EnfoPath: Energy-Informed Analysis of Generative Trajectories in Flow Matching

Ziyun Li*
KTH Royal Institute of Technology
Stockholm, Sweden
`ziyli@kth.se; liziyun2014@gmail.com`

Ben Dai
Chinese University of hong kong
Hong Kong, China
`ben.dai@cuhk.edu.hk`

Huancheng Hu
Hasso Plattner Institute
Potsdam, Germany
`huancheng.hu@hpi.de`

Henrik Boström
KTH Royal Institute of Technology
Stockholm, Sweden
`bostromh@kth.se`

Soon Hoe Lim
KTH Royal Institute of Technology & Nordita
Stockholm, Sweden
`shlim@kth.se`

Abstract

Flow-based generative models synthesize data by integrating a learned velocity field from a reference distribution to the target data distribution. Prior work has focused on endpoint metrics (e.g., fidelity, likelihood, perceptual quality) while overlooking a deeper question: what do the sampling trajectories reveal? Motivated by classical mechanics, we introduce kinetic path energy (KPE), a simple yet powerful diagnostic that quantifies the total kinetic effort along each generation path of ODE-based samplers. Through comprehensive experiments on CIFAR-10 and ImageNet-256, we uncover two key phenomena: (*i*) higher KPE predicts stronger semantic quality, indicating that semantically richer samples require greater kinetic effort, and (*ii*) higher KPE inversely correlates with data density, with informative samples residing in sparse, low-density regions. Together, these findings reveal that semantically informative samples naturally reside on the sparse frontier of the data distribution, demanding greater generative effort. Our results suggest that trajectory-level analysis offers a physics-inspired and interpretable framework for understanding generation difficulty and sample characteristics.

Code: `https://github.com/ziyunli-2023/EnfoPath-FM`.

1 Introduction

Flow-based generative models [1–5] synthesize high-fidelity images by continuously transforming noise into data. Despite extensive research on architectures [6–8], training objectives [9, 10], and sampling efficiency [11–13], existing work focuses almost exclusively on endpoint metrics [14, 15], such as FID, likelihood, perceptual quality, while ignoring the sampling trajectory itself. Only a few works consider trajectory behavior: AMED-Solver [16] exploits trajectory smoothness to allow very large integration steps, PFDiff [17] uses past and future score information to approximate local trajectory curvature, and FlowGuard [18] monitors velocity and curvature in flow-matching

---

*Corresponding author

trajectories to reject unreliable paths. Yet in deterministic ODE-based samplers [11, 13] of flow matching, each generated sample is obtained as the result of a realization of a path $\{x(t)\}_{t\in[0,1]}$ governed by $\dot{x}(t) = v_\theta(x(t), t)$, where v_θ is the learnt velocity field and $x(0) \sim p_0$ for some reference (noise) distribution p_0 (typically standard Gaussian). This raises a fundamental question: *Can the trajectory itself reveal sample quality, generation difficulty, and data characteristics?*

To explore this, we adopt a trajectory-wise perspective and draw inspiration from classical mechanics [19, 20] and recent path integral approaches to generative modeling [21, 22]. The central quantity of interest is the *kinetic path energy (KPE)*:

$$E = \tfrac{1}{2} \int_0^1 \|v_\theta(x(t), t)\|^2 \, dt,$$

which we also refer to as simply *trajectory energy*. Viewing each sample as a free particle traversing from noise to data, we can interpret E as a measure of the total kinetic cost of this journey. Though not physically grounded in the strict sense, this lens reveals striking patterns. Through comprehensive experiments on CIFAR-10 [23] and ImageNet-256 [24], we have two key findings: *(1) Higher energy, stronger semantics.* Higher-E trajectories consistently produce semantically richer, more class-specific images (CLIP score/margin [25], subsection 3.2). In physics, reaching a precise target requires sustained, directed force, incurring a high kinetic cost. Analogously, generating semantically rich, class-specific images demands that the model maintain high velocity magnitudes along the trajectory, accumulating more energy. *(2) Higher energy, lower density.* High-E samples fall in sparse, low-density regions of the data manifold (subsection 3.3). A particle with higher kinetic energy travels farther from its origin. Analogously, reaching atypical, lower-density regions, distant from the training distribution's center, requires the model to have higher velocities throughout the trajectory, thus higher energy. These findings establish KPE as a dual indicator: it predicts both semantic strength and sample rarity, revealing that high-quality, informative samples naturally inhabit low-density regions.

We summarize our contributions as follows:

- We introduce KPE as a simple yet powerful diagnostic that quantifies the kinetic cost of ODE-based generation process in flow matching through a physics-inspired lens.
- Our empirical results show that: (i) higher KPE predicts higher semantic quality: class-specific samples demand greater kinetic cost to generate, (ii) higher KPE indicates lower data density: rare samples occupy regions that require greater kinetic cost to reach.

2 Kinetic Analogy and Trajectory Energy

2.1 Physical Motivation

In classical mechanics, the evolution of a system is characterized by the action functional [19, 20]

$$S[x(\cdot)] = \int_{t_0}^{t_1} L(x(t), \dot{x}(t), t) \, dt,$$

defined over the time interval $[t_0, t_1]$, where $L(x(t), \dot{x}(t), t) = T(\dot{x}) - V(x(t))$ is the Lagrangian, given by the difference between the kinetic energy $T(\dot{x})$ and the potential energy $V(x)$. This formulation embodies Hamilton's principle of least action [19] and Feynman's path integral framework [20]. For a free particle (i.e., when $V(x) = 0$), the action functional reduces to the kinetic term:

$$S_{\text{free}} = \int_{t_0}^{t_1} \tfrac{1}{2} \|\dot{x}(t)\|^2 \, dt.$$

This kinetic form is a fundamental example of an action functional in physics. Inspired by similar analogies in generative modeling [21, 22], we adopt a kinetic framework to define trajectory-level diagnostics to gain insight into the generation process in flow matching.

2.2 Kinetic Cost Definition

We interpret the flow matching sampling process as a particle moving through a velocity field. The sampling trajectory is governed by the learned velocity field $v_\theta(x, t)$ via the ODE [1, 2, 11, 5]:

$$\frac{dx}{dt} = v_\theta(x(t), t), \quad t \in [0, 1], \quad x(0) \sim \mathcal{N}(0, I),$$

which describes the probability flow [26] from noise to data distribution. The trajectory represents the path of a particle driven by the velocity field. Inspired by classical mechanics [19], we define *kinetic path energy E* as:

$$E := \tfrac{1}{2} \int_0^1 \|v_\theta(x(t), t)\|^2 \, dt,$$

where we adopt the convention of *unit mass (m = 1)*, standard in the free particle action formulation.

