# Neural Field Thermal Tomography: A Differentiable Physics Framework for Non-Destructive Evaluation

**Tao Zhong**[1], **Yixun Hu**[1], **Dongzhe Zheng**[1], **Aditya Sood**[1], **Christine Allen-Blanchette**[1]
[1]Princeton University
{tzhong, ca15}@princeton.edu

## Abstract

Inverse problems for stiff parabolic partial differential equations (PDEs), such as the inverse heat conduction problem (IHCP), are severely ill-posed: the forward map rapidly damps high-frequency interior structure before it reaches the boundary. Soft-constrained physics-informed neural networks (PINNs), which embed the PDE as a residual penalty, decouple the diffusivity from the surface data in their standard two-network inverse formulation, and the evaluated variants fit boundary measurements while leaving the interior field essentially untouched. We propose Neural Field Thermal Tomography (NeFTY), a hard-constrained neural field framework for label-free three-dimensional inverse heat conduction. NeFTY represents the unknown diffusivity as a continuous coordinate-based neural network, and at every optimization step passes the candidate field through a differentiable implicit-Euler heat solver with harmonic-mean interface flux, so that the simulated temperature satisfies the discretized heat equation up to the solver tolerance. Adjoint gradients propagate the surface reconstruction error back to the network weights at solver-level memory cost, making test-time inversion tractable on a single GPU. Across synthetic 3D benchmarks, NeFTY outperforms soft-constrained PINN variants, a voxel-grid baseline, and classical L-BFGS voxel inversion through the same solver on label-free volumetric recovery, and it transfers to real thermography data, surpassing classical signal-processing baselines in both defect segmentation and depth estimation. Additional details at `cab-lab-princeton.github.io/nefty`.

## 1 Introduction

Heat diffusion blurs interior structure before it reaches the surface: high spatial frequencies inside a specimen are damped on their way to the boundary, so the map from internal material properties to surface measurements discards much of the information needed to invert it [18, 22]. Inverse problems for such stiff parabolic partial differential equations (PDEs) are therefore severely ill-posed. Soft physics-informed losses can fit boundary data without recovering the underlying field, because their gradients become poorly conditioned and dominated by the surface signal in this stiff regime [51, 65]. Classical adjoint-based inverse heat conduction [2, 30] and differentiable physics [13, 26, 27, 43] build the discretized solver into the optimization loop, so the governing equations hold as a hard constraint at every step. We combine this hard constraint with a neural-field prior for the three-dimensional inverse heat conduction problem (IHCP).

Non-destructive testing (NDT) by pulsed thermography is the working example throughout (Figure 1). A pulsed laser deposits energy on the specimen surface, and a high-speed camera records the subsequent surface temperature decay. Subsurface delaminations, voids, and inclusions perturb the diffusive flux and leave faint thermal contrast on the boundary [32, 48, 54]. Reconstructing the volumetric thermal diffusivity field $\alpha(x, y, z)$ from these surface signals is the inverse heat conduction problem. Unlike hyperbolic ultrasonic or radar imaging [5, 64], the parabolic forward map damps

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

internal features rapidly with depth, so small surface perturbations can correspond to large interior variations and the inverse map becomes severely ill-posed [34, 50].

Existing thermography and learning approaches each address part of this volumetric inversion. Classical thermography pipelines enhance defect contrast in transformed signal domains: Thermographic Signal Reconstruction (TSR) [58, 59] and Pulsed Phase Thermography (PPT) [38] reduce to 1D pixel-wise inversions that ignore lateral diffusion and bias estimates for shallow or low-aspect-ratio defects [49], and the Virtual Wave Concept (VWC) [4] images defect geometry. None recovers a volumetric diffusivity field. Soft-constrained physics-informed neural networks (PINNs) target the volumetric problem in principle, yet in their standard two-network form they meet a gradient pathology in stiff transient diffusion: the PDE residual is high and rugged, so the temperature surrogate fits the boundary measurements while the diffusivity network saturates at a near-trivial field [65]. Supervised volumetric learning sidesteps these issues but requires 3D ground-truth labels that are unavailable in real NDT.

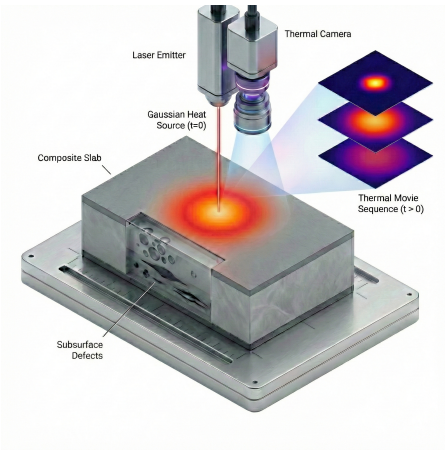



Figure 1: **IHCP setup.** A high-speed camera measures the transient surface temperature following pulsed laser heating. NeFTY reconstructs the 3D subsurface diffusivity field from these measurements.

To this end, we introduce Neural Field Thermal Tomography (NeFTY), a hard-constrained neural field framework for label-free 3D inverse heat conduction. NeFTY represents the unknown diffusivity as a continuous coordinate-based neural network [40, 62] with positional encoding, frequency annealing, and a bounded output. At every optimization step, the candidate field is passed through a differentiable implicit-Euler heat solver with harmonic-mean interface flux, so the temperature satisfies the discretized heat equation up to the iterative solver tolerance. Adjoint gradients propagate the surface reconstruction error back to the network weights at solver-level memory cost, making test-time inversion tractable on a single GPU. On synthetic homogeneous and layered settings, NeFTY reaches volumetric IoU of $0.45$ and $0.37$, while soft-constrained PINN variants and a voxel-grid baseline remain at or near zero, and it outperforms classical thermography baselines on real PVC thermal data in both 2D defect segmentation and 2.5D depth recovery.

We summarize our contributions as follows:

- We propose a hard-constrained neural field formulation for label-free 3D inverse heat conduction from surface-only thermal sequences, combining a continuous coordinate-based prior with a differentiable implicit heat solver and adjoint gradients. To our knowledge, NeFTY is the first label-free ML-based method to recover a full 3D diffusivity field from surface-only thermal measurements.
- We give a theoretical and empirical account of soft-constrained PINNs on stiff parabolic inverse problems: in the standard two-network formulation with data, residual, and initial-condition terms, surface data reaches the diffusivity only through the residual, and across four evaluated PINN variants low surface error does not imply correct volumetric recovery.
- We carry out an evaluation that pairs synthetic 3D label-free reconstruction with real PVC thermal data, soft-constrained PINN and classical L-BFGS baselines, robustness studies on geometry, noise, and contrast, frequency- and edge-domain diagnostics, component ablations, and a wall-clock, memory, and scaling analysis of the adjoint solver.

## 2 Related Work

**Thermographic Non-Destructive Testing** traditionally extracts subsurface defect information through signal-processing transforms applied to surface temperature decays. Thermographic Signal Reconstruction (TSR) [58, 59] fits log-time polynomials to enhance defect contrast, Pulsed Phase Thermography (PPT) [10, 38] maps the decays into a frequency phase domain that suppresses noise and emissivity variations, and the Virtual Wave Concept (VWC) [4, 57] deconvolves the diffusive forward map into a pseudo-wave for depth estimation. Quantitative-thermography refinements add

calibrated depth correlations and effective-diffusivity estimates [1, 37, 49], while TSR/PPT-style depth inversions remain pixel-wise and amplify high-frequency measurement noise. In parallel, supervised convolutional pipelines based on U-Net [17, 42, 53, 60] learn defect masks directly from thermal sequences, and several recent surveys [32, 35, 48, 54] catalogue this growing detection-focused literature. The closest prior systems reconstruct 3D defect geometry [61], learn defect depth [70], or recover internal temperature fields [34]. None recovers a material-property volume. Pixel-wise transforms such as TSR and PPT ignore lateral heat flux, and supervised volumetric models require 3D diffusivity labels that are unavailable in real NDT. NeFTY recovers a full 3D diffusivity field directly from the surface sequence, without volumetric labels and without 1D pixel-wise approximations.

**Physics-Informed Neural Networks** embed governing equations as soft penalties alongside data terms in scientific machine learning. Standard PINNs [6, 34, 51] jointly minimize boundary fit and a PDE residual, and a series of variants attack the resulting optimization pathologies: separable architectures in SPINN [9] scale up collocation throughput, temporal-causality weighting in Causal-PINN [66] stabilizes long-horizon transients, and dual-cone gradient descent in DCGD [28] resolves residual-versus-data gradient conflicts. Diagnostic studies document conditioning, spectral-bias, and loss-balancing difficulties of PDE-residual penalties in stiff regimes [23, 33, 65], even with adaptive loss balancing such as GradNorm [8]. Section 3.3 analyzes the gradient decoupling of the two-network inverse formulation evaluated here. A complementary line replaces the penalty with a differentiable solver: discretize-then-optimize [43], differentiable PDE engines [3, 26, 27], and end-to-end differentiable physics for control and robotics [13, 14, 63, 72] all enforce the discretized governing equations at every optimization step. Classical IHCP inversion follows the same principle: adjoint-based optimization [30], iterative regularization [2], and textbook treatments [44] enforce the heat equation through a numerical solver, within the broader theory of PDE-constrained optimization [25]. NeFTY belongs to this discretize-then-optimize adjoint family and contributes a neural-field prior over the diffusivity, which enables label-free 3D recovery from surface-only measurements.

**Coordinate-based Neural Fields** parameterize a continuous signal as a multilayer perceptron (MLP) over spatial coordinates, typically combined with positional encoding or sinusoidal activations. SIREN [62] and NeRF [40] established neural fields as a high-frequency representation prior for visual signals, and deformable extensions such as Nerfies [45] broadened the class of recoverable scenes. Scientific imaging has begun to adopt the same representation for tomographic and field-reconstruction problems, including X-ray computed tomography [71, 73], dynamic medical imaging via ProxNF [36], and fluid scalar fields with FluidNeRF [31]. These works contribute a continuous, mesh-free function space, and the parameterization itself does not enforce a governing PDE. NeFTY uses the neural field as the diffusivity prior and couples it with a differentiable heat solver. The two are complementary, and with the solver and adjoint held fixed, the neural-field prior carries the largest share of NeFTY's gain over voxel-field inversion among the evaluated configurations (Section 5.2, Appendix G.8).

## 3 Forward Model, Ill-Posedness, and Soft-Constraint Pathology

### 3.1 Forward model and effective-diffusivity parameterization

The conservation form of transient heat conduction in an isotropic medium is posed on a bounded Lipschitz domain $\Omega \subset \mathbb{R}^3$, with boundary $\partial\Omega$ and observation window $\Gamma_{\text{obs}} \subseteq \partial\Omega$, as follows:

$$\rho(\mathbf{x})C_p(\mathbf{x})\,\partial_t T \;=\; \nabla\cdot\big(k(\mathbf{x})\,\nabla T\big) + Q(\mathbf{x},t), \qquad \mathbf{x}\in\Omega,\; t\in(0,t_{\text{end}}], \tag{1}$$

with mass density $\rho$, specific heat capacity $C_p$, thermal conductivity $k$, and source $Q$ ($Q \equiv 0$ post flash). Dividing (1) by $\rho C_p$ and applying the product rule gives the equivalent diffusivity form

$$\partial_t T \;=\; \nabla\cdot\big(\alpha(\mathbf{x})\nabla T\big) \;-\; k(\mathbf{x})\,\nabla T\cdot\nabla\big(\tfrac{1}{\rho C_p}\big), \qquad \alpha(\mathbf{x}) := \frac{k(\mathbf{x})}{\rho(\mathbf{x})C_p(\mathbf{x})}, \tag{2}$$

for smooth, strictly positive $\rho C_p$ (Appendix A.1). When $\rho C_p$ is uniform within each phase, as in a host material with embedded defects, the second term vanishes inside each phase and acts only on the interfaces $\Sigma$ between them, where a finite-volume grid cannot represent it directly [47] (Appendix A.2). NeFTY therefore parameterizes a single effective field $\alpha(\mathbf{x})$ obeying $\partial_t T = \nabla\cdot(\alpha\nabla T)$ within each phase and restores flux continuity across interfaces through the harmonic mean at cell faces.

**Proposition 1** (Harmonic mean as discrete interface continuity for the effective-$\alpha$ flux)**.** *For a one-dimensional cell pair with constant cell-centered diffusivities $\alpha_i, \alpha_{i+1}$, cell spacing $\Delta x$, cell-centered temperatures $T_i, T_{i+1}$, and effective $\alpha$-flux $j_\alpha := -\alpha \partial_x T$, the unique flux satisfying continuity across the cell face $x_{i+1/2}$ is $j_\alpha = -\bar{\alpha}_{i+1/2}(T_{i+1} - T_i)/\Delta x$ with $\bar{\alpha}_{i+1/2} = 2\alpha_i\alpha_{i+1}/(\alpha_i + \alpha_{i+1})$.*

The proof, multi-dimensional consequence, and high-contrast behavior are in Appendix A. Synthetic and real-PVC initial and boundary condition settings are specified in Appendix A.5.

### 3.2 The inverse problem and its ill-posedness

The IHCP recovers $\alpha$ from noisy surface measurements $\hat{T}(\mathbf{x}, t_i)$, $\mathbf{x} \in \Gamma_{\text{obs}}$, $t_i \in \{t_1, \ldots, t_M\} \subset (0, t_{\text{end}}]$. Let $\mathcal{K}(\alpha) := T_\alpha|_{\Gamma_{\text{obs}} \times (0, t_{\text{end}})}$ denote the parameter-to-observation operator (with $T_\alpha$ the solution of (2) under fixed initial and lateral boundary conditions) acting on the admissible set $\mathcal{A} \subset H^1(\Omega) \cap L^\infty(\Omega)$. The reconstruction objective is

$$\begin{aligned} \alpha^* &= \arg\min_{\alpha \in \mathcal{A}} \mathcal{J}(\alpha), \\ \mathcal{J}(\alpha) &= \sum_{i=1}^{M} \left\| \mathcal{K}(\alpha)(\cdot, t_i) - \hat{T}(\cdot, t_i) \right\|^2_{\Gamma_{\text{obs}}} + \lambda \mathcal{R}(\alpha), \end{aligned} \tag{3}$$

with $\|\cdot\|_{\Gamma_{\text{obs}}}$ the $L^2$ norm over $\Gamma_{\text{obs}}$, $\mathcal{R}$ a regularization functional, and $\lambda > 0$. The IHCP is widely recognized as one of the most severely ill-posed problems in mathematical physics [16, 22, 39]. We quantify it by bounding the singular values of the linearized observation map.

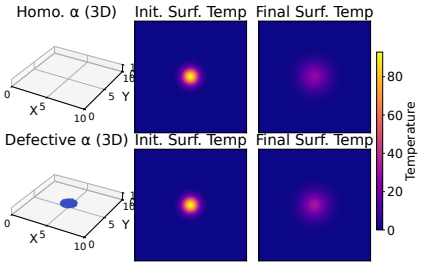



Figure 2: **Empirical ill-posedness.** Distinct interiors yield nearly indistinguishable surface profiles, illustrating Prop. 2.

**Proposition 2** (Compactness and algebraic singular-value decay)**.** *Let $\alpha_0 > 0$ be a constant background diffusivity with $\nabla T_{\alpha_0} \in L^\infty(\Omega \times (0, t_{\text{end}}))$. The Fréchet derivative $d\mathcal{K}_{\alpha_0} : H^1(\Omega) \to L^2(\Gamma_{\text{obs}} \times (0, t_{\text{end}}))$ is bounded and compact, with singular values $\sigma_n \le C\,(1 + \lambda_n)^{-1/2}$, where $\{\lambda_j\}_{j \ge 0}$ are the Laplacian eigenvalues on $\Omega$ under the imposed boundary conditions. On the slab, Weyl's law gives $\sigma_n \lesssim n^{-1/3}$.*

**Corollary 1** (Hadamard ill-posedness)**.** *The pseudo-inverse $d\mathcal{K}^\dagger_{\alpha_0}$ is unbounded, with $\sigma_n^{-1} \ge C^{-1}(1 + \lambda_n)^{1/2}$ amplifying noise along the $n$-th left singular vector.*

In practical terms, finer interior structure leaves progressively weaker surface signatures, and inverting the map amplifies measurement noise without bound. Proof in Appendix B. Figure 2 visualizes the smoothing. NeFTY counters this with a neural-field prior plus frequency annealing (Section 4) and a total-variation penalty $\mathcal{R}(\alpha)$.

### 3.3 Soft-constraint pathology in stiff parabolic inversion

PINNs are the dominant and broadly successful paradigm for physics-informed learning. Our analysis concerns their standard two-network inverse formulation, which jointly optimizes a temperature surrogate $T_\phi(\mathbf{x}, t)$ with parameters $\phi$ and a diffusivity surrogate $\alpha_\theta(\mathbf{x})$ with separate parameters $\theta$, against

$$\mathcal{L}_{\text{PINN}} = \underbrace{\|T_\phi - \hat{T}\|^2_{\Gamma_{\text{obs}}}}_{\mathcal{L}_{\text{data}}} + \lambda_{\text{PDE}} \underbrace{\|\partial_t T_\phi - \nabla \cdot (\alpha_\theta \nabla T_\phi)\|^2_\Omega}_{\mathcal{L}_{\text{PDE}}} + \lambda_{\text{IC}} \underbrace{\|T_\phi(\cdot, 0) - T_0\|^2_\Omega}_{\mathcal{L}_{\text{IC}}}, \tag{4}$$

with weights $\lambda_{\text{PDE}}, \lambda_{\text{IC}} > 0$ and $T_0$ the prescribed initial temperature field. Because $T_\phi$ is parameterized by $\phi$ alone, $\nabla_\theta \mathcal{L}_{\text{data}} = \nabla_\theta \mathcal{L}_{\text{IC}} = 0$, so the diffusivity update reduces to $\nabla_\theta \mathcal{L}_{\text{PINN}} = \lambda_{\text{PDE}} \nabla_\theta \mathcal{L}_{\text{PDE}}$, with magnitude controlled pointwise by the Laplacian $\Delta T_\phi$ of the under-fitted temperature surrogate. Appendix C gives the full derivation and shows that adaptive loss balancing (GradNorm), separable temperature parameterizations (SPINN), causal weighting (Causal-PINN), and gradient-conflict resolution (DCGD) all preserve this decoupling.

In this formulation, surface observations reach $\theta$ only indirectly through the residual penalty, independently of the optimizer. The derivation establishes this decoupling. Combined with the unbounded amplification in Corollary 1, it gives a plausible mechanism for the PINN training pathology in which

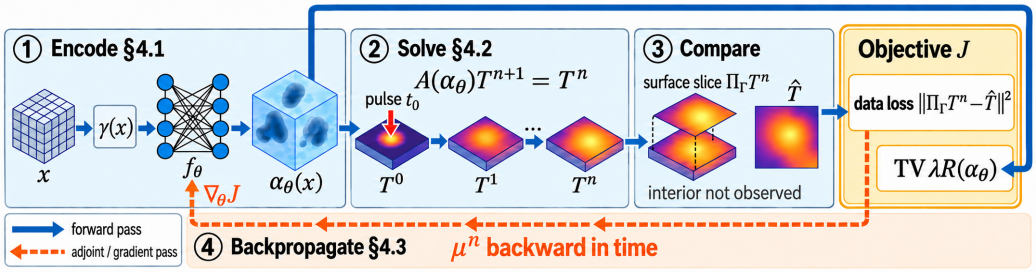



Figure 3: **NeFTY pipeline**, read left to right. (1) Encode: a coordinate MLP $f_\theta$ maps each grid node $\mathbf{x}$ through a frequency-annealed positional encoding $\gamma(\cdot)$ into a bounded diffusivity $\alpha_\theta(\mathbf{x})$ (§4.1). (2) Solve: a differentiable implicit-Euler solver with harmonic-mean coefficients evolves the temperature $\mathbf{T}^n$ on a Cartesian grid (§4.2). (3) Compare: only the surface slice of $\mathbf{T}^n$ is compared to the measurements. (4) Backpropagate: the adjoint state $\boldsymbol{\mu}^n$ is propagated backward through the same sparse system to assemble the gradient on $\theta$ without retaining Jacobi iterates (§4.3).

