal density $\bar{\rho}(x, t)$ converges to the stationary Boltzmann distribution. However, the Gaussian perturbation kernel $\rho(x | \bar{x}, t) = \mathcal{N}(\mu(t) \bar{x}, \sigma(t)^2 I)$ employed by the flow-based probability paths is defined only over a finite time interval $t \in [0, t_{\text{max}}]$. Additionally, the marginal homotopy $\bar{\rho}(x, t)$ is not guaranteed to reach equilibrium within this prescribed time window.

To resolve these limitations of the flow-based probability paths, we explicitly enforce stationarity in our training implementation, by imposing a steady-state equilibrium $p_\infty(x) = \bar{\rho}(x, t \geq t_{\text{max}})$ beyond some cutoff time $t_{\text{max}} < t_{\text{end}}$ close to the terminal time. This steady-state equilibrium $p_\infty(x) \equiv p_B(x)$ thus corresponds to the stationary Boltzmann distribution, parameterized by the energy function derived in (20) with $f_\infty = f(t_{\text{max}})$ and $g_\infty = g(t_{\text{max}})$. Given that a steady-state equilibrium is enforced via $p_\infty(x) = \bar{\rho}(x, t \geq t_{\text{max}}) \approx p_{\text{data}}(\bar{x})$, the stationary Boltzmann distribution approximates the true data likelihood by design.

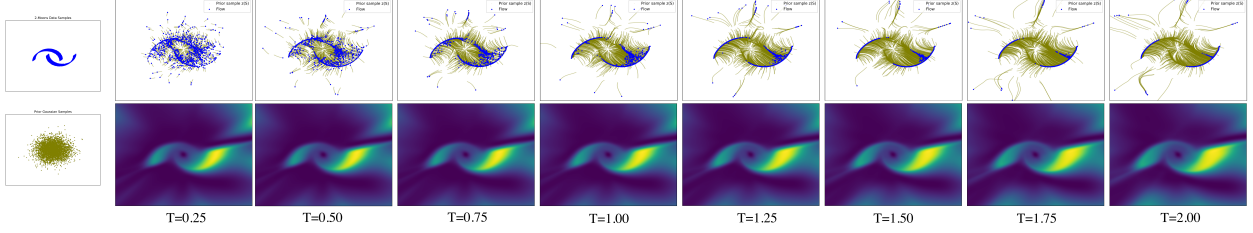


Figure 1: 2D potential flow. Top: Sample trajectories from the Gaussian prior noise distribution (black) to the target 2-Moons distribution (blue), driven by the potential energy $\Phi(x, t)$ and sampled using an ODE solver. Bottom: Time evolution of the learned potential energy landscape $\Phi(x, t)$.

Finally, our VPFB loss function is implemented as follows:

$$\mathcal{L}^{\text{VPFB}}(\Phi) = \int_0^{t_{\text{end}}} \mathcal{L}(\Phi, t) dt = \mathbb{E}_{\mathcal{U}(0, t_{\text{end}})} [\mathcal{L}(\Phi, t)] \quad (24)$$

where

$$\begin{aligned} \mathcal{L}(\Phi, t) = & \text{Cov}_{\rho(x|\bar{x}, t) p_{\text{data}}(\bar{x})} \left[\Phi(x, t), w(t) \gamma(x, \bar{x}, t) \right] - \frac{\nabla_x \Phi(x, t) \cdot v(x | \bar{x}, t)}{\|\nabla_x \Phi(x, t)\| \|v(x | \bar{x}, t)\|} \\ & + \mathbb{E}_{\rho(x|\bar{x}, t) p_{\text{data}}(\bar{x})} \left[\|\nabla_{(x, t)} \Phi(x, t)\|^2 + \eta \|\Phi(x, t)\|^2 \right] \end{aligned} \quad (25)$$

Here, we incorporate an additional cosine distance between the potential gradient $\nabla_x \Phi$ and the conditional vector field in (18) to the loss function. While this cosine distance does not influence the learning of the potential energy’s magnitude (magnitude learning is entirely supervised by the covariance loss), it enforces directional alignment between the gradient and the vector field. To further enforce convergence toward a steady-state potential $\Phi_{\infty}(x)$, we additionally encourage quasi-static dynamics by minimizing the Euclidean norm of the time derivative $\left| \frac{\partial \Phi}{\partial t} \right|^2$ alongside the gradient norm during training. Also, a weighting $w(t) = (1 - t)^{\kappa}$ with decay exponent $\kappa > 1$ is applied to the innovation term to balance the covariance loss across time to stabilize training.

Considering that the marginal homotopy may not satisfy the Poincaré inequality (23), we include the right-hand side of this inequality in the loss function to enforce the uniqueness of the minimizer. To empirically validate the existence of a positive Poincaré constant η , Figure 10 plots the ratio between the mean gradient norm $\mathbb{E}[\|\nabla_x \Phi\|^2]$ and the mean energy norm $\mathbb{E}[\|\Phi\|^2]$ over training iterations on CIFAR-10, without applying the additional Poincaré regularization loss. It shows that the ratio is bounded below by $\eta = 6.81 \times 10^{-5}$, thereby confirming the existence of a positive Poincaré constant during training.

Nonetheless, our experiments indicate that the existence and magnitude of such an unenforced Poincaré constant vary across different neural architectures. For completeness, we incorporate the Poincaré regularization with a small η for both the WideResNet and U-Net models, which we fine-tune during training for optimal results. The cutoff time t_{max} , terminal time t_{end} , decay exponent κ , and spectral gap constant η are hyperparameters to be determined during training. Algorithm 1 summarizes the training procedure of our proposed VPFB framework.

4 Experiments

In this section, we validate the energy-based generative modeling capabilities of VPFB across several key tasks. Section 4.1 explores 2D density estimation. Section 4.2 presents the unconditional generation and spherical interpolation results on CIFAR-10 and CelebA. Section 4.4 evaluates mode coverage and model generalization through energy histograms of train and test data and the nearest neighbors of generated samples. Section 4.5 examines unsupervised OOD detection performance on various datasets. Section 4.6 verifies the convergence of long-run ODE samples to a Boltzmann equilibrium. Additional results on ablation

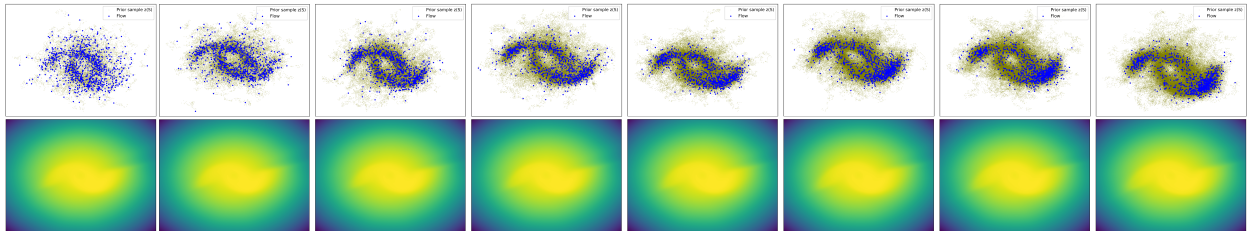


Figure 2: 2D Boltzmann density estimation. Top: Sample trajectories from the Gaussian prior noise distribution (black) to the target 2-Moons distribution (blue), driven by the Boltzmann energy and sampled via SGLD. Middle: Visualization of the log-density estimation (up to an additive constant) $\log p_B(x) = \Phi_B(x)$ parameterized by Boltzmann energy.

study and computational efficiency are provided in Appendix A. Additional discussions of the results are also provided in Appendix B. Finally, implementation details, including architecture, training, numerical solvers, datasets, and FID evaluation, are provided in Appendix D.

4.1 Density Estimation on 2D Data

To verify the convergence properties of the potential energy and to assess the validity of the Boltzmann energy (20), we conduct density estimation on 2D synthetic datasets. Specifically, we learn a potential flow that transforms an unimodal Gaussian prior distribution into a 2-Moons target distribution. Figure 1 shows the sample trajectories driven by the potential flow $dx(t) = \nabla_x \Phi(x, t) dt$, obtained via the deterministic Euler solver. Figure 2 presents the sample trajectories and density estimation of the Boltzmann distribution $p_B \propto e^{\Phi_B(x)}$, obtained via the Stochastic Gradient Langevin Dynamics (SGLD). Notably, both the potential energy $\Phi(x)$ and the Boltzmann energy $\Phi_B(x)$ exhibit stable convergence toward their steady-state equilibrium. Furthermore, the results indicate that the estimated Boltzmann density closely aligns with the ground-truth 2-Moons distribution. These results highlight the effectiveness of our variational principle approach in learning the Boltzmann stationary distribution through homotopy matching against the stationary-enforced marginal $p_\infty(x)$.

Nonetheless, a standard formulation of the forward-time SDE (15), or equivalently, the marginal probability flow ODE (17), is valid only in the case of a unimodal Gaussian prior, e.g., $q(x) = \rho(x | \bar{x}, t = 0) = \mathcal{N}(0, \omega^2 I)$, as discussed in Kingma & Gao (2023). This assumption underpins the consistency of the Fokker–Planck dynamics with the continuous-time diffusion framework, ensuring the validity of the stationary Boltzmann energy in the limit. We acknowledge this limitation of our current framework. As a direction for future work, we propose extending the forward-time SDE or ODE formulation of continuous-time diffusion to be admissible for more general prior distributions, such as mixtures of Gaussians or learned priors, to accommodate multi-modal data while maintaining consistency with our proposed energy-based framework.

4.2 Unconditional Image Generation

For image generation, we consider three VPFB model variants: an autonomous (independent of time) energy model $\Phi(x)$ parameterized by Zagoruyko & Komodakis (2016), and a time-varying energy model $\Phi(x, t)$ parameterized by U-Net (Ronneberger et al., 2015). Figure 3 shows the uncured and unconditional image samples generated using the time-varying energy model on CIFAR-10 32×32 and CelebA 64×64 . The generated samples are of decent quality and resemble the original datasets, despite not having the highest fidelity as achieved by state-of-the-art models. Table 1 summarizes the quantitative evaluations of our framework in terms of FID (Heusel et al., 2017) scores on the CIFAR-10. In particular, the VPFB models achieved FID scores competitive to existing generative models. Figures 6 and 7 show additional uncured samples of unconditional image generation on CIFAR-10 and CelebA, respectively.



Figure 3: Uncurated and unconditional samples generated for CIFAR-10 (left) and CelebA (right).

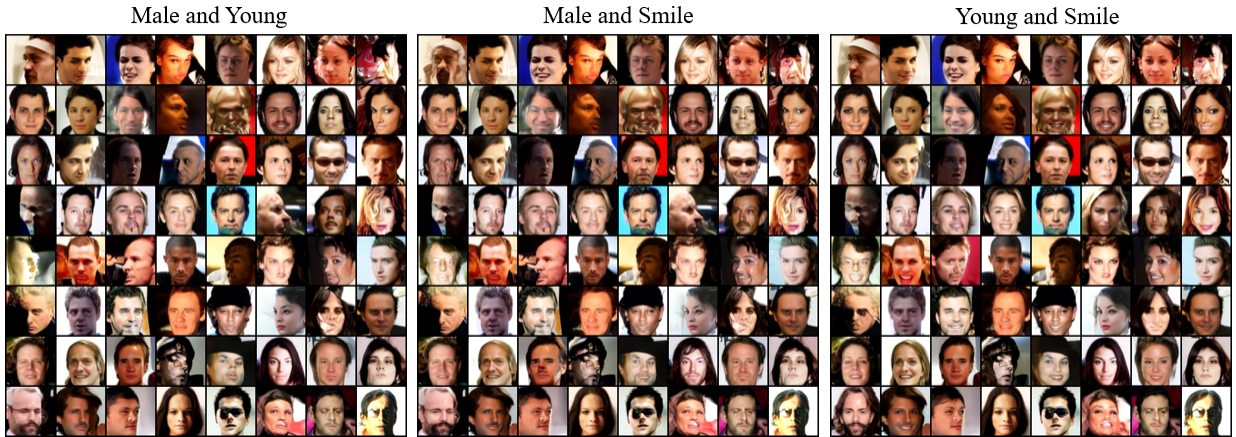


Figure 4: Compositional and conditional CelebA samples generated based on three attribute pairs.

4.3 Image Interpolation and Compositional Generation

To achieve smooth and semantically coherent image interpolation, we perform spherical interpolation between two Gaussian noises and subsequently apply ODE sampling to the interpolated noises. Figures 8 and 9 show additional interpolation results on CIFAR-10 and CelebA, respectively. For compositional sample generation, we first train a class-conditioned energy model $\Phi(x, c)$, and then sample by averaging the conditional energies across selected classes. Figure 4 presents compositional generation results conditioned on composite CelebA attributes, specifically *(Male, Young)*, *(Male, Smile)*, and *(Young, Smile)*. However, certain compositional samples show limited variation across attribute pairs, suggesting that incorporating composition weights could improve attribute-specific conditioning.



Figure 5: Generated CIFAR-10 samples and their five nearest neighbors in train set based on pixel distance.

Table 1: FID scores on unconditional CIFAR-10 image generation.

Energy-based Models	FID ↓	Other Likelihood-based Models	FID ↓
EBM-IG (Du & Mordatch, 2019)	38.2	ResidualFlow (Chen et al., 2019a)	47.4
EBM-FCE (Gao et al., 2020)	37.3	Glow (Kingma & Dhariwal, 2018)	46.0
CoopVAEBM (Xie et al., 2021b)	36.2	DC-VAE (Parmar et al., 2021)	17.9
CoopNets (Xie et al., 2020)	33.6	GAN-based Models	
Divergence Triangle (Han et al., 2019)	30.1	SN-GAN (Miyato et al., 2018)	21.7
VERA (Grathwohl et al., 2021)	27.5	SNGAN-DDLS Che et al. (2020)	15.4
EBM-CD (Du et al., 2021)	25.1	BigGAN (Brock et al., 2019)	14.8
GBM (Arbel et al., 2021)	19.3	Score-based and Diffusion Models	
HAT-EBM (Hill et al., 2022)	19.3	NCSN-v2 (Song & Ermon, 2020)	10.9
CF-EBM (Zhao et al., 2020)	16.7	DDPM Distil (Luhman et al., 2021)	9.36
CoopFlow (Xie et al., 2022)	15.8	DDPM (Ho et al., 2020)	3.17
VAEBM (Xiao et al., 2021a)	12.2	NCSN++ (Song et al., 2021)	2.20
DRL (Gao et al., 2021)	9.58	Flow-based Models	
CLEL (Lee et al., 2022)	8.61	Action Matching (Neklyudov et al., 2023)	10.0
DDAEBM (Geng et al., 2024)	4.82	Flow Matching (Lipman et al., 2023)	6.35
CDRL (Zhu et al., 2024)	3.68	Rectified Flow (Liu et al., 2023b)	4.85
VPFB (Autonomous)	14.5	DSBM (Shi et al., 2023)	4.51
VPFB (Time-varying)	6.72	PFGM (Xu et al., 2022)	2.35

4.4 Model Generalization and Mode Evaluation

To evaluate the generalization capability of the VPFB model, Figure 5 presents the nearest neighbors of the generated samples in the CIFAR-10 training set. The results show that nearest neighbors differ significantly from the generated samples, suggesting that our model does not overfit the training data and generalizes well to the underlying data distribution. To validate the mode coverage and over-fitting ability, Figure 11 presents a histogram of the energy outputs for both the CIFAR-10 training and test datasets. The histogram shows that the learned energy model assigns similar energy values to images from both sets. This indicates that the VPFB model generalizes well to unseen test data while maintaining broad mode coverage of the training distribution.

