Energy-Guided Transport for Projection-Free Physics-Informed Flow Matching

Giuliano Galadini **Aurelio Uncini** **Danilo Comminiello**
Department of Information Engineering, Electronics and Telecommunications
Sapienza University of Rome
Rome, Italy
{giuliano.galadini, aurelio.uncini, danilo.comminiello}@uniroma1.it

Abstract

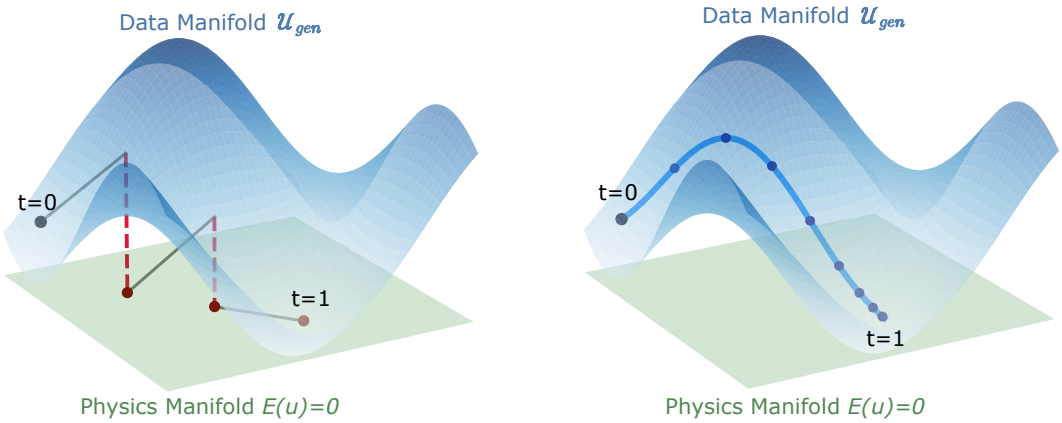
Deep generative models, particularly Flow Matching frameworks, have emerged as powerful tools for synthesizing complex physical systems. However, standard architectures remain purely statistically driven and agnostic to the underlying governing equations, often violating fundamental physical constraints during inference. Existing zero-shot approaches designed to enforce these constraints typically rely on rigid algebraic projections. While effective for sparse boundary conditions, applying such discrete geometric corrections to dense, full-domain PDEs may disrupt the probability path, potentially deviating from the learned manifold and introducing substantial computational overhead. To address these challenges, we introduce Energy-Guided Transport, a novel training-free framework that formulates non-linear physical constraints as a smooth potential energy landscape. Rather than interrupting the ODE integration, our method applies a continuous dissipative force directly to the intermediate generative states, steering them toward the physical manifold. By evaluating this force implicitly via reverse-mode automatic differentiation, our framework preserves the natural dynamics of the flow while maintaining an efficient $\mathcal{O}(N)$ algorithmic complexity. We extensively evaluate our framework across various 1D and 2D physical systems, including the chaotic Navier-Stokes equations. Our method achieves stable generation without structural divergence, significantly improves the stability-accuracy trade-off compared to state-of-the-art methods, and substantially accelerates inference with respect to exact projection strategies. Code available at: https://github.com/ispamm/Projection-free-energy-guided-flow-matching.

1 Introduction

The accurate simulation of complex spatiotemporal dynamics has traditionally relied on computationally expensive numerical solvers. To bypass the curse of dimensionality inherent in high-dimensional PDEs, deep neural networks were initially introduced as direct physics solvers [1, 2]. Building on this foundation, deep generative models, ranging from Score-Based Diffusion [3, 4] to Flow Matching frameworks [5, 6], have recently emerged as scalable data-driven surrogates for modeling physical systems [7–10]. These models have achieved breakthrough successes in global weather forecasting and fluid dynamics [11–16].

By learning the intrinsic data manifold from high-fidelity solvers, these generative architectures offer unprecedented scalability, producing complex, high-dimensional physical states in a fraction of the time required by traditional simulators [8, 11, 12]. However, standard generative models remain purely statistical; they synthesize visually convincing fields but lack an inherent understanding of the governing physics. This purely statistical nature often leads to severe violations of conservation laws, boundary conditions, and internal dynamic consistencies [17]. Ultimately, violations of these

40th Conference on Neural Information Processing Systems (NeurIPS 2026).



(a) Explicit Hard Projections (b) EGT-FFM (Ours)

Figure 1: Conceptual comparison of generative probability flows. (a) Algebraic projections interrupt continuous integration, potentially displacing trajectories from the learned data manifold. (b) Our dissipative guidance smoothly steers the trajectory along the data manifold towards the physical equilibrium.

macroscopic constraints severely degrade out-of-distribution generalization [18] and the reliability of foundation models for PDEs [19].

To bridge this gap, historical techniques attempted injecting physical priors into the generative process. Early approaches primarily focused on data-driven inverse problems and inpainting [20–22]. In these settings, models rely heavily on dense interior observations (e.g., sparse sensor data) to anchor the generative trajectory, reducing the physics problem to a conditioned interpolation task. While highly effective for data assimilation, these observation-dependent methods fail to act as rigorous, independent PDE solvers. Consequently, the field recently shifted toward Zero-Shot Physics Guidance, a much stricter paradigm. In this regime, a pre-trained base model is guided exclusively by the analytical PDE and its formal Boundary/Initial Conditions (BC/IC) during inference, without access to interior ground-truth data or computationally prohibitive retraining.

Despite the theoretical elegance of zero-shot inference, modern guidance architectures suffer from critical structural and computational limitations. Existing approaches often face trade-offs between numerical stability and computational scalability: gradient-based approximations can become unstable in non-linear regimes, while explicit algebraic projections (e.g., PCFM [23]), though effective for sparse boundary conditions, incur significant computational overhead and may perturb the generative trajectory when applied to dense, full-domain PDE constraints.

To address these limitations, we propose Energy-Guided Transport (EGT), a continuous guidance framework for physics-informed Functional Flow Matching. Rather than relying on unstable approximations or discrete geometric projections that intermittently halt the generative process, we formulate the physical PDE residual as a smooth potential energy landscape. Central to our approach are the Dissipation properties of Flow Matching, which proves that modulating the generative vector field with the gradient of this physical energy applies a Lyapunov-like dissipative force. As conceptualized in Figure 1, this allows our guidance to natively harmonize with the probability flow: it smoothly counteracts physical violations along the trajectory, steering samples toward the true physical manifold without the structural discontinuities or $\mathcal{O}(N^2)$ memory bottlenecks of exact algebraic projections.

To empirically validate our theoretical claims, we conduct an exhaustive experimental campaign across four distinct physical systems of increasing non-linearity and dimensionality: the 1D Heat equation, 1D Reaction-Diffusion (RD), 1D Burgers' equation, and the chaotic 2D Navier-Stokes equations. Most importantly, for each physical system, we rigorously evaluate our framework under two distinct regimes: partial physics guidance (enforcing only boundary and initial conditions) and a highly constrained full-PDE stress test. Furthermore, we provide extensive ablation studies investigating the hyperparameter dynamics of the dissipative force and the structural impact of

generative priors, comparing standard White Noise against spatially correlated Gaussian Random Fields (Smooth Kernels).

Our primary contributions are summarized as follows:

- **Theoretical Formulation of Dissipative Guidance:** we establish the dissipation properties of Flow Matching, demonstrating via a Lagrangian perspective how a continuous dissipative force ensures the minimization of physical energy functionals. By steering the vector field natively rather than interrupting the ODE to compute discrete projections, our framework seamlessly preserves the foundational continuity of Flow Matching.
- **Analysis of Projection-Based Guidance:** we provide a rigorous empirical analysis exposing the structural limitations of explicit projection methods. We demonstrate that enforcing rigid algebraic constraints can disrupt the generative flow, highlighting the trade-off between exact boundary matching and internal physical coherence (the waterbed effect).
- **Computational Scalability:** by completely avoiding explicit Jacobian materialization and employing VJP-based dissipative guidance coupled with gradient norm clipping, our method circumvents the $\mathcal{O}(N^2)$ memory bottlenecks, enabling scalable full-PDE guidance on 2D fluid dynamics.
- **Favorable Speed–Stability–Accuracy Trade-Off:** through a strict zero-shot evaluation protocol across four benchmark datasets, our method is capable of maintaining consistent numerical stability (100% success rate) while delivering competitive distributional fidelity and low physical residuals at up to two orders of magnitude faster inference speeds than explicit projection methods.

2 Related Work

Generative Models for Physical Systems. While early advancements in modeling complex physical dynamics relied heavily on Score-Based Diffusion Models [4], recent attention has shifted toward Flow Matching (FM) frameworks [5, 6, 24], drawing deep connections with computational optimal transport [25–27]. By formulating generation as an Ordinary Differential Equation (ODE) [28] and learning a continuous vector field, FM provides faster convergence and structurally smoother trajectories, seamlessly adapting even to complex geometries and Riemannian manifolds [29–32]. In the context of spatiotemporal physics, Functional Flow Matching (FFM) combines continuous normalizing flows with Neural Operators (e.g., Fourier Neural Operators [33], DeepONet [34], and general operator frameworks [35, 36]). Recently, the extension of these score-based and flow paradigms to functional spaces has gained significant traction [37, 38]. To overcome this purely statistical nature, the literature has historically turned to physics-informed training.

Physics-Informed Training. The historical paradigm for injecting physics into neural networks incorporates analytical residuals directly into the training objective (e.g., PINNs [39–41]). While highly successful for specific forward and inverse regression tasks, applying this paradigm to generative modeling fundamentally limits generalizability. Enforcing physics during training structurally couples the generative prior to specific boundary and initial conditions. Consequently, adapting the model to novel constraints requires computationally prohibitive retraining, rendering these approaches inflexible. Furthermore, directly optimizing physical residuals alongside data objectives during network training often suffers from severe gradient flow pathologies, where competing loss terms prevent convergence [42, 43]. This limitation has motivated the recent paradigm shift toward Zero-Shot guidance.

Zero-Shot Physics Guidance. Zero-shot inference methods attempt to correct generative trajectories using analytical PDE constraints without altering the pre-trained weights. Drawing inspiration from classifier-free [44] and universal guidance [45] in vision tasks, these approaches generally fall into three categories:

- *Optimization-Based Guidance:* Methods such as DiffusionPDE [46] and D-Flow [47] utilize gradients of physical residuals to shift the generative trajectory. For instance, D-Flow optimizes the initial noise to align the final state with the constraints. However, because diffusion backbones are heavily corrupted by noise, computing physical gradients directly on intermediate states is unstable. Consequently, these methods often rely on a "predict-

and-correct" mechanism using denoised terminal predictions (e.g., via Tweedie's formula), which can lead to overshooting in highly non-linear dynamics.

- *Projection-Based Guidance:* To avoid gradient-based instability, frameworks like Extrapolation, Correction, Interpolation (ECI) [48] and Physics-Constrained Flow Matching (PCFM) [23] apply explicit geometric corrections. ECI is a gradient-free approach that interleaves extrapolation, orthogonal projection, and interpolation to strictly enforce linear constraints. PCFM extends this concept to arbitrary nonlinear constraints using Gauss-Newton updates. Notably, while PCFM is highly memory-efficient ($\mathcal{O}(n)$ complexity) for sparse constraints by solving a small $m \times m$ Schur complement system, applying its Jacobian-based projections to dense, full-domain PDE residuals incurs prohibitive computational costs. Furthermore, rigid geometric projections can sometimes force the intermediate states off the learned data manifold.
- *Proximal-Based Guidance:* To reconcile strict physical consistency with the pre-trained prior, recent frameworks like ProFlow [49] reframe the correction step as a proximal optimization problem. ProFlow alternates between a terminal optimization step, which anchors the prediction to the physically consistent set via proximal minimization, and an interpolation step that restores compatibility with the learned flow probability path. This implicitly balances trust in the generative prior against physical and data fidelity, avoiding both the intractability of full Jacobian inversions and the trajectory disruption caused by pure algebraic projections.

Unlike these paradigms, our Energy-Guided FFM avoids both unrolled terminal-state estimations and discrete intermediate projections. By applying a continuous dissipative force directly to the intermediate states, we natively harmonize physical consistency with the probability flow while maintaining strict algorithmic scalability.

3 Methodology

3.1 Problem Formulation

We consider physical systems governed by time-dependent Partial Differential Equations (PDEs) [50] defined over a bounded spatial domain $\Omega \subset \mathbb{R}^d$ and a time interval $[0, T]$. Let $u : \Omega \times [0, T] \to \mathbb{R}^n$ denote the spatio-temporal solution field. A generalized physical system is strictly characterized by its internal dynamics, initial conditions (IC), and boundary conditions (BC):

$$\partial_t u + \mathcal{N}(u; \theta) = 0 \text{ in } \Omega \times (0, T], \quad u(\cdot, 0) = u_0 \text{ in } \Omega, \quad \mathcal{B}(u) = 0 \text{ on } \partial\Omega \times [0, T]. \tag{1}$$

where $\mathcal{N}$ is a spatial differential operator, θ denotes the physical parameters, and $\mathcal{B}$ is the boundary operator. We denote the manifold of all physically admissible functions that exactly satisfy Eq. (1) as $\mathcal{U}_{true}$.

In generative scientific machine learning, we assume access to a pre-trained base model (e.g., an unguided Flow Matching network) that approximates the empirical data distribution. However, due to finite data and neural approximations, this unguided model generates solutions from a broader, unconstrained support $\mathcal{U}_{gen}$, frequently violating the strict structural properties of $\mathcal{U}_{true}$.

Our objective is to formulate a zero-shot inference mechanism that smoothly steers the generative trajectories toward the physically admissible manifold $\mathcal{U}_{true}$. To achieve this, we encapsulate the targeted physical laws into a general constraint operator $h(u) = 0$. Depending on the downstream task, $h(u)$ can represent partial constraints (e.g., IC/BC enforcement) or the highly non-linear, full microscopic PDE residual. A comprehensive theoretical taxonomy of these constraint operators and their energy landscapes is provided in Appendix B.

3.2 Functional Flow Matching Foundations

Our framework builds upon Functional Flow Matching (FFM) [38], which generalizes diffusion paradigms to infinite-dimensional Hilbert spaces $\mathcal{H}$ (e.g., $L^2(\Omega)$). Let $p_0(u)$ be a tractable prior distribution over functions (e.g., Gaussian random fields) and $p_1(u)$ be the empirical target distribution of valid physical states. FFM learns a neural functional vector field $v_\theta(u, t) : \mathcal{H} \times [0, 1] \to \mathcal{H}$ to

Algorithm 1 Energy-Guided Transport for Functional Flow Matching

Require: Pre-trained vector field v_θ, Constraint $h(u)$, Energy $E(u) = \frac{1}{2}\|h(u)\|_2^2$
Require: Integration steps N, Max guidance γ_{max}, Clip threshold c
Sample initial state $u_0 \sim p_0(u)$ ▷ e.g., Smooth Gaussian Kernel
Set step size $\Delta t = 1/N$
$u \leftarrow u_0$
for $i = 0$ to $N - 1$ **do** ▷ Continuous Guided Flow
  $t \leftarrow i \cdot \Delta t$
  $\gamma_t \leftarrow \text{ComputeSchedule}(\gamma_{max}, t)$ ▷ e.g., Parabolic or Sine
  $u \leftarrow \text{RequireGradient}(u)$
  $v_{\text{base}} \leftarrow v_\theta(u, t)$ ▷ Evaluate neural prior
  $g \leftarrow \nabla_u E(u)$ ▷ Compute implicit VJP via reverse-mode AD
  $\tilde{g} \leftarrow g \cdot \min\left(1, \frac{c}{\|g\|_2}\right)$ ▷ Apply gradient clipping
  $u \leftarrow \text{StopGradient}(u) + \Delta t\,(v_{\text{base}} - \gamma_t \tilde{g})$ ▷ Euler ODE update
end for
return Generated physical field u

parameterize the probability-flow ODE:

$$\frac{du}{dt} = v_\theta(u, t), \quad u(0) \sim p_0, \tag{2}$$

which continuously transports probability mass such that the marginal density $p_t(u)$ interpolates between p_0 and p_1. Utilizing an Optimal Transport (OT) formulation, the conditional vector field is defined along deterministic paths as $v_t(u|u_1) = u_1 - u_0$.

Operating directly in function space yields two structural advantages for PDE-constrained generation. First, it ensures resolution invariance: the learned field v_θ can be evaluated at arbitrary spatial resolutions during inference, mitigating numerical discretization errors when computing the physical constraint $h(u)$. Second, intermediate states u_t evolve as continuous, structurally coherent fields rather than unstructured discrete noise arrays, enabling the stable evaluation of spatial differential operators at any integration timestep.

3.3 Energy-Guided Transport

Formulating physical dynamics through energy or Hamiltonian landscapes has shown great success in forward modeling [51, 52]. We extend this principle to generative guidance, leveraging implicit differentiable physics [53]. To embed physical knowledge directly into the probability flow without breaking its continuous-time trajectory, we define a physical energy functional $E(u) = \frac{1}{2}\|h(u)\|_2^2$, where $h(u) \in \mathbb{R}^m$ is the discretized non-linear PDE residual. We augment the baseline neural vector field $v_\theta(u, t)$ by injecting the functional gradient of this energy, effectively interpreting it as a time-annealed, Gibbs-like log-prior:

$$\bar{v}_t(u) = v_\theta(u, t) - \gamma(t)\nabla_u E(u), \quad u(0) \sim p_0 \tag{3}$$

where $\gamma(t) \geq 0$ is a scheduling function acting as the inverse temperature $1/\tau(t)$. This thermodynamic cooling schedule allows the generative prior to dictate the macroscopic structure at high temperatures ($t \approx 0$) before strictly locking onto the physical manifold ($E(u) = 0$) at low temperatures ($t \to 1$).

