MONASH University

ICLR

Published in Transactions on Machine Learning Research (TMLR 2025)

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

Junn Yong Loo, Fang Yu Leong, Michelle Adeline,
Julia Kaiwen Lau, Hwa Hui Tew, Arghya Pal,
Vishnu Monn Baskaran, Chee-Ming Ting, Raphaël C.-W. Phan

School of Information Technology, Monash University Malaysia

Introduction and Motivation

① Energy-based generative models

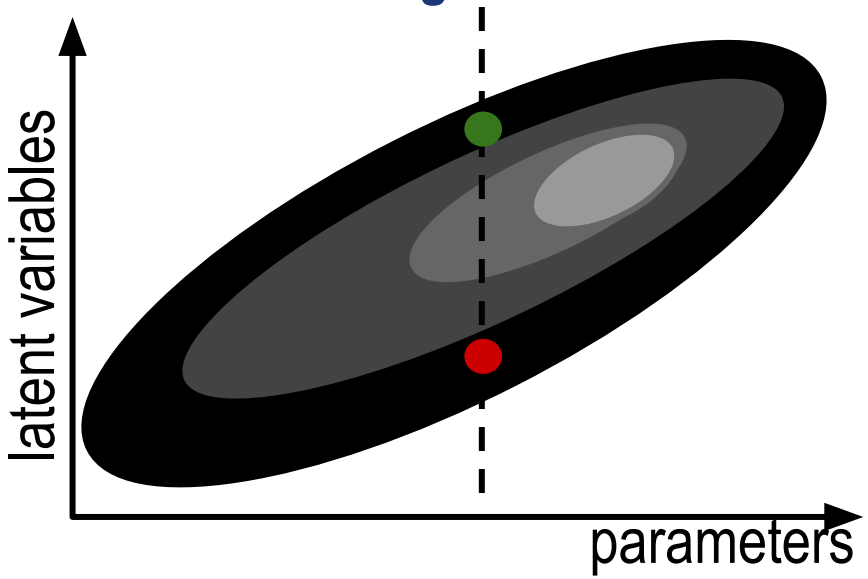
➔ Energy-based models are interpretable as the underlying energy function can be visualized in terms of energy surfaces.
➔ Can flexibly shape its energy surface to match wide range complex data pattern (Flexibility).
➔ EBMs exhibit inherent robustness to Out-of-Distribution (OOD) inputs, given that regions with low likelihood are naturally penalized

② Flow-based generative models

➔ Leveraging continuous normalizing flows and optimal transport techniques to surpass diffusion models in sample quality and efficiency.
➔ Fast sampling (one-shot inference) without use of Markov Chain Monte Carlo (MCMC).