$\mathcal{L}_{\text{PDE}}$ shrinks while $\alpha_\theta$ remains near a trivial constant [23, 33, 65], the collapse we observe in the four evaluated variants (Section 5.2). Architectures whose temperature network depends on $\theta$ lie outside this derivation. Appendix C.4 studies such a directly coupled variant. In NeFTY, surface data propagates to $\alpha_\theta$ through the adjoint of the discretized heat operator (Section 4).

# 4 Method

NeFTY recovers the diffusivity field by directly minimizing the reconstruction objective $\mathcal{J}$ of equation (3), simulating each candidate field with a differentiable heat solver whose temperature satisfies the discretized heat equation up to the chosen spatial and temporal discretization and the iterative linear-solver tolerance. Each step (Figure 3) encodes the diffusivity as a coordinate-based neural field $\alpha_\theta$ (§4.1), solves the implicit-Euler heat equation with harmonic-mean interface flux (§4.2), and propagates the surface-fidelity gradient back to $\theta$ with a discrete adjoint (§4.3). Because the discrete temperature is a function of $\alpha_\theta$, the gradient decoupling of Section 3.3 does not arise.

## 4.1 Neural diffusivity field

NeFTY parameterizes the unknown effective diffusivity $\alpha_\theta : \Omega \to \mathbb{R}_{>0}$ as a coordinate multilayer perceptron (MLP) $f_\theta : \mathbb{R}^{d_\gamma} \to \mathbb{R}$ whose input is lifted through a Fourier feature map $\gamma : \mathbb{R}^3 \to \mathbb{R}^{d_\gamma}$:

$$\gamma(\mathbf{x}) \;=\; \big(\sin(2^0\pi\mathbf{x}), \cos(2^0\pi\mathbf{x}), \ldots, \sin(2^{N-1}\pi\mathbf{x}), \cos(2^{N-1}\pi\mathbf{x})\big), \tag{5}$$

of bandwidth $N$. The encoding combats the spectral bias of plain MLPs and lets the network represent the sharp interfaces characteristic of subsurface defects [40, 62]. To stabilize the high-frequency components against the algebraic noise amplification of Corollary 1, we adopt the cosine frequency-annealing schedule of Park et al. [45], gradually unmasking encoding bands during training. The network output is hard-bracketed to the physically admissible range

$$\alpha_\theta(\mathbf{x}) \;=\; \alpha_{\min} + (\alpha_{\max} - \alpha_{\min})\,\sigma\big(f_\theta(\gamma(\mathbf{x}))\big), \tag{6}$$

with $\sigma(\cdot)$ the logistic sigmoid. The bounds $\alpha_{\min}, \alpha_{\max} > 0$ are chosen to span the bulk-to-defect contrast of the application without inducing ill-conditioning in the discrete system below. We instantiate the regularizer in (3) as the isotropic total variation $\mathcal{R}(\alpha_\theta) = \int_\Omega \|\nabla\alpha_\theta\|$ [55], consistent with the piecewise-constant defect structure of the effective-diffusivity model in Appendix A. The full architecture, the cosine annealing schedule, and the discrete TV form are given in Appendix D.1.

## 4.2 Hard-constrained differentiable physics

We discretize the bulk equation $\partial_t T = \nabla \cdot (\alpha \nabla T)$ on a uniform Cartesian grid covering $\Omega$ with spacings $\Delta x, \Delta y, \Delta z$ along the three axes and $N_t$ time steps of size $\Delta t$ aligned with the camera frame rate. Let $\mathbf{T}^n \in \mathbb{R}^{N_g}$ denote the discrete temperature at step $n$ on the $N_g$ grid nodes, and let $\bar{\alpha}_{i+1/2}$ denote the harmonic-mean face-centered diffusivity from Proposition 1. The second-order

finite-volume diffusion operator $\mathbf{L}(\alpha_\theta) \in \mathbb{R}^{N_g \times N_g}$ acts at an interior node by

$$[\mathbf{L}(\alpha_\theta)\mathbf{T}]_{i,j,k} \;=\; \frac{1}{\Delta x^2}\big[\bar{\alpha}_{i+1/2}(T_{i+1,j,k} - T_{i,j,k}) - \bar{\alpha}_{i-1/2}(T_{i,j,k} - T_{i-1,j,k})\big] + (y\text{-, } z\text{-terms}), \tag{7}$$

with lateral periodic and through-thickness adiabatic conditions (Appendix A.5) closing the stencil at the boundary. The harmonic-mean coefficient is the unique discrete realization of effective-flux continuity at interior faces for the single-$\alpha$ model (Proposition 1), whereas an arithmetic mean would spuriously allow heat to leak across the interface [47]. All discrete temperatures here are interpreted as deviations from the ambient temperature (Appendix A.5), so the implicit-Euler step is affine-free.

To decouple the time step from the diffusivity contrast, we integrate in time with the unconditionally stable implicit Euler scheme [12], advancing $\mathbf{T}^n$ to $\mathbf{T}^{n+1}$ through the sparse linear system

$$\mathbf{A}(\alpha_\theta)\,\mathbf{T}^{n+1} \;=\; \mathbf{T}^n, \qquad \mathbf{A}(\alpha_\theta) \;:=\; \mathbf{I} - \Delta t\,\mathbf{L}(\alpha_\theta). \tag{8}$$

The matrix $\mathbf{A}(\alpha_\theta)$ is sparse, symmetric positive-definite for our boundary conditions, and depends on $\theta$ only through $\alpha_\theta$. We solve (8) on the GPU through a fixed number $K$ of unrolled Jacobi iterations, expressed as stencil convolutions within standard autodiff frameworks [46]. The full discretization, choice of $K$, and Jacobi update rule are given in Appendix D.2. Solver validation and the distinction between numerical and inversion error are reported in Appendix I.

### 4.3 Adjoint optimization

Composing the implicit Euler recurrence (8) from $\mathbf{T}^0$ defines $\mathbf{T}^n = \mathbf{T}^n(\alpha_\theta)$ implicitly, hence the discrete analogue $\mathcal{K}_{\text{disc}}$ of the parameter-to-observation map. Let $n_i \in \{1, \dots, N_t\}$ be chosen so that $t_{n_i} = t_i$ for each measurement time $t_i$, and let $\mathbf{\Pi}_\Gamma : \mathbb{R}^{N_g} \to \mathbb{R}^{N_\Gamma}$ denote the row-selector matrix that extracts the $N_\Gamma$ surface samples on $\Gamma_{\text{obs}}$. NeFTY minimizes the discrete reconstruction objective

$$\min_\theta \;\; \mathcal{J}_{\text{disc}}(\theta) \;=\; \sum_{i=1}^{M} \big\|\mathbf{\Pi}_\Gamma \mathbf{T}^{n_i}(\alpha_\theta) - \hat{\mathbf{T}}_i\big\|^2 \;+\; \lambda\,\mathcal{R}(\alpha_\theta), \tag{9}$$

the discretization of (3) with $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^{N_\Gamma}$ and $\mathcal{R}$ the discrete total-variation regularizer of Appendix D.1. Because $\mathbf{T}^n$ is defined implicitly by the state equation (8), gradients $d\mathcal{J}_{\text{disc}}/d\theta$ obtained by backpropagation through the $K$ unrolled Jacobi iterates of every step cost $\mathcal{O}(KN_gN_t)$ memory, which is prohibitive for the $N_g \approx 10^5$, $N_t \approx 10^2$ regime of pulsed thermography. We use the discrete adjoint method instead [7, 43].

Treat (8) as the constraint $\mathbf{F}^n(\mathbf{T}^n, \mathbf{T}^{n-1}, \alpha_\theta) := \mathbf{A}(\alpha_\theta)\mathbf{T}^n - \mathbf{T}^{n-1} = \mathbf{0}$ for $n = 1, \dots, N_t$, and introduce adjoint variables $\boldsymbol{\mu}^n$ as Lagrange multipliers for each step. Setting the total derivative of the augmented Lagrangian with respect to $\mathbf{T}^n$ to zero yields the backward-in-time recurrence

$$\mathbf{A}(\alpha_\theta)^\top \boldsymbol{\mu}^n \;=\; \boldsymbol{\mu}^{n+1} + \big(\partial \ell^n / \partial \mathbf{T}^n\big)^\top, \qquad \boldsymbol{\mu}^{N_t+1} \;=\; \mathbf{0}, \tag{10}$$

where $\ell^n$ collects the data-fidelity terms in (9) that depend on $\mathbf{T}^n$ (nonzero only at measurement frames $n \in \{n_1, \dots, n_M\}$). The adjoint at step $n$ uses the transpose of the same sparse $\mathbf{A}$ that drives the forward solve, so a single backward sweep with the same Jacobi inner loop yields all $\boldsymbol{\mu}^n$ at the cost of one forward pass. The parameter gradient is then assembled by

$$\frac{d\mathcal{J}_{\text{disc}}}{d\theta} \;=\; \frac{\partial(\lambda\mathcal{R})}{\partial\theta} \;+\; \Delta t \sum_{n=1}^{N_t} (\boldsymbol{\mu}^n)^\top \frac{\partial(\mathbf{L}(\alpha_\theta)\mathbf{T}^n)}{\partial \alpha_\theta}\,\frac{\partial \alpha_\theta}{\partial \theta}, \tag{11}$$

where $\partial\alpha_\theta/\partial\theta$ is the MLP Jacobian. Because (10) requires only $\boldsymbol{\mu}^{n+1}$ and $\mathbf{T}^n$, the adjoint's solver working set is independent of $N_t$. Only the forward states $\mathbf{T}^n$ are stored. The full Lagrangian construction, the derivation of (10)–(11), and the consistency with the continuous adjoint of the bulk equation are given in Appendix D.3. The implementation hyperparameters (depth, width, encoding bandwidth $N$, grid resolution, Jacobi count $K$, and optimizer schedule) are listed in Appendix D.4. The data term depends directly on $\theta$ through the chain $\theta \to \alpha_\theta \to \mathbf{T}^n \to \mathcal{J}_{\text{disc}}$, the dependency absent from the two-network formulation of Section 3.3.

Table 1: **Label-free 3D thermal diffusivity reconstruction on synthetic $1{,}000$-sample benchmark** (mean $\pm$ 95% CI). $\uparrow$ higher is better, $\downarrow$ lower is better. Best label-free entry per metric in **bold**. PINN [51], Causal-PINN [66], and DCGD [28] $\alpha_\theta$ rows coincide because all three collapse to the same near-trivial constant. See Section 5.2 and Appendix C.

| | Homogeneous | | | | Layered | | | |
|---|---|---|---|---|---|---|---|---|
| Method | MSE ($10^{-4}$) $\downarrow$ | PSNR $\uparrow$ | SSIM $\uparrow$ | IoU $\uparrow$ | MSE ($10^{-4}$) $\downarrow$ | PSNR $\uparrow$ | SSIM $\uparrow$ | IoU $\uparrow$ |
| *Data-driven (zero-shot)* | | | | | | | | |
| U-Net (Sound-Only) [53] | $14.73 \pm 3.83$ | $14.83 \pm 0.71$ | $0.83 \pm 0.02$ | $0.00 \pm 0.00$ | $10.17 \pm 2.21$ | $15.42 \pm 0.98$ | $0.88 \pm 0.02$ | $0.00 \pm 0.00$ |
| *Label-free* | | | | | | | | |
| PINN [8, 51] | $208.4 \pm 26.2$ | $-0.24 \pm 0.06$ | $0.04 \pm 0.00$ | $0.015 \pm 0.003$ | $200.4 \pm 17.8$ | $1.42 \pm 0.27$ | $0.04 \pm 0.00$ | $0.018 \pm 0.004$ |
| SPINN [9] | $32.71 \pm 10.02$ | $9.12 \pm 0.93$ | $0.45 \pm 0.06$ | $0.001 \pm 0.001$ | $33.84 \pm 8.57$ | $9.71 \pm 0.77$ | $0.53 \pm 0.07$ | $0.000 \pm 0.000$ |
| Causal-PINN [66] | $208.4 \pm 26.2$ | $-0.24 \pm 0.06$ | $0.04 \pm 0.00$ | $0.015 \pm 0.003$ | $200.4 \pm 17.8$ | $1.42 \pm 0.27$ | $0.04 \pm 0.00$ | $0.018 \pm 0.004$ |
| DCGD [28] | $208.4 \pm 26.2$ | $-0.24 \pm 0.06$ | $0.04 \pm 0.00$ | $0.015 \pm 0.003$ | $200.4 \pm 17.8$ | $1.42 \pm 0.27$ | $0.04 \pm 0.00$ | $0.018 \pm 0.004$ |
| Grid Opt. | $12.61 \pm 4.20$ | $13.99 \pm 1.18$ | $0.56 \pm 0.04$ | $0.04 \pm 0.02$ | $15.01 \pm 2.97$ | $13.27 \pm 0.88$ | $0.57 \pm 0.04$ | $0.03 \pm 0.01$ |
| **NeFTY (Ours)** | $\mathbf{3.66 \pm 1.31}$ | $\mathbf{18.48 \pm 0.53}$ | $\mathbf{0.77 \pm 0.02}$ | $\mathbf{0.45 \pm 0.04}$ | $\mathbf{9.26 \pm 2.81}$ | $\mathbf{15.88 \pm 0.79}$ | $\mathbf{0.74 \pm 0.03}$ | $\mathbf{0.37 \pm 0.06}$ |

# 5 Experiments

## 5.1 Setup

**Datasets.** Synthetic data are generated with PhiFlow [26], a finite-volume engine independent of the implicit-Euler Jacobi solver of Section 4.2, so reconstruction is not subject to inverse crime. The real PVC datasets [68, 69] provide only 2D defect masks and discrete defect depths from CAD, so real-world evaluation uses projected 2D masks and 2.5D depth annotations rather than 3D volumetric labels. The synthetic benchmark contains $1{,}000$ samples on the unitless slab $\Omega = [0, 10]^2 \times [0, 1]$ at $64 \times 64 \times 16$ resolution, 100 frames at $\Delta t = 0.05$, and one to four subsurface defects per sample, split into a homogeneous and a layered configuration. Full data-generation procedure, dimensional analysis, and compute resources are in Appendix E.

**Baselines.** On synthetic data we compare to direct grid optimization (Grid Opt.; a voxel field through the same solver, adjoint, and TV regularizer, without the neural prior), classical L-BFGS voxel inversion with TV or $H^1$ regularization (Appendix G.8), PINN [51] with GradNorm [8], SPINN [9], Causal-PINN [66], DCGD [28], and a Sound-Only U-Net [53] that trains on only defect-free samples and probes data-driven generalization to out-of-distribution defects. On real PVC data we additionally compare to the classical thermography heuristics PPT [29, 38] and TSR [58, 59] and to a Swin-UNETR [24] trained on the synthetic benchmark and transferred zero-shot. Implementation details for every baseline are in Appendix F.

**Metrics.** Synthetic 3D reconstruction is evaluated by Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) [67], and volumetric Intersection over Union (IoU; defect mask $\alpha_\theta < 0.03$). Real PVC results are evaluated by 2D IoU and Dice for defect segmentation and by Absolute Relative error (Abs Rel), Root Mean Squared Error (RMSE), and the standard depth thresholds $\delta < 1.25^k$ ($k = 1, 2, 3$) [15]. Defect Edge F1 and the radial power spectrum of the recovered field, both targeting frequency-domain fidelity, are reported as diagnostics in Appendix G. Precise definitions for all metrics are catalogued in Appendix E.

## 5.2 Synthetic 3D reconstruction

Table 1 reports the full benchmark. Figures 4 and 5 show representative qualitative reconstructions in the homogeneous and layered settings (per-defect-count and per-layer breakdowns are in Appendix G). Grid Opt. inverts the same hard PDE constraint as NeFTY but lacks a coordinate-based regularizer: it lowers the diffusivity MSE relative to every PINN variant yet caps IoU at $0.04$/$0.03$ with ringing artifacts at defect interfaces. The four soft-constrained PINN variants of Section 3.3 fit the surface temperature observations through their $T_\phi$ network but fail volumetrically: their $\alpha_\theta$ networks saturate to near-trivial fields with IoU $\leq 0.018$ (SPINN's separable surrogate lowers the homogeneous diffusivity MSE from $208.4$ to $32.71$ ($\times 10^{-4}$) without resolving defects), in line with the structural decoupling $\nabla_\theta \mathcal{L}_{\text{data}} = 0$ proven in Appendix C (the bit-identical PINN, Causal-PINN, and DCGD $\alpha_\theta$ rows in Table 1 reflect this same collapse despite differing $T_\phi$ surrogates). NeFTY balances the physical consistency of the hard solver with the prior of a continuous neural field, reaching IoU $0.45$/$0.37$ and PSNR $18.48$/$15.88$ dB.

**Data-fit paradox.** The evaluated PINN variants fit the surface observations while recovering little volumetric structure: vanilla PINN attains a surface PSNR of $\sim 63$ dB on $T_\phi$ while its volumetric

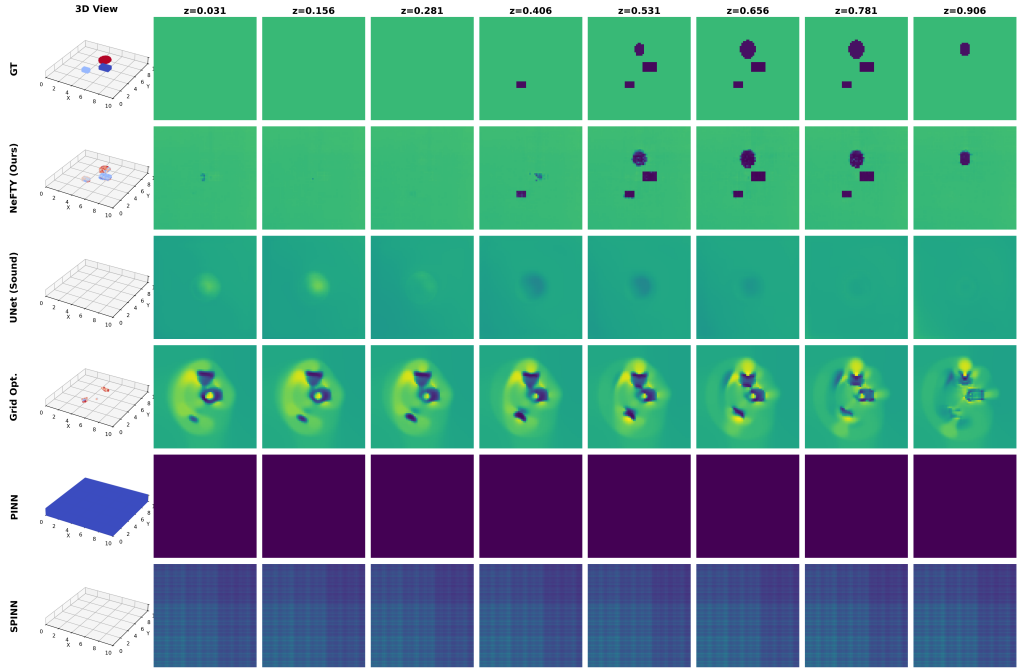



Figure 4: **Homogeneous-setting reconstructions:** depth-wise $z$-slices show NeFTY isolating defect outlines at the correct depth while Grid Opt. yields a noisy field and PINN variants converge to near-uniform diffusivity.