Scalable Dissipative Force. Computing the guidance field $\nabla_u E(u)$ requires differentiating through complex, non-linear spatial operators. By the chain rule, $\nabla_u E(u) = J_h(u)^T h(u)$, where $J_h \in \mathbb{R}^{m \times N}$ is the constraint Jacobian. Algebraic projection methods (e.g., PCFM [23]) enforce constraints by explicitly materializing and solving systems involving $J_h J_h^T$. While highly memory-efficient for sparse boundaries (where $m \ll N$), this geometric approach incurs an intractable $\mathcal{O}(N^2)$ memory bottleneck when scaled to dense, full-domain PDE residuals ($m \approx N$). Our energy-based formulation naturally circumvents this limitation: rather than materializing J_h, the dissipative force is evaluated implicitly via standard reverse-mode automatic differentiation (Vector-Jacobian Products). This guarantees strict $\mathcal{O}(N)$ space and time complexity, effortlessly scaling full-PDE guidance to

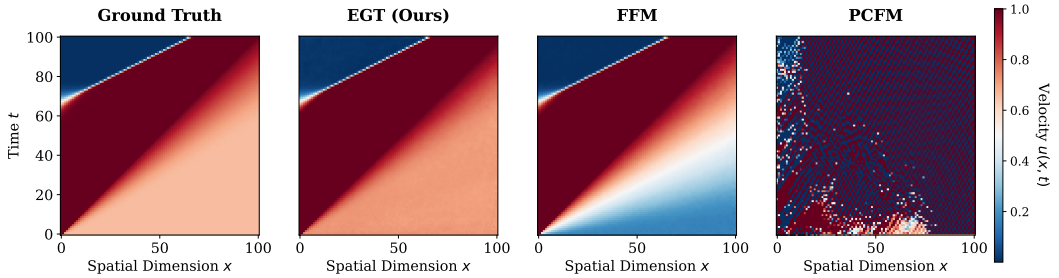


Figure 2: Qualitative evaluation on the 1D Burgers' equation (Full-PDE Stress Test). Spatiotemporal velocity fields $u(x,t)$ generated from the same initial conditions. Explicit algebraic projections (PCFM) can be sensitive to solver step sizes under non-linear shock dynamics, leading to numerical instabilities and high-frequency artifacts. In comparison, continuous Energy-Guided Transport provides smooth updates along the trajectory, accurately resolving sharp shockfronts while maintaining distributional fidelity.

high-dimensional fields without structural bottlenecks (see Appendix H for a complexity analysis). Furthermore, by relying exclusively on matrix-vector multiplications via VJPs rather than inverting local linear systems, EGT naturally handles ill-conditioned or rank-deficient Jacobians without numerical instability (producing zero gradients along rank-deficient directions). Unlike approximate Gauss-Newton or preconditioned iterative solvers that still require multiple inner-loop evaluations per step to approximate a projection, EGT embeds the physical correction directly into the transport dynamics via a single VJP evaluation.

Theoretical Guarantees. To show that Eq. (3) introduces a principled dissipative correction without arbitrarily perturbing the underlying data manifold, we establish the following propositions governing the continuous transport of probability mass (full analytical proofs are provided in Appendix A).

Proposition. ***(Continuous Energy Dissipation under Ideal Transport)*** *Let $\mathbb{E}_{p_t}[E(u)]$ denote the expected physical energy over the marginal distribution at time t. The instantaneous variation of the physical error under the augmented field decomposes into the baseline variation minus a strictly non-negative dissipative term:*

$$\frac{d\mathbb{E}_{total}}{dt} = \mathbb{E}_{p_t}\left[\langle \nabla_u E, v_\theta \rangle\right] - \gamma(t)\mathbb{E}_{p_t}\left[\|\nabla_u E\|_2^2\right] \tag{4}$$

Consequently, for any $\gamma(t) \geq 0$, the continuous injection strictly dampens the physical violation relative to the baseline ODE trajectory, effectively acting as a Lyapunov-like dissipative mechanism for the physical constraint.

While Proposition 3.3 holds identically for the generative flow, in practice, the neural field v_θ exhibits an approximation error $\delta(u,t)$ relative to the optimal transport field. By bounding this error such that $\|\delta(u,t)\| \leq \epsilon_t$, we formulate the global descent guarantees for the practical implementation:

Corollary. ***(Error-Bounded Global Energy Descent)*** *The true rate of change of the expected physical energy is bounded by the ideal transport variation and an error term proportional to the network's inaccuracy. Integrating this instantaneous rate over the generation interval $t \in [0,1]$ yields the macroscopic physical variation:*

$$\mathbb{E}[E(u_1)] - \mathbb{E}[E(u_0)] \leq \Delta E_{base} - \int_0^1 \gamma(t)\mathbb{E}_{p_t}\left[\|\nabla_u E\|_2^2\right] dt + \mathcal{O}(\bar{\epsilon}) \tag{5}$$

The instantaneous and cumulative forms of this bound are stated and proved as Corollaries 1 and 2 in Appendix A.

Scope of the Theoretical Guarantee. Importantly, Proposition 3.3 and Corollary 3.3 establish a continuous dissipative correction relative to the unguided flow under finite-step inference, rather than asymptotic convergence to $E(u_1) = 0$. Instantaneous energy decrease is guaranteed whenever the dissipative term $\gamma(t)\|\nabla_u E\|_2^2$ dominates the baseline alignment $\langle \nabla_u E, v_\theta \rangle$; if the learned flow strongly opposes the physical gradient, the residual may temporarily increase. When machine-precision constraint satisfaction is strictly required, a lightweight terminal refinement can be applied at $t = 1$, motivating our Hybrid EGT variant (see Section 3.5).

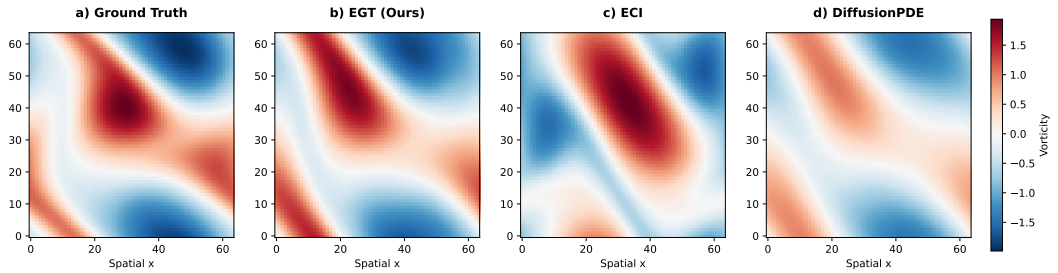


Figure 3: Qualitative evaluation on 2D Navier-Stokes vorticity fields. While all guided frameworks attempt to satisfy the underlying physical constraints, (c) ECI and (d) DiffusionPDE tend to introduce noticeable structural blurring and dampen the vorticity intensity. In contrast, our proposed (b) EGT better preserves the macroscopic structural coherence and the sharp gradients of the flow, offering a closer approximation to the (a) Ground Truth dynamics.

3.4 Architectural Synergies: FFM and Smooth Priors

While Proposition 3.3 theoretically justifies continuous guidance across any probability flow, its practical success is deeply synergistic with the structural properties of Functional Flow Matching. Unlike diffusion models, where intermediate states are heavily corrupted by stochastic noise, FFM constructs a deterministic ODE path. Consequently, intermediate states u_t rapidly emerge as structurally coherent fields. This coherence allows for the stable, direct evaluation of spatial differential operators (e.g., $\nabla u_t, \Delta u_t$) and the physical energy $E(u_t)$ along the generative trajectory, completely bypassing the need for terminal-state approximations.

Furthermore, this synergy can be naturally extended through the choice of the generative prior. While our dissipative guidance is highly effective under standard White Noise, utilizing a spatially correlated Gaussian Random Field (Smooth Kernel) ensures inherent spatial differentiability right from $t = 0$. By providing a well-behaved energy landscape from the very first integration step, this formulation allows for the stable evaluation of physical derivatives across the entire generative trajectory, fully unlocking the potential of continuous guidance without early-stage numerical truncation.

3.5 Practical Implementation and Sampling Algorithm

The complete generative inference process is summarized in Algorithm 1. To ensure stability across highly non-linear PDE landscapes without sacrificing the $\mathcal{O}(N)$ computational complexity, our implementation incorporates three core mechanisms: L^2-norm gradient clipping, dynamic guidance scheduling, and first-order Euler integration.

We formally define two variants of our framework. Our primary, natively continuous approach is denoted as EGT. Additionally, we evaluate Hybrid EGT, which incorporates an optional post-hoc refinement phase (pure energy descent at $t = 1$) to strictly anchor the final sample to the physical manifold. Rather than replacing terminal refinement, pure EGT replaces expensive repeated projections along the flow, steering the trajectory into a favorable physical "basin of attraction" that serves as an optimal initialization for lightweight post-hoc refinement. Unless otherwise specified, our primary visualizations highlight pure EGT. This isolates our core hypothesis: continuous guidance during the flow is fundamentally superior to discrete algebraic corrections. Comprehensive implementation details, hyperparameter configurations, and the exact formulation of the guidance schedules are provided in Appendix G.

4 Experimental Results

To rigorously evaluate our Energy-Guided Transport, we benchmark against state-of-the-art physics-guided generative models: ProFlow [49], ECI [48], DiffusionPDE [46], D-Flow [47], and PCFM [23]. We test across four physical systems: 1D Heat, 1D Reaction-Diffusion (RD), 1D Burgers', and 2D Navier-Stokes. To isolate their capabilities as pure PDE solvers and guarantee strict fairness, we adopt the zero-shot evaluation framework established by PCFM [23]. Specifically, all competing methods are denied access to interior ground-truth data and share the identical pre-trained FFM

Table 1: Quantitative comparison across PDE datasets (White Noise Prior). Panel A shows performance with weak constraints (BC/IC only). Panel B shows the stress test constrained by the Full PDE. For each dataset, we report Success Rate (**SR**, %), average inference speed (**Spd**, sec/sample), Sample MSE (**MSE**), and physical residual (**Phys**). DNC (Did Not Complete) denotes trajectories that encountered numerical divergence or memory exhaustion (NaN/OOM). Best results are highlighted in **bold**, and second-best are underlined (excluding the unguided Vanilla baseline). Values marked with † fall below the discrete ground-truth stencil precision limit (stencil overfitting/waterbed effect). See Appendix I for an analysis of the Waterbed Effect.

Method	Heat				RD				Burgers				Navier-Stokes			
	SR↑	Spd↓	MSE↓	Phys↓	SR↑	Spd↓	MSE↓	Phys↓	SR↑	Spd↓	MSE↓	Phys↓	SR↑	Spd↓	MSE↓	Phys↓
Panel A: Partial Physics Guidance (Boundary & Initial Conditions)																
Vanilla FFM	100	0.19	0.08	0.26	100	0.38	0.08	0.20	100	0.42	0.21	0.10	100	0.69	0.54	0.29
ProFlow	**100**	8.13	0.06	0.19	**100**	25.65	7.38e+05	1.16e+03	**100**	21.34	0.23	0.13	**100**	0.85	0.55	0.20
ECI	**100**	8.31	**0.01**	**0.01**	**100**	24.47	0.25	0.05	**100**	21.87	**0.11**	0.05	**100**	3.56	0.63	3.75e-05
DiffusionPDE	**100**	0.95	**0.01**	0.03	**100**	2.00	**0.02**	0.04	**100**	1.17	0.15	0.06	**100**	1.40	0.54	0.12
D-Flow	**100**	15.58	0.35	0.09	**100**	32.07	1.98e+02	0.74	**100**	39.62	0.31	0.08	14.2	DNC	DNC	DNC
PCFM	0.2	DNC	DNC	DNC	**100**	2.15	0.04	**1.11e-08**†	**100**	1.29	0.19	**1.19e-08**†	**100**	93.97	0.76	**1.18e-10**†
EGT (Ours)	**100**	**0.37**	0.02	0.05	**100**	**1.13**	0.05	0.07	**100**	**0.48**	0.21	0.05	**100**	**0.71**	**0.53**	0.01
Hybrid EGT (Ours)	**100**	0.40	0.02	0.02	**100**	1.24	0.05	8.51e-03	**100**	0.52	0.20	0.01	**100**	0.90	0.55	1.41e-04
Panel B: Full PDE Guidance (Stress Test)																
Vanilla FFM	100	0.14	0.08	6.24e-03	100	0.38	0.08	3.57e-03	100	0.27	0.20	5.13e-03	100	0.86	0.56	7.52e-03
ProFlow	0.4	DNC	DNC	DNC	**100**	25.05	5.00e+11	2.91e+05	**100**	18.68	0.24	6.27e-03	**100**	1.21	**0.55**	5.63e-03
ECI	**100**	10.28	**0.01**	**1.85e-03**	**100**	24.34	0.25	7.16e-05	**100**	18.68	**0.12**	1.95e-04	**100**	4.40	0.61	1.80e-03
DiffusionPDE	**100**	0.69	18.15	0.99	**100**	1.86	**0.02**	6.58e-04	**100**	1.14	0.15	9.46e-04	**100**	1.79	**0.55**	4.19e-03
D-Flow	**100**	18.06	23.67	0.03	**100**	31.44	0.17	6.58e-04	**100**	32.70	0.35	2.91e-03	1.6	DNC	DNC	DNC
PCFM	0	DNC	DNC	DNC	**100**	76.72	6.69e+16	1.34e+05	**100**	67.11	1.14e+17	4.76e+14	0	DNC	DNC	DNC
EGT (Ours)	**100**	**0.24**	0.05	5.12e-03	**100**	**0.89**	0.05	1.18e-03	**100**	**0.46**	0.19	1.92e-03	**100**	**0.94**	**0.55**	1.63e-03
Hybrid EGT (Ours)	**100**	**0.24**	0.05	3.97e-03	**100**	0.93	0.05	**1.95e-05**	**100**	0.54	0.19	**1.49e-04**	**100**	1.02	**0.55**	**2.89e-05**

backbone and initial noise realizations. Performance is evaluated via Success Rate (SR) for numerical stability, Inference Speed, Sample MSE for distributional fidelity, and Physical Residual (Phys) MAE. Full formulations and extended metrics are deferred to Appendices C and J. Table 1 summarizes quantitative performance under a standard White Noise prior, while Table 2 details the corresponding performance under a spatially correlated Smooth Kernel prior. We evaluate all methods under two constraint regimes: Partial Guidance (Boundary/Initial Conditions only) and the Full-PDE Stress Test. Comprehensive visualizations are provided in our Github repository.

Partial Physics Guidance. Under relaxed boundary constraints, geometric projection methods (e.g., ECI, PCFM) achieve near-zero physical errors on specific equations (RD, Burgers). However, their stability and efficiency are inconsistent across domains: PCFM drops to a 0.2% SR on the Heat equation and exhibits substantial latency on 2D Navier-Stokes (~ 94s per sample), while concurrently degrading the Sample MSE to 0.76 (compared to 0.54 for the Vanilla model). Similarly, terminal-state optimization via D-Flow struggles with high dimensionality, achieving only a 14.2% SR on Navier-Stokes. While other methods like DiffusionPDE and ProFlow maintain stability, they occasionally degrade the baseline distributional fidelity (e.g., DiffusionPDE increasing the Sample MSE to 0.74 on Navier-Stokes). In contrast, our continuous guidance maintains a 100% Success Rate across all environments, completing the 2D Navier-Stokes generation in just 0.71s while delivering competitive physical and visual fidelity.

Full PDE Guidance (The Stress Test). Constraining generation to satisfy the complete spatiotemporal PDE presents a challenging regime for guidance methods. Proximal formulations like ProFlow show sensitivity in this regime, yielding a 0.4% SR on the Heat equation and noticeable sample degradation on Reaction-Diffusion (Sample MSE reaching 5.0×10^{11}). Gradient-based optimization approaches (DiffusionPDE, D-Flow) technically avoid crashing on 1D grids but severely distort the data manifold, causing the Sample MSE on the Heat equation to spike to 18.15 and 23.67, respectively. For exact projection methods like PCFM, explicitly materializing the Jacobian for full-domain constraints results in Out-of-Memory errors or immediate numerical divergence (0% SR) on both Heat and Navier-Stokes systems. More broadly, comparing operating regimes across Tables 1 and 2 reveals a clear structural trade-off: explicit projection methods (ECI, PCFM) excel in low-dimensional, projection-friendly regimes (e.g., 1D Heat or Burgers Sample MSE), but incur heavy computational costs (4.40s–18.68s for ECI) or numerical divergence when constraints become dense and non-linear. Conversely, our EGT framework, stabilized by Gradient Norm Clipping and dynamic scheduling, reliably scales to the Full-PDE regime while preserving sharp shockfronts and vortex structures without high-frequency artifacts or blurring, as visually confirmed on 1D Burgers (Figure 2) and 2D Navier–Stokes (Figure 3). EGT maintains a 100% Success Rate globally with strictly sub-second inference speeds. On the chaotic 2D Navier-Stokes system, it safely navigates

Table 2: Quantitative comparison across PDE datasets using the **Smooth Kernel** prior. *Note*†*:* Evaluation on the 2D Navier-Stokes equations is omitted as the smooth kernel base distribution is currently not formulated for this specific environment. Values marked with † fall below the discrete ground-truth stencil precision limit (stencil overfitting/waterbed effect); see Appendix I. Best results are highlighted in **bold**, and second-best are underlined (excluding the unguided Vanilla).

Method	Heat				RD				Burgers			
	SR↑	Spd↓	MSE↓	Phys↓	SR↑	Spd↓	MSE↓	Phys↓	SR↑	Spd↓	MSE↓	Phys↓
Panel A: Partial Physics Guidance (Boundary & Initial Conditions)												
Vanilla FFM	100	0.19	0.26	0.43	100	0.38	0.09	0.22	100	0.41	0.16	0.13
ProFlow	**100**	8.16	0.05	0.18	**100**	25.60	0.19	0.24	**100**	21.41	0.16	0.08
ECI	0	DNC	DNC	DNC	**100**	24.43	0.36	0.09	**100**	21.94	0.14	0.07
DiffusionPDE	**100**	0.95	**0.04**	0.06	**100**	1.98	**9.37e-03**	0.03	**100**	1.16	**0.13**	0.05
D-Flow	**100**	15.64	24.89	0.97	**100**	31.43	0.09	0.03	**100**	39.29	10.58	0.37
PCFM	1.1	DNC	DNC	DNC	**100**	2.11	0.09	**1.11e-08**†	**100**	1.29	0.22	**1.23e-08**†
EGT (Ours)	**100**	**0.37**	0.05	0.05	**100**	**1.09**	0.01	0.03	**100**	**0.48**	0.14	0.02
Hybrid EGT (Ours)	**100**	0.43	0.05	**0.02**	**100**	1.25	9.96e-03	8.48e-03	**100**	0.52	0.14	4.58e-03
Panel B: Full PDE Guidance (Stress Test)												
Vanilla FFM	100	0.20	0.25	0.17	100	0.37	0.09	3.63e-03	100	0.38	0.16	5.76e-03
ProFlow	0	DNC	DNC	DNC	**100**	25.10	0.19	4.04e-03	**100**	17.42	0.16	4.49e-03
ECI	0	DNC	DNC	DNC	**100**	24.40	0.36	5.24e-04	**100**	17.08	0.14	3.08e-04
DiffusionPDE	**100**	1.04	**0.04**	0.08	**100**	1.86	**9.38e-03**	2.35e-04	**100**	1.74	**0.12**	1.46e-03
D-Flow	**100**	16.79	17.62	0.66	**100**	31.31	1.53	1.60e-03	**100**	28.46	14.45	0.05
PCFM	0	DNC	DNC	DNC	**100**	77.71	2.14e+17	5.91e+04	**100**	50.87	2.56e+16	5.74e+13
EGT (Ours)	**100**	**0.52**	**0.04**	7.19e-03	**100**	**0.89**	9.80e-03	3.60e-04	**100**	**0.95**	0.13	1.46e-03
Hybrid EGT (Ours)	**100**	0.55	**0.04**	**4.50e-03**	**100**	0.95	9.79e-03	**8.00e-05**	**100**	1.06	0.13	**1.85e-04**

the complex physical manifold in just 0.94s, reducing the physical residual to 1.63×10^{-3} (over $4\times$ lower than the Vanilla unguided model) while preserving the baseline distributional fidelity (Sample MSE 0.55).

