One Flow Fits All! A Scale-Aware Generative Framework for Diverse Data

Hubin Cao[1], **Jun Ma**[2], **Yusupu Ainiwaer**[1], **Hanquan Zhang**[1], **Yanjun Qin**[1*], **Zixuan Wang**[1] and **Xiaoming Tao**[1,3]
[1]School of Computer Science and Technology, Xinjiang University, Urumqi 830017, China
[2]Department of Systems Science, Faculty of Arts and Sciences, Beijing Normal University, Zhuhai 519087, China
[3]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
qinyanjun@xju.edu.cn, caohubin@stu.xju.edu.cn

Abstract

Real-world systems increasingly require coherent reasoning and generation over diverse data modalities simultaneously. Current generative frameworks rely on complex, multi-stage training, resulting in low efficiency due to iterative inference and high computational cost. They also struggle with unified multimodal representation, failing to balance fine-grained details with global structures and long-term dependencies, which limits its generative quality and practical usability. To overcome these limitations, we introduce the Inverse Heat Mean Flow (IHMF), a general-purpose solver that is compatible with a wide range of model backbones. Without requiring pre-training, IHMF directly learns an average velocity field through an inverse heat formul ation. By exploiting the inherent scale-space properties of the inverse heat process, IHMF explicitly decomposes multi-scale complexities, thereby simplifying trajectory learning and enabling adaptive topological alignment. Extensive experiments show that IHMF achieves a competitive performance with state-of-the-art methods, providing a robust and unified mathematical framework for various generative tasks. The code is publicly available at https://github.com/CHBonline/IHMF.

1 Introduction

The development of a unified generative model that can explicitly accommodate the simultaneous generation of diverse data modalities modalities—including temporal sequences, non-Euclidean structured data (e.g. graphs and manifolds), and Euclidean image data—constitutes a fundamental yet insufficiently investigated problem in multimodal generative representation learning. The central research objective extends beyond the construction of shared latent representations; It is to design a single generative framework capable of accurately modeling, synthesizing, and supporting downstream inference and decision-making tasks across these intrinsically distinct data types. Achieving such unification would mitigate modality-specific architectural fragmentation, facilitate systematic cross-modal knowledge transfer, and offer a principled basis for general-purpose generative models that exhibit coherent behavior across a broad spectrum of data domains.

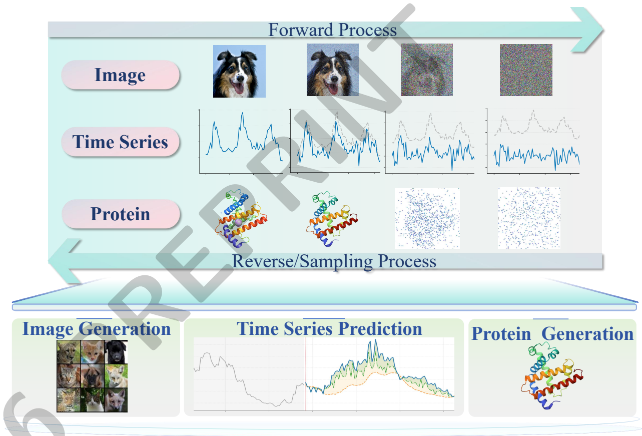

Figure 1: Temporal Unification via Thermodynamic Evolution.

Achieving this capability is critical, as real-world systems increasingly require coherent reasoning and generation over time-evolving signals, relational structures, and high-dimensional perceptual data simultaneously. Generative modeling has witnessed a fundamental paradigm shift, with Diffusion Models [Ho *et al.*, 2020; Sohl-Dickstein *et al.*, 2015; Song and Ermon, 2019] and Flow Matching [Lipman *et al.*, 2023] emerging as dominant frameworks for synthesizing high-dimensional data[Liu *et al.*, 2023a; Albergo and Vanden-Eijnden, 2023]. This capability allows for the injection of structured inductive biases directly into the generation trajectory. The significant differences in data distributions and inherent structures across modalities fundamentally constrain the development of a unified generative or representation learning model.

However, achieving high-fidelity "one-step generation" directly on such diverse distributions remains a formidable challenge. While the Heat Equation naturally describes a Gaussian scale-space that evolves from fine details to coarse structures [Witkin, 1987; Rissanen *et al.*, 2023], simply reversing this process is mathematically ill-posed: recovering

*Corresponding author.

high-frequency details from diffused states is extremely sensitive to numerical precision and observation noise. Existing one-step solutions, such as Consistency Models[Song *et al.*, 2023] or Rectified Flow [Liu *et al.*, 2023a], often struggle to navigate this complexity without guidance. Consequently, they are frequently forced to resort to cumbersome multi-stage distillation pipelines[Salimans and Ho, 2022] to trade training complexity for inference speed. These methods lack a self-contained mechanism to model the structural evolution of data from scratch and fail to explicitly exploit the multi-scale nature of complex distributions, often resulting in blurred outputs or training instability.

To bridge this gap, we propose Inverse Heat Mean Flow (IHMF), a unified algorithmic. As illustrated in Figure 1, our framework ensures that they can all be governed by a unified thermodynamic evolution process. Instead of attempting to numerically solve the unstable inverse heat equation , IHMF directly learns an Average Velocity Field within the Flow Matching framework. We leverage the scale-space properties of the Inverse Heat mechanism as a strong inductive bias. By explicitly decomposing the data complexity into a "coarse-to-fine" hierarchy, we guide the velocity field to learn a trajectory that naturally aligns global structures before refining local geometries. This reformulation bypasses the numerical instability of direct inversion, enabling the model to adaptively learn high-fidelity, one-step generation paths across distinct data topologies without any pre-training or distillation.