Table 2: AUROC scores \uparrow for OOD detection on several datasets.

Models	CIFAR-10 Interpolation	CIFAR-100	CelebA	SVHN
PixelCNN (Salimans et al., 2017)	0.71	0.63	-	0.32
GLOW (Kingma & Dhariwal, 2018)	0.51	0.55	0.57	0.24
NVAE (Vahdat & Kautz, 2020)	0.64	0.56	0.68	0.42
EBM-IG (Du & Mordatch, 2019)	0.70	0.50	0.70	0.63
VAEBM (Xiao et al., 2021a)	0.70	0.62	0.77	0.83
CLEL (Lee et al., 2022)	0.72	0.72	0.77	0.98
DRL (Gao et al., 2021)	-	0.44	0.64	0.88
CDRL (Zhu et al., 2024)	0.75	0.78	0.84	0.82
VPFB (Ours)	0.78	0.67	0.84	0.61

4.5 Out-of-Distribution Detection

Given that the potential flow corresponds to a stationary Boltzmann distribution, the Boltzmann energy Φ_B from (20) can be used to distinguish between in-distribution and OOD samples based on their assigned energy values. Specifically, the potential energy model trained on the CIFAR-10 training set assigns energy values to both in-distribution samples (CIFAR-10 test set) and OOD samples from various other image datasets. We evaluate OOD detection performance using the AUROC metric, where a higher score reflects better model’s efficacy in accurately assigning lower energy values to OOD samples. Table 2 compares the AUROC scores of VPFB with those of various likelihood-based and EBMs. The results show that our model performs exceptionally well on interpolated CIFAR-10 and CelebA 32×32 while achieving moderate performance on CIFAR-100 and SVHN.

4.6 Long-Run Steady-State Equilibrium

Figure 12 illustrates long-run ODE sampling over an extended time horizon $t \in [0, 20]$ using the autonomous energy model parameterized by WideResNet. Additionally, Figure 13 illustrates long-run ODE sampling using the time-varying energy model parameterized by U-Net. The results indicate a similar deterioration in image quality over extended time periods, albeit to a greater extent compared to the autonomous model. Figure 14 plots the mean gradient norm $\mathbb{E}[\|\nabla_x \Phi\|^2]$ and the mean energy norm $\mathbb{E}[\|\Phi\|^2]$, neither of which exhibit convergence. These results are consistent with those observed in EBMs trained using non-convergent short-run MCMC (Agoritsas et al., 2023; Nijkamp et al., 2020). This issue arises from the inherent difficulty neural network models face in learning complex energy landscapes in high-dimensional spaces. Regions that remain unseen during training can correspond to poorly modeled areas of the energy landscape, often resulting in the emergence of sharp local minima. Consequently, ODE-based sampling may become trapped in these local minima, leading to mode collapse and poor mixing, which manifest as visual artifacts such as excessive saturation and loss of background details.

To resolve these issues, we replace the deterministic ODE solver with the conventional SGLD sampler for image generation, enabling sampling from the Boltzmann energy via $x_{t+1} = x_t + \Delta_t \nabla_x \Phi_B(x_t) + \sqrt{2\Delta_t} \epsilon$ where $\epsilon \sim \mathcal{N}(0, \lambda^2 I)$ denotes isotropic Gaussian noise with temperature scale λ (standard deviation), and Δ_t is the step size. The injected stochasticity from the diffusive noise in SGLD facilitates escape from local minima and enhances mixing efficiency during sampling. As shown in Figures 15 and 17, SGLD mitigates mode collapse and the long-run image samples converge well to the stationary equilibrium. Furthermore, Figures 16 and 18 demonstrate that the gradient norm converges to zero, while the energy norm asymptotically stabilizes, indicating steady-state thermalization. These SDE-based sampling results confirm that equilibrium convergence is achievable with a stochastic sampler. Nonetheless, the temperature scale λ must be carefully tuned to balance convergence speed and sample quality. Moreover, our experiments show that ODE-based sampling consistently yields better FID scores, potentially due to the deterministic nature of the proposed potential flow and the straightness of the linearly interpolated OT-FM trajectories, which contribute to sharper and more consistent sample generation.

5 Conclusion

We propose VPFB, a novel energy-based potential flow framework designed to reduce the computational cost and instability typically associated with EBM training. Empirical results demonstrate that VPFB outperforms several existing EBMs in unconditional image generation and achieves competitive performance in OOD detection, highlighting its versatility across diverse generative modeling tasks. Despite these promising results, future work will aim to refine the training strategy to improve scalability to higher-resolution images and other data modalities, while addressing the limitations outlined in this work. Additionally, exploring generative models that inherently incorporate Neumann boundary conditions into the design of their blurring perturbation kernels (Rissanen et al., 2023; Hooeboom & Salimans, 2023; Daras et al., 2023) presents a promising direction for improving energy landscape modeling and enhancing sample diversity without incurring the computational burden of long-run MCMC sampling.

Broader Impact Statement

Generative models represent a rapidly growing field of study with overarching implications in science and society. Our work proposes a new generative model designed for efficient data generation and OOD detection, with potential applications in fields such as medical imaging, entertainment, and content creation. However, as with any powerful technology, generative models come with substantial risks, including the potential misuse in creating deepfakes or misleading content that could undermine social security and trust. Given this dual-use nature, it is essential to implement safeguards, such as classifier-based guidance, to prevent the generation of biased or harmful content. Moreover, generative models are vulnerable to backdoor adversarial attacks and can inadvertently amplify biases present in the training data, reinforcing social inequalities. Although our work uses standard datasets, it is important to address how such biases are handled. We are actively exploring methods to identify and mitigate biases during both the training and generation phases. This includes employing fairness-aware training algorithms and evaluating the model’s output for biased patterns. One potential solution is incorporating privacy-preserving encryption techniques to safeguard sensitive data and ensure that generative models do not expose private information. Furthermore, while this work demonstrates the potential benefits of generative models, the ethical concerns surrounding their deployment must be considered. Addressing these issues will require ongoing collaboration to develop frameworks for responsible use, including transparency, model interpretability, and robust safeguards against malicious applications. By proactively engaging with these ethical concerns, the broader community can contribute to the responsible advancement of generative modeling technologies.

References

- Elisabeth Agoritsas, Giovanni Catania, Aurélien Decelle, and Beatriz Seoane. Explaining the effects of non-convergent MCMC in the training of energy-based models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 322–336. PMLR, 2023.
- Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023.
- Brian D.O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. ISSN 0304-4149.
- Michael Arbel, Liang Zhou, and Arthur Gretton. Generalized energy based models. In *International Conference on Learning Representations*, 2021.
- Sam Bond-Taylor, Adam Leach, Yang Long, and Chris G. Willcocks. Deep generative modelling: A comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7327–7347, 2022.
- Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2019.

-
- Tong Che, Ruixiang ZHANG, Jascha Sohl-Dickstein, Hugo Larochelle, Liam Paull, Yuan Cao, and Yoshua Bengio. Your gan is secretly an energy-based model and you should use discriminator driven latent sampling. In *Advances in Neural Information Processing Systems*, volume 33, pp. 12275–12287. Curran Associates, Inc., 2020.
- Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, 2018.
- Ricky TQ Chen, Jens Behrmann, David K Duvenaud, and Jörn-Henrik Jacobsen. Residual flows for invertible generative modeling. *Advances in Neural Information Processing Systems*, 32, 2019a.
- Xinshi Chen, Hanjun Dai, and Le Song. Particle flow Bayes’ rule. In *International Conference on Machine Learning (ICML)*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1022–1031, 2019b.
- Giannis Daras, Mauricio Delbracio, Hossein Talebi, Alex Dimakis, and Peyman Milanfar. Soft diffusion: Score matching with general corruptions. *Transactions on Machine Learning Research*, 2023.
- Fred Daum and Jim Huang. Nonlinear filters with log-homotopy. In *Signal and Data Processing of Small Targets*, volume 6699, pp. 669918, 2007. doi: 10.1117/12.725684.
- Yuntian Deng, Anton Bakhtin, Myle Ott, Arthur Szlam, and Marc’Aurelio Ranzato. Residual energy-based models for text generation. In *International Conference on Learning Representations*, 2020.
- Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, 2021.
- J.R. Dormand and P.J. Prince. A family of embedded runge-kutta formulae. *Journal of Computational and Applied Mathematics*, 6(1):19–26, 1980. ISSN 0377-0427. doi: [https://doi.org/10.1016/0771-050X\(80\)90013-3](https://doi.org/10.1016/0771-050X(80)90013-3).
- Yilun Du and Igor Mordatch. Implicit generation and modeling with energy based models. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- Yilun Du, Toru Lin, and Igor Mordatch. Model-based planning with energy-based models. In *Proceedings of the Conference on Robot Learning*, volume 100, pp. 374–383, 2020.
- Yilun Du, Shuang Li, Joshua Tenenbaum, and Igor Mordatch. Improved contrastive divergence training of energy-based models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 2837–2848, 2021.
- Weinan E and Bing Yu. The deep ritz method: A deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, 6(1):1–12, 2018.
- Christoph Feinauer and Carlo Lucibello. Reconstruction of pairwise interactions using energy-based models*. *Journal of Statistical Mechanics: Theory and Experiment*, 2021(12):124007, 2021.
- Ruiqi Gao, Erik Nijkamp, Diederik P. Kingma, Zhen Xu, Andrew M. Dai, and Ying Nian Wu. Flow contrastive estimation of energy-based models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- Ruiqi Gao, Yang Song, Ben Poole, Ying Nian Wu, and Diederik P Kingma. Learning energy-based models by diffusion recovery likelihood. In *International Conference on Learning Representations*, 2021.
- C. Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer Series in Synergetics. Springer Berlin Heidelberg, 2009. ISBN 9783540707127.
- Cong Geng, Tian Han, Peng-Tao Jiang, Hao Zhang, Jinwei Chen, Søren Hauberg, and Bo Li. Improving adversarial energy-based model via diffusion process. In *Forty-first International Conference on Machine Learning*, 2024.

-
- Will Grathwohl, Kuan-Chieh Wang, Joern-Henrik Jacobsen, David Duvenaud, Mohammad Norouzi, and Kevin Swersky. Your classifier is secretly an energy based model and you should treat it like one. In *International Conference on Learning Representations*, 2020a.
- Will Grathwohl, Kuan-Chieh Wang, Joern-Henrik Jacobsen, David Duvenaud, and Richard Zemel. Learning the stein discrepancy for training and evaluating energy-based models without sampling. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 3732–3747, 2020b.
- Will Sussman Grathwohl, Jacob Jin Kelly, Milad Hashemi, Mohammad Norouzi, Kevin Swersky, and David Duvenaud. No {mcmc} for me: Amortized sampling for fast and stable training of energy-based models. In *International Conference on Learning Representations*, 2021.
- Tian Han, Erik Nijkamp, Xiaolin Fang, Mitch Hill, Song-Chun Zhu, and Ying Nian Wu. Divergence triangle for joint training of generator model, energy-based model, and inferential model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8670–8679, 2019.
- Dan Hendrycks and Kevin Gimpel. Bridging nonlinearities and stochastic regularizers with gaussian error linear units, 2017.
- Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- Mitch Hill, Erik Nijkamp, Jonathan Craig Mitchell, Bo Pang, and Song-Chun Zhu. Learning probabilistic models from generator latent spaces with hat EBM. In *Advances in Neural Information Processing Systems*, 2022.
- Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020.
- Emiel Hoogeboom and Tim Salimans. Blurring diffusion models. In *The Eleventh International Conference on Learning Representations*, 2023.
- Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577. Curran Associates, Inc., 2022.
- Dongjun Kim, Seungjae Shin, Kyungwoo Song, Wanmo Kang, and Il-Chul Moon. Soft truncation: A universal training technique of score-based diffusion model for high precision score estimation. *arXiv preprint arXiv:2106.05527*, 2021.
- Diederik Kingma and Ruiqi Gao. Understanding diffusion objectives as the elbo with simple data augmentation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 65484–65516. Curran Associates, Inc., 2023.
- Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, 31, 2018.
- Alex Krizhevsky. Learning multiple layers of features from tiny images, 2009.
- Richard S. Laugesen, Prashant G. Mehta, Sean P. Meyn, and Maxim Raginsky. Poisson’s equation in nonlinear filtering. *SIAM Journal on Control and Optimization*, 53(1):501–525, 2015.
- Hankook Lee, Jongheon Jeong, Sejun Park, and Jinwoo Shin. Guiding energy-based models via contrastive latent variables. In *The Eleventh International Conference on Learning Representations*, 2022.

-
- Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.
- Meng Liu, Keqiang Yan, Bora Oztekin, and Shuiwang Ji. GraphEBM: Molecular graph generation with energy-based models. In *Energy Based Models Workshop - ICLR 2021*, 2021.
- Min Liu, Zhiqiang Cai, and Karthik Ramani. Deep ritz method with adaptive quadrature for linear elasticity. *Computer Methods in Applied Mechanics and Engineering*, 415:116229, 2023a.
- Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023b.
- Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.
- Eric Luhman, Troy Luhman, and Troy Luhman. Knowledge distillation in iterative generative models for improved sampling speed. *arXiv preprint arXiv:2101.02388*, 2021.
- Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018.
- Johannes Müller and Marius Zeinhofer. Deep ritz revisited. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2019.
- Kirill Neklyudov, Rob Brekelmans, Daniel Severo, and Alireza Makhzani. Action matching: Learning stochastic dynamics from samples. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25858–25889. PMLR, 2023.
- Erik Nijkamp, Mitch Hill, Song-Chun Zhu, and Ying Nian Wu. Learning non-convergent non-persistent short-run mcmc toward energy-based model. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- Erik Nijkamp, Mitch Hill, Tian Han, Song-Chun Zhu, and Ying Nian Wu. On the anatomy of mcmc-based maximum likelihood learning of energy-based models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04):5272–5280, 2020.
- Erik Nijkamp, Ruiqi Gao, Pavel Sountsov, Srinivas Vasudevan, Bo Pang, Song-Chun Zhu, and Ying Nian Wu. MCMC should mix: Learning energy-based model with neural transport latent space MCMC. In *International Conference on Learning Representations*, 2022.
- S. Yagiz Olmez, Amirhossein Taghvaei, and Prashant G. Mehta. Deep pf: Gain function approximation in high-dimensional setting. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pp. 4790–4795, 2020.
- Arghya Pal, Raphael C.-W. Phan, and KokSheik Wong. Synthesize-it-classifier: Learning a generative classifier through recurrent self-analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5161–5170, June 2021a.
- Soumyasundar Pal, Liheng Ma, Yingxue Zhang, and Mark Coates. Rnn with particle flow for probabilistic spatio-temporal forecasting. In *International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8336–8348, 2021b.
- Bo Pang, Tianyang Zhao, Xu Xie, and Ying Nian Wu. Trajectory prediction with latent belief energy-based model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11814–11824, June 2021.
- Gaurav Parmar, Dacheng Li, Kwonjoon Lee, and Zhuowen Tu. Dual contradistinctive generative autoencoder. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 823–832, 2021.