The Smooth Kernel Prior: A Test of Robustness. To evaluate model stability against prior shifts, we replace standard Gaussian noise with a spatially correlated Smooth Kernel prior (Table 2). Learning the transport map from correlated noise is empirically harder for the base network, leading to less accurate unguided trajectories. Under this more challenging baseline landscape, projection- and proximal-based methods show increased sensitivity: PCFM, ECI, and ProFlow do not complete (0% Success Rate) on the Full-PDE Heat equation. Furthermore, discrete corrections can compound early-trajectory inaccuracies, resulting in degraded sample quality (Sample MSE reaching 2.1×10^{17} for PCFM on Reaction-Diffusion). Among the other compared guidance techniques, only DiffusionPDE maintains consistent stability across these 1D tasks, whereas D-Flow suffers from heavily degraded sample quality (Sample MSE spiking to 17.62 on the Heat equation). Our continuous framework demonstrates robust corrective power under these adverse conditions. In the Heat Full-PDE benchmark, Hybrid EGT effectively guides the less accurate baseline trajectory, reducing the physical residual by over an order of magnitude (from 0.17 down to 4.50e-3) and lowering the Sample MSE to 0.04. This confirms that continuous energy guidance provides substantial resilience against both complex physical constraints and variations in the generative prior.

4.1 Ablation Studies

To isolate the operational dynamics of our framework, we conduct targeted ablation studies. Figure 4 illustrates the sensitivity of Energy-Guided Transport (EGT) to the number of ODE integration steps (N) on the highly non-linear Full-PDE Burgers' equation, evaluated across 256 samples for $N \in \{10, 25, 50, 100, 250, 500\}$. Notably, the analysis reveals that EGT achieves its optimal stability-accuracy trade-off at a significantly lower discretization budget (e.g., $N \approx 50$). For the primary evaluations in Section 4, we deliberately adopted a higher, suboptimal integration budget (i.e., $N = 200$) to maintain a standardized evaluation against projection-based baselines (e.g., PCFM), which gen-

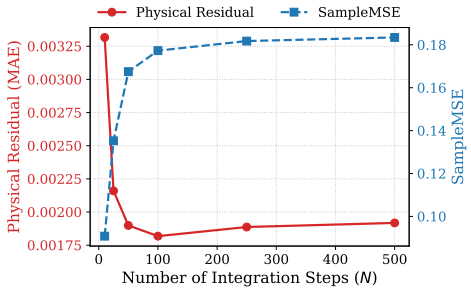


Figure 4: Impact of the number of ODE integration steps (N) on the 1D Burgers' equation (Full-PDE guidance). The dual-axis plot tracks Physical Residual MAE (red) and Sample MSE (blue). EGT stabilizes at low step counts ($N \approx 50$).

erally require finer temporal discretizations for numerical stability. This finding underscores that, in practical deployments, our continuous guidance can be substantially accelerated while concurrently improving distributional fidelity (Sample MSE).

Accuracy–Stability–Speed Pareto Analysis. To further verify whether EGT improves the overall operational trade-off rather than representing a single static operating point, we perform a comprehensive Pareto analysis across guidance strengths (γ) and scheduling profiles on the 1D Burgers and 1D Heat systems (detailed in Appendix K). The results confirm that EGT lies on a highly favorable accuracy–stability–speed frontier: it strictly dominates Vanilla FFM and DiffusionPDE by simultaneously improving physical accuracy and distributional stability at ~ 0.23–0.26 s/sample. While ECI achieves lower absolute residuals in projection-friendly regimes, it incurs a 40–$80\times$ higher computational cost. Importantly, EGT occupies a compact region in the runtime dimension, allowing practitioners to freely navigate the trade-off between physical residual and Sample MSE by tuning γ or the scheduler at an essentially constant, sub-second inference time.

Isolating Continuous Transport vs. Terminal Refinement. To verify that the performance of Hybrid EGT stems from our continuous transport rather than the terminal optimization itself, we apply the exact same post-hoc refinement procedure to the baseline trajectories on 1D Burgers and 2D Navier–Stokes (Table 16 in Appendix K). On 2D Navier–Stokes, refining unconditional FFM samples only reduces the physical MAE from 7.50e-3 to 3.96e-3, whereas Pure EGT already reaches 1.60e-3 prior to any refinement and drops to 2.90e-5 when refined, confirming that continuous guidance steers trajectories into a markedly superior physical basin of attraction.

5 Conclusion

In this work, we introduced Energy-Guided Transport (EGT), a zero-shot inference framework for physics-informed Functional Flow Matching. By formulating physical constraints as a smooth potential energy landscape, our approach maintains the continuous-time probability flow without requiring discrete geometric interruptions. Our evaluations demonstrate that while explicit algebraic projections can face numerical instability and computational overhead in non-linear domains, continuous dissipative guidance provides a scalable and stable alternative. We achieved up to a $130\times$ speedup over explicit projection techniques and a consistent 100% success rate across chaotic environments like 2D Navier-Stokes, achieving a highly favorable balance between physical consistency, computational efficiency, and distributional fidelity. Although evaluating the guidance field introduces a marginal per-step overhead compared to the vanilla unguided model, the framework's strict $\mathcal{O}(N)$ scalability ensures its applicability to arbitrary continuous domains.

Limitations and Future Work. While EGT provides a scalable alternative to explicit projections, several limitations define its operational scope. First, as a continuous dissipative mechanism rather than a hard algebraic constraint, EGT does not guarantee exact $E(u_1) = 0$ convergence without optional terminal refinement, and relies on the pretrained FFM to provide a structurally reasonable transport prior. Second, excessively large guidance scales (γ) increase ODE stiffness (e.g., $\gamma > 30$ on 2D Navier–Stokes). Third, because our current energy functional minimizes pointwise L^2 PDE residuals, it does not explicitly constrain global macroscopic statistics such as shock gradient distributions or energy spectra (analyzed in Appendix J.6). Since EGT natively supports any differentiable objective and interacts with the domain geometry solely via $h(u)$, future work will incorporate spectral and conservation-law penalties into $E(u)$ and leverage geometry-aware operators (e.g., GNO, SFNO) to scale EGT to unstructured 3D turbulence.

References

[1] Jiequn Han, Arnulf Jentzen, and Weinan E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 2018.

[2] Justin Sirignano and Konstantinos Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 2018.

[3] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

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

[5] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

[6] Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 2025.

[7] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 2021.

[8] Dmitrii Kochkov, Jamie A Smith, Ayya Alieva, Qing Wang, Michael P Brenner, and Stephan Hoyer. Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, 2021.

[9] Johannes Brandstetter, Daniel E Worrall, and Max Welling. Message passing neural pde solvers. In *International Conference on Learning Representations*, 2022.

[10] Steven L Brunton, Bernd R Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 2020.

[11] Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 2023.

[12] Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, et al. Learning skillful medium-range global weather forecasting. *Science*, 2023.

[13] Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

[14] C Jiang, R Wang, et al. Physics-informed machine learning: case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2021.

[15] Rui Wang, Karthik Kashinath, Mustafa Mustafa, Adrian Albert, and Rose Yu. Towards physics-informed deep learning for turbulent flow prediction. In *Conference on Knowledge Discovery and Data Mining*, 2020.

[16] Shengze Cai, Zhiping Mao, Zhicheng Wang, Minglang Yin, and George Em Karniadakis. Physics-informed neural networks (pinns) for fluid mechanics: A review. *Acta Mechanica Sinica*, 2021.

[17] Derek Hansen, Danielle C. Maddix, Shima Alizadeh, Gaurav Gupta, and Michael W. Mahoney. Learning physical models that can respect conservation laws. In *International Conference on Machine Learning*, 2023.

[18] S Chandra Mouli, Danielle C Maddix, Shima Alizadeh, Gaurav Gupta, Andrew Stuart, Michael W Mahoney, and Yuyang Wang. Using uncertainty quantification to characterize and improve out-of-domain learning for PDEs. In *International Conference on Machine Learning*, 2024.

[19] Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel De Bezenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for PDEs. In *Advances in Neural Information Processing Systems*, 2024.

[20] Georgios Batzolis, Jan Stanczuk, Carola-Bibiane Schönlieb, and Christian Etmann. Conditional image generation with score-based diffusion models. *arXiv preprint arXiv:2111.13606*, 2021.

[21] Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. In *Advances in Neural Information Processing Systems*, 2021.

[22] Ruizhi Deng, Marcus A Brubaker, Greg Mori, and Andreas Lehrmann. Continuous latent process flows. In *Advances in Neural Information Processing Systems*, 2021.

[23] Utkarsh Utkarsh, Pengfei Cai, Alan Edelman, Rafael Gomez-Bombarelli, and Christopher Vincent Rackauckas. Physics-constrained flow matching: Sampling generative models with hard constraints. In *Advances in Neural Information Processing Systems*, 2025.

[24] Xingchao Liu, Chengyue Gong, et al. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023.

[25] Gabriel Peyré and Marco Cuturi. *Computational optimal transport: With applications to data science*. Now Foundations and Trends, 2019.

[26] 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*, 2024.

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

[28] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, 2018.

[29] Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, 2024.

[30] Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. In *Advances in Neural Information Processing Systems*, 2020.

[31] Mevlana C Gemici, Danilo Rezende, and Shakir Mohamed. Normalizing flows on riemannian manifolds. *arXiv preprint arXiv:1611.02304*, 2016.

[32] Valentin De Bortoli, Emile Mathieu, Michael J. Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian score-based generative modelling. In *Advances in Neural Information Processing Systems*, 2022.

[33] Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M. Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

[34] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 2021.

[35] 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*, 2023.

[36] Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 2024.

[37] Jae Hyun Lim, Nikola B Kovachki, Ricardo Baptista, Christopher Beckham, Kamyar Azizzadenesheli, Jean Kossaifi, Vikram Voleti, Jiaming Song, Karsten Kreis, Jan Kautz, et al. Score-based diffusion models in function space. *Journal of Machine Learning Research*, 2025.

[38] Gavin Kerrigan, Giosue Migliorini, and Padhraic Smyth. Functional flow matching. In *International Conference on Artificial Intelligence and Statistics*, 2024.

[39] Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 2019.

[40] Maziar Raissi, Alireza Yazdani, and George Em Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 2020.

[41] Xiaowei Jin, Shengze Cai, Hui Li, and George Em Karniadakis. NSFnets (Navier-Stokes flow nets): Physics-informed neural networks for the incompressible navier-stokes equations. *Journal of Computational Physics*, 2021.

[42] Aditi Krishnapriyan, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems*, 2021.

[43] Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 2021.

[44] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

[45] Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *International Conference on Learning Representations*, 2024.

[46] Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. Diffusionpde: Generative PDE-solving under partial observation. In *Advances in Neural Information Processing Systems*, 2024.

[47] Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-flow: Differentiating through flows for controlled generation. In *International Conference on Machine Learning*, 2024.

[48] Chaoran Cheng, Boran Han, Danielle C Maddix, Abdul Fatir Ansari, Andrew Stuart, Michael W Mahoney, and Bernie Wang. Gradient-free generation for hard-constrained systems. In *International Conference on Learning Representations*, 2025.

[49] Zichao Yu, Ming Li, Wenyi Zhang, Difan Zou, and Weiguo Gao. Proflow: Zero-shot physics-consistent sampling via proximal flow guidance. *arXiv preprint arXiv:2601.20227*, 2026.

[50] Lawrence C Evans. *Partial Differential Equations*. American Mathematical Society, 2022.

[51] Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, 2019.

[52] Jack Richter-Powell, Yaron Lipman, and Ricky TQ Chen. Neural conservation laws: A divergence-free perspective. In *Advances in Neural Information Processing Systems*, 2022.

[53] Philipp Holl, Nils Thuerey, and Vladlen Koltun. Learning to control pdes with differentiable physics. In *International Conference on Learning Representations*, 2020.

A Proofs of Energy Dissipation and Error Bounds

In this Appendix, we provide the formal proofs for the Dissipation Proposition and its associated corollaries presented in Section 3.3. The derivations are formulated within the deterministic Flow Matching framework, utilizing a Lagrangian (pathwise) perspective to track the continuous probability mass transport in infinite-dimensional spaces.

A.1 Proof of Proposition 1: Energy Dissipation (Ideal Case)

Let $\mathcal{H}$ be the infinite-dimensional Hilbert space representing the state space of the generated physical fields. Under the Functional Flow Matching (FFM) paradigm, the continuous temporal evolution of a generated state is governed by a deterministic ordinary differential equation (ODE) $\dot{u}_t = \bar{v}_t(u_t)$. Let $\Phi_t : \mathcal{H} \to \mathcal{H}$ denote the corresponding flow map, such that $u_t = \Phi_t(u_0)$ with the initial state drawn from the base prior $u_0 \sim p_0$.

In our continuous guidance framework, the augmented velocity field is defined as:

$$\bar{v}_t(u) = v_\theta(u, t) - \gamma(t)\nabla_u E(u) \tag{6}$$

where $\nabla_u E(u)$ denotes the Fréchet derivative (gradient) of the physical energy functional with respect to the field u.

We assume that the neural vector field v_θ and the energy gradient $\nabla_u E(u)$ satisfy mild local Lipschitz conditions such that the ODE admits a unique global solution $u_t \in \mathcal{H}$ almost everywhere. Furthermore, we assume sufficient regularity to permit the exchange of expectations and time derivatives via the Dominated Convergence Theorem.

We define the expected physical energy at time t over the marginal density $p_t(u)$. Using the Law of the Unconscious Statistician, we express this expectation by pulling it back to the base prior p_0 via the flow map:

$$\mathbb{E}_{p_t}[E(u)] = \mathbb{E}_{p_0}[E(\Phi_t(u_0))] \tag{7}$$

Applying the functional chain rule in $\mathcal{H}$, the derivative of the expected energy with respect to the generative time t is:

$$\frac{d}{dt}\mathbb{E}_{p_0}[E(\Phi_t(u_0))] = \mathbb{E}_{p_0}\left[\left\langle \nabla_u E(\Phi_t(u_0)), \frac{d}{dt}\Phi_t(u_0)\right\rangle_{\mathcal{H}}\right] \tag{8}$$

Substituting the ODE $\frac{d}{dt}\Phi_t(u_0) = \bar{v}_t(\Phi_t(u_0))$ and pushing the expectation forward back to the marginal distribution p_t, we obtain the instantaneous variation of the expected energy:

$$\frac{d}{dt}\mathbb{E}_{p_t}[E(u)] = \mathbb{E}_{p_t}\left[\langle \nabla_u E(u), \bar{v}_t(u)\rangle_{\mathcal{H}}\right] \tag{9}$$

Substituting the explicit formulation of our augmented vector field into the inner product:

$$\frac{d}{dt}\mathbb{E}_{p_t}[E(u)] = \mathbb{E}_{p_t}\left[\langle \nabla_u E(u), v_\theta(u, t) - \gamma(t)\nabla_u E(u)\rangle_{\mathcal{H}}\right] \tag{10}$$

$$= \mathbb{E}_{p_t}\left[\langle \nabla_u E(u), v_\theta(u, t)\rangle_{\mathcal{H}}\right] - \gamma(t)\mathbb{E}_{p_t}\left[\|\nabla_u E(u)\|^2_{\mathcal{H}}\right] \tag{11}$$

Let the first term be defined as $\frac{d\mathbb{E}_{\text{base}}}{dt} \triangleq \mathbb{E}_{p_t}[\langle \nabla_u E(u), v_\theta(u, t)\rangle_{\mathcal{H}}]$, representing the natural energy variation induced by the unguided model. Since $\gamma(t) \geq 0$ and the expected squared norm is strictly non-negative, the second term acts as a strictly dissipative force:

$$\frac{d\mathbb{E}_{\text{total}}}{dt} \leq \frac{d\mathbb{E}_{\text{base}}}{dt} \tag{12}$$

This concludes the proof.

A.2 Proof of Corollary 1: Error-Bounded Dissipation

In practice, the neural vector field $v_\theta(u, t)$ is an approximation of the ideal target field $v^*(u, t)$. We define the approximation error as $\delta(u, t) = v_\theta(u, t) - v^*(u, t)$. In the context of Flow Matching, the

network is trained to minimize the expected mean squared error; thus, we bound the error in $L^2(p_t)$ norm: $\epsilon_t^2 \triangleq \mathbb{E}_{p_t}[\|\delta(u,t)\|_{\mathcal{H}}^2]$.

Substituting $v_\theta(u,t) = v^*(u,t) + \delta(u,t)$ into the expectation of the base variation defined in Proposition 1:

$$\mathbb{E}_{p_t}\left[\langle \nabla_u E(u), v_\theta(u,t)\rangle_{\mathcal{H}}\right] = \mathbb{E}_{p_t}\left[\langle \nabla_u E(u), v^*(u,t)\rangle_{\mathcal{H}}\right] + \mathbb{E}_{p_t}\left[\langle \nabla_u E(u), \delta(u,t)\rangle_{\mathcal{H}}\right] \tag{13}$$

We now bound the spurious error term by applying the generalized Cauchy-Schwarz inequality for expectations:

$$\mathbb{E}_{p_t}\left[\langle \nabla_u E(u), \delta(u,t)\rangle_{\mathcal{H}}\right] \leq \sqrt{\mathbb{E}_{p_t}\left[\|\nabla_u E(u)\|_{\mathcal{H}}^2\right]}\sqrt{\mathbb{E}_{p_t}\left[\|\delta(u,t)\|_{\mathcal{H}}^2\right]} \tag{14}$$

Defining the expected gradient energy landscape as $V(t) \triangleq \mathbb{E}_{p_t}[\|\nabla_u E(u)\|_{\mathcal{H}}^2]$, we obtain:

$$\mathbb{E}_{p_t}\left[\langle \nabla_u E(u), \delta(u,t)\rangle_{\mathcal{H}}\right] \leq \sqrt{V(t)}\epsilon_t \tag{15}$$

Substituting this bound back into the full dissipation equation establishes Corollary 1:

$$\frac{d\mathbb{E}_{\text{total}}}{dt} \leq \frac{d\mathbb{E}_{v^*}}{dt} - \gamma(t)V(t) + \sqrt{V(t)}\epsilon_t \tag{16}$$

A.3 Proof of Corollary 2: Global Energy Descent

To evaluate the macroscopic effect of the continuous guidance, we integrate the instantaneous rate of change derived in Corollary 1 over the entire generation process, from $t = 0$ to $t = 1$:

$$\int_0^1 \frac{d}{dt}\mathbb{E}_{p_t}[E(u)]dt \leq \int_0^1 \frac{d\mathbb{E}_{v^*}}{dt}dt - \int_0^1 \gamma(t)V(t)dt + \int_0^1 \sqrt{V(t)}\epsilon_t dt \tag{17}$$

By the Fundamental Theorem of Calculus, the left-hand side evaluates exactly to the total change in expected physical energy:

$$\mathbb{E}_{p_1}[E(u_1)] - \mathbb{E}_{p_0}[E(u_0)] \tag{18}$$

Defining the cumulative natural energy variation under the ideal transport field as $\Delta E_{\text{base}} = \int_0^1 \frac{d\mathbb{E}_{v^*}}{dt}dt$, and the exact cumulative neural approximation error as $\mathcal{E}_{\text{err}} = \int_0^1 \sqrt{V(t)}\epsilon_t dt$, we obtain Corollary 2:

$$\mathbb{E}_{p_1}[E(u_1)] - \mathbb{E}_{p_0}[E(u_0)] \leq \Delta E_{\text{base}} - \int_0^1 \gamma(t)V(t)dt + \mathcal{E}_{\text{err}} \tag{19}$$

This confirms that the cumulative dissipative force must be scaled via the scheduling parameter $\gamma(t)$ to robustly mitigate the cumulative model error $\mathcal{E}_{\text{err}}$, ensuring a net reduction in the overall PDE violation. Note that if the neural approximation error is uniformly bounded by $\sup_t \epsilon_t \leq \epsilon$, the cumulative error scales as $\mathcal{O}(\epsilon)$. while the physical energy landscape $E(u)$ may be highly non-convex for non-linear PDEs (as discussed in Appendix B), the local dissipativity guaranteed by the inner product $\langle \nabla_u E, -\nabla_u E\rangle \leq 0$ holds universally. Unlike explicit algebraic projections that diverge in non-convex topologies, our continuous gradient injection safely ensures energy descent along the probability path without requiring global convexity.

