AMBIENT PHYSICS: TRAINING NEURAL PDE SOLVERS WITH PARTIAL OBSERVATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

In many scientific settings, acquiring complete observations of PDE coefficients and solutions can be expensive, hazardous, or impossible. Recent diffusion-based methods can reconstruct fields given partial observations, but require complete observations for training. We introduce Ambient Physics, a framework for learning to reconstruct coefficient-solution pairs directly from partial observations, without requiring a single complete observation. Unlike self-supervised methods, which create supervision by masking complete observations and then predicting the held-out content; Ambient Physics randomly masks a subset of already-observed measurements and supervises only where measurements exist, so the model cannot distinguish "truly unobserved" from "artificially unobserved", and must produce plausible predictions everywhere. Compared with prior diffusion-based methods, Ambient Physics achieves state-of-the-art reconstruction performance: a 61.82% reduction in average overall error while using 500× fewer function evaluations. We provide theoretical justification for learning directly from partial observations and identify a "one-measurement transition": masking a single already-observed measurement enables learning from partial observations across architectures and measurement patterns. We also extend our empirical evaluation to super-resolution tasks and real-world ERA5 weather fields. Ambient Physics thus enables scientific progress in settings where complete observations are unavailable.

1 INTRODUCTION

Partial differential equations are fundamental to science and engineering (Evans, 2022; Higham et al., 2015). Problems are often categorized as either the *forward problem*: to predict the solution from a given set of coefficients, or the *inverse problem*: to infer the coefficients from a given solution. In practice, measurements of coefficients and solutions are often incomplete. Climate scientists approximate atmospheric states from sparse weather stations (Lynch, 2006), and geophysicists determine properties of the Earth's subsurface from limited borehole data (Aki & Richards, 2002).

Partial observations present significant challenges. Classical numerical solvers require defined coefficients and boundary conditions, and inverse methods require regularization or strong priors; thus, predicting solutions from partial observations of coefficients or inferring coefficients from partial observations of solutions can elicit large uncertainties. To tackle this challenge, researchers have proposed various learning-based methods to reconstruct partial observations into physically plausible complete observations. In particular, founded on Diffusion Posterior Sampling (DPS) (Chung et al., 2022), DiffusionPDE and FunDPS make use of diffusion-based methods to learn the joint distribution of coefficient-solution pairs $p(a, u)$, where a and u denote the coefficients and solution respectively, and then apply DPS to sample from the posterior $p(a, u|A_a a, A_u u)$, where A_a and A_u denote measurement operators, and $A_a a$ and $A_u u$ are the corresponding partial observations (Huang et al., 2024; Yao et al., 2025).

Both DiffusionPDE and FunDPS exhibit exceptional reconstruction performance. With partial observations of only 3%, these diffusion-based methods can sample physically plausible complete observations from the posterior distribution; and across diverse PDEs, achieve relative L_2 errors in the range of 2–20%. In contrast, more traditional neural PDE solvers (DeepONet (Lu et al., 2019), FNO (Li et al., 2020), PINN (Raissi et al., 2019), PINO (Li et al., 2024)) struggle with partial observations and achieve errors that are often an order of magnitude larger. Still, there is a fundamental motivational

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

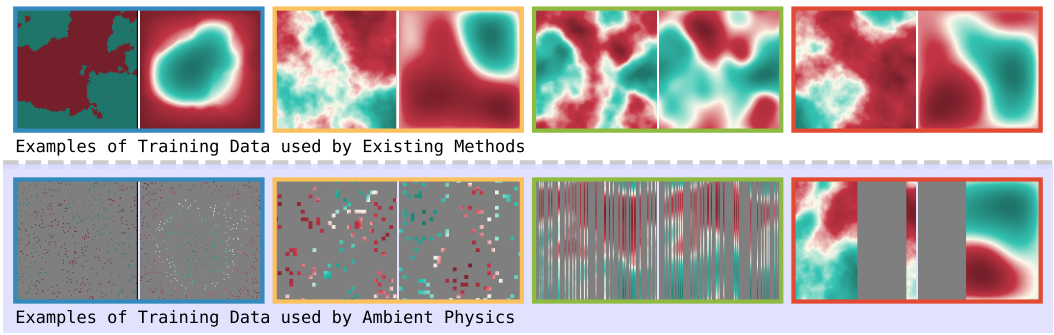


Figure 1: **Examples of Training Data used by Existing Methods vs Ambient Physics. Top:** Existing diffusion-based methods require access to complete observations to learn the joint distribution of coefficient-solution pairs. **Bottom:** Ambient Physics learns directly from partial observations, without requiring a single complete observation. **Left to Right:** Darcy flow with point observations, Helmholtz with patch observations, Navier-Stokes with slice observations, and Poisson with window observations. Within each panel, the left field represents the coefficients (or initial condition) and the right field represents the solution (or final state).

misalignment: DiffusionPDE and FunDPS require complete observations to learn the joint distribution of coefficient-solution pairs, having been pre-trained on 50,000 complete observations. For many real-world problems—especially those that would most benefit from reconstruction—acquiring 50,000 complete measurements is infeasible. One idealistic workaround is *naive training*: training the model directly on available partial observations. However, naive training fails catastrophically (see Figure 4)—in this setting, the model learns only the distribution of partial observations $p(A_a a, A_u u)$, predicting accurately at observed locations and arbitrarily elsewhere.

To address the motivational misalignment described above, we propose Ambient Physics, a framework for learning to reconstruct coefficient-solution pairs directly from partial observations. The key idea is to additionally mask a subset of already-observed measurements and supervise on the full set of available measurements, so that the model cannot distinguish "truly unobserved" from "artificially unobserved" and must produce physically plausible predictions everywhere to minimize error. [1] Concretely, our contributions are as follows:

- We introduce Ambient Physics, a framework for learning to reconstruct coefficient-solution pairs directly from partial observations, and provide theoretical justification for learning directly from partial observations (see section 3 and Appendix C).
- We demonstrate that, across Darcy flow, Helmholtz, Navier-Stokes, and Poisson, Ambient Physics achieves a 61.82% reduction in average overall error while using 500× fewer function evaluations compared with prior diffusion-based methods (see section 4).
- We demonstrate that the Ambient Physics objective is not tied to diffusion or flow matching by adapting traditional neural PDE solvers to learn from partial observations (see section 4).
- We study Ambient Physics across different measurement patterns where we identify a "one-measurement transition" in which masking a single already-observed measurement enables learning from partial observations. We also study different observation rates, test-time observation shifts, sample sizes, and extend our empirical evaluation to super-resolution tasks and real-world ERA5 weather fields (see section 4 and Appendix).

[1] This differs from self-supervised methods such as masked autoencoders (He et al., 2022), which create supervision by masking complete observations and then predicting the held-out content; Ambient Physics randomly masks a subset of already-observed measurements and supervises only where measurements exist. We provide a comparison in Appendix I.

2 BACKGROUND

Partial Differential Equations. We consider partial differential equations of the form:

$$\mathcal{L}_a u = 0, \quad \text{on } \Omega,$$
$$u_{|\partial\Omega} = g,$$

where $\mathcal{L}_a$ represents a differential operator, $a : \Omega \to \mathbb{R}^n$ represents the coefficients, $u : \Omega \to \mathbb{R}^n$ represents the solution, and $g : \partial\Omega \to \mathbb{R}^{n_b}$ represents the boundary conditions. For time-dependent PDEs, we may view a as the initial condition $u(0)$ and u as the final state $u(T)$ at time T.

The forward problem is to map a given set of coefficients (or initial conditions) a to the solution u, while the inverse problem is to map the solution u to a given set of coefficients a. These problems induce corresponding operators: the forward operator $\mathcal{M}_a : a \to u$, and the inverse operator $\mathcal{M}_u : u \to a$.

Neural PDE Solvers. A neural PDE solver approximates the forward or inverse operator $\mathcal{M}$ with a learned model $\mathcal{M}^\theta$. These methods are often trained by minimizing a data loss,

$$\|x_{\text{true}} - x_{\text{pred}}\|_2,$$

where x denotes the coefficients a or solution u, sometimes combined with a PDE loss that penalizes violations of the governing equation. However, in many cases, incorporating the PDE loss only yields marginal improvements (Krishnapriyan et al., 2021; Wang et al., 2022), and with incomplete inputs, the required derivatives become difficult to compute (Huang et al., 2024).

Neural PDE solvers are evaluated by measuring the error between ground truth and predictions. In practice, specific requirements of the problem or task may dictate the choice of the metric. For example, climate models may prioritize long-term accuracy and stability over pointwise accuracy, while engineering designs may prioritize maximum tolerances in critical regions over average performance. Following prior work, we report the *relative L_2 error*

$$\frac{\|x_{\text{true}} - x_{\text{pred}}\|_2}{\|x_{\text{true}}\|_2}$$

for continuous-valued fields, which is particularly useful for comparing performance across fields with different magnitudes and across different PDEs. For discrete-valued fields, we report the *misclassification rate*

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left[x_{\text{true}}^{(i)} \neq x_{\text{pred}}^{(i)}\right].$$

Neural PDE solvers are commonly grouped into two categories: physics-informed methods and operator-learning methods. Physics-Informed Neural Networks (PINNs) incorporate physical laws, typically described by partial differential equations, into the loss functions to guide training (Raissi et al., 2019). Several extensions improve parallelization and scalability: conservative PINNs (cPINNs) apply separate neural networks on subdomains and enforce flux continuity at subdomain interfaces (Jagtap et al., 2020); eXtended PINNs (XPINNs) generalize this domain decomposition approach to arbitrary domains, geometries, and PDE types (Jagtap & Karniadakis, 2020). In contrast, neural operators learn mappings between infinite-dimensional function spaces (Kovachki et al., 2023). Deep Operator Networks (DeepONets) use two sub-networks that combine via a dot product to produce predictions at queried locations (Lu et al., 2019). Fourier Neural Operators (FNOs) parameterize the solution operator via learnable integral kernels in Fourier space (Li et al., 2020). Many extensions improve expressivity and transferability: Factorized-FNO (F-FNO) uses separable spectral layers and modern residual connections and training strategies (Tran et al., 2021); Geometry-aware FNO (Geo-FNO) uses message passing to adapt operator-learning to irregular geometries (Li et al., 2023). Physics-Informed Neural Operator (PINO) incorporates PDE residuals into the loss function to bridge the gap between physics-informed and operator-learning methods (Li et al., 2024). These methods have been applied across a wide range of domains, including heat transfer (Zobeiry & Humfeld, 2021; Cai et al., 2021; Lu & Wang, 2024; Cao et al., 2024), electromagnetic fields (Baldan et al., 2023; Piao et al., 2024; Peng et al., 2022; Augenstein et al., 2023), and fluid mechanics (Jin et al., 2021; Eivazi et al., 2022; Luo et al., 2023; Azizzadenesheli et al., 2024).

Reconstruction Under Partial Observations. Several works study reconstruction under partial observations, particularly in fluid mechanics where real-world data are often collected by a limited number of probes, yielding observations only at sparse locations or along a small number of slices, and sometimes at low-resolution. For instance, transformer-based architectures have been used to reconstruct two- and three-dimensional flows from sparse locations (Zhang et al., 2025), and convolutional neural networks have been used to reconstruct three-dimensional flows from two-dimensional cross-sectional slices (Matsuo et al., 2024). More recently, diffusion models have been adopted for this problem because they generate high-quality samples and support uncertainty quantification. A number of works modify diffusion training or sampling by incorporating PDE residuals, and demonstrate applications to field reconstruction and super-resolution tasks for two- and three-dimensional flows (Yousif et al., 2021; Shu et al., 2023; Jacobsen et al., 2025; Amorós-Trepat et al., 2026). DiffusionPDE (Huang et al., 2024) learns the joint distribution of coefficient-solution pairs $p(a, u)$ and then applies Diffusion Posterior Sampling (Chung et al., 2022) to guide the denoising process using partial observations and PDE residuals, while FunDPS (Yao et al., 2025) extends this framework to function-space diffusion models. Importantly, while these works are motivated by scientific settings in which complete observations may be unavailable, their experimental setups use complete observations for training and validation, with datasets ranging from 1,000–50,000 complete observations. These supervised reconstruction settings are therefore closely related to masked prediction and masked autoencoders (He et al., 2022): complete observations are masked to act as inputs, and losses can be evaluated against complete observations; we provide further discussion in Appendix I. This motivates methods that can learn directly from partial observations. Closest to our setting, Incomplete Data, Complete Dynamics studies diffusion-based learning of physical dynamics from incomplete observations (Zhou et al., 2025); however, its focus is time-dependent dynamics imputation, where multiple partially observed frames provide complementary coverage over time—a much easier problem. We provide a comparison in Appendix I, including differences in assumptions, masking objectives, baselines, and empirical performance.

Generative Priors for Inverse Problems. Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020) are effective generative priors for inverse problems, enabling reconstruction by generating samples that are consistent with a learned prior and the available observations. This is often done by training an unconditional diffusion model on complete observations and then performing approximate posterior sampling by guiding the denoising process with a measurement-consistency objective, optionally augmented with additional constraints. Examples include DDRM for linear inverse problems (Kawar et al., 2022) and DPS for more general inverse problems (Chung et al., 2022). However, these methods assume access to fully observed training samples. In contrast, "ambient" methods aim to learn a generative model of the underlying data distribution given only incomplete or noisy measurements. AmbientGAN shows that learning the target distribution from partially observed training samples is possible (Bora et al., 2018), and Ambient Diffusion extends this perspective to diffusion models (Daras et al., 2023). This enables learning generative priors directly from partial observations, without requiring a single complete observation.

3 METHOD

The Reconstruction Problem. The PDE literature typically distinguishes between the *classical forward problem* $a \to u$ and the *classical inverse problem* $u \to a$. In this work, we propose a generalized formulation that treats both as special cases and explicitly addresses real-world partial observations.

Let $A_a : \mathbb{R}^n \to \mathbb{R}^n$ and $A_u : \mathbb{R}^n \to \mathbb{R}^n$ represent (possibly non-invertible) measurement operators (e.g., binary masks) that act on coefficients and solutions, respectively. Given partial observations $(A_a a, A_u u)$, the *reconstruction problem* is to predict a coefficient-solution pair (a, u) that agrees with the partial observations and is physically plausible. This perspective is closely analogous to image inpainting with multiple channels and per-channel masks, as well as many other tasks in computational imaging (e.g., X-ray computed tomography).

This generalized formulation treats the classical forward and inverse problems as special cases. Let $\mathbf{I}$ denote the identity operator and $\mathbf{0}$ denote the null operator. When $A_a = \mathbf{I}$ and $A_u = \mathbf{0}$, the reconstruction problem reduces to the classical forward problem:

$$(\mathbf{I}a, \mathbf{0}u) = (a, \emptyset) \to (a, u).$$

Conversely, when $A_a = \mathbf{0}$ and $A_u = \mathbf{I}$, the reconstruction problem reduces to the classical inverse problem:

$$(\mathbf{0}a, \mathbf{I}u) = (\emptyset, u) \to (a, u).$$

It also captures settings not covered by the classical dichotomy. For example, when $A_a \notin \{\mathbf{0}, \mathbf{I}\}$ and $A_u = \mathbf{0}$, the reconstruction problem becomes *coefficient inpainting*, where the goal is to predict the coefficients given partial coefficients. More generally, both the coefficients and solution may be incomplete, i.e., $A_a, A_u \notin \{\mathbf{0}, \mathbf{I}\}$. All of these cases fall under our proposed generalized formulation. In the remainder of this work, we use "reconstruction problem" to refer to this unified task of predicting coefficient-solution pairs (a, u) given partial observations $(A_a a, A_u u)$, including the classical forward and inverse problems as special cases.

Objective. Our objective is to solve the reconstruction problem by learning from the joint distribution of coefficient-solution pairs $p(a, u)$, and then, at inference, estimating (a, u) from the posterior $p(a, u \mid A_a a, A_u u)$. Prior diffusion-based methods adopt this strategy, but they require access to complete observations (a, u) during training. In many scientific and engineering settings, however, only partial observations $(A_a a, A_u u)$ are available. This makes learning the joint distribution strictly more challenging: we must learn $p(a, u)$ directly from partial observations $(A_a a, A_u u)$.