Physical interpretation. The quantity E encapsulates the cumulative kinetic cost incurred during the sampling process. Conceptually, it quantifies the "energy" required to transport a sample from the noise distribution to the data manifold. A higher E signifies that the model employs greater velocity magnitudes on average, reflecting a more energetically demanding generation process.

Important clarification. We emphasize that E is not strictly a physical energy, but rather a *kinetic-inspired metric* designed to quantify trajectory complexity. This metric is straightforward to compute: it only requires recording the velocity field magnitudes during standard ODE integration, incurring negligible computational overhead. Moreover, the expected KPE matches the Benamou–Brenier dynamic formulation [27] of the 2-Wasserstein transport cost when the learned flow realizes the optimal transport [3, 10], grounding our diagnostic in optimal transport theory [28] rather than mere physical analogy.

3 Experiments

3.1 Experimental Setup

To explore the properties of the sampling trajectory, we evaluate on CIFAR-10 [23] using a pretrained OTCFM model [3], and on ImageNet-256 [24] using two pretrained models: DiT-B [6] (shortcut [6]) and SiT-XL/2 [7] with Stable Diffusion VAE for latent encoding.

3.2 KPE E vs. Semantic Strength

Setup and Metrics. We generate 5K ImageNet-256 samples using the forward Euler sampler with Classifier-Free Guidance (CFG) [9] scales of 1.0, 1.5, and 4.0. CFG is a common technique in diffusion and flow-based models that interpolates between unconditional and class-conditional generation to control the strength of semantic conditioning—larger scales typically yield sharper, more class-consistent images at the cost of diversity. While our initial setup considers trajectories from the unconditional vector field $v_\theta(x, t)$, we also examine CFG sampling, where a constructed mixture field amplifies conditional guidance – allowing us to study how stronger semantic alignment influences KPE and sample quality. Trajectories are grouped into low (0–33%), mid (33–67%), and high (67–100%) energy bins, and evaluated using E, CLIP score, and CLIP margin [25]. **CLIP score** quantifies semantic alignment as the maximum cosine similarity (scaled by 100) between normalized image and text features for the true class (e.g., "a photo of a golden retriever"). **CLIP margin** measures semantic discriminability as the difference between the similarity to the true class and the highest similarity to competing classes: $\text{Margin} = \text{Sim}_{\text{true}} - \max_{c \in \mathcal{C}_{\text{others}}} \text{Sim}(c)$, where $\mathcal{C}_{\text{others}}$ represents competing classes. Higher margins indicate stronger class-specific semantics.

Results. Figure 1 and Figure 2 show that both CLIP score and CLIP margin increase with E across different CFG settings. For instance, at CFG=1.5, the median CLIP score increases from 23.52 (low-energy) to 25.12 (high-energy), and the median CLIP margin rises from 7.05 to 10.02. Additionally, Table 1 and Table 2 compare low-energy and high-energy groups. Independent two-sample t-tests ($p < 0.01$) confirm statistical significance, with large effect sizes (Cohen's d) emphasizing practical relevance. These findings demonstrate that higher-energy trajectories produce images with clearer and more accurate semantics. More visualization examples are provided in the Appendix.

Conclusion. *Higher E consistently correlates with stronger semantic alignment and discriminability.*
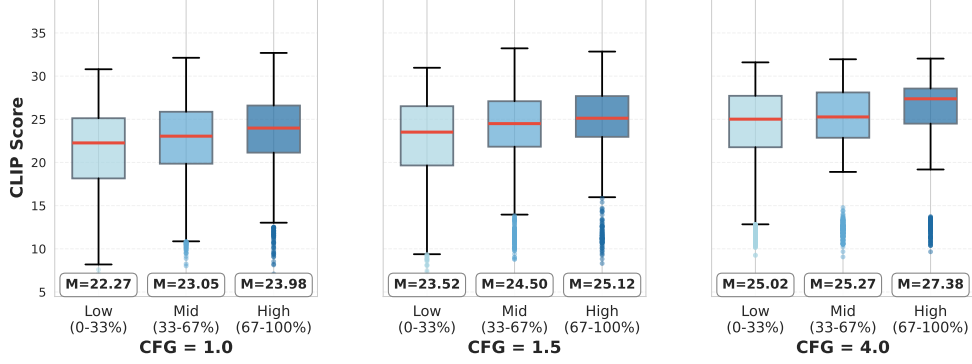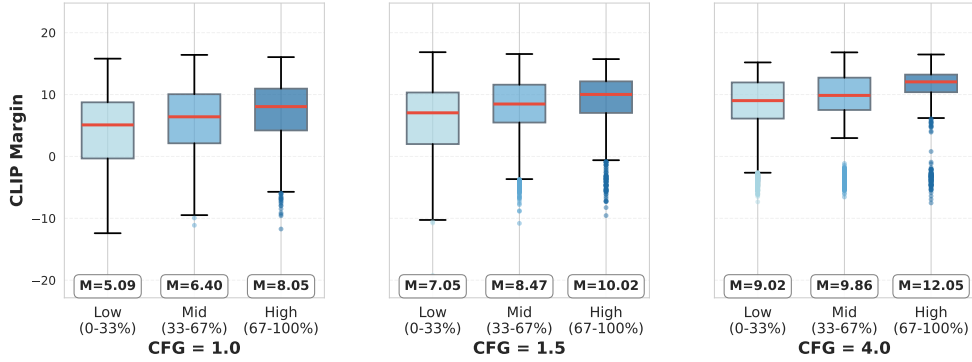
Figure 1: Trajectory energy (E) positively correlates with CLIP score across CFG settings.

Table 1: CLIP Score: Energy Group Comparison across CFG Scales (High vs. Low)

CFG Scale	Low Energy $\mu \pm \sigma$	High Energy $\mu \pm \sigma$	$\Delta\mu$	t-statistic	p-value	Cohen's d
1.0	21.22 ± 5.35	23.43 ± 4.44	$+2.21$	11.55	3.75×10^{-30}	0.450
1.5	21.87 ± 5.99	24.62 ± 4.29	$+2.75$	13.54	2.02×10^{-40}	0.527
4.0	23.23 ± 5.89	25.87 ± 4.39	$+2.64$	13.06	7.99×10^{-38}	0.509

Notes: All comparisons are statistically significant at $p < 0.001$. Cohen's d values indicate medium to large effect sizes. Low energy: 0–33% percentile; High energy: 67–100% percentile. $n = 1320$ per group.

Figure 2: Trajectory energy (E) positively correlates with CLIP margin across CFG settings.