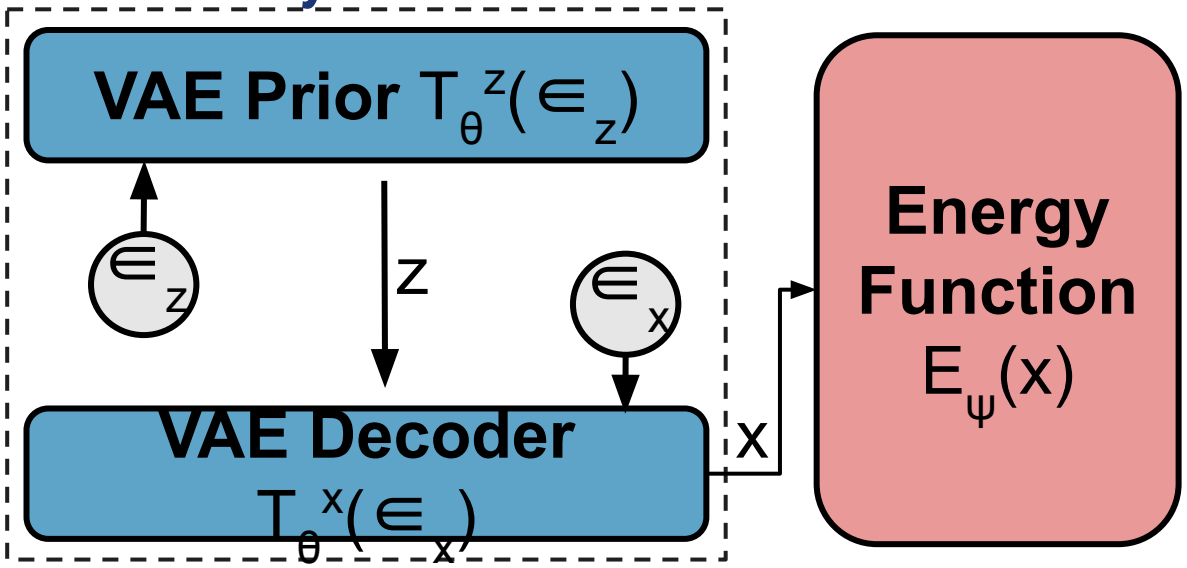
Challenges of Energy-Based Methods

➔ *Poor mixing*



(Grathwohl et al., 2020, Nijkamp et al., 2022, Bond-Taylor et al., 2022)

➔ *Auxiliary model*



(Xie et al., 2020, Grathwohl et al., 2021, Hill et al., 2022)

Key Observations:

1. Often relies on implicit Markov Chain Monte Carlo (MCMC) sampling for contrastive divergence.
2. High-dimensional MCMC suffers from poor and slow mixing that cause biased estimates.
3. Implementation of auxiliary strategies/models that complicates training procedures.

Solution: Variational Potential (VAPO) Flow Bayes

Direct connection between energy-parameterized flow models and explicit (marginal) EBMs remains unexplored.

01 We propose Variational Potential (VAPO) Flow Bayes, a novel energy-based generative framework grounded in variational principles that eliminates the need for auxiliary models and implicit MCMC sampling.

02 We incorporate Deep Ritz method to learn an energy-parameterized potential flow, ensuring alignment between the flow-driven density homotopy and the data-recovery likelihood homotopy.

03 To address the intractability of homotopy matching, we formulate a variational loss function that minimizes the Kullback-Leibler (KL) divergence between these density homotopies.

04 Our results demonstrate that VAPO achieves 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.

Methods

A. Log-homotopy Bayesian transport

Builds a continuous probability path between a Gaussian prior q(x) and the data likelihood by tempering the two densities.

$$\rho(x \mid \bar{x}, t) = \frac{e^{h(x|\bar{x},t)}}{\int_\Omega e^{h(x|\bar{x},t)}\, dx} \qquad h(x \mid \bar{x}, t) = \alpha(t)\, \log q(x) + \beta(t)\, \log p(\bar{x} \mid x)$$

B. Potential flow for homotopy matching

Drives particles with a potential-generated velocity field so that their density ρ_Φ tracks the marginal homotopy $\bar{\rho}$

$$\frac{dx(t)}{dt} = \nabla_x \Phi(x,t) \qquad \frac{\partial \rho_\Phi(x,t)}{\partial t} = -\nabla_x \cdot \Big(\rho_\Phi(x,t)\, \nabla_x \Phi(x,t)\Big) \qquad \frac{\partial \bar{\rho}(x,t)}{\partial t} = -\frac{1}{2}\, \mathbb{E}_{p_{data}(\bar{x})} \Big[\rho(x \mid \bar{x}, t) \Big(\gamma(x,\bar{x},t) - \bar{\gamma}(x,\bar{x},t)\Big)\Big]$$

To achieve homotopy matching, $\rho\Phi \equiv \bar{\rho}$, by aligning their respective time evolutions. This leads to the following PDE, which takes the form of a density-weighted Poisson equation:

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

Methods

C. Links to diffusion & EBMs

When the flow-driven homotopy equals the marginal homotopy, its Fokker–Planck dynamics converge to a stationary Boltzmann distribution with energy.

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

- $\Phi_\infty(x)$ is the learned potential at t = 1.
- The quadratic $\| x \|^2$ term comes from the diffusion coefficient g(t) "freezing" at t=1.

D. Deep-Ritz variational loss

Solving the Poisson equation directly is hard; turn it as a variational problem that SGD can tackle.

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

- Covariance term: Cov[Φ,γ] pushes the vector field $\nabla\Phi$ to match the homotopy's "force" γ(x,$\bar{x}$,t).
- Dirichlet energy E $\| \nabla\Phi \|^2$ smoothes Φ, avoids overly spiky energies.