B Mathematical Taxonomy of Constraints

Integrating physical knowledge into zero-shot generative models requires defining a set of constraint functions $h(u) = 0$. As referenced in Section 3.1, we categorize these constraints based on their spatial scale, linearity, and the topological complexity of their resulting energy landscape $E(u) = \frac{1}{2}\|h(u)\|^2$. Table 3 summarizes the mathematical properties of the constraints evaluated in this work.

1. **Initial and Boundary Conditions (IC/BC):** These constraints define the admissibility and uniqueness of the solution. While essential, they only constrain the field at $t = 0$ or on the spatial boundary $\partial\Omega$, leaving the internal spatio-temporal evolution entirely to the learned prior of the generative model. Mathematically, they are typically formulated as linear constraints (e.g., Dirichlet or standard Neumann conditions), though non-linear boundary physics can occur.

2. **Macroscopic Constraints (Conservation Laws):** Often expressed as global integrals over the domain, such as $\int_\Omega u(\mathbf{x}, t)d\mathbf{x} = C$, these constraints enforce the conservation of macroscopic invariants (e.g., total mass or momentum). Because the integral operator is a strictly linear functional mapping, its associated energy landscape is strictly convex. This guarantees that explicit algebraic projection methods can find a unique, global minimum efficiently. However, acting globally, they lack the spatial resolution required to correct local structural dynamics.
3. **Full PDE Residual:** This represents the most rigorous microscopic constraint, requiring the governing differential equation to hold locally at every coordinate $(\mathbf{x}, t)$. The topological complexity of this constraint is entirely dictated by the nature of the spatial differential operator $\mathcal{N}(u)$. For linear differential equations (e.g., the Heat equation), the resulting energy functional remains strictly convex. Conversely, for non-linear advective systems (e.g., Burgers' or Navier-Stokes equations), the residual forms a highly non-convex energy landscape. In these non-linear regimes, exact explicit projections become computationally intractable and structurally fragile, necessitating the continuous dissipative guidance proposed in our framework.

Constraint Type	**Analytical Form** $h(u) = 0$	**Spatial Scope**	**Linearity / Landscape**	**Generative Impact**
Initial/Boundary	$\mathcal{B}(u) - g(\mathbf{x}, t) = 0$	$\partial\Omega \cup \{t = 0\}$	Typically Linear (Convex)	Anchors the domain; ensures uniqueness.
Macroscopic	$\int_\Omega u\, d\mathbf{x} - C = 0$	Global (Ω)	Strictly Linear (Convex)	Enforces global invariants; blind to local topology.
Full PDE	$\partial_t u + \mathcal{N}(u) = 0$	Microscopic ($\forall \mathbf{x}, t$)	System-Dependent	Dictates true local dynamics; implicitly ensures global conservation.

Table 3: Theoretical taxonomy of physical constraints in generative modeling. The structural linearity of the physical operator dictates the convexity of the resulting energy landscape. This topological property drastically impacts the computational feasibility of applying exact algebraic corrections during the inference process.

C Dataset Discretization and Numerical Generation

To ensure a fair and rigorous baseline comparison, all datasets used in this work follow the exact numerical generation protocols established by the Physics-Constrained Flow Matching (PCFM) framework [23]. The datasets are mathematically identical to those in PCFM, ensuring that performance gains stem strictly from the generative methodology.

1. Heat Equation (1D): We consider the linear diffusion equation $\partial_t u = \alpha \partial_{xx} u$. Instead of a numerical approximation, we generate samples using the exact analytical solution:

$$u(x, t) = \sin(x + \varphi)e^{-\alpha t} \tag{20}$$

where the diffusion coefficient is $\alpha \sim \mathcal{U}(1.0, 5.0)$ and the phase is $\varphi \sim \mathcal{U}(0, \pi)$. The domain is discretized on a 100×100 spatiotemporal grid ($x \in [0, 2\pi], t \in [0, 1]$) with periodic boundary conditions. This provides a high-fidelity ground truth with zero discretization error.

2. Reaction-Diffusion (1D): This system is governed by the non-linear equation $\partial_t u - \nu \partial_{xx} u - \rho u(1 - u) = 0$, with parameters set to $\rho = 0.01$ and $\nu = 0.005$. Data is generated via a semi-implicit finite difference scheme on a 128×100 grid over $x \in [0, 1]$. We construct 6,400 solutions by pairing 80 initial conditions with 80 boundary conditions.

3. Inviscid Burgers' Equation (1D): We solve the inviscid form $\partial_t u + u\partial_x u = 0$, which is highly prone to shock formation. To handle discontinuities, we employ a Godunov finite volume method. The initial conditions are smoothed step functions $u(x, 0) = (1 + \exp((x - p)/\epsilon))^{-1}$ with steepness $\epsilon = 0.02$ and the center location sampled uniformly as $p \sim \mathcal{U}(0.2, 0.8)$. The grid is 101×101 over $x, t \in [0, 1]$.

4. Navier-Stokes Equations (2D): We use the incompressible Navier-Stokes equations in vorticity form: $\partial_t \omega + (u \cdot \nabla)\omega = \nu \Delta \omega + f$. We set the viscosity to $\nu = 10^{-3}$ and use a Crank-Nicolson spectral solver on a 64×64 periodic grid, recording 50 temporal snapshots. The dataset comprises 10,000 simulations generated by sampling 100 random initial vorticities and 100 forcing phases.

D Generative Kernels and Prior Distributions

The choice of the prior distribution $p_0(u)$ is critical in Functional Flow Matching, as it defines the starting point of the probability flow and dictates the initial geometry of the sample paths. While white noise is the standard setting for discrete diffusion models, embedding spatial smoothness directly into the functional prior is fundamental for the stable evaluation of high-order spatial physical operators. In our experiments, we compare two distinct classes of priors:

1. White Noise Prior: This represents the uncorrelated generative setting, where the initial state is sampled as independent and identically distributed (i.i.d.) Gaussian noise across all spatial grid points, effectively mimicking $u_0 \sim \mathcal{N}(0, \mathbf{I})$. For the 2D Navier-Stokes equations, we exclusively utilize this prior due to the extreme dimensionality of the full spatio-temporal grid ($64 \times 64 \times 50$), which makes the instantiation and inversion of full dense covariance matrices computationally prohibitive for spatially correlated priors. Visualized in Figure 5a, it presents as high-frequency noise lacking any spatial structure.

2. Smooth Kernel Prior (Gaussian Random Fields): To investigate the impact of spatial correlation on continuous physical guidance, we employ Gaussian Random Fields (GRFs) as functional priors for the 1D equations. Unlike white noise, GRFs ensure that the initial state u_0 possesses a predefined degree of spatial correlation, guaranteeing spatial differentiability right from $t = 0$. This property provides a well-behaved energy landscape for guiding derivatives ($\partial_x, \partial_{xx}$).

- **Heat Equation:** We utilize a Matérn covariance kernel ($\nu = 3/2$) with a characteristic correlation length $\ell = 0.2$, promoting smoother and larger initial spatial features (Figure 5b).
- **Reaction-Diffusion and Burgers' Equations:** We utilize a Squared Exponential kernel with a tighter correlation length $\ell = 0.05$ (Figure 5c), allowing for the representation of sharper initial functional shapes.

Figure 5 provides a qualitative 2D visualization of samples drawn from these functional prior distributions, clearly highlighting how the characteristic correlation length dictates the smoothness and frequency spectrum of the initial generative state.

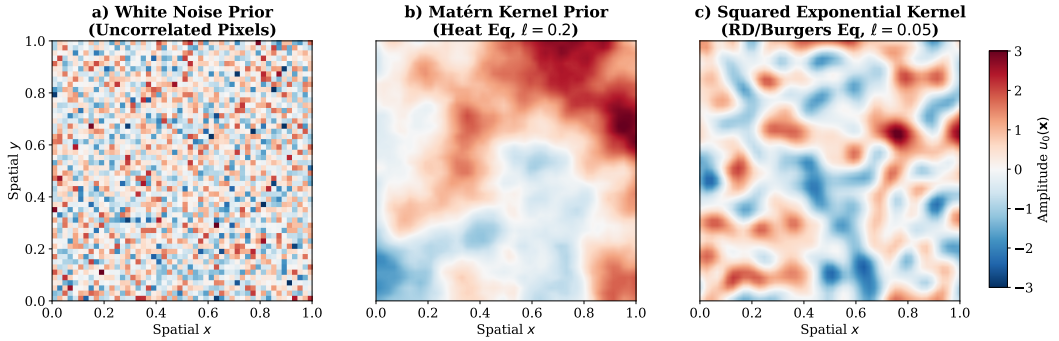


Figure 5: Qualitative 2D visual comparison of standard samples drawn from different generative functional prior distributions $p_0(u)$, visualized as spatial maps over the domain $[0, 1] \times [0, 1]$. Panel a) shows white noise, characterized by uncorrelated pixels that create local high-frequency content. Panels b) and c) show functional smooth kernel priors (Gaussian Random Fields), highlighting how the characteristic length scale ℓ dictates the typical size and smoothness of initial functional structures.

E Architecture and Training Details

FNO Backbone. All generative models in this work parameterize the underlying time-dependent vector field $v_\theta(u_t, t)$ using a Fourier Neural Operator (FNO) backbone. The FNO encoder takes as input the concatenation of the current state u_t, a sinusoidal positional embedding for the spatial grid, and a Fourier time embedding to capture the non-stationary nature of the probability flow.

To maintain strict consistency with the baseline setups, we tailor the architecture capacity to the dimensionality of the problem:

- **1D Systems (Heat, Reaction-Diffusion, Burgers):** The architecture consists of 4 Fourier layers preserving 32 modes across spatial and temporal dimensions, with 64 hidden channels and 256 projection channels.
- **2D Navier-Stokes:** Due to the high-dimensional nature of the $64 \times 64 \times 50$ spatio-temporal grid, we utilize a 3D FNO equipped with 2 Fourier layers, preserving 16 modes per dimension, with 32 hidden channels and 256 projection channels.

Training Regimen. We pretrained 7 distinct Functional Flow Matching (FFM) models to comprehensively evaluate our continuous guidance against unconstrained baselines under different prior assumptions (White Noise vs. Smooth GRFs). All models were trained using the Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$, zero weight decay) with an initial learning rate of 3×10^{-4}. We employed a Reduce-LROnPlateau scheduler with a decay factor of 0.5, a patience of 10 validation steps, and a minimum learning rate of 1×10^{-4}.

The batch sizes were set to 256 for the 1D environments and 24 for the 2D Navier-Stokes environment. Training was performed on clusters equipped with NVIDIA RTX 8000, A6000 and L40 GPUs. The 1D models converged after 20,000 iterations, while the Navier-Stokes model required 500,000 training steps to adequately capture the complex fluid dynamics. Table 4 summarizes the specific configurations for each of the 7 pretrained models.

Computational Bottleneck of Spatially Correlated Priors. While smooth GRFs were utilized for 1D systems, the 2D Navier-Stokes environment (64×64 spatial grid with 50 temporal snapshots) presents a prohibitive computational bottleneck. Instantiating a spatially correlated Gaussian prior requires computing and decomposing a dense covariance matrix of size $N \times N$, where N is the total number of grid points per sample. For Navier-Stokes, $N = 64 \times 64 \times 50 = 204,800$, resulting in a dense covariance matrix with over 4.19×10^{10} elements. Even in single precision (FP32), storing this matrix for a single instance would require approximately 167 GB of VRAM, far exceeding the capacity of standard modern GPUs, thereby restricting our evaluation to the uncorrelated White Noise prior for this dataset.

Table 4: Architectural and optimization hyperparameters for the 7 pre-training environments.

PDE System	Prior Type	Layers	Modes	Channels	Batch Size	Steps
Heat 1D	White / Smooth	4	32	64	256	20k
RD 1D	White / Smooth	4	32	64	256	20k
Burgers 1D	White / Smooth	4	32	64	256	20k
Navier-Stokes 2D	White	2	16	32	24	500k

F Baseline Descriptions and Implementation Details

To rigorously evaluate the efficacy of our proposed continuous dissipative guidance, we benchmarked it against five representative state-of-the-art methodologies for physical constraint enforcement and generative guidance. The theoretical and implementation details of each baseline are outlined below.

1. Physics-Constrained Flow Matching (PCFM): PCFM [23] enforces nonlinear physical constraints by applying an explicit algebraic projection at each solver step. The framework relies on a first-order Taylor expansion of the physical constraint functional $h(u)$. Given the unconstrained terminal state prediction $\hat{u}_1$, PCFM finds a corrected state u by solving the linearized system $h(\hat{u}_1) + J(u - \hat{u}_1) = 0$, where $J = \nabla h(\hat{u}_1)$ is the constraint Jacobian. This maps the sample onto the tangent space of the constraint manifold via the Moore-Penrose pseudo-inverse. While theoretically exact for linear equations, this projection requires computing and applying $J^\dagger$, leading to the severe $\mathcal{O}(N^2)$ memory bottlenecks and computational overheads discussed in Section 4 during our full-PDE stress tests.

2. ProFlow (Proximal Flow Guidance): ProFlow [49] is a recent framework (introduced in Jan 2026) specifically designed to address the computational weight and numerical instability of exact projection-based methods like PCFM. Instead of solving a strict linearized system via matrix inversion, it utilizes K unrolled proximal gradient steps on the predicted terminal state $\hat{u}_1$ to softly enforce physics consistency. Most importantly, it then applies a linear interpolation step to project the refined state back onto the probability path, attempting to avoid entirely disrupting the learned prior. *Note:* At the time of writing, the official source code for ProFlow was not publicly available; therefore, we provided an independent and faithful implementation based strictly on the algorithmic description and hyperparameter settings ($K = 3$ proximal steps) detailed in the original manuscript.

3. DiffusionPDE: DiffusionPDE [46] utilizes the gradient of a physical residual or observation loss to softly guide the generative process. Originally formulated for score-based diffusion models, it relies on obtaining a denoised estimate $\hat{u}_0$ and computing the guidance gradient directly on it. In the context of Flow Matching, this baseline is adapted by computing the gradient of the physical energy functional directly on the deterministic prediction of the terminal state $\hat{u}_1$. Unlike our proposed dissipative guidance, which applies a continuous local force inherently coupled with the intermediate state u_t, this method relies entirely on the accuracy of the one-step prediction to anchor the physical correction.

4. Extrapolation, Correction, and Interpolation (ECI): ECI [48] is a gradient-free, zero-shot framework designed for hard-constrained generation. It iteratively alternates between three stages during flow sampling: extrapolating the intermediate state to a terminal prediction, applying an exact algebraic correction (e.g., orthogonal or oblique projection) to strictly satisfy the constraint, and interpolating the corrected state back to the flow probability path. While highly effective for simple linear or non-overlapping constraints where closed-form projections exist, its reliance on analytical corrections restricts its extensibility to dense, highly non-linear constraints, such as full-domain PDE residuals.

5. D-Flow: D-Flow [47] utilizes a flow-based generative prior coupled with an inner-loop optimization routine to satisfy physical constraints. Instead of modifying the intermediate sampling states, it attempts to optimize the initial noise u_0 by backpropagating the constraint violation loss through the entire unrolled ODE trajectory.

G Inference Implementation Details

To maintain stability across the highly non-convex PDE landscapes discussed in Section 3.5, our guided inference algorithm incorporates specific numerical mechanisms and carefully tuned hyperparameters to establish a fair evaluation against the baselines.

G.1 Numerical Stabilization Mechanisms

Gradient Clipping The non-linear operators within the full PDE residual can produce gradients of highly variable magnitude, particularly in the presence of unrefined early-stage structures or sharp physical shocks. To prevent these gradients from dominating the neural vector field v_θ and destabilizing the ODE solver, we employ L^2-norm gradient clipping. For a given clip threshold c, the gradient g is scaled as $\tilde{g} = g \cdot \min(1, c/\|g\|_2)$. Unlike global gradient normalization, this strategy bounds the maximum corrective force while preserving the natural decay of the gradient magnitude as the sample converges toward the physical manifold.

Transient Guidance Scheduling At $t \approx 0$, the state u_t is highly noisy, making the evaluation of spatial physical operators uninformative. Conversely, as $t \to 1$, the formation of sharp physical features can lead to numerical instability in spatial PDE gradients. To address both extremes, we employ a dynamic scheduling function $\gamma(t)$ that implements a transient guidance strategy. In our main experiments, we utilize a sine schedule ($\gamma(t) = \gamma_{max} \sin(\pi t)$, as illustrated in Figure 6). This schedule explicitly zeroes out the physics penalty at $t = 0$ to allow the unguided neural prior to establish macroscopic structures, applies the strongest physical correction at $t = 0.5$, and smoothly decays to zero as $t \to 1$ to prevent gradient explosion and preserve the generative fine-tuning. Empirically, the choice of scheduler significantly influences generative fidelity, particularly for

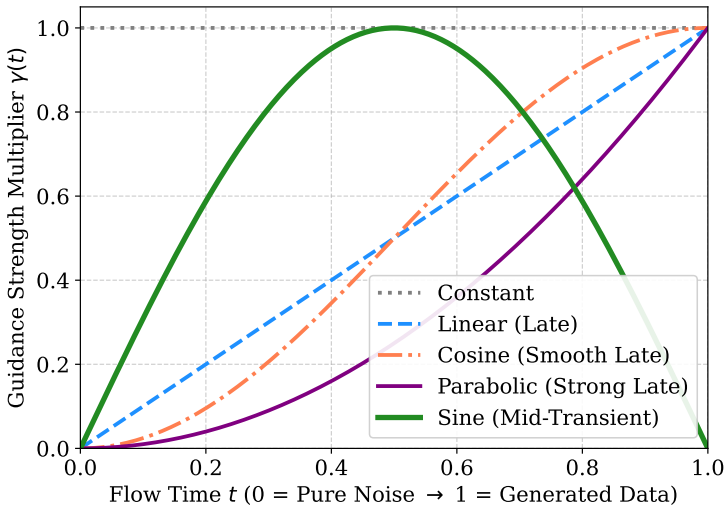


Figure 6: Comparison of guidance scheduling profiles $\gamma(t)$ as a function of the probability flow time $t \in [0, 1]$. While traditional 'Late Guidance' schedules (Linear, Parabolic, Cosine) monotonically increase to apply maximum physical correction at the terminal state, our proposed 'Transient' Sine schedule peaks at $t = 0.5$ and decays to zero at $t = 1$. This mitigates numerical instabilities caused by evaluating non-linear physical operators on sharp fluid structures during the final stages of the generative process.

complex PDE dynamics. A comprehensive ablation study comparing different scheduling profiles is provided in Appendix K.