Flow Matching. In this work, we adopt flow matching as our generative modeling framework (Lipman et al., 2022). Flow matching learns a transport from a simple base distribution (e.g., Gaussian noise $x_0 \sim \mathcal{N}(0, I)$) to the data distribution (e.g., $x_1 \sim p_{\text{data}}$) by sampling an interpolation between noise x_0 and data x_1 and training a neural network to predict the corresponding flow field. Flow matching also supports conditional generation (e.g., posterior sampling).

We use rectified flow (Liu et al., 2022),[2] which has a particularly simple training and sampling procedure and typically performs competitively with more complex score-based diffusion models. While common practice for rectified flow is to parameterize the network to predict the instantaneous velocity, we parameterize our network to predict the final state x_1.

Naive Training. With only partial observations available, a natural baseline is to train a conditional flow matching model directly on the available measurements.

Let $\epsilon_a, \epsilon_u \sim \mathcal{N}(0, I)$ denote independent Gaussian noise and let $t \sim \mathcal{U}([0, 1))$ denote the flow time. Following rectified flow, one forms *mid-flow states* (i.e., noised partial observations) by linearly interpolating between noise and the measurements:

$$a_t = t \cdot (A_a a) + (1 - t) \cdot \epsilon_a, \qquad u_t = t \cdot (A_u u) + (1 - t) \cdot \epsilon_u.$$

Then, one can train a conditional model $\mathcal{M}^\theta$ to predict the coefficient-solution pair (a, u) from the mid-flow states (a_t, u_t) while conditioning on partial observations $(A_a a, A_u u)$, masks (A_a, A_u), and time t:

$$\mathcal{M}^\theta(A_a a_t, A_u u_t, A_a a, A_u u, A_a, A_u, t) = (\hat{a}, \hat{u}).$$

Conditioning on partial observations $(A_a a, A_u u)$ is necessary so that, at inference, generated fields depend on, and agree with, the available measurements; removing it would yield unconditional samples that are not constrained to match partial observations $(A_a a, A_u u)$. Since only partial observations $(A_a a, A_u u)$ are available for training, supervision is restricted to the available measurements:

$$\mathcal{L}_{\text{meas}} = \|A_a a - A_a \hat{a}\|_2^2 + \|A_u u - A_u \hat{u}\|_2^2.$$

This naive training procedure leads to catastrophic failure: the model predicts accurately only at the available measurements and arbitrarily elsewhere (Figure 4). Since the loss $\mathcal{L}_{\text{meas}}$ penalizes errors only at observed locations while the full partial observations $(A_a a, A_u u)$ are also provided as conditioning, the loss can be minimized by simply copying the observed values into $(\hat{a}, \hat{u})$, yielding near-perfect predictions at observed locations. Everywhere else, however, the model receives no training signal. Predictions at unobserved locations do not affect $\mathcal{L}_{\text{meas}}$, so the model can make arbitrary predictions.

[2]Note that we use rectified flow for our main instantiation, but the Ambient Physics framework is completely independent of this choice and can be applied to other generative models or deterministic neural PDE solvers.

Ambient Physics Training. We propose Ambient Physics training to learn the joint distribution of coefficient-solution pairs directly from partial observations $(A_a a, A_u u)$. Naive training provides no training signal beyond the measurements used for conditioning. Ambient Physics creates such a signal by withholding some already-observed measurements from the model input while still supervising on them. The key idea is to randomly mask a subset of already-observed measurements and supervise on these artificially masked measurements. As a result, the model cannot distinguish "truly unobserved" from "artificially unobserved" and must produce physically plausible predictions across the entire domain to minimize error.

Let $B_a : \mathbb{R}^n \to \mathbb{R}^n$ and $B_u : \mathbb{R}^n \to \mathbb{R}^n$ denote additional masking operators that further mask the partial observations. In particular, B_a and B_u are chosen so that they mask only a subset of already-observed measurements (i.e., they zero-out some entries that are observed under A_a and A_u, and never reveal new ones). Consequently, the inputs $(B_a A_a a,\ B_u A_u u)$ contain strictly less information than the training observations $(A_a a,\ A_u u)$. For notational convenience, we define $\tilde{B}_x := B_x A_x$.

Let $\epsilon_a, \epsilon_u \sim \mathcal{N}(0, I)$ denote independent Gaussian noise and let $t \sim \mathcal{U}([0, 1))$ denote the flow time. Following rectified flow, we form mid-flow states by linearly interpolating between noise and the masked partial observations:

$$a_t = t \cdot (\tilde{B}_a a) + (1 - t) \cdot \epsilon_a, \qquad u_t = t \cdot (\tilde{B}_u u) + (1 - t) \cdot \epsilon_u.$$

Then, we train a conditional model $\mathcal{M}^\theta$ to predict the clean fields (a, u) from (a_t, u_t) while conditioning on the masked partial observations $(\tilde{B}_a a, \tilde{B}_u u)$, masks $(\tilde{B}_a, \tilde{B}_u)$, and time t:

$$\mathcal{M}^\theta(\tilde{B}_a a_t, \tilde{B}_u u_t, \tilde{B}_a a, \tilde{B}_u u, \tilde{B}_a, \tilde{B}_u, t) = (\hat{a}, \hat{u}).$$

Critically, the model is given strictly less information than the training observations, but it is still supervised on all available measurements $(A_a a, A_u u)$:

$$\mathcal{L}_{\text{meas}} = \|A_a a - A_a \hat{a}\|_2^2 + \|A_u u - A_u \hat{u}\|_2^2.$$

In contrast to naive training, which can minimize $\mathcal{L}_{\text{meas}}$ by simply copying the observed values and predicting arbitrarily elsewhere, Ambient Physics withholds a random subset of the available measurements from the conditioning using (B_a, B_u). The model is still supervised on the full set of available measurements $(A_a a, A_u u)$ through $\mathcal{L}_{\text{meas}}$, so it must predict the held-out ("artificially unobserved") measurements from the remaining context. Since the model cannot distinguish truly unobserved points from artificially unobserved ones, the minimizer of $\mathcal{L}_{\text{meas}}$ is the posterior mean of the clean fields across the entire domain given the model inputs, the Bayes-optimal reconstruction under squared error. In Appendix C, we provide a population-level justification showing that, under standard coverage assumptions, the Ambient Physics objective recovers the conditional Bayes predictor of the clean coefficient-solution fields from further-corrupted observations.

Ambient Physics Sampling. During training, the Ambient Physics model receives as input the masked mid-flow states $(\tilde{B}_a a_t, \tilde{B}_u u_t)$ and conditions on the masked partial observations $(\tilde{B}_a a, \tilde{B}_u u)$, the corresponding masks $(\tilde{B}_a, \tilde{B}_u)$, and time t. At inference, we are given partial observations $(A_a a, A_u u)$ and our goal is to estimate (a, u) from these partial observations. Crucially, sampling must replicate the same masking and conditioning structure used during training.

We first sample Gaussian noise $a_0, u_0 \sim \mathcal{N}(0, I)$ and masking operators (B_a, B_u) from the same distribution used during training, i.e., $B_a \sim p(B_a \mid A_a)$ and $B_u \sim p(B_u \mid A_u)$. We then numerically integrate the rectified-flow ODE from $t = 0$ to $t \approx 1$. To do so, we discretize time as $t_k = k/K$ for $k = 0, \ldots, K - 1$, where K is the number of sampling steps (function evaluations), and we maintain states (a_k, u_k) that approximate (a_{t_k}, u_{t_k}). At each step k, we predict the corresponding clean fields by applying the same masking/conditioning structure used during training:

$$\mathcal{M}^\theta(\tilde{B}_a a_k, \tilde{B}_u u_k, \tilde{B}_a a, \tilde{B}_u u, \tilde{B}_a, \tilde{B}_u, t_k) = (\hat{a}_k, \hat{u}_k).$$

We then compute the rectified-flow velocity at t_k as

$$v_a(t_k) = \frac{\hat{a}_k - a_k}{(1 - t_k)}, \qquad v_u(t_k) = \frac{\hat{u}_k - u_k}{(1 - t_k)},$$

and step forward with an ODE solver (e.g., Euler or Heun). The final state (a_K, u_K) is returned as a reconstruction sample $(\hat{a}, \hat{u})$.

Sampling different additional masks yields an ensemble of reconstructions; in Appendix D, we use this to estimate consistency and pointwise uncertainty on held-out measurements.

Table 1: **Reconstruction Performance Under Partial Observations.** Relative L_2 error (%) (†misclassification rate) between ground truth and reconstructed coefficients ($\hat{a}$) and solutions ($\hat{u}$) from partial observations $(A_a a, A_u u)$ with 3% uniformly random measurements. *Note: In prior work these are reported as "Inverse" for coefficients and "Forward" for solutions.* "Partial Obs. Training" indicates whether training uses partial observations (✓) or complete observations (✗); "NFE" is the number of function evaluations. Across all PDEs and both fields, Ambient Flow achieves the lowest relative L_2 error despite training only on partial observations and sampling with just 1 NFE.

	Partial Obs. Training	NFE	Darcy Flow		Helmholtz		Navier-Stokes		Poisson	
			Coefficient†	Solution	Coefficient	Solution	Coefficient	Solution	Coefficient	Solution
DeepONet	✗	–	41.1%	38.3%	132.8%	123.1%	97.2%	103.2%	105.8%	155.5%
FNO	✗	–	49.3%	28.2%	218.2%	98.2%	96.0%	101.4%	232.7%	100.9%
PINN	✗	–	59.7%	48.8%	160.0%	142.3%	146.8%	142.7%	130.0%	128.1%
PINO	✗	–	49.2%	35.2%	216.9%	106.5%	96.0%	101.4%	231.9%	107.1%
DiffusionPDE	✗	2000	7.87%	6.07%	19.07%	12.64%	9.63%	3.78%	21.10%	4.88%
FunDPS	✗	200	6.78%	2.88%	20.07%	2.20%	9.87%	3.99%	24.04%	2.04%
		500	5.18%	2.49%	17.16%	2.13%	8.48%	3.32%	20.47%	1.99%
Ambient Flow	✓	1	**2.08%**	**0.76%**	**7.73%**	**0.47%**	**6.41%**	**0.98%**	**7.86%**	**0.48%**

4 RESULTS

Reconstruction Under Partial Observations. For the reconstruction problem, we compare traditional neural PDE solvers and prior state-of-the-art diffusion-based methods against Ambient Flow—our instantiation of Ambient Physics. The fundamental question is *whether Ambient Physics can learn to reconstruct coefficient-solution pairs directly from partial observations.*

In this experimental setup, we are given partial observations $(A_a a, A_u u)$, where A_a and A_u denote independently sampled binary masks that retain only 3% uniformly random measurements. We consider four diverse PDEs: Darcy flow, Helmholtz, Navier-Stokes, and Poisson. Reconstruction performance is measured using the relative L_2 error between ground truth and prediction, computed over the entire field for each coefficient/solution and averaged across all test examples. We compare against traditional neural PDE solvers (DeepONet (Lu et al., 2019), FNO (Li et al., 2020), PINN (Raissi et al., 2019), PINO (Li et al., 2024)), which learn directional mappings and therefore require two separate models per PDE. We also include diffusion-based methods (DiffusionPDE (Huang et al., 2024), FunDPS (Yao et al., 2025)), which learn the joint distribution of coefficient-solution pairs and then perform posterior sampling using the available partial observations. Importantly, all baselines and Ambient Flow are evaluated on the same reconstruction task; however, the baselines require complete observations for training, while Ambient Flow is trained only on partial observations.

In Ambient Flow, we follow the Ambient Physics procedure from section 3. Starting from partial observations $(A_a a, A_u u)$, we sample additional masks (B_a, B_u) that withhold one-third of the observed points. Thus, the model is conditioned on $(B_a A_a a,\ B_u A_u u)$, which retain only 2% uniformly random measurements overall, while we supervise on all available measurements $(A_a a, A_u u)$. We use the UNet of Ho et al. (2020) and sample with an Euler scheme. Implementation details are provided in Appendix A.

Table 1 summarizes reconstruction performance across PDEs and fields. Traditional neural PDE solvers perform poorly, with relative L_2 errors ranging from 28.2% to 232.7%—as expected, since these solvers are not designed for reconstruction, particularly under extreme sparsity (3% uniformly random measurements). Diffusion-based methods perform substantially better, with errors on reconstructed coefficients ranging from 5.18% to 24.04% and on reconstructed solutions ranging from 1.99% to 12.64%. Across all PDEs and both fields, Ambient Flow achieves state-of-the-art reconstruction performance, with errors on coefficients ranging from 2.08% to 7.86% and errors on solutions ranging from 0.47% to 0.98%. Compared to the prior state-of-the-art method (FunDPS at 500 NFE), Ambient Flow achieves a 61.82% reduction in average overall error while using 500$\times$ fewer function evaluations, despite being trained directly on partial observations. Specifically, it reduces error by 64.66% on Darcy flow, 66.44% on Helmholtz, 47.45% on Navier–Stokes, and 68.74% on Poisson, with reductions of 50.20% on coefficients and 73.44% on solutions.

These results address the fundamental question: *Ambient Physics can learn to reconstruct coefficient-solution pairs directly from partial observations*, achieving state-of-the-art reconstruction performance without any complete observations.

Table 2: **Reconstruction Performance Under Partial Observations.** Relative L_2 error (%) (†misclassification rate) between ground truth and reconstructed coefficients ($\hat{a}$) and solutions ($\hat{u}$) from partial observations $(A_a a, A_u u)$ with 3% uniformly random measurements. *Note: In prior work these are reported as "Inverse" for coefficients and "Forward" for solutions.* "Partial Obs. Training" indicates whether training uses partial observations (✓) or complete observations (✗). Red entries indicate settings that exhibit visible artifacts (Figure 2). Ambient FNO and Ambient UNet perform competitively with diffusion-based methods despite training only on partial observations.

	Partial Obs. Training		Darcy Flow		Helmholtz		Navier-Stokes		Poisson	
			Coefficient†	Solution	Coefficient	Solution	Coefficient	Solution	Coefficient	Solution
DeepONet	✗	–	41.1%	38.3%	132.8%	123.1%	97.2%	103.2%	105.8%	155.5%
FNO	✗	–	49.3%	28.2%	218.2%	98.2%	96.0%	101.4%	232.7%	100.9%
PINN	✗	–	59.7%	48.8%	160.0%	142.3%	146.8%	142.7%	130.0%	128.1%
PINO	✗	–	49.2%	35.2%	216.9%	106.5%	96.0%	101.4%	231.9%	107.1%
Ambient FNO	✓	–	6.79%	24.90%	22.86%	5.85%	25.78%	45.92%	22.16%	5.76%
		+Mask Conditioning	6.11%	27.20%	20.69%	5.35%	26.99%	41.93%	21.62%	7.15%
		+Joint Dist. Modeling	**5.38%**	**2.56%**	**16.57%**	**2.86%**	**8.84%**	**2.12%**	**17.20%**	**2.99%**
Ambient UNet	✓	–	2.55%	2.59%	9.16%	2.00%	8.25%	3.57%	9.65%	1.94%
		+Mask Conditioning	**2.21%**	2.57%	**7.97%**	1.99%	**7.35%**	3.36%	**8.32%**	1.87%
		+Joint Dist. Modeling	2.46%	**0.64%**	10.81%	**0.49%**	8.33%	**1.25%**	10.53%	**0.48%**

Beyond Diffusion and Flow Matching. Prior work observed that DeepONet and FNO perform poorly with sparse inputs (Huang et al., 2024; Yao et al., 2025), and therefore use complete observations during training and partial observations only at inference. We ask *whether Ambient Physics can adapt traditional neural PDE solvers to learn directly from partial observations.*

We consider the same experimental setup as before and start from an FNO backbone, which we adapt for reconstruction under partial observations. The goal is to learn directional reconstruction maps, $A_a a \to u$ and $A_u u \to a$, while supervising only on the available measurements. This objective can be viewed as an instance of Ambient Physics training: the model is conditioned on masked partial observations $(B_a A_a a, B_u A_u u)$ and supervised on the available measurements $(A_a a, A_u u)$. For the forward-direction model $A_a a \to u$, the model conditions only on partial coefficients, corresponding to $B_a = I$ and $B_u = 0$; for the inverse-direction model $A_u u \to a$, the model conditions only on partial solutions, corresponding to $B_a = 0$ and $B_u = I$. We refer to this model as Ambient FNO.