Experiments: Density Estimation on 2D Data

Fig.1 2D potential flow

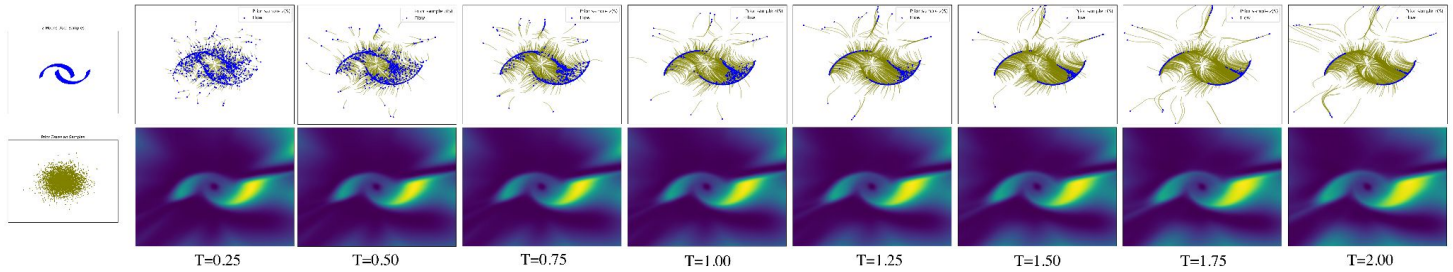
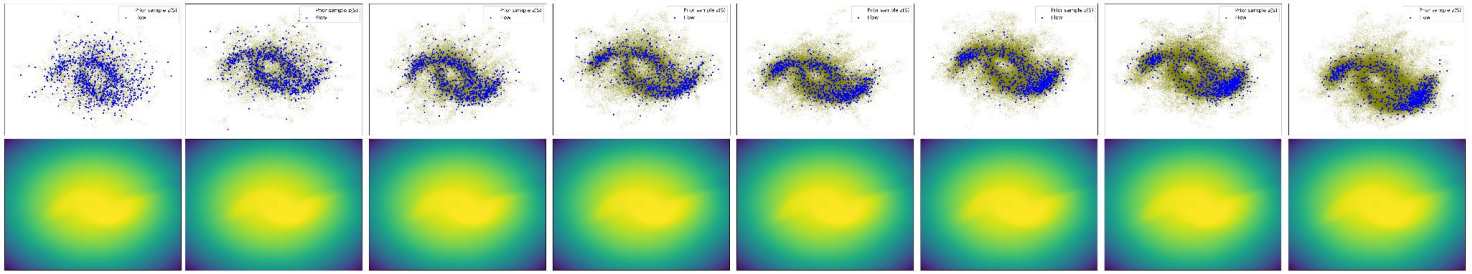


Fig.2 2D Boltzmann density estimation



①To verify the convergence properties of the potential energy and to assess the validity of the Boltzmann energy. We conduct density estimation on 2D synthetic datasets: 2-Moons target distribution.

②*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.

③*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).

④The results indicate that the estimated Boltzmann density closely aligns with the ground-truth and exhibits convergence toward steady-state equilibrium.

Experiments: Unconditional Image Generation

CIFAR-10 (left) and CelebA (right).

FID scores on CIFAR-10 image generation.