Efficient ODE Integration While higher-order ODE solvers (e.g., Runge-Kutta) are commonly used in unguided flow matching, they require multiple evaluations of the vector field per time step. In an energy-guided framework, each evaluation necessitates a forward-backward pass through complex PDE operators to compute the implicit VJP ($\nabla_u E(u_t)$), making higher-order solvers computationally expensive. By leveraging the inherent smoothness of FFM alongside gradient clipping and transient scheduling, the augmented probability flow is sufficiently stabilized to be integrated using a first-order Euler method. This configuration ensures an $\mathcal{O}(N)$ computational complexity, maintaining efficient inference times compared to projection-based baselines.

G.2 Inference Hyperparameters and Baseline Tuning

To ensure a fair zero-shot evaluation, all sampling techniques are initialized with identical noise realizations. The foundational integration parameters are summarized in Table 5. The hyperparameters for PCFM, ECI, and D-Flow follow the configurations established by Utkarsh et al. [23]. ProFlow is configured according to its original manuscript [49], and DiffusionPDE [46] is tuned specifically to maintain numerical stability across the evaluated PDE systems.

Table 5: Shared integration hyperparameters across all physical systems. NFE denotes the Number of Function Evaluations (forward integration steps).

Shared Parameter	**Heat (1D)**	**RD (1D)**	**Burgers (1D)**	**Navier-Stokes (2D)**
ODE Solver Scheme	Forward Euler	Forward Euler	Forward Euler	Forward Euler
NFE (Steps)	100	200	200	100
Base Flow Prior	White / Smooth	White / Smooth	White / Smooth	White

Energy guided transport (Ours): The dissipative energy scale γ follows a sine schedule. For 1D systems, we set $\gamma_{max} = 75$, while for the 2D Navier-Stokes equations, we reduce the scale to $\gamma_{max} = 25$ to ensure stability. An optional post-hoc refinement phase is applied at $t = 1$ to further

anchor the sample to the physical manifold. By default, this involves 100 gradient descent steps with a learning rate of $\eta = 0.1$. For the "Heat Full PDE" experiment, we use 50 steps with $\eta = 0.01$, and for the "Heat Partial Guidance" experiment, we use 200 steps with $\eta = 0.1$. The gradient norm clipping threshold is fixed at $c = 1.0$ across all experiments.

PCFM Specifics: We strictly adopt the official configurations from [23]. Specifically, guided interpolation ($\lambda = 1.0$, step size 0.01, 20 refinement steps) is restricted exclusively to the linear Heat equation, whereas non-linear PDEs employ a single Gauss-Newton step per integration step without guided interpolation. To maximize stability, we retain the official implementation's default Tikhonov regularization parameter ($\epsilon = 10^{-6}$) for inverting the Schur complement and execute projections in `float64` precision. As noted independently by [49], explicit Jacobian-based projections can still exhibit numerical divergence when scaled to dense, non-linear full-PDE constraints.

Optimization-Based Baselines (DiffusionPDE & D-Flow): These baselines optimize a composite loss function $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda_{\text{phys}}\mathcal{L}_{\text{phys}}$. Following the evaluation protocol in [23], we set $\lambda_{\text{phys}} = 0.01$. We observe that these methods are highly sensitive to this hyperparameter: larger values can lead to severe manifold degradation, while smaller values fail to adequately enforce the PDE constraint.

- **DiffusionPDE:** Uses a guidance step size of $\eta = 0.01$. We adopt a more conservative step size than the original formulation to improve numerical stability and prevent divergence across the benchmarked PDEs.
- **D-Flow:** Performs $N_{iter} = 20$ L-BFGS inner optimization steps with a learning rate of $\eta = 1.0$.

Proximal and Projection Baselines (ECI & ProFlow):

- **ECI:** Configured with $n_{mix} = 5$ mixing updates per step and a prior resampling frequency of 5 steps, strictly mirroring the benchmark settings in [23].
- **ProFlow:** Employs $K = 3$ proximal steps with a time-dependent learning rate schedule starting from $\eta_0 = 0.01$ following [49]. When encountering numerical instabilities in our reimplementation, we escalated execution to `float64` precision and performed a best-effort stabilization search while strictly preserving the original methodology.

H Computational Complexity and Efficiency Analysis

A fundamental advantage of the proposed Energy-Guided Transport is its computational efficiency and structural scalability to high-dimensional domains. In this section, we formally analyze the time and memory complexity introduced by our physics-informed guidance mechanism, contrasting it with state-of-the-art baselines.

Let N denote the total number of spatial degrees of freedom (e.g., $N = H \times W$ for a 2D physical field).

H.1 Theoretical Complexity: VJP vs. Explicit Jacobian

The primary computational bottleneck in physics-guided generation is the interaction with the constraint operator $h(u)$.

Memory Complexity: $\mathcal{O}(N)$ To compute a hard projection, methods like PCFM must explicitly materialize the Jacobian matrix $J = \nabla h(u) \in \mathbb{R}^{M \times N}$. For a full-PDE constraint on a 2D field of size 64×64, $N = 4096$, resulting in a dense Jacobian with over 16 million entries per sample. Storing and inverting this matrix imposes an intractable $\mathcal{O}(N^2)$ memory footprint, severely limiting its application in 3D systems or high-resolution 2D fluids. Conversely, our framework leverages reverse-mode automatic differentiation to compute implicit Vector-Jacobian Products (VJPs). By directly computing the gradient of the scalar energy ($\nabla_u E(u_t)$), the memory requirement is strictly bounded by the storage of the intermediate forward activation tensors. Consequently, the spatial complexity of our physical guidance remains $\mathcal{O}(N)$, matching the theoretical minimum required to represent the state itself.

Time Complexity: $\mathcal{O}(N)$ The computational overhead per ODE integration step in our method consists of evaluating the analytical PDE residual $h(u_t)$ (forward pass) and computing the VJP (backward pass). Assuming the differential operators are discretized via standard local finite differences, the forward evaluation scales linearly, $\mathcal{O}(N)$. According to the Baur-Strassen theorem of algorithmic differentiation, the time complexity of the reverse-mode backward pass is mathematically guaranteed to be bounded by a small constant factor $c \leq 5$ of the forward pass. Therefore, the total time complexity for the physical guidance step is strictly $\mathcal{O}(N)$. This linear scaling ensures that the physical guidance does not become the computational bottleneck: while the base generative backbone (e.g., Fourier Neural Operator) operates at $\mathcal{O}(N \log N)$ due to Fast Fourier Transforms, our dissipative guidance preserves the overall asymptotic complexity of the framework.

H.2 Empirical Computational Budget (NFE and Operator Evaluations)

In flow-based and diffusion-based generative models, the practical computational cost is typically measured by the Number of Function Evaluations (NFE) of the neural vector field v_θ, combined with the number of physical operator evaluations.

Baseline Bottlenecks Optimization and projection-based baselines inherently multiply the computational budget to satisfy constraints. For instance, D-Flow performs global optimization over the initial noise, requiring a forward and an adjoint-backward pass through the entire ODE for each of its 20 L-BFGS iterations. For a standard 100-step solver, this demands up to 4,000 NFE per generated sample. ECI multiplies its integration cost by repeating its mixing step $n_{mix} = 5$ times per ODE step. PCFM and ProFlow compute computationally demanding proximal or projection updates at every intermediate time step (e.g., ProFlow computes $K = 3$ proximal PDE steps per integration step, while PCFM evaluates explicit Jacobians alongside 20 refinement iterations per step).

Energy-Guided Transport Unlike the aforementioned baselines, our standard guided inference requires exactly 1 NFE and 1 VJP evaluation per integration step. The continuous dynamic provided by the transient scheduler smoothly steers the generation without requiring expensive inner-loop optimizations during the flow. For a 100-step Euler integration, this strictly bounds the budget to 100 NFE and 100 PDE gradient evaluations.

Hybrid Energy-Guided Transport When maximum physical adherence is required, our algorithm optionally incorporates a terminal refinement phase at $t = 1$. Importantly, this refinement minimizes the physical energy strictly via gradient descent on the final tensor without querying the neural network. Therefore, performing R refinement steps (e.g., $R = 50$ or $R = 200$) adds 0 NFE to the overall budget. It incurs only R lightweight $\mathcal{O}(N)$ evaluations of the explicit physical operator. This design strategically uncouples the heavy generative modeling from the physical fine-tuning, resulting in rigorous physical consistency, without compromising the learned data manifold, achieved at a fraction of the computational cost of competing methodologies.

I The Ground Truth Floor, Stencil Overfitting, and the Waterbed Effect

To rigorously contextualize the physical residuals achieved by the generative models, we explicitly compute the evaluation metrics on the testing split of the Ground Truth (GT) simulated dataset. As shown in Table 6, the GT errors for Initial Conditions, Boundary Conditions, and Global Mass evaluate to exactly zero (up to machine precision limits). This is expected, as these macroscopic constraints are explicitly hard-coded or globally conserved during the numerical simulation process.

Conversely, the interior PDE residual and, notably, the Global Mass error in non-linear systems (such as Burgers and Reaction-Diffusion) evaluate to strictly positive values (> 0). This occurs because simulating continuous physical laws on a discrete grid inherently introduces numerical truncation errors and temporal drift (e.g., $\mathcal{O}(\Delta x^2)$ for standard spatial finite difference stencils). These GT residuals serve as the theoretical numerical precision limit, or the "floor," for our evaluation framework.

Overfitting the Evaluation Stencil (Discrete Floor Violations) An interesting phenomenon arises when analyzing optimization-heavy baselines: they occasionally report global scaled PDE or mass

residuals that are strictly lower than the GT floor. As observed in the broader literature on Physics-Informed Neural Networks [42, 43], breaching the GT floor does not necessarily indicate the discovery of a structurally superior physical solution. Rather, it is often a manifestation of stencil overfitting.

When a generative baseline heavily relies on a specific discrete evaluation operator to guide its generation (e.g., via hard algebraic projections or continuous gradient penalties), the model treats the evaluation stencil itself as the absolute optimization objective. Most importantly, because the base generative prior is trained exclusively on the GT dataset, any sample whose discrete residual is forced significantly below the GT floor is, by definition, pushed outside the support of the training distribution. The model is essentially generating out-of-distribution samples that it has never seen, finding a local minimum where localized, high-frequency perturbations zero out the discrete arithmetic of the stencil while drifting from the true continuous physics.

The Waterbed Effect We term this topological trade-off the Waterbed Effect: artificially suppressing the scaled PDE error below the theoretical floor inevitably causes a complementary degradation in unconstrained metrics. While the generated samples generally maintain their macroscopic coherence, this subtle overfitting is empirically validated by two corresponding statistical signals in our evaluation:

1. **An increase in Raw PDE Error:** The unscaled spatial derivatives become abnormally large compared to the vanilla unguided model, revealing the presence of subtle high-frequency artifacts injected across the generated field.
2. **A degradation of Sample MSE:** Because the generated sample is pulled artificially off the true underlying data manifold to satisfy the discrete stencil, it results in a measurable worsening of the core distributional metrics.

Our EGT framework mitigates this effect structurally. By guiding the generative process smoothly and deferring explicit physics optimizations, the framework steers the generation toward the GT floor without over-constraining the features to the specific discrete stencil. This approach preserves the continuous physical manifold, successfully minimizing the physical residual without artificially inflating the complementary distributional metrics.

Table 6: Intrinsic physical residuals of the Ground Truth simulated datasets. These values represent the numerical precision limit (lower bound) dictated by grid discretization. To maintain strict consistency with our main evaluations, we report the global mean absolute error (MAE) computed using both the Partial and Full generative configurations. (*Note: The Heat equation dataset is evaluated exclusively on Initial Conditions, hence the absence of BC error*).

Dataset	IC Error	BC Error	Mass Error	PDE Error (Scaled)	PDE Error (Raw)	Partial Phys (MAE)	Full Phys (MAE)
Heat Equation	0.00e0	-	1.32e-7	3.57e-5	9.04e-3	5.44e-8	3.53e-5
Reaction-Diff.	0.00e0	0.00e0	1.90e-3	3.07e-7	5.03e-3	8.30e-4	3.04e-7
Burgers' Eq.	0.00e0	5.99e-9	4.70e-3	6.24e-5	6.24e-1	1.56e-3	6.05e-5
Navier-Stokes	0.00e0	0.00e0	1.43e-7	6.98e-6	2.86e-2	1.66e-9	6.63e-6

J Extended Results and Disaggregated Metrics

In this section, we provide the exhaustive evaluation metrics for all tested generative methods across our suite of PDE benchmarks. To facilitate a clear and comprehensive analysis, we consolidate the results into dedicated subsections for each physical system (Heat, Reaction-Diffusion, Burgers', and Navier-Stokes).

Each physical system is evaluated under two distinct observation regimes:

- **Partial Physics Guidance (Panel A):** The generative models are constrained solely by partial observations, typically Initial Conditions (IC), Boundary Conditions (BC), and/or global integral quantities.
- **Full PDE Guidance (Panel B):** The generative models are subjected to a rigorous "stress test", where they are constrained by the full underlying differential equation across the entire spatiotemporal domain, alongside the boundary constraints.

Furthermore, we disaggregate the results based on the underlying generative prior: the standard Gaussian White Noise prior and the structured Smooth Gaussian Kernel prior.

J.1 Evaluation Metrics Formulation

To rigorously assess the performance of the generative models, we track a diverse set of metrics encompassing computational stability, distributional fidelity, and physical consistency.

Computational Stability and Efficiency:

- **Success Rate (SR, %):** Represents the percentage of samples successfully generated without encountering numerical divergence (e.g., exploding gradients resulting in `NaN` values) or Out-Of-Memory (OOM) errors. Trajectories that encountered numerical divergence or memory exhaustion (NaN/OOM) are denoted as *DNC* (Did Not Complete).
- **Inference Speed (Spd, sec/sample):** The average computational wall-clock time required to generate a single physical sample, amortized over the total batch size. *Hardware Disclaimer:* To ensure a strictly fair and unbiased comparison, all competing generative methods evaluated within a specific PDE benchmark were executed on the identical hardware (e.g., NVIDIA RTX 8000, A6000, L40 GPU). Consequently, while absolute inference times may vary across different physical systems due to hardware assignments and grid dimensionalities, the relative speed comparisons within each individual experiment are strictly standardized.

Distributional and Observational Fidelity: Following established practices in generative modeling for PDEs [23, 49], we evaluate the capability of the models to capture the true underlying data distribution. Let $U_{pred} \in \mathbb{R}^{B\times N_x\times N_t}$ and $U_{true} \in \mathbb{R}^{B\times N_x\times N_t}$ represent the batched generated and ground truth tensors.

- **Mean MSE (MMSE):** Measures the mean squared error between the sample mean of the generated batch and the ground truth batch: $\text{MSE}(\mu_{pred}, \mu_{true})$. It indicates the accuracy of the estimated posterior mean.
- **Standard Deviation MSE (SMSE):** Measures the mean squared error between the sample standard deviations: $\text{MSE}(\sigma_{pred}, \sigma_{true})$. It quantifies how well the method captures the predictive uncertainty.
- **Sample-wise MSE (Sample MSE):** The standard point-wise mean squared error evaluated directly between the predicted and true spatiotemporal fields.

Physical Consistency (Constraint Errors): To evaluate adherence to the governing physical laws, we compute absolute residuals for various constraints. All metrics are reported as Mean Absolute Errors (MAE) averaged over the generated samples.

- **IC Error & BC Error:** Measure the absolute deviation of the generated solution from the prescribed Initial Conditions (at $t = 0$) and boundary values (e.g., Dirichlet or Neumann fluxes).
- **Mass Error & Conservation Laws (CL):** Quantify the violation of global integral invariants. **Mass** explicitly refers to the conservation of mass over time (e.g., balancing total state sum with boundary fluxes and reaction sources), while **CL** refers to additional physical invariants or local multi-step flux balances (such as Godunov flux constraints in the Burgers' equation).
- **Raw PDE Error:** The unscaled absolute error of the interior partial differential equation evaluated via discrete finite differences. As discussed in Appendix I, observing this metric is key to detect high-frequency artifacts and "stencil overfitting".
- **Total Phys (MAE):** This is the global Mean Absolute Error of the specific composite guidance function, $\mathcal{H}(u)$, employed during inference. It is important to notice that the composition of this metric depends on the evaluation regime:
  - In the **Partial Guidance** task, the MAE is computed exclusively over the boundary constraints (IC, BC) and the integral mass/CL residuals.

– In the **Full PDE Guidance** task, the MAE concatenates the boundary constraints with the internal PDE residual. In this full configuration, the PDE residual is strictly scaled by the spatial grid discretization factor (e.g., Δx^2). This matches the exact $\mathcal{H}(u)$ loss formulation structurally required by optimization-based baselines to maintain numerical stability during gradient-based guidance.

J.2 1D Heat Equation Results

Table 7 presents the comprehensive evaluation for the 1D Heat Equation. As the simplest linear PDE in our benchmark, this environment serves as a foundational test to measure the models' capability to enforce basic spatial and temporal constraints before moving to non-linear dynamics.

Table 7: Comprehensive evaluation metrics for the **1D Heat Equation**. The table compares zero-shot guidance performance across Partial (Boundary/IC) and Full-PDE tasks using both White Noise and Smooth Gaussian priors. ↓ indicates lower is better, ↑ indicates higher is better. All physical metrics are averaged over the test set. Instances with a success rate lower than 100% are marked as DNC. Best results are in **bold**, second-best are underlined (excluding Vanilla).