-
- Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pp. 1530–1538, 2015.
- Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. In *The Eleventh International Conference on Learning Representations*, 2023.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Cham, 2015. Springer International Publishing.
- Tim Salimans and Jonathan Ho. Should EBMs model the energy or the score? In *Energy Based Models Workshop - ICLR 2021*, 2021.
- Tim Salimans and Durk P Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.
- Tim Salimans, Andrej Karpathy, Xi Chen, and Diederik P Kingma. Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. *arXiv preprint arXiv:1701.05517*, 2017.
- Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. *Advances in neural information processing systems*, 33:12438–12448, 2020.
- Yang Song and Diederik P. Kingma. How to train your energy-based models, 2021.
- Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.
- Simone Carlo Surace, Anna Kutschireiter, and Jean-Pascal Pfister. How to avoid the curse of dimensionality: Scalability of particle filters with and without importance weights. *SIAM Review*, 61(1):79–91, 2019.
- Amirhossein Taghvaei, Prashant G. Mehta, and Sean P. Meyn. Diffusion map-based algorithm for gain function approximation in the feedback particle filter. *SIAM/ASA Journal on Uncertainty Quantification*, 8(3):1090–1117, 2020.
- Arash Vahdat and Jan Kautz. Nvae: A deep hierarchical variational autoencoder. In *Advances in Neural Information Processing Systems*, volume 33, pp. 19667–19679, 2020.
- Dafeng Wang, Hongbo Liu, Naiyao Wang, Yiyang Wang, Hua Wang, and Seán McLoone. Seem: A sequence entropy energy-based model for pedestrian trajectory all-then-one prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1070–1086, 2023.
- Zhisheng Xiao, Karsten Kreis, Jan Kautz, and Arash Vahdat. {VAEBM}: A symbiosis between variational autoencoders and energy-based models. In *International Conference on Learning Representations*, 2021a.
- Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. In *International Conference on Learning Representations*, 2021b.
- Jianwen Xie, Yang Lu, Ruiqi Gao, Song-Chun Zhu, and Ying Nian Wu. Cooperative training of descriptor and generator networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):27–45, 2020.
- Jianwen Xie, Yifei Xu, Zilong Zheng, Song-Chun Zhu, and Ying Nian Wu. Generative pointnet: Deep energy-based learning on unordered point sets for 3d generation, reconstruction and classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14976–14985, 2021a.

-
- Jianwen Xie, Zilong Zheng, and Ping Li. Learning energy-based model with variational auto-encoder as amortized sampler. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(12):10441–10451, 2021b.
- Jianwen Xie, Yaxuan Zhu, Jun Li, and Ping Li. A tale of two flows: Cooperative learning of langevin flow and normalizing flow toward energy-based model. *arXiv preprint arXiv:2205.06924*, 2022.
- Yilun Xu, Ziming Liu, Max Tegmark, and Tommi Jaakkola. Poisson flow generative models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 16782–16795, 2022.
- Tao Yang, Prashant G. Mehta, and Sean P. Meyn. Feedback particle filter. *IEEE Transactions on Automatic Control*, 58(10):2465–2480, 2013.
- Tao Yang, Henk A. P. Blom, and Prashant G. Mehta. The continuous-discrete time feedback particle filter. In *American Control Conference (ACC)*, pp. 648–653, 2014.
- Tao Yang, Richard S. Laugesen, Prashant G. Mehta, and Sean P. Meyn. Multivariable feedback particle filter. *Automatica*, 71:10–23, 2016. ISSN 0005-1098.
- Xiulong Yang, Qing Su, and Shihao Ji. Towards bridging the performance gaps of joint energy-based models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15732–15741, 2023.
- Sangwoong Yoon, Young-Uk Jin, Yung-Kyun Noh, and Frank C. Park. Energy-based models for anomaly detection: A manifold diffusion recovery approach. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- Yang You, Jing Li, Sashank Reddi, Jonathan Hseu, Sanjiv Kumar, Srinadh Bhojanapalli, Xiaodan Song, James Demmel, Kurt Keutzer, and Cho-Jui Hsieh. Large batch optimization for deep learning: Training bert in 76 minutes. In *International Conference on Learning Representations*, 2020.
- Sergey Zagoruyko and Nikos Komodakis. Wide Residual Networks. In *British Machine Vision Conference 2016*, York, France, January 2016. British Machine Vision Association. doi: 10.48550/arXiv.1605.07146.
- Yang Zhao, Jianwen Xie, and Ping Li. Learning energy-based generative models via coarse-to-fine expanding and sampling. In *International Conference on Learning Representations*, 2020.
- Yaxuan Zhu, Jianwen Xie, Ying Nian Wu, and Ruiqi Gao. Learning energy-based models by cooperative diffusion recovery likelihood. In *The Twelfth International Conference on Learning Representations*, 2024.

A Additional Results

In this section, we present additional experiments that further validate the effectiveness and efficiency of the proposed VPFB framework. We first conduct an ablation study to evaluate the contribution of key loss components and architectural choices, demonstrating their impact on both FID performance and convergence to the Boltzmann equilibrium. Then, we compare the computational efficiency of VPFB against recent EBM baselines, highlighting its advantages in training and inference time while maintaining competitive generative performance across model variants.

A.1 Ablation Study

To isolate and quantify the impact of individual components in the proposed VPFB loss function. Table 3 presents an ablation study conducted on a smaller (VPFB-Base) model and a reduced training batch size to accelerate training. Notably, the FID scores increase without (B) the covariance loss and (C) the cosine distance gradient alignment, indicating that these loss components are essential to the VPFB training. We note that since (B) learns only the normalized gradient and not the energy magnitude, it requires careful tuning of denormalization during ODE sampling.

Subsequently, (D) replaces the cosine distance with inner product, and (E) replaces the entire VPFB loss with the flow matching loss of Lipman et al. (2023). Although these loss configurations yield better FID performances, the sampling results and norm plots in Figures 19 - 22 show that neither of these configurations achieves Boltzmann equilibrium under SGLD sampling. Removing the cosine distance in (D) eliminates the scale invariance of cosine similarity, leading to large variations in gradient magnitudes that disrupt the stability of energy required for steady-state convergence. Similarly, the flow matching loss in (E) does not inherently enforce Boltzmann stationarity. Applying the same weighting $w(t)$ to the flow matching loss would halt the learning of the gradient field cutoff time t_{\max} .

For these reasons, we do not adopt the loss configurations (D) and (E) in our loss framework, despite the lower FID scores. In contrast, the variational nature of the covariance loss allows it to enforce that the energy values $\Phi(x, t)$ remain stationary near $t = 1$ without impeding learning. This covariance loss is fundamental as it corresponds to the Fokker-Planck dynamics underlying a probability path. More importantly, its adaptability to weighting ensures a proper establishment of the stationary Boltzmann distribution. Finally, by (F) incorporating a larger VPFB-Base model and (G) increasing the training batch size, we obtain improved FID scores with (A) the proposed VPFB loss function.

A.2 Computational Efficiency

Table 4 compares the training-time computational efficiency of VPFB against the recent EBM baselines. Here, we included an additional smaller model (VPFB-Base) with fewer parameters but a higher FID score. The training time and GPU memory footprint are measured on a single A100 GPU of 80G memory. Italicized values represent estimates for the VAEBM model based on experiments conducted using smaller batch size, as the model cannot be trained on a single GPU using the prescribed batch sizes. Additionally, Table 5 compares the inference-time computational efficiency of VPFB against the recent EBM baselines. Overall, these results suggest that our method provides improved computational efficiency in both training and

Table 3: Ablation study across different training losses and model configurations.

Loss Configuration	Model Variant	Parameter Count	Batch Size	FID ↓
(A) VPFB loss	Base	38.3M	64	9.45
(B) VPFB loss without covariance	Base	38.3M	64	13.1
(C) VPFB loss without cosine distance	Base	38.3M	64	11.3
(D) cosine distance → inner product	Base	38.3M	64	9.21
(E) VPFB loss → flow matching loss	Base	38.3M	64	8.90
(F) VPFB loss	Large	55.7M	64	7.66
(G) VPFB loss	Large	55.7M	128	6.72

Table 4: Comparison of training-time computational efficiency against EBM baselines.

Method	Sampling/Perturbation Approach	Parameter Count	Memory Usage (GB)	Training Iterations	Training Time (hrs)	FID ↓
VAEBM Xiao et al. (2021a)	SGLD + Variational Inference + Replay Buffer	135.9M	129	25K	414	12.2
DRL (Gao et al. (2021))	SGLD + Diffusion	38.6M	56	240K	172	9.58
CLEL (Lee et al. (2022))	SGLD + Replay Buffer	30.7M	10	100K	133	8.61
CDRL (Zhu et al. (2024))	SGLD + Amortized Inference + Diffusion	34.8M	69	400K	144	4.31
VPFB-Base (Ours)	Stationary-Enforced OT-FM	38.3M	35	300K	48	9.33
VPFB-Large (Ours)	Stationary-Enforced OT-FM	55.7M	60	300K	112	6.72

Table 5: Comparison of inference-time computational efficiency against EBM baselines.

Method	Numerical Approach	Parameter Count	Sampling Steps	Inference Time (s)	Training Latency (ms)	FID ↓
VAEBM (Xiao et al. (2021a))	SGLD	135.1M	16	21.3	13.31	12.2
CoopFlow (Xie et al. (2022))	SGLD	45.9M	30	2.5	0.833	15.8
DRL (Gao et al. (2021))	SGLD	34.8M	180	23.9	1.328	9.58
CDRL (Zhu et al. (2024))	SGLD	38.6M	90	12.2	1.356	4.31
VPFB (Ours)	ODE Solver	55.7M	74	14.6	1.968	6.72

inference compared to most strong EBM baselines while maintaining competitive FID scores, particularly for the base model, which has a parameter count similar to the baselines. Nonetheless, there remains a gap in FID performance relative to the state-of-the-art generative models.

B Additional Discussions

In this section, we discuss the strengths and limitations of our proposed VPFB framework in the broader context of energy-based generative modeling. We first highlight the advantages of VPFB over conventional diffusion and flow-based approaches, emphasizing its interpretability, theoretical alignment with Boltzmann energy, and improved sampling efficiency. Next, we examine the critical role of MCMC in achieving Boltzmann-convergent training and sampling. While our method avoids the high cost of long-run MCMC by leveraging structured probability paths, we show that such paths may fail to explore low-density regions in high-dimensional spaces. This limitation can result in mode collapse and degraded sample quality when using deterministic ODE-based sampling. Finally, we analyze the remaining performance gap between VPFB and state-of-the-art EBMs, attributing it to architectural simplicity, trade-offs between convergence and generative sharpness, and the added complexity of modeling marginal rather than conditional distributions.

B.1 Advantages over Diffusion and Flow-based Generative Models

Our proposed energy-parameterized potential flow offers several advantages over conventional diffusion and flow matching models, where vector fields are directly parameterized by neural networks rather than derived as the gradients of scalar-valued energy functions. Specifically, these benefits include:

Interpretable Energy Landscape via Marginal Density Modeling By explicitly parameterizing the vector field as the gradient of a scalar potential energy, our method provides a natural energy-based representation of data dynamics. Such a formulation supports key energy-based modeling tasks, including explicit (marginal) density estimation, composable generation, and OOD detection capabilities not inherently provided by conventional diffusion or flow matching approaches. As shown in Table 2, VPFB demonstrates strong OOD detection performance due to our proposed energy-based formulation.

Theoretical Connection to Boltzmann Energy Our Boltzmann energy formulation in Proposition 5 rigorously connects the deterministic potential flow to a stationary Boltzmann distribution characterized by the Boltzmann energy Φ_B . This theoretical foundation firmly situates our approach within the energy-based modeling framework, offering theoretical coherence that is lacking in existing diffusion and flow matching models. As a result, it allows our approach to combine the efficiency of flow-based probability paths with the interpretability and rigor of the Boltzmann energy representation. As shown in Figure 2, the potential flow converges effectively to the stationary Boltzmann distribution.

Optimality of Conservative Vector Field Our approach learns a purely gradient-based vector field $\nabla_x \Phi$, in contrast to diffusion and flow matching methods, which may incorporate divergence-free components, as noted in Neklyudov et al. (2023). By enforcing a conservative energy function through Helmholtz decomposition, our method reduces the dynamical cost associated with these divergence-free components, enabling more efficient particle transport and improving training efficiency, as demonstrated by the comparative benchmark in Table 4.

Efficient Deterministic ODE Sampling By eliminating the reliance on implicit MCMC sampling, our approach reduces computational overhead and avoids common convergence issues encountered in traditional EBMs. Furthermore, our potential flow formulation enables deterministic ODE-based sampling that is generally more stable and efficient for generating high-quality samples with fewer steps than stochastic sampling methods, as demonstrated by the comparative results in Table 5.

B.2 Incorporating Langevin Dynamics for Boltzmann-Convergent Sampling

Conventional EBM training often relies on convergent (long-run) MCMC sampling to thoroughly explore the data space and assign appropriate energy values across the landscape, particularly in low-density regions. This process helps smooth out sharp local minima and mitigates overfitting to high-density areas. In contrast, our framework employs a conditional homotopy $p(x | \bar{x}, t)$ (perturbation kernel) to guide training samples along a structured dynamic probability path. As noted by Nijkamp et al. (2019), such probability paths resemble short-run MCMC behavior, which limits their capacity to explore low-density regions in high-dimensional spaces. Consequently, many of these regions remain unseen during training and are thus poorly modeled in the resulting energy landscape. ODE-based samplers further exacerbate this issue due to their lack of stochasticity, making them more susceptible to becoming trapped in sharp local minima and suffering from poor mixing. Compared to stochastic samplers like Stochastic Gradient Langevin Dynamics (SGLD), deterministic sampling is inherently more sensitive to gaps in energy coverage. This limitation is not unique to our approach - it is a general challenge for diffusion-based and flow-based models that rely on time-dependent Gaussian perturbations to construct their training trajectories.

To achieve proper Boltzmann convergence and reduce mode collapse under deterministic ODE sampling, it is necessary to improve data space coverage beyond what is provided by structured Gaussian perturbations. This could be addressed by incorporating long-run MCMC during training, although doing so incurs the substantial computational overhead characteristic of traditional EBMs. As a potential direction for future work, we propose integrating long-run MCMC sampling within the reverse-diffusion conditional path, following techniques developed by Gao et al. (2021), Zhu et al. (2024), and Geng et al. (2024). However, these methods primarily model the conditional distribution $p(x_{t-1} | x_t)$ rather than the marginal data distribution $p(x)$, which remains the focus of our current work. As a result, these conditional EBMs do not establish a direct connection to the Boltzmann distribution $p(x) \propto e^{\Phi(x)}$, which forms the theoretical foundation of marginal energy-based modeling. Adapting such conditional modeling techniques to the marginal EBM setting would require substantial methodological developments. To offset the training cost of long-run MCMC, future work may also explore hybrid strategies involving pre-sampled replay buffers or distillation from a pre-trained convergent EBM.

B.3 Addressing Performance Gaps: Modeling Trade-offs and Theoretical Challenges

While our proposed VPFB framework achieves competitive FID performance, there remains a noticeable gap compared to state-of-the-art EBMs. In the following discussion, we analyze the underlying factors contributing to this discrepancy.

Difference in Model Architecture DDAEBM (Geng et al., 2024) utilizes a multi-model architecture comprising three distinct components: (1) a generator model parameterized by the modified U-Net architecture of Xiao et al. (2021b), (2) an energy model parameterized by the NCSN++ architecture from Song et al. (2021), and (3) a CNN-based encoder. This multi-component architecture enables specialized modules to collaboratively refine each other’s behavior through adversarial training, thereby enhancing generative performance. In contrast, our VPFB model adopts a simpler framework design, employing only a single EBM parameterized by the U-Net architecture described in Dhariwal & Nichol (2021). Although this single-

component architecture has a comparable parameter count to that of the NCSN++ used in DDAEBM, it may lack the collaborative optimization and mutual refinement advantages that arise from multi-model training setups. Consequently, our VPFB’s FID scores align more closely with those of single-energy or joint-energy EBMs (Salimans & Ho, 2021; Gao et al., 2021; Lee et al., 2022). We hypothesize that incorporating additional auxiliary model components with adversarial training strategies, as utilized by DDAEBM, could enhance the representational capacity and further sharpen the energy landscape of our VPFB model, potentially leading to improved FID performance. Investigating this hybrid approach, which balances computational efficiency with adversarial training strategy, will be reserved for our future work.