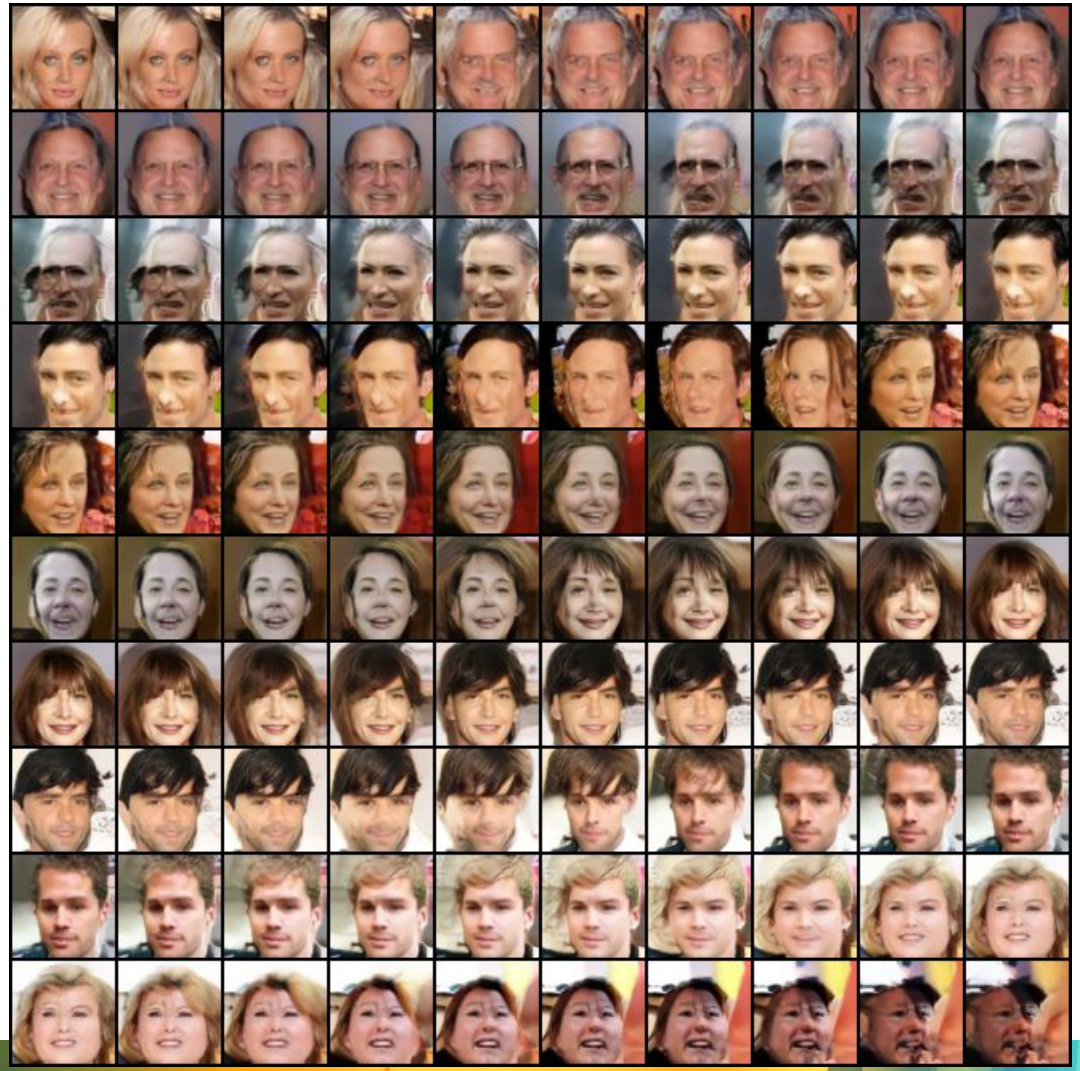
Energy-based Models	FID ↓	Other Likelihood-based Models	FID ↓
IGEBM (Du & Mordatch, 2019)	38.2	ResidualFlow (Chen et al., 2019a)	47.4
EBM-FCE (Gao et al., 2020)	37.3	Glow (Kingma & Dhariwal, 2018)	46.0
CoopVAEBM (Xie et al., 2021b)	36.2	DC-VAE (Parmar et al., 2021)	17.9
CoopNets (Xie et al., 2020)	33.6	**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
GEBM (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
VAPO (Autonomous)	**14.5**	DSBM (Shi et al., 2023)	4.51
VAPO (Time-varying)	**6.72**	PFGM (Xu et al., 2022)	2.35

Figure (left) shows the uncurated and unconditional image samples generated using the time-varying energy model on CIFAR-10 32 × 32 and Celeb A 64 × 64. The generated samples are of decent quality and resemble the original datasets.

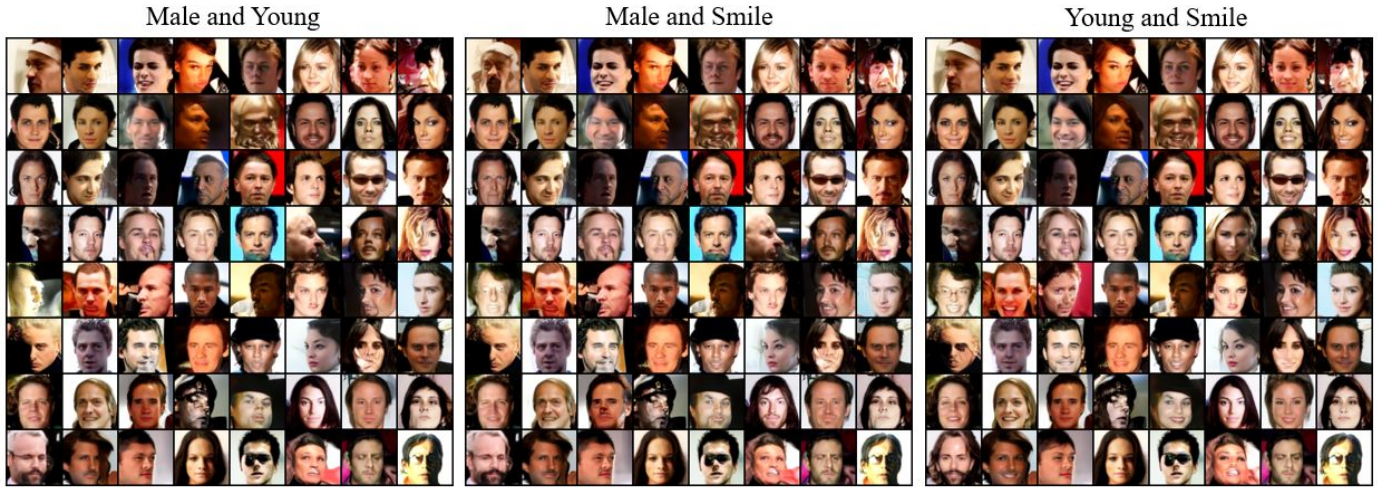
Table (right) summarizes the quantitative evaluations of our framework in terms of FID scores on the CIFAR-10. Our model achieved FID scores competitive to existing generative models.