Table 2: CLIP Margin: Energy Group Comparison across CFG Scales (High vs. Low)

CFG Scale	Low Energy $\mu \pm \sigma$	High Energy $\mu \pm \sigma$	$\Delta\mu$	t-statistic	p-value	Cohen's d
1.0	4.15 ± 5.99	7.09 ± 5.00	$+2.94$	13.70	2.42×10^{-41}	0.534
1.5	5.66 ± 6.17	8.93 ± 4.54	$+3.27$	15.48	8.40×10^{-52}	0.603
4.0	7.44 ± 5.95	10.82 ± 4.40	$+3.38$	16.59	7.21×10^{-59}	0.646

Notes: All comparisons are statistically significant at $p < 0.001$. Cohen's d values indicate medium to large effect sizes. Low energy: 0–33% percentile; High energy: 67–100% percentile. $n = 1320$ per group.

3.3 KPE E vs. Data Density

Setup and Metrics. We analyze the correlation between E and estimated sample densities using the OTCFM [3] (on CIFAR-10) and Shortcut [29] (on ImageNet256). We generate 2,000 samples via

4

Euler solver ($N = 10, 50, 150$ steps) and extract *low-dimensional* image features, including channel statistics, texture, and edge descriptors, resulting in 22 features per image. PCA reduces these to 2D embeddings (explained variance $> 85\%$). Density is estimated using k-NN ($k = 50$) and Gaussian KDE with Scott's bandwidth. We use Spearman's ρ to quantify monotonic correlation, and Cliff's δ to measure effect size between high-energy (top 20%) and low-energy (bottom 20%) groups, validated using the Mann-Whitney U test on CIFAR10 data.

Qualitative Results. We observe *a negative correlation between E and data density*. Figure 3a shows 3D visualizations of $\log(\text{density})$ (left) and E (right). The surfaces exhibit opposite shapes: high-density areas align with low energy, and low-density areas align with high energy. In Figure 3b, the top 10% highest-energy generated samples are projected onto the training density surface, consistently falling in low-density regions. This empirically illustrates the inverse energy–density relationship.

Quantitative Results. Table 3 shows consistent negative correlations between trajectory energy and data density. For CIFAR-10, Spearman's ρ drops from -0.54 to -0.65 (k-NN) and -0.54 to -0.64 (KDE) as N increases, with similar trends for Cliff's δ and ImageNet-256. Figure 4 shows a scatter plot of E vs. k-NN's log-density, with blue dots as samples and a red trend line. More visualizations and detailed statistics are provided in the Appendix.

Conclusion. E inversely correlates with the estimated data densities, suggesting that high-energy trajectories tend to concentrate in sparse regions of the data manifold.

(a) KPE vs. density KDE surface   (b) High-energy region highlights
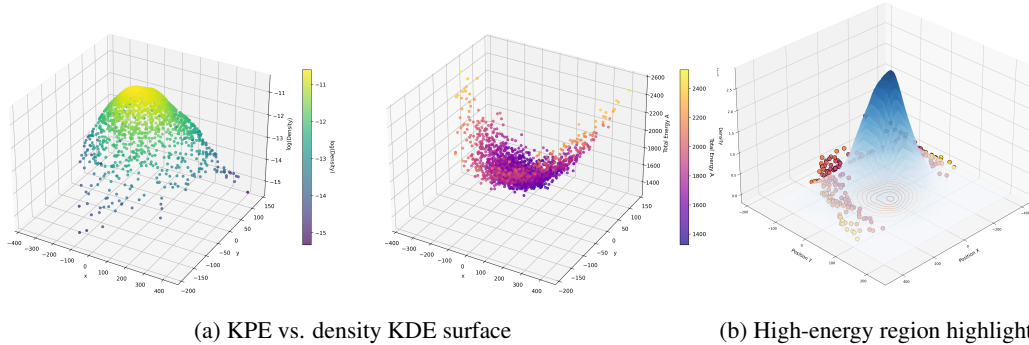
Figure 3: (a) 3D surfaces of $\log(\text{density})$ (left) and trajectory energy E (right) (CIFAR-10, 150 steps) show high-density regions align with low energy. (b) Top 10% highest-energy samples consistently fall in low-density regions, confirming the inverse energy-density correlation.

Metric		CIFAR-10			ImageNet-256		
		$N{=}10$	$N{=}50$	$N{=}150$	$N{=}10$	$N{=}50$	$N{=}150$
$\rho \downarrow$	k-NN	-0.54	-0.61	-0.65	-0.38	-0.42	-0.38
$\delta \downarrow$		-0.83	-0.89	-0.93	-0.55	-0.58	-0.55
$\rho \downarrow$	KDE	-0.54	-0.61	-0.64	-0.31	-0.33	-0.31
$\delta \downarrow$		-0.82	-0.88	-0.92	-0.43	-0.47	-0.43

Table 3: Correlation metrics of k-NN and KDE methods on CIFAR-10 and ImageNet-256 for different N.
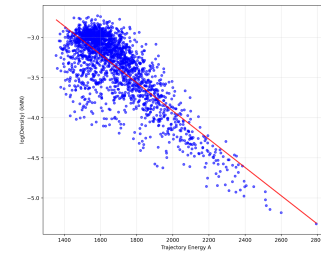
Figure 4: E vs. k-NN log-density ($\rho = -0.6314$, $p = 6.56 \times 10^{-223}$).

4 Conclusion

We introduced Kinetic Path Energy (KPE), a physics-inspired diagnostic measuring the kinetic cost of flow-based generation. Our experiments reveal that higher KPE correlates with stronger semantic quality and lower data density, showing that informative samples naturally lie in sparse regions. This trajectory-level view offers interpretable insights beyond traditional endpoint-based evaluation, and motivates future theoretical work on the connection between KPE and generative dynamics, as well as extensions to stochastic samplers through the lens of stochastic thermodynamics [30, 31].

Acknowledgments

References

[1] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

[2] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[3] Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, pages 1–34, 2024.

[4] Michael S Albergo and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023.

[5] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in neural information processing systems*, volume 31, 2018.

[6] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4195–4205, 2023.

[7] Nanye Ma, Mark Goldstein, Michael S Albergo, Nicholas M Boffi, Eric Vanden-Eijnden, and Saining Xie. Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. In *European Conference on Computer Vision*, pages 23–40. Springer, 2024.

[8] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[9] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

[10] Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky Chen. Multisample flow matching: Straightening flows with minibatch couplings. *International Conference on Machine Learning*, 2023.

[11] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

[12] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022.

[13] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, 35: 26565–26577, 2022.

[14] 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, 2017.

[15] Gaurav Parmar, Richard Zhang, and Jun-Yan Zhu. Aliased image inception: The importance of alignment when fine-tuning inception. *arXiv preprint arXiv:2203.06490*, 2022.