IoU is $0.01$, SPINN reaches $\sim 61$ dB at IoU $0.000$, and NeFTY $\sim 82$ dB at IoU $0.45$. Under the soft-constraint formulation of Eq. (4), surface observations influence the diffusivity only indirectly, through the residual term (Appendix C.2), and in this stiff parabolic regime the low surface error of the evaluated variants does not translate into correct volumetric structure. Frequency-domain diagnostics in Appendix G (Edge F1 $0.470 \pm 0.137$ for NeFTY versus $0.000$ for every PINN variant; full radial power spectrum) confirm that NeFTY is the only label-free method in Table 1 whose recovered field carries detectable defect-scale gradients.

**Classical PDE-constrained inversion.** We also optimize a bounded voxel diffusivity field with L-BFGS under TV or $H^1$ regularization through the same solver and adjoint, with weights tuned on a disjoint calibration set (Appendix G.8). On 64 paired benchmark specimens these arms reach IoU $0.129/0.112$ (TV) and $0.125/0.127$ ($H^1$) on homogeneous/layered data, against $0.45/0.37$ for NeFTY and $0.04/0.03$ for Grid Opt. (Table 1, full benchmark). On the same specimens (ratio of means), L-BFGS improves on Grid Opt. by about $3.2\times$ and the neural-field prior adds a further $3.3\times$, so among the evaluated configurations the representation carries the largest share of NeFTY's gain.

**Robustness.** On new irregular-geometry specimens, NeFTY reaches 3D IoU $0.316$ on irregular blobs and $0.222$ on curved strata interfaces, where Grid Opt. reaches $0.000$ (Appendix G.9). Under Gaussian observation noise at $0.5$, $1$, $2$, and $5\%$ of peak signal, NeFTY retains $96$, $82$, $64$, and $37\%$ of its noise-free IoU (Appendix G.10). Across defect-to-bulk contrast from 1:5 to 1:50, IoU shows no detectable change (Appendix G.11).

### 5.3 Real-world PVC validation

The same NeFTY solver is applied to the PVC datasets after the three physical adaptations of Appendix A.5 (near-uniform front-face flash initial condition, centimeter and second-scale Fourier-number rescaling, convective Robin back-face boundary), with the projection from $\hat{\alpha}_\theta$ to 2D masks and 2.5D depth documented in Appendix H. As reported in Table 2, NeFTY is the best label-free method on every metric, with $0.431$ IoU and $0.597$ Dice for 2D segmentation and $0.465$ Abs Rel and $0.385$ $\delta < 1.25$ accuracy for 2.5D depth. The classical heuristics PPT and TSR are competitive on segmentation but degrade sharply on depth, since their inversion formulas assume a semi-infinite one-dimensional geometry that does not hold on the finite PVC slab (Appendix F.4). The supervised

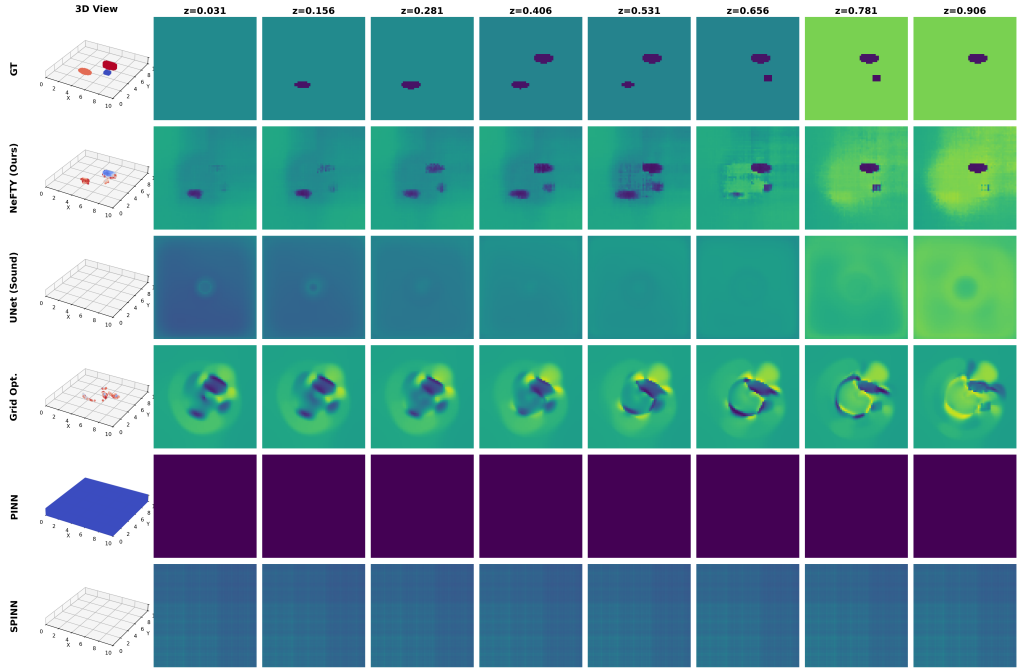



Figure 5: **Layered-setting reconstructions:** NeFTY recovers defects against the stratified bulk, whereas Grid Opt. exhibits ringing artifacts and PINN variants again collapse to featureless fields.

Table 2: **Real PVC evaluation:** NeFTY is the best label-free method on every metric for both 2D segmentation [69] and 2.5D depth [68].

| | 2D SEGM. | | 2.5D DEPTH | |
|---|---|---|---|---|
| METHOD | IoU ↑ | DICE ↑ | ABS REL ↓ | $\delta < 1.25$ ↑ |
| SWIN-UNETR [24] | 0.085 | 0.151 | 3.131 | 0.052 |
| PPT [38] | 0.419 | 0.563 | 3.416 | 0.005 |
| TSR [59] | 0.346 | 0.464 | 13.366 | 0.002 |
| SPINN [9] | 0.050 | 0.086 | 6.737 | 0.002 |
| GRID OPT. | 0.395 | 0.557 | 0.941 | 0.104 |
| **NEFTY (OURS)** | **0.431** | **0.597** | **0.465** | **0.385** |

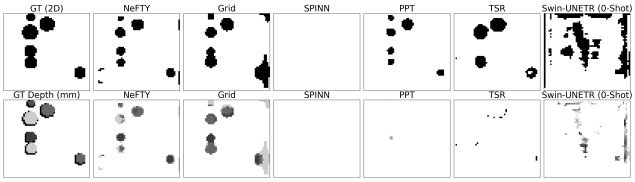



Figure 6: **Real PVC reconstruction:** NeFTY recovers both the defect layout (Row 1) and the depth ordering (Row 2) from a single surface IR sequence, while PPT/TSR yield coarse depth and SPINN collapses to a near-uniform field.

Swin-UNETR collapses under the synthetic-to-real domain shift, in contrast to NeFTY's per-specimen test-time optimization which adapts to the observed data without training labels (Figure 6; full depth-metric breakdown in Appendix H.4). These projected 2D and 2.5D metrics validate segmentation and depth recovery under real acquisition. The full volumetric claim rests on the synthetic benchmark with exact 3D ground truth (Section 5.2).

### 5.4 Ablations

Table 3 traces the contribution of each NeFTY component. Positional encoding (PE) and frequency annealing (FA) push IoU from $0.03$ to $0.14$ by stabilizing high-frequency optimization against the noise amplification of Corollary 1. The bounded sigmoid output ($\sigma$) prevents drift into the ill-conditioned regime $\alpha_\theta \to 0$, and the harmonic-mean coefficient (HM, Proposition 1) restores effective-flux continuity at high-contrast interfaces. The TV regularizer doubles IoU again to $0.45$/$0.37$ by aligning the prior with the piecewise-constant defect structure of the effective-diffusivity model. The full ablation with $95\%$ CIs and a per-difficulty breakdown is in Appendix G.

### 5.5 Computational efficiency

Table 4 reports a per-step solver benchmark on a single specimen. The discrete adjoint method (AM) of Section 4.3 matches the simulation accuracy of standard Autograd (AD) bit-for-bit while reducing

Table 3: **Ablation of NeFTY:** each row adds one component (H/L = homogeneous/Layered).

| VARIANT | PSNR (H/L) ↑ | IOU (H/L) ↑ |
|---|---|---|
| BASE | 0.47 / 2.89 | 0.03 / 0.02 |
| + PE | 4.17 / 6.15 | 0.09 / 0.09 |
| + PE, FA | 8.33 / 8.81 | 0.14 / 0.14 |
| + PE, FA, $\sigma$ | 9.19 / 9.27 | 0.18 / 0.14 |
| + PE, FA, $\sigma$, HM | 10.95 / 11.26 | 0.24 / 0.22 |
| **FULL (+ TV)** | **18.48 / 15.88** | **0.45 / 0.37** |

Table 4: **Solver-level efficiency benchmark:** our discrete adjoint method (AM) matches the simulation accuracy of autograd (AD) while reducing peak GPU memory by orders of magnitude on a single $64 \times 64 \times 16$ sample over 50 time steps.

| METHOD | FWD TIME (S) ↓ | BWD TIME (S) ↓ | PEAK MEM ↓ | SIM. ERROR ↓ |
|---|---|---|---|---|
| PHIFLOW (EX) | $26.36 \pm 0.10$ | $0.87 \pm 0.01$ | 3.26 GB | $7.91 \times 10^{-4}$ |
| PHIFLOW (IM) | $3.26 \pm 0.40$ | $0.76 \pm 0.01$ | 275.5 MB | $8.47 \times 10^{-4}$ |
| OURS (AD) | $1.43 \pm 0.04$ | $1.30 \pm 0.02$ | 18.63 GB | $\mathbf{3.73 \times 10^{-8}}$ |
| **OURS (AM)** | $\mathbf{0.46 \pm 0.00}$ | $\mathbf{0.50 \pm 0.00}$ | **21.9 MB** | $\mathbf{3.73 \times 10^{-8}}$ |

peak GPU memory by three orders of magnitude (from 18.63 GB to 21.9 MB) and halving the backward time, in agreement with the solver-level $\mathcal{O}(N_g)$ versus $\mathcal{O}(KN_gN_t)$ analysis of Appendix D.3. The same AM forward pass is also $\sim 7\times$ faster than PhiFlow's implicit solver. The training-level wall-clock comparison across all baselines (NeFTY converges in 9.6 min versus 18.2-58.3 min for the soft-PINN family on the same hardware) is in Table 10 of Appendix G. Up to $128 \times 128 \times 32$ grids and 200 frames, peak memory scales with exponents 0.985 in voxels and 0.694 in frames, with a near-flat marginal cost per additional frame (Appendix G.12).

## 6 Conclusion

We presented NeFTY, a hard-constrained neural field framework for three-dimensional inverse heat conduction. NeFTY parameterizes the unknown thermal diffusivity as a continuous coordinate-based field, enforces the governing PDE through a differentiable implicit-Euler solver with harmonic-mean interface flux, and inverts the resulting forward map with adjoint gradients at solver-level memory cost. On synthetic homogeneous and layered settings it reaches IoU 0.45 and 0.37, while four soft-constrained PINN variants and a voxel-grid baseline remain at or near zero and classical L-BFGS voxel inversion through the same solver trails by about $3.3\times$ on the same specimens. On real PVC pulsed-thermography data it segments defects at IoU 0.431 and recovers depth at Abs Rel 0.465 from surface measurements alone.

**Limitations and Broader Impact.** NeFTY shares the limitations common to test-time inverse-problem solvers: per-specimen optimization (9.6 min and 4.3 GB on one workstation GPU, suited to offline inspection), attenuated quantitative recovery inside very high-contrast voids, and evaluation against projected 2D and 2.5D supervision because public NDT benchmarks do not yet provide 3D volumetric ground truth. Future directions include amortized initialization through meta-learning or hypernetworks, preconditioned solvers in the differentiable loop, and paired CT and material-characterization labels for full volumetric evaluation (Appendix J). We primarily target defects that form discrete regions in a host material, imaged with the same one-sided, non-contact acquisition as PPT and TSR. Smoothly varying fields, non-Cartesian geometries, anisotropy, and temperature-dependent coefficients are important future scope. The intended use of NeFTY is non-destructive evaluation, where it can support safer infrastructure inspection and earlier defect detection without destructive testing or ionizing X-ray exposure. As with any automated inspection pipeline, its outputs are best treated as decision support for human qualification of safety-critical components, and reconstructions in regions of low surface sensitivity should be reported with uncertainty alongside the recovered field.

## Acknowledgments and Disclosure of Funding

This project is partially supported by the AI Lab Seed Grant from the Princeton Laboratory for Artificial Intelligence. The authors declare no competing interests.

## References

[1] Zain Ali, Sri Addepalli, and Yifan Zhao. Effective thermal diffusivity measurement using through-transmission pulsed thermography: Extending the current practice by incorporating multi-parameter optimisation. *Sensors*, 25(4):1139, 2025.

[2] O M Alifanov. *Inverse Heat Transfer Problems*. International Series in Heat and Mass Transfer. Springer-Verlag, 1994.

[3] Nacime Bouziani, David A Ham, and Ado Farsi. Differentiable programming across the pde and machine learning barrier. *arXiv preprint arXiv:2409.06085*, 2024.

[4] Peter Burgholzer, Michael Thor, Jürgen Gruber, and Günther Mayr. Three-dimensional thermographic imaging using a virtual wave concept. *Journal of Applied Physics*, 121(10), 2017.

[5] Peter Burgholzer, Gregor Stockner, and Guenther Mayr. Acoustic reconstruction for photothermal imaging. *Bioengineering*, 5(3):70, 2018.

[6] 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.

[7] Jean Céa. Conception optimale ou identification de formes, calcul rapide de la dérivée directionnelle de la fonction coût. *ESAIM: Modélisation mathématique et analyse numérique*, 20(3):371–402, 1986.

[8] Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International conference on machine learning*, pages 794–803. PMLR, 2018.

[9] Junwoo Cho, Seungtae Nam, Hyunmo Yang, Seok-Bae Yun, Youngjoon Hong, and Eunbyung Park. Separable physics-informed neural networks. *Advances in Neural Information Processing Systems*, 36: 23761–23788, 2023.

[10] Yoonjae Chung, Seungju Lee, and Wontae Kim. Latest advances in common signal processing of pulsed thermography for enhanced detectability: A review. *Applied Sciences*, 11(24):12168, 2021.

[11] S F Cobos, C J Norley, S I Pollmann, and D W Holdsworth. Cost-effective micro-CT system for non-destructive testing of titanium 3D printed medical components. *PLOS ONE*, 17(10):e0275732, 2022.

[12] Richard Courant, Kurt Friedrichs, and Hans Lewy. Über die partiellen differenzengleichungen der mathematischen physik. *Mathematische annalen*, 100(1):32–74, 1928.

[13] Filipe de Avila Belbute-Peres, Kevin Smith, Kelsey Allen, Josh Tenenbaum, and J Zico Kolter. End-to-end differentiable physics for learning and control. *Advances in neural information processing systems*, 31, 2018.

[14] Jonas Degrave, Michiel Hermans, Joni Dambre, and Francis Wyffels. A differentiable physics engine for deep learning in robotics. *Frontiers in neurorobotics*, 13:6, 2019.

[15] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. *Advances in neural information processing systems*, 27, 2014.

[16] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[17] Qiang Fang, Clemente Ibarra-Castanedo, Iván Garrido, Yuxia Duan, and Xavier Maldague. Automatic detection and identification of defects by deep learning algorithms from pulsed thermography data. *Sensors*, 23(9):4444, 2023.

[18] Lukas Gahleitner, Gregor Thummerer, Bernhard Plank, Johannes Wiedemann, Günther Mayr, Christian Hühne, Peter Burgholzer, and Umut Cakmak. Photothermal defect imaging in hybrid fiber metal laminates using the virtual wave concept. *Journal of Applied Physics*, 135(7), 2024.

[19] Serafina C Garcea, Ying Wang, and Philip J Withers. X-ray computed tomography of polymer composites. *Composites Science and Technology*, 156:305–319, 2018.

[20] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[21] E Haber, U M Ascher, and D Oldenburg. On optimization techniques for solving nonlinear inverse problems. *Inverse Problems*, 16(5):1263–1280, 2000.

[22] Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52, 1902.

[23] Baoli Hao, Ulisses Braga-Neto, Chun Liu, Lifan Wang, and Ming Zhong. Training pinns with hard constraints and adaptive weights: An ablation study. *arXiv preprint arXiv:2404.16189*, 2024.

[24] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI brainlesion workshop*, pages 272–284. Springer, 2021.

[25] M Hinze, R Pinnau, M Ulbrich, and S Ulbrich. *Optimization with PDE Constraints*. Mathematical Modelling: Theory and Applications. Springer, 2009.

[26] Philipp Holl and Nils Thuerey. $\phi$-flow: Differentiable simulations for pytorch, tensorflow and jax. In *Proceedings of the Forty-first International Conference on Machine Learning*, 2024.

[27] Philipp Holl, Vladlen Koltun, and Nils Thuerey. Learning to control pdes with differentiable physics. *arXiv preprint arXiv:2001.07457*, 2020.

[28] Youngsik Hwang and Dong-Young Lim. Dual cone gradient descent for training physics-informed neural networks. *Advances in Neural Information Processing Systems*, 37:98563–98595, 2024.

[29] Clemente Ibarra-Castanedo and Xavier Maldague. Pulsed phase thermography reviewed. *Quantitative Infrared Thermography Journal*, 1(1):47–70, 2004.

[30] Y Jarny, M N Özişik, and J P Bardon. A general optimization method using adjoint equation for solving multidimensional inverse heat conduction. *International Journal of Heat and Mass Transfer*, 34(11): 2911–2919, 1991.

[31] Dustin L Kelly and Brian S Thurow. Fluidnerf: A scalar-field reconstruction technique for flow diagnostics using neural radiance fields. In *AIAA SciTech 2023 Forum*, page 0412, 2023.

[32] Péter Kovács, Bernhard Lehner, Gregor Thummerer, Günther Mayr, Peter Burgholzer, and Mario Huemer. Deep learning approaches for thermographic imaging. *Journal of Applied Physics*, 128(15), 2020.

[33] A S Krishnapriyan, A Gholami, S Zhe, R M Kirby, and M W Mahoney. Characterizing possible failure modes in physics-informed neural networks. *Advances in Neural Information Processing Systems*, pages 26548–26560, 2021.

[34] Theodoros Leontiou, Anna Frixou, Marios Charalambides, Efstathios Stiliaris, Costas N Papanicolas, Sofia Nikolaidou, and Antonis Papadakis. Three-dimensional thermal tomography with physics-informed neural networks. *Tomography*, 10(12):1930, 2024.

[35] Y Liu, Y Yao, F Wang, S Sfarra, and K Liu. Review of unsupervised machine learning methods in active infrared thermography for defect detection and analysis. *Quantitative InfraRed Thermography Journal*, 23 (3):223–250, 2025.