Image Interpolation and Compositional Generation

Interpolation results on CelebA 64x64.



Compositional Generation



To achieve coherent image interpolation, we perform spherical interpolation between two Gaussian noises and apply ODE sampling to the interpolated noises. Figure (left) shows the interpolation results.

We showcase compositional generation by averaging class-conditioned energies for attribute pairs, results show limited variation across attribute pairs, incorporating composition weights could improve attribute-specific conditioning.

Out-of-Distribution Detection

AUROC scores ↑ 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
VAPO (Ours)	0.78	0.67	0.84	0.61

➔ We evaluate OOD detection performance using the AUROC metric, results show that our model performs exceptionally well on interpolated CIFAR-10 and CelebA 32×32, while performing moderately on CIFAR-100 and SVHN.

Long-Run Steady-State Equilibrium

① ODE Sampling

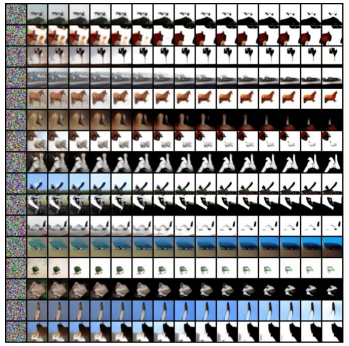


Fig.A

Fig.B

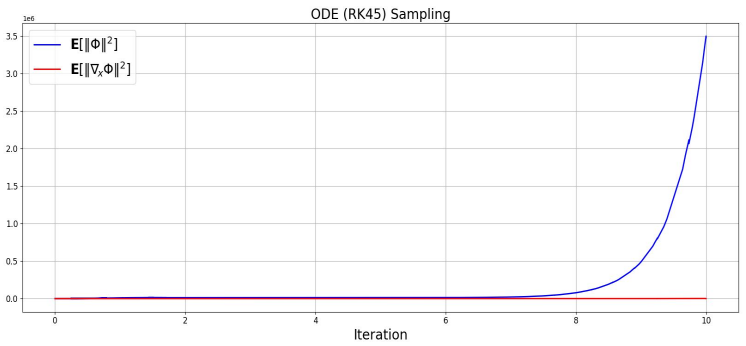


Fig.C

- ➔ Fig.A illustrates long-run ODE sampling over an extended time horizon t ∈ [0, 20] using the autonomous energy model parameterized by WideResNet.
- ➔ Fig.B shows that ODE sampling with the U-Net time-varying energy model degrades image quality over long runs more severely than the autonomous model.
- ➔ Fig.C shows that neither the mean gradient norm E // ∇xΦ // ^{2}, nor the mean energy norm, E // Φ // ^{2}, converge. These findings mirror those from EBMs trained with non-convergent short-run MCMC (Agoritsas et al., 2023; Nijkamp et al., 2020) and highlight the fundamental challenge neural networks face in accurately modeling complex, high-dimensional energy landscapes.