[16] Zhenyu Zhou, Defang Chen, Can Wang, and Chun Chen. Fast ode-based sampling for diffusion models in around 5 steps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[17] Guangyi Wang, Yuren Cai, Lijiang Li, Wei Peng, and Songzhi Su. Pfdiff: Training-free acceleration of diffusion models combining past and future scores. In *The Thirteenth International Conference on Learning Representations*, 2025.

[18] Ziyun Li and Henrik Boström. Flowguard: Guarding flow matching via conformal sampling. *Proceedings of Machine Learning Research*, 2025.

[19] Herbert Goldstein, Charles Poole, and John Safko. *Classical Mechanics*. Addison Wesley, 3rd edition, 2002. ISBN 0-201-65702-3.

[20] Richard P Feynman and Albert R Hibbs. *Quantum Mechanics and Path Integrals*. McGraw-Hill, New York, 1965.

[21] Yuji Hirono, Akinori Aoki, Kenji Kusakabe, and Yuichi Nakai. Understanding diffusion models by feynman's path integral. *arXiv preprint arXiv:2403.11262*, 2024.

[22] Qinsheng Zhang and Yongxin Chen. Path integral sampler: a stochastic control approach for sampling. *International Conference on Learning Representations*, 2023.

[23] Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[24] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. IEEE, 2009.

[25] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, 2021.

[26] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2021.

[27] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[28] Cédric Villani. *Optimal Transport: Old and New*. Springer, 2009. ISBN 978-3-540-71050-9.

[29] Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. In *International Conference on Learning Representations*, 2025.

[30] Udo Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Reports on Progress in Physics*, 75(12):126001, 2012.

[31] Kotaro Ikeda, Tomoya Uda, Daisuke Okanohara, and Sosuke Ito. Speed-accuracy relations for diffusion models: Wisdom from nonequilibrium thermodynamics and optimal transport. *Physical Review X*, 15(3):031031, 2025.

5 Appendix

5.1 Additional Experiments on Kinetic Path Energy (KPE) vs. Semantic Strength

Qualitative Examples We provide qualitative visual comparisons across diverse ImageNet-256 classes to demonstrate the consistent semantic quality differences between high-energy and low-energy trajectories.
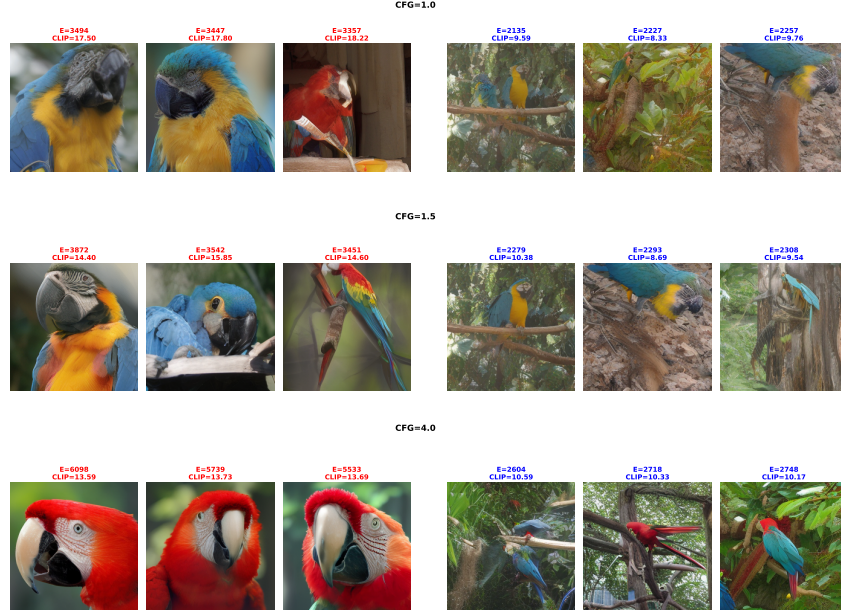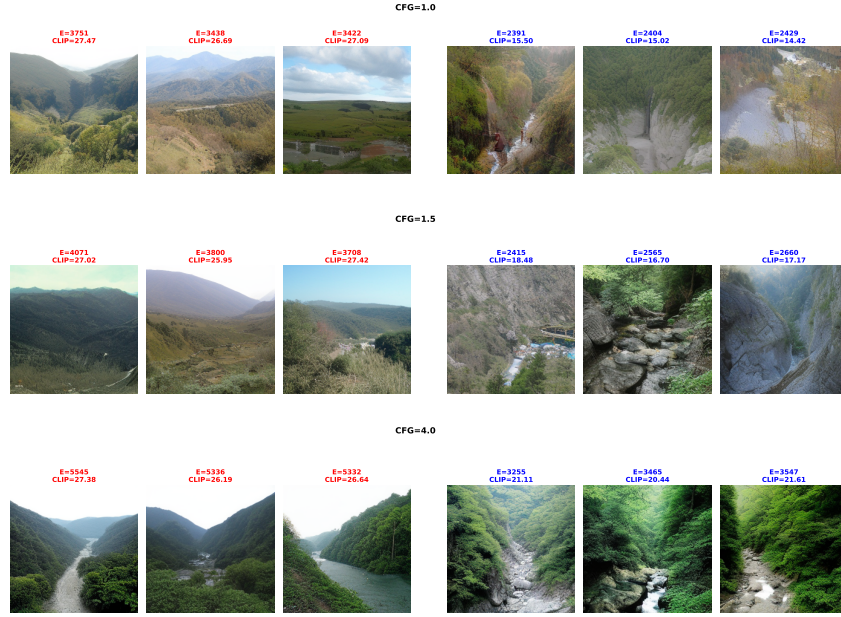

Figure 5: **Macaw** (ImageNet-256): High-KPE (left) vs. low-KPE (right) across CFG scales 1.0, 1.5, 4.0. Higher KPE yields richer semantic details, vibrant colors, and sharper textures.
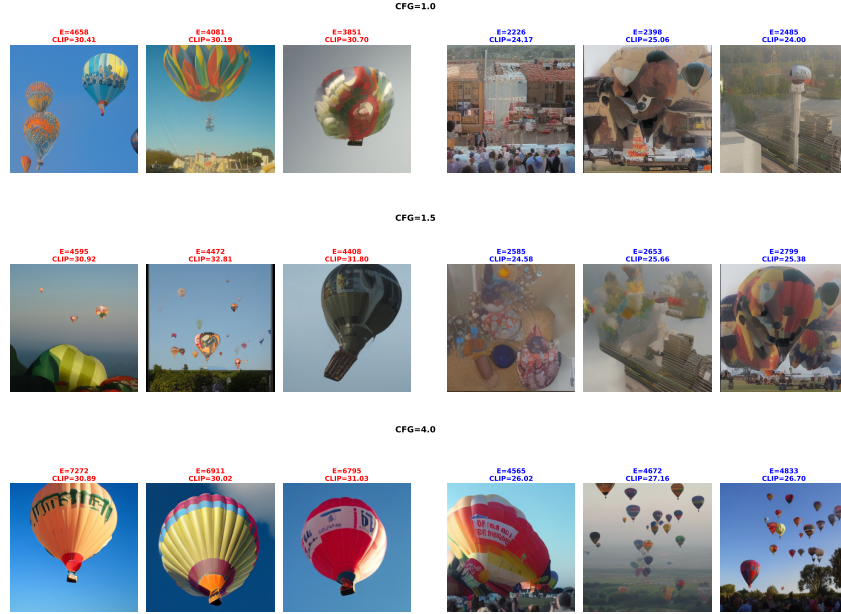
Figure 6: **Hot Air Balloon** (ImageNet-256): High-KPE (left) vs. low-KPE (right) across CFG scales 1.0, 1.5, 4.0. Higher KPE shows clearer structures and better color saturation.