Boltzmann stationarity and FID Trade-off In our previous ablation study, we observed a notable trade-off between Boltzmann stationarity and FID performance. This observation aligns with insights from Agoritsas et al. (2023) and Nijkamp et al. (2020), which show that non-convergent EBMs outperform convergent EBMs trained with long-run MCMC sampling in image generation. Models such as DDAEBM fall within this class of non-convergent EBMs. We hypothesize that this is due to the inherent tension between accurate equilibrium modeling and sharp generative quality. Specifically, imposing strict Boltzmann convergence tends to smooth out the energy landscape, inadvertently flattening local minima that correspond to meaningful data modes. Although such smoothing enhances theoretical interpretability by faithfully approximating the true equilibrium of the Boltzmann distribution, it compromises the sharpness and detail of generated samples, leading to higher FID scores. To mitigate this limitation, we propose exploring advanced sampling techniques such as the Metropolis-Adjusted Langevin Algorithm (MALA) used in Pal et al. (2021a) or other adaptive short-run samplers in future work. The gradient-informed proposal and acceptance steps of MALA could potentially enhance local mode exploration efficiency without fully abandoning the Boltzmann equilibrium.

Conditional vs Marginal EBMs Another critical contributing factor is the distinction between conditional EBMs as modeled by DDAEBM, and marginal EBMs considered by VPFB. In particular, Geng et al. (2024) articulate that modeling conditional distributions simplifies the learning task, as these conditional distributions are inherently less multi-modal compared to complex marginal distributions. DDAEBM leverages this insight by decomposing the generation process into discrete diffusion steps, each focusing on simpler, conditional distributions that are easier to model effectively. In contrast, our VPFB explicitly models a global marginal distribution, thus inherently facing greater complexity due to increased modality. Consequently, achieving comparable generative performance poses additional challenges. To address this limitation, future work could explore hierarchical or multi-stage conditional modeling strategies to simplify explicit marginal modeling. Nevertheless, conditional EBMs inherently lack a direct relationship to the marginal Boltzmann distribution, which is fundamental to the theoretical underpinnings of our VPFB framework. Therefore, adapting conditional modeling techniques to fully marginal EBMs would require substantial development.

C Proofs and Derivations

C.1 Proof of Proposition 1

Proof. Based on the definitions of $q(x)$ and $p(\bar{x} \mid x)$, we can expand their logarithms (ignoring additive constants) as follows:

$$\log q(x) = -\frac{1}{2\omega^2} \frac{1}{2\omega^2} \|x\|^2 + (\text{terms independent of } x) \quad (26)$$

$$\log p(\bar{x} \mid x) = -\frac{1}{2\nu^2} \|\bar{x} - x\|^2 + (\text{terms independent of } x) \quad (27)$$

Substituting these into (5), we obtain:

$$h(x \mid \bar{x}, t) = -\frac{\alpha(t)}{2\omega^2} \|x\|^2 - \frac{\beta(t)}{2\nu^2} \|\bar{x} - x\|^2 \quad (28)$$

Expanding the squared term:

$$\|\bar{x} - x\|^2 = \|x\|^2 - 2x^T \bar{x} + \|\bar{x}\|^2 \quad (29)$$

and substituting it back into $h(x \mid \bar{x}, t)$:

$$h(x \mid \bar{x}, t) = -\left(\frac{\alpha(t)}{\omega^2} + \frac{\beta(t)}{\nu^2}\right) \|x\|^2 + \frac{\beta(t)}{\nu^2} x^T \bar{x} \quad (30)$$

Recognizing the quadratic form in terms of x , we identify that $\rho(x \mid \bar{x}, t)$ is a Gaussian density:

$$\rho(x \mid \bar{x}, t) = \mathcal{N}(x; \mu(t)\bar{x}, \sigma(t)^2 I) \quad (31)$$

whose mean $\mu(t)$ and variance $\sigma(t)^2$ can be obtained by completing the square.

Define

$$A := \frac{\alpha(t)}{\omega^2} + \frac{\beta(t)}{\nu^2}, \quad B := \frac{\beta(t)}{\nu^2} \quad (32)$$

Then the exponent becomes

$$h(x \mid \bar{x}, t) = -\frac{1}{2} \left[A \|x\|^2 - 2B x^T \bar{x} \right] \quad (33)$$

and we wish to express this quadratic form as follows

$$A \|x - \mu(t)\bar{x}\|^2 + (\text{terms independent of } x) \quad (34)$$

Expanding $A \|x - \mu(t)\bar{x}\|^2$, we obtain

$$A \|x - \mu(t)\bar{x}\|^2 = A \|x\|^2 - 2A\mu(t)x^T \bar{x} + A\mu(t)^2 \|\bar{x}\|^2 \quad (35)$$

To match the linear term, the mean of the Gaussian is thus given by

$$\mu(t) = \frac{B}{A} = \frac{\beta(t)/\nu^2}{\alpha(t)/\omega^2 + \beta(t)/\nu^2} = \text{sigmoid} \left(\log \left(\frac{\beta(t)}{\alpha(t)} \frac{\omega^2}{\nu^2} \right) \right) \quad (36)$$

where $\text{sigmoid}(z) = \frac{1}{1+e^{-z}}$ denotes the standard logistic (sigmoid) function.

By comparing with the standard Gaussian exponent

$$-\frac{1}{2\sigma^2} \|x - \mu(t)\bar{x}\|^2, \quad (37)$$

we deduce that the variance is given by

$$\sigma(t)^2 = \frac{1}{A} = \frac{1}{\alpha(t)/\omega^2 + \beta(t)/\nu^2}. \quad (38)$$

Using the expression obtained for $\mu(t)$, the standard deviation can also be written as

$$\sigma(t) = \sqrt{\frac{\nu^2}{\beta(t)}} \mu(t). \quad (39)$$

□

C.2 Proof of Proposition 2

Proof. Differentiating the conditional homotopy $\rho(x | \bar{x}, t)$ in (4) with respect to t , we have:

$$\begin{aligned} \frac{\partial \rho(x | \bar{x}, t)}{\partial t} &= \frac{1}{\int_{\Omega} e^{h(x|\bar{x},t)} dx} \frac{\partial [e^{h(x|\bar{x},t)}]}{\partial t} - \frac{e^{h(x|\bar{x},t)}}{[\int_{\Omega} e^{h(x|\bar{x},t)} dx]^2} \frac{\partial [\int_{\Omega} e^{h(x|\bar{x},t)} dx]}{\partial t} \\ &= \frac{1}{\int_{\Omega} e^{h(x|\bar{x},t)} dx} \frac{\partial [e^{h(x|\bar{x},t)}]}{\partial f} \frac{\partial h(x | \bar{x}, t)}{\partial t} - \frac{e^{h(x|\bar{x},t)}}{[\int_{\Omega} e^{h(x|\bar{x},t)} dx]^2} \int_{\Omega} \frac{\partial [e^{h(x|\bar{x},t)}]}{\partial f} \frac{\partial h(x | \bar{x}, t)}{\partial t} dx \\ &= \frac{e^{h(x|\bar{x},t)}}{\int_{\Omega} e^{h(x|\bar{x},t)} dx} \frac{\partial h(x | \bar{x}, t)}{\partial t} - \frac{e^{h(x|\bar{x},t)}}{\int_{\Omega} e^{h(x|\bar{x},t)} dx} \int_{\Omega} \frac{e^{h(x|\bar{x},t)}}{\int_{\Omega} e^{h(x|\bar{x},t)} dx} \frac{\partial h(x | \bar{x}, t)}{\partial t} dx \\ &= \rho(x | \bar{x}, t) \left(\frac{\partial h(x | \bar{x}, t)}{\partial t} - \int_{\Omega} \rho(x | \bar{x}, t) \frac{\partial h(x | \bar{x}, t)}{\partial t} dx \right) \\ &= -\frac{1}{2} \rho(x | \bar{x}, t) \left(\frac{d\alpha(t)}{dt} \frac{x^T x}{\omega^2} + \frac{d\beta(t)}{dt} \frac{(x - \bar{x})^T (x - \bar{x})}{\nu^2} \right. \\ &\quad \left. - \int_{\Omega} \rho(x | \bar{x}, t) \frac{d\alpha(t)}{dt} \frac{x^T x}{\omega^2} + \frac{d\beta(t)}{dt} \frac{(x - \bar{x})^T (x - \bar{x})}{\nu^2} dx \right) \end{aligned} \quad (40)$$

where we have applied the quotient rule in the first equation and the chain rule in the second equation.

Subsequently, define

$$\gamma(x, \bar{x}, t) = \frac{d\alpha(t)}{dt} \frac{\|x\|^2}{\omega^2} + \frac{d\beta(t)}{dt} \frac{\|x - \bar{x}\|^2}{\nu^2} \quad (41)$$

and using the fact that:

$$\frac{\partial \bar{\rho}(x, t)}{\partial t} = \frac{\partial \int_{\Omega} \rho(x | \bar{x}, t) p_{\text{data}}(\bar{x}) d\bar{x}}{\partial t} = \int_{\Omega} \frac{\partial \rho(x | \bar{x}, t)}{\partial t} p_{\text{data}}(\bar{x}) d\bar{x} \quad (42)$$

we can substitute (40) into (42) to obtain:

$$\frac{\partial \bar{\rho}(x, t)}{\partial t} = - \int_{\Omega} p_{\text{data}}(\bar{x}) \rho(x | \bar{x}, t) \left(\gamma(x, \bar{x}, t) - \int_{\Omega} \rho(x | \bar{x}, t) \gamma(x, \bar{x}, t) dx \right) d\bar{x} \quad (43)$$

Given that both $\rho(x | \bar{x}, t)$ and $p_{\text{data}}(\bar{x})$ are normalized (proper) density functions, writing (43) in terms of expectations yields the PDE in (9).

□

C.3 Proof of Proposition 3

Here, we used the Einstein tensor notation interchangeably with the conventional notation for vector dot product and matrix-vector multiplication in PDE. Also, we omit the time index t of $\Phi(x, t)$ in this section for brevity.

Proof. Applying forward Euler to the particle flow ODE (11) using step size Δ_t , we obtain:

$$x_{t+\Delta_t} = \alpha(x_t) = x_t + \Delta_t u(x_t) \quad (44)$$

where

$$u(x_t) = \nabla_x \Phi(x_t) \quad (45)$$

where x_t denotes the discretization of $x(t)$. Hereafter, we abbreviate $x_t, \alpha(x_t), \nu(x_t)$ as x, α, ν , respectively.

Assuming that the $\alpha : \Omega \rightarrow \Omega$ is a diffeomorphism (bijective function with differentiable inverse), the push-forward operator $\alpha_\# : \mathbb{R} \rightarrow \mathbb{R}$ defines the density transformation $\rho^\Phi(\alpha, t + \Delta_t) := \alpha_\# \rho^\Phi(x, t)$. Associated with this change of variable formula is the following density transformation:

$$\rho^\Phi(\alpha, t + \Delta_t) = \frac{1}{|D_x \alpha|} \rho^\Phi(x, t) \quad (46)$$

where $|D_x \alpha|$ denotes the Jacobian determinant of α , where the the Jacobian is taken with respect to x .

From (9) and (43), we obtain:

$$\frac{\partial \log \bar{\rho}(x, t)}{\partial t} = \frac{1}{\bar{\rho}(x, t)} \frac{\partial \bar{\rho}(x, t)}{\partial t} = - \frac{1}{\bar{\rho}(x, t)} \frac{1}{2} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] \quad (47)$$

Applying the forward Euler method to (47), we obtain:

$$\log \bar{\rho}(x, t + \Delta_t) \geq \log \bar{\rho}(x, t) - \frac{\Delta_t}{2} \frac{1}{\bar{\rho}(x, t)} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] \quad (48)$$

Applying the change-of-variables formula (46) and then substituting (48) into the KL divergence $\mathcal{D}_{\text{KL}} [\rho(x, t + \Delta_t) \| \bar{\rho}(x, t + \Delta_t)]$ at time $t + \Delta_t$, we obtain:

$$\begin{aligned} \mathcal{D}_{\text{KL}} [\rho^\Phi(x, t + \Delta_t) \| \bar{\rho}(x, t + \Delta_t)] &= \int_{\Omega} \rho^\Phi(x, t) \log \left(\frac{\rho^\Phi(\alpha, t + \Delta_t)}{\bar{\rho}(\alpha, t + \Delta_t)} \right) dx \\ &= \int_{\Omega} \rho^\Phi(x, t) \left(\log \rho^\Phi(x, t) - \log |D_x \alpha| - \log \bar{\rho}(\alpha, t) \right. \\ &\quad \left. + \frac{\Delta_t}{2} \frac{1}{\bar{\rho}(\alpha, t)} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(\alpha | \bar{x}, t) (\gamma(\alpha, \bar{x}, t) - \bar{\gamma}(\alpha, \bar{x}, t)) \right] + C \right) dx \end{aligned} \quad (49)$$

Consider minimizing the KL divergence (49) with respect to α as follows:

$$\begin{aligned} \min_{\alpha} \mathcal{D}_{\text{KL}}(\alpha) &= \min_{\alpha} \underbrace{\frac{\Delta_t}{2} \int_{\Omega} \rho^\Phi(x, t) \frac{1}{\bar{\rho}(\alpha, t)} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(\alpha | \bar{x}, t) (\gamma(\alpha, \bar{x}, t) - \bar{\gamma}(\alpha, \bar{x}, t)) \right] dx}_{\mathcal{D}_1^{\text{KL}}(\alpha)} \\ &\quad - \underbrace{\int_{\Omega} \rho^\Phi(x, t) \log \bar{\rho}(\alpha, t) dx}_{\mathcal{D}_2^{\text{KL}}(\alpha)} - \underbrace{\int_{\Omega} \rho^\Phi(x, t) \log |D_x \alpha| dx}_{\mathcal{D}_3^{\text{KL}}(\alpha)} \end{aligned} \quad (50)$$

where we have neglected the constant terms that do not depend on α .

To solve the optimization (50), we consider the following optimality condition in the first variation of \mathcal{D}_{KL} :

$$\mathcal{I}(\alpha, \nu) = \frac{d}{d\epsilon} \mathcal{D}_{\text{KL}}(\alpha + \epsilon \nu) \Big|_{\epsilon=0} = 0 \quad (51)$$

This condition must hold for all trial functions ν .