[36] Luke Lozenski, Refik Mert Cam, Mark D Pagel, Mark A Anastasio, and Umberto Villa. Proxnf: neural field proximal training for high-resolution 4d dynamic image reconstruction. *IEEE Transactions on Computational Imaging*, 2024.

[37] Botao Ma, Shupeng Sun, and Lin Zhang. Quantitative depth estimation in lock-in thermography: Modeling and correction of lateral heat conduction effects. *Materials*, 18(22):5247, 2025.

[38] Xavier Maldague, François Galmiche, and Adel Ziadi. Advances in pulsed phase thermography. *Infrared physics & technology*, 43(3-5):175–181, 2002.

[39] Andrea Carolina Martinez Mundarain. Artificial neural networks as the solution of inverse heat conduction problems in multidimensional domains. 2024.

[40] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65 (1):99–106, 2021.

[41] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[42] BCF Oliveira, AA Seibert, VK Borges, A Albertazzi, and RH Schmitt. Employing a u-net convolutional neural network for segmenting impact damages in optical lock-in thermography images of cfrp plates. *Nondestructive Testing and Evaluation*, 36(4):440–458, 2021.

[43] Derek Onken and Lars Ruthotto. Discretize-optimize vs. optimize-discretize for time-series regression and continuous normalizing flows. *arXiv preprint arXiv:2005.13420*, 2020.

[44] M N Özişik and H R B Orlande. *Inverse Heat Transfer: Fundamentals and Applications*. CRC Press, 2nd edition, 2021.

[45] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5865–5874, 2021.

[46] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[47] S. Patankar. *Numerical Heat Transfer and Fluid Flow*. Series in computational methods in mechanics and thermal sciences. Taylor & Francis, 1980. ISBN 9780891165224.

[48] Shaoyang Peng, Sri Addepalli, and Maryam Farsi. Machine learning in thermography non-destructive testing: a systematic review. *Applied Sciences*, 15(17):9624, 2025.

[49] Eduardo Pérez, Cemil Emre Ardıç, Ozan Çakıroğlu, Kevin Jacob, Sayako Kodera, Luca Pompa, Mohamad Rachid, Han Wang, Yiming Zhou, Cyril Zimmer, et al. Integrating ai in nde: techniques, trends, and further directions. *NDT & E International*, 156:103442, 2025.

[50] Weijia Qian, Xin Hui, Bosen Wang, Zongwei Zhang, Yuzhen Lin, and Siheng Yang. Physics-informed neural network for inverse heat conduction problem. *Heat Transfer Research*, 54(4), 2023.

[51] 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.

[52] A Rashidi, T Olfatbakhsh, B Crawford, and A S Milani. A review of current challenges and case study toward optimizing micro-computed X-ray tomography of carbon fabric composites. *Materials*, 13(16): 3606, 2020.

[53] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[54] Renan Garcia Rosa, Bruno Pereira Barella, Iago Garcia Vargas, José Ricardo Tarpani, Hans-Georg Herrmann, and Henrique Fernandes. Advanced thermal imaging processing and deep learning integration for enhanced defect detection in carbon fiber-reinforced polymer laminates. *Materials*, 18(7):1448, 2025.

[55] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[56] F Santosa. A level-set approach for inverse problems involving obstacles. *ESAIM: Control, Optimisation and Calculus of Variations*, 1:17–33, 1996.

[57] Alexander Schager, Gerald Zauner, Günther Mayr, and Peter Burgholzer. Extension of the thermographic signal reconstruction technique for an automated segmentation and depth estimation of subsurface defects. *Journal of Imaging*, 6(9):96, 2020.

[58] Steven M Shepard and Maria Frendberg Beemer. Advances in thermographic signal reconstruction. In *Thermosense: thermal infrared applications XXXVII*, volume 9485, pages 204–210. SPIE, 2015.

[59] Steven M Shepard, D Wang, James R Lhota, Bruce A Rubadeux, and Tasdiq Ahmed. Reconstruction and enhancement of thermographic sequence data. In *Nondestructive evaluation and health monitoring of aerospace materials and civil infrastructures*, volume 4704, pages 74–77. SPIE, 2002.

[60] Xijin Shi and Sheng-Jen Hsieh. Infrared imaging and machine learning techniques for plant root location and depth prediction. In *Thermosense: Thermal Infrared Applications XLIII*, volume 11743, page 1174303. SPIE, 2021.

[61] A Sirikham, Y Zhao, H Liu, Y Xu, S Williams, and J Mehnen. Three-dimensional subsurface defect shape reconstruction and visualisation by pulsed thermography. *Infrared Physics & Technology*, 104:103151, 2020.

[62] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

[63] Dylan Turpin, Tao Zhong, Shutong Zhang, Guanglei Zhu, Eric Heiden, Miles Macklin, Stavros Tsogkas, Sven Dickinson, and Animesh Garg. Fast-grasp'd: Dexterous multi-finger grasp generation through differentiable simulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

[64] V Vavilov, Xavier Maldague, J Picard, RL Thomas, and LD Favro. Dynamic thermal tomography: new nde technique to reconstruct inner solids structure using multiple ir image processing. In *Review of progress in quantitative nondestructive evaluation*, pages 425–432. Springer, 1992.

[65] 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.

[66] Sifan Wang, Shyam Sankaran, and Paris Perdikaris. Respecting causality for training physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 421:116813, 2024.

[67] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[68] Ziang Wei, Ahmad Osman, Bernd Valeske, and Xavier Maldague. A dataset of pulsed thermography for automated defect depth estimation. *Applied Sciences*, 13(24):13093, 2023.

[69] Ziang Wei, Ahmad Osman, Bernd Valeske, and Xavier Maldague. Pulsed thermography dataset for training deep learning models. *Applied Sciences*, 13(5):2901, 2023.

[70] Z Wu, S Chen, F Feng, J Qi, L Feng, N Tao, and C Zhang. Automatic defect detection and three-dimensional reconstruction from pulsed thermography images based on a bidirectional long-short term memory network. *Engineering Applications of Artificial Intelligence*, 124:106574, 2023.

[71] Di Xu, Yang Yang, Hengjie Liu, Qihui Lyu, Martina Descovich, Dan Ruan, and Ke Sheng. Tomograf: An x-ray physics-driven generative radiance field framework for extremely sparse view ct reconstruction. *Plos one*, 20(8):e0330463, 2025.

[72] Tao Zhong and Christine Allen-Blanchette. Gagrasp: Geometric algebra diffusion for dexterous grasping. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6771–6778, 2025. doi: 10.1109/ICRA55743.2025.11127957.

[73] Li Zhou, Changsheng Fang, Bahareh Morovati, Yongtong Liu, Shuo Han, Yongshun Xu, and Hengyong Yu. $\rho$-nerf: leveraging attenuation priors in neural radiance field for 3d computed tomography reconstruction. In *2025 IEEE International Conference on Image Processing (ICIP)*, pages 1636–1641. IEEE, 2025.

# Appendix

## A Effective Diffusivity Model and Interface Flux

### A.1 Derivation of the diffusivity form

Let $\rho, C_p \in C^1(\Omega)$ with $\rho C_p > 0$ on $\bar{\Omega}$, and let $T \in C^{2,1}(\Omega \times (0, t_{\text{end}}])$ satisfy the conservation form (1) with $Q \equiv 0$:

$$\rho C_p \, \partial_t T \;=\; \nabla \cdot (k \nabla T).$$

Dividing both sides by the strictly positive function $\rho C_p$ and applying the product rule $\nabla \cdot (\gamma \mathbf{v}) = \gamma \nabla \cdot \mathbf{v} + \nabla \gamma \cdot \mathbf{v}$ with $\gamma = 1/(\rho C_p)$ and $\mathbf{v} = k \nabla T$ gives

$$\begin{aligned} \partial_t T \;&=\; \frac{1}{\rho C_p} \nabla \cdot (k \nabla T) \;=\; \nabla \cdot \left( \frac{k}{\rho C_p} \nabla T \right) \;-\; \nabla \left( \frac{1}{\rho C_p} \right) \cdot (k \nabla T) \\ &=\; \nabla \cdot (\alpha \nabla T) \;-\; k \nabla T \cdot \nabla \left( \frac{1}{\rho C_p} \right), \end{aligned}$$

with $\alpha := k/(\rho C_p)$. The reduction is reversible by multiplying through by $\rho C_p$, so the conservation form (1) (under $Q \equiv 0$) and the diffusivity form (2) are equivalent on $\Omega \times (0, t_{\text{end}}]$ whenever $\rho C_p \in C^1$.

### A.2 Distributional treatment of the heat-capacity correction

When $\rho C_p$ is piecewise-constant with values $c_b > 0$ on a bulk subdomain $\Omega_b$ and $c_d > 0$ on a defect subdomain $\Omega_d$, separated by a smooth interface $\Sigma = \partial\Omega_b \cap \partial\Omega_d$, the gradient $\nabla(1/(\rho C_p))$ is a vector-valued distribution supported on $\Sigma$:

$$\nabla \left( \frac{1}{\rho C_p} \right) \;=\; \left( \frac{1}{c_d} - \frac{1}{c_b} \right) \mathbf{n}_\Sigma \, \delta_\Sigma, \tag{12}$$

where $\mathbf{n}_\Sigma$ is the unit normal to $\Sigma$ pointing from $\Omega_b$ to $\Omega_d$ and $\delta_\Sigma$ is the surface measure on $\Sigma$. The correction term in (2) reads

$$-\, k \nabla T \cdot \nabla \left( \frac{1}{\rho C_p} \right) \;=\; -\left( \frac{1}{c_d} - \frac{1}{c_b} \right) (k \nabla T \cdot \mathbf{n}_\Sigma) \, \delta_\Sigma,$$

which is identically zero on $\Omega \setminus \Sigma$ and concentrates as a jump in the normal physical Fourier flux $q := -k \, \partial_n T$ across $\Sigma$. A finite-volume discretization on a uniform Cartesian grid of spacing $\Delta x$ cannot resolve a Dirac measure on a measure-zero interface [47]. Dropping this distributional term yields the bulk equation $\partial_t T = \nabla \cdot (\alpha \nabla T)$ on $\Omega \setminus \Sigma$, which is the model NeFTY actually solves.

The single-$\alpha$ effective model has its own interface conservation law: across a face where $\alpha$ jumps, the effective $\alpha$-flux $j_\alpha := -\alpha \partial_n T$ must be continuous, since otherwise temperature would lose mass at that face. The remainder of this appendix shows that the harmonic-mean stencil enforces this $j_\alpha$-continuity exactly: it is the unique discrete analogue of the jump condition appropriate to the single-$\alpha$ model that NeFTY parameterizes, while the dropped distributional term belongs to the full $\rho C_p$ model.

### A.3 Proof of Proposition 1 (Harmonic mean as discrete interface continuity)

Consider two adjacent one-dimensional cells with center coordinates $x_i$ and $x_{i+1} = x_i + \Delta x$, cell-centered diffusivities $\alpha_i$ and $\alpha_{i+1}$, and cell-centered temperatures $T_i$ and $T_{i+1}$. Assume each cell has constant diffusivity, with the discontinuity located at the cell face $x_{i+1/2} = x_i + \Delta x/2$. Let $T_{i+1/2}$ denote the (a priori unknown) face temperature.

Within the left half-cell $[x_i, x_{i+1/2}]$, Fourier's law for the effective $\alpha$-flux $j_\alpha = -\alpha \partial_x T$ gives a constant flux and a linear temperature profile, so

$$j_\alpha \;=\; -\alpha_i \cdot \frac{T_{i+1/2} - T_i}{\Delta x/2}. \tag{13}$$

Within the right half-cell $[x_{i+1/2}, x_{i+1}]$, by the same argument,

$$j_\alpha \;=\; -\alpha_{i+1} \cdot \frac{T_{i+1} - T_{i+1/2}}{\Delta x/2}. \tag{14}$$

Continuity of the effective flux at $x_{i+1/2}$ in the single-$\alpha$ model imposes the same value of $j_\alpha$ in (13) and (14). From (13), $T_{i+1/2} = T_i - j_\alpha \Delta x/(2\alpha_i)$. From (14), $T_{i+1/2} = T_{i+1} + j_\alpha \Delta x/(2\alpha_{i+1})$. Equating and rearranging,

$$T_i - T_{i+1} \;=\; j_\alpha \, \frac{\Delta x}{2}\left(\frac{1}{\alpha_i} + \frac{1}{\alpha_{i+1}}\right) \;=\; j_\alpha \, \frac{\Delta x \, (\alpha_i + \alpha_{i+1})}{2\alpha_i \alpha_{i+1}}.$$

Solving for $j_\alpha$,

$$j_\alpha \;=\; -\frac{2\alpha_i \alpha_{i+1}}{\alpha_i + \alpha_{i+1}} \cdot \frac{T_{i+1} - T_i}{\Delta x} \;=:\; -\bar{\alpha}_{i+1/2} \cdot \frac{T_{i+1} - T_i}{\Delta x}.$$

The expression $\bar{\alpha}_{i+1/2} = 2\alpha_i \alpha_{i+1}/(\alpha_i + \alpha_{i+1})$ is the harmonic mean. Uniqueness follows from the linearity of the two-equation system in $(j_\alpha, T_{i+1/2})$ with non-degenerate coefficient matrix. □

### A.4 Multi-dimensional consequence and physical interpretation

In ambient spatial dimension $D \in \{2, 3\}$ on a uniform Cartesian grid, the same argument applied along each principal direction gives the second-order finite-volume diffusion operator $\mathbf{L}(\alpha)$ used by NeFTY, defined by its action on a discrete temperature field $\mathbf{T}$ at grid node $(i, j, k)$:

$$[\mathbf{L}(\alpha)\mathbf{T}]_{i,j,k} \;:=\; \frac{1}{\Delta x^2}\left[\bar{\alpha}_{i+1/2}(T_{i+1,j,k} - T_{i,j,k}) - \bar{\alpha}_{i-1/2}(T_{i,j,k} - T_{i-1,j,k})\right] \;+\; (y\text{-, } z\text{-terms}).$$

The harmonic mean has two physically important properties for thermal NDT. First, $\bar{\alpha}_{i+1/2}$ is dominated by the smaller of $\alpha_i, \alpha_{i+1}$ when the two are well separated, so an insulating defect cell $\alpha_{i+1} \ll \alpha_i$ throttles flux at its boundary as physical thermal resistance does. Second, $\bar{\alpha}$ is exact for piecewise-constant $\alpha$ at cell faces and second-order accurate for smooth $\alpha$. Using the arithmetic mean instead is first-order accurate and overestimates flux across high-contrast interfaces, leading to systematic underestimation of defect depth in inverse reconstruction [47]. NeFTY's adoption of the harmonic mean is therefore the unique discrete analogue of the effective-flux continuity condition for the single-$\alpha$ bulk equation in (2).

### A.5 Initial and boundary conditions: synthetic and real PVC datasets

The bulk diffusivity equation $\partial_t T = \nabla \cdot (\alpha \nabla T)$ is closed by an initial condition and by lateral and through-thickness boundary conditions. The synthetic dataset and the real PVC datasets [68, 69] require different choices, all consistent with the same underlying physics but adapted to the specimen scale, the heating modality, and the dominant non-dimensional groups.

**Synthetic dataset.** The synthetic data are generated on a unitless slab $\Omega = [0, L]^2 \times [0, H]$ with $L = 10$, $H = 1$, grid resolution $64 \times 64 \times 16$ and 100 transient frames, on micron-scale specimens at microsecond timescales for which the Biot number is small and convective dissipation is negligible. We accordingly impose:

- *Initial condition.* A spatially Gaussian post-flash temperature distribution localized near the front face,

$$T_0(\mathbf{x}) \;=\; T_{\mathrm{amb}} + A\exp\big(-((x-x_c)^2+(y-y_c)^2)/(2w_{xy}^2)\big)\exp\big(-z^2/(2w_z^2)\big),$$

  where $T_{\mathrm{amb}}$ is the ambient temperature, $A$ is the deposited pulse amplitude, $(x_c, y_c)$ is the lateral footprint center, $w_{xy}$ is the lateral pulse width, and $w_z \ll H$ is the absorption depth set by the focused-laser pulse profile.
- *Lateral boundaries ($\partial\Omega_{xy}$).* Periodic conditions on $\partial\Omega_{xy} := [0,L]^2 \times \{0,L\}_x \times \{0,L\}_y$ (the $x = 0/L$ and $y = 0/L$ faces), modeling a semi-infinite slab and removing reflections from the truncated simulation grid.
- *Through-thickness boundaries ($\partial\Omega_z$).* Adiabatic conditions $\mathbf{n}\cdot(\alpha\nabla T) = 0$ on $\partial\Omega_z$ (the front $z = 0$ and back $z = H$ faces), modeling negligible convective loss during the short inspection window.

**Real PVC dataset.** The PVC-Infrared [69] and PVC-Depth [68] benchmarks consist of pulsed thermography sequences on $100 \times 100 \times 5$ mm PVC specimens with cylindrical flat-bottom holes of varying size and depth, recorded over second-scale dynamics. Three physical adaptations are required so that the same NeFTY solver applies without retraining the ground-truth simulator:

- *Initial condition.* The Gaussian pulse is replaced by a near-uniform front-face flash $T_0(\mathbf{x}) = T_{\mathrm{amb}} + A\exp(-z^2/(2w_z^2))$, matching the broad flash-lamp excitation used in the benchmark setups (lateral profile uniform over the inspection footprint, absorption depth $w_z$ small relative to specimen thickness).
- *Centimeter and second-scale Fourier-number rescaling.* The PVC specimens are centimeter-scale at second-scale dynamics rather than the synthetic micron / microsecond regime, so the unitless solver is rescaled through the Fourier number $\mathrm{Fo} = \alpha_{\mathrm{phys}}\, t_{\mathrm{total}}/L_0^2$ (full scaling below).
- *Convective Robin back-face boundary.* The centimeter-scale specimen and second-scale dynamics make convective dissipation non-negligible. On the back face we therefore impose a Robin condition $\mathbf{n}\cdot(\alpha\nabla T) = -h\,(T - T_{\mathrm{amb}})$ with $h > 0$ a calibrated convective coefficient, recovering the asymptotic Biot regime appropriate for the PVC scale. The front face retains the adiabatic Neumann condition, and lateral conditions remain periodic on interior crops or are replaced by adiabatic Neumann depending on the recording.

The unitless and dimensional scales are aligned through the Fourier number $\mathrm{Fo} = \alpha_{\mathrm{phys}}\, t_{\mathrm{total}}/L_0^2$, with $\alpha_{\mathrm{phys}}$ the physical diffusivity (in $\mathrm{m}^2/\mathrm{s}$), $t_{\mathrm{total}}$ the physical experiment duration (in s), and $L_0$ a characteristic specimen length scale (in m). The unitless solver evolves in $(t/t_{\mathrm{total}}, \mathbf{x}/L_0)$ coordinates with effective diffusivity $\alpha = \alpha_{\mathrm{phys}}\, t_{\mathrm{total}}/L_0^2$. A single set of unitless solver parameters represents widely different material classes by reinterpreting $L_0$ and $t_{\mathrm{total}}$, so the synthetic and real PVC settings differ in their initial-condition shape and back-face Robin coefficient but share the same forward operator $\mathcal{K}$ and the same harmonic-mean stencil derived above. Full physical scaling, calibrated $h$ values, and projection from a reconstructed 3D $\alpha$ field to the 2D defect masks and 2.5D depth labels supplied by the benchmarks are documented in Appendix H.