Figure 7: **Golden Retriever** (ImageNet-256): High-KPE (left) vs. low-KPE (right) across CFG scales 1.0, 1.5, 4.0. Higher KPE produces finer textures and clearer facial features.

Figure 8: **African Elephant** (ImageNet-256): High-KPE (left) vs. low-KPE (right) across CFG scales 1.0, 1.5, 4.0. Higher KPE shows more defined features and better skin texture.

Figure 9: **Valley** (ImageNet-256): High-KPE (left) vs. low-KPE (right) across CFG scales 1.0, 1.5, 4.0. Higher KPE generates more detailed terrain and better depth perception.
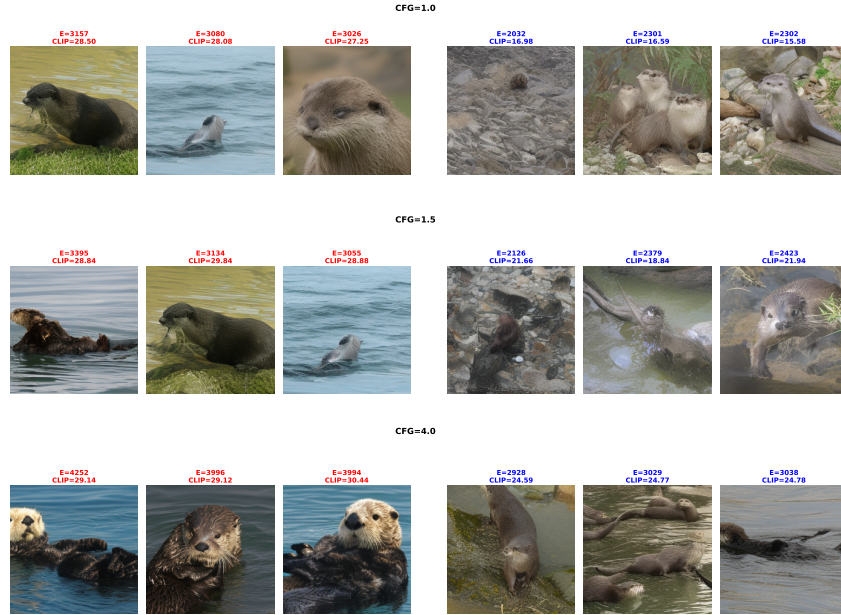
Figure 10: **Otter** (ImageNet-256): High-KPE (left) vs. low-KPE (right) across CFG scales 1.0, 1.5, 4.0. Higher KPE shows sharper outlines and more realistic details.

5.2 Additional Experiments on Kinetic Path Energy (KPE) vs. Data Density

This section provides comprehensive evidence supporting our second key finding: higher KPE strongly correlates with lower data density. We present correlation analysis, effect size quantification, spatial distribution patterns, and robustness validation across different sampling configurations.