Subsequently, taking the variational derivative of the first functional $\mathcal{D}_1^{\text{KL}}$ in (50), we obtain:

$$\begin{aligned}
\mathcal{I}^1(\alpha, \nu) &= \left. \frac{d}{d\epsilon} \mathcal{D}_1^{\text{KL}}(\alpha + \epsilon\nu) \right|_{\epsilon=0} \\
&= \frac{\Delta_t}{2} \int_{\Omega} \rho^{\Phi}(x, t) \frac{d}{d\epsilon} \left\{ \frac{1}{\bar{\rho}(\alpha + \epsilon\nu, t)} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(\alpha + \epsilon\nu \mid \bar{x}, t) (\gamma(\alpha + \epsilon\nu, \bar{x}, t) - \bar{\gamma}(\alpha + \epsilon\nu, \bar{x}, t)) \right] \right\} \Big|_{\epsilon=0} dx \\
&= \frac{\Delta_t}{2} \int_{\Omega} \rho^{\Phi}(x, t) \frac{\partial}{\partial \alpha} \left\{ \frac{1}{\bar{\rho}(\alpha, t)} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(\alpha \mid \bar{x}, t) (\gamma(\alpha, \bar{x}, t) - \bar{\gamma}(\alpha, \bar{x}, t)) \right] \right\} \nu dx \\
&= \frac{\Delta_t}{2} \int_{\Omega} \rho^{\Phi}(x, t) D_x \left\{ \frac{1}{\bar{\rho}(\alpha, t)} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(\alpha \mid \bar{x}, t) (\gamma(\alpha, \bar{x}, t) - \bar{\gamma}(\alpha, \bar{x}, t)) \right] \right\} (D_x \alpha)^{-1} \nu dx
\end{aligned} \tag{52}$$

where the last equation is due to chain rule $\frac{\partial f}{\partial \alpha} = D_x f (D_x \alpha)^{-1}$.

Applying the Taylor series expansion to the derivative $\frac{\partial g}{\partial x_i}(\alpha)$ with respect to x_i yields:

$$\frac{\partial g(\alpha)}{\partial x_i} = \frac{\partial g(x + \Delta_t u)}{\partial x_i} = \frac{\partial g(x)}{\partial x_i} + \Delta_t \sum_j \frac{\partial^2 g(x)}{\partial x_i \partial x_j} u_j + O(\Delta_t^2) \tag{53}$$

In addition, the inverse of Jacobian $D_x \alpha^{-1}$ can be expanded via the Neuman series to obtain:

$$D_x \alpha^{-1} = (I + \Delta_t D_x u)^{-1} = I - \Delta_t D_x u + O(\Delta_t^2) \tag{54}$$

Using the Taylor series and Neuman series expansions in (53) and (54), we can write (52) in tensor notation, as follows:

$$\mathcal{I}^1(\alpha, \nu) = \frac{\Delta_t}{2} \int_{\Omega} \rho^{\Phi}(x, t) \sum_i \frac{\partial}{\partial x_i} \left\{ \frac{1}{\bar{\rho}(x, t)} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x \mid \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] \right\} \nu_i dx + O(\Delta_t^2) \tag{55}$$

Taking the variational derivative of the second functional $\mathcal{D}_2^{\text{KL}}$ in (50) yields:

$$\begin{aligned}
\mathcal{I}^2(\alpha, \nu) &= \left. \frac{d}{d\epsilon} \mathcal{D}_2^{\text{KL}}(\alpha + \epsilon\nu) \right|_{\epsilon=0} = \int_{\Omega} \rho^{\Phi}(x, t) \frac{d}{d\epsilon} \log \bar{\rho}(\alpha + \epsilon\nu) \Big|_{\epsilon=0} dx \\
&= \int_{\Omega} \rho^{\Phi}(x, t) \frac{1}{\bar{\rho}(\alpha, t)} \nabla_x \bar{\rho}(\alpha, t) \cdot \nu dx = \int_{\Omega} \rho^{\Phi}(x, t) \nabla_x \log \bar{\rho}(\alpha, t) \cdot \nu dx
\end{aligned} \tag{56}$$

where we have used the derivative identity $d \log g = \frac{1}{g} dg$ to obtain the second equation.

Using the Taylor series expansion (53), we can write (56) in tensor notation, as follows:

$$\begin{aligned}
\mathcal{I}^2(\alpha, \nu) &= - \int_{\Omega} \rho^{\Phi}(x, t) \sum_i \left(\frac{\partial \log \bar{\rho}(x, t)}{\partial x_i} - \Delta_t \sum_j \frac{\partial^2 \log \bar{\rho}(x, t)}{\partial x_i \partial x_j} u_j \right) \nu_i dx + O(\Delta_t^2) \\
&= - \int_{\Omega} \rho^{\Phi}(x, t) \sum_i \left(\frac{\partial \log \bar{\rho}(x, t)}{\partial x_i} - \Delta_t \sum_j \frac{\partial^2 \log \bar{\rho}(x, t)}{\partial x_i \partial x_j} u_j \right) \nu_i dx + O(\Delta_t^2)
\end{aligned} \tag{57}$$

Similarly, taking the variational derivative of the $\mathcal{D}_3^{\text{KL}}$ term in (50), we obtain:

$$\begin{aligned}\mathcal{I}^3(\alpha, \nu) &= \frac{d}{d\epsilon} \mathcal{D}_3^{\text{KL}}(\alpha + \epsilon\nu) \Big|_{\epsilon=0} = \int_{\Omega} \rho^{\Phi}(x, t) \frac{d}{d\epsilon} \log |D(\alpha + \epsilon\nu)| \Big|_{\epsilon=0} dx \\ &= \int_{\Omega} \rho^{\Phi}(x, t) \frac{1}{|D_x \alpha|} \frac{d}{d\epsilon} |D(\alpha + \epsilon\nu)| \Big|_{\epsilon=0} dx = \int_{\Omega} \rho^{\Phi}(x, t) \operatorname{tr} (D_x \alpha^{-1} D\nu) dx\end{aligned}\quad (58)$$

where we have used the following Jacobi's formula:

$$\frac{d}{d\epsilon} |D(\alpha + \epsilon\nu)| \Big|_{\epsilon=0} = |D_x \alpha| \operatorname{tr} (D_x \alpha^{-1} D\nu) \quad (59)$$

to obtain the last equation in (58).

Substituting in (54) and using the Taylor series expansion (53), (56) can be written in tensor notation as follows:

$$\begin{aligned}\mathcal{I}^3(\alpha, \nu) &= \int_{\Omega} \sum_i \left(\rho^{\Phi}(x, t) \frac{\partial \nu_i}{\partial x_i} - \Delta_t \sum_j \rho^{\Phi}(x, t) \frac{\partial u_j}{\partial x_i} \frac{\partial \nu_i}{\partial x_j} \right) dx + O(\Delta_t^2) \\ &= \int_{\Omega} \sum_i \left(\frac{\partial \rho^{\Phi}(x, t)}{\partial x_i} \nu_i - \Delta_t \sum_j \frac{\partial}{\partial x_j} \left\{ \rho^{\Phi}(x, t) \frac{\partial u_j}{\partial x_i} \right\} \nu_i \right) dx + O(\Delta_t^2) \\ &= \int_{\Omega} \sum_i \left(\frac{\partial \rho^{\Phi}(x, t)}{\partial x_i} - \Delta_t \sum_j \frac{\partial}{\partial x_j} \left\{ \rho^{\Phi}(x, t) \frac{\partial u_j}{\partial x_i} \right\} \right) \nu_i dx + O(\Delta_t^2)\end{aligned}\quad (60)$$

where we have used integration by parts to obtain the second equation.

Taking the limit $\lim \Delta_t \rightarrow 0$, the terms $O(\Delta_t^2)$ that approach zero exponentially vanish. Subtracting (55) by (57) and (60), then equating to zero, we obtain the first-order optimality condition (51) as follows:

$$\begin{aligned}\int_{\Omega} \bar{\rho}(x, t) \sum_i \sum_j - \frac{\partial}{\partial x_i} \left\{ \frac{1}{\bar{\rho}(x, t)} \frac{\partial}{\partial x_j} \left\{ \bar{\rho}(x, t) u_j \right\} \right\} \\ + \frac{1}{2} \frac{\partial}{\partial x_i} \left\{ \frac{1}{\bar{\rho}(x, t)} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] \right\} \nu_i dx = 0\end{aligned}\quad (61)$$

where we have assumed that $\rho^{\Phi}(x, t) \equiv \bar{\rho}(x, t)$ holds and have used the following identities:

$$\begin{aligned}\frac{\partial \log \bar{\rho}(x, t)}{\partial x_i} &= \frac{1}{\bar{\rho}(x, t)} \frac{\partial \bar{\rho}(x, t)}{\partial x_i} \\ \frac{\partial^2 \log \bar{\rho}(x, t)}{\partial x_i \partial x_j} &= \frac{\partial}{\partial x_i} \left(\frac{1}{\bar{\rho}(x, t)} \frac{\partial \bar{\rho}(x, t)}{\partial x_j} \right)\end{aligned}\quad (62)$$

Given that ν_i can take any value, equation (61) holds (in the weak sense) only if the terms within the round bracket vanish. Integrating this term with respect to the x_i , we obtain:

$$\sum_j \frac{\partial}{\partial x_j} \left\{ \bar{\rho}(x, t) u_j \right\} = \frac{1}{2} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] + \bar{\rho}(x, t) C \quad (63)$$

which can also be written in vector notation as follows:

$$\nabla_x \cdot (\bar{\rho}(x, t) u) = \frac{1}{2} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] + \bar{\rho}(x, t) C \quad (64)$$

To find the scalar constant C , we integrate both sides of (64) to obtain:

$$\begin{aligned} \int_{\Omega} \nabla_x \cdot (\bar{\rho}(x, t) u) dx &= \frac{1}{2} \int_{\Omega} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] dx + \int_{\Omega} \bar{\rho}(x, t) C dx \\ &= \frac{1}{2} \int_{\Omega} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] dx + C \end{aligned} \quad (65)$$

Applying the divergence theorem to the left-hand side of (65), we obtain:

$$\int_{\Omega} \nabla_x \cdot (\bar{\rho}(x, t) u) dx = \int_{\partial\Omega} \bar{\rho}(x, t) u \cdot \hat{n} dx \quad (66)$$

where \hat{n} is the outward unit normal vector to the boundary $\partial\Omega$ of Ω .

Given that $\bar{\rho}(x, t)$ is a normalized (proper) density with compact support (vanishes on the boundary), the term (66) becomes zero, leading to $C = 0$. Substituting this result along with $u(x) = \nabla_x \Phi(x)$ into (64), we arrive at the following PDE:

$$\nabla_x \cdot (\bar{\rho}(x, t) \nabla_x \Phi(x)) = \frac{1}{2} \mathbb{E}_{p_{\text{data}}(\bar{x})} \left[\rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] \quad (67)$$

Therefore, assuming that the base case $\rho_0(x) \equiv \bar{\rho}_0(x)$ holds and that a solution to (67) exists at every t , the proposition follows by the principle of induction. \square

C.4 Proof of Proposition 4

To show that the conditional and marginal homotopies satisfy the reverse diffusion process, we first express the forward-time SDE and ODE of Song et al. (2021):

$$\begin{aligned} dx(\tau) &= f(\tau) x(\tau) d\tau + g(\tau) dW(\tau) \\ \frac{dx(\tau)}{d\tau} &= f(\tau) x(\tau) - \frac{1}{2} g(\tau)^2 \nabla_x \log p(x, \tau) \end{aligned} \quad (68)$$

in terms of reverse time $t = 1 - \tau$, via applying the change of variable $dt = -d\tau$ as follows:

$$\begin{aligned} dx(t) &= -f(t) x(t) dt + g(t) dW(t) \\ \frac{dx(t)}{dt} &= -f(t) x(t) + \frac{1}{2} g(t)^2 \nabla_x \log \bar{\rho}(x, t) \end{aligned} \quad (69)$$

which gives (15) and (17).

Substituting the marginal score $\nabla_x \log \bar{\rho}(x, t)$ with the conditional score:

$$\nabla_x \log \rho(x | \bar{x}, t) = \frac{1}{\rho(x | \bar{x}, t)} \nabla_x \rho(x | \bar{x}, t) = -\frac{\epsilon}{\sigma(t)} \quad (70)$$

and applying reparameterization $x(t) = \mu(t) \bar{x} + \sigma(t) \epsilon$ and (16), we can write the conditional ODE as follows:

$$\begin{aligned} \frac{dx(t)}{dt} &= v(x | \bar{x}, t) \\ &= -f(t) x(t) + \frac{1}{2} g(t)^2 \nabla_x \log \rho(x | \bar{x}, t) \\ &= -f(t) x(t) + \frac{1}{2} g(t)^2 \frac{\epsilon}{\sigma(t)} \\ &= -f(t) x(t) + \sigma(t) \left(\dot{\sigma}(t) + f(t) \sigma(t) \right) \frac{\epsilon}{\sigma(t)} \\ &= \frac{\dot{\mu}(t)}{\mu(t)} \left(x(t) - \sigma(t) \epsilon \right) + \dot{\sigma}(t) \epsilon \\ &= \dot{\mu}(t) \bar{x} + \dot{\sigma}(t) \epsilon \end{aligned} \quad (71)$$

and thus corresponds to the conditional vector field defined in flow matching (Lipman et al., 2023).

Marginalizing (71) with respect to

$$p_{\text{data}}(\bar{x} | x) = \frac{\rho(x | \bar{x}, t) p_{\text{data}}(\bar{x})}{\bar{\rho}(x, t)} \quad (72)$$

and substituting (19) and applying (70), we obtain

$$\begin{aligned} v(x, t) &= \int_{\Omega} \left(-f(t) x(t) + \frac{1}{2} g(t)^2 \nabla_x \log \rho(x | \bar{x}, t) \right) \frac{\rho(x | \bar{x}, t) p_{\text{data}}(\bar{x})}{\bar{\rho}(x, t)} d\bar{x} \\ &= -f(t) x(t) + \frac{1}{2} g(t)^2 \int_{\Omega} \nabla_x \log \rho(x | \bar{x}, t) \frac{\rho(x | \bar{x}, t) p_{\text{data}}(\bar{x})}{\bar{\rho}(x, t)} d\bar{x} \\ &= -f(t) x(t) + \frac{1}{2} g(t)^2 \int_{\Omega} \frac{1}{\rho(x | \bar{x}, t)} \frac{\rho(x | \bar{x}, t) p_{\text{data}}(\bar{x})}{\bar{\rho}(x, t)} \nabla_x \rho(x | \bar{x}, t) d\bar{x} \\ &= -f(t) x(t) + \frac{1}{2} g(t)^2 \frac{1}{\bar{\rho}(x, t)} \nabla_x \bar{\rho}(x, t) \\ &= -f(t) x(t) + \frac{1}{2} g(t)^2 \frac{1}{\bar{\rho}(x, t)} \nabla_x \log \bar{\rho}(x, t) \end{aligned} \quad (73)$$

and thus corresponds to the marginal probability flow ODE 17.

C.5 Proof of Proposition 5

Proof. Based on the result of Proposition 4 and using (12), we can express the homotopy matching problem

$$\frac{\partial \rho_{\Phi}(x, t)}{\partial t} = \frac{\partial \bar{\rho}(x, t)}{\partial t} \quad (74)$$

equivalently as

$$\nabla_x \cdot \left(\rho_{\Phi} \nabla_x \Phi(x, t) \right) = \nabla_x \cdot \left(\rho_{\Phi} \left(-f(t) x(t) + \frac{1}{2} g(t)^2 \nabla_x \log \bar{\rho}(x, t) \right) \right) \quad (75)$$

Given that this matching holds identically, we have

$$\nabla_x \Phi(x, t) = -f(t) x(t) + \frac{1}{2} g(t)^2 \nabla_x \log \bar{\rho}(x, t) \quad (76)$$

Furthermore, given that both the forward-time ODE and SDE of Song et al. (2021) exhibit the same marginal probability density $\bar{\rho}(x, t)$, it is shown that they satisfy the following reverse-time SDE:

$$dx(\tau) = \left(f(\tau) x(\tau) - g(\tau)^2 \nabla_x \log \bar{\rho}(x, t) \right) d\tau + g(\tau) dW(\tau) \quad (77)$$

which reverses the diffusion process as outlined by Anderson (1982) and Song et al. (2021). Applying the change of variable $dt = -d\tau$, this reverse-time SDE can similarly be written in terms of $t = 1 - \tau$ as

$$dx(t) = - \left(f(t) x(t) - g(t)^2 \nabla_x \log \bar{\rho}(x, t) \right) dt + g(t) dW(t) \quad (78)$$

where $dW(t)$ does not change sign, since the Wiener process is invariant under time reversal.