# B Ill-Posedness of the Inverse Heat Conduction Problem

This appendix establishes the compactness and algebraic singular-value decay claimed in Proposition 2. The argument is consistent with the standard treatment of parameter identification in parabolic inverse problems [16, 22, 39]. We present a self-contained derivation on the slab geometry used in our experiments. The proof factorizes $d\mathcal{K}_{\alpha_0} = B \circ J$, where $B : L^2(\Omega) \to L^2(\Gamma_{\mathrm{obs}} \times (0, t_{\mathrm{end}}))$ is bounded by an energy estimate plus a parabolic boundary trace, and $J : H^1(\Omega) \hookrightarrow L^2(\Omega)$ is the compact Rellich–Kondrachov embedding whose approximation numbers along the Laplacian eigenbasis decay as $(1+\lambda_n)^{-1/2}$. Truncating the parameter side via $J$ rather than the source side avoids the mode-mixing introduced by multiplication with $\nabla T_{\alpha_0}$.

## B.1 Linearization of the parameter-to-observation map

Let $\alpha_0 > 0$ be a constant background diffusivity and let $T_{\alpha_0}(\mathbf{x}, t)$ solve the bulk heat equation $\partial_t T = \alpha_0 \Delta T$ with the prescribed initial condition $T_0$ and boundary conditions on the slab $\Omega = [0, L]^2 \times [0, H]$. Throughout the proof we assume the background satisfies $\nabla T_{\alpha_0} \in L^\infty(Q)$ with

$Q := \Omega \times (0, t_{\mathrm{end}})$. Consider a perturbation $\alpha = \alpha_0 + \varepsilon\,\delta\alpha$ with $\delta\alpha \in H^1(\Omega)$ and $\varepsilon > 0$ small. To first order in $\varepsilon$, the perturbation $\delta T = (T_\alpha - T_{\alpha_0})/\varepsilon$ solves the linearized PDE in the weak sense

$$\langle \partial_t(\delta T), v\rangle + \alpha_0\big(\nabla(\delta T), \nabla v\big)_{L^2(\Omega)} \;=\; -\int_\Omega \delta\alpha\, \nabla T_{\alpha_0}(\cdot,t)\cdot\nabla v\, d\mathbf{x}, \tag{15}$$

for all $v \in H^1(\Omega)$ compatible with the imposed boundary conditions, with vanishing initial condition $\delta T(\cdot,0) = 0$. We define the source distribution $F_{\delta\alpha}(t) \in H^{-1}(\Omega)$ by $\langle F_{\delta\alpha}(t), v\rangle = -\int_\Omega \delta\alpha\, \nabla T_{\alpha_0}(\cdot,t)\cdot\nabla v\, d\mathbf{x}$, satisfying

$$\|F_{\delta\alpha}(t)\|_{H^{-1}(\Omega)} \;\le\; \|\nabla T_{\alpha_0}(\cdot,t)\|_{L^\infty(\Omega)}\, \|\delta\alpha\|_{L^2(\Omega)}. \tag{16}$$

The Fréchet derivative is the linear map $d\mathcal{K}_{\alpha_0} : H^1(\Omega) \to L^2(\Gamma_{\mathrm{obs}} \times (0,t_{\mathrm{end}}))$ defined by $d\mathcal{K}_{\alpha_0}(\delta\alpha) = \delta T|_{\Gamma_{\mathrm{obs}}\times(0,t_{\mathrm{end}})}$.

### B.2 Proof of Proposition 2 (Compactness and algebraic singular-value decay)

The proof has four steps: (i) an energy estimate in $L^2(0,t_{\mathrm{end}};H^1(\Omega))$ for $\delta T$; (ii) a parabolic trace inequality giving boundedness of the auxiliary map $B$; (iii) a finite-rank approximation through the Rellich–Kondrachov embedding $J$; (iv) Weyl asymptotics on the slab.

**Step 1: energy estimate.** Existence and uniqueness of a weak solution $\delta T \in L^2(0,t_{\mathrm{end}};H^1(\Omega)) \cap H^1(0,t_{\mathrm{end}};H^{-1}(\Omega))$ to (15) with vanishing initial data and source $F_{\delta\alpha} \in L^2(0,t_{\mathrm{end}};H^{-1}(\Omega))$ follow from standard Galerkin / Lions–Magenes parabolic theory [16]. With this regularity, testing (15) with $v = \delta T(\cdot,t)$ is permissible and gives

$$\tfrac{1}{2}\tfrac{d}{dt}\|\delta T(\cdot,t)\|^2_{L^2(\Omega)} + \alpha_0\|\nabla\delta T(\cdot,t)\|^2_{L^2(\Omega)} \;=\; -\int_\Omega \delta\alpha\, \nabla T_{\alpha_0}(\cdot,t)\cdot\nabla(\delta T)(\cdot,t)\, d\mathbf{x}.$$

Applying Young's inequality $|ab| \le \frac{1}{2\alpha_0}a^2 + \frac{\alpha_0}{2}b^2$ to the right-hand side and integrating over $(0,t_{\mathrm{end}})$ with the vanishing initial condition,

$$\tfrac{1}{2}\|\delta T(\cdot,t_{\mathrm{end}})\|^2_{L^2(\Omega)} + \tfrac{\alpha_0}{2}\int_0^{t_{\mathrm{end}}} \|\nabla\delta T\|^2_{L^2(\Omega)}\, dt \;\le\; \tfrac{1}{2\alpha_0}\int_0^{t_{\mathrm{end}}} \|\delta\alpha\, \nabla T_{\alpha_0}(\cdot,t)\|^2_{L^2(\Omega)}\, dt.$$

With $\|\delta\alpha\,\nabla T_{\alpha_0}(\cdot,t)\|_{L^2(\Omega)} \le \|\nabla T_{\alpha_0}(\cdot,t)\|_{L^\infty(\Omega)}\,\|\delta\alpha\|_{L^2(\Omega)}$, this yields

$$\|\delta T\|_{L^2(0,t_{\mathrm{end}};H^1(\Omega))} \;\le\; C_E(\Omega,t_{\mathrm{end}},\alpha_0)\,\|\nabla T_{\alpha_0}\|_{L^\infty(Q)}\,\|\delta\alpha\|_{L^2(\Omega)}, \tag{17}$$

with $C_E$ an absolute constant depending only on $\Omega$, $t_{\mathrm{end}}$, $\alpha_0$ (a Poincaré-type constant absorbed into the bound).

**Step 2: boundary trace and the auxiliary operator $B$.** The Sobolev trace inequality $\|u|_{\partial\Omega}\|_{L^2(\partial\Omega)} \le C_\Gamma\|u\|_{H^1(\Omega)}$ on a Lipschitz domain combined with (17) gives

$$\|\delta T|_{\Gamma_{\mathrm{obs}}\times(0,t_{\mathrm{end}})}\|^2_{L^2(\Gamma_{\mathrm{obs}}\times(0,t_{\mathrm{end}}))} \;\le\; C_\Gamma^2\,\|\delta T\|^2_{L^2(0,t_{\mathrm{end}};H^1(\Omega))} \;\le\; C_B^2\,\|\delta\alpha\|^2_{L^2(\Omega)},$$

with $C_B := C_\Gamma\, C_E\,\|\nabla T_{\alpha_0}\|_{L^\infty(Q)}$. The map

$$B \,:\, L^2(\Omega) \to L^2(\Gamma_{\mathrm{obs}}\times(0,t_{\mathrm{end}})), \qquad B(\delta\alpha) \,=\, \delta T|_{\Gamma_{\mathrm{obs}}\times(0,t_{\mathrm{end}})},$$

is therefore bounded, $\|B\|_{\mathrm{op}} \le C_B$. The Fréchet derivative is $d\mathcal{K}_{\alpha_0} = B\circ J$, where $J : H^1(\Omega) \hookrightarrow L^2(\Omega)$ is the inclusion.

**Step 3: compactness and singular-value decay via Rellich–Kondrachov.** The embedding $J$ is compact by the Rellich–Kondrachov theorem on bounded Lipschitz domains. Equipping $H^1(\Omega)$ with the spectral norm $\|\delta\alpha\|^2_{H^1} = \sum_{j\ge0}(1+\lambda_j)|a_j|^2$ where $\delta\alpha = \sum_j a_j\phi_j$, the approximation numbers of $J$ along the Laplacian eigenbasis are exactly $(1+\lambda_n)^{-1/2}$: for the $L^2$-orthogonal projector $P_N$ onto $\mathrm{span}(\phi_0,\dots,\phi_{N-1})$,

$$\|(I-P_N)\delta\alpha\|^2_{L^2(\Omega)} \;=\; \sum_{j\ge N} a_j^2 \;\le\; (1+\lambda_N)^{-1}\sum_{j\ge N}(1+\lambda_j)a_j^2 \;\le\; (1+\lambda_N)^{-1}\,\|\delta\alpha\|^2_{H^1(\Omega)}.$$

Define the rank-$N$ operator $A_N := B\circ P_N\circ J$. For $\|\delta\alpha\|_{H^1}\le 1$,

$$\|(d\mathcal{K}_{\alpha_0} - A_N)(\delta\alpha)\|_{L^2(\Gamma_{\mathrm{obs}}\times(0,t_{\mathrm{end}}))} \;\le\; \|B\|_{\mathrm{op}}\,\|(I-P_N)(J\delta\alpha)\|_{L^2(\Omega)} \;\le\; C_B\,(1+\lambda_N)^{-1/2}.$$

Since approximation numbers and singular values coincide for compact operators between Hilbert spaces [16], this yields

$$\sigma_{N+1}(d\mathcal{K}_{\alpha_0}) \;\le\; C_B\,(1+\lambda_N)^{-1/2}, \qquad N \ge 0. \tag{18}$$

Compactness of $d\mathcal{K}_{\alpha_0}$ follows from $\sigma_n \to 0$, equivalently the convergence $A_N \to d\mathcal{K}_{\alpha_0}$ in operator norm. The mode-mixing introduced by multiplication with $\nabla T_{\alpha_0}$ in the source $F_{\delta\alpha}$ does not affect this argument: the truncation $P_N$ acts on the parameter $\delta\alpha$ before it enters the linearized PDE, and the small-$L^2$-tail of $\delta\alpha$ for $\delta\alpha \in H^1$-unit-ball produces a small output in $L^2(\Gamma_{\mathrm{obs}} \times (0, t_{\mathrm{end}}))$ via the $L^2$-bounded map $B$.

**Step 4: slab Weyl asymptotics.** For $\Omega = [0, L]^2 \times [0, H]$ with periodic conditions on the lateral faces and Neumann conditions on the top/bottom, the eigenfunctions are tensor products $\phi_{m,\ell,p}(x, y, z) = e^{2\pi i(mx/L+\ell y/L)} \cos(p\pi z/H)$ with eigenvalues $\lambda_{m,\ell,p} = (2\pi m/L)^2 + (2\pi\ell/L)^2 + (p\pi/H)^2$, indexed by $(m, \ell, p) \in \mathbb{Z}^2 \times \mathbb{Z}_{\ge 0}$. By Weyl's law in three dimensions, enumerated in increasing order $\lambda_n \asymp n^{2/3}$, so (18) reads $\sigma_n \lesssim n^{-1/3}$ on the slab. □

**Remark 1.** *The constant Laplacian eigenmode $\phi_0 \equiv |\Omega|^{-1/2}$ has $\lambda_0 = 0$, and adding a constant to $\alpha$ only rescales time without changing the inverse problem. Restricting to mean-zero perturbations $\delta\alpha \in \{a \in H^1(\Omega) : \int_\Omega a\, d\mathbf{x} = 0\}$ with norm $\|\nabla\delta\alpha\|_{L^2}$ replaces the bound by $\sigma_n \le C\,\lambda_n^{-1/2}$ for $n \ge 1$, recovering exactly the algebraic rate $\nu = 1/2$.*

### B.3 Proof of Corollary 1 (Hadamard ill-posedness)

By Proposition 2, the singular values of $d\mathcal{K}_{\alpha_0}$ tend to zero, so the Moore–Penrose pseudo-inverse $d\mathcal{K}_{\alpha_0}^\dagger$ is unbounded on the closure of the range. Concretely, the singular value decomposition $d\mathcal{K}_{\alpha_0}(\delta\alpha) = \sum_{n:\sigma_n>0} \sigma_n \langle \delta\alpha, v_n\rangle u_n$ has right singular vectors $v_n \in H^1(\Omega)$ in the parameter space and left singular vectors $u_n \in L^2(\Gamma_{\mathrm{obs}} \times (0, t_{\mathrm{end}}))$ in the data space, with $d\mathcal{K}_{\alpha_0}^\dagger u_n = \sigma_n^{-1} v_n$ for the indices $n$ with $\sigma_n > 0$ and $d\mathcal{K}_{\alpha_0}^\dagger u_n = 0$ on the kernel of $d\mathcal{K}_{\alpha_0}^*$. A measurement perturbation aligned with a non-degenerate $n$-th left singular vector $u_n$ therefore produces an $H^1$-norm perturbation in the recovered diffusivity of magnitude $\sigma_n^{-1} \ge C_B^{-1}(1+\lambda_n)^{1/2}$. The forward problem is well-posed (existence, uniqueness, continuous dependence in the forward direction follow from standard parabolic theory), but inversion fails the continuous-dependence criterion in the sense of Hadamard. □

### B.4 Discussion: pointwise damping versus operator decay

Two distinct smoothing rates appear in this analysis and should not be conflated. *Pointwise* in time, the free-space heat kernel $G_t(\mathbf{x}) = (4\pi\alpha_0 t)^{-3/2} \exp(-|\mathbf{x}|^2/(4\alpha_0 t))$ governs the impulse response: a localized defect at depth $z$ contributes a surface temperature signal whose amplitude decays Gaussian-fast in $z$, suppressed by $\exp(-z^2/(4\alpha_0 t))$ at observation time $t > 0$. Equivalently in the spectral picture, the heat semigroup $S(t) = e^{\alpha_0 \Delta t}$ damps the $j$-th Laplacian eigenmode by $e^{-\alpha_0\lambda_j t}$, so a defect supported at frequency $\lambda_j$ is exponentially attenuated at any fixed time. These pointwise statements are the depth- and frequency-decay heuristics that motivate the original ill-posedness intuition for IHCP.

The *operator* decay of $d\mathcal{K}_{\alpha_0}$, in contrast, integrates the boundary trace against the entire time interval $(0, t_{\mathrm{end}})$, and the bound (18) is only algebraic in $\lambda_n$. The two rates coincide in the degenerate observation model where the trace is taken at a single final time $t = t_{\mathrm{end}}$, in which case $\sigma_n(S(t_{\mathrm{end}})) = e^{-\alpha_0\lambda_n t_{\mathrm{end}}}$ recovers the exponential rate. Integrating over $(0, t_{\mathrm{end}})$ averages out this exponential factor and yields the algebraic bound proven above.

The practical consequences for IHCP follow from the algebraic operator-rate. Longer observation windows $t_{\mathrm{end}}$ and higher-conductivity bulk materials improve early-time signal-to-noise ratio but preserve the algebraic singular-value rate. Regularization remains mandatory at any rate, since any consistent reconstruction method must implicitly or explicitly bound the high-frequency content of admissible $\alpha$. NeFTY combines two such regularizers, namely a continuous neural-field prior with frequency annealing (Section 4) and an isotropic total-variation penalty $\mathcal{R}(\alpha) = \int_\Omega \|\nabla\alpha\|$ aligned with the piecewise-constant defect structure assumed by the effective-diffusivity model of Section 3.1.

# C Gradient Decoupling in the Two-Network Soft-Constrained PINN Formulation

PINNs remain the dominant and broadly successful paradigm for physics-informed learning [51]. This appendix analyzes one formulation of the inverse problem, the standard two-network formulation of Section 3.3: a temperature surrogate $T_\phi$ whose parameters $\phi$ are separate from the parameters $\theta$ of the diffusivity surrogate $\alpha_\theta$, trained on the data, residual, and initial-condition objective (4), applied to stiff parabolic inverse problems with surface-only data. Under these assumptions the decoupling follows from the parameterization alone (Appendix C.1), so optimizer quality and tuning play no role in the derivation. This identity is distinct from the conditioning and optimization difficulties that PDE-residual penalties can introduce [33]. The derivation establishes decoupling. The collapse of $\alpha_\theta$ is the observed behavior of the four evaluated variants (Section 5.2). Appendix C.4 examines a soft variant outside these assumptions.

## C.1 Derivation of the gradient decoupling

Write $\mathcal{L}_{\mathrm{PINN}} = \mathcal{L}_{\mathrm{data}}(T_\phi) + \lambda_{\mathrm{PDE}}\mathcal{L}_{\mathrm{PDE}}(T_\phi, \alpha_\theta) + \lambda_{\mathrm{IC}}\mathcal{L}_{\mathrm{IC}}(T_\phi)$ as in equation (4). The data term $\mathcal{L}_{\mathrm{data}}(T_\phi) = \|T_\phi - \hat{T}\|^2_{\Gamma_{\mathrm{obs}}}$ depends on $\theta$ only if $T_\phi$ does. Because $T_\phi$ is parameterized by $\phi$ alone, $\partial T_\phi/\partial\theta \equiv 0$, hence $\nabla_\theta\mathcal{L}_{\mathrm{data}} = 0$. The same argument applied to $\mathcal{L}_{\mathrm{IC}}(T_\phi)$ gives $\nabla_\theta\mathcal{L}_{\mathrm{IC}} = 0$. Therefore

$$\nabla_\theta\mathcal{L}_{\mathrm{PINN}} \;=\; \lambda_{\mathrm{PDE}}\,\nabla_\theta\mathcal{L}_{\mathrm{PDE}}(T_\phi, \alpha_\theta),$$

which establishes the gradient decoupling. To bound the residual-gradient magnitude, expand $\mathcal{L}_{\mathrm{PDE}}$ pointwise:

$$\mathcal{L}_{\mathrm{PDE}}(T_\phi, \alpha_\theta) \;=\; \frac{1}{N_c}\sum_{i=1}^{N_c}\big(r_i(T_\phi, \alpha_\theta)\big)^2, \qquad r_i \;=\; \partial_t T_\phi(\mathbf{x}_i, \tau_i) - \nabla\cdot(\alpha_\theta\nabla T_\phi)(\mathbf{x}_i, \tau_i),$$

where $\{(\mathbf{x}_i, \tau_i)\}_{i=1}^{N_c}$ are the PINN collocation points (using $\tau_i$ for collocation times to distinguish them from the measurement times $t_i$ of Section 3.2). Differentiating with respect to $\theta$ and applying the product rule to the divergence,

$$\nabla\cdot(\alpha_\theta\nabla T_\phi) \;=\; \alpha_\theta\,\Delta T_\phi \;+\; \nabla\alpha_\theta\cdot\nabla T_\phi,$$

so

$$\nabla_\theta r_i \;=\; -\Delta T_\phi(\mathbf{x}_i, \tau_i)\,\nabla_\theta\alpha_\theta(\mathbf{x}_i) \;-\; \nabla T_\phi(\mathbf{x}_i, \tau_i)\cdot\nabla_\theta\nabla\alpha_\theta(\mathbf{x}_i). \tag{19}$$

The Cauchy–Schwarz bound $\|\nabla_\theta\mathcal{L}_{\mathrm{PDE}}\| \le \frac{2}{N_c}\sum_i |r_i|\,\|\nabla_\theta r_i\|$ combined with (19) gives a residual-gradient magnitude controlled by the spatial derivatives $\Delta T_\phi$ and $\nabla T_\phi$ of the temperature surrogate at the collocation points, dominated by $\|\Delta T_\phi\|$ when $\nabla\alpha_\theta$ is bounded. Empirically and analytically, $\|\Delta T_\phi\|$ is large whenever $T_\phi$ has not yet converged, because second-order spatial derivatives of an under-fitted neural temperature field are noisy and large in magnitude.