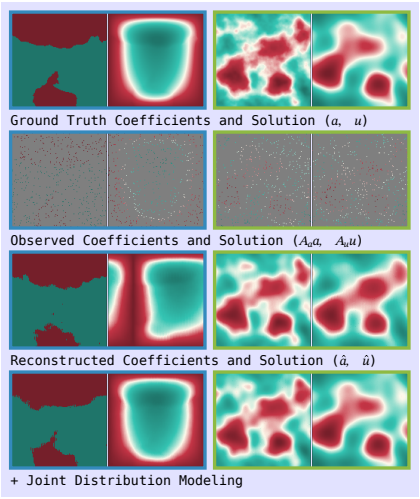


Figure 2: **Examples of Ambient FNO Reconstructions From Partial Observations. Top to Bottom:** Ground truth coefficients and solution (a, u), partial observations $(A_a a, A_u u)$ with 3% uniformly random measurements, Ambient FNO reconstructed coefficients and solutions $(\hat{a}, \hat{u})$, Ambient FNO +Joint Dist. Modeling reconstructed coefficients and solutions $(\hat{a}, \hat{u})$. **Left:** Darcy flow. **Right:** Navier-Stokes. Within each panel, the left field represents the coefficients (or initial condition) and the right field represents the solution (or final state).

Table 2 summarizes reconstruction performance across PDEs and fields. Compared to a vanilla FNO, Ambient FNO achieves a 74.27% reduction in average overall error despite training only on partial observations. However, the reconstructed solutions $\hat{u}$ occasionally exhibit visible artifacts: Darcy flow solutions can appear spatially shifted, and Navier-Stokes final states can resemble blurred initial conditions (see Figure 2). We test two modifications: +Mask Conditioning, which yields minimal improvement, and +Joint Dist. Modeling, where B_a and B_u mask only a subset of already-observed values ($B_a, B_u \notin \{I, 0\}$), which yields substantial improvements and eliminates visible artifacts. Compared to DiffusionPDE, Ambient FNO +Joint Dist. Modeling achieves a 36.16% reduction in average overall error and is competitive with FunDPS at 200 NFE. We also repeat these ablations using the UNet of Ho et al. (2020).

These results address our question: *Ambient Physics can adapt standard neural PDE architectures to learn directly from partial observations*, showing that the framework is not tied to diffusion or flow matching.

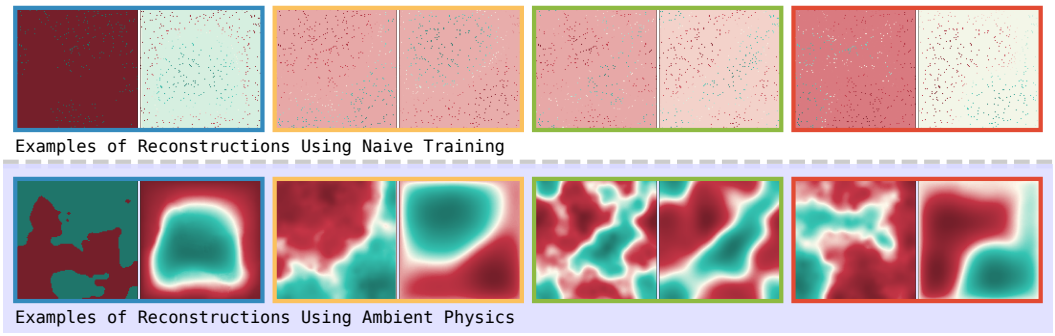


Figure 4: **Naive Training vs Ambient Physics Predictions from Partial Observations. Top:** Naively training on partial observations yields a model that predicts accurately only at observed locations and arbitrarily elsewhere. **Bottom:** Ambient Physics, *trained on the same partial observations,* yields a model that predicts physically plausible fields everywhere. **Left to Right:** Darcy flow with point observations, Helmholtz with patch observations, Navier-Stokes with slice observations, and Poisson with window observations. Within each panel, the left field represents the coefficients (or initial condition) and the right field represents the solution (or final state).

One-Measurement Transition to Ambient Physics. In Ambient Physics training (section 3), we randomly mask a subset of already-observed measurements (defined by the pattern) but still supervise on all available measurements. We ask how large this subset must be. We consider the same experimental setup as before. Starting from naive training ($B = I$), we sample masking operators that withhold $\{0, 1, 2, 4, 8, 16, 32, 64, 128, 256\}$ observed measurements. Figure 3 shows a sharp threshold: masking a single observed measurement yields an $\sim$10$\times$ reduction in coefficient error and an $\sim$100$\times$ reduction in solution error; performance improves up to 128 extra masked points. This "one-measurement transition" provides evidence that the improvements are primarily attributable to the Ambient Physics framework, rather than incidental choices such as architecture or optimization (see Appendix E).

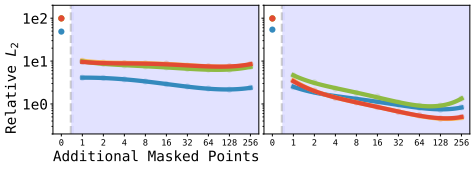


Figure 3: **One-Measurement Transition to Ambient Physics.** Relative L_2 error (log scale) versus the number of additional masked already-observed points used during training, across Darcy flow, Helmholtz, Navier-Stokes, and Poisson. **Left:** Coefficient error. **Right:** Solution error. The point at 0 additional masked points corresponds to naive training ($B = I$); masking even a single already-observed point leads to a $\sim$ 10–100$\times$ reduction in error.

Robustness Across Patterns and Super-Resolution. Beyond 3% uniformly random measurements, we study Darcy flow under structured observation patterns—*patch*, *column*, and *window* measurements—across multiple sparsity levels. Ambient Physics accurately reconstructs both coefficients and solutions across these patterns, showing that its performance is not limited to uniformly random sampling (see Appendix F). We also treat low-resolution measurements as partial observations of an underlying high-resolution field and show that Ambient Physics can super-resolve a 32×32 measurement into a 128×128 reconstruction without a single 128×128 training measurement.

5 Conclusions

Complete observations of PDE coefficients and solutions are often unavailable, yet existing reconstruction methods typically require complete observations to learn useful priors. We introduced Ambient Physics, a framework for learning joint coefficient–solution models directly from partial observations by additionally masking already-observed measurements and supervising only where measurements exist. Across four diverse PDEs, Ambient Flow achieves state-of-the-art reconstruction under extreme sparsity while reducing function evaluations by orders of magnitude. We further showed that Ambient Physics is not tied to diffusion or flow matching, and identified a "one-measurement transition" in which masking a single already-observed measurement is sufficient to move from catastrophic failure to effective learning. We discuss future directions and limitations in Appendix J.

AI USE STATEMENT

In this work, we used generative AI tools for implementing methods (generating code) and creating or modifying scientific figures or images (generating code). We have not used generative AI tools for generating synthetic data sets, helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, assisting with translation, cleaning and reformatting dataset, supporting qualitative and thematic data analysis, or interpreting results. Additionally, we used generative AI tools for editing to improve readability (checking grammar and spelling). We have reviewed all AI-assisted work. For example, LLM-generated code was manually verified and tested for correctness. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

We adhere to the ICLR Code of Ethics. Our work does not contain studies that involve human subjects, potentially harmful insights, methodologies, and applications, potential conflicts of interest and sponsorship, discrimination/bias/fairness concerns, privacy and security issues, or legal compliance and research integrity issues. Our work involves data sets generated by numerical methods (which we describe in Appendix A and Appendix B and are publicly available).

REPRODUCIBILITY STATEMENT

We have made significant effort to ensure reproducibility. Section 3 describes the Reconstruction Problem, the objective, naive training, and Ambient Physics training and sampling using clear and consistent terminology. Section 4 describes each experiment, its motivation and practical implementation. Appendix A describes hardware details, data details, architecture details and optimization details. Appendix B describes each partial differential equation. Source code, with configuration files for each experiment, will be made public upon acceptance. Together, our descriptions and source code enable independent researchers to reproduce and verify all of our findings.

REFERENCES

Keiiti Aki and Paul Richards. *Quantitative seismology*. MIT Press, 2002.

Marc Amorós-Trepat, Luis Medrano-Navarro, Qiang Liu, Luca Guastoni, and Nils Thuerey. Guiding diffusion models to reconstruct flow fields from sparse data. *Physics of Fluids*, 38(1), 2026.

Yannick Augenstein, Taavi Repan, and Carsten Rockstuhl. Neural operator-based surrogate solver for free-form electromagnetic inverse design. *ACS Photonics*, 10(5):1547–1557, 2023.

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, 6(5):320–328, 2024.

Marco Baldan, Paolo Di Barba, and David A Lowther. Physics-informed neural networks for inverse electromagnetic problems. *IEEE Transactions on Magnetics*, 59(5):1–5, 2023.

Ashish Bora, Eric Price, and Alexandros G Dimakis. Ambientgan: Generative models from lossy measurements. In *International conference on learning representations*, 2018.

Shengze Cai, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks for heat transfer problems. *Journal of Heat Transfer*, 143(6):060801, 2021.

Qianying Cao, Somdatta Goswami, and George Em Karniadakis. Laplace neural operator for solving differential equations. *Nature Machine Intelligence*, 6(6):631–640, 2024.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alex Dimakis, and Adam Klivans. Ambient diffusion: Learning clean distributions from corrupted data. *Advances in Neural Information Processing Systems*, 36:288–313, 2023.

Hamidreza Eivazi, Mojtaba Tahani, Philipp Schlatter, and Ricardo Vinuesa. Physics-informed neural networks for solving reynolds-averaged navier–stokes equations. *Physics of Fluids*, 34(7), 2022.

Lawrence C Evans. *Partial differential equations*, volume 19. American mathematical society, 2022.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

D Hendrycks. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly journal of the royal meteorological society*, 146(730):1999–2049, 2020.

Nicholas J Higham, Mark R Dennis, Paul Glendinning, Paul A Martin, Fadil Santosa, and Jared Tanner. *Princeton companion to applied mathematics*. Princeton University Press, 2015.

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

Jiahe Huang, Guandao Yang, Zichen Wang, and Jeong Joon Park. Diffusionpde: Generative pde-solving under partial observation. *Advances in Neural Information Processing Systems*, 37: 130291–130323, 2024.

Christian Jacobsen, Yilin Zhuang, and Karthik Duraisamy. Cocogen: Physically consistent and conditioned score-based generative models for forward and inverse problems. *SIAM Journal on Scientific Computing*, 47(2):C399–C425, 2025.

Ameya D Jagtap and George Em Karniadakis. Extended physics-informed neural networks (xpinns): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations. *Communications in Computational Physics*, 28(5), 2020.

Ameya D Jagtap, Ehsan Kharazmi, and George Em Karniadakis. Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering*, 365:113028, 2020.

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*, 426:109951, 2021.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in neural information processing systems*, 35:23593–23606, 2022.

Jean Kossaifi, Nikola Kovachki, Kamyar Azizzadenesheli, and Anima Anandkumar. Multi-grid tensorized fourier neural operator for high-resolution pdes. *arXiv preprint arXiv:2310.00120*, 2023.

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

Aditi Krishnapriyan, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W Mahoney. Characterizing possible failure modes in physics-informed neural networks. *Advances in neural information processing systems*, 34:26548–26560, 2021.

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

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023.

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*, 1(3):1–27, 2024.

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

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

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Lu Lu, Pengzhan Jin, and George Em Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.

Xiaoyi Lu and Yi Wang. Surrogate modeling for radiative heat transfer using physics-informed deep neural operator networks. *Proceedings of the Combustion Institute*, 40(1-4):105282, 2024.

Yining Luo, Yingfa Chen, and Zhen Zhang. Cfdbench: A large-scale benchmark for machine learning methods in fluid dynamics. *arXiv preprint arXiv:2310.05963*, 2023.

Peter Lynch. *The emergence of numerical weather prediction: Richardson's dream*. Cambridge University Press, 2006.

Mitsuaki Matsuo, Kai Fukami, Taichi Nakamura, Masaki Morimoto, and Koji Fukagata. Reconstructing three-dimensional bluff body wake from sectional flow fields with convolutional neural networks. *SN Computer Science*, 5(3):306, 2024.

Zhong Peng, Bo Yang, Yixian Xu, Feng Wang, Lian Liu, and Yi Zhang. Rapid surrogate modeling of electromagnetic data in frequency domain using neural operator. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12, 2022.

Shiyuan Piao, Hong Gu, Aina Wang, and Pan Qin. A domain-adaptive physics-informed neural network for inverse problems of maxwell's equations in heterogeneous media. *IEEE Antennas and Wireless Propagation Letters*, 23(10):2905–2909, 2024.

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*, 378:686–707, 2019.

Dule Shu, Zijie Li, and Amir Barati Farimani. A physics-informed diffusion model for high-fidelity flow field reconstruction. *Journal of Computational Physics*, 478:111972, 2023.

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

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. *arXiv preprint arXiv:2111.13802*, 2021.

Sifan Wang, Xinling Yu, and Paris Perdikaris. When and why pinns fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2022.

Jiachen Yao, Abbas Mammadov, Julius Berner, Gavin Kerrigan, Jong Chul Ye, Kamyar Azizzadenesheli, and Anima Anandkumar. Guided diffusion sampling on function spaces with applications to pdes. *arXiv preprint arXiv:2505.17004*, 2025.

Mustafa Z Yousif, Linqi Yu, and Hee-Chang Lim. High-fidelity reconstruction of turbulent flow from spatially limited data using enhanced super-resolution generative adversarial network. *Physics of Fluids*, 33(12), 2021.

Qian Zhang, Dmitry Krotov, and George Em Karniadakis. Operator learning for reconstructing flow fields from sparse measurements: an energy transformer approach. *Journal of Computational Physics*, pp. 114148, 2025.

Zihan Zhou, Chenguang Wang, Hongyi Ye, Yongtao Guan, and Tianshu Yu. Incomplete data, complete dynamics: A diffusion approach. *arXiv preprint arXiv:2509.20098*, 2025.

Navid Zobeiry and Keith D Humfeld. A physics-informed machine learning approach for solving heat transfer equation in advanced manufacturing and engineering applications. *Engineering Applications of Artificial Intelligence*, 101:104232, 2021.

A Implementation Details

Hardware Details. All models were trained on NVIDIA H100 GPUs. In total, we trained roughly 200 models for the main results and roughly 100 models for early-stage experimentation (e.g., preliminary exploration and hyperparameter tuning), for a total of $\sim$300 models. Across experiments, the average training time was approximately 250–300 minutes per model (including data loading and evaluation under our standard experimental setup).

Data Details. We use the same PDE datasets as DiffusionPDE (Huang et al., 2024) and FunDPS (Yao et al., 2025): Darcy flow, Helmholtz, unbounded Navier–Stokes, and Poisson. For each PDE, we use 50,000 training samples and 1,000 test samples at a spatial resolution of 128×128.

DiffusionPDE and FunDPS also include a bounded Navier–Stokes dataset; however, we found that approximately 10% of its samples contain non-physical artifacts, and we therefore deemed it unsuitable for our evaluation.

We emphasize that Ambient Physics models never receive fully observed fields as inputs during training or inference. Instead, partial observations $(A_a a, A_u u)$ are generated in the data loader by applying the corresponding measurement operators. In our implementation, the observation pattern is fixed per sample throughout training (i.e., from one epoch to the next), so unobserved locations remain unobserved for that sample.

Architecture Details. For Ambient Flow and Ambient UNet, we use the UNet backbone from DDPM (Ho et al., 2020). For Ambient FNO, we use a modern Fourier Neural Operator (FNO) backbone based on Kossaifi et al. (2023). Architecture hyperparameters are listed in Table 3.