Our main contributions are summarized as follows:

- We introduce IHMF, a domain-agnostic solver that unifies the training paradigm for distinct data modalities under a single mathematical formulation based on average velocity modeling.
- We propose a novel training mechanism that reformulates the ill-posed inverse heat problem into a stable regression task. By injecting Gaussian scale-space priors, we enable the model to learn valid "coarse-to-fine" generation trajectories from scratch.
- We demonstrate IHMF's effectiveness across three distinct tasks: disentangling frequency components in Image Generation, aligning global-local geometry in Protein Backbone Generation, and refining dynamics via a two-stage residual strategy in Time-Series Forecasting.

2 Related Work

2.1 Continuous Normalizing Flows and Unified Flow Matching

Continuous Normalizing Flows (CNFs) establish a continuous mapping between prior and data distributions via Ordinary Differential Equations (ODEs)[Chen *et al.*, 2018; Grathwohl *et al.*, 2019]. Recently, Flow Matching [Lipman *et al.*, 2023; Liu *et al.*, 2023a] has emerged as a dominant paradigm, offering an efficient alternative to diffusion models by directly regressing target velocity fields. This framework shifts the modeling focus from inverting noise to designing optimal probability paths, enabling greater flexibility in trajectory design. Driven by this flexibility, recent works have explored unified architectures across modalities. For instance, INRFlow [Wang *et al.*, 2024] unifies the spatial representation of diverse data (e.g.images, point clouds) using Implicit Neural Representations. However, existing unified approaches primarily focus on spatial alignment, often lacking a universal physical inductive bias for "temporal dynamics". Consequently, when applied to non-Euclidean data with complex topologies (such as protein backbones), generic FM methods often struggle to maintain geometric consistency without explicit physical guidance.

2.2 Physics-Driven Generation and the Ill-Posedness of Inverse Heat

Integrating physical laws into generative processes is a promising avenue for enhancing sample fidelity and interpretability. A prominent example is the Inverse Heat Dissipation Model (IHDM)[Rissanen *et al.*, 2023], which leverages the Gaussian Scale-Space naturally induced by the heat equation[Witkin, 1987; Koenderink, 1984]. This mechanism introduces a valuable "coarse-to-fine" inductive bias, determining global structures early in the generation process before refining local details. Nevertheless, explicitly reversing the heat equation is mathematically ill-posed: recovering high-frequency details from diffused states is exponentially sensitive to numerical precision and observation noise. To render the reverse process probabilistically well-defined, IHDM is forced to inject auxiliary noise, causing it to degenerate into a discrete, diffusion-like stochastic process. This reliance on small step sizes fundamentally limits its inference efficiency, making real-time application unfeasible.

2.3 One-Step Generation and Distillation-Free Paradigms

To overcome the inference bottleneck of ODE-based models, accelerating generation to a single step has become a critical objective. One category of solutions relies on distillation[Salimans and Ho, 2022] or recursive rectification, such as Consistency Models[Song *et al.*, 2023] and InstaFlow[Liu *et al.*, 2023b]. However, these methods typically require pretrained teacher models or multiple rounds of "reflow" training to compress trajectories, incurring high computational costs. Another category, exemplified by MeanFlow[Geng *et al.*, 2025], represents the state-of-the-art for direct one-step generation. While statistically optimal for minimizing transport costs, this intrinsic "averaging" mechanism inevitably introduces a non-negligible low-frequency spectral bias.

Distinct from these approaches, our proposed IHMF bridges the gap between physical priors and training stability. We do not attempt to solve the unstable inverse heat PDE, nor do we rely on distillation. Instead, by learning the Average Velocity Field of heat dissipation paths, IHMF converts the ill-posed high-frequency recovery into a stable supervised regression task. This effectively "straightens" the curved physical trajectories via the heat kernel, enabling high-fidelity one-step generation across images, proteins, and time-series from scratch.

3 Method

In this section, we formally present the IHMF framework. IHMF is a universal generative flow architecture grounded in Spectral Dynamics[Chung, 1997; Belkin and Niyogi, 2003], designed to internalize the intrinsic evolution laws of signals into the Mean Flow prediction paradigm[Geng *et al.*, 2025] through physics-inspired path construction. Unlike traditional isotropic flow models, IHMF leverages generalized operators[Li *et al.*, 2020; Kovachki *et al.*, 2023] to characterize the intrinsic topology of data, demonstrating exceptional versatility across tasks such as image synthesis, protein backbone generation, and multi-scale time-series forecasting.

3.1 Theoretical Foundation: Generalized Spectral Dynamics

Operator-Driven Signal Degradation:IHMF assumes that the evolution of a high-dimensional signal $x \in \mathbb{R}^D$ is governed by a domain-specific operator $\mathcal{L}$. We define its structured evolution via a generalized heat equation: $\frac{\partial u}{\partial \tau} = \mathcal{L}u$ [Widder, 1976; Rissanen *et al.*, 2023]. Utilizing spectral decomposition $\mathcal{L} = V\Lambda V^T$, the state u_t at time $t \in [0,1]$ is analytically expressed as:

$$u_t \triangleq V(1-t)^{-\Lambda}V^T x_1 \tag{1}$$

Generality and Interpretation:The choice of $\mathcal{L}$ determines the topological dimensions perceived by the model. For images, $\mathcal{L}$ is the discrete Laplacian utilizing the Discrete Cosine Transform (DCT) basis V [Ahmed *et al.*, 2006]; for protein or graph signals, $\mathcal{L}$ evolves into a Graph Laplacian to capture geometric connectivity[Shuman *et al.*, 2013]; for time-series, it corresponds to difference operators. This hierarchical processing ensures that high-frequency details dissipate faster than global structures.