(a) 10 sampling steps

(b) 50 sampling steps

(c) 100 sampling steps

(d) 150 sampling steps
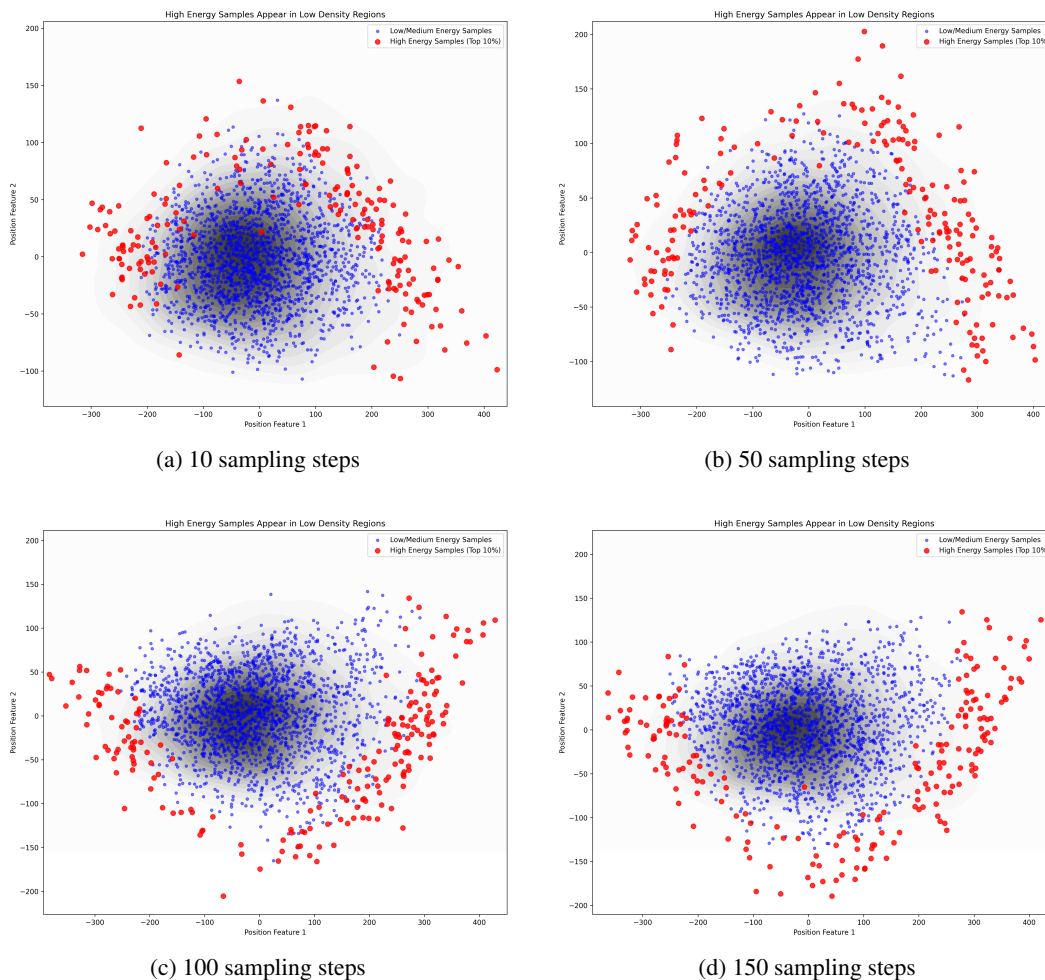
Figure 11: 2D visualization of kinetic path energy (KPE) and data density. Red points represent the top 10% highest-energy samples, which consistently fall in low-density regions. Blue points represent low/medium-energy samples. Contours indicate density levels, highlighting the inverse correlation between KPE and density.
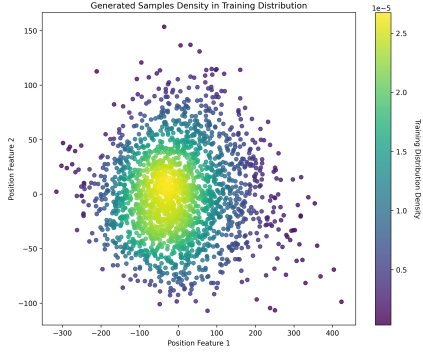
(a) 10 steps - Density distribution in 2D
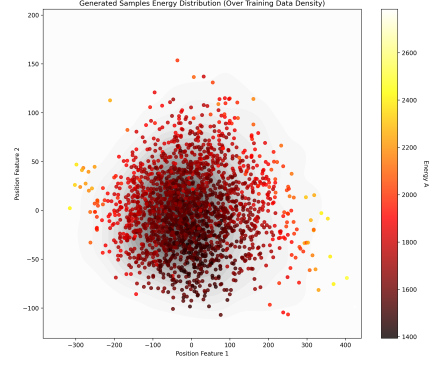
(b) 10 steps - KPE distribution in 2D
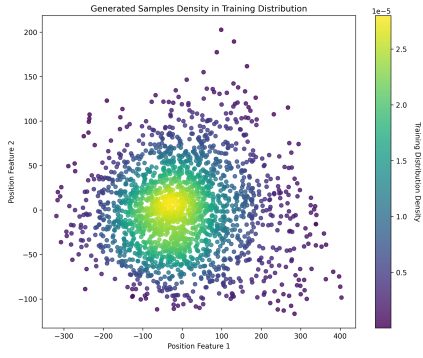
(c) 50 steps - Density distribution in 2D
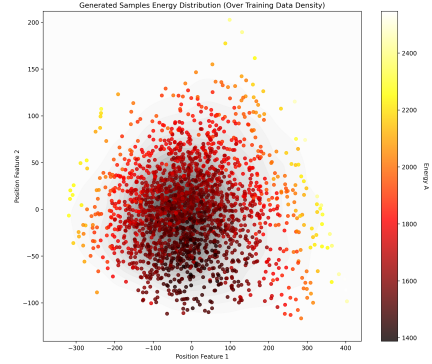
(d) 50 steps - KPE distribution in 2D
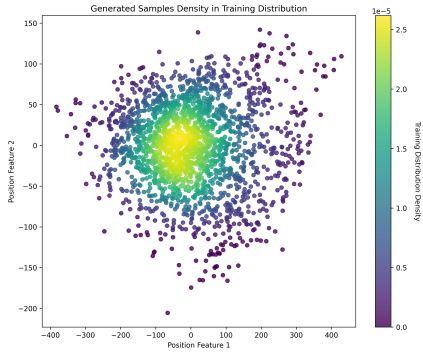
(e) 100 steps - Density distribution in 2D
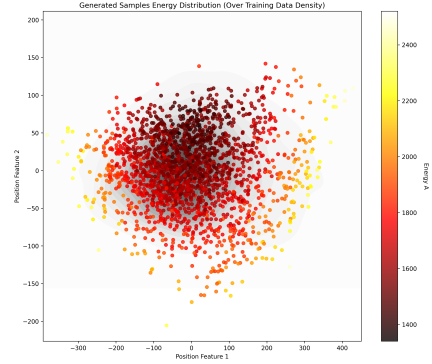
(f) 100 steps - KPE distribution in 2D
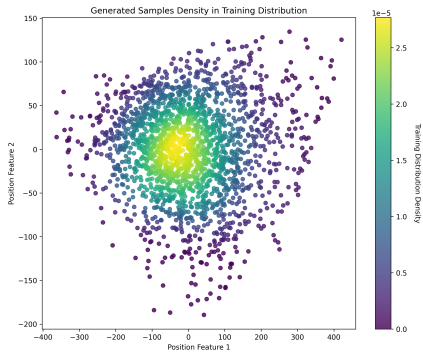
(g) 150 steps - Density distribution in 2D
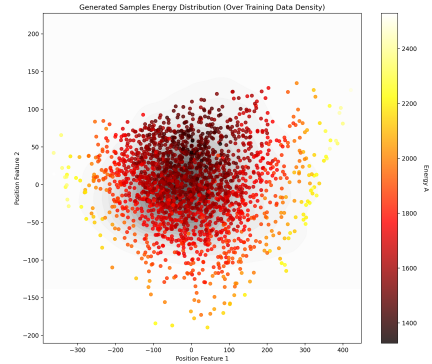
(h) 150 steps - KPE distribution in 2D