Optimization Details. For all models, we use the AdamW optimizer (Loshchilov & Hutter, 2017). Optimization hyperparameters are listed in Table 3.

Table 3: Hyperparameters used by Ambient Flow, Ambient UNet, and Ambient FNO.

Hyperparameter	Ambient Flow	Ambient UNet	Ambient FNO
Architecture			
`fourier_channels`	64	–	–
`hidden_channels`	192	64	64
`ch_mult`	[1, 2, 3, 4]	[1, 2, 3, 4]	–
`is_attn`	[`False`, `False`, `True`, `True`]	[`False`, `False`, `False`, `True`]	–
`md_attn`	`True`	`True`	–
`dropout`	0.1	0.0	–
`attn_heads`	4	4	–
`attn_head_channels`	32	32	–
`depth`	–	–	8
`fft_modes`	–	–	16
`activation`	–	–	`GELU` (Hendrycks, 2016)
Optimization			
`epochs`	25	100	100
`clip_grad_max_norm`	1.0	1.0	1.0
`mixed_precision`	`True`	`False`	`False`
`lr`	1.0×10^{-4}	1.0×10^{-4}	1.0×10^{-4}
`weight_decay`	1.0×10^{-5}	1.0×10^{-5}	1.0×10^{-5}
`betas`	[0.9, 0.95]	[0.9, 0.999]	[0.9, 0.999]
`eps`	1.0×10^{-8}	1.0×10^{-8}	1.0×10^{-8}

B PARTIAL DIFFERENTIAL EQUATIONS

Darcy Flow Equation. The Darcy flow equation is a second-order elliptic partial differential equation used to model the flow of viscous incompressible fluid through a porous medium. Its two-dimensional variant can be expressed as:

$$-\nabla \cdot (a(\mathbf{x}) \nabla u(\mathbf{x})) = f(\mathbf{x}), \quad \mathbf{x} \in (0,1)^2,$$

with zero boundary conditions $u|_{\partial\Omega} = 0$. Here, the coefficient $a(\mathbf{x})$ represents the spatially varying permeability of the medium and typically takes binary values to model heterogeneous materials, while $f(\mathbf{x})$ (often set to a constant) represents the forcing term. The binary nature of $a(\mathbf{x})$ introduces sharp discontinuities in permeability, and learning-based methods must represent these discontinuities while keeping the solution field physically consistent.

Inhomogeneous Helmholtz Equation. The inhomogeneous Helmholtz equation is a second-order elliptic partial differential equation used to model time-harmonic wave propagation in heterogeneous media. Its two-dimensional variant can be expressed as:

$$\nabla^2 u(\mathbf{x}) + k^2 u(\mathbf{x}) = a(\mathbf{x}), \quad \mathbf{x} \in (0,1)^2,$$

with zero boundary conditions $u|_{\partial\Omega} = 0$. Here, k denotes the wave number (typically set to $k = 1$), and $a(\mathbf{x})$ represents the spatially varying source term, often modeled as a piecewise-constant function. The Laplacian term $\nabla^2 u$ governs diffusion and the $k^2 u$ term introduces oscillations, which can yield interference patterns and local resonances. When $k = 0$, the equation reduces to the Poisson equation.

Unbounded Incompressible Navier-Stokes Equation. The incompressible Navier-Stokes equations are a system of nonlinear partial differential equations that govern the dynamics of viscous fluid flow. For the unbounded velocity formulation, the two-dimensional variant can be expressed as:

$$\begin{aligned} \partial_t w(\mathbf{x},\tau) + \mathbf{v}(\mathbf{x},\tau) \cdot \nabla w(\mathbf{x},\tau) &= \nu \Delta w(\mathbf{x},\tau) + q(\mathbf{x}), \quad \mathbf{x} \in (0,1)^2,\ \tau \in (0,T], \\ \nabla \cdot \mathbf{v}(\mathbf{x},\tau) &= 0, \end{aligned}$$

where $w = \nabla \times \mathbf{v}$ is the vorticity, $\mathbf{v}(\mathbf{x},\tau)$ is the velocity field, $q(\mathbf{x})$ is a fixed force field, and ν is the kinematic viscosity. With $\nu = 10^{-3}$, the corresponding Reynolds number $Re = 1/\nu = 1000$ places the flow in the turbulent regime, where nonlinear advection and viscous diffusion generate vortices, shear layers, and energy cascades. The divergence-free constraint $\nabla \cdot \mathbf{v} = 0$ enforces incompressibility, coupling the velocity and vorticity fields through the stream function. Learning-based methods must model the time evolution while respecting these constraints.

Poisson Equation. The Poisson equation is a second-order elliptic partial differential equation that describes steady-state diffusion processes. Its two-dimensional variant can be expressed as:

$$\nabla^2 u(\mathbf{x}) = a(\mathbf{x}), \quad \mathbf{x} \in (0,1)^2,$$

with homogeneous Dirichlet boundary conditions $u|_{\partial\Omega} = 0$. Here, $a(\mathbf{x})$ represents the source term, which often exhibits piecewise-constant structure to model localized sources or sinks in heterogeneous domains. Although the solution field is typically smooth, discontinuities in $a(\mathbf{x})$ influence $u(\mathbf{x})$ through the elliptic operator, and models must respect the boundary conditions.

C THEORY

This section gives a population-level justification for Ambient Physics training. The argument adapts the Ambient Diffusion theory to the joint coefficient–solution setting. Informally, because the model is shown only a further-corrupted version of the available measurements but is supervised on all measurements that were originally observed, the population minimizer is the conditional Bayes estimator under the Ambient Physics measurement loss. For squared L_2 measurement loss, this estimator reduces to the conditional mean.

Let (a, u) denote a coefficient–solution pair, and let A_a and A_u be random measurement operators acting on a and u, respectively. Let B_a and B_u be additional random masking operators drawn conditionally on A_a and A_u, respectively, such that they only remove already-observed measurements and never reveal new measurements. Define the further-corrupted measurement operators

$$\widetilde{B}_a := B_a A_a, \qquad \widetilde{B}_u := B_u A_u.$$

Let

$$\epsilon_a, \epsilon_u \sim \mathcal{N}(0, I)$$

be independent Gaussian noise variables, independent of all measurement operators and of (a, u). For $t \in [0, 1)$, define the masked mid-flow states

$$a_t = t(\widetilde{B}_a a) + (1 - t)\epsilon_a, \qquad u_t = t(\widetilde{B}_u u) + (1 - t)\epsilon_u.$$

Let

$$Z := \big(\widetilde{B}_a a_t, \widetilde{B}_u u_t, \widetilde{B}_a a, \widetilde{B}_u u, \widetilde{B}_a, \widetilde{B}_u, t\big).$$

denote the information provided to the predictor, and consider measurable predictors

$$M(Z) = \big(m_a(Z), m_u(Z)\big).$$

Let ℓ_a and ℓ_u be nonnegative measurable losses applied to the measured coefficient and solution errors. The population Ambient Physics objective is

$$\mathcal{L}(M) = \mathbb{E}[\ell_a(A_a(m_a(Z) - a)) + \ell_u(A_u(m_u(Z) - u))] \,.$$

Theorem 1 (Population Bayes estimator of Ambient Physics)**.** *Assume that the conditional risks below are well-defined and admit measurable minimizers. Then any population minimizer $M^\star$ of $\mathcal{L}(M)$, up to almost-sure equality, is a conditional Bayes estimator under the Ambient Physics measurement loss. In particular, for almost every $Z = z$,*

$$M^\star(z) \in \arg\min_{(m_a, m_u)} \mathbb{E}[\ell_a(A_a(m_a - a)) + \ell_u(A_u(m_u - u)) \,|\, Z = z] \,.$$

Equivalently, Ambient Physics learns the Bayes-optimal predictor of the clean coefficient and solution fields under the loss induced by the available measurements and the information supplied to the model.

Proof. For any measurable predictor M, define the conditional risk

$$R(M; Z) := \mathbb{E}[\ell_a(A_a(m_a(Z) - a)) + \ell_u(A_u(m_u(Z) - u)) \,|\, Z] \,.$$

By the tower property,

$$\mathcal{L}(M) = \mathbb{E}[R(M; Z)] \,.$$

For each fixed value $Z = z$, the quantity $R(M; z)$ depends on the predictor only through the value $M(z) = (m_a(z), m_u(z))$. Therefore, the global population objective is minimized by choosing, for almost every z, a pointwise minimizer of the conditional risk:

$$M^\star(z) \in \arg\min_{(m_a, m_u)} \mathbb{E}[\ell_a(A_a(m_a - a)) + \ell_u(A_u(m_u - u)) \,|\, Z = z] \,.$$

If M fails to minimize this conditional risk on a set of z's with positive probability, then replacing $M(z)$ by a conditional-risk minimizer on that set strictly decreases $\mathcal{L}(M)$, contradicting population optimality. Hence every population minimizer must agree almost surely with a conditional Bayes estimator, and any measurable conditional Bayes estimator is population-optimal. This proves the claim. □

Corollary 1 (Squared L_2 measurement loss)**.** *Suppose the Ambient Physics measurement loss is the squared L_2 loss,*

$$\ell_a(r) = \|r\|_2^2, \qquad \ell_u(r) = \|r\|_2^2.$$

Assume further that the original measurement operators carry no additional information about the clean fields beyond the information given to the predictor,

$$\mathbb{E}[a \mid Z, A_a, A_u] = \mathbb{E}[a \mid Z], \qquad \mathbb{E}[u \mid Z, A_a, A_u] = \mathbb{E}[u \mid Z],$$

and that the conditional Gram matrices

$$G_a(Z) := \mathbb{E}[A_a^\top A_a \mid Z], \qquad G_u(Z) := \mathbb{E}[A_u^\top A_u \mid Z]$$

are positive definite almost surely. The positive-definiteness condition formalizes conditional coverage: every coordinate receives positive measurement probability given the information provided to the model. Then the unique population minimizer is the conditional mean predictor,

$$M^\star(Z) = (\mathbb{E}[a \mid Z], \mathbb{E}[u \mid Z]).$$

Thus, for squared L_2 measurement loss, the conditional Bayes estimator reduces to the conditional expectation of the clean coefficient and solution fields given the further-corrupted observations and masks supplied to the model.

Proof. Let

$$\bar{a} := \mathbb{E}[a \mid Z], \qquad \bar{u} := \mathbb{E}[u \mid Z],$$

and define the residuals

$$r_a(Z) := m_a(Z) - \bar{a}, \qquad r_u(Z) := m_u(Z) - \bar{u}.$$

Then

$$m_a(Z) - a = (\bar{a} - a) + r_a(Z), \qquad m_u(Z) - u = (\bar{u} - u) + r_u(Z).$$

Substituting these decompositions into the squared measurement loss gives

$$\begin{aligned}\mathcal{L}(M) = {} & \mathbb{E}\|A_a(\bar{a} - a)\|_2^2 + \mathbb{E}\|A_u(\bar{u} - u)\|_2^2 \\ & + \mathbb{E}\|A_a r_a(Z)\|_2^2 + \mathbb{E}\|A_u r_u(Z)\|_2^2 \\ & + 2\mathbb{E}[(\bar{a} - a)^\top A_a^\top A_a r_a(Z)] + 2\mathbb{E}[(\bar{u} - u)^\top A_u^\top A_u r_u(Z)].\end{aligned}$$

We first show that the cross terms vanish. For the coefficient term, since $r_a(Z)$ is measurable with respect to Z, the tower property gives

$$\mathbb{E}[(\bar{a} - a)^\top A_a^\top A_a r_a(Z)] = \mathbb{E}\big[\mathbb{E}[\bar{a} - a \mid Z, A_a, A_u]^\top A_a^\top A_a r_a(Z)\big].$$

By the no-additional-information assumption,

$$\mathbb{E}[a \mid Z, A_a, A_u] = \mathbb{E}[a \mid Z] = \bar{a},$$

so $\mathbb{E}[\bar{a} - a \mid Z, A_a, A_u] = 0$. Hence the coefficient cross term is zero. The solution cross term vanishes by the same argument. Therefore,

$$\mathcal{L}(M) = C + \mathbb{E}\|A_a r_a(Z)\|_2^2 + \mathbb{E}\|A_u r_u(Z)\|_2^2,$$

where

$$C = \mathbb{E}\|A_a(\bar{a} - a)\|_2^2 + \mathbb{E}\|A_u(\bar{u} - u)\|_2^2$$

does not depend on M. It remains only to minimize the residual terms. Applying the tower property again,

$$\mathbb{E}\|A_a r_a(Z)\|_2^2 = \mathbb{E}[r_a(Z)^\top G_a(Z) r_a(Z)], \qquad \mathbb{E}\|A_u r_u(Z)\|_2^2 = \mathbb{E}[r_u(Z)^\top G_u(Z) r_u(Z)].$$

Since $G_a(Z)$ and $G_u(Z)$ are positive definite almost surely, these quadratic forms are nonnegative and equal zero only when

$$r_a(Z) = 0, \qquad r_u(Z) = 0 \qquad \text{almost surely.}$$

Consequently, the unique minimizer satisfies

$$m_a(Z) = \bar{a} = \mathbb{E}[a \mid Z], \qquad m_u(Z) = \bar{u} = \mathbb{E}[u \mid Z],$$

which proves the claim. □

D CONSISTENCY AND POINTWISE UNCERTAINTY

In inherently ill-posed settings, there may be multiple valid reconstructions that are consistent with the available partial observations. We therefore evaluate validity by checking consistency with measurements that are intentionally withheld from the model, and quantify uncertainty using variability across multiple reconstructions, which we refer to as mask-resampling reconstruction uncertainty. Ambient Physics naturally supports this procedure: for the same partial observation, we can generate multiple reconstructions by varying the additional masking operator at inference. We then use the reconstruction error on the additionally masked, but known, measurements as a direct consistency metric, and use the pointwise variance across reconstructions as an uncertainty estimate.

To demonstrate this procedure, we consider an extreme ill-posed setting in which each example is observed only outside a contiguous window, so that 25% of the field is missing in a structured window pattern. This is substantially more difficult than uniformly random point or patch measurements, since the missing region is spatially contiguous and provides no local measurements. During training, we apply an additional window mask that removes a further 25% of the already observed region. Importantly, this additional mask is fixed for each sample throughout training rather than resampled every epoch; otherwise, the model could gradually recover the additionally masked region across epochs.

At inference, we sample multiple different additional masks and generate multiple reconstructions for the same partial observation. In Figure 5, we generate 16 reconstructions and compute both the ensemble mean and the pointwise variance. The figure shows the partially observed input, several reconstructions, the ensemble mean, and the pointwise variance. Table 4 reports representative consistency metrics for three individual predictions and for the ensemble mean, evaluated on measurements that are intentionally withheld from the model.