Metric	Vanilla	EGT (Ours)	Hybrid EGT (Ours)	PCFM	ProFlow	DiffPDE	ECI	D-Flow
PART 1: PARTIAL/BOUNDARY GUIDANCE								
Prior: White Noise								
Success Rate (↑)	100%	**100%**	**100%**	0.2%	**100%**	**100%**	**100%**	**100%**
Raw PDE Residual (↓)	0.354	0.399	0.483	DNC	0.402	**0.318**	0.473	5.676
Physical Res. MAE (↓)	0.256	0.051	0.019	DNC	0.189	0.026	**0.013**	0.085
IC Error (↓)	0.433	0.015	3.1e-5	DNC	0.270	4.3e-3	**2.2e-20**	0.011
Mass Error (↓)	1.8e-3	3.8e-3	0.033	DNC	**1.6e-3**	3.3e-3	0.027	0.152
MMSE (↓)	5.2e-3	1.1e-3	**9.3e-4**	DNC	4.9e-3	9.4e-4	**3.3e-4**	8.6e-3
SMSE (↓)	1.5e-2	6.1e-3	6.7e-3	DNC	5.0e-3	6.4e-3	**5.3e-5**	0.177
Sample MSE (↓)	0.083	0.015	0.017	DNC	0.062	**0.013**	0.014	0.348
Speed (s/sample) (↓)	0.19	**0.37**	0.46	DNC	8.12	0.94	8.30	15.58
Prior: Smooth Kernel								
Success Rate (↑)	100%	**100%**	**100%**	1.1%	**100%**	**100%**	0%	**100%**
Raw PDE Residual (↓)	40.83	40.68	40.30	DNC	**3.728**	39.23	DNC	1.3e3
Physical Res. MAE (↓)	0.425	0.047	**0.018**	DNC	0.179	0.062	DNC	0.973
IC Error (↓)	0.984	0.010	**2.7e-5**	DNC	0.226	0.010	DNC	0.092
Mass Error (↓)	0.053	**0.014**	0.032	DNC	0.030	0.046	DNC	1.817
MMSE (↓)	1.3e-2	5.0e-4	3.7e-4	DNC	1.2e-2	**2.6e-4**	DNC	1.361
SMSE (↓)	3.4e-2	1.0e-2	1.0e-2	DNC	7.4e-3	**6.2e-3**	DNC	21.66
Sample MSE (↓)	0.256	0.049	0.049	DNC	0.052	**0.043**	DNC	24.88
Speed (s/sample) (↓)	0.19	**0.37**	0.43	DNC	8.16	0.94	DNC	15.63
PART 2: FULL-PDE GUIDANCE								
Prior: White Noise								
Success Rate (↑)	100%	**100%**	**100%**	0%	0.4%	**100%**	**100%**	**100%**
Raw PDE Residual (↓)	0.345	0.530	0.305	DNC	DNC	**0.025**	0.472	7.875
Physical Res. MAE (↓)	6.2e-3	5.1e-3	3.9e-3	DNC	DNC	0.985	**1.8e-3**	0.032
IC Error (↓)	0.410	0.174	0.156	DNC	DNC	18.05	**2.0e-20**	28.12
Mass Error (↓)	1.8e-3	0.017	**0.012**	DNC	DNC	0.671	0.026	0.571
MMSE (↓)	4.9e-3	4.0e-3	4.0e-3	DNC	DNC	0.137	**1.0e-4**	0.029
SMSE (↓)	1.4e-2	0.015	0.016	DNC	DNC	16.81	**3.8e-5**	22.11
Sample MSE (↓)	0.078	0.046	0.049	DNC	DNC	18.14	**0.014**	23.66
Speed (s/sample) (↓)	0.13	**0.23**	0.27	DNC	DNC	0.68	10.28	18.06
Prior: Smooth Kernel								
Success Rate (↑)	100%	**100%**	**100%**	0%	0%	**100%**	0%	**100%**
Raw PDE Residual (↓)	40.48	1.529	**0.914**	DNC	DNC	20.16	DNC	1.6e2
Physical Res. MAE (↓)	0.165	7.1e-3	**4.5e-3**	DNC	DNC	0.079	DNC	0.657
IC Error (↓)	0.941	0.027	0.018	DNC	DNC	**9.8e-3**	DNC	0.117
Mass Error (↓)	0.052	**0.066**	0.044	DNC	DNC	0.024	DNC	3.298
MMSE (↓)	0.013	1.2e-3	1.2e-3	DNC	DNC	**3.1e-4**	DNC	0.620
SMSE (↓)	0.034	2.3e-3	**2.1e-3**	DNC	DNC	6.6e-3	DNC	15.47
Sample MSE (↓)	0.248	**0.035**	**0.035**	DNC	DNC	0.045	DNC	17.61
Speed (s/sample) (↓)	0.19	**0.51**	0.55	DNC	DNC	1.04	DNC	16.79

J.2.1 Discussion

The empirical results for the 1D Heat Equation highlight the distinct operational characteristics of the evaluated methods, particularly regarding their sensitivity to the underlying generative prior and the complexity of the imposed constraints.

Performance under White Noise Prior Given the linear nature of the Heat Equation, the uncorrelated White Noise regime favors baseline methods equipped with exact algebraic correction mechanisms. Most notably, the ECI framework achieves state-of-the-art performance across almost all metrics in both Partial and Full-PDE tasks. It successfully enforces the Initial Condition to near machine precision (2.2×10^{-20}) while maintaining excellent distributional fidelity (Sample MSE 0.014), albeit at a higher computational cost (~ 10 seconds per sample). DiffusionPDE demonstrates strong performance in the Partial task (Sample MSE 0.013), but its distributional fidelity degrades significantly under Full-PDE guidance, where the Sample MSE rises to 18.14 despite maintaining numerical stability. ProFlow operates reliably under Partial guidance (100% SR) but struggles to complete the Full-PDE stress test (0.4% SR). Finally, while D-Flow maintains a 100% success rate, it incurs significant computational overhead (~ 18s) and struggles to preserve the underlying data distribution, yielding elevated Sample MSE and physical errors. PCFM encounters severe numerical instabilities in this environment, resulting in Did Not Complete (DNC) in all Heat experiments.

Prior Shift to Smooth Gaussian Kernels The operational landscape shifts considerably when transitioning to the spatially correlated Smooth Kernel prior. Methods reliant on hard algebraic projections struggle to adapt to this structured kernel: ECI drops to a 0% success rate, being unable to interpolate smoothly without diverging, and PCFM remains predominantly DNC. ProFlow continues to stably solve the Partial task but again fails to scale to the Full-PDE formulation (0% SR). While DiffusionPDE safely completes all trajectories (100% SR), its physical correction capability is limited in the Partial task (yielding a Raw PDE residual of 39.23, comparable to the unguided Vanilla model) and introduces distributional drift in the Full-PDE task.

Robustness and Efficiency of the Proposed Method The proposed EGT framework and its Hybrid variant demonstrate consistent robustness, maintaining a 100% success rate across all generative priors and constraint complexities. While EGT slightly trails ECI's exact numerical matching in the simpler White Noise Partial task, it exhibits superior adaptability in the structured prior regime. Under the challenging Smooth Kernel Full-PDE task, Hybrid EGT effectively balances physical adherence and manifold preservation: it reduces the Raw PDE residual to 0.914 (down from the Vanilla baseline of 40.48) while maintaining a highly accurate Sample MSE of 0.035. Most importantly, this consistency is achieved with minimal computational overhead. Hybrid EGT requires only 0.55 seconds per sample, making it highly efficient compared to optimization-heavy baselines like D-Flow (~ 16.7s) or ECI's White Noise performance (~ 10.2s).

J.3 1D Burgers' Equation Results

Table 8 presents the comprehensive evaluation for the 1D Burgers' Equation. As a canonical nonlinear PDE that develops steep gradients and shock waves over time, this dataset introduces significant complexity, stress-testing the models' ability to transport the generative flow accurately while avoiding numerical divergence in highly stiff regions.

J.3.1 Discussion

The empirical results for the 1D Burgers' Equation illustrate the varying scalability of guidance methods when subjected to non-linear physical dynamics and the formation of steep spatial gradients (shock waves).

Performance under White Noise Prior In the uncorrelated prior regime, optimization and proximal methods demonstrate distinct trade-offs. ECI achieves the best distributional fidelity (Sample MSE 0.110 in Partial, 0.116 in Full) and enforces Initial Conditions to machine precision, though it requires higher inference times ($\sim 18 - 21$s). DiffusionPDE provides a highly efficient balance, lowering the Sample MSE compared to the Vanilla baseline while running in near real-time (~ 1.1s). ProFlow successfully minimizes the Raw PDE Residual (0.949) more effectively than other methods,

Table 8: Comprehensive evaluation metrics for the **1D Burgers' Equation**. The table compares zero-shot guidance performance across Partial (Boundary/IC) and Full-PDE tasks using both White Noise and Smooth Gaussian priors. ↓ indicates lower is better, ↑ indicates higher is better. All physical metrics are averaged over the test set. Instances with a success rate lower than 100% are marked as DNC. † indicates values that artificially fall below the discrete ground-truth stencil precision limit (stencil overfitting/waterbed effect).

Metric	Vanilla	EGT (Ours)	Hybrid EGT (Ours)	PCFM	ProFlow	DiffPDE	ECI	D-Flow
PART 1: PARTIAL/BOUNDARY GUIDANCE								
Prior: White Noise								
Success Rate (↑)	100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Raw PDE Residual (↓)	0.955	1.247	1.441	1.266	**0.949**	1.128	1.205	4.370
Physical Res. MAE (↓)	0.102	0.051	0.013	**1.2e-8**†	0.134	0.059	0.049	0.083
IC Error (↓)	0.175	0.167	0.166	0.187	0.184	8.1e-3	**6.0e-20**	0.174
BC Error (↓)	0.064	7.0e-3	2.5e-5	**5.5e-16**†	0.098	6.2e-4	1.2e-4	0.160
Mass Error (↓)	4.9e-3	0.030	0.032	**3.6e-8**†	2.2e-3†	0.141	0.138	0.083
CL Error (↓)	0.077	0.107	0.083	**0.066**	0.172	0.132	0.140	0.090
MMSE (↓)	0.021	0.039	0.038	0.028	0.075	0.017	**4.2e-3**	0.023
SMSE (↓)	5.1e-3	5.7e-3	5.8e-3	6.2e-3	8.0e-3	4.9e-3	**1.5e-3**	0.104
Sample MSE (↓)	0.205	0.206	0.205	0.193	0.235	0.155	**0.110**	0.309
Speed (s/sample) (↓)	0.41	**0.48**	0.52	1.29	21.34	1.17	21.87	39.61
Prior: Smooth Kernel								
Success Rate (↑)	100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Raw PDE Residual (↓)	1.543	1.684	1.722	1.856	**1.052**	1.733	1.512	3.6e2
Physical Res. MAE (↓)	0.127	0.021	4.6e-3	**1.2e-8**†	0.084	0.053	0.068	0.372
IC Error (↓)	0.130	0.131	0.131	0.206	0.156	0.019	**2.5e-18**†	14.673
BC Error (↓)	0.080	1.3e-3	4.6e-6	**5.5e-16**†	0.042	1.4e-3	2.9e-4	8.939
Mass Error (↓)	0.021	0.014	0.012	**3.7e-8**†	6.8e-3	0.106	0.188	0.532
CL Error (↓)	0.065	0.077	0.068	0.066	**0.048**	0.111	0.205	0.109
MMSE (↓)	4.4e-3	6.0e-3	**5.9e-3**	0.026	0.074	**3.8e-3**	0.057	0.037
SMSE (↓)	4.7e-3	**4.9e-3**	**4.9e-3**	0.015	0.065	**4.9e-3**	0.060	9.320
Sample MSE (↓)	0.16	0.137	0.137	0.223	0.164	**0.126**	0.142	10.584
Speed (s/sample) (↓)	0.41	0.48	0.52	1.28	21.41	1.16	21.94	39.29
PART 2: FULL-PDE GUIDANCE								
Prior: White Noise								
Success Rate (↑)	100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Raw PDE Residual (↓)	0.955	1.328	1.529	4.9e+18	**0.947**	1.131	1.221	5.116
Physical Res. MAE (↓)	5.1e-3	1.9e-3	**1.5e-4**	4.8e+14	6.3e-3	9.5e-4	2.0e-4	2.9e-3
IC Error (↓)	0.175	0.011	1.5e-9†	1.1e+15	0.184	7.6e-3	**6.0e-20**†	0.170
BC Error (↓)	0.062	6.6e-3	**1.8e-9**†	4.4e+5	0.101	6.2e-4	8.7e-5	0.095
Mass Error (↓)	4.9e-3	0.147	0.196	2.9e+14	**3.2e-3**	0.142	0.143	0.108
CL Error (↓)	0.074	0.156	0.200	3.0e+5	0.174	0.134	0.140	**0.096**
MMSE (↓)	0.021	0.033	0.033	1.3e+14	0.073	0.018	**5.7e-3**	0.022
SMSE (↓)	4.7e-3	4.8e-3	5.2e-3	1.1e+17	6.8e-3	4.5e-3	**1.7e-3**	0.140
Sample MSE (↓)	0.204	0.189	0.189	1.1e+17	0.239	0.153	**0.116**	0.346
Speed (s/sample) (↓)	0.26	0.45	0.54	67.11	18.68	1.14	18.67	32.69
Prior: Smooth Kernel								
Success Rate (↑)	100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Raw PDE Residual (↓)	1.538	1.819	1.900	5.9e+17	**1.053**	1.727	1.513	4.0e2
Physical Res. MAE (↓)	5.8e-3	1.5e-3	**1.8e-4**	5.7e+13	4.5e-3	1.5e-3	3.1e-4	0.048
IC Error (↓)	0.127	0.018	1.6e-9†	3.3e+15	0.153	0.018	**2.5e-18**†	16.369
BC Error (↓)	0.082	1.3e-3	**3.4e-10**†	9.1e+4	0.043	1.4e-3	2.9e-4	16.783
Mass Error (↓)	0.022	0.108	0.174	8.3e+14	**7.2e-3**	0.107	0.188	0.369
CL Error (↓)	0.065	0.125	0.184	2.7e+5	**0.048**	0.112	0.205	0.129
MMSE (↓)	4.0e-3	4.9e-3	4.9e-3	2.8e+13	0.074	**3.5e-3**	0.057	0.052
SMSE (↓)	4.8e-3	5.2e-3	5.1e-3	2.6e+16	0.066	**5.0e-3**	0.060	12.724
Sample MSE (↓)	0.158	0.128	0.128	2.6e+16	0.164	**0.124**	0.142	14.448
Speed (s/sample) (↓)	0.37	**0.95**	1.06	50.87	17.41	1.74	17.08	28.45

but doing so slightly increases the Sample MSE (0.239) and incurs significant computational overhead ($\sim 18 - 21$s). D-Flow avoids numerical divergence but consistently yields elevated physical errors and requires the longest inference times ($\sim 32 - 39$s). Notably, while PCFM minimizes boundary errors in the Partial task, its explicit Jacobian formulation struggles fundamentally with the non-linear shock dynamics under Full-PDE constraints, resulting in severe numerical overflow (Physical MAE $> 10^{14}$).

Prior Shift to Smooth Gaussian Kernels Transitioning to the spatially correlated prior exacerbates the challenges for certain baselines while favoring others. DiffusionPDE and ECI remain highly stable, effectively lowering the Sample MSE (0.124 and 0.142 in Full-PDE, respectively) compared to the Vanilla baseline (0.158). ProFlow maintains the lowest Raw PDE Residual (1.053) but remains computationally demanding. Conversely, D-Flow exhibits high sensitivity to the correlated prior, severely degrading the visual fidelity and causing its Sample MSE to spike above 10. PCFM continues to experience massive numerical overflow in the Full-PDE task. In the Partial task, PCFM artificially drives boundary errors below the Ground Truth numerical limit (stencil overfitting, marked with †), which structurally disconnects the flow from the learned manifold and degrades the Sample MSE (0.223) relative to Vanilla (0.160).

Robustness of the Proposed Method Our continuous frameworks (EGT and Hybrid EGT) robustly handle both the non-linear shocks of the Burgers' equation and the complexities of the prior shift. EGT consistently improves the Sample MSE over the Vanilla baseline while maintaining sub-second inference speeds. Hybrid EGT proves highly effective under Full-PDE constraints, achieving the lowest overall Physical Residual MAE (1.5×10^{-4} for White Noise and 1.8×10^{-4} for Smooth Kernel) without relying on computationally expensive unrolled trajectory optimizations. Objectively, it must be noted that under the intense terminal refinement phase of Hybrid EGT in the Full-PDE task, the local IC and BC errors are driven below the numerical floor (marked with †). However, unlike strict algebraic projections, this localized boundary optimization does not negatively impact the global manifold, as Hybrid EGT still achieves a highly competitive Sample MSE (e.g., 0.128 under Smooth Kernel) at a fraction of the inference time (~ 1.06s) required by ECI or ProFlow.

J.4 1D Reaction-Diffusion Equation Results

Table 9 presents the comprehensive evaluation for the 1D Reaction-Diffusion (RD) Equation. By introducing a non-linear reaction source term that actively competes with spatial diffusion, this PDE dictates complex transient dynamics and localized pattern formations. Consequently, it rigorously tests the solvers' ability to balance competing physical forces across the generative trajectory without falling into numerical instability.

J.4.1 Discussion

The empirical results for the Reaction-Diffusion Equation heavily expose the structural limitations of rigid algebraic projections when faced with coupled non-linear PDE constraints over dense temporal grids.

Performance under White Noise Prior Under standard White Noise and simple Partial (Boundary/IC) constraints, most baseline methods maintain reasonable stability. However, scaling to the Full-PDE task exposes the scalability limits of explicit projection frameworks. Both PCFM and ProFlow encounter severe numerical instabilities along the flow; although they technically complete the generation without throwing NaN exceptions (100% Success Rate), their Raw PDE Residuals experience severe overflow, reaching 2.2×10^9 and 4.8×10^9, respectively. In contrast, DiffPDE and ECI handle the White Noise environment more reliably, maintaining sub-optimal but stable physical residuals without diverging.

Prior Shift and stencil overfitting The introduction of the spatially correlated Smooth Kernel prior further amplifies the divergence between continuous guidance and post-hoc approximations. The Waterbed Effect is starkly visible in PCFM during Partial tasks: it drives the physical and boundary residuals down to precision limits (marked with †), but this aggressive optimization of the discrete stencil severely degrades the underlying manifold, resulting in poor Sample MSE scores. Furthermore, the prior shift heavily impacts optimization-based methods on Full-PDE tasks. For

Table 9: Comprehensive evaluation metrics for the **1D Reaction-Diffusion Equation**. The table compares zero-shot guidance performance across Partial (Boundary/IC) and Full-PDE tasks using both White Noise and Smooth Gaussian priors. ↓ indicates lower is better, ↑ indicates higher is better. All physical metrics are averaged over the test set. Instances with a success rate lower than 100% are marked as DNC.. † indicates values that artificially fall below the discrete ground-truth stencil precision limit (stencil overfitting/waterbed effect).