The derivation establishes that surface observations reach $\theta$ only indirectly through the residual term, so the only avenue for $\hat{T}$ to influence $\alpha_\theta$ is via residual gradients that scale with the under-fitted Laplacian of $T_\phi$. The collapse of $\alpha_\theta$ to a near-trivial constant is the behavior observed for the four evaluated variants (Section 5). It is consistent with this channel and with PINN training analyses [23, 65], and it is the experimental complement to the present derivation.

## C.2 Coupling with the ill-posedness of Section 3.2

Even if the gradient decoupling derived in Appendix C.1 were absent, recovering $\alpha_\theta$ through any chain that begins at the surface data would require inverting the parameter-to-observation map $\mathcal{K}$. By Corollary 1, this inverse algebraically amplifies the $n$-th left singular vector by $\sigma_n^{-1} \ge C_B^{-1}(1+\lambda_n)^{1/2}$. In the two-network formulation, the surface data influences $\alpha_\theta$ only second-hand, by first updating $T_\phi$ through $\mathcal{L}_{\mathrm{data}}$ and then propagating to $\alpha_\theta$ through the residual term: the high-frequency information needed to localize defects is already attenuated by the smoothing of the forward map before the residual gradient sees it, and must then be re-amplified by an unbounded pseudo-inverse. This mechanism is consistent with the experimentally observed pattern in which $\mathcal{L}_{\mathrm{data}}$ converges (the temperature net fits the surface) but $\alpha_\theta$ remains near a trivial constant, the soft-constraint analogue of the data-fit paradox documented in our experimental study (Section 5) and in independent analyses of PINN training [23, 65].

### C.3 Standard remedies within the two-network formulation

Several SciML methods address documented PINN training pathologies: loss imbalance, collocation cost and residual variance, causality, and gradient conflict. The four evaluated here all keep the two-network parameterization, so we examine one question for each: whether it creates a path from $\hat{T}$ to $\theta$ that bypasses the residual term. In each case it does not, and the gradient decoupling continues to apply.

**GradNorm [8].** GradNorm rescales the loss weights $\lambda_{\mathrm{PDE}}, \lambda_{\mathrm{IC}}$ to balance gradient magnitudes during training. Because the rescaling acts on $\lambda_{\mathrm{PDE}}$ and not on the parameterization, the identity $\nabla_\theta \mathcal{L}_{\mathrm{PINN}} = \lambda_{\mathrm{PDE}} \nabla_\theta \mathcal{L}_{\mathrm{PDE}}$ is unchanged. GradNorm changes the relative size of the residual gradient but cannot manufacture a data-side gradient on $\theta$.

**SPINN [9].** SPINN replaces the joint $T_\phi(\mathbf{x}, t)$ network with a separable product $T_\phi(\mathbf{x}, t) = \sum_k T_\phi^{(1)}(x)\, T_\phi^{(2)}(y)\, T_\phi^{(3)}(z)\, T_\phi^{(4)}(t)$, accelerating collocation and reducing the variance of the residual term. Among the four evaluated variants, SPINN attains the highest aggregate SSIM (Table 1), a partial success of this parameterization. The diffusivity network $\alpha_\theta$ remains independent of $\phi$, so the gradient decoupling is unchanged.

**Causal-PINN [66].** Causal-PINN reweights the residual integrand by a causal weight that emphasizes early times before late times. The weight enters $\mathcal{L}_{\mathrm{PDE}}$ through a non-negative pointwise factor and again does not change the parameterization. The decoupling result holds.

**DCGD [28].** Dual-cone gradient descent projects the data and residual gradients onto a common acceptable cone to avoid destructive interference. It operates on already-computed gradients of $\mathcal{L}_{\mathrm{data}}$ and $\mathcal{L}_{\mathrm{PDE}}$ separately. Because $\nabla_\theta \mathcal{L}_{\mathrm{data}} = 0$ identically, the only gradient applied to $\theta$ remains the (projected) residual gradient.

In NeFTY the temperature is the implicit solution of a discretized PDE driven by $\alpha_\theta$, so the chain of dependencies is $\theta \to \alpha_\theta \to T \to \mathcal{L}_{\mathrm{data}}$ and the surface data term depends directly on $\theta$, with its gradient computed through the adjoint of the discrete heat operator (Section 4).

### C.4 A directly coupled soft variant

The analysis above requires exactly (i) the two-network parameterization with $T_\phi$ independent of $\theta$ and (ii) the data, residual, and initial-condition objective. Under these, $\partial T_\phi / \partial \theta \equiv 0$ (Appendix C.1). A soft-constrained architecture that lets surface data influence the diffusivity parameters directly falls outside this scope. We built such a variant to test whether a live data-to-$\theta$ path suffices on its own.

**Construction.** The temperature network is conditioned on the diffusivity network's scalar output through an input channel $c(\mathbf{x}, t) = w(t)\, \alpha_n(\mathbf{x})$, where $\alpha_n$ is the normalized bounded diffusivity and $w(t) = t/t_{\mathrm{end}}$ ramps from $w(0) = 0$, so the diffusivity-independent initial condition injects no gradient. Thus $\partial T / \partial \theta \neq 0$ for $t > 0$. The control is parameter-identical, with the channel zeroed.

**Verification of the coupling.** We verified the coupling directly: the surface-data-to-$\theta$ path is live, and a predefined virtual data-gradient perturbation induces a nonzero change in the predicted diffusivity field while that field is still interior to its parameterized range. Table 5 summarizes the paired comparison.

Table 5: **Directly coupled soft variant versus its parameter-identical control** (channel zeroed) on paired synthetic specimens.

| | CONTROL (CHANNEL ZEROED) | COUPLED |
|---|---|---|
| SURFACE-DATA-TO-$\theta$ GRADIENT | EXACTLY ZERO, EVERY LOGGED STEP | NONZERO, EVERY LOGGED STEP |
| MEAN-FIELD TRAJECTORY | REFERENCE | DIVERGES UP TO 1.97% PRE-COLLAPSE |
| FINAL FIELD | DEGENERATE CONSTANT AT LOWER BOUND | IDENTICAL; PER-SPECIMEN IOU DIFFERENCES EXACTLY ZERO |

**Outcome.** Each specimen's final IoU equals its own ground-truth defect volume fraction (arm mean $0.014$): the constant field at the lower bound marks every voxel as defect and recovers no structure.

For context, NeFTY reaches $0.158$ on the same synthetic specimens. On this problem class, the directly coupled architecture did not by itself close the gap under the matched training protocol.

**Interpretation and scope.** Direct coupling does not remove the underlying ill-posedness of Section 3.2, but this experiment does not identify ill-posedness as the cause of the coupled variant's result: optimization, local conditioning, the absent boundary-condition term, the absent TV prior, and gradient conflict all remain live explanations.

# D Method Details: Neural-Field Architecture, Forward Solver, and Adjoint Gradients

This appendix records the implementation of the three components introduced in Section 4: the neural diffusivity field of Section 4.1, the discrete forward solver of Section 4.2, and the discrete adjoint of Section 4.3. The harmonic-mean stencil derived in Appendix A (Proposition 1) and the boundary conditions described in Appendix A.5 are reused without further comment. Notation matches Section 4 throughout: $\theta$ denotes the MLP parameter set, $\boldsymbol{\mu}^n$ the discrete adjoint variable at step $n$ (kept distinct from the Laplacian eigenvalues $\lambda_n$ of Proposition 2 and from the regularization weight $\lambda$ in (9)), $\mathbf{\Pi}_\Gamma$ the observation projector onto $\Gamma_{\mathrm{obs}}$, and $\mathbf{A}(\alpha_\theta) := \mathbf{I} - \Delta t\, \mathbf{L}(\alpha_\theta)$ the implicit-Euler system matrix. The boldface discrete diffusion operator $\mathbf{L}(\alpha_\theta)$ is distinct from the italic lateral side length $L$ of Appendix A.5, the boldface Jacobi diagonal $\mathbf{D}$ and off-diagonal $\mathbf{R}$ of Appendix D.2 are distinct from the italic spatial dimension $D$ and the calligraphic regularizer $\mathcal{R}$ of (9), and the boldface state residual $\mathbf{F}^n$ is distinct from the italic source $F_{\delta\alpha}$ of Appendix B.1.

## D.1 Neural diffusivity-field architecture

**Coordinate MLP.** The diffusivity field is modeled as a fully connected network $f_\theta : \mathbb{R}^{d_\gamma} \to \mathbb{R}$ with depth $L_{\mathrm{MLP}}$, hidden width $W_{\mathrm{MLP}}$, ReLU activations [41], and a single skip connection that re-injects the encoded coordinate $\gamma(\mathbf{x})$ at the middle layer to preserve gradient flow [40]. Weights are initialized with Xavier scaling [20]. Concrete values for $L_{\mathrm{MLP}}, W_{\mathrm{MLP}}$, and the skip-layer index are tabulated in Appendix D.4.

**Positional encoding.** Each input coordinate $\mathbf{x} \in \Omega \subset \mathbb{R}^D$ is normalized to $[-1, 1]^D$ and mapped to

$$\gamma(\mathbf{x}) \;=\; \big(\sin(2^0\pi\mathbf{x}), \cos(2^0\pi\mathbf{x}), \ldots, \sin(2^{N-1}\pi\mathbf{x}), \cos(2^{N-1}\pi\mathbf{x})\big) \;\in\; \mathbb{R}^{2DN}, \tag{20}$$

yielding $d_\gamma = 2DN$. The bandwidth $N$ controls the maximum spatial frequency the network can express. Raising $N$ improves the network's ability to resolve sharp interfaces but exposes it to the algebraic noise amplification of Corollary 1 along high-eigenmode left singular vectors.

**Frequency annealing.** To suppress this amplification during the early phase of optimization, we apply the cosine schedule of Park et al. [45] to the encoding bands. Let $\beta(t) \in [0, N]$ be a training-step-dependent annealing parameter. The $k$-th band ($k = 0, \ldots, N-1$) of $\gamma(\mathbf{x})$ is multiplied by

$$w_k(\beta) \;=\; \tfrac{1}{2}\big[1 - \cos\big(\pi\, \mathrm{clamp}(\beta - k, 0, 1)\big)\big]\,, \tag{21}$$

so that $w_k = 0$ before the schedule reaches band $k$, ramps smoothly to $w_k = 1$ over a unit-width interval in $\beta$, and remains at one thereafter. We linearly increase $\beta$ from 0 to $N$ over the first $T_{\mathrm{FA}}$ optimization steps (frequency-annealing horizon, value in Appendix D.4), after which all bands are fully unmasked. Compared with binary masking, the cosine schedule avoids abrupt gradient bursts when a new band becomes active and lets the network grow high-frequency detail only after the low-frequency reconstruction has stabilized.

**Bounded output.** The MLP output is hard-bracketed to a physically admissible range,

$$\alpha_\theta(\mathbf{x}) \;=\; \alpha_{\min} + (\alpha_{\max} - \alpha_{\min})\, \sigma\big(f_\theta(\gamma(\mathbf{x}))\big),$$

with $\sigma(\cdot)$ the logistic sigmoid and bounds $0 < \alpha_{\min} < \alpha_{\max}$. The bracketing serves three purposes: (i) it enforces $\alpha_\theta(\mathbf{x}) > 0$ identically, satisfying the positivity required by the diffusivity form (2); (ii) it bounds the diffusivity contrast so that the system matrix $\mathbf{A}(\alpha_\theta)$ remains well-conditioned for the Jacobi inner solver of Appendix D.2; (iii) it prevents the early-training optimizer from probing extreme values that would yield zero gradients through the saturation of the sigmoid or through stiff numerical regimes. Concrete numerical values for $\alpha_{\min}, \alpha_{\max}$ are calibrated once per dataset using the bulk-to-defect contrast described in Appendix A.5, and are reported in Appendix D.4.

**Discrete total-variation regularizer.** The continuous regularizer $\mathcal{R}(\alpha_\theta) = \int_\Omega \|\nabla \alpha_\theta(\mathbf{x})\|\, d\mathbf{x}$ is implemented on the same Cartesian grid used by the forward solver, with per-axis dimensions $N_x, N_y, N_z$ and spacings $\Delta x, \Delta y, \Delta z$ formally defined in Appendix D.2. With $\alpha_{i,j,k} := \alpha_\theta(\mathbf{x}_{i,j,k})$ at grid node $(i, j, k)$, we use the isotropic forward-difference discretization

$$\mathcal{R}(\alpha_\theta) \;=\; \Delta x\,\Delta y\,\Delta z \sum_{i,j,k=0}^{N_x-1,\,N_y-1,\,N_z-1} \Bigg[ \big(\delta_x \alpha_{i,j,k}/\Delta x\big)^2 + \big(\delta_y \alpha_{i,j,k}/\Delta y\big)^2 + \big(\delta_z \alpha_{i,j,k}/\Delta z\big)^2 + \epsilon_{\mathrm{TV}}^2 \Bigg]^{1/2}, \quad (22)$$

where the per-axis forward differences are $\delta_x \alpha_{i,j,k} := \alpha_{(i+1) \bmod N_x, j, k} - \alpha_{i,j,k}$, $\delta_y \alpha_{i,j,k} := \alpha_{i,(j+1) \bmod N_y, k} - \alpha_{i,j,k}$ (lateral periodic wrap, matching the lateral periodic BCs of Appendix A.5), and $\delta_z \alpha_{i,j,k} := \alpha_{i,j,k+1} - \alpha_{i,j,k}$ for $k < N_z - 1$ with $\delta_z \alpha_{i,j,N_z-1} := 0$ (the adiabatic and Robin BCs do not constrain $\alpha$ at the front/back faces, so the through-thickness forward difference at the top slice vanishes, while the lateral $\delta_x, \delta_y$ contributions on that slice are retained). The smoothing constant $\epsilon_{\mathrm{TV}} > 0$ has dimension $[\alpha]/[\mathrm{length}]$ and ensures differentiability at vanishing gradients (concrete value in Appendix D.4, distinct from the perturbation parameter $\varepsilon$ used in Appendix B.1). The penalty promotes piecewise-constant $\alpha_\theta$ aligned with the model assumed in Appendix A.2 and supplements the high-eigenmode suppression provided by frequency annealing.

### D.2 Discrete forward solver

**Ambient shift.** Throughout this appendix and Section 4, the discrete temperature field $\mathbf{T}^n$ denotes the deviation from the ambient temperature $T_{\mathrm{amb}}$ defined in Appendix A.5. Under this convention the synthetic post-flash and PVC near-uniform initial conditions of Appendix A.5 have ambient floor zero and the convective Robin condition $\mathbf{n} \cdot (\alpha \nabla T) = -h(T - T_{\mathrm{amb}})$ becomes the homogeneous form $\mathbf{n} \cdot (\alpha \nabla \mathbf{T}) = -h\,\mathbf{T}$. All measured frames $\hat{\mathbf{T}}_i$ in Section 4.3 are likewise interpreted as ambient-shifted observations $\hat{T}_i - T_{\mathrm{amb}}$. The state equation (8) therefore retains the homogeneous form $\mathbf{A}(\alpha_\theta)\mathbf{T}^{n+1} = \mathbf{T}^n$ across both datasets, with no affine right-hand side.

**Spatial discretization.** The bulk equation $\partial_t T = \nabla \cdot (\alpha \nabla T)$ on the slab $\Omega = [0, L]^2 \times [0, H]$ is discretized on a uniform Cartesian grid with $N_x \times N_y \times N_z$ nodes and spacings $\Delta x, \Delta y, \Delta z$ aligned with the lateral and through-thickness axes. Continuing the one-dimensional argument of Proposition 1, the second-order finite-volume diffusion operator $\mathbf{L}(\alpha_\theta) \in \mathbb{R}^{N_g \times N_g}$ acts at an interior node $(i, j, k)$ by

$$\begin{aligned}
[\mathbf{L}(\alpha_\theta)\mathbf{T}]_{i,j,k} \;&=\; \tfrac{1}{\Delta x^2}\big[\bar{\alpha}_{i+1/2,j,k}(T_{i+1,j,k} - T_{i,j,k}) - \bar{\alpha}_{i-1/2,j,k}(T_{i,j,k} - T_{i-1,j,k})\big] \\
&+\; \tfrac{1}{\Delta y^2}\big[\bar{\alpha}_{i,j+1/2,k}(T_{i,j+1,k} - T_{i,j,k}) - \bar{\alpha}_{i,j-1/2,k}(T_{i,j,k} - T_{i,j-1,k})\big] \\
&+\; \tfrac{1}{\Delta z^2}\big[\bar{\alpha}_{i,j,k+1/2}(T_{i,j,k+1} - T_{i,j,k}) - \bar{\alpha}_{i,j,k-1/2}(T_{i,j,k} - T_{i,j,k-1})\big],
\end{aligned}$$

with each face-centered coefficient $\bar{\alpha}$ computed as the harmonic mean of its two neighbouring cell-centered values. By Proposition 1 this is the unique discrete realization of effective-flux continuity for the single-$\alpha$ model.

**Boundary conditions.** Lateral periodicity is implemented by circular padding so that the stencil at $i = 0$ wraps to $i = N_x - 1$ and similarly for $j$. Through-thickness adiabatic conditions (zero-flux Neumann) are implemented by replicate padding on the front $z = 0$ and back $z = H$ faces, which enforces $\bar{\alpha}_{i,j,-1/2}(T_{i,j,0} - T_{i,j,-1}) = 0$ at the boundary face by zeroing the cross-face temperature increment. The PVC convective Robin condition described in Appendix A.5 replaces the back-face replicate padding by a ghost-cell update consistent with $\mathbf{n} \cdot (\alpha \nabla T) = -h(T - T_{\mathrm{amb}})$.

**Implicit-Euler step.** The state at frame $n + 1$ is the solution of the sparse linear system

$$\mathbf{A}(\alpha_\theta)\,\mathbf{T}^{n+1} \;=\; \mathbf{T}^n, \qquad \mathbf{A}(\alpha_\theta) \;:=\; \mathbf{I} - \Delta t\,\mathbf{L}(\alpha_\theta), \quad (23)$$

which is unconditionally stable [12] so that $\Delta t$ is set by the camera frame rate rather than the CFL bound $\Delta t \le \Delta x^2/(2D\,\alpha_{\max})$.