The Mean Flow Paradigm: To achieve superior inference efficiency, IHMF introduces a mean predictor $u_\theta(z_t, t, h)$, which estimates the average displacement $\frac{z_t - z_{t-h}}{h}$ over an interval h [Geng *et al.*, 2025; Liu *et al.*, 2023b].

This paradigm shifts the generative task from simulating microscopic instantaneous trajectories to predicting macroscopic evolution trends, significantly enhancing the robustness of one-step inference.

3.2 IHMF Model

The core innovation of IHMF lies in the construction of a nonlinear, frequency-aware coupling path z_t, which dynamically bridges stochastic Gaussian noise with systematic structural evolution[Albergo and Vanden-Eijnden, 2023].

Path Construction and Physical Significance: We define the interpolation path of IHMF as:

$$z_t = (1-t)u_t + t\epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \tag{2}$$

This path incorporates a dual-guidance mechanism. In the regime where t is small, the trajectory is primarily governed by the $(1-t)u_t$ term, which prioritizes the global topology of the signal (e.g. a protein's backbone or a time series' trend). Conversely, as t increases, the path seamlessly transitions towards high-frequency priors. This "global-to-local" evolutionary strategy preserves structural consistency across multiple scales[Witkin, 1987].

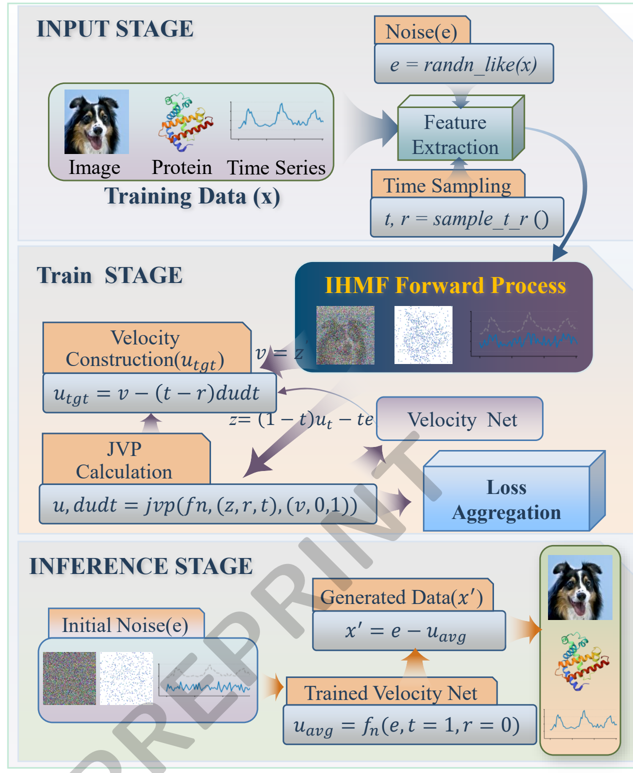

Figure 2: Geometric Principles of IHMF.

Path Construction and Physical Significance:By taking the full derivative of Equation (2) with respect to t, we derive the instantaneous velocity field v_t of IHMF:

$$v_t = -u_t + \mathcal{L}u_t + \epsilon \tag{3}$$

The $\mathcal{L}u_t$ term in the velocity field acts as a structural diffusion component, automatically introducing smoothing constraints during the flow matching process. This ensures that the model maintains the inherent physical consistency of the signal space, whether completing missing residues in proteins or synthesizing high-definition images.

3.3 Numerical Stability and Time Mapping

To mitigate the risk of numerical overflow as $t \to 1$, we introduce a physical time mapping $t = 1 - e^{-s}$. The reparameterized velocity field $\tilde{v}_s = e^{-s}(-u_t + \mathcal{L}u_t + \epsilon)$ ensures stable convergence of the evolution as $s \to \infty$.

This design not only enhances training stability but also establishes a dual relationship between flow matching and classical physical diffusion processes[Song *et al.*, 2020].

3.4 Training and Sampling

As illustrated in Figure 2, IHMF unifies training and sampling by constructing a physically consistent Geometric Flow Path.

During training, we employ domain-specific operators to generate intermediate states u_t adhering to inverse heat flow laws, defining smooth evolutionary trajectories in function space. To precisely capture manifold dynamics, the model utilizes the Jacobian-Vector Product (JVP) to efficiently compute tangential derivatives $\frac{du}{dt}$, explicitly modeling path curvature. Consequently, the training objective is reformulated as regressing the Average Velocity Field, enabling the network to learn a "global transport vector" that effectively "straightens" complex diffusion trajectories.

In the inference phase, this global prediction mechanism translates into a minimalist One-Step Denoising protocol: starting from Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$, the model achieves an instantaneous leap across the time axis via a single forward pass, ensuring geometric consistency while maximizing generation efficiency.

4 Experiments

To comprehensively verify the effectiveness of IHMF as a universal solver, we conducted extensive experiments across three modalities with distinct topological characteristics: Image Generation (high-dimensional Euclidean grids), Protein Backbone Generation (non-Euclidean geometric graphs), and Time-Series Forecasting (long-range dependent sequences).

4.1 Dataset Description

Image Generation. We evaluate our method on the standard CIFAR-10 benchmark [Krizhevsky *et al.*, 2009]. This dataset comprises 60,000 32×32 RGB images across 10 distinct classes. With its complex background semantics and rich textural variations, CIFAR-10 serves as an ideal testbed to verify high-frequency detail preservation and structural coherence under the one-step generation setting.