Metric	Vanilla	EGT (Ours)	Hybrid EGT (Ours)	PCFM	ProFlow	DiffPDE	ECI	D-Flow
PART 1: PARTIAL/BOUNDARY GUIDANCE								
Prior: White Noise								
Success Rate (↑)	100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Raw PDE Residual (↓)	0.086	0.222	0.423	0.716	7.3e+3	**0.184**	1.185	4.258
Physical Res. MAE (↓)	0.200	0.075	8.5e-3	**1.1e-8**†	1.2e+3	0.043	0.054	0.739
IC Error (↓)	0.181	0.024	5.5e-6	**1.4e-44**	4.5e+5	7.8e-3	1.5e-20	2.0e+2
BC Error (↓)	6.9e-4	1.7e-3	0.032	0.154	7.8e+6	**7.8e-4**	0.224	3.040
Mass Error (↓)	4.5e-3	0.011	0.017	**2.5e-8**†	2.6e+3	0.016	0.124	1.077
CL Error (↓)	0.025	0.052	0.109	0.047	17.208	**0.026**	0.175	0.054
MMSE (↓)	2.0e-3	9.1e-3	8.9e-3	3.1e-3	3.3e+3	**1.2e-3**	0.041	0.245
SMSE (↓)	4.6e-3	**5.0e-3**	5.2e-3	6.7e-3	7.4e+5	5.1e-3	0.093	1.9e+2
Sample MSE (↓)	0.083	0.051	0.050	0.041	7.4e+5	**0.023**	0.246	2.0e+2
Speed (s/sample) (↓)	0.38	**1.13**	1.24	2.16	25.65	2.00	24.47	32.07
Prior: Smooth Kernel								
Success Rate (↑)	100%	**100%**	100%	100%	100%	100%	100%	100%
Raw PDE Residual (↓)	1.274	1.697	1.460	3.636	2.604	**1.309**	8.682	10.993
Physical Res. MAE (↓)	0.220	0.028	8.5e-3	**1.1e-8**†	0.239	0.027	0.086	0.026
IC Error (↓)	0.182	1.1e-3	4.6e-6	**5.7e-44**	0.219	4.6e-4	1.4e-18	5.2e-3
BC Error (↓)	0.018	0.117	0.039	0.164	0.084	**0.017**	1.118	0.088
Mass Error (↓)	0.053	0.027	0.017	**2.5e-8**†	0.056	0.043	0.198	0.038
CL Error (↓)	0.031	**0.036**	0.053	0.100	0.041	0.041	0.091	0.071
MMSE (↓)	1.4e-3	1.1e-3	1.0e-3	**3.5e-4**	0.058	4.3e-4	0.044	1.7e-3
SMSE (↓)	3.8e-3	7.0e-4	6.8e-4	0.036	0.043	**6.3e-4**	0.179	0.015
Sample MSE (↓)	0.088	0.010	9.9e-3	0.087	0.191	**9.3e-3**	0.365	0.094
Speed (s/sample) (↓)	0.38	1.09	1.25	2.11	25.60	1.98	24.43	31.43
PART 2: FULL-PDE GUIDANCE								
Prior: White Noise								
Success Rate (↑)	100%	**100%**	100%	100%	100%	100%	100%	100%
Raw PDE Residual (↓)	0.086	0.209	0.322	2.2e+9	4.8e+9	**0.183**	1.185	0.703
Physical Res. MAE (↓)	3.5e-3	1.2e-3	**2.0e-5**	1.3e+5	2.9e+5	6.5e-3	7.2e-5	6.6e-3
IC Error (↓)	0.181	0.021	1.1e-10	1.5e+5	5.0e+11	7.7e-3	**1.5e-20**	0.168
BC Error (↓)	6.9e-4	**5.6e-4**	5.6e-3	7.6e+18	3.0e+11	7.8e-4	0.224	0.085
Mass Error (↓)	4.5e-3	0.034	0.085	6.5e+14	2.5e+9	**0.015**	0.124	0.104
CL Error (↓)	0.025	0.056	0.106	1.0e+5	8.8e+2	**0.026**	0.175	0.045
MMSE (↓)	1.9e-3	8.0e-3	8.0e-3	8.6e+13	3.3e+9	**1.2e-3**	0.041	2.2e-3
SMSE (↓)	4.6e-3	5.2e-3	5.2e-3	6.7e+16	5.0e+11	**5.0e-3**	0.093	0.065
Sample MSE (↓)	0.083	0.049	0.049	6.7e+16	5.0e+11	**0.023**	0.246	0.169
Speed (s/sample) (↓)	0.376	**0.891**	0.928	76.720	25.055	1.860	24.343	31.443
Prior: Smooth Kernel								
Success Rate (↑)	100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Raw PDE Residual (↓)	1.274	1.335	1.324	9.8e+8	2.609	**1.310**	8.682	12.475
Physical Res. MAE (↓)	3.6e-3	3.6e-4	**8.0e-5**	5.9e+4	4.0e-3	2.3e-4	5.2e-4	1.5e-3
IC Error (↓)	0.182	1.1e-3	9.9e-12	1.7e+6	0.219	4.6e-4	**1.4e-18**	6.214
BC Error (↓)	0.018	0.017	**0.017**	2.4e+19	0.088	0.018	1.118	3.369
Mass Error (↓)	0.053	0.052	**0.047**	2.1e+15	0.082	0.051	0.198	0.153
CL Error (↓)	0.031	**0.036**	0.052	2.0e+5	0.041	0.041	0.091	0.107
MMSE (↓)	1.3e-3	1.0e-3	1.0e-3	2.1e+14	0.058	**4.3e-4**	0.044	3.7e-3
SMSE (↓)	3.8e-3	6.6e-4	6.6e-4	2.1e+17	0.044	**6.3e-4**	0.179	1.138
Sample MSE (↓)	0.088	0.010	9.8e-3	2.1e+17	0.191	**9.4e-3**	0.365	1.530
Speed (s/sample) (↓)	0.375	**0.889**	0.951	77.715	25.099	1.857	24.395	31.313

example, ECI, which performed admirably under White Noise, struggles to interpolate the correlated noise, causing its Sample MSE to degrade to 0.365 under the Smooth Kernel Full-PDE task.

Robustness of the Proposed Method Our continuous guidance architectures (EGT and Hybrid EGT) stand out as the most robust and computationally efficient frameworks across all configurations. Hybrid EGT consistently excels in the challenging Full-PDE tasks, achieving the best Physical Residual MAE under the White Noise prior (2.0×10^{-5}) and maintaining exceptional fidelity under the Smooth Kernel (8.0×10^{-5}). Most importantly, our framework enforces these complex non-linear dynamics without resorting to stencil overfitting and without compromising the macroscopic visual coherence of the fields (maintaining top-tier Sample MSE). Beyond raw accuracy, EGT significantly outperforms optimization-based baselines in computational efficiency: it resolves the Full-PDE Smooth Kernel task in just ~ 0.95 seconds per sample, bypassing the severe bottlenecks observed in ECI (~ 24 seconds) and PCFM (~ 77 seconds).

J.5 2D Navier-Stokes Equations Results

Table 10 presents the comprehensive evaluation for the 2D Navier-Stokes (NS) Equations. As the only two-dimensional macroscopic fluid dynamics environment in our benchmark, this task requires enforcing complex vorticity structures, non-linear advection, and strict incompressibility over a dense spatial grid. This environment pushes the structural and computational limits of guided generative models.

Table 10: Comprehensive evaluation metrics for the **2D Navier-Stokes Equations**. The table evaluates zero-shot guidance performance across Partial (Boundary/IC) and Full-PDE tasks using a White Noise prior. ↓ indicates lower is better, ↑ indicates higher is better. All physical metrics are averaged over the test set. Instances with a success rate lower than 100% are marked as DNC. † indicates values that artificially fall below the discrete ground-truth stencil precision limit (stencil overfitting/waterbed effect).

Metric	Vanilla	EGT (Ours)	Hybrid EGT (Ours)	PCFM	ProFlow	DiffPDE	ECI	D-Flow
PART 1: PARTIAL/BOUNDARY GUIDANCE								
Prior: White Noise								
Success Rate (↑)	100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**	14.2%
Raw PDE Residual (↓)	0.063	0.060	0.060	0.075	**0.057**	0.070	0.071	DNC
Physical Res. MAE (↓)	0.291	0.010	1.4e-4	**1.2e-10**†	0.197	0.124	3.8e-5	DNC
IC Error (↓)	0.136	1.7e-4	2.4e-8	**0.0**	0.063	0.025	8.9e-17	DNC
BC Error (↓)	0.058	**0.056**	0.057	0.066	0.058	0.057	0.060	DNC
Mass Error (↓)	0.011	7.6e-3	6.2e-3	**1.0e-8**†	3.1e-3	0.039	3.2e-3	DNC
CL Error (↓)	0.012	7.6e-3	6.1e-3	**1.0e-8**	3.3e-3	0.038	3.1e-3	DNC
MMSE (↓)	0.018	**0.018**	0.018	0.037	0.020	0.036	0.116	DNC
SMSE (↓)	0.011	0.011	0.011	0.023	**9.9e-3**	0.011	0.011	DNC
Sample MSE (↓)	0.542	**0.529**	0.548	0.756	0.550	0.541	0.630	DNC
Speed (s/sample) (↓)	0.690	**0.714**	0.898	93.972	0.847	1.402	3.559	DNC
PART 2: FULL-PDE GUIDANCE								
Prior: White Noise								
Success Rate (↑)	100%	**100%**	**100%**	0%	**100%**	**100%**	**100%**	1.6%
Raw PDE Residual (↓)	0.062	0.080	0.124	DNC	**0.057**	0.069	0.070	DNC
Physical Res. MAE (↓)	7.5e-3	1.6e-3	**2.9e-5**	DNC	5.6e-3	4.2e-3	1.8e-3	DNC
IC Error (↓)	0.136	1.3e-3	4.9e-14	DNC	0.064	0.025	**8.9e-17**	DNC
BC Error (↓)	0.059	0.034	**1.0e-6**	DNC	0.057	0.057	0.059	DNC
Mass Error (↓)	9.4e-3	7.5e-3	6.4e-3	DNC	**3.1e-3**	0.035	3.2e-3	DNC
CL Error (↓)	9.9e-3	7.6e-3	6.2e-3	DNC	3.3e-3	0.035	**3.1e-3**	DNC
MMSE (↓)	0.019	**0.018**	**0.018**	DNC	0.022	0.032	0.098	DNC
SMSE (↓)	0.011	0.011	0.011	DNC	**9.8e-3**	0.010	0.010	DNC
Sample MSE (↓)	0.559	**0.547**	**0.547**	DNC	0.549	0.553	0.611	DNC
Speed (s/sample) (↓)	0.858	**0.937**	1.021	DNC	1.209	1.789	4.400	DNC

J.5.1 Discussion

The empirical results for the Navier-Stokes equations decisively illustrate the scalability limits and structural fragilities of explicit projection methods when applied to high-dimensional physical systems.

Baseline Performance and Guidance Collapse Scaling the constraint complexity to a 2D grid exposes severe instabilities in baseline frameworks. D-Flow experiences significant guidance degradation, achieving a mere 14.2% Success Rate on Partial constraints and dropping to 1.6% on the Full-PDE task. More critically, PCFM experiences total numerical divergence under Full-PDE constraints (0% SR, yielding NaN metrics), indicating that its reliance on explicit Jacobian approximations is computationally and mathematically restrictive for complex macroscopic fluid flows. While methods like ProFlow, DiffPDE, and ECI technically complete the Full-PDE generation, they struggle to simultaneously satisfy the physics and preserve the underlying data distribution.

High-Dimensional stencil overfitting The Waterbed Effect is exceptionally prominent in this 2D setting. Under Partial constraints, PCFM algebraically forces the Initial Condition error to absolute zero (0.0) and the Physical Residual to precision limits (1.2×10^{-10}, marked with †). However, this aggressive structural constraint severely distorts the generative trajectory, heavily degrading the Sample MSE to 0.756 (markedly worse than the unguided Vanilla model at 0.542). ECI similarly exhibits stencil overfitting on the IC Error (8.9×10^{-17}) while suffering from visual and structural degradation (Sample MSE of 0.611 on the Full-PDE task).

Robustness of the Proposed Method Our Continuous Dissipative Guidance architectures (EGT and Hybrid EGT) successfully navigate the high-dimensional fluid manifold without succumbing to numerical divergence or stencil overfitting. On the highly constrained Full-PDE task, Hybrid EGT decisively outperforms all baselines, achieving a state-of-the-art Physical Residual MAE of 2.9×10^{-5} and driving boundary errors down to 1.0×10^{-6}. Crucially, our framework accomplishes this while perfectly preserving the structural integrity of the fluid (Sample MSE of 0.547). Beyond accuracy, EGT demonstrates unmatched scalability: it strictly satisfies the complex 2D Navier-Stokes physics in just ~ 1.02 seconds per sample, completely bypassing the extreme computational bottlenecks of algebraic methods (e.g., PCFM requiring nearly 94 seconds per sample on Partial tasks).

J.6 Macroscopic Physical Diagnostics

To evaluate how pointwise PDE residual minimization affects broader macroscopic statistics, we compute two additional holistic diagnostics on the generated samples from the main evaluation (White Noise, Full-PDE configuration): the 1D Wasserstein Distance between spatial gradient distributions (capturing shock statistics on 1D Burgers) and the Energy Spectrum L^1 Error (assessing spectral consistency on 2D Navier–Stokes).

Table 11: Macroscopic physical diagnostics on 1D Burgers and 2D Navier–Stokes (White Noise, Full-PDE).

Benchmark	Model	Diagnostic Metric	Mean $\pm$ Std $\downarrow$
Burgers 1D	Vanilla	Gradient Wasserstein Dist.	5.855e-1 $\pm$ 2.577e-1
	DiffusionPDE	Gradient Wasserstein Dist.	4.441e-1 $\pm$ 2.239e-1
	ECI	Gradient Wasserstein Dist.	**3.819e-1 $\pm$ 2.327e-1**
	Ours (Pure EGT)	Gradient Wasserstein Dist.	6.870e-1 $\pm$ 3.010e-1
	Ours (Hybrid EGT)	Gradient Wasserstein Dist.	8.399e-1 $\pm$ 3.687e-1
Navier–Stokes 2D	Vanilla	Energy Spectrum L^1 Error	9.630e-2 $\pm$ 13.06e-2
	DiffusionPDE	Energy Spectrum L^1 Error	3.299e-1 $\pm$ 2.722e-1
	ECI	Energy Spectrum L^1 Error	**2.740e-2 $\pm$ 1.870e-2**
	Ours (Pure EGT)	Energy Spectrum L^1 Error	9.840e-2 $\pm$ 13.09e-2
	Ours (Hybrid EGT)	Energy Spectrum L^1 Error	8.570e-2 $\pm$ 11.43e-2

Discussion. Table 11 highlights an important distinction between local pointwise consistency and global statistical realism. Because the current EGT formulation explicitly minimizes the pointwise L^2 norm of the PDE, IC, and BC residuals, it substantially reduces local physical errors (and improves spectral consistency over DiffusionPDE on 2D Navier–Stokes) but does not directly optimize gradient distributions on 1D Burgers. Because EGT can accommodate any differentiable energy functional, augmenting $E(u)$ with differentiable spectral or gradient-matching penalties represents a natural extension for jointly optimizing local and macroscopic realism.

J.7 Statistical Significance and Seed-to-Seed Variability

To verify the statistical robustness of our evaluation, we first compare the empirical standard deviations computed across 100 and $1,000$ generated samples on the 1D Burgers and 2D Navier–Stokes benchmarks (White Noise, Full-PDE). As shown in Table 12, the standard deviations are virtually identical, confirming that 100 samples provide reliable variance estimates. We therefore evaluate seed-to-seed variability by generating 100 samples across 5 independent random seeds (Table 13, using the linear scheduler for EGT based on our scheduler ablation). EGT exhibits consistently low seed-to-seed variance across both systems, achieving a lower physical MAE (1.49e-3 vs. 1.81e-3) and more than $2\times$ lower Sample MSE variance (1.42e-2 vs. 3.06e-2) than ECI on 2D Navier–Stokes.

Table 12: Variance convergence analysis comparing standard deviations across 100 and $1,000$ generated samples (White Noise, Full-PDE).

Model	Metric	Std NS (100)	Std NS (1000)	Std Burgers (100)	Std Burgers (1000)
Ours (EGT)	Sample MSE	2.292e-1	2.321e-1	9.750e-2	1.208e-1
Ours (EGT)	Physical MAE	3.000e-4	3.000e-4	5.000e-4	6.000e-4
ECI	Sample MSE	3.746e-1	4.020e-1	7.380e-2	9.500e-2
ECI	Physical MAE	6.000e-4	6.000e-4	1.000e-4	1.000e-4
DiffusionPDE	Sample MSE	2.448e-1	2.297e-1	9.700e-2	1.147e-1
DiffusionPDE	Physical MAE	6.000e-4	8.000e-4	4.000e-4	4.000e-4
Vanilla	Sample MSE	2.302e-1	2.339e-1	1.059e-1	1.329e-1
Vanilla	Physical MAE	1.000e-3	1.400e-3	2.100e-3	2.500e-3

Table 13: Seed-to-seed variability across 5 independent random seeds (100 samples per seed, White Noise, Full-PDE).

Model	Burgers (Sample MSE) ↓	Burgers (Phys MAE) ↓	NS (Sample MSE) ↓	NS (Phys MAE) ↓
Vanilla	0.163 ± 9.82e-3	4.54e-3 ± 2.18e-4	0.541 ± 1.43e-2	6.78e-3 ± 9.76e-5
DiffusionPDE	0.139 ± 1.04e-2	1.71e-3 ± 5.89e-5	0.547 ± 1.70e-2	3.72e-3 ± 7.40e-5
ECI	**0.127 ± 6.12e-3**	**2.05e-4 ± 4.98e-6**	0.620 ± 3.06e-2	1.81e-3 ± 4.52e-5
Ours (EGT)	0.146 ± 9.13e-3	$\underline{6.04\text{e-}4 \pm 1.20\text{e-}5}$	**0.528 ± 1.42e-2**	**1.49e-3 ± 2.50e-5**

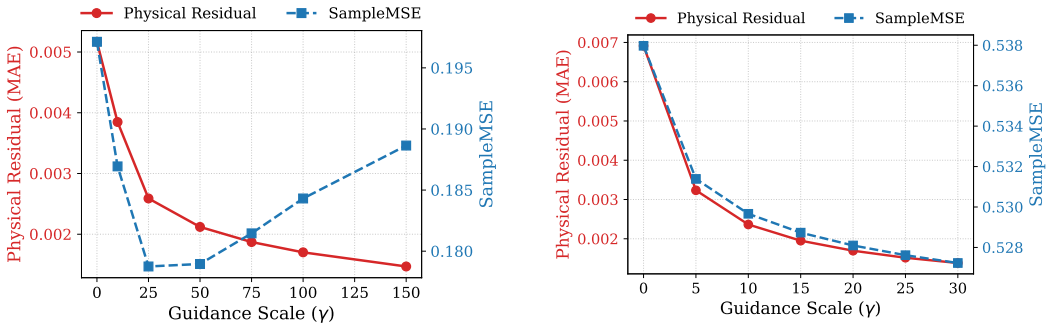
K Ablation Studies

In this appendix, we provide an extensive sensitivity analysis of the core hyperparameters governing Continuous Dissipative Guidance. To evaluate the robustness of our framework across different dimensionalities, we conduct all ablation studies on the 1D Burgers' (Full-PDE) and 2D Navier-Stokes (Full-PDE) datasets. These cases represent the challenging scenarios where the unguided model's physics violation is most prominent. For consistency across all ablation trials, we utilize the standard White Noise prior p_0. All tests are evaluated on 256 samples.