Subsequently, the Fokker-Plank dynamic that governs the time evolution of the marginal density homotopy $\bar{\rho}(x, t)$ is given by

$$\frac{\partial \bar{\rho}(x, t)}{\partial t} = - \nabla_x \cdot \left(\bar{\rho}(x, t) \left(-f(t) x(t) + g(t)^2 \nabla_x \log \bar{\rho}(x, t) \right) \right) + \frac{1}{2} g(t)^2 \Delta_x \bar{\rho}(x, t) \quad (79)$$

where $\Delta_x = \nabla_x \cdot \nabla_x$ denotes the Laplacian. By substituting (76) into this Fokker-Planck equation, we then have

$$\frac{\partial \bar{\rho}(x, t)}{\partial t} = -\nabla_x \cdot \left(\bar{\rho}(x, t) \left(2 \nabla_x \Phi(x, t) + f(t) x(t) \right) \right) + \frac{1}{2} g(t)^2 \Delta_x \bar{\rho}(x, t) \quad (80)$$

At equilibrium $\frac{\partial \bar{\rho}(x, t)}{\partial t} = 0$, the Fokker-Planck equation admits a unique normalized steady-state solution, given by the Boltzmann distribution:

$$p_B(x) \propto \exp \left(\frac{2}{g_\infty^2} \left(2 \Phi_\infty(x) + \frac{f_\infty}{2} x(t)^T x(t) \right) \right) \quad (81)$$

when the potential energy function, the drift coefficient, and the diffusion coefficient reach their time-independent steady states $\Phi_\infty(x)$, f_∞ and g_∞ at equilibrium. The Boltzmann distribution can then be written in terms of a coherent Boltzmann energy Φ_B considered in EBMs, as follows:

$$p_B(x) = \frac{e^{\Phi_B}}{Z} \quad (82)$$

where

$$\Phi_B(x) = \frac{4 \Phi_\infty(x) + f_\infty \|x\|^2}{g_\infty^2} \quad (83)$$

□

and $Z = \int_\Omega e^{\Phi_B(x)} dx$ is the normalizing constant.

C.6 Proof of Proposition 6

Proof. The variational loss function in (22) can be written as follows:

$$\mathcal{L}(\Phi, t) = \frac{1}{2} \mathbb{E}_{\rho(x|\bar{x}, t) p_{\text{data}}(\bar{x})} \left[\Phi(x) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \right] + \frac{1}{2} \mathbb{E}_{\bar{\rho}(x, t)} \left[\|\nabla_x \Phi(x)\|^2 \right] \quad (84)$$

where we have assumed, without loss of generality, that a normalized energy $\bar{E}_\theta(x, t) = 0$. For an unnormalized solution $\Phi(x)$, we can always obtain a normalization by subtracting its mean.

The optimal solution Φ of the functional (84) is given by the first-order optimality condition:

$$\mathcal{I}(\Phi, \Psi) = \frac{d}{d\epsilon} \mathcal{L}(\Phi(x) + \epsilon \Psi(x), t) \Big|_{\epsilon=0} = 0 \quad (85)$$

which must hold for all trial function Ψ .

Taking the variational derivative of the particle flow objective (85) with respect to ϵ , we have:

$$\begin{aligned} \mathcal{I}(\Phi, \Psi) &= \frac{d}{d\epsilon} \mathcal{L}(\Phi + \epsilon \Psi) \Big|_{\epsilon=0} \\ &= \frac{1}{2} \int_{\Omega \times \Omega} p_{\text{data}}(\bar{x}) \rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \frac{d}{d\epsilon} (\Phi + \epsilon \Psi) d\bar{x} dx \\ &\quad + \frac{1}{2} \int_{\Omega} \bar{\rho}(x, t) \frac{d}{d\epsilon} \|\nabla_x (\Phi + \epsilon \Psi)\|^2 dx \\ &= \frac{1}{2} \int_{\Omega \times \Omega} p_{\text{data}}(\bar{x}) \rho(x | \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \Psi d\bar{x} dx + \int_{\Omega} \bar{\rho}(x, t) \nabla_x \Phi \cdot \nabla_x \Psi dx \end{aligned} \quad (86)$$

Given that $\Phi \in \mathcal{H}_0^1(\Omega; \rho)$, its value vanishes on the boundary $\partial\Omega$. Therefore, the second summand of the last expression in (86) can be written, via multivariate integration by parts, as

$$\int_{\Omega} \bar{\rho}(x, t) \nabla_x \Phi \cdot \nabla_x \Psi = - \int_{\Omega} \nabla_x \cdot (\bar{\rho}(x, t) \nabla_x \Phi) \Psi \, dx \quad (87)$$

By substituting (87) into (86), we get

$$\mathcal{I}(\Phi, \Psi) = \int_{\Omega} \left(\frac{1}{2} \int_{\Omega} p_{\text{data}}(\bar{x}) \rho(x \mid \bar{x}, t) (\gamma(x, \bar{x}, t) - \bar{\gamma}(x, \bar{x}, t)) \, d\bar{x} - \int_{\Omega} \nabla_x \cdot (\bar{\rho}(x, t) \nabla_x \Phi) \right) \Psi \, dx \quad (88)$$

and equating it to zero, we obtain the weak formulation (21) of the density-weighted Poisson's equation.

Given that the Poincaré inequality (23) holds, Theorem 2.2 of Laugesen et al. (2015) presents a rigorous proof of existence and uniqueness for the solution of the weak formulation (21), based on the Hilbert-space form of the Riesz representation theorem. \square

D Experimental Details

D.1 Model architecture

Our network architectures for the autonomous and time-varying VPFB models are based on the WideResNet (Zagoruyko & Komodakis, 2016) and the U-Net (Ronneberger et al., 2015), respectively. For WideResNet, we include a spectral regularization loss during model training to penalize the spectral norm of the convolutional layer. Also, we apply weight normalization with data-dependent initialization (Salimans & Kingma, 2016) on the convolutional layers to further regularize the model’s output. Our WideResNet architecture adopts the model hyperparameters reported by Xiao et al. (2021a). For U-Net, we remove the final scale-by-sigma operation (Kim et al., 2021; Song et al., 2021) and replace it with the Euclidean norm $\frac{1}{2}\|x - f_\theta(x)\|^2$ computed between the input $x(t)$ and the output of the U-Net $f_\theta(x)$. Our U-Net architecture adopts the hyperparameters used by Lipman et al. (2023). In both the WideResNet and U-Net models, we replace LeakyReLU activations with Gaussian Error Linear Unit (GELU) activations (Hendrycks & Gimpel, 2017), which we find improves training stability and convergence.

D.2 Training

We use the Lamb optimizer (You et al., 2020) and a learning rate of 10^{-3} for all the experiments. We find that Lamb performs better than Adam over large learning rates. We use a batch size of 128 and 64 for training CIFAR-10 and CelebA, respectively. For all experiments, we set a cutoff time of $t_{\max} = 1 - 10^{-5}$, a terminal time of $t_{\text{end}} = 1$, a decay exponent of $\kappa = 1.5$, and a spectral gap constant of $\eta = 10^{-4}$ during training. Here, the mean and standard deviation scheduling functions $\mu(t) = t$ and $\sigma(t) = 1 - t$ follow those defined by the OT-FM path. All models are trained on a single NVIDIA A100 (80GB) GPU until the FID scores, computed on 5k samples, no longer show improvement. We observe that the models converge within 300k training iterations.

D.3 Numerical Solver

In our experiments, the default ODE solver is the black-box solver from the SciPy library using the RK45 method (Dormand & Prince, 1980), following the approach of Xu et al. (2022). We allow additional ODE iterations to further refine the samples in regions of high likelihood, which we observe improves the quality of the generated images. This is achieved by extending the time horizon; our experiments indicate that setting the terminal time to $t_{\text{end}} = 1.575$ yields the best ODE sampling results.

D.4 Datasets

We conduct our experiments using the CIFAR-10 (Krizhevsky, 2009) and CelebA (Liu et al., 2015) datasets. CIFAR-10 consists of 50,000 training images and 10,000 test images at a resolution of 32×32 . The CelebA dataset contains 202,599 face images, with 162,770 used for training and 19,962 for testing. Each image is first cropped to 178×178 before being resized to 64×64 . During resizing, we enable anti-aliasing by setting

Algorithm 1 VPFB Training

input: Initial energy model Φ_θ , mean and standard deviation scheduling functions $\mu(t)$, $\sigma(t)$, cutoff time t_{\max} , terminal time t_{end} , decay exponent κ , spectral gap constant η , and batch size B .
repeat
 Sample observed data $\bar{x}_i \sim p_{\text{data}}(\bar{x})$, $t_i \sim \mathcal{U}(0, t_{\text{end}})$, and $\epsilon_i \sim \mathcal{N}(0, I)$
 Set $t_i = \min(t_i, t_{\max})$
 Sample $x_i \sim \rho(x \mid \bar{x}, t_i)$ via reparameterization $x_i = \mu(t_i)\bar{x}_i + \sigma(t_i)\epsilon_i$
 Compute gradient $\nabla_x \Phi_\theta(x_i, t_i)$ w.r.t. x_i via backpropagation
 Calculate VPFB loss $\frac{1}{B} \sum_{i=1}^B \mathcal{L}(\Phi_\theta, t_i)$
 Backpropagate and update model parameters θ
until FID converges

the antialias parameter to True. Additionally, we apply random horizontal flipping as a data augmentation technique.

D.5 Quantitative Evaluation

We employ the FID and inception scores as quantitative evaluation metrics for assessing the quality of generated samples. For CIFAR-10, the FID is computed between 50,000 samples and the pre-computed statistics from the training set, following Heusel et al. (2017). For CelebA 64×64 , we adopt the setting from Song & Ermon (2020), computing the FID between 5,000 samples and the pre-computed statistics from the test set. For model selection, we follow Song et al. (2021), selecting the checkpoint with the lowest FID score, computed on 2,500 samples every 10,000 iterations.

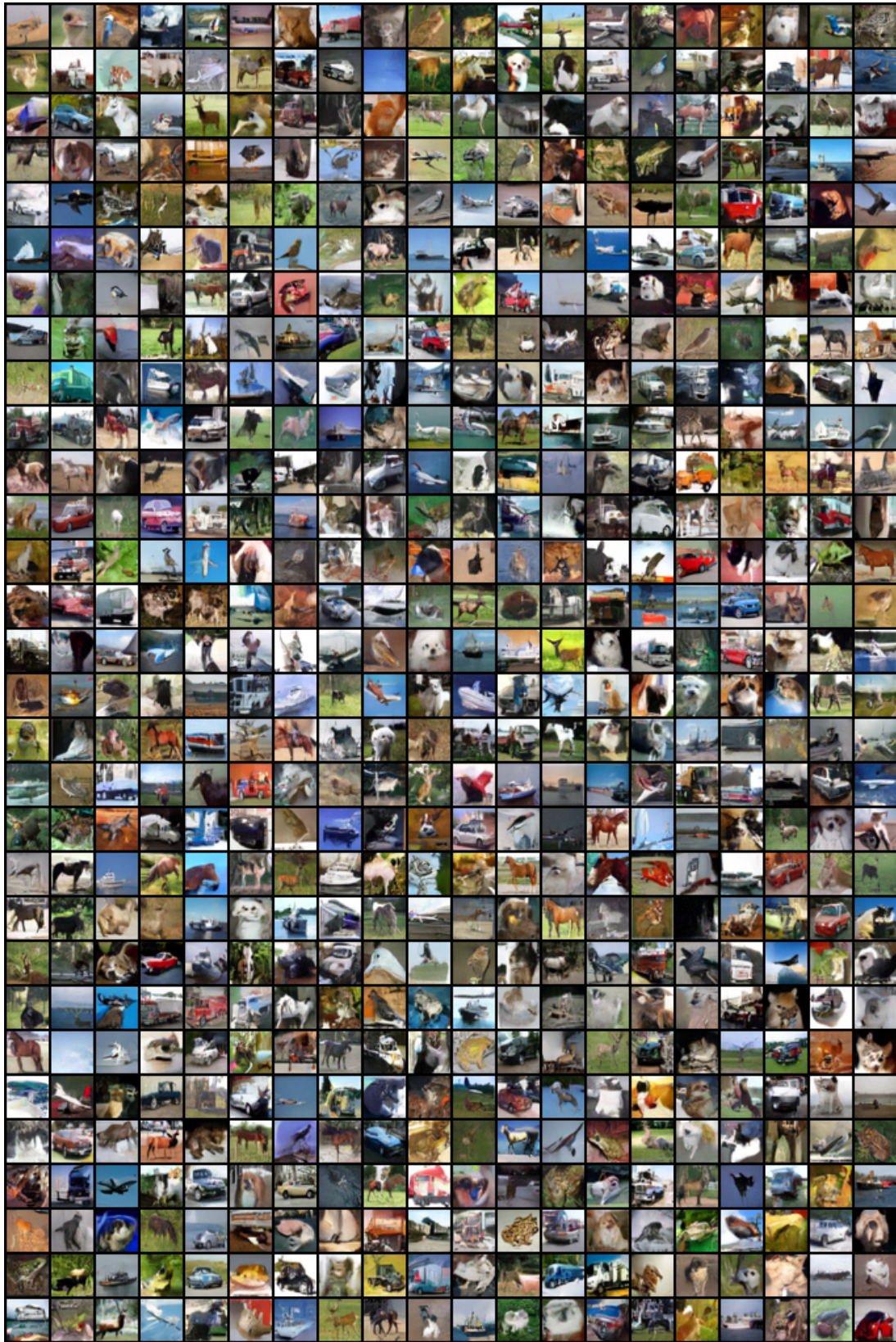


Figure 6: Additional uncured samples on unconditional CIFAR-10 32×32 .