Protein Backbone Generation. Targeting protein structure generation subject to complex geometric constraints, we utilized the CATH dataset[Sillitoe *et al.*, 2019]. This dataset classifies proteins based on structural domains and serves as a standard benchmark in geometric deep learning. Following common protocols [Ingraham *et al.*, 2019; Trippe *et al.*, 2022], we partitioned the data according to CATH topology classification, selecting non-redundant domains for training, validation, and testing.

Time-Series Forecasting. We evaluated our model on widely adopted long-term forecasting benchmarks, with a primary focus on the **ETT** (Electricity Transformer Temperature) series (e.g., ETTh1) [Zhou *et al.*, 2021], which exhibits strong seasonality and long-range dependencies, as well as multivariate datasets including **Weather**. These benchmarks require the model to precisely capture multi-scale temporal dynamics.

4.2 Experiment Setup

The IHMF framework is implemented in PyTorch. All models were trained on a high-performance cluster equipped with six NVIDIA A100 GPUs (80GB) using data parallelism to ensure stability during large-batch optimization. Detailed configurations are provided in Supplementary Material(A.Model Configuration and Training Settings).

Evaluation Metrics. We adopted standard metrics for each domain to ensure fair comparison:

- Image: We use FID [Heusel *et al.*, 2017] to measure generation quality and diversity.
- Protein: We use scTM [Trippe *et al.*, 2022; Wu *et al.*, 2024] to evaluate the geometric foldability and validity of the generated structures.
- Time-Series: We use MSE and MAE to quantify prediction errors.

4.3 Image Generation

Image generation serves as a cornerstone benchmark for evaluating generative models and has long been dominated by Diffusion Models and multi-step Flow Matching methods[Lipman *et al.*, 2023]. However, state-of-the-art models often grapple with a trade-off between generation quality and sampling efficiency. Achieving high-fidelity results typically requires tens to hundreds of iterative denoising steps (i.e.Number of Function Evaluations, NFE), which severely hinders their potential for real-time deployment. Conversely, existing one-step generation approaches, such as Consistency Models[Song *et al.*, 2023] or distillation techniques[Salimans and Ho, 2022], often come at the cost of high-frequency details and texture sharpness, leading to "over-smoothed" results or blurry artifacts.

In this section, we evaluate the performance of IHMF on image generation tasks. Our primary objective is to verify whether the "inverse heat" mechanism proposed in IHMF can effectively disentangle the multi-scale complexities of images. Specifically, we investigate its ability to precisely capture high-frequency details while maintaining global structural coherence under the challenging regime of One-Step Inference.

We benchmark IHMF against prevailing generative frameworks, primarily utilizing the FID to quantify the discrepancy between the generated images and the real data distribution (where lower values indicate better performance).

As reported in Table 1, we compare the FID scores of various generative models on CIFAR-10. IHMF achieves a remarkable FID of 3.79, significantly outperforming mainstream baselines such as Consistency Models (CT)[Song *et al.*, 2023] and Flow Matching. Notably, our method demonstrates performance highly comparable to the state-of-the-art MeanFlow (3.65)[Geng *et al.*, 2025], verifying the strong distribution fitting capability of IHMF in the one-step generation setting. Figure 3 showcases the visual quality of samples generated by IHMF in a single step. For more extensive sampling results, please refer to Supplementary Materials (B. CIFAR Sampling and C. AFHQ Sampling).

4.4 Protein Backbone Generation

To validate the effectiveness of IHMF in handling non-Euclidean data subject to complex biophysical constraints, we evaluated it on the protein backbone generation task. Following standard evaluation protocols in geometric generative modeling [Trippe *et al.*, 2022; Watson *et al.*, 2023], we adopt the Self-Consistency TM-score (scTM) as the core metric for measuring the biological validity of generated backbones. The evaluation pipeline consists of the following key steps:

- Sequence Design: We use the pre-trained inverse folding model ProteinMPNN[Dauparas *et al.*, 2022] to design amino acid sequences for the generated backbones.

CIFAR-10 (32 × 32)	
Model	FID ↓
Cold Diffusion[Bansal *et al.*, 2023]	80.08
ResidualFlow[Chen *et al.*, 2019]	46.4
DenseFlow[Wang *et al.*, 2020]	34.9
Shortcut Model [Song *et al.*, 2023]	27.28
Flow Matching [Lipman *et al.*, 2023]	20.63
DC-VAE[Frans *et al.*, 2024]	20.63
IHDM[Rissanen *et al.*, 2023]	18.96
CT [Song *et al.*, 2023]	8.7
MeanFlow [Geng *et al.*, 2025]	3.65
IHMF(Ours)	**3.79**

Table 1: Quantitative comparison of image generation performance.

Figure 3: One-Steps Geometric Image of IHMF.

- Structure Prediction:The designed sequences are fed into the high-precision folding model AlphaFold2 [Jumper *et al.*, 2021] to predict their corresponding 3D structures, followed by calculating the consistency between the predicted structures and the original backbones.

Empirically, an scTM > 0.5 indicates that the generated backbone possesses a correct folding topology and can be effectively recognized and encoded by existing sequence design algorithms. Figure 4 presents the scTM scores achieved by IHMF across varying chain lengths, ranging from short peptides to long chains. The results demonstrate that IHMF maintains competitive designability across all length intervals. Benefiting from the “coarse-to-fine” scale-space prior introduced by the Inverse Heat process, our method helps preserve global topological coherence during generation, thereby mitigating issues such as structural breaks or unnatural distortions. This corroborates that the generated backbones possess not only geometric coherence but also high biological realizability. For additional visualizations of the generated high-designability protein samples, please refer to Supplementary Material (D.CATH Samplings).