K.1 Sensitivity to Guidance Scale γ

The guidance scale γ controls the magnitude of the dissipative force injected into the generative vector field. Figure 7 illustrates the trade-off between physical adherence and distributional fidelity as a function of γ.

Discussion. As shown in the dual-axis plots, increasing γ triggers a rapid dissipative behaviour of the physical residual. In 1D Burgers', we observe that the physical MAE reaches a stable floor already at $\gamma \approx 25$, with negligible impact on the Sample MSE. This confirms that our continuous guidance effectively steers trajectories toward the physical manifold without causing the "manifold collapse" commonly observed in rigid algebraic projection methods. For 2D Navier-Stokes, the energy landscape is significantly more complex; while EGT successfully navigates this space and improves physical fidelity up to $\gamma = 30$, pushing the guidance scale beyond this threshold disrupts the generative trajectory and results in numerical instability. Importantly, selecting γ in a genuine zero-shot regime does not require access to ground-truth test solutions: since the physical energy $E(u)$ depends exclusively on spatial differential operators evaluated on the generated samples, practitioners



(a) Burgers' Dual-Axis

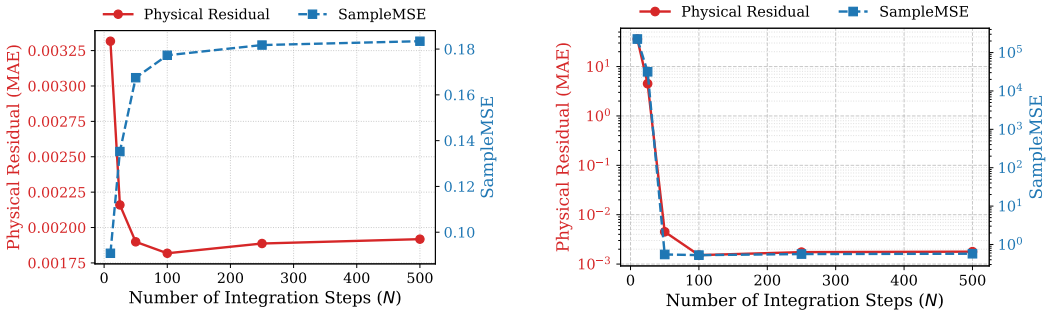
(b) NS' Dual-Axis

Figure 7: Impact of the guidance scale γ. Both dual-axis plots simultaneously track the Physical Residual MAE (physics) and Sample MSE (fidelity) on 1D Burgers' (a) and 2D Navier-Stokes (b).

can simply perform a validation sweep on a small batch of initial noise, progressively increasing γ and selecting the largest value before numerical stiffness or gradient divergence (NaNs) occurs.

K.2 Number of Flow Matching Steps N

We investigate the performance and computational efficiency of our framework under different discretization budgets for the ODE solver, ranging from $N = 10$ to $N = 250$ generative steps (Figure 8).



(a) Burgers' Budget Ablation

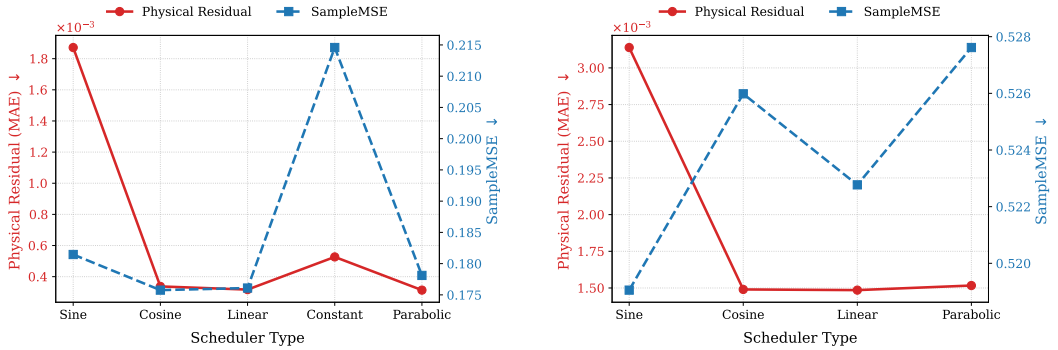
(b) NS' Budget Ablation

Figure 8: Evaluation of ODE integration steps N. Both environments demonstrate that EGT stabilizes quickly. Note that for 2D Navier-Stokes (b), the y-axis for the Sample MSE is strictly logarithmic to accommodate the severe error variance at highly constrained integration budgets.

Discussion. Unlike standard diffusion or heavy explicit projection models that often require hundreds of steps for stability, our energy-guided FFM maintains rigorous physical consistency even at highly accelerated integration budgets ($N \approx 50$). For instance, in the complex 2D Navier-Stokes setting (where the Sample MSE is plotted on a logarithmic y-axis to capture the extreme variance), lowering N below 25 leads to severe numerical divergence. However, at just $N = 50$, the framework instantly stabilizes, achieving a Sample MSE perfectly aligned with the unguided baseline. While the physical residual further decreases with finer discretization, the relative gain plateaus around $N = 100$. This scalability is a significant practical advantage for real-time physics simulators where sub-second inference is required.

K.3 Effect of Guidance Scheduling

Finally, we compare different scheduling functions $\gamma(t)$ for the temporal application of the dissipative force. We categorize these into a Constant baseline, "Late Guidance" schedules (Linear, Parabolic, Cosine), and a "Transient" schedule (Sine).



(a) Burgers' Scheduler Ablation

(b) NS' Scheduler Ablation

Figure 9: Impact of the temporal guidance scheduler $\gamma(t)$. Both "Late Guidance" and "Transient" schedules systematically prevent trajectory collapse by allowing the base neural prior to establish macroscopic coherent structures before rigid physical constraints are fully applied.

Discussion. Our empirical analysis strongly validates the necessity of "Late Guidance" schedules (Linear, Parabolic, Cosine) and the "Transient" schedule (Sine). By allowing the unguided neural prior to freely establish the macroscopic structure of the field near pure noise ($t \approx 0$) before smoothly ramping up the physical constraints, these dynamic schedules avoid interfering with the inherently high variance of the initial noise. Most notably, applying a Constant guidance schedule throughout the flow results in a complete divergence (0% Success Rate) on the 2D Navier-Stokes environment, and a noticeable degradation of Sample MSE in the Burgers' equation. Consequently, smoothly scheduling the dissipative force acts as a necessary annealing process, leading to both optimal Sample MSE and lower final physical residuals.

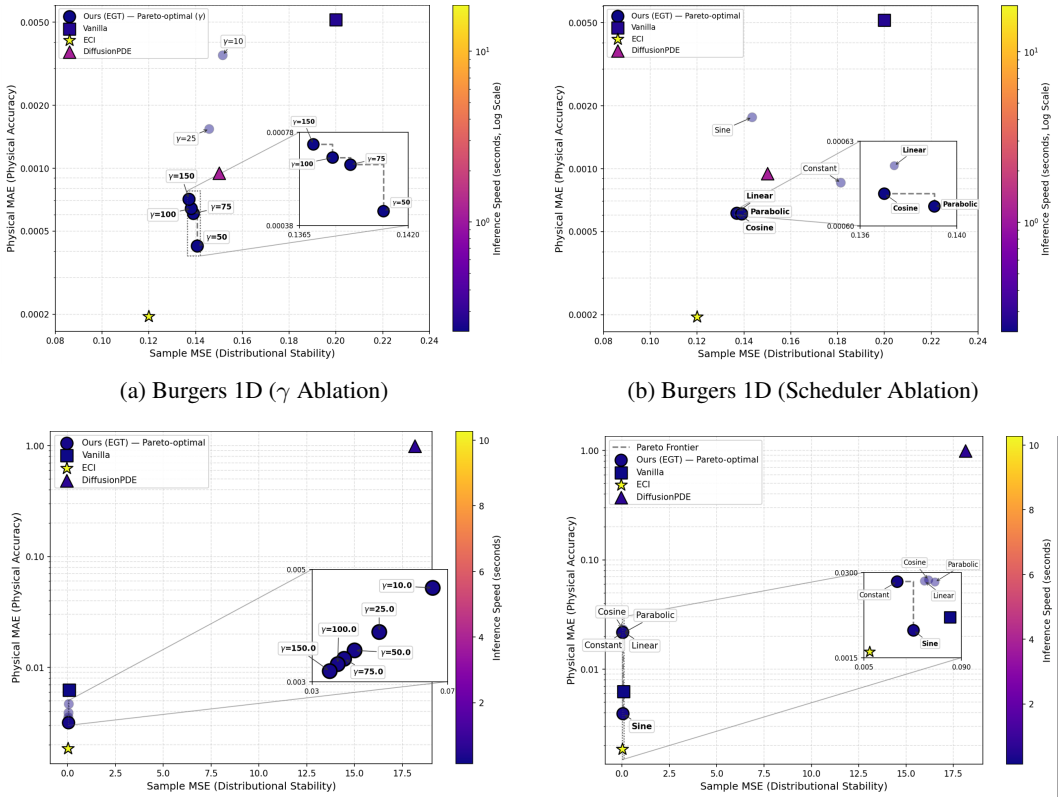
K.4 Pareto Frontier Analysis: Accuracy vs. Stability vs. Speed

To evaluate how EGT navigates the accuracy–stability–speed trade-off compared to projection- and gradient-based baselines, we compute the Pareto frontiers on the 1D Burgers and 1D Heat benchmarks (Full-PDE, $N = 100$) across guidance scales (γ) and schedulers. Figure 10 visualizes the four resulting trade-off curves, while Tables 14 and 15 report the exact numerical operating points.

Table 14: Guidance strength (γ) and scheduler Pareto ablation on the **1D Burgers** system. EGT provides a controllable trade-off between physical accuracy and distributional fidelity while maintaining a constant sub-second runtime (0.23 s/sample).

Method (γ Ablation)	Speed (s) ↓	Phys MAE ↓	Sample MSE ↓	Method (Scheduler Ablation)	Speed (s) ↓	Phys MAE ↓	Sample MSE ↓
Vanilla	0.13	5.53e-3	2.17e-1	Vanilla	0.13	5.53e-3	2.17e-1
ECI	9.30	1.99e-4	1.66e-1	ECI	9.30	1.99e-4	1.66e-1
DiffusionPDE	0.63	1.27e-3	1.88e-1	DiffusionPDE	0.63	1.27e-3	1.88e-1
EGT ($\gamma = 10$)	0.23	3.47e-3	1.52e-1	EGT (Sine)	0.23	1.76e-3	1.44e-1
EGT ($\gamma = 25$)	0.23	1.54e-3	1.46e-1	EGT (Cosine)	0.23	6.11e-4	1.37e-1
EGT ($\gamma = 50$)	0.23	4.24e-4	1.41e-1	EGT (Linear)	0.23	6.21e-4	1.37e-1
EGT ($\gamma = 75$)	0.23	6.07e-4	1.39e-1	EGT (Constant)	0.23	8.56e-4	1.81e-1
EGT ($\gamma = 100$)	0.23	6.41e-4	1.38e-1	EGT (Parabolic)	0.23	6.07e-4	1.39e-1
EGT ($\gamma = 150$)	0.23	7.10e-4	1.37e-1				

Discussion. The Pareto analysis highlights two key operational insights. First, the guidance scale γ smoothly controls the operating point of EGT in a problem-dependent manner: on Burgers, increasing γ continuously improves Sample MSE while the physical residual reaches its minimum around $\gamma = 50$; on Heat, both metrics monotonically improve across the evaluated range. Second, the



(a) Burgers 1D (γ Ablation)

(b) Burgers 1D (Scheduler Ablation)

(c) Heat 1D (γ Ablation)

(d) Heat 1D (Scheduler Ablation)

Figure 10: Pareto frontiers of distributional stability (Sample MSE) versus physical accuracy (Physical MAE), with inference speed (sec/sample) encoded via the colormap. EGT consistently dominates Vanilla and DiffusionPDE while operating 40–80$\times$ faster than ECI.

Table 15: Guidance strength (γ) and scheduler Pareto ablation on the **1D Heat** system.

Method (γ Ablation)	**Speed (s)** $\downarrow$	**Phys MAE** $\downarrow$	**Sample MSE** $\downarrow$	**Method (Scheduler Ablation)**	**Speed (s)** $\downarrow$	**Phys MAE** $\downarrow$	**Sample MSE** $\downarrow$
Vanilla	0.13	7.38e-3	1.12e-1	Vanilla	0.13	7.38e-3	1.12e-1
ECI	8.80	1.95e-3	2.80e-2	ECI	8.80	1.95e-3	2.80e-2
DiffusionPDE	0.63	6.90e-1	9.40e+1	DiffusionPDE	0.63	6.90e-1	9.40e+1
EGT ($\gamma = 10$)	0.26	4.67e-3	6.56e-2	EGT (Sine)	0.27	3.91e-3	4.82e-2
EGT ($\gamma = 25$)	0.26	3.88e-3	4.98e-2	EGT (Cosine)	0.27	2.30e-2	6.14e-2
EGT ($\gamma = 50$)	0.26	3.56e-3	4.25e-2	EGT (Linear)	0.27	2.21e-2	5.78e-2
EGT ($\gamma = 75$)	0.26	3.41e-3	3.95e-2	EGT (Constant)	0.27	2.17e-2	3.40e-2
EGT ($\gamma = 100$)	0.26	3.31e-3	3.76e-2	EGT (Parabolic)	0.27	2.14e-2	6.69e-2
EGT ($\gamma = 150$)	0.26	3.19e-3	3.52e-2				

optimal temporal schedule is problem-dependent: cosine, linear, and parabolic schedules yield the strongest overall balance on Burgers, whereas the transient sine schedule achieves by far the lowest physical residual on Heat. Across all configurations, EGT reduces inference cost by one to two orders of magnitude compared to explicit projections (ECI) while avoiding any computational penalty when tuning hyperparameters.

K.5 Contribution of Continuous Guidance vs. Post-Hoc Refinement

To isolate the distinct contribution of our continuous transport mechanism (Pure EGT) from the optional terminal energy descent (Hybrid EGT), we apply the exact same post-hoc optimization procedure (identical learning rate and number of refinement steps) to the terminal states generated by Vanilla FFM, DiffusionPDE, and ECI under the White Noise Full-PDE setting ($\gamma = 75$ for Burgers, $\gamma = 25$ for Navier–Stokes).

Table 16: Ablation study applying identical post-hoc terminal refinement across baselines on 1D Burgers and 2D Navier–Stokes (White Noise, Full-PDE).

Model	**Burgers (Sample MSE)** $\downarrow$	**Burgers (Phys MAE)** $\downarrow$	**NS (Sample MSE)** $\downarrow$	**NS (Phys MAE)** $\downarrow$
Vanilla (Pure FFM)	2.04e-1 $\pm$ 1.33e-1	5.10e-3 $\pm$ 2.50e-3	5.60e-1 $\pm$ 2.34e-1	7.50e-3 $\pm$ 1.40e-3
Vanilla (Refined)	2.01e-1 $\pm$ 1.31e-1	1.54e-4 $\pm$ 1.00e-4	5.40e-1 $\pm$ 2.34e-1	3.96e-3 $\pm$ 1.42e-3
DiffusionPDE (Refined)	1.53e-1 $\pm$ 1.15e-1	1.16e-4 $\pm$ 2.22e-5	5.41e-1 $\pm$ 2.30e-1	4.42e-4 $\pm$ 5.49e-4
ECI (Refined)	1.16e-1 $\pm$ 9.50e-2	1.18e-4 $\pm$ 1.80e-5	6.20e-1 $\pm$ 4.02e-1	3.16e-5 $\pm$ 1.17e-5
Ours (Pure EGT)	1.89e-1 $\pm$ 1.21e-1	1.90e-3 $\pm$ 6.00e-4	5.47e-1 $\pm$ 2.32e-1	1.60e-3 $\pm$ 3.00e-4
Ours (Hybrid EGT)	1.89e-1 $\pm$ 1.18e-1	1.50e-4 $\pm$ 2.90e-5	5.47e-1 $\pm$ 2.32e-1	2.90e-5 $\pm$ 7.00e-6

Discussion. Table 16 demonstrates that post-hoc refinement alone cannot replace continuous guidance during the flow. While terminal optimization reduces residuals across all methods on 1D Burgers, the non-linear 2D Navier–Stokes system exposes a clear gap: refining unconditional samples (Vanilla Refined) plateaus at a physical MAE of 3.96e-3, whereas Pure EGT already achieves 1.60e-3 before any refinement. Applying the identical terminal optimization to EGT (Hybrid EGT) drives the error down to 2.90e-5. Although ECI + Refinement reaches a comparable residual (3.16e-5), it relies on repeated projections that are nearly $5\times$ slower and degrade Sample MSE (0.620 vs. 0.547).

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately summarize the paper's primary contributions: the Energy-Guided Transport framework, its theoretical foundations, and its empirical advantages over projection-based baselines.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations regarding the sensitivity to the guidance scale γ in high-dimensional settings and the reliance on discrete stencils for energy evaluation are discussed in the Conclusion and Appendix J.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All mathematical assumptions required for the derivation of the dissipative force and the stability analysis are explicitly stated in Appendix A.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We fully disclose all necessary information to reproduce our experiments. The model architectures, continuous guidance hyperparameters, ODE integration budgets, and evaluation metrics are meticulously detailed in Appendices E, G, and J. Additionally, we provide a GitHub repository containing the complete source code to seamlessly reproduce all baseline and proposed framework results.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We provide a link to an GitHub repository containing the implementation, dataset generation scripts, and instructions for reproduction in the introduction.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Detailed training regimes, FNO architectures, and optimization hyperparameters are documented in Appendices E and G.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We report Sample MSE, MMSE, and SMSE across all benchmarks, and provide a dedicated statistical significance and multi-seed variability analysis (across 5 independent random seeds) in Appendix J.7.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Hardware specifications (NVIDIA RTX 8000, A6000, L40) and average inference times per sample are provided in Appendices E and J.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This research focuses on fundamental computational physics and does not have foreseeable direct negative societal impacts.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: Our study uses synthetic physical datasets (Heat, Burgers, Navier-Stokes) and does not involve human subjects or data derived from them.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: This research focuses on solving standard continuous physical systems (e.g., Navier-Stokes, Burgers' equations) using synthetic data, which poses no high risks for societal misuse or dual-use.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All existing assets (e.g., baseline model architectures, PDE simulation environments) used or adapted in our experiments are properly cited in the main text and Appendices, respecting their original open-source licenses.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We introduce a new codebase for the Energy-Guided Transport framework. It is thoroughly documented and provided via an GitHub repository linked in the submission, complete with instructions for data generation and reproducibility.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This research does not involve crowdsourcing or human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: This research does not involve human subjects, making IRB approval non-applicable.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs were used for writing assistance, structural editing, and formatting purposes.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.