Long-Run Steady-State Equilibrium

② Stochastic Gradient Langevin Dynamics (SGLD) Sampling

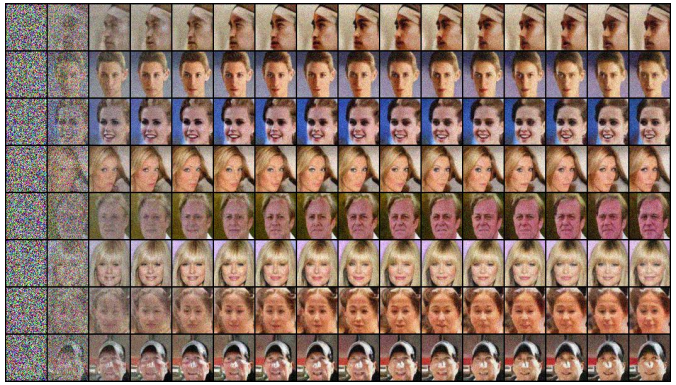


Fig.A

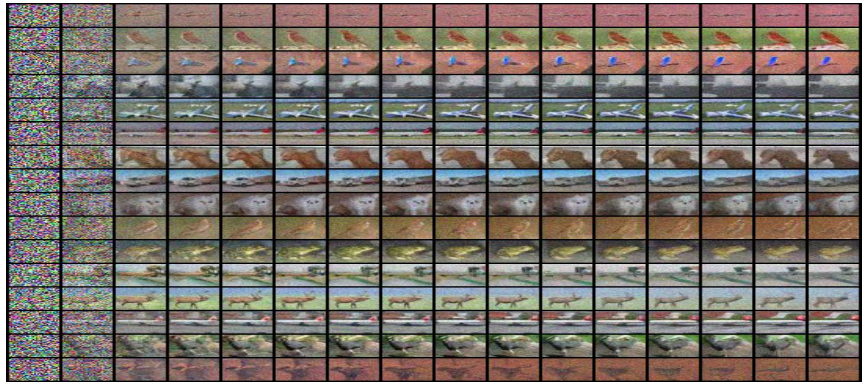


Fig.B

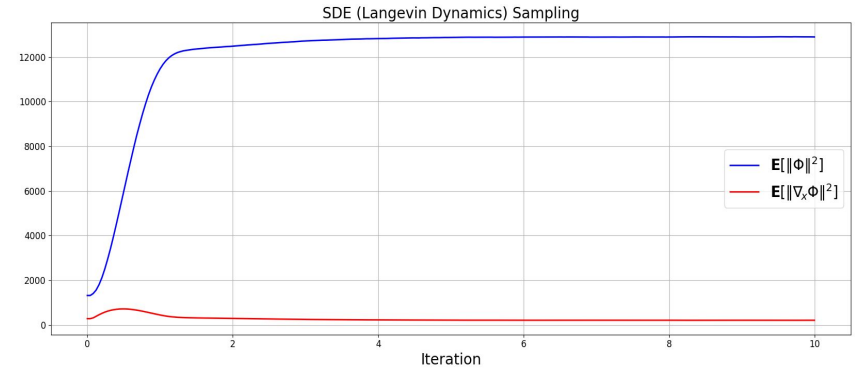


Fig.C

➔ Figures A & B show that SGLD long-run sampling mitigates mode collapse seen under deterministic ODE—the injected Gaussian noise lets trajectories hop out of sharp local minima and preserve background detail. Figure C confirms true equilibrium convergence: the mean gradient norm decays to zero and the mean energy norm plateaus after a brief burn-in, signaling steady-state thermalization under the Boltzmann distribution. Properly tuning the temperature λ balances sampling speed and image quality, yielding asymptotic stability that ODE alone cannot achieve.

Conclusion and Future Work

01 Key Contributions

- ➔ We designed a novel energy-based potential flow framework (VAPO) to reduce the computational cost and instability typically associated with EBM training.
- ➔ VAPO outperform existing EBMs in unconditional image generation and achieved competitive performance in OOD detection.

02 Future Directions

- ➔ Aim to refine the training strategy to improve scalability to higher-resolution images and other data modalities.
- ➔ Incorporate boundary-aware blurring perturbations (Rissanen et al., 2023; Hoogeboom & Salimans, 2023; Daras et al., 2023) to enrich energy landscapes without extra sampling steps.

Thanks for listening

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

Junn Yong Loo, Fang Yu Leong , Michelle Adeline,
Julia Kaiwen Lau , Hwa Hui Tew , Arghya Pal,
Vishnu Monn Baskaran , Chee-Ming Ting , Raphaël C.-W. Phan

School of Information Technology, Monash University Malaysia

Paper: https://arxiv.org/pdf/2504.16262
Code: https://github.com/ljun0004/VAPO