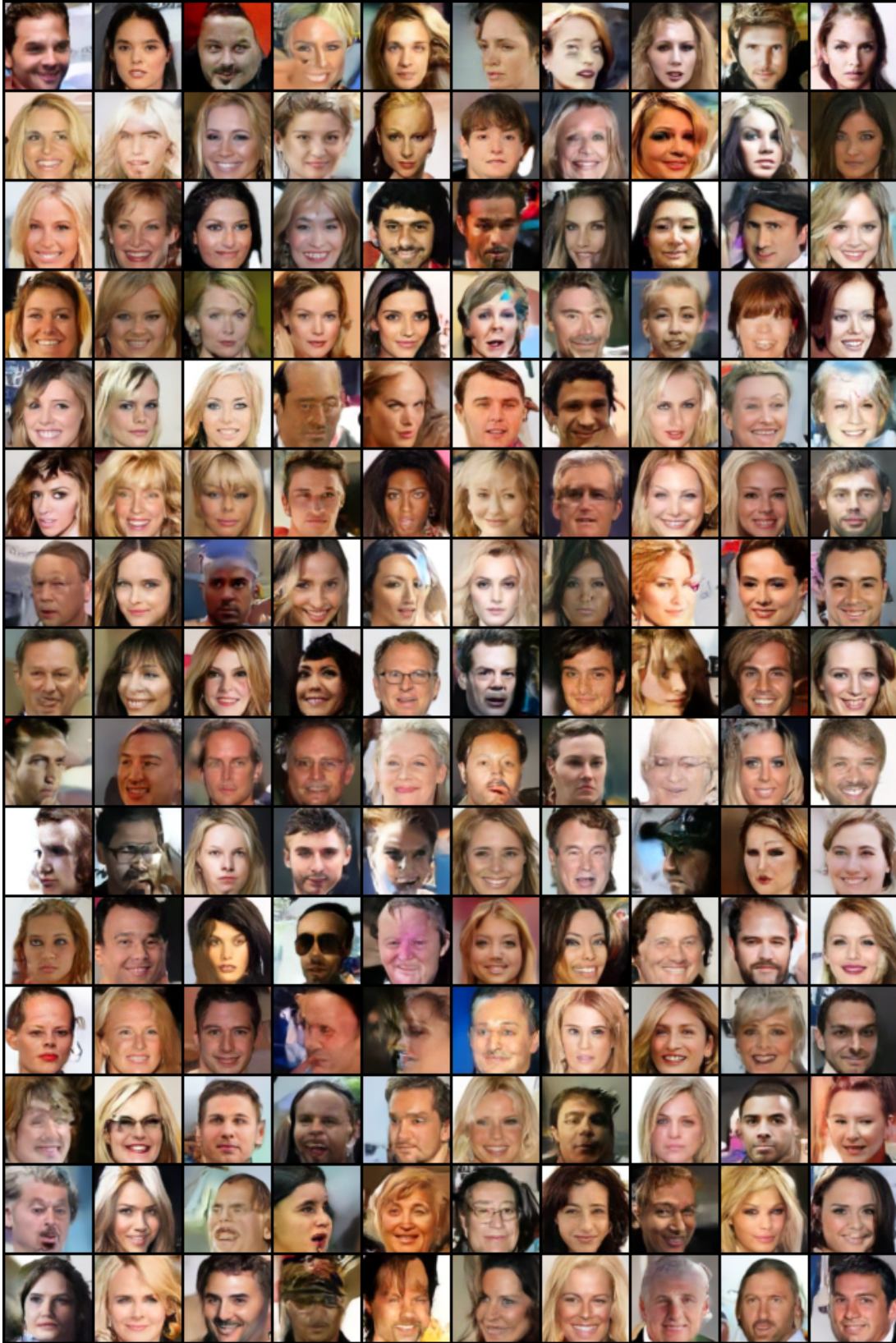


Figure 7: Additional uncured samples on unconditional CelebA 64×64 .

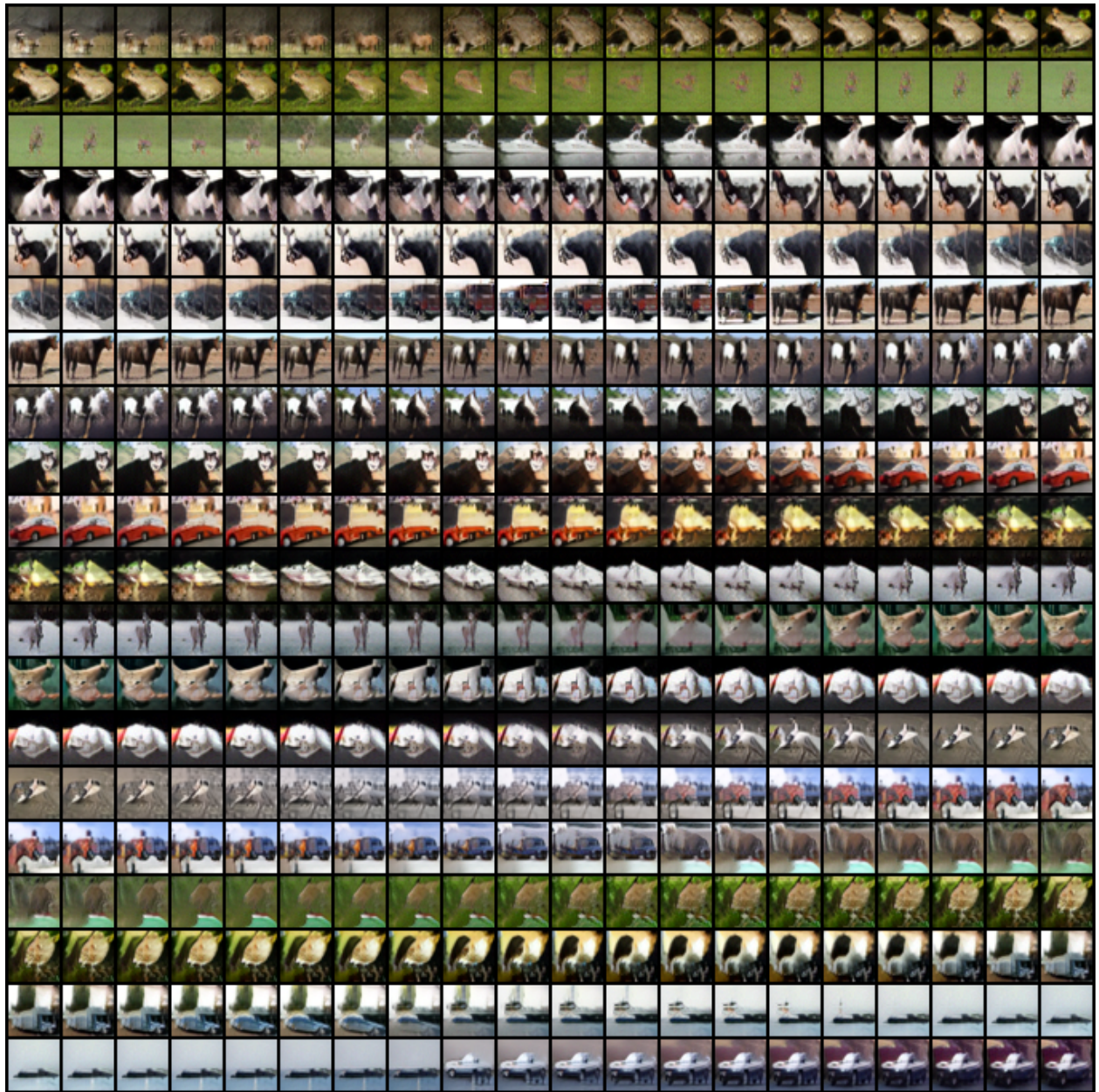


Figure 8: Additional interpolation results on unconditional CIFAR-10 32×32 .



Figure 9: Additional interpolation results on unconditional CelebA 64×64 .

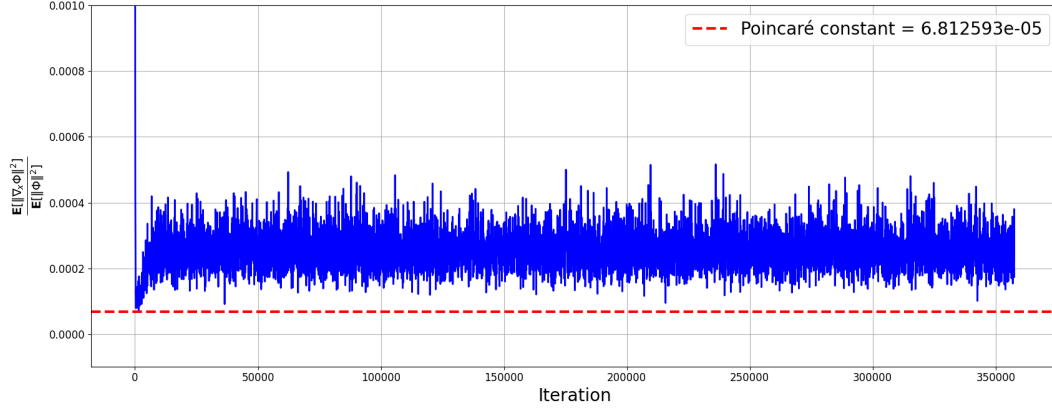


Figure 10: Validation of a Poincaré lower bound using the ratio of the gradient norm to the energy norm on CIFAR-10.

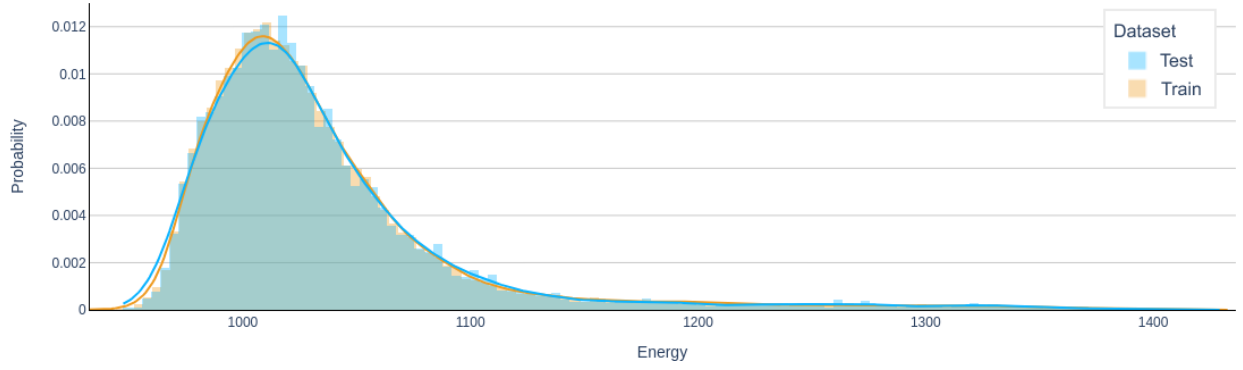


Figure 11: Histogram of the energy-parameterized density estimates for the CIFAR-10 training and test datasets.

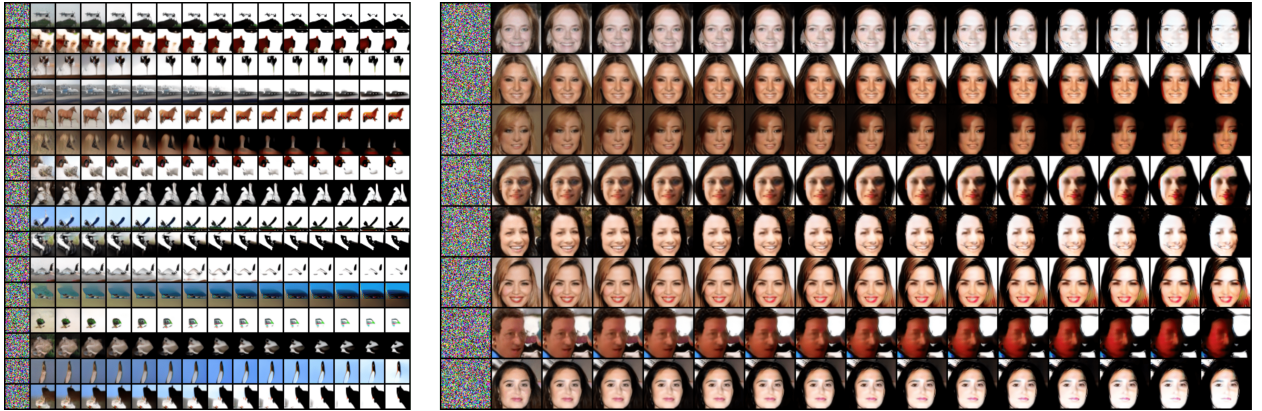


Figure 12: Long-run ODE (RK45) sampling using autonomous potential energy $\Phi(x)$ on CIFAR-10 (left) and CelebA (right).



Figure 13: Long-run ODE (RK45) sampling using time-varying potential energy $\Phi(x, t)$ on CelebA.

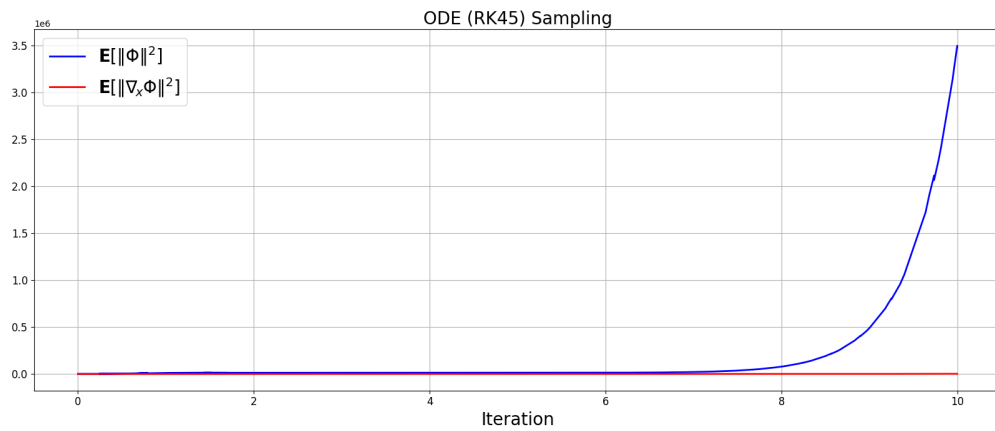


Figure 14: Validation of the convergence of gradient norm and energy norm in long-run ODE (RK45) sampling on CelebA.

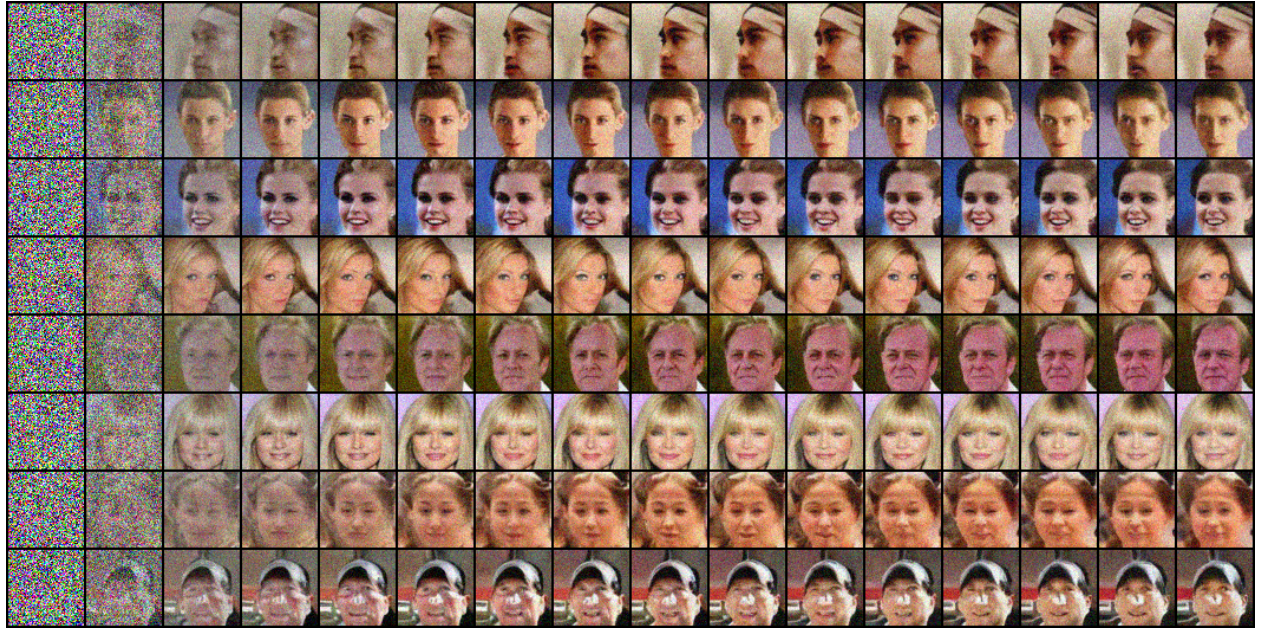


Figure 15: Long-run SGLD sampling using the Boltzmann energy with $\lambda = 0.35$ on CelebA.