4.5 Time-series Forecasting

To verify the effectiveness of IHMF in handling data with complex temporal dependencies and multi-scale dynamic characteristics, we conducted a systematic evaluation on the Long-term Time Series Forecasting task. Following standard evaluation protocols in the field[Wu *et al.*, 2021; Nie, 2022], we adopt Mean Squared Error (MSE) and Mean Absolute Error (MAE) as the core metrics.

To address the challenge of forecasting temporal fluctuations, we propose a two-stage prediction framework based on Spectral Residual Refinement. First, an encoder is used to obtain the base trend prediction. Then, the residual distribution is efficiently learned in the frequency domain via the Discrete Cosine Transform (DCT)[Ahmed *et al.*, 2006] and the Inverse Heat Dissipation (IHD) process. Finally, the refined residual is recovered through one-step ODE sampling and superimposed onto the base prediction. This mechanism effectively enhances the model’s capability to capture complex dynamics and local variations.

data	metric	stage1		stage2		improve(%)	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.365	0.392	0.361	0.388	1.10%	1.02%
	192	0.403	0.416	0.399	0.411	0.99%	1.20%
	336	0.423	0.427	0.417	0.421	1.42%	1.41%
	720	0.423	0.447	0.419	0.443	0.95%	0.89%
	avg	0.404	0.421	0.399	0.416	1.12%	1.13%
ETTh2	96	0.266	0.330	0.264	0.327	0.75%	0.91%
	192	0.341	0.378	0.337	0.374	1.17%	1.06%
	336	0.375	0.404	0.371	0.400	0.99%	1.07%
	720	0.406	0.442	0.401	0.436	1.23%	1.36%
	avg	0.347	0.388	0.343	0.384	1.08%	1.09%
ETTm1	96	0.286	0.339	0.283	0.335	1.05%	1.18%
	192	0.324	0.362	0.321	0.359	0.93%	0.83%
	336	0.353	0.381	0.351	0.378	0.57%	0.79%
	720	0.399	0.411	0.395	0.407	1.00%	0.97%
	avg	0.340	0.373	0.338	0.370	0.88%	0.94%
ETTm2	96	0.085	0.204	0.084	0.202	1.18%	0.98%
	192	0.188	0.307	0.186	0.304	1.06%	0.98%
	336	0.351	0.428	0.348	0.424	0.85%	0.93%
	720	0.352	0.378	0.349	0.375	0.85%	0.79%
	avg	0.244	0.329	0.242	0.326	0.92%	0.91%
Weather	96	0.145	0.198	0.143	0.196	1.38%	1.01%
	192	0.185	0.233	0.183	0.230	1.08%	1.29%
	336	0.242	0.276	0.239	0.273	1.24%	1.09%
	720	0.318	0.329	0.314	0.326	1.25%	0.85%
	avg	0.223	0.259	0.220	0.256	1.23%	1.04%

Table 2: Multivariate Time Series Forecasting Results

As shown in Table 2, we quantitatively analyze the performance difference between the Base Prediction (Stage 1) and the Refined Sampling (Stage 2). The results indicate that Stage 2 consistently outperforms Stage 1 across all datasets and prediction horizons. This provides strong evidence that our “residual refinement” strategy effectively rectifies the bias of deterministic predictions and improves generation fidelity. Furthermore, Table 3 presents a comprehensive comparison between IHMF and other baseline methods. The results show that IHMF achieves the best (red) or second-best (blue) performance in the vast majority of datasets and prediction windows. For qualitative visualizations of the forecasting results, please refer to Supplementary Material(E.Time-series Fore-

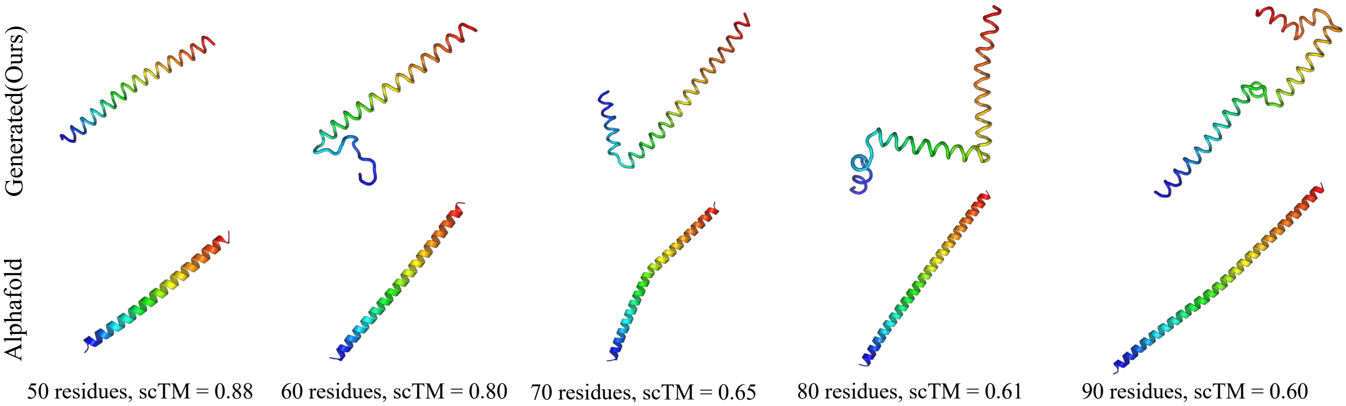

Figure 4: Selected generated protein backbones of varying length that are approximately designable.

casting).