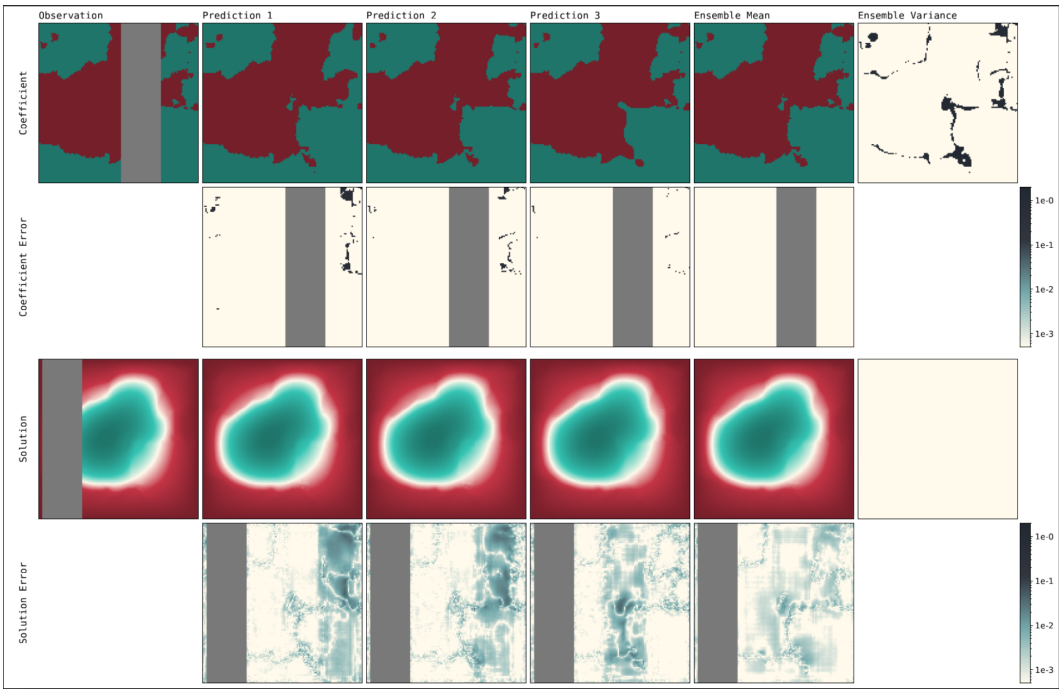


Figure 5: Consistency and pointwise uncertainty in an ill-posed structured-window setting without relying on full ground truth. Each example is observed only outside a contiguous window, and multiple reconstructions are generated by varying the additional masking operator at inference. The ensemble mean provides a consensus reconstruction, while the pointwise variance highlights regions where reconstructions disagree.

Table 4: Consistency metrics in the structured-window setting. Errors are computed on measurements that are intentionally withheld from the model. The ensemble mean achieves the lowest error, while variability across samples provides a pointwise uncertainty estimate.

Prediction	Relative L_2 (Coefficient)	Relative L_2 (Solution)
1	13.61%	0.96%
2	3.24%	0.82%
3	6.67%	0.60%
Ensemble mean	**0.00%**	**0.31%**

In this extreme ill-posed setting, the ensemble mean achieves the lowest error on measurements that are intentionally withheld from the model. In addition, the pointwise variance provides a natural uncertainty map, highlighting regions where multiple valid reconstructions disagree most.

For completeness, we repeat the experiment in a setting where full ground truth is available. Figure 6 shows qualitative examples, and Table 5 reports representative errors for three individual predictions and for the ensemble mean. We observe the same trend: the ensemble mean improves over individual reconstructions, and the pointwise variance identifies regions of high reconstruction ambiguity.

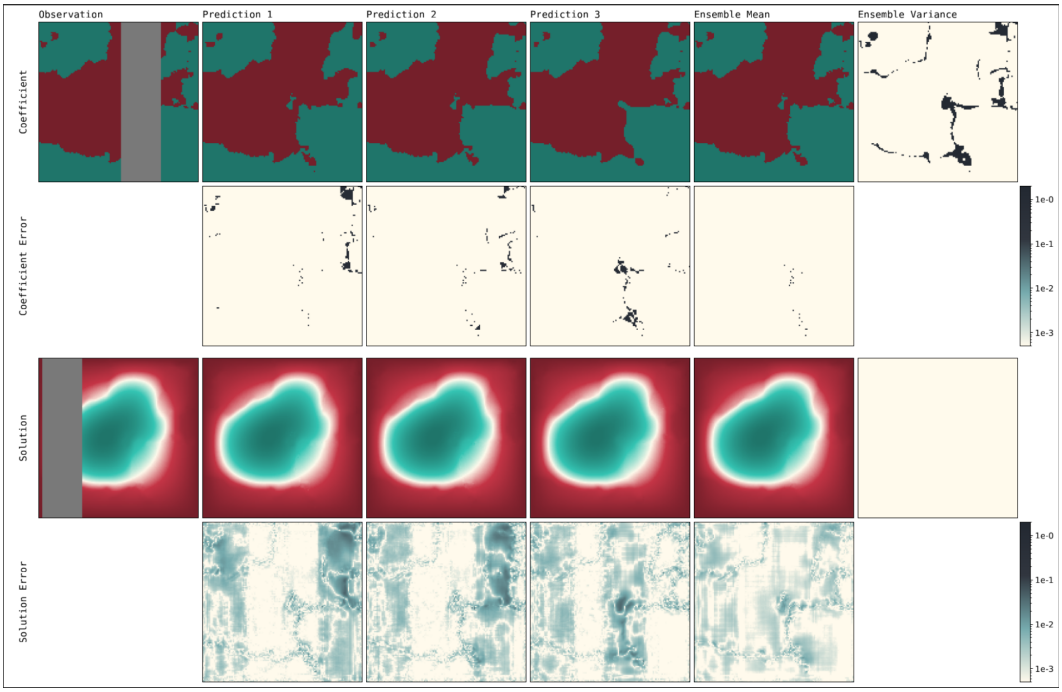


Figure 6: Consistency and pointwise uncertainty in the structured-window setting with full ground truth available for evaluation. Multiple reconstructions are generated for the same partial observation by varying the additional masking operator at inference. The ensemble mean reduces reconstruction error, while the pointwise variance captures spatial uncertainty.

This experiment provides a practical way to assess consistency and uncertainty even when full ground truth is unavailable at training or inference time. Consistency can be measured using additionally withheld observations, while uncertainty can be estimated from the pointwise variability of multiple reconstructions.

Table 5: Reconstruction errors in the structured-window setting when full ground truth is available for evaluation. The ensemble mean again improves over individual reconstructions.

Prediction	Relative L_2 (Coefficient)	Relative L_2 (Solution)
1	12.64%	0.95%
2	10.71%	0.80%
3	7.36%	0.61%
Ensemble mean	**3.43%**	**0.33%**

E One-Measurement Transition to Ambient Physics

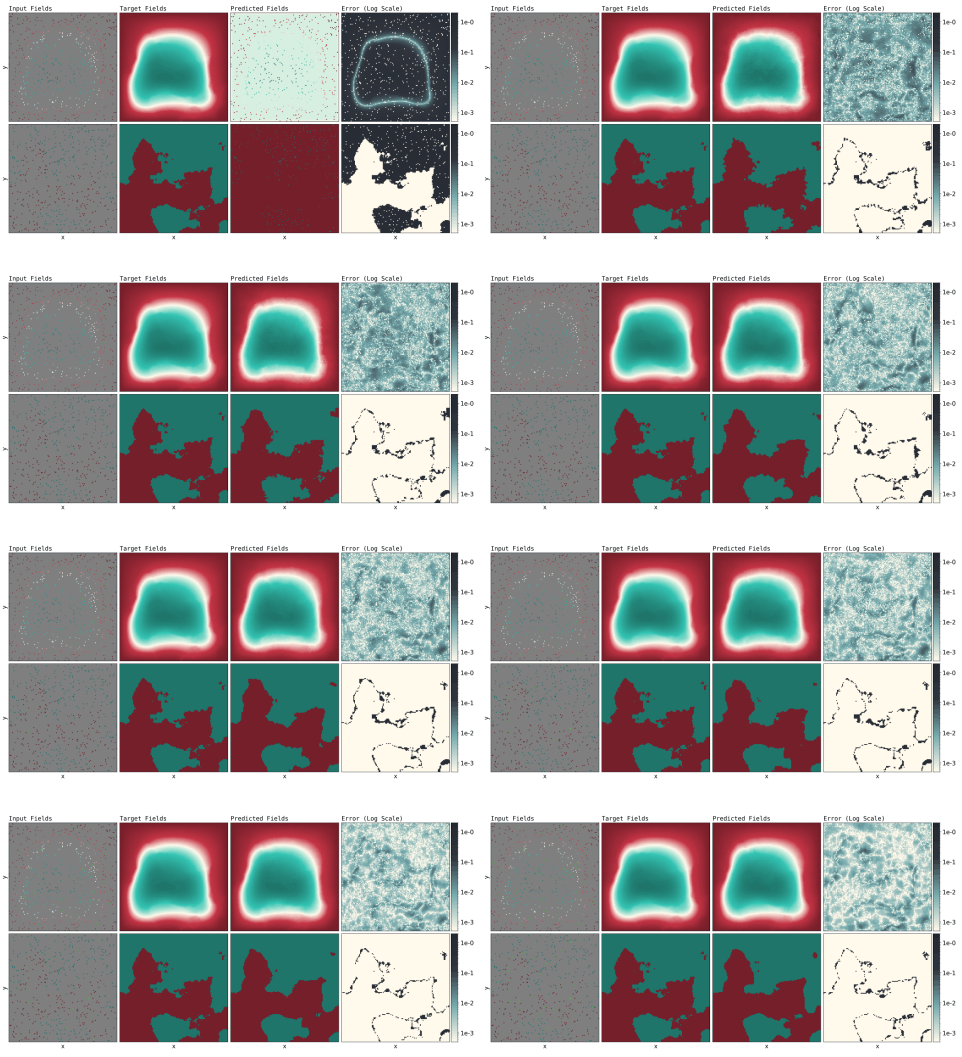


Figure 7: **Example of "One-Measurement Transition" (Darcy Flow).** Each subplot shows an Ambient Flow reconstruction from partial observations (3% uniformly random points), varying the number of additional masked (already-observed) points used during training (increasing across the grid; top-left: 0, which corresponds to naive training). **Within Each Subplot: Top:** Solution field. **Bottom:** Coefficient Field. **Left to Right:** Input Fields, Target Fields, Reconstructed Fields, Error (Log Scale). Within each input panel, gray pixels represent unobserved locations and green pixels represent masked measurements withheld from the model.

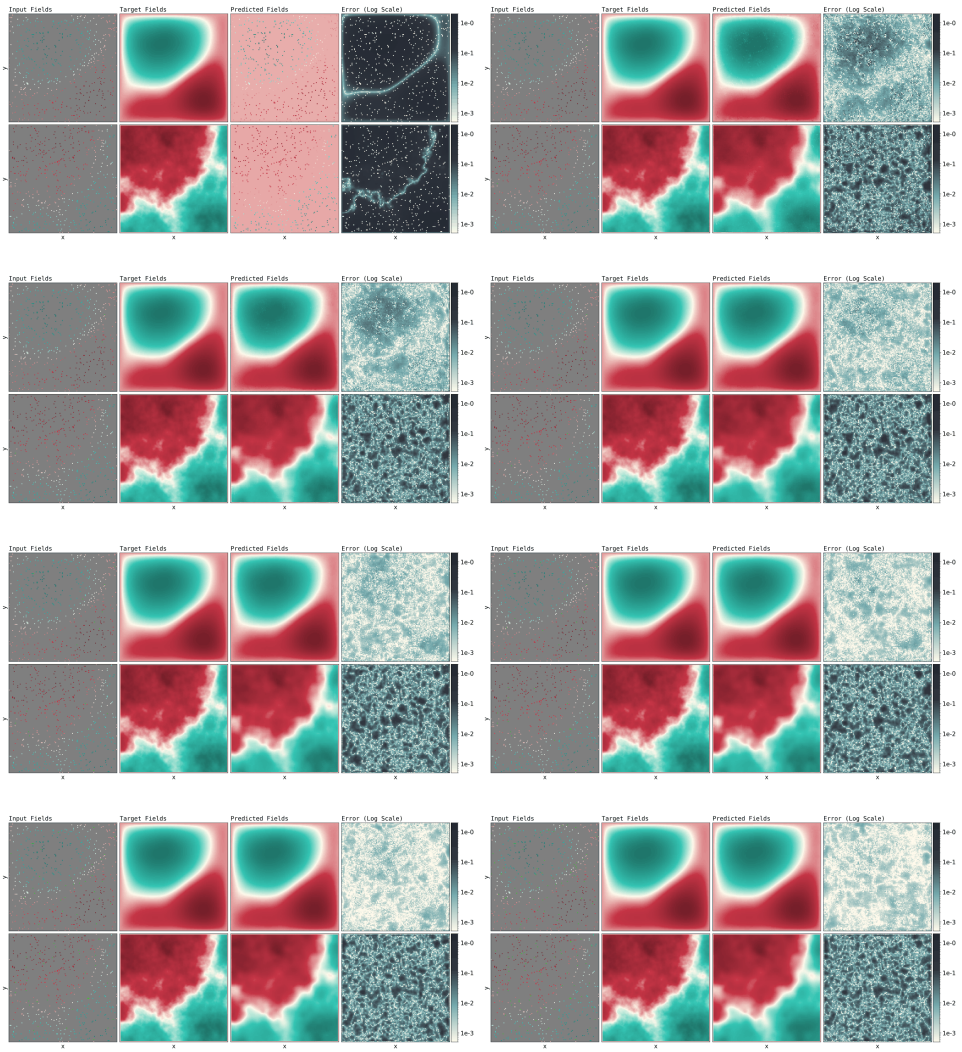


Figure 8: **Example of "One-Measurement Transition" (Helmholtz).** Each subplot shows an Ambient Flow reconstruction from partial observations (3% uniformly random points), varying the number of additional masked (already-observed) points used during training (increasing across the grid; top-left: 0, which corresponds to naive training). **Within Each Subplot: Top:** Solution field. **Bottom:** Coefficient Field. **Left to Right:** Input Fields, Target Fields, Reconstructed Fields, Error (Log Scale). Within each input panel, gray pixels represent unobserved locations and green pixels represent masked measurements withheld from the model.

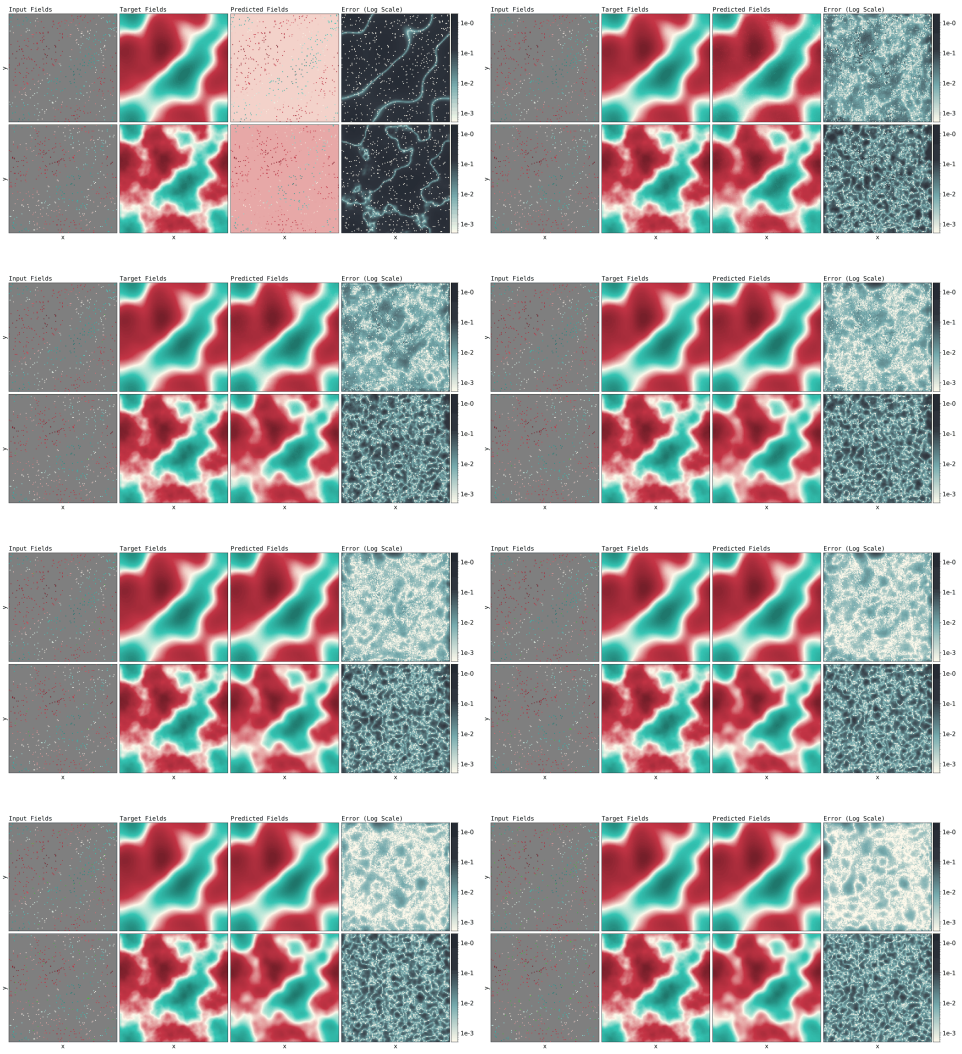


Figure 9: **Example of "One-Measurement Transition" (Navier-Stokes).** Each subplot shows an Ambient Flow reconstruction from partial observations (3% uniformly random points), varying the number of additional masked (already-observed) points used during training (increasing across the grid; top-left: 0, which corresponds to naive training). **Within Each Subplot: Top:** Solution field. **Bottom:** Coefficient Field. **Left to Right:** Input Fields, Target Fields, Reconstructed Fields, Error (Log Scale). Within each input panel, gray pixels represent unobserved locations and green pixels represent masked measurements withheld from the model.