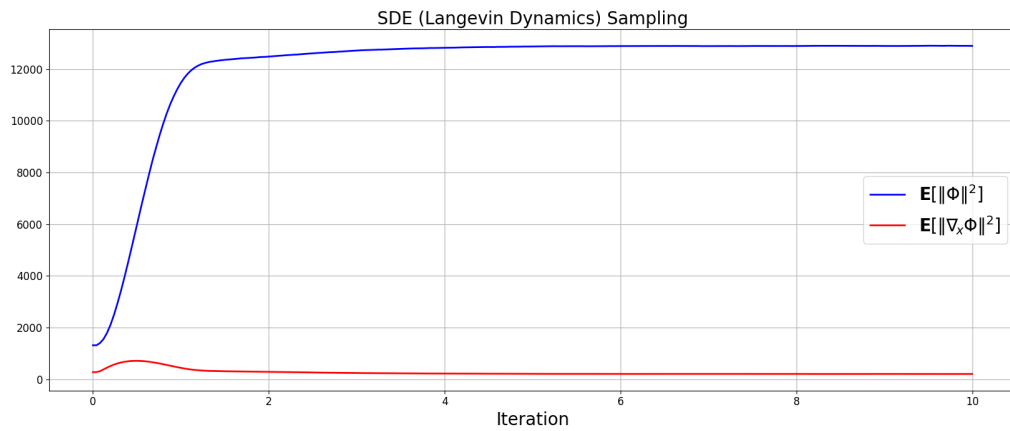


Figure 16: Validation of the convergence of the gradient norm and the energy norm in long-run SGLD sampling on CelebA.

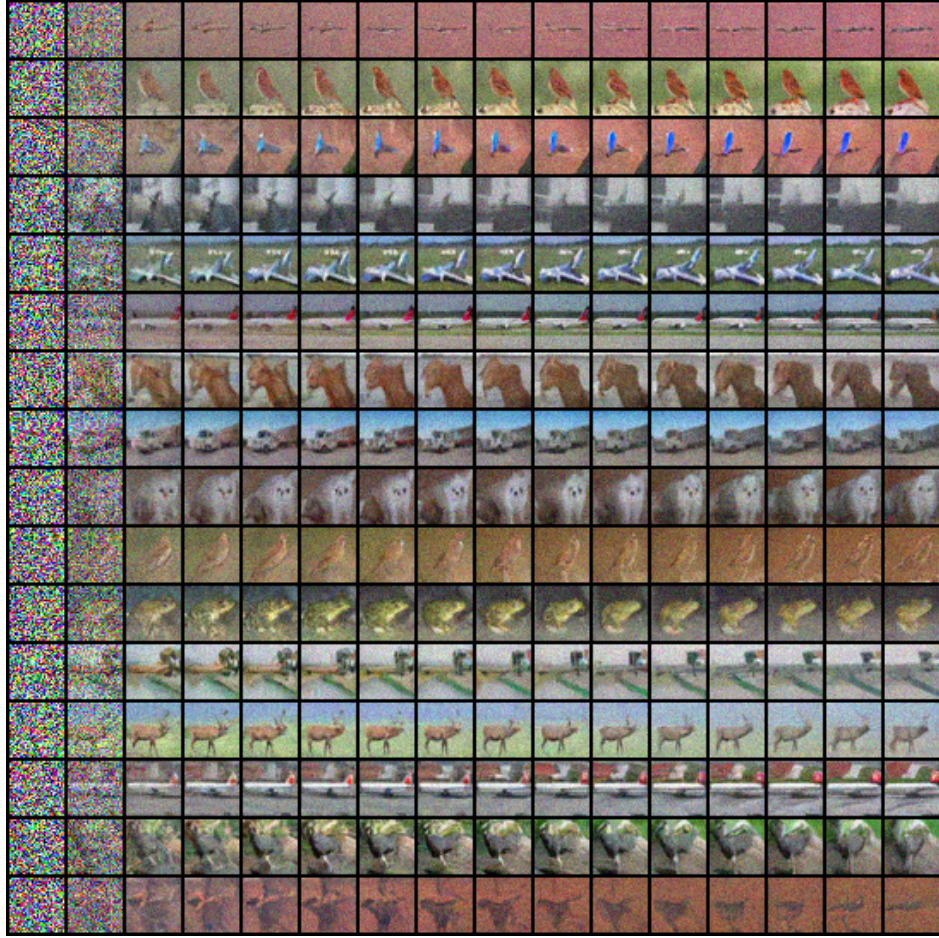


Figure 17: Long-run SGLD sampling using the Boltzmann energy with $\lambda = 0.35$ on CIFAR-10.

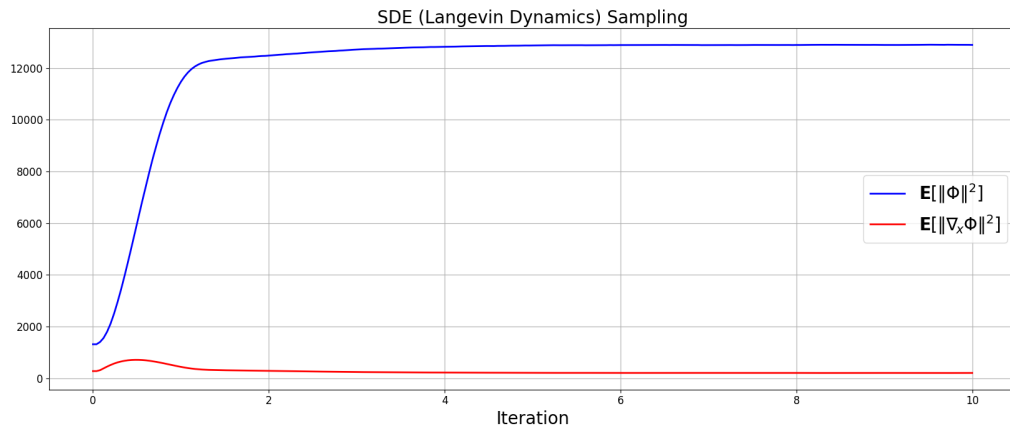


Figure 18: Validation of the convergence of the gradient norm and the energy norm in long-run SGLD sampling on CIFAR-10.

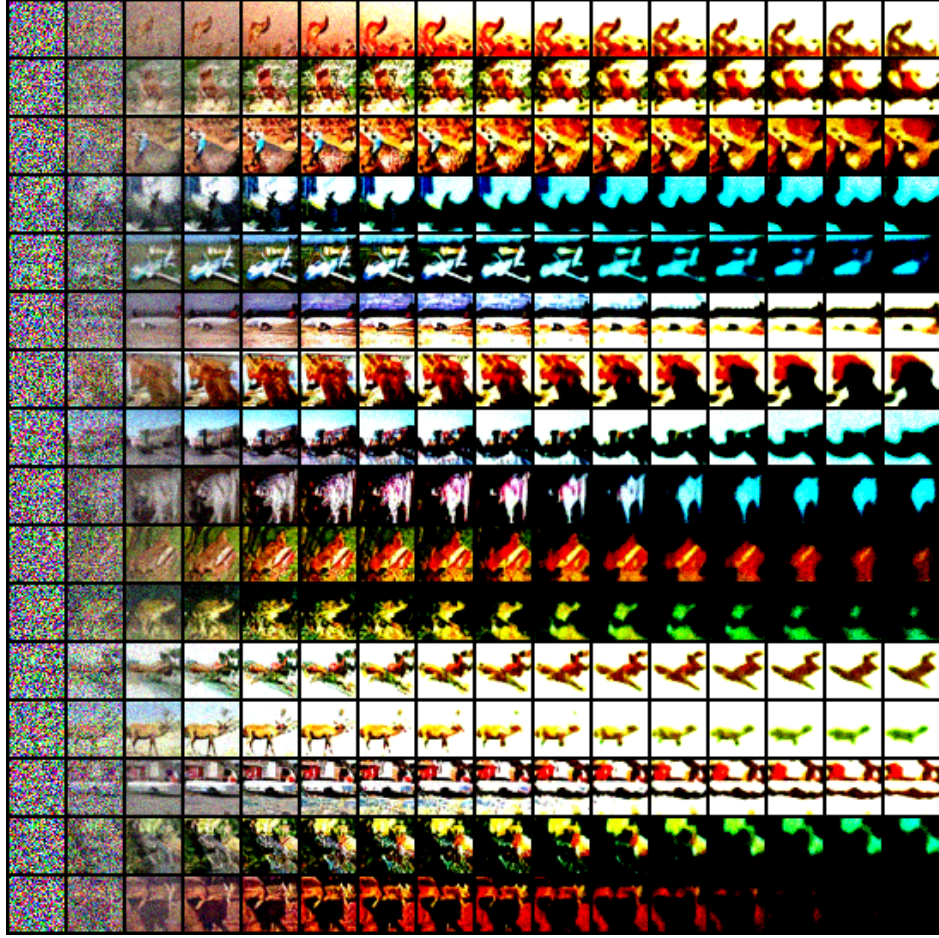


Figure 19: Long-run SGLD sampling using the Boltzmann energy on CIFAR-10 for loss configuration (D).

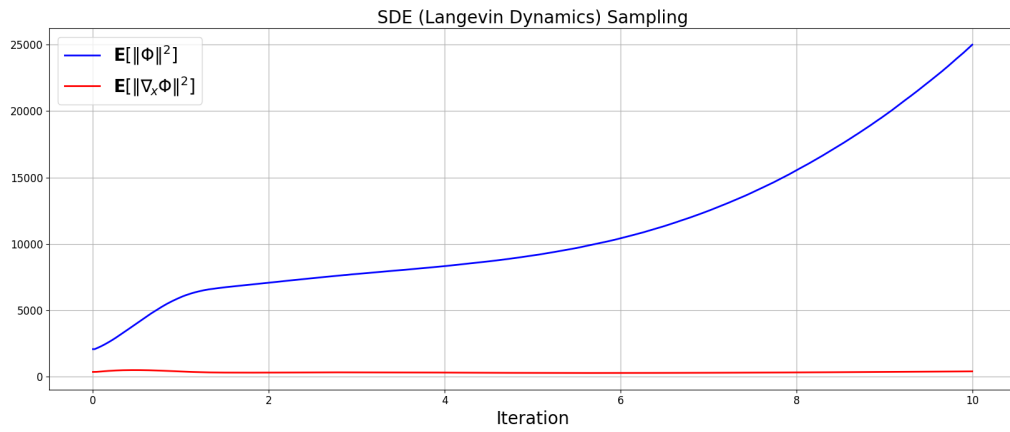


Figure 20: Validation of the convergence of gradient norm and energy norm in long-run SGLD sampling with loss configuration (D) on CIFAR-10.

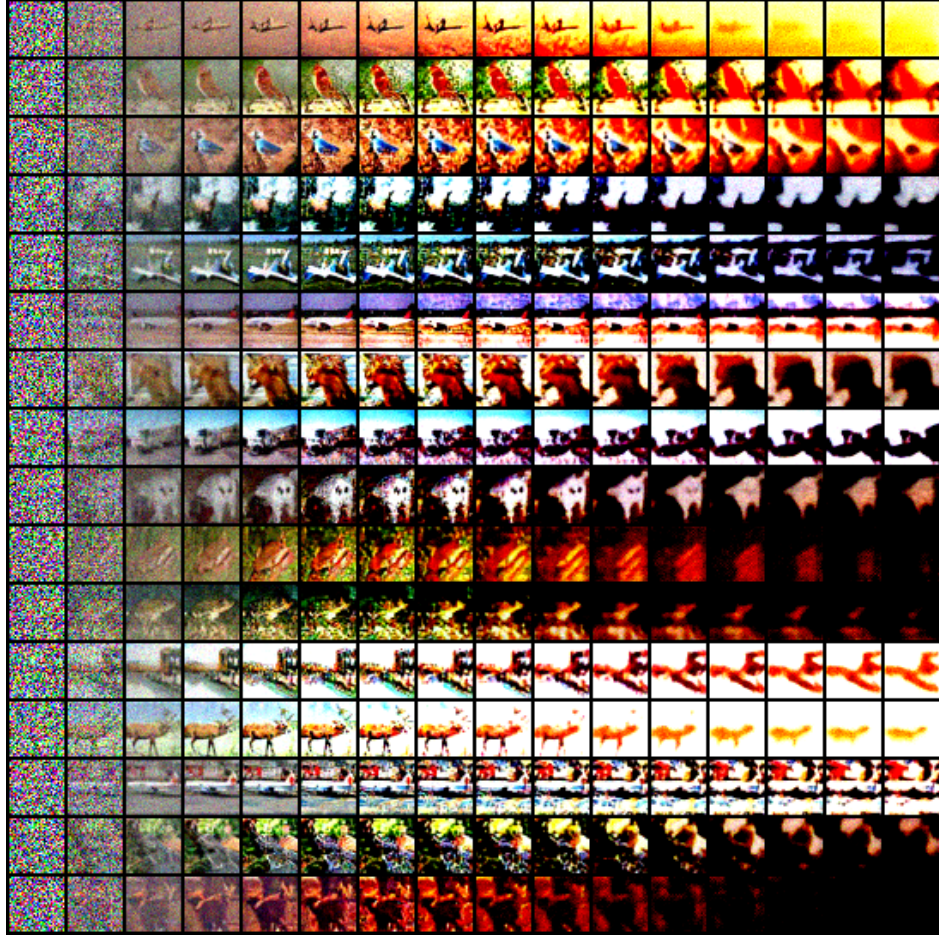


Figure 21: Long-run SGLD sampling using the Boltzmann energy with loss configuration (E) on CIFAR-10.

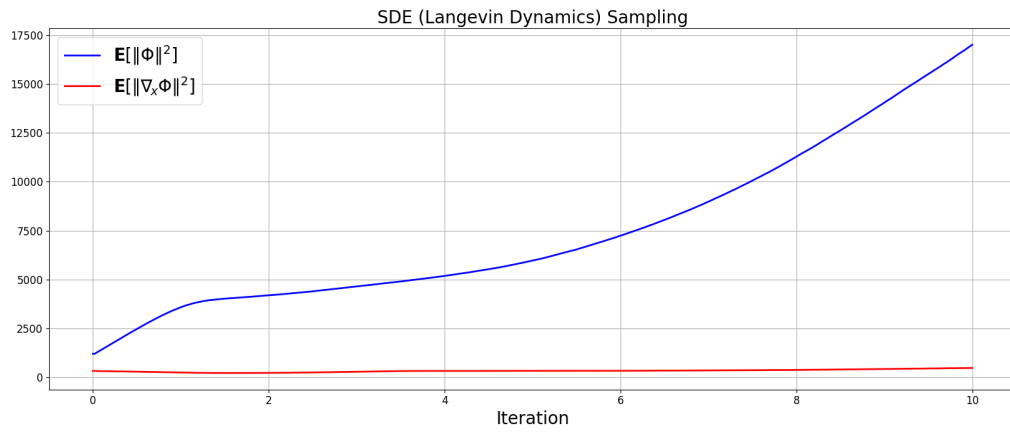


Figure 22: Validation of the convergence of gradient norm and energy norm in long-run SGLD sampling with loss configuration (E) on CIFAR-10.