**Symmetry, sign, and diagonal dominance.** Under the homogeneous boundary conditions enforced by the ambient shift above (lateral periodic, adiabatic through-thickness, and the homogeneous

Robin form $\mathbf{n} \cdot (\alpha \nabla \mathbf{T}) = -h\,\mathbf{T}$ on the PVC back face), the harmonic-mean construction makes $\mathbf{L}(\alpha_\theta)$ symmetric: each interior face conductance $a_{pq} := \bar{\alpha}_{pq}/\Delta x_d^2$ ($d \in \{x, y, z\}$ the axis joining grid nodes $p, q$) appears symmetrically in row $p$ and row $q$, lateral periodic wrapping preserves this symmetry, and a finite-volume back-face flux balance $\mathbf{n} \cdot (\alpha \nabla \mathbf{T}) \cdot \Delta x \Delta y = -h\,T_p \cdot \Delta x \Delta y$ divided by the nodal volume $\Delta x \Delta y \Delta z$ yields a strictly positive diagonal sink $\beta_p := h/\Delta z$ on each back-face Robin row. Writing $q \sim p$ for the set of interior-face neighbours of node $p$ and $\beta_p \geq 0$ for the per-row Robin sink (zero for non-Robin rows), the entries of $\mathbf{L}(\alpha_\theta)$ are

$$[\mathbf{L}(\alpha_\theta)]_{pp} \;=\; -\beta_p - \sum_{q \sim p} a_{pq}, \qquad [\mathbf{L}(\alpha_\theta)]_{pq} \;=\; a_{pq} \geq 0 \;\text{ for }\; q \sim p,$$

$$\sum_q [\mathbf{L}(\alpha_\theta)]_{pq} \;=\; -\beta_p \;\leq\; 0,$$

so the constant-temperature mode is annihilated on rows without a Robin sink and damped at rate $\beta_p$ on Robin boundary rows. Therefore $\mathbf{L}(\alpha_\theta)$ is symmetric and negative semi-definite under purely periodic/adiabatic BCs, and negative definite once at least one $\beta_p > 0$. Strict diagonal dominance of $\mathbf{A}(\alpha_\theta)$ follows from the identity term and the (non-negative) Robin sink:

$$[\mathbf{A}(\alpha_\theta)]_{pp} - \sum_{q \neq p} \big|[\mathbf{A}(\alpha_\theta)]_{pq}\big| \;=\; \big(1 + \Delta t \beta_p + \Delta t \sum_{q \sim p} a_{pq}\big) - \Delta t \sum_{q \sim p} a_{pq} \;=\; 1 + \Delta t \beta_p \;\geq\; 1 \;>\; 0,$$

holding strictly on every row independently of the conductance pattern, so $\mathbf{A}(\alpha_\theta)$ is symmetric positive-definite and the Jacobi iteration below converges geometrically.

**Jacobi inner solver.** For autodiff compatibility on the GPU we avoid sparse-matrix factorizations and solve (23) by a fixed unrolled number $K$ of Jacobi iterations. With the splitting $\mathbf{A}(\alpha_\theta) = \mathbf{D} + \mathbf{R}$ where $\mathbf{D}$ is the diagonal of $\mathbf{A}$ and $\mathbf{R}$ the off-diagonal remainder, the iteration with right-hand side $\mathbf{b} := \mathbf{T}^n$ is

$$\mathbf{T}^{(\kappa+1)} \;=\; \mathbf{D}^{-1}\Big(\mathbf{b} - \mathbf{R}\,\mathbf{T}^{(\kappa)}\Big), \qquad \kappa = 0, 1, \ldots, K-1, \tag{24}$$

initialized from $\mathbf{T}^{(0)} := \mathbf{T}^n$ (warm-started by the previous frame's solution). Equation (24) reduces to a 7-point stencil convolution, expressible as a single fused kernel in standard autodiff frameworks [46] and accelerated with `torch.compile` just-in-time autotuning. The iteration count $K$ is chosen large enough that the residual $\|\mathbf{A}\mathbf{T}^{(K)} - \mathbf{b}\|$ falls below the noise floor of the camera measurements (Appendix D.4). Empirically, $K = 50$ suffices for our grid resolution and $\Delta t$. Because the Jacobi iteration converges to the unique exact solution of (23) under the strict diagonal dominance established above, we treat $\mathbf{T}^{(K)}$ as the (numerical) exact state $\mathbf{T}^{n+1}$, and the discrete adjoint of Appendix D.3 below is derived for this exact linear solve. The resulting parameter gradients are exact for the converged objective up to the same iterative tolerance.

### D.3 Adjoint recurrence and gradient assembly

**Constraint and augmented Lagrangian.** Set $\mathbf{T}^0$ from the prescribed initial condition (Appendix A.5) and define the state residual at step $n$ by

$$\mathbf{F}^n(\mathbf{T}^n, \mathbf{T}^{n-1}, \alpha_\theta) \;:=\; \mathbf{A}(\alpha_\theta)\,\mathbf{T}^n - \mathbf{T}^{n-1} \;=\; \mathbf{0}, \qquad n = 1, \ldots, N_t. \tag{25}$$

Define the per-step data term

$$\ell^n(\mathbf{T}^n) \;:=\; \begin{cases} \big\|\mathbf{\Pi}_\Gamma \mathbf{T}^n - \hat{\mathbf{T}}_i\big\|^2 & \text{if } n = n_i \in \{n_1, \ldots, n_M\}, \\ 0 & \text{otherwise,} \end{cases}$$

so the data part of the objective is $\sum_{n=1}^{N_t} \ell^n(\mathbf{T}^n)$ and matches the data-fidelity sum in (9). The squared Euclidean data term in (9) differs from the discrete squared $L^2(\Gamma_{\text{obs}})$ term of (3) by the lateral quadrature factor $\Delta x \Delta y$, and rescaling the objective absorbs this factor into the TV weight $\lambda$. We introduce adjoint variables $\boldsymbol{\mu}^n \in \mathbb{R}^{N_g}$ for $n = 1, \ldots, N_t$ as Lagrange multipliers of the state residuals, giving the augmented functional

$$\widetilde{\mathcal{L}}(\theta; \{\mathbf{T}^n\}, \{\boldsymbol{\mu}^n\}) \;=\; \sum_{n=1}^{N_t} \ell^n(\mathbf{T}^n) + \lambda \mathcal{R}(\alpha_\theta) - \sum_{n=1}^{N_t} (\boldsymbol{\mu}^n)^\top \mathbf{F}^n(\mathbf{T}^n, \mathbf{T}^{n-1}, \alpha_\theta).$$

For any $\{\mathbf{T}^n\}$ that satisfy the state equation (25), $\widetilde{\mathcal{L}} = \mathcal{J}_{\mathrm{disc}}(\theta)$ regardless of the multipliers $\boldsymbol{\mu}^n$.

**Stationarity and adjoint recurrence.** Computing $d\mathcal{J}_{\mathrm{disc}}/d\theta$ via the chain rule and discarding terms that are eliminated by the state equation, it suffices to compute $d\widetilde{\mathcal{L}}/d\theta$ in the way that does not require the implicit derivatives $d\mathbf{T}^n/d\theta$. To eliminate them, we choose the multipliers so that the partial derivative of $\widetilde{\mathcal{L}}$ with respect to each $\mathbf{T}^n$ vanishes. Differentiating,

$$\frac{\partial \widetilde{\mathcal{L}}}{\partial \mathbf{T}^n} \;=\; \frac{\partial \ell^n}{\partial \mathbf{T}^n} - (\boldsymbol{\mu}^n)^\top \frac{\partial \mathbf{F}^n}{\partial \mathbf{T}^n} - (\boldsymbol{\mu}^{n+1})^\top \frac{\partial \mathbf{F}^{n+1}}{\partial \mathbf{T}^n} \;=\; 0,$$

where we set $\boldsymbol{\mu}^{N_t+1} = \mathbf{0}$ because no constraint $\mathbf{F}^{N_t+1}$ exists. Reading off the partials from (25),

$$\frac{\partial \mathbf{F}^n}{\partial \mathbf{T}^n} \;=\; \mathbf{A}(\alpha_\theta), \qquad \frac{\partial \mathbf{F}^{n+1}}{\partial \mathbf{T}^n} \;=\; -\mathbf{I},$$

the stationarity condition becomes

$$\mathbf{A}(\alpha_\theta)^\top \boldsymbol{\mu}^n \;=\; \left(\partial \ell^n / \partial \mathbf{T}^n\right)^\top + \boldsymbol{\mu}^{n+1}, \qquad \boldsymbol{\mu}^{N_t+1} \;=\; \mathbf{0}, \tag{26}$$

which is (10). The recurrence is integrated backwards in time from $n = N_t$ down to $n = 1$. The same Jacobi inner solver of Appendix D.2 applies because the finite-volume discretization with diagonal Robin sink $\beta_p := h/\Delta z$ makes $\mathbf{A}(\alpha_\theta)^\top = \mathbf{A}(\alpha_\theta)$ on every dataset (lateral periodic, adiabatic through-thickness, and homogeneous Robin), so the forward and backward sweeps share the same SPD system matrix.

**Gradient assembly.** With $\boldsymbol{\mu}^n$ chosen so that $\partial\widetilde{\mathcal{L}}/\partial\mathbf{T}^n = 0$ for $n = 1, \ldots, N_t$, the total derivative of $\mathcal{J}_{\mathrm{disc}}$ with respect to $\theta$ is the partial derivative of $\widetilde{\mathcal{L}}$ with respect to $\theta$ at fixed $\{\mathbf{T}^n\}$:

$$\frac{d\mathcal{J}_{\mathrm{disc}}}{d\theta} \;=\; \frac{\partial \widetilde{\mathcal{L}}}{\partial \theta} \;=\; \frac{\partial(\lambda \mathcal{R})}{\partial \theta} - \sum_{n=1}^{N_t} (\boldsymbol{\mu}^n)^\top \frac{\partial \mathbf{F}^n}{\partial \alpha_\theta} \frac{\partial \alpha_\theta}{\partial \theta}.$$

Because $\mathbf{F}^n = \mathbf{A}(\alpha_\theta)\mathbf{T}^n - \mathbf{T}^{n-1}$ and $\mathbf{A}(\alpha_\theta) = \mathbf{I} - \Delta t\, \mathbf{L}(\alpha_\theta)$ depends on $\alpha_\theta$ only through the diffusion operator,

$$\frac{\partial \mathbf{F}^n}{\partial \alpha_\theta} \;=\; -\Delta t\, \frac{\partial(\mathbf{L}(\alpha_\theta)\mathbf{T}^n)}{\partial \alpha_\theta},$$

so the assembled gradient becomes

$$\frac{d\mathcal{J}_{\mathrm{disc}}}{d\theta} \;=\; \frac{\partial(\lambda \mathcal{R})}{\partial \theta} \;+\; \Delta t \sum_{n=1}^{N_t} (\boldsymbol{\mu}^n)^\top \frac{\partial(\mathbf{L}(\alpha_\theta)\mathbf{T}^n)}{\partial \alpha_\theta} \frac{\partial \alpha_\theta}{\partial \theta}, \tag{27}$$

which is (11). The Jacobian $\partial\alpha_\theta/\partial\theta$ is obtained by ordinary backpropagation through the MLP at each grid node, and the term $\partial(\mathbf{L}(\alpha_\theta)\mathbf{T}^n)/\partial\alpha_\theta$ is a sparse local stencil whose nonzeros are determined by the harmonic-mean derivatives $\partial\bar{\alpha}_{i+1/2}/\partial\alpha_i = 2\alpha_{i+1}^2/(\alpha_i + \alpha_{i+1})^2$ and similarly for $\partial/\partial\alpha_{i+1}$.

**Memory cost.** The forward sweep stores $\mathbf{T}^1, \ldots, \mathbf{T}^{N_t}$, about one field per frame, which the gradient assembly (27) consumes one frame at a time during the backward sweep that produces $\boldsymbol{\mu}^n$. Because (26) requires only $\boldsymbol{\mu}^{n+1}$ and $\mathbf{T}^n$ at any given step, the adjoint's solver working set (one current and one trailing slice plus the running gradient accumulator) is $\mathcal{O}(N_g)$, independent of $N_t$. The naive backpropagation-through-time alternative would store the entire computational graph of the $K$ unrolled Jacobi iterations at every one of the $N_t$ time steps, with peak memory scaling as $\mathcal{O}(KN_gN_t)$, whereas the adjoint replaces these retained iterates with the $\mathcal{O}(N_g)$ solver working set plus the stored states, at the cost of one backward solve, an exact restatement of the implicit-function-theorem trick in a sequential setting [7].

**Sensitivity to initial conditions.** If the initial temperature $\mathbf{T}^0$ is itself uncertain or learnable, the same calculation under the augmented Lagrangian above (which carries a $-\sum_n (\boldsymbol{\mu}^n)^\top \mathbf{F}^n$ term) gives

$$\frac{d\mathcal{J}_{\mathrm{disc}}}{d\mathbf{T}^0} \;=\; -(\boldsymbol{\mu}^1)^\top \frac{\partial \mathbf{F}^1}{\partial \mathbf{T}^0} \;=\; -(\boldsymbol{\mu}^1)^\top(-\mathbf{I}) \;=\; +\boldsymbol{\mu}^1,$$

where only $\mathbf{F}^1 = \mathbf{A}\mathbf{T}^1 - \mathbf{T}^0$ depends on $\mathbf{T}^0$. The sign is consistent with the adjoint recurrence (26) and the gradient assembly (27) under the same Lagrangian convention. The experiments of Section 5 keep $\mathbf{T}^0$ fixed at the post-flash initial profile calibrated externally per Appendix A.5.

**Consistency with the continuous adjoint.** The recurrence (26) is the implicit-Euler discretization, applied backwards in time, of the continuous adjoint problem associated with the bulk equation $\partial_t T = \nabla \cdot (\alpha \nabla T)$ and the cumulative observation functional $\sum_i \|\mathbf{\Pi}_\Gamma T(\cdot, t_i) - \hat{T}_i\|^2$: namely, $-\partial_t p = \nabla \cdot (\alpha \nabla p) + \sum_i 2\mathbf{\Pi}_\Gamma^*(\mathbf{\Pi}_\Gamma T(\cdot, t_i) - \hat{T}_i)\,\delta(t - t_i)$ with terminal condition $p(\cdot, t_{\mathrm{end}}) = 0$. The discretize-then-optimize formulation we use thus inherits the same backward-parabolic structure as the continuous setting [43], while delivering exact gradients of the discrete objective $\mathcal{J}_{\mathrm{disc}}$ rather than discretized gradients of the continuous one.

### D.4 Implementation hyperparameters

The hyperparameters used in the synthetic experiments of Section 5 are listed in Table 6. Adam is used with the listed learning rate and step decay. Gradients are accumulated through a single forward and a single backward solver sweep per iteration. Compute resources, including the GPUs used for benchmarking, are documented in Appendix E.

Table 6: Hyperparameter configuration for NeFTY on the synthetic benchmark.

| Category | Parameter | Value |
|---|---|---|
| **Network architecture** | MLP depth $L_{\mathrm{MLP}}$ | 10 |
| | MLP hidden width $W_{\mathrm{MLP}}$ | 512 |
| | Encoding bandwidth $N$ | 12 |
| | Skip-connection layer | 4 |
| | Output activation | scaled sigmoid |
| **Domain and grid** | Domain size $(L, L, H)$ | $10.0 \times 10.0 \times 1.0$ |
| | Grid resolution $(N_x, N_y, N_z)$ | $64 \times 64 \times 16$ |
| | Lateral spacing $\Delta x = \Delta y$ | 0.156 |
| | Through-thickness spacing $\Delta z$ | 0.0625 |
| **Forward solver** | Time step $\Delta t$ | 0.05 |
| | Number of frames $N_t$ | 100 |
| | Time integrator | implicit Euler |
| | Inner solver | Jacobi unrolling |
| | Jacobi iterations $K$ | 50 |
| | Face coefficient | harmonic mean (Prop. 1) |
| | Diffusivity bounds $(\alpha_{\min}, \alpha_{\max})$ | $(0.003,\ 0.25)$ |
| **Optimization** | Optimizer | Adam |
| | Learning rate | $5 \times 10^{-5}$ |
| | Step decay $\gamma$ | 0.1 per 1000 steps |
| | Total iterations | 10,000 |
| | Frequency-annealing horizon $T_{\mathrm{FA}}$ | 2,500 |
| **Regularization** | TV weight $\lambda$ | $1 \times 10^{-3}$ |
| | TV smoothing $\epsilon$ | $1 \times 10^{-6}$ |

## E Experimental Details

### E.1 Synthetic dataset generation

**Inverse-crime mitigation.** To avoid evaluating the reconstruction against data generated by its own discretization, we use PhiFlow [26], an independent finite-volume engine, for ground-truth generation. PhiFlow integrates the bulk equation (2) with an explicit scheme. To overcome the explicit stability constraint at fine grid resolution, we apply an adaptive substepping routine. The maximum stable step size is

$$\Delta t_{\mathrm{stable}} \approx \frac{\Delta x^2}{2D\,\alpha_{\max}},$$

with $D \in \{2, 3\}$ the spatial dimension and $\alpha_{\max}$ the per-batch maximum diffusivity, and the number of solver substeps per recorded frame is set to

$$N_{\text{sub}} \;=\; \max\!\left(10,\; \left\lceil \frac{\Delta t}{\Delta t_{\text{stable}}} \times 2.0 \right\rceil\right),$$

so that the forward simulation often executes dozens of substeps per recorded frame.

**Simulation configuration.** All synthetic samples are generated on the unitless slab $\Omega = [0, 10]^2 \times [0, 1]$ at $64 \times 64 \times 16$ grid resolution. The lateral $(x, y)$ faces use periodic boundary conditions and the through-thickness $(z)$ faces use adiabatic Neumann conditions, both consistent with Appendix A.5. The transient is recorded over 100 frames with $\Delta t = 0.05$. Material properties are sampled from uniform distributions: bulk diffusivity $\alpha_{\text{base}} \sim \mathcal{U}(0.1, 0.2)$, defect diffusivity $\alpha_{\text{defect}} \sim \mathcal{U}(0.005, 0.015)$. Each sample contains one to four ellipsoidal, cylindrical, or box-shaped subsurface defects buried at varying depth. The dataset of 1,000 samples is split into a homogeneous configuration (constant $\alpha_{\text{base}}$) and a layered configuration (three to four bulk strata along $z$), and we evaluate every method on the full 1,000-sample benchmark per configuration.

**Dimensional analysis.** The unitless solver evolves in $(t/t_{\text{total}}, \mathbf{x}/L_0)$ coordinates with effective diffusivity $\alpha_{\text{sim}} = \alpha_{\text{phys}}\, t_{\text{total}}/L_0^2$, the Fourier-number scaling of Appendix A.5. A simulation with $\alpha_{\text{sim}} \approx 0.1$ corresponds to highly conductive silicon ($\alpha_{\text{phys}} \approx 10^{-4}\,\text{m}^2/\text{s}$) on a microsecond timescale at $L_0 = 10\,\mu\text{m}$, or to a resistive polymer ($\alpha_{\text{phys}} \approx 10^{-7}\,\text{m}^2/\text{s}$) on a millisecond timescale at the same length scale. A single set of unitless parameters thus represents widely different material classes by reinterpreting $L_0$ and $t_{\text{total}}$.

**Defect contrast scaling.** The defect-to-bulk diffusivity ratio is set to roughly $1:20$. We choose this contrast over the realistic air-to-solid ratio of $> 1:1000$ for numerical stability: at $1:1000$ the linear system $\mathbf{A}(\alpha_\theta) = \mathbf{I} - \Delta t\, \mathbf{L}(\alpha_\theta)$ becomes severely ill-conditioned and the iterative solver stalls. At $1:20$ the surface-temperature signature already saturates against the perfect-insulator limit (further reductions in $\alpha_{\text{defect}}$ leave the boundary thermograms unchanged), while the condition number remains compatible with efficient gradient-based optimization.