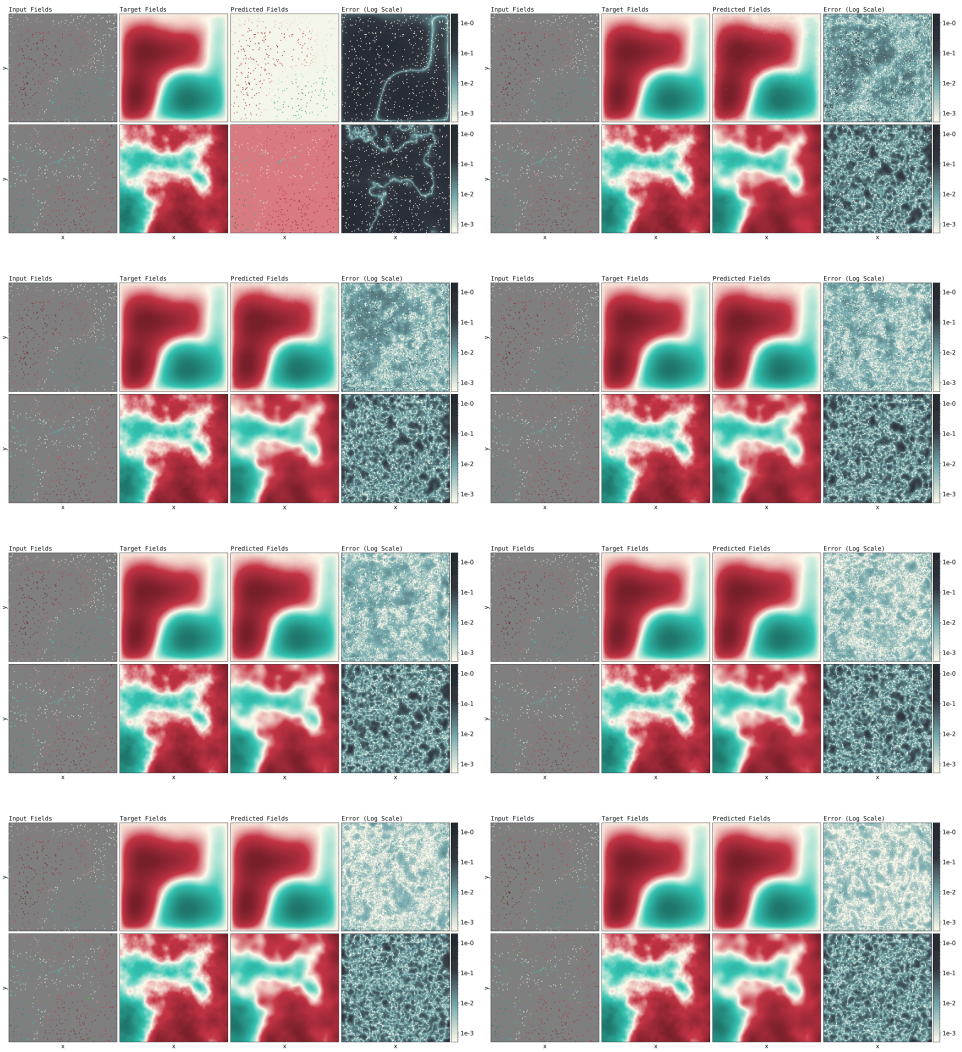


Figure 10: **Example of "One-Measurement Transition" (Poisson).** Each subplot shows an Ambient Flow reconstruction from partial observations (3% uniformly random points), varying the number of additional masked (already-observed) points used during training (increasing across the grid; top-left: 0, which corresponds to naive training). **Within Each Subplot: Top:** Solution field. **Bottom:** Coefficient Field. **Left to Right:** Input Fields, Target Fields, Reconstructed Fields, Error (Log Scale). Within each input panel, gray pixels represent unobserved locations and green pixels represent masked measurements withheld from the model.

F Robustness Across Patterns and Super-Resolution

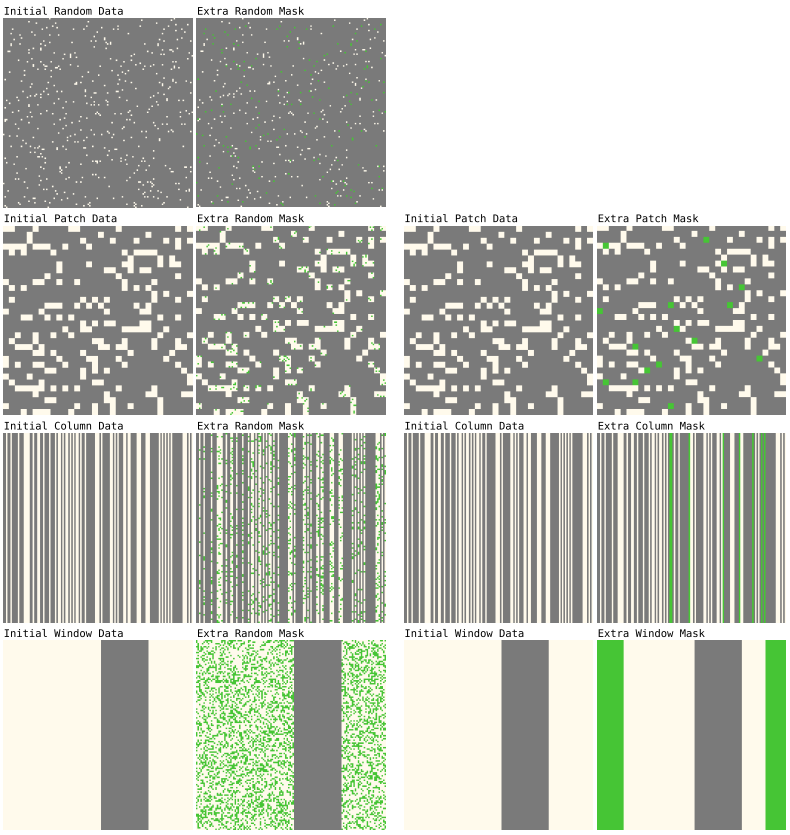


Figure 11: **Measurement patterns and extra masking operators.** Visualization of measurement patterns A and additional masking operators B used during Ambient Physics training. Additional masking B may match the measurement pattern (pattern-matched) or differ from it (mismatched); when mismatched, stronger additional masking is typically required to achieve similar training effects. **Top to bottom:** Random, Patch, Column, Window. Gray pixels represent unobserved locations and green pixels represent masked measurements withheld from the model.

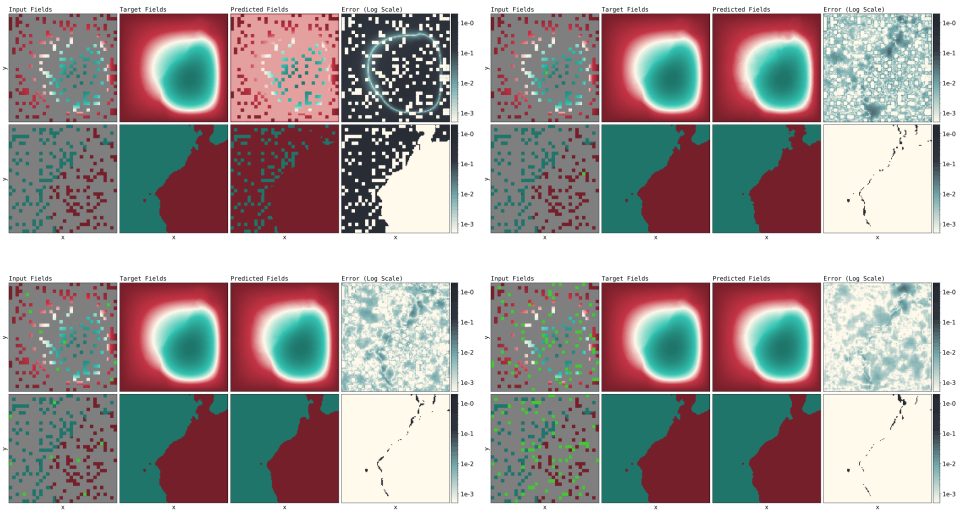


Figure 12: **Example of Patch Measurements (Darcy Flow).** Each subplot shows an Ambient Flow reconstruction from partial observations (20% uniformly random patches), varying the number of additional masked (already-observed) patches used during training (increasing across the grid; top-left: 0, which corresponds to naive training). **Within Each Subplot: Top:** Solution field. **Bottom:** Coefficient Field. **Left to Right:** Input Fields, Target Fields, Reconstructed Fields, Error (Log Scale). Within each input panel, gray pixels represent unobserved locations and green pixels represent masked measurements withheld from the model.

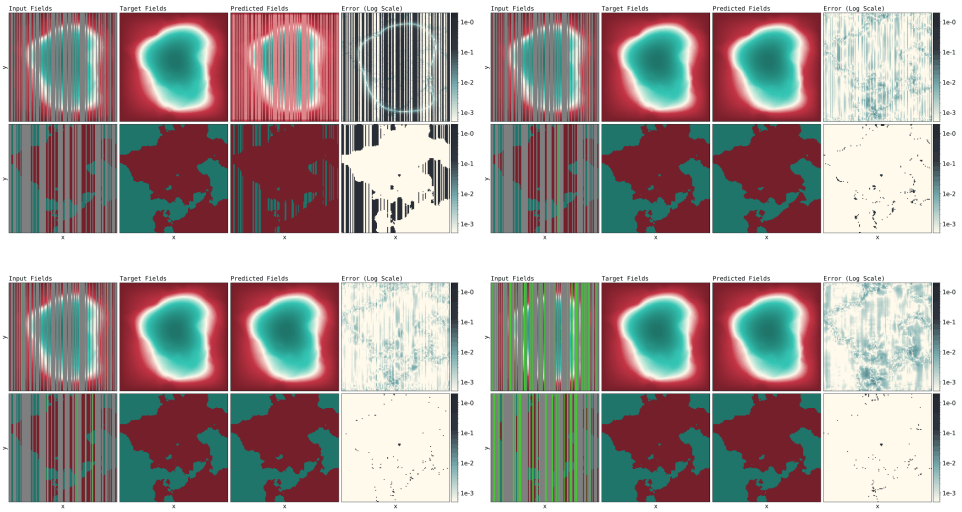


Figure 13: **Example of Column Measurements (Darcy Flow).** Each subplot shows an Ambient Flow reconstruction from partial observations (40% uniformly random columns), varying the number of additional masked (already-observed) columns used during training (increasing across the grid; top-left: 0, which corresponds to naive training). **Within Each Subplot: Top:** Solution field. **Bottom:** Coefficient Field. **Left to Right:** Input Fields, Target Fields, Reconstructed Fields, Error (Log Scale). Within each input panel, gray pixels represent unobserved locations and green pixels represent masked measurements withheld from the model.

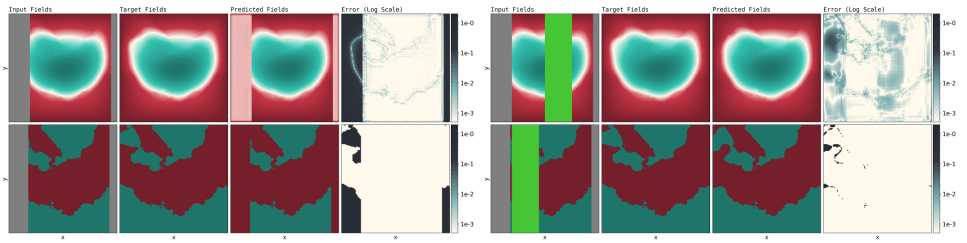


Figure 14: **Example of Window Measurements (Darcy Flow).** Each subplot shows an Ambient Flow reconstruction from partial observations (75% contiguous window). **Left:** No additional mask (naive training). **Right:** One additional mask.**Within Each Subplot: Top:** Solution field. **Bottom:** Coefficient Field. **Left to Right:** Input Fields, Target Fields, Reconstructed Fields, Error (Log Scale). Within each input panel, gray pixels represent unobserved locations and green pixels represent masked measurements withheld from the model.

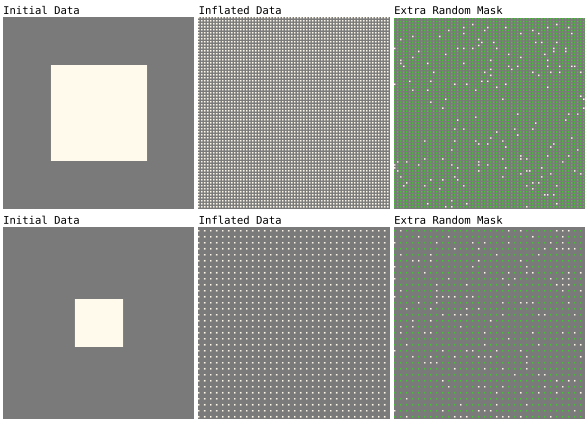


Figure 15: **Low-Resolution Observations as Partial Measurements for Super-Resolution.** A low-resolution measurement can be interpreted as a partially observed high-resolution field by inflating (upsampling) the observed values onto a higher-resolution (possibly shifted) grid. However, the inflated observations lie on a regular checkerboard-like lattice; if supervised only on these fixed locations, a model can learn to be accurate on the lattice while predicting unreliably elsewhere. We therefore apply an additional random mask over the inflated grid. Crucially, this mask must be sufficiently strong to hide the lattice structure, so the supervised locations no longer reveal a fixed checkerboard pattern and the model is encouraged to produce coherent high-resolution predictions across the domain. Ambient Physics can then learn a generative prior over high-resolution fields directly from low-resolution observations and perform super-resolution (see Table 6).

Table 6: **Super-resolution ablation on additional random masking.** Relative L_2 error (%) for coefficient/solution reconstruction after inflating low-resolution observations and applying an extra random mask. Intermediate masking (about 98%–99% total unobserved) performs best, consistent with the need to obscure the checkerboard observation lattice.

Initial Resolution	Extra Masked	Total Unobserved	Darcy Flow		Helmholtz		Navier-Stokes		Poisson	
			Coefficients	Solutions	Coefficients	Solutions	Coefficients	Solutions	Coefficients	Solutions
64×64	22.00%	97.00%	3.57%	2.20%	9.55%	19.34%	8.23%	2.28%	10.63%	37.20%
	23.00%	98.00%	**1.55%**	1.57%	**8.67%**	6.04%	**7.40%**	**1.71%**	**9.77%**	3.65%
	24.00%	99.00%	2.26%	**1.13%**	10.02%	**1.60%**	9.32%	2.48%	11.39%	**1.47%**
	24.50%	99.50%	3.43%	1.36%	14.29%	3.06%	14.82%	6.28%	14.85%	1.59%
32×32	3.25%	97.00%	1.72%	5.31%	24.74%	16.22%	14.41%	5.40%	27.33%	23.96%
	4.25%	98.00%	3.05%	3.17%	14.25%	6.00%	**8.77%**	1.50%	14.12%	5.83%
	5.25%	99.00%	**2.39%**	**1.82%**	**10.82%**	2.34%	9.05%	**1.46%**	**11.27%**	2.29%
	5.75%	99.50%	3.48%	1.91%	13.73%	**1.92%**	14.38%	2.51%	13.89%	**2.05%**

G Additional Studies

Observation Rates. Our main experiments use only 3% observed measurements in order to match the setting used by prior state-of-the-art diffusion-based PDE reconstruction methods and to evaluate Ambient Physics in an extremely sparse regime. This setting is substantially more difficult than settings with larger observation rates, such as 20% observed measurements. To study how performance varies with the amount of available information, we train and evaluate Ambient Flow across multiple uniformly random observation rates on all four PDE datasets.