4.6 Ablation Study

To thoroughly investigate the underlying mechanisms of Inverse Heat Mean Flow (IHMF) and quantify the specific contributions of its components, we conduct comprehensive ablation studies on the CIFAR-10 dataset. Our primary objective is to disentangle the factors driving IHMF's superior performance in single-step generation and to verify its robustness under varying hyperparameter configurations. We focus our analysis on three critical dimensions: the impact of sampling steps on trajectory linearity, the robustness against noise perturbations, and the pivotal role of blur intensity in high-frequency guidance.

Sampling Steps and Trajectory Linearity

A core premise of IHMF is its ability to construct a near-linear velocity field via the inverse heat process, thereby facilitating high-quality one-step generation. To validate this hypothesis, we evaluated the model's performance in terms of FID across different numbers of function evaluations (NFE = 1, 5, 10, 15, 20) during inference.

As shown in Table 4, the results reveal a significant phenomenon: IHMF establishes a highly competitive performance baseline in the single-step (NFE=1) setting with an FID of 3.79. While increasing the sampling steps to 5 brings a notable improvement (FID drops to 3.06), further increasing NFE beyond 10 yields diminishing returns (e.g.FID only decreases marginally to 2.80 at NFE=20). In stark contrast, traditional Flow Matching or Diffusion Models typically suffer from severe mode collapse or excessive blurring at NFE=1.

This observation strongly suggests that the probability flow ODE trajectories learned by IHMF exhibit extremely low curvature [Liu *et al.*, 2023a]. Although multi-step sampling (e.g., NFE=5) can further refine high-frequency details, the fact that a single Euler step captures the majority of the distribution quality confirms that the inverse heat equation effectively "straightens" the mapping path. Consequently, IHMF serves as an efficient "one-step solver", relying on the geometric optimization of the flow manifold rather than mere memorization.

Robustness to Noise Perturbation

Noise serves not only as a random seed but also as a crucial regularizer for the ill-posed inverse heat process. We investigated the impact of introducing additive Gaussian noise (σ_{noise}) of varying scales on IHMF performance (Table 5).

To efficiently explore the hyperparameter space, the models in this and the subsequent section were trained for 1,000 epochs. While not fully converged, this duration is sufficient to establish stable relative trends for sensitivity analysis. We observe distinct behaviors across different noise regimes:

- **Low-Noise Regime** ($\sigma_{noise} = 0.1$)**:** The model performance drops significantly (FID rises to 19.41). This indicates that insufficient noise injection fails to regularize the inverse heat equation effectively, leading to numerical instabilities and high-frequency artifacts rather than clean generation.
- **Optimal Regime** ($\sigma_{noise} = 0.5$)**:** With moderate noise perturbation, the model achieves the best performance (FID 10.45). This suggests that appropriate stochasticity effectively fills the sparse regions of the manifold and smooths the velocity field, striking an optimal balance between stability and detail.
- **High-Noise Regime** ($\sigma_{noise} = 1.0$)**:** Notably, as noise intensity doubles from 0.5 to 1.0, the FID only degrades marginally (from 10.45 to 10.98). This demonstrates IHMF's exceptional *robustness*. Unlike standard methods where performance often decays rapidly with excessive noise, the intrinsic "focusing" property of the inverse heat operator allows the model to preserve structural integrity even in high-noise environments.

Impact of Blur Intensity and High-Frequency Guidance

This component is critical to our ablation study. We verify the necessity of "High-Frequency Guidance" by adjusting the heat diffusion coefficient (Blur Intensity, σ_{blur}) to control the degree of frequency decomposition. We categorized the experiments into three settings: Under-Regularization, Optimal Balance, and Over-Smoothing (Table 5).

The results reveal a distinct U-shaped relationship between blur intensity and generation quality, highlighting a fundamental trade-off between geometric regularization and information preservation.