Figure 12: 2D visualizations. Left column shows the distribution-density visualization of generated samples projected to 2D using PCA. Right column presents the corresponding kinetic path energy (KPE) in 2D across different sampling steps (10, 50, 100, 150 rows).
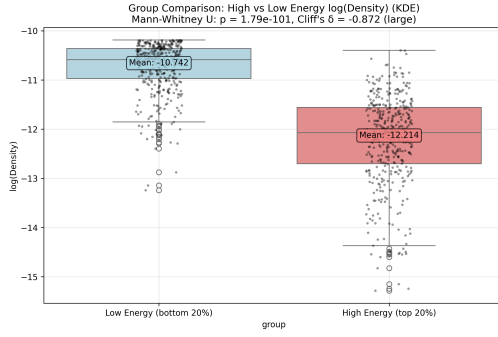
(a) 10 steps - KDE ($\delta = -0.872$)
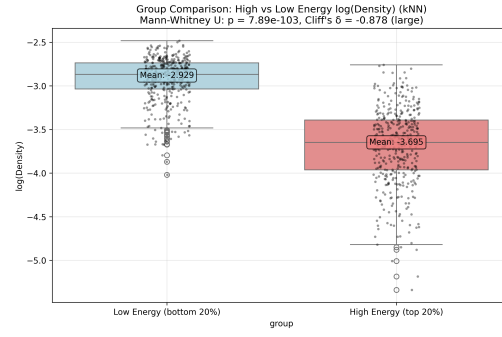
(b) 10 steps - k-NN ($\delta = -0.878$)
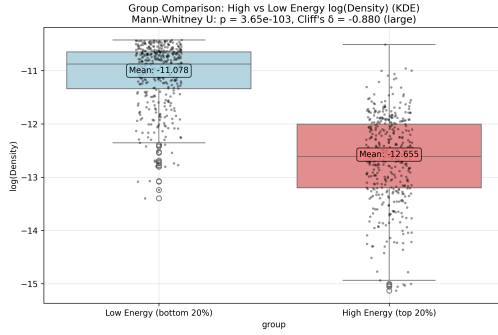
(c) 50 steps - KDE ($\delta = -0.880$)
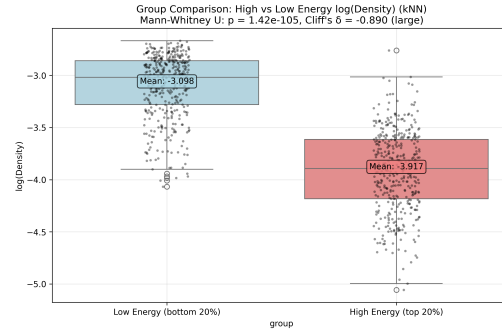
(d) 50 steps - k-NN ($\delta = -0.890$)
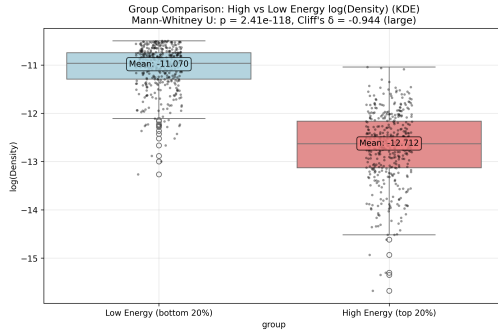
(e) 100 steps - KDE ($\delta = -0.944$)
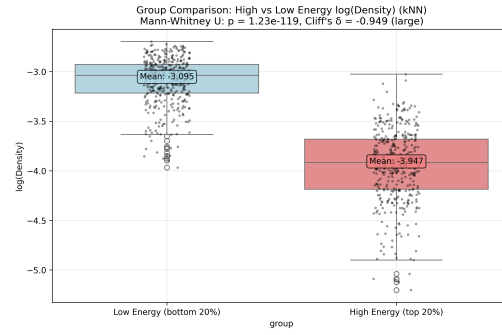
(f) 100 steps - k-NN ($\delta = -0.949$)
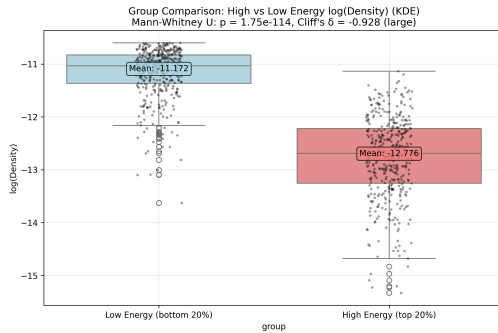
(g) 150 steps - KDE ($\delta = -0.928$)
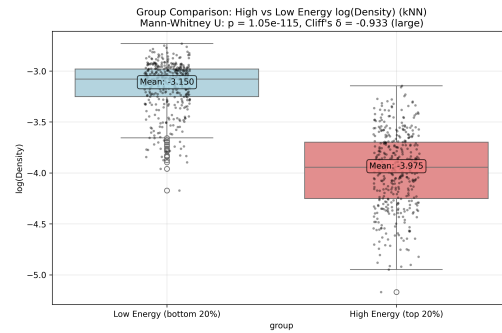
(h) 150 steps - k-NN ($\delta = -0.933$)