Table 7 reports reconstruction performance for observation rates ranging from 3% to 25%. As expected, reconstruction error decreases as the number of observed measurements increases. At 20% observed measurements, Ambient Flow achieves low reconstruction error across all four PDEs: coefficient errors range from 1.23% to 3.14%, and solution errors range from 0.19% to 0.27%. This confirms that Ambient Physics learns effectively when 20% of the measurements are available.

Importantly, Ambient Flow remains effective even in the much harder 3% observation setting used in the main experiments. Although performance improves with additional measurements, the method does not require dense observations to learn meaningful structure: even with only 3% uniformly random measurements, Ambient Flow reconstructs both coefficients and solutions accurately across Darcy flow, Helmholtz, Navier–Stokes, and Poisson. These results suggest that Ambient Physics degrades gracefully as observations become sparser and benefits predictably from additional measurements.

Table 7: **Observation Rate Ablation.** Relative L_2 error (%) between ground truth and reconstructed coefficients ($\hat{a}$) and solutions ($\hat{u}$) from partial observations $(A_a a, A_u u)$ with uniformly random measurements and varying observation rates. Ambient Flow improves as the number of observed measurements increases, while remaining effective even in the sparse 3% setting used in the main experiments.

	Observed Measurements	Darcy Flow		Helmholtz		Navier-Stokes		Poisson	
		Coefficient	Solution	Coefficient	Solution	Coefficient	Solution	Coefficient	Solution
Ambient Flow	25%	1.16%	0.25%	2.77%	0.18%	1.79%	0.23%	2.70%	0.19%
	20%	1.23%	0.27%	3.14%	0.19%	2.10%	0.24%	3.10%	0.20%
	15%	1.28%	0.31%	3.68%	0.23%	2.52%	0.27%	3.61%	0.23%
	10%	1.34%	0.38%	4.59%	0.27%	3.16%	0.34%	4.61%	0.28%
	5%	1.65%	0.54%	6.39%	0.39%	4.64%	0.52%	6.47%	0.38%
	3%	2.08%	0.76%	7.73%	0.47%	6.41%	0.98%	7.86%	0.48%

Test-Time Observation Shifts. In the main experiments, Ambient Flow is trained and evaluated using the same observation process: 3% uniformly random measurements. In practice, however, the observation pattern or observation rate available at inference may differ from the one used during training. We therefore study the robustness of Ambient Physics to test-time observation shift by training Ambient Flow once on 3% uniformly random observations and then evaluating it under different observation quantities and measurement patterns.

Specifically, at test time we vary the percentage of observed measurements and evaluate four observation patterns: uniformly random point measurements, patch measurements, column measurements, and contiguous window measurements. This evaluates both quantitative shifts, where the number of observations changes, and structural shifts, where the geometry of the measurement operator differs from the one seen during training.

Figure 16 summarizes the results. Ambient Flow is highly robust to changes in the number of uniformly random point observations: performance remains essentially unchanged across a wide range of observation rates. The method is also robust to moderate structural shifts. Patch and column observations remain stable when the observed fraction is sufficiently large, with low error above roughly 20% observed measurements. Performance degrades primarily in the most ill-posed regimes, especially for contiguous window observations, where a large spatial region is entirely unobserved and the model receives little local information from the missing region.

These results suggest that Ambient Physics is not overly sensitive to small or moderate shifts in the observation process. At the same time, they highlight an expected limitation: when the test-time observation pattern differs substantially from the training distribution and removes large contiguous regions, reconstruction becomes more ambiguous and error increases. In real-world applications, this

suggests that the training mask distribution should be chosen to reflect the expected sensor geometry or measurement process whenever possible.

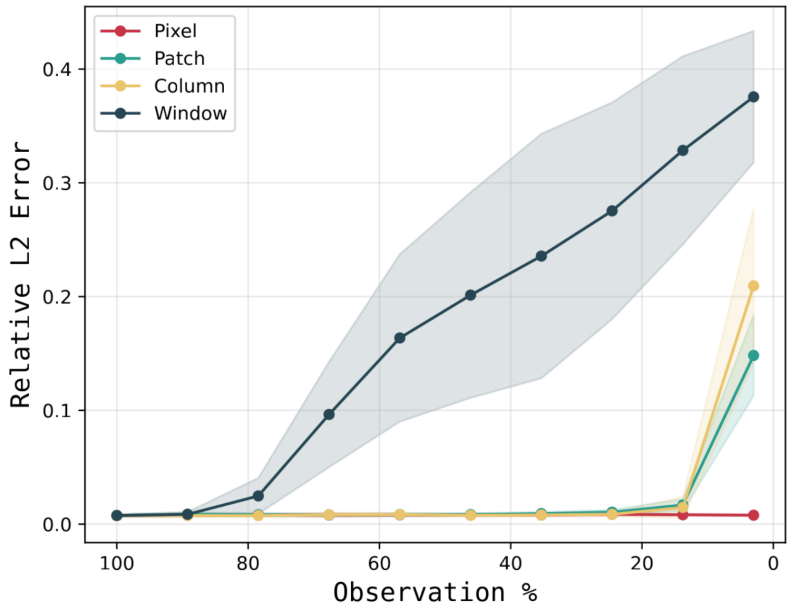


Figure 16: **Test-time observation shift.** Ambient Flow is trained on 3% uniformly random observations and evaluated under different test-time observation rates and patterns. Performance remains stable for uniformly random point observations across a wide range of observation rates, remains robust for patch and column observations above roughly 20% observed measurements, and degrades mainly in highly ill-posed structured-window settings where large contiguous regions are unobserved.

Sample Sizes. Because Ambient Physics trains only from partial observations, a natural question is whether it requires substantially more training samples than methods trained on complete observations. We therefore study sample efficiency by varying both the number of training samples and the observation rate per sample.

For each observation rate, we train Ambient Flow with different numbers of training examples and report reconstruction performance as a function of the total number of observed points available during training. Figure 17 shows that performance improves rapidly as the number of training samples increases, but begins to saturate beyond roughly 1024–2048 training samples. This trend is consistent across observation rates. Increasing the observation rate per sample also improves performance, indicating that reconstruction quality depends strongly on the amount of information contained in each partially observed example.

These results suggest that Ambient Physics does not require unusually large datasets to learn from partial observations. Instead, performance is governed by two complementary factors: the number of partially observed examples and the number of measurements available within each example. Once the model has seen enough examples to learn the underlying coefficient-solution distribution, additional samples provide diminishing returns, while denser observations per sample continue to improve reconstruction accuracy.

ERA5 Weather Fields. Although our main experiments focus on learning the joint distribution of PDE coefficients and solutions, the Ambient Physics framework is more general: it can be applied to other physical reconstruction problems in which only partial observations are available. Weather modeling provides a natural example. In practice, atmospheric fields are often observed through sparse and heterogeneous measurement systems, such as weather stations, satellites, or sensor networks, and one may wish to reconstruct complete physical fields from these incomplete measurements.

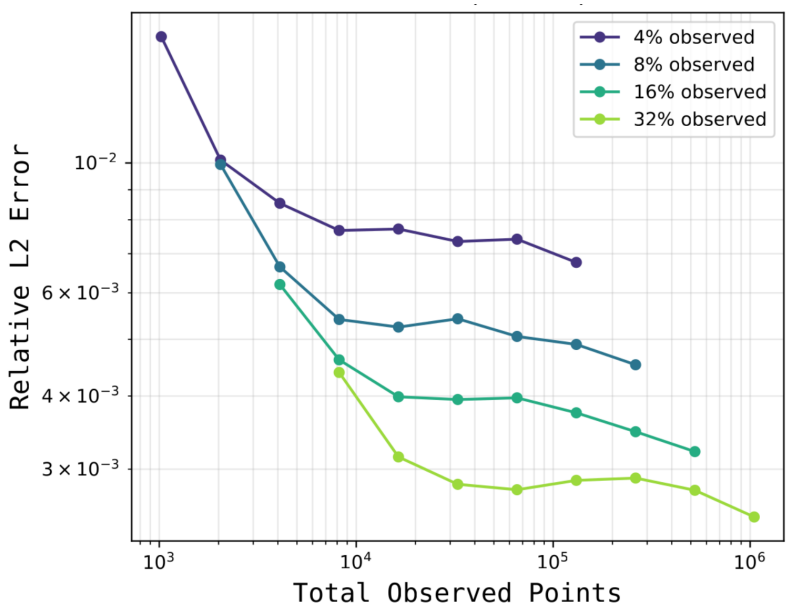


Figure 17: **Sample efficiency of Ambient Flow.** Reconstruction error as a function of the total number of observed points available during training, for different observation rates per sample. Performance improves rapidly with additional training samples and begins to saturate beyond roughly 1024–2048 samples. Higher observation rates per sample yield lower error, suggesting that Ambient Physics does not require unusually large datasets and benefits most from informative partially observed examples.

To demonstrate that Ambient Physics generalizes beyond the coefficient-solution reconstruction setting studied in the main paper, we evaluate Ambient Physics on ERA5 reanalysis fields (Hersbach et al., 2020), using data from the Copernicus Climate Data Store.[3] We emphasize that this experiment is intended as a proof-of-concept extension rather than the main focus of the paper. Instead of modeling a pair (a, u) of PDE coefficients and solutions, we treat the weather variables as a multi-channel physical field and train Ambient Physics to reconstruct the complete field from sparse observations.

We compare against Incomplete Data, Complete Dynamics (IDCD), which is the closest prior work studying diffusion-based learning of physical dynamics from incomplete observations. As discussed in Appendix I, IDCD assumes a time-dependent dynamics-imputation setting in which multiple partially observed frames can provide complementary coverage over time. In contrast, Ambient Physics does not require complementary temporal coverage: it learns to reconstruct complete fields from partial observations by additionally masking observed measurements and supervising only where measurements are available.

Figure 18 shows a qualitative comparison between IDCD and Ambient Physics on ERA5 reconstruction. Ambient Physics produces substantially sharper and more physically plausible reconstructions, while IDCD exhibits noticeable artifacts and loses fine-scale structure. Quantitatively, Ambient Physics achieves approximately an order of magnitude lower reconstruction error than IDCD in this setting. This improvement is consistent with the qualitative visualization: Ambient Physics better preserves coherent spatial structures and avoids the noisy, speckled artifacts visible in the IDCD prediction.

We also evaluate Ambient Physics on a multi-channel ERA5 reconstruction task involving nine weather fields. Figure 19 shows the sparse input fields, target fields, and predicted fields for all variables. Across variables, Ambient Physics reconstructs the large-scale spatial structure of the target fields and preserves many fine-scale features despite observing only a sparse subset of measurements.

[3]ERA5 is an existing reanalysis dataset provided by the Copernicus Climate Change Service/ECMWF; we use it in accordance with the applicable Copernicus/ECMWF terms.

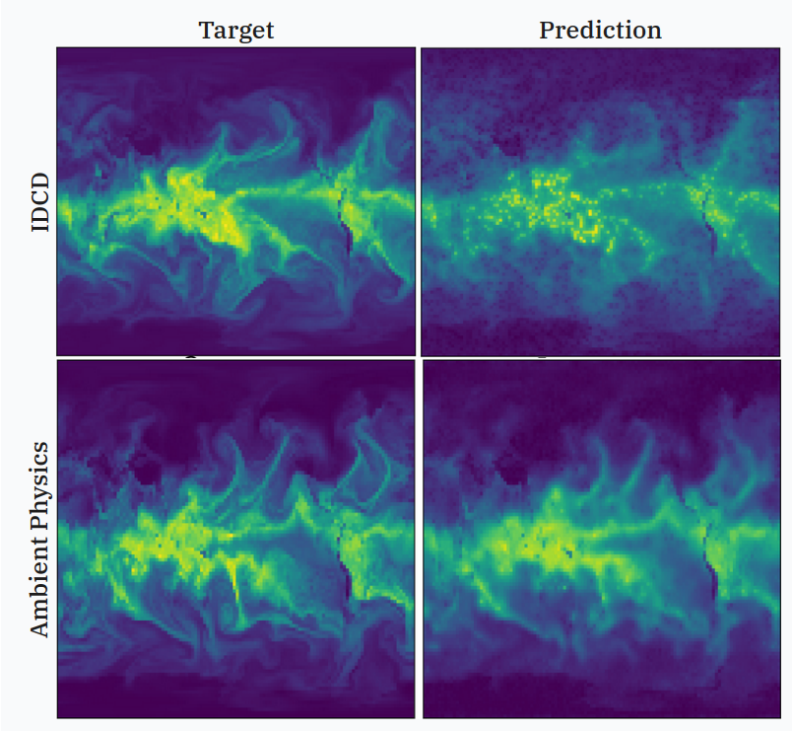


Figure 18: **ERA5 qualitative comparison with Incomplete Data, Complete Dynamics (IDCD).** We compare target fields and predicted fields from IDCD and Ambient Physics. Ambient Physics produces a sharper and more coherent reconstruction, while IDCD exhibits visible artifacts and degraded fine-scale structure. Quantitatively, Ambient Physics achieves approximately an order of magnitude lower reconstruction error in this setting.

This suggests that the same masking principle used for joint coefficient-solution modeling can extend naturally to multi-channel physical field reconstruction.

These results show that Ambient Physics is not limited to the PDE coefficient-solution setting. The same framework can be used for broader physical reconstruction problems where complete observations are unavailable, including weather and climate applications with sparse or heterogeneous measurement systems. A systematic study of large-scale weather reconstruction, data assimilation, and forecasting is beyond the scope of this work, but these ERA5 results indicate that Ambient Physics provides a promising foundation for such settings.

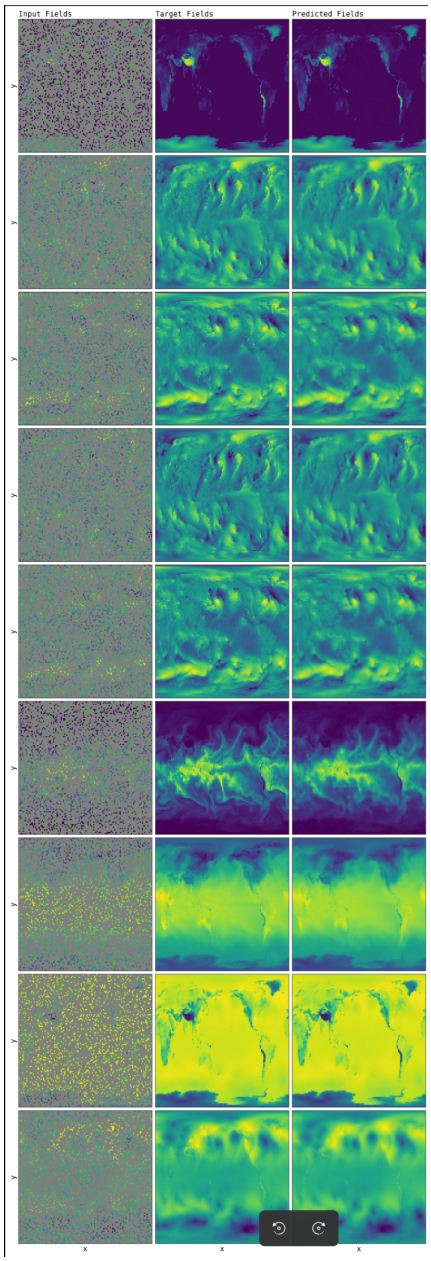


Figure 19: **ERA5 multi-channel reconstruction.** Ambient Physics is applied to reconstruct nine weather fields from sparse observations. Each row corresponds to one physical variable, and columns show the sparse input field, target field, and predicted field. Ambient Physics accurately recovers the large-scale structure of the target fields and preserves fine-scale spatial patterns across variables.