Models		IHMF(Ours)		TimeMixer++		TimeMixer		iTransformer		PatchTST		Crossformer		TiDE		TimesNet		Dlinear	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.361	0.388	0.361	0.403	0.375	0.400	0.386	0.405	0.460	0.447	0.423	0.448	0.479	0.464	0.384	0.402	0.397	0.412
	192	0.399	0.411	0.416	0.441	0.429	0.421	0.441	0.512	0.477	0.429	0.471	0.474	0.525	0.492	0.436	0.429	0.446	0.441
	336	0.417	0.421	0.430	0.434	0.484	0.458	0.487	0.458	0.546	0.496	0.570	0.546	0.565	0.515	0.491	0.469	0.489	0.467
	720	0.419	0.443	0.467	0.451	0.498	0.482	0.503	0.491	0.544	0.517	0.653	0.621	0.594	0.558	0.521	0.500	0.513	0.510
	avg	0.399	0.416	0.419	0.432	0.447	0.440	0.454	0.447	0.516	0.484	0.529	0.522	0.541	0.507	0.458	0.450	0.461	0.457
ETTh2	96	0.264	0.327	0.276	0.328	0.289	0.341	0.297	0.349	0.308	0.355	0.745	0.584	0.400	0.440	0.340	0.374	0.340	0.394
	192	0.337	0.374	0.342	0.379	0.372	0.392	0.380	0.400	0.393	0.405	0.877	0.656	0.528	0.509	0.402	0.414	0.482	0.479
	336	0.371	0.400	0.346	0.398	0.386	0.414	0.428	0.432	0.427	0.436	1.043	0.731	0.643	0.571	0.452	0.452	0.591	0.541
	720	0.401	0.436	0.392	0.415	0.412	0.434	0.427	0.445	0.436	0.450	1.104	0.763	0.874	0.679	0.462	0.468	0.839	0.661
	avg	0.343	0.384	0.339	0.380	0.364	0.395	0.383	0.407	0.391	0.411	0.942	0.684	0.611	0.550	0.414	0.427	0.563	0.519
ETTm1	96	0.283	0.335	0.310	0.334	0.320	0.357	0.334	0.368	0.352	0.374	0.404	0.426	0.364	0.387	0.338	0.375	0.346	0.374
	192	0.321	0.359	0.348	0.362	0.361	0.381	0.390	0.393	0.374	0.387	0.450	0.451	0.398	0.404	0.374	0.387	0.382	0.391
	336	0.351	0.378	0.376	0.391	0.390	0.404	0.426	0.420	0.421	0.414	0.532	0.515	0.428	0.425	0.410	0.411	0.415	0.415
	720	0.395	0.407	0.440	0.423	0.454	0.441	0.491	0.459	0.462	0.449	0.666	0.589	0.487	0.461	0.478	0.450	0.473	0.451
	avg	0.338	0.370	0.369	0.378	0.381	0.395	0.407	0.410	0.406	0.407	0.513	0.495	0.419	0.419	0.400	0.406	0.404	0.408
ETTm2	96	0.084	0.202	0.170	0.245	0.175	0.258	0.180	0.264	0.183	0.270	0.287	0.366	0.207	0.305	0.187	0.267	0.193	0.293
	192	0.186	0.304	0.229	0.291	0.237	0.299	0.250	0.309	0.255	0.314	0.414	0.492	0.290	0.364	0.249	0.309	0.284	0.361
	336	0.348	0.424	0.303	0.343	0.298	0.340	0.311	0.348	0.309	0.347	0.597	0.542	0.377	0.422	0.321	0.351	0.382	0.429
	720	0.349	0.375	0.373	0.399	0.391	0.396	0.412	0.407	0.412	0.404	1.730	1.042	0.558	0.524	0.408	0.403	0.558	0.525
	avg	0.242	0.326	0.269	0.320	0.275	0.323	0.288	0.332	0.290	0.334	0.757	0.610	0.358	0.404	0.291	0.333	0.354	0.402
Weather	96	0.143	0.196	0.155	0.205	0.163	0.209	0.174	0.214	0.186	0.227	0.195	0.271	0.202	0.261	0.172	0.220	0.195	0.252
	192	0.183	0.230	0.201	0.245	0.208	0.250	0.221	0.254	0.234	0.265	0.209	0.277	0.242	0.298	0.219	0.261	0.237	0.295
	336	0.239	0.273	0.237	0.265	0.251	0.287	0.278	0.296	0.284	0.301	0.273	0.332	0.287	0.335	0.280	0.306	0.282	0.331
	720	0.314	0.326	0.312	0.334	0.339	0.341	0.358	0.347	0.356	0.349	0.379	0.401	0.351	0.386	0.365	0.359	0.345	0.382
	avg	0.220	0.256	0.226	0.262	0.240	0.271	0.258	0.278	0.265	0.285	0.264	0.320	0.271	0.320	0.259	0.287	0.265	0.315

Table 3: Multivariate time-series forecasting results comparison (MSE and MAE).

CIFAR-10 (32×32)	
NFE	FID $\downarrow$
1	3.79
5	3.06
10	2.98
15	2.87
20	2.80

Table 4: NFE Ablation

CIFAR-10 (32×32)		
Blur	Noise	FID $\downarrow$
1.0	0.1	19.41
1.0	0.5	10.45
1.0	1.0	10.98
0.7	1.0	10.10
0.3	1.0	10.45

Table 5: Noise,Blur Ablation

- **Under-Regularization** ($\sigma_{blur} = 0.3$)**:** When the blur intensity is too low, the performance drops slightly (FID rises to 10.45). Although minimal blur preserves the maximum amount of original information, it fails to provide sufficient geometric smoothing. Consequently, the model cannot fully benefit from the simplified trajectory offered by the heat equation, leading to suboptimal reconstruction compared to the moderate setting.
- **Optimal Balance** ($\sigma_{blur} = 0.7$)**:** The model achieves its peak performance at a moderate blur intensity, yielding the lowest FID of 10.10. This suggests that $\sigma_{blur} \approx 0.7$ acts as a "sweet spot" for the CIFAR-10 dataset. In this regime, the heat diffusion is strong enough to effectively simplify the data manifold and guide the generation trajectory, yet mild enough to retain sufficient high-frequency details for the inverse recovery process.
- **Over-Smoothing** ($\sigma_{blur} = 1.0$)**:** Conversely, increasing the blur intensity excessively leads to performance degradation (FID 10.98). Strong diffusion obliterates a significant portion of fine-grained texture information, rendering the inverse problem increasingly ill-posed. This confirms that while high-frequency guidance is beneficial, maintaining a reasonable Signal-to-Noise Ratio (SNR) in the guidance term is essential for stability.

5 Conclusion

In this work, we present the Inverse Heat Mean Flow (IHMF), a unified generative solver specifically engineered to efficiently model heterogeneous data modalities. By reformulating the generative task as learning the average velocity field of an inverse heat process, our framework effectively circumvents the mathematical ill-posedness inherent in reverse diffusion while leveraging Gaussian scale-space properties to establish a stable "coarse-to-fine" trajectory. Crucially, this physics-driven paradigm enables high-fidelity, single-step generation directly from noise, eliminating the computational burden of pre-training or multi-stage distillation pipelines required by traditional flow matching methods.