Figure 13: Density comparison between high-energy vs. low-energy groups using Cliff's Delta across sampling steps (10, 50, 100, 150 rows) and density estimation methods (KDE left, k-NN right). Large negative effect sizes ($\delta \in [-0.95, -0.87]$) confirm high-energy samples consistently reside in lower-density regions.
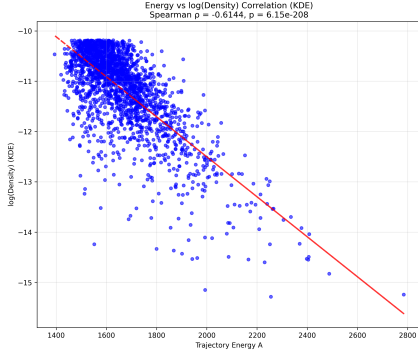
(a) 10 steps - KDE ($\rho = -0.6144$, $p = 6.15 \times 10^{-208}$)
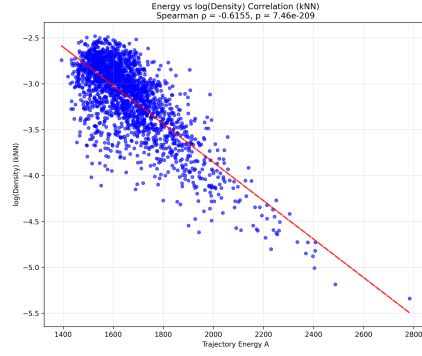
(b) 10 steps - k-NN ($\rho = -0.6155$, $p = 7.46 \times 10^{-209}$)
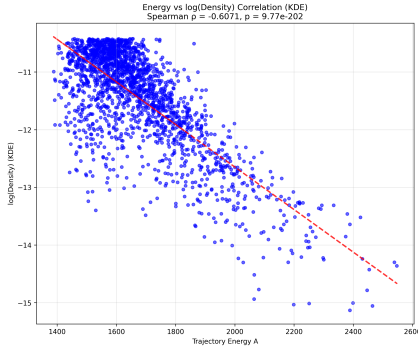
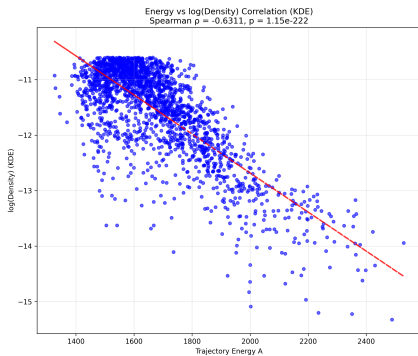(c) 50 steps - KDE ($\rho = -0.6071$, $p = 9.77 \times 10^{-202}$)
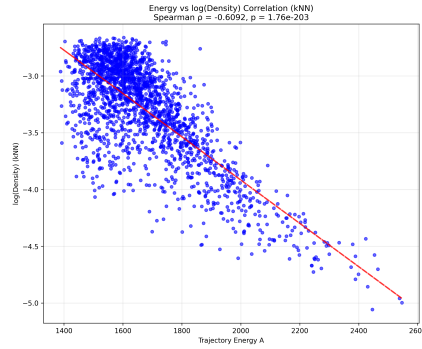
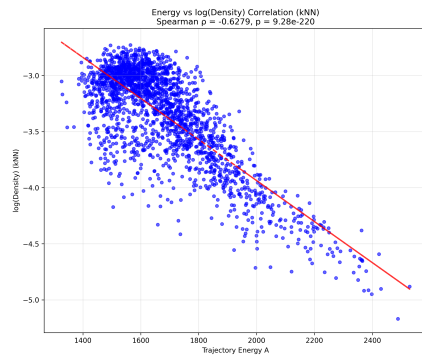(d) 50 steps - k-NN ($\rho = -0.6092$, $p = 1.76 \times 10^{-203}$)
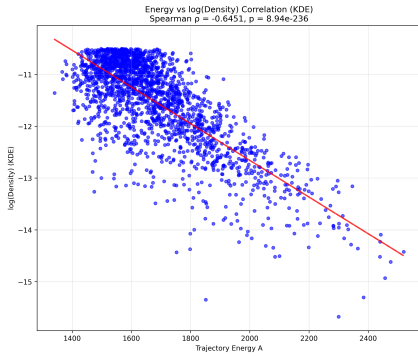
(e) 100 steps - KDE ($\rho = -0.6451$, $p = 8.94 \times 10^{-236}$)
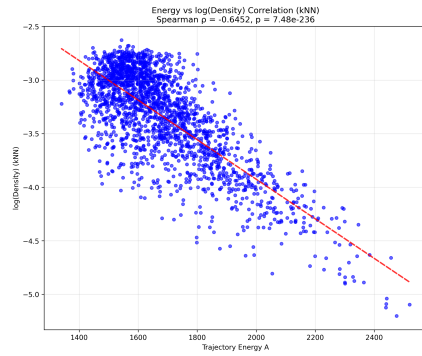
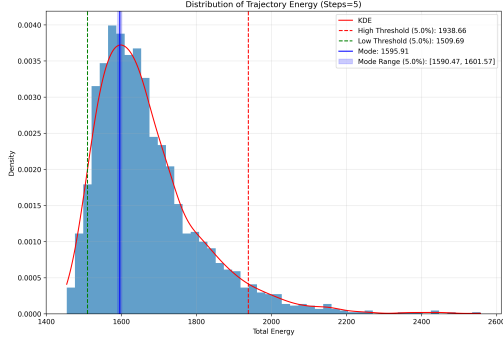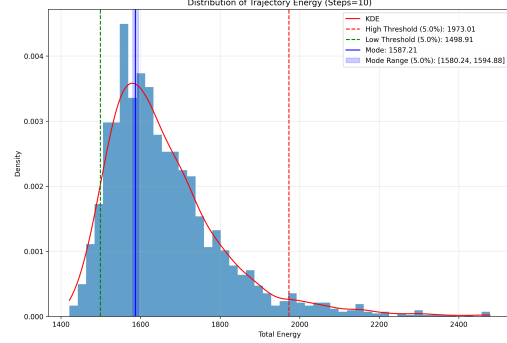(f) 100 steps - k-NN ($\rho = -0.6452$, $p = 7.48 \times 10^{-236}$)

(g) 150 steps - KDE ($\rho = -0.6311$, $p = 1.15 \times 10^{-222}$)

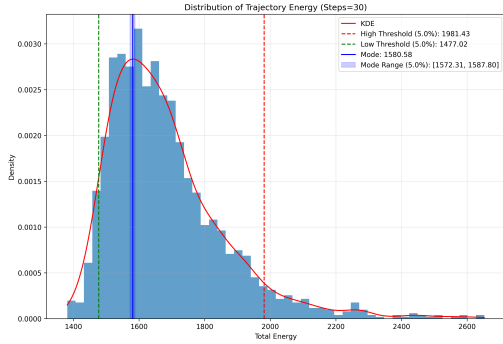(h) 150 steps - k-NN ($\rho = -0.6279$, $p = 9.28 \times 10^{-220}$)

Figure 14: KPE E vs. Data Density correlation across steps (10, 50, 100, 150) using KDE and k-NN. Strong negative correlations (Spearman $\rho \in [-0.65, -0.61]$, $p < 10^{-200}$) confirm higher energy aligns with lower density.
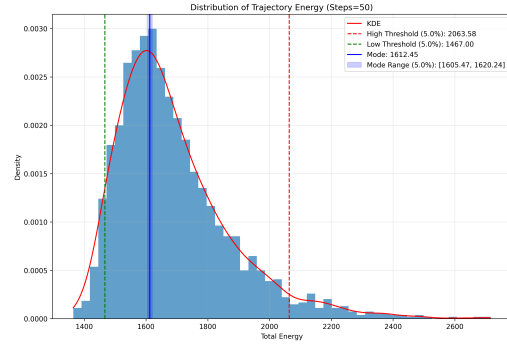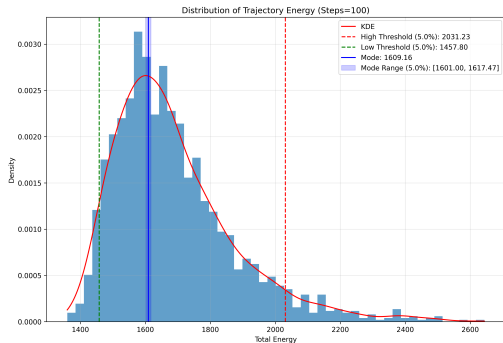
(a) 5 sampling steps
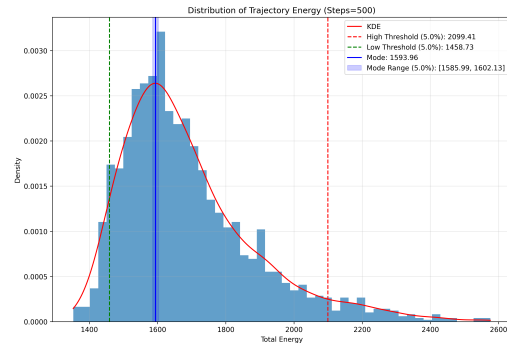

(b) 10 sampling steps

(c) 30 sampling steps

(d) 50 sampling steps

(e) 100 sampling steps

(f) 500 sampling steps

Figure 15: Kinetic path energy (KPE) distributions on CIFAR-10 across varying sampling steps (5, 10, 30, 50, 100, 500). The distributions remain stable with consistent mean ($\mu \approx 1850 \pm 100$) and standard deviation ($\sigma \approx 350 \pm 50$). These observations suggest that KPE is relatively stable across sampling configurations.