H Additional Controls and Diagnostics

This section reports additional controls on the experimental protocol and diagnostics of the trained models. Unless stated otherwise, the setup follows section 4: partial observations $(A_a a, A_u u)$ with 3% uniformly random measurements, and relative L_2 error (%) between ground truth and reconstructed coefficients and solutions.

Training Seeds and Checkpoint Selection. We repeat the Ambient Flow experiments of Table 1 with two additional training seeds and report the mean and standard deviation across three training runs (Table 8). The results show low variation across training seeds.

In the main experiments, the best checkpoint is selected using the relative L_2 error on complete validation fields. Because complete fields are unavailable in the settings that motivate this work, we repeat these experiments with fixed hyperparameters while selecting the best checkpoint exclusively through the normalized relative error on artificially masked points. Complete fields are never used for training or checkpoint selection, only to report final performance. Checkpoint selection using artificially masked points mostly matches checkpoint selection using complete fields; the largest difference is the Darcy flow solution error (Table 8).

Table 8: **Training Seeds and Checkpoint Selection.** Relative L_2 error (%) of Ambient Flow, reported as mean $\pm$ standard deviation over three training seeds. The best checkpoint is selected using complete validation fields (as in the main experiments) or using only artificially masked points, in which case complete fields are used only to report final performance.

Checkpoint Selection	Darcy Flow		Helmholtz		Navier-Stokes		Poisson	
	Coefficient	Solution	Coefficient	Solution	Coefficient	Solution	Coefficient	Solution
Complete validation fields	$2.08 \pm 0.01\%$	$0.76 \pm 0.03\%$	$7.73 \pm 0.02\%$	$0.46 \pm 0.05\%$	$6.41 \pm 0.03\%$	$1.06 \pm 0.04\%$	$7.86 \pm 0.02\%$	$0.48 \pm 0.03\%$
Artificially masked points only	$2.10 \pm 0.02\%$	$0.98 \pm 0.06\%$	$7.70 \pm 0.02\%$	$0.47 \pm 0.04\%$	$6.40 \pm 0.01\%$	$1.04 \pm 0.06\%$	$7.71 \pm 0.02\%$	$0.48 \pm 0.04\%$

Conditional and Unconditional Ambient Training. To separate the contribution of ambient training from that of conditional amortization, we train an unconditional Ambient Flow model with the Ambient Physics objective but without measurement conditioning, and evaluate the same model using two reconstruction methods: (1) one-step reconstruction, where the noise-free measurements are inserted at the final flow time, followed by a single model evaluation, which corresponds to the Ambient Diffusion setting (Daras et al., 2023); and (2) DPS-style guidance, where the same unconditional Ambient Flow model is sampled using measurement guidance for 100 NFE. Conditional training provides an additional accuracy improvement over both one-step reconstruction and DPS-style guidance (Table 9).

Table 9: **Conditional and Unconditional Ambient Training.** Relative L_2 error (%) of conditional Ambient Flow and of an unconditional Ambient Flow model trained with the same Ambient Physics objective but without measurement conditioning, reconstructed with one-step reconstruction or DPS-style guidance.

Method	NFE	Darcy Flow		Helmholtz		Navier-Stokes		Poisson	
		Coefficient	Solution	Coefficient	Solution	Coefficient	Solution	Coefficient	Solution
Conditional Ambient Flow	1	**2.08%**	**0.76%**	**7.73%**	**0.46%**	**6.41%**	**1.06%**	**7.86%**	**0.48%**
Unconditional Ambient Flow + one-step reconstruction	1	3.63%	1.85%	12.26%	2.69%	10.36%	3.82%	11.29%	2.54%
Unconditional Ambient Flow + DPS-style guidance	100	4.55%	3.38%	12.48%	2.42%	10.35%	3.52%	11.79%	2.32%

Using All Available Measurements at Inference. By default, we sample the additional masks at inference from the same distribution used during training to avoid a train–test distribution shift, so a single reconstruction withholds roughly one-third of the available measurements from the model. To test the setting in which every available measurement is supplied to the model at inference, we evaluate with no additional mask, i.e., $B = I$ (Table 10). Using all available measurements improves the coefficient error on all four datasets and the Darcy flow solution error, but worsens the solution error for Helmholtz, Navier–Stokes, and Poisson. Alternatively, averaging reconstructions over resampled additional masks improves relative L_2 error and can recover part of the information withheld by a single additional mask (Appendix D).

Nested Masks. In the default design, the noisy flow state and the conditioning tensor are masked with the same further-corrupted measurement operator, so every clean observation available under that mask is supplied to the model. In a preliminary experiment, we instead use strictly nested masks: supervision on all available measurements (3% of pixels), a noisy flow state that retains 2.5% of pixels, and a conditioning tensor that retains 2% of pixels, so that the noisy state contains measurements that are absent from the conditioning tensor. This substantially worsens relative L_2 error; on Darcy flow coefficients, the error increases from 2.08% to 14.71% (Table 10).

Table 10: **Inference and State Masking.** Relative L_2 error (%) of Ambient Flow with the default design (additional masks sampled from the training distribution at inference; noisy state and conditioning share the same mask), with no additional mask at inference ($B = I$), and with nested masks during training (noisy state retains 2.5% and conditioning retains 2% of pixels; preliminary experiment).

Variant	Darcy Flow		Helmholtz		Navier-Stokes		Poisson	
	Coefficient	Solution	Coefficient	Solution	Coefficient	Solution	Coefficient	Solution
Default	2.08%	0.76%	7.73%	**0.46%**	6.41%	**1.06%**	7.86%	**0.48%**
No additional mask at inference ($B = I$)	**1.80%**	**0.69%**	**6.98%**	0.96%	**5.73%**	1.21%	**7.00%**	0.94%
Nested state and conditioning masks	14.71%	0.83%	7.87%	0.54%	6.37%	1.17%	8.05%	0.54%

Sensitivity to the Initial Noise. For each of 100 test conditions on Darcy flow, we fix the physical example, observation mask, additional mask, and conditioning tensor, and vary only the initial noise x_0 across 64 Gaussian draws. We compute the pixelwise sample variance of the reconstructed coefficients and report its scale-normalized square root, the relative RMS spread (Table 11). Ambient Flow has nonzero fixed-condition sample variance, and its relative spread is $15.9\times$ greater at unobserved locations than at locations whose measurements are supplied as conditioning. Thus, the trained model is not empirically invariant to x_0, and its coefficient variation is concentrated primarily where measurements are unavailable. Unconditional Ambient Flow with one-step reconstruction and Ambient UNet are deterministic at fixed conditioning.

Table 11: **Sensitivity to the Initial Noise.** Relative RMS spread (%) of the reconstructed coefficients across 64 draws of the initial noise x_0 at fixed conditioning, averaged over 100 test conditions, at locations whose measurements are supplied as conditioning (observed) and elsewhere (unobserved).

Method	Locations	Coefficient Relative RMS Spread
Ambient Flow	Observed	0.083%
	Unobserved	1.32%
Unconditional Ambient Flow + one-step reconstruction	Observed	0.00%
	Unobserved	0.00%
Ambient UNet	Observed	0.00%
	Unobserved	0.00%

Darcy Flow Coefficient Predictions. The Darcy flow coefficient is binary, taking the values 3 and 12. We bin the raw predicted coefficients at unobserved locations into four bins: near 3 ($|a-3| \leq 0.5$), intermediate ($3.5 < a < 11.5$), near 12 ($|a - 12| \leq 0.5$), and outside (all other values) (Table 12). For Ambient Flow, 93.97% of the raw coefficient predictions lie within 0.5 of one of the two physical values, and only 6.03% are intermediate-valued. Thus, Ambient Flow does not exhibit widespread collapse toward values between the two physical modes. Ambient UNet also preserves the binary coefficient structure, but unlike Ambient Flow it is deterministic at fixed conditioning.

Table 12: **Darcy Flow Coefficient Predictions.** Distribution (%) of the raw predicted Darcy flow coefficients at unobserved locations.

Method	Near 3	Intermediate	Near 12	Outside
Ambient Flow	46.79%	6.03%	47.18%	0.00%
Unconditional Ambient Flow + one-step reconstruction	42.28%	14.75%	42.97%	0.00%
Ambient UNet	46.74%	5.86%	47.40%	0.00%

I Related Methods

Masked Autoencoders. A natural question is how Ambient Physics differs from masked self-supervised learning methods such as Masked Autoencoders (MAE). The key distinction is whether

Table 13: **Comparison with Masked Autoencoders.** Relative L_2 error (%) between ground truth and reconstructed coefficients ($\hat{a}$) and solutions ($\hat{u}$). MAE performs much better when trained with complete observations than when trained only with partial observations, confirming that complete observations provide a substantial training advantage. Ambient Flow is trained only from partial observations and achieves substantially lower error than MAE trained under the same partial-observation constraint.

Method	Complete Obs. Training	Darcy Flow		Helmholtz		Navier-Stokes		Poisson	
		Coefficient	Solution	Coefficient	Solution	Coefficient	Solution	Coefficient	Solution
MAE	Yes	3.69%	3.01%	14.09%	4.39%	13.92%	7.60%	13.68%	4.18%
MAE	No	36.47%	53.39%	98.55%	98.81%	98.49%	98.39%	98.49%	98.70%
Ambient Flow	No	**2.08%**	**0.76%**	**7.73%**	**0.47%**	**6.41%**	**0.98%**	**7.86%**	**0.48%**

complete observations are available during training. MAE assumes access to complete training examples. It masks a subset of each complete example, passes the masked example as input, and computes the reconstruction loss using the complete unmasked example. Thus, although part of the input is hidden from the model, the supervision still comes from complete observations.

Ambient Physics addresses a different setting. We assume that complete observations are unavailable during training. The model receives only partial observations $(A_a a, A_u u)$, additionally masks a subset of the already-observed measurements, and computes loss only where measurements are actually available. Therefore, Ambient Physics cannot supervise on arbitrary masked locations in the full field; it can only supervise on measurements that were observed in the first place. This distinction is central to the problem setting: standard masked self-supervised objectives rely on complete data for supervision, while Ambient Physics is designed to learn directly from incomplete data.

To empirically test whether a standard MAE objective can solve our setting, we train MAE on the same PDE reconstruction datasets. We use the ViT architecture and hyperparameters from the original MAE setup, and evaluate two variants. First, we train MAE with complete observations, corresponding to the standard setting where MAE has access to full ground-truth fields during training. Second, we train MAE using only partial observations, matching the information available to Ambient Physics. For comparison, we also report Ambient Flow, which is trained only from partial observations.

Table 13 shows the results. As expected, MAE performs substantially better when trained with complete observations than when trained with partial observations, confirming that complete observations provide a significant training advantage. When MAE is restricted to partial observations, performance degrades sharply across all PDEs and fields. In contrast, Ambient Flow learns effectively from partial observations and achieves much lower error. Moreover, Ambient Flow is competitive with or better than MAE even when MAE is trained with complete observations, despite Ambient Flow never observing complete fields during training.

We note that a different architecture or a heavily tuned masked-autoencoding pipeline may improve MAE performance. However, the purpose of this comparison is not to claim that MAE architectures are inherently weak; rather, it is to clarify the difference between the learning settings. Ambient Physics is designed for the regime where complete observations are unavailable, while standard MAE relies on complete observations to define its reconstruction target.

Incomplete Data, Complete Dynamics. The closest prior work to our setting is Incomplete Data, Complete Dynamics (IDCD), which studies diffusion-based learning of physical dynamics from incomplete observations. While IDCD is related in spirit, there are important differences in both the algorithmic setup and empirical evaluation.

From an algorithm-design perspective, IDCD can be viewed as an instantiation of ambient diffusion for time-dependent physical dynamics. However, IDCD allows the case in which all observed measurements are passed to the model as input, so that no observed measurements are withheld for supervision. In the terminology of this paper, this corresponds to the *naive training* regime. As shown in this paper, naive training can minimize the measurement loss by copying the observed inputs while receiving no training signal at unobserved locations. This leads to accurate predictions only where measurements are provided and arbitrary predictions elsewhere. Therefore, for learning

complete fields from partial observations, it is essential to withhold at least some already-observed measurements from the model input and supervise on them.

This distinction is central to Ambient Physics. Our method explicitly separates the observed measurements used as conditioning from the observed measurements used for supervision. The additional masking operator creates a nontrivial prediction problem on measurements that are known during training but hidden from the model input. This prevents the model from solving the training objective by simple copying and forces it to learn structure in the underlying physical fields. In contrast, when no measurements are withheld, the objective degenerates to naive training. Moreover, IDCD does not study how the ratio between conditioning measurements and withheld measurements affects reconstruction performance. Our one-measurement transition and additional masking ablations show that this ratio is important: even withholding a single observed measurement can sharply improve reconstruction, while the naive zero-withheld-measurement case fails.

There are also important empirical differences. IDCD focuses on time-dependent dynamics imputation, where multiple partially observed frames provide complementary information over time. In their setting, each state may observe a different subset of the field, so a sequence of partially observed states can collectively contain substantially more information than a single partial observation. This is a different and generally easier setting than the one considered in our main experiments, where the goal is to reconstruct coefficient-solution pairs or physical fields directly from partial observations without relying on complementary temporal coverage. Despite this additional temporal information, IDCD exhibits poor reconstruction quality in our ERA5 comparison. As shown in Figure 18, IDCD produces noisy reconstructions with visible artifacts and degraded spatial structure, whereas Ambient Physics produces sharper and more coherent fields.

Finally, we note that IDCD is difficult to use as a reproducible baseline in our main PDE reconstruction experiments. The paper does not provide code or a complete description of the data source and preprocessing pipeline, and its reported baselines are not designed for reconstructing physical coefficient or solution fields from partial observations. Instead, the baselines are primarily drawn from tabular imputation or generic image-sampling settings. For this reason, we focus our main comparisons on DiffusionPDE and FunDPS, which are prior state-of-the-art methods for the same PDE reconstruction task, datasets, and evaluation protocol. We include the ERA5 comparison with IDCD to show that Ambient Physics also improves over the closest incomplete-observation dynamics method in a real-world physical-field reconstruction setting.

Overall, IDCD and Ambient Physics address related but distinct problems. IDCD studies time-dependent dynamics imputation from incomplete sequences, while Ambient Physics studies learning complete physical-field distributions directly from partial observations. The key algorithmic difference is the additional masking of already-observed measurements: Ambient Physics avoids the naive copying solution by withholding observed measurements from the input and supervising on them, which we find to be essential for accurate reconstruction.

J Future Directions and Limitations

Ambient Physics provides, to our knowledge, the first framework for learning the joint distribution of coefficient-solution pairs in physical-science problems directly from partial observations, without requiring complete training examples. We demonstrate strong performance across several PDE-based datasets and show that the same principle extends to real-world ERA5 weather fields. Nevertheless, an important next step is to deploy Ambient Physics in a real-world scientific setting where complete ground-truth fields are genuinely unavailable, rather than synthetically hidden for evaluation. Such a setting would provide a stronger test of whether the method can support scientific reconstruction problems in practice.

A second limitation is that the current experiments focus on a fixed set of observation patterns and corruption levels. In real applications, sensor layouts, satellite coverage, missingness patterns, and observation densities may vary across time and deployment conditions. A natural future direction is therefore to train Ambient Physics over a broader distribution of corruption levels and measurement patterns, including mixtures of sparse observations, structured masks, and clean samples when they are available. Extending the framework to three-dimensional physical fields, irregular meshes, and

unstructured geometries is also important for applications in climate, geophysics, fluid dynamics, and engineering.

Finally, this work focuses primarily on incomplete measurements, where the main corruption is missing data. Real scientific observations are often noisy, biased, calibrated with uncertainty, or collected from heterogeneous instruments with different error profiles. Extending Ambient Physics beyond missingness to handle noisy or uncertain measurements is a promising direction. In such settings, the measurement operator could incorporate both the observation pattern and a noise or uncertainty model, allowing the framework to learn from partial, noisy, and potentially biased scientific data.