Empirical evaluations across diverse topologies demonstrate IHMF's superiority as a general-purpose solver. From decoupling high-frequency textures in image synthesis and ensuring geometric validity in non-Euclidean protein folding, to capturing long-term dependencies in time-series forecasting, the model consistently outperforms specialized baselines through its spectral dynamics alignment. These results not only validate IHMF as a realization of the "One Flow Fits All" vision but also establish a robust mathematical foundation for physics-driven AI. Future work will extend this scalable framework to broader scientific domains and large-scale multimodal pre-training.

Ethical Statement

There are no ethical issues.

Acknowledgments

This work was supported by the "Tianchi Yingcai" Introduction Program, the Basic Research Project of the Autonomous Region's Universities' Basic Research Operating Funds (XJEDU2025J001), the Open Research Fund Program of Beijing National Research Center for Information Science and Technology, the Key Research and Development Project of the Autonomous Region (2024B03028), the National Key Research and Development Program of China (Grant No. 2025YFF0515600), and the National Natural Science Foundation of China (Grant No. 62595730).

Contribution Statement

Hubin Cao, Jun Ma, Yusupu Ainiwaer and Hanquan Zhang contributed equally.

References

[Ahmed *et al.*, 2006] Nasir Ahmed, T_ Natarajan, and Kamisetty R Rao. Discrete cosine transform. *IEEE transactions on Computers*, 100(1):90–93, 2006.

[Albergo and Vanden-Eijnden, 2023] Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *ICLR*, 2023.

[Bansal *et al.*, 2023] Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie Li, Hamid Kazemi, Furong Huang, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. *Advances in Neural Information Processing Systems*, 36:41259–41282, 2023.

[Belkin and Niyogi, 2003] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

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

[Chen *et al.*, 2019] 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, 2019.

[Chung, 1997] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[Dauparas *et al.*, 2022] Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.

[Frans *et al.*, 2024] Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One step diffusion via shortcut models. *arXiv preprint arXiv:2410.12557*, 2024.

[Geng *et al.*, 2025] Zhengyang Geng, Mingyang Deng, Xingjian Bai, J Zico Kolter, and Kaiming He. Mean flows for one-step generative modeling. *arXiv preprint arXiv:2505.13447*, 2025.

[Grathwohl *et al.*, 2019] Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. In *International Conference on Learning Representations (ICLR)*, 2019.

[Heusel *et al.*, 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. *Advances in neural information processing systems*, 30, 2017.

[Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[Ingraham *et al.*, 2019] John Ingraham, Vikas Garg, Regina Barzilay, and Tommi Jaakkola. Generative models for graph-based protein design. *Advances in neural information processing systems*, 32, 2019.

[Jumper *et al.*, 2021] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *nature*, 596(7873):583–589, 2021.

[Koenderink, 1984] Jan J Koenderink. The structure of images. *Biological cybernetics*, 50(5):363–370, 1984.

[Kovachki *et al.*, 2023] Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

[Krizhevsky *et al.*, 2009] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[Li *et al.*, 2020] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

[Lipman *et al.*, 2023] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

[Liu *et al.*, 2023a] Xingchao Liu, Chenyang Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2023.

[Liu *et al.*, 2023b] Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, and Qiang Liu. Instaflow: One step is enough for high-quality diffusion-based text-to-image generation. *arXiv preprint arXiv:2309.06380*, 2023.

[Nie, 2022] Y Nie. A time series is worth 64words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

[Rissanen *et al.*, 2023] Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. In *International Conference on Learning Representations*, 2023.

[Salimans and Ho, 2022] Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022.

[Shuman *et al.*, 2013] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.

[Sillitoe *et al.*, 2019] Ian Sillitoe, Natalie Dawson, Tony E Lewis, Sayoni Das, Jonathan G Lees, Paul Ashford, Adeyelu Tolulope, Harry M Scholes, Ilya Senatorov, Andra Bujan, et al. Cath: expanding the horizons of structure-based functional annotations for genome sequences. *Nucleic acids research*, 47(D1):D280–D284, 2019.

[Sohl-Dickstein *et al.*, 2015] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. pmlr, 2015.

[Song and Ermon, 2019] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

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

[Song *et al.*, 2023] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. 2023.

[Trippe *et al.*, 2022] Brian L Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*, 2022.

[Wang *et al.*, 2020] Shiguang* Wang, Zhizhong* Li, Yue Zhao, Yuanjun Xiong, Limin Wang, and Dahua Lin. denseflow. https://github.com/open-mmlab/denseflow, 2020.

[Wang *et al.*, 2024] Yuyang Wang, Anurag Ranjan, Josh Susskind, and Miguel Angel Bautista. Inrflow: Flow matching for inrs in ambient space. *arXiv preprint arXiv:2412.03791*, 2024.

[Watson *et al.*, 2023] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

[Widder, 1976] David Vernon Widder. *The heat equation*, volume 67. Academic Press, 1976.

[Witkin, 1987] Andrew P Witkin. Scale-space filtering. In *Readings in computer vision*, pages 329–332. Elsevier, 1987.

[Wu *et al.*, 2021] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 2021.

[Wu *et al.*, 2024] Kevin E Wu, Kevin K Yang, Rianne van den Berg, Sarah Alamdari, James Y Zou, Alex X Lu, and Ava P Amini. Protein structure generation via folding diffusion. *Nature communications*, 15(1):1059, 2024.

[Zhou *et al.*, 2021] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.