### E.2 Metrics

**Synthetic 3D reconstruction.** For each test sample, let $\alpha^\star$ denote the ground-truth diffusivity field on the $N_g = N_x N_y N_z$ grid and $\hat{\alpha}$ the recovered field. We report:

- *MSE* $= \frac{1}{N_g}\|\hat{\alpha} - \alpha^\star\|_2^2$.
- *PSNR* $= 10 \log_{10}\big((\alpha_{\max} - \alpha_{\min})^2/\text{MSE}\big)$.
- *SSIM* [67] computed slice-by-slice along $z$ and averaged.
- *IoU* $= |\,(\hat{\alpha} < \tau) \cap (\alpha^\star < \tau)\,| / |\,(\hat{\alpha} < \tau) \cup (\alpha^\star < \tau)\,|$ with defect threshold $\tau = 0.03$, chosen as twice the upper end of the defect distribution to provide a clean binary defect mask.

The metrics are computed per sample and averaged over the full 1,000-sample benchmark per configuration.

**Real PVC: 2D segmentation.** Following the convention of the PVC-Infrared benchmark [69], the recovered 3D $\hat{\alpha}_\theta$ is projected onto a 2D defect mask through the depth-averaged anomaly contrast described in Appendix H, and IoU and Dice are computed against the 2D ground-truth annotation.

**Real PVC: 2.5D depth.** For each pixel $p$ assigned to a defect, let $d_p^\star$ denote the ground-truth depth (in mm) supplied by the PVC-Depth benchmark [68] and $\hat{d}_p$ the predicted depth from the spatial-median anomaly contrast along $z$ (Appendix H). Let $P$ denote the set of pixels assigned to a defect by the 2D mask above and $\delta_p := \max(\hat{d}_p/d_p^\star, d_p^\star/\hat{d}_p)$ the per-pixel depth ratio. We report the standard depth metrics of Eigen et al. [15]: Absolute Relative error $\text{Abs Rel} = \frac{1}{|P|}\sum_{p\in P} |\hat{d}_p - d_p^\star|/d_p^\star$, RMSE $= \big(\frac{1}{|P|}\sum_{p\in P}(\hat{d}_p - d_p^\star)^2\big)^{1/2}$, and the threshold accuracies $\delta < 1.25^k := \frac{1}{|P|}|\{p \in P : \delta_p < 1.25^k\}|$ for $k \in \{1, 2, 3\}$.

**Frequency-domain diagnostics.** *Defect Edge F1* thresholds the gradient magnitude $\|\nabla\alpha\|$ at $50\%$ of the ground-truth maximum, restricts to a dilated defect-boundary mask, and reports the harmonic

mean of precision and recall. *Radial power spectrum* extracts a tight $16 \times 16 \times 8$ crop centered on the defect cluster, computes the 3D Fourier transform, and radially averages the power, with frequencies normalized so that the through-thickness Nyquist limit equals $\sim 8$ cycles per unit length.

### E.3 Hyperparameters and compute resources

The full NeFTY hyperparameter table (architecture, solver, optimization, regularization) is given in Table 6 of Appendix D.4. Our experimental framework was executed on a hybrid infrastructure comprising local workstations for controlled benchmarking and a high-performance computing (HPC) cluster for large-scale training. The local development environment consists of servers equipped with a 32-core CPU and two NVIDIA RTX PRO 6000 Blackwell GPUs (96 GB VRAM each). For consistency across the efficiency analysis, all hardware-sensitive metrics, specifically the wall-clock times and peak GPU memory usage detailed in Table 4 of the main body and Table 10 of this appendix, were benchmarked exclusively on this local server with reconstruction confined to a single GPU per specimen. For the large-scale synthetic dataset generation under PhiFlow, we utilized a compute cluster where each node is provisioned with dual 26-core CPUs and eight NVIDIA L40 GPUs.

## F Baseline Implementations

This appendix gives the configuration of every baseline used in Section 5. Unless otherwise stated, all baselines are tuned with the same compute budget as NeFTY on the same hardware (Appendix E.3).

### F.1 Soft-constrained PINN family

**PINN with GradNorm.** The vanilla PINN baseline instantiates two networks, a temperature surrogate $T_\phi(\mathbf{x}, t)$ and a diffusivity surrogate $\alpha_\theta(\mathbf{x})$, and minimizes the composite loss of equation (4). The PDE residual is evaluated through automatic differentiation at $N_c$ space-time collocation points sampled uniformly per iteration, and the initial-condition penalty matches $T_\phi(\mathbf{x}, 0)$ to the prescribed Gaussian post-flash profile of Appendix A.5. The two physics weights $\lambda_{\mathrm{PDE}}$ and $\lambda_{\mathrm{IC}}$ are dynamically balanced through GradNorm [8], which normalizes per-loss gradient magnitudes to a common scale. We use $T_\phi$ depth 5 and width 128 (4.91 M total parameters across both networks), $N_c = 24{,}576$ collocation points per iteration, Adam with learning rate $10^{-3}$, and 22,000 iterations: $2.2\times$ more iterations and $2\times$ more parameters than NeFTY. Despite this advantage, the recovered $\alpha_\theta$ saturates at a near-trivial constant (Section 5.2), in line with the structural decoupling proven in Appendix C.

**SPINN, Causal-PINN, DCGD.** For the three modern PINN variants we follow the official implementations and hyperparameter settings released by the original authors [9, 28, 66], paired with the same diffusivity-network $\alpha_\theta$ and the same 24,576 collocation points and 22,000 iterations as the vanilla PINN above so that compute is matched across the family.

### F.2 Voxel-grid optimization

The Grid Opt. baseline replaces the neural diffusivity field $\alpha_\theta$ of Section 4.1 by a learnable tensor $\boldsymbol{\alpha} \in \mathbb{R}^{N_x \times N_y \times N_z}$ (initialized at the bulk mean) and optimizes it through the same differentiable solver and adjoint of Section 4. This isolates the contribution of the continuous coordinate-based prior. Grid Opt. uses the same Adam learning-rate schedule, the same number of iterations (10,000), and the same TV regularizer as NeFTY. Without the neural prior, it converges to a noisy field with ringing artifacts (Section 5.2).

### F.3 Supervised U-Net baselines

**U-Net (Full).** A 3D U-Net [53] adapted for spatiotemporal regression. The input temperature sequence of shape $(B, 1, 100, 64, 64)$ is interpolated along the temporal axis from 100 to 16 depth slices and treated as a $(B, 1, 16, 64, 64)$ volume. The architecture follows a four-level encoder-decoder with channel sizes $\{32, 64, 128, 256\}$, double 3D convolutions and max-pooling on the contracting path, and trilinear upsampling with skip connections on the expansive path. The output passes through a sigmoid scaled to $[\alpha_{\min}, \alpha_{\max}]$. Training uses MSE between predicted and ground-truth $\alpha$ on the full synthetic dataset (defects included), which represents an upper bound consuming the volumetric labels that the label-free baselines never see.

**U-Net (Sound-Only).** The same architecture trained exclusively on defect-free samples. This baseline probes the susceptibility of purely data-driven inversion to out-of-distribution defects, and is included in the main Table 1 as a data-driven zero-shot reference.

### F.4 Classical thermography heuristics

**PPT.** Pulsed Phase Thermography [29, 38] computes the per-pixel discrete Fourier transform of the surface-temperature decay, identifies the dominant phase peak at frequency $f_{\mathrm{peak}}$, and inverts depth through the diffusion-length formula $d = 1.82\,\sqrt{\alpha_{\mathrm{phys}}/(\pi f_{\mathrm{peak}})}$. We use the implementation of the original authors with six target frequencies spanning $0.01$-$0.5$ Hz, matching the PVC-Depth recording rate.

**TSR.** Thermographic Signal Reconstruction [58, 59] fits a polynomial in $\log t$ to each pixel's temperature decay, identifies the time $t_{\mathrm{peak}}$ at which the second log-derivative peaks, and inverts depth as $d = 1.8\,\sqrt{\alpha_{\mathrm{phys}}\, t_{\mathrm{peak}}}$. We use the standard fifth-order polynomial fit and the per-pixel $t_{\mathrm{peak}}$ extraction.

### F.5 Swin-UNETR (real PVC, zero-shot transfer)

The supervised Swin-UNETR baseline of Hatamizadeh et al. [24] is trained on the synthetic dataset of Appendix E.1 for $500$ epochs under MSE supervision (loss converges to $4 \times 10^{-6}$) and transferred zero-shot to the real PVC datasets. Inputs are temporally resampled from the PVC frame rate to the synthetic $100$-frame format and amplitude-normalized to the synthetic peak-temperature scale. The predicted 3D $\hat{\alpha}$ is projected to 2D masks and 2.5D depth through the same procedure used for NeFTY (Appendix H).

## G Additional Results

This appendix collects the supporting evidence referenced from Section 5: the per-difficulty robustness breakdown including the supervised U-Net (Full) upper bound (Table 7), the surface-fidelity decomposition that quantifies the data-fit paradox of Section 5.2 (Table 8 and Figure 7), the cumulative ablation in full (Table 9), the training-level wall-clock comparison referenced from Section 5.5 (Table 10), the frequency-domain diagnostics (Edge F1 and radial power spectrum), additional per-defect-count and per-layer-count qualitative reconstructions (Figures 10–13), and a representative failure mode (Figure 14). It then reports a classical PDE-constrained inversion baseline (Appendix G.8), robustness studies on irregular geometry, observation noise, and material contrast (Appendices G.9–G.11), and runtime and memory scaling with grid size and frame count (Appendix G.12).

### G.1 Robustness across defect count and layer count

Table 7: **Robustness breakdown of Table 1.** PSNR↑ and IoU↑ stratified by defect count (Homogeneous, 1 to 4 defects) and layer count (Layered Composite, 3 and 4 layers, each with 1 to 4 defects). The supervised U-Net (Full) upper bound is included here, as it consumes the volumetric labels the label-free baselines never see. Best label-free entry per column in **bold**.

| | Homogeneous | | | | | | | | Layered Composite | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 Defect | | 2 Defects | | 3 Defects | | 4 Defects | | 3 Layers | | 4 Layers | |
| **Method** | PSNR | IoU | PSNR | IoU | PSNR | IoU | PSNR | IoU | PSNR | IoU | PSNR | IoU |
| *Supervised* | | | | | | | | | | | | |
| U-Net (Full) | 27.98 | 0.72 | 24.99 | 0.72 | 22.97 | 0.69 | 20.84 | 0.66 | 20.81 | 0.67 | 19.26 | 0.68 |
| U-Net (Sound-Only) | 18.96 | 0.00 | 15.95 | 0.00 | 13.19 | 0.00 | 11.23 | 0.00 | 15.62 | 0.00 | 15.22 | 0.00 |
| *Label-free* | | | | | | | | | | | | |
| Grid Opt. | 17.41 | 0.03 | 14.01 | 0.01 | 13.04 | 0.06 | 11.51 | 0.07 | 13.15 | 0.04 | 13.40 | 0.02 |
| PINN | −0.37 | 0.01 | −0.30 | 0.01 | −0.18 | 0.02 | −0.13 | 0.02 | 1.19 | 0.02 | 1.64 | 0.02 |
| SPINN | 10.48 | 0.000 | 9.42 | 0.000 | 9.53 | 0.000 | 7.04 | 0.004 | 9.59 | 0.000 | 9.84 | 0.000 |
| **NeFTY (Ours)** | **19.99** | **0.40** | **19.36** | **0.51** | **17.55** | **0.43** | **17.04** | **0.44** | **16.69** | **0.41** | **15.07** | **0.34** |

NeFTY's IoU stays in the $0.34$-$0.51$ band across defect-count and layer-count slices, while Grid Opt. collapses below $0.07$ and PINN stays at $\sim 0.02$. SPINN, the strongest of the four PINN variants on the aggregate SSIM (Table 1), still recovers IoU $\leq 0.004$ in every slice and its PSNR drifts down with

defect density (from $10.48$ at one defect to $7.04$ at four defects), so the soft-PINN collapse persists across every defect-count and layer-count slice. The supervised U-Net (Full) peaks at IoU $0.72$ on the simplest setting and degrades only mildly with complexity, providing the upper bound referenced from Section 5.2. NeFTY narrows the gap between the label-free baselines and this upper bound in every slice without using labels.

### G.2 Surface fidelity and the data-fit paradox

Table 8: **Surface-temperature reconstruction error.** For each method we re-simulate the temperature with the recovered $\alpha_\theta$ (or directly read off $T_\phi$ for the soft-PINN family) and compare to the ground-truth surface frames. Comparing with the volumetric IoU of Table 7 surfaces the data-fit paradox of Section 5.2: PINN attains a surface PSNR above $62$ dB while its IoU stays at $\sim 0.01$. MSE is scaled by $10^{-4}$.

| | Homogeneous | | | | | | | | Layered Composite | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 Defect | | 2 Defects | | 3 Defects | | 4 Defects | | 3 Layers | | 4 Layers | |
| **Method** | MSE↓ | PSNR↑ | MSE↓ | PSNR↑ | MSE↓ | PSNR↑ | MSE↓ | PSNR↑ | MSE↓ | PSNR↑ | MSE↓ | PSNR↑ |
| Grid Opt. | 4.82 | 75.51 | 7.02 | 71.56 | 9.14 | 71.03 | 29.84 | 67.36 | 15.94 | 71.14 | 6.61 | 71.95 |
| PINN | 43.89 | 63.04 | 45.10 | 62.82 | 52.12 | 62.20 | 52.65 | 62.19 | 54.28 | 62.05 | 55.39 | 61.95 |
| **NeFTY (Ours)** | **0.50** | **82.33** | **0.52** | **82.26** | **0.73** | **81.34** | **0.56** | **82.17** | **0.54** | **82.10** | **0.50** | **82.42** |

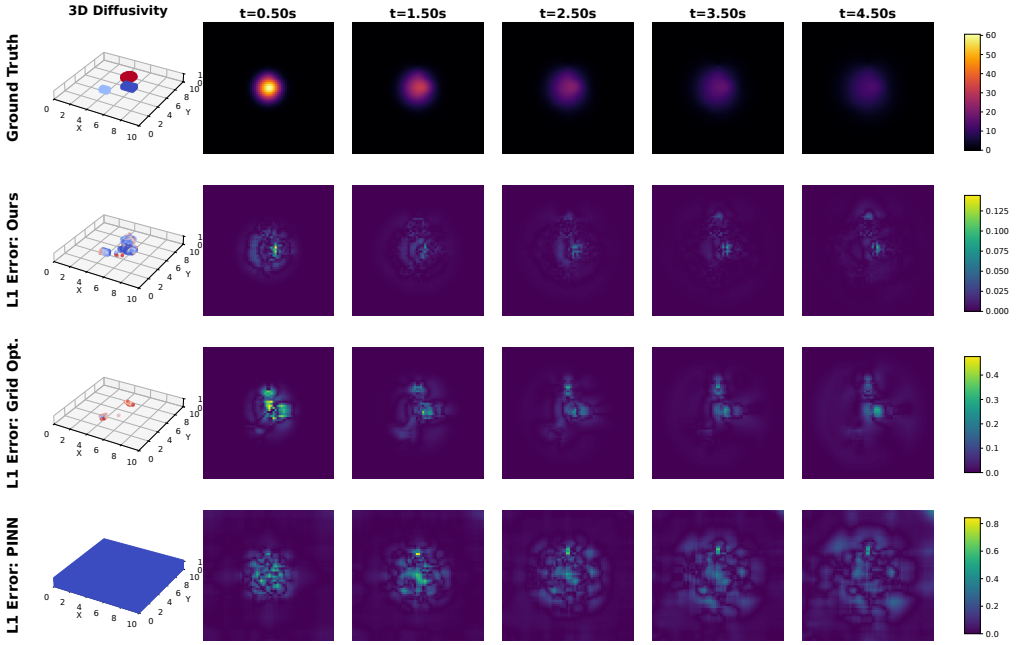



Figure 7: **Surface-temperature error maps over time** for a representative three-defect specimen. NeFTY (Row 2) attains the lowest residual error throughout the decay, while Grid Opt. (Row 3) and PINN (Row 4) exhibit structured residuals that persist and diffuse outward, signaling that their reconstructed interiors are inconsistent with the observed boundary thermograms.

The PINN baseline attains a surface MSE of $43.89 \times 10^{-4}$ and a PSNR of $63.04$ dB on the single-defect setting yet a volumetric IoU of $0.01$, the data-fit paradox anticipated by Appendix C.2: the soft-constraint formulation can fit the boundary thermograms with a non-physical interior. NeFTY is forced through the hard solver to keep its interior consistent with the boundary, and consequently attains the lowest surface MSE ($0.50 \times 10^{-4}$) and the highest PSNR ($82.33$ dB) while also recovering the correct internal structure.

### G.3 Cumulative ablation, complete

Table 9 expands the compact ablation of Table 3: each row reports MSE, PSNR, SSIM, and IoU as mean $\pm$ 95% CI across the $1{,}000$-sample synthetic benchmark, separately for the homogeneous

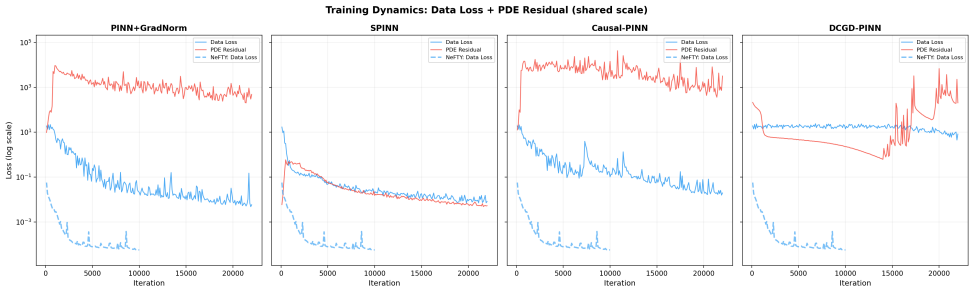



Figure 8: **Training dynamics of the four PINN variants compared to NeFTY** (representative homogeneous specimen, identical $y$-axis scale across panels). Solid blue: PINN data loss; red: PDE residual; dashed blue: NeFTY data loss. Across all four variants the PDE residual is orders of magnitude larger than the data term and highly noisy. The optimizer drives the data loss down while the diffusivity network saturates at a near-trivial solution. NeFTY's data loss (dashed) reaches orders of magnitude below all PINN data losses, since the hard solver couples surface fitting and diffusivity recovery directly.

Table 9: **Cumulative ablation in full.** Each row adds one component to the previous configuration, mean $\pm$ 95% CI across the 1,000-sample synthetic benchmark. The compact summary version is Table 3 in the main body.

| | Homogeneous | | | | Layered | | | |
|---|---|---|---|---|---|---|---|---|
| Variant | MSE ($10^{-4}$)↓ | PSNR↑ | SSIM↑ | IoU↑ | MSE ($10^{-4}$)↓ | PSNR↑ | SSIM↑ | IoU↑ |
| Base | $127.4 \pm 92.2$ | $0.47 \pm 3.17$ | $0.28 \pm 0.05$ | $0.03 \pm 0.02$ | $221.6 \pm 130.5$ | $2.89 \pm 1.88$ | $0.24 \pm 0.05$ | $0.02 \pm 0.02$ |
| + PE | $126.1 \pm 59.4$ | $4.17 \pm 1.36$ | $0.12 \pm 0.02$ | $0.09 \pm 0.03$ | $87.16 \pm 29.71$ | $6.15 \pm 1.11$ | $0.13 \pm 0.02$ | $0.09 \pm 0.03$ |
| + PE, FA | $29.80 \pm 4.78$ | $8.33 \pm 0.59$ | $0.16 \pm 0.02$ | $0.14 \pm 0.03$ | $44.33 \pm 15.94$ | $8.81 \pm 0.87$ | $0.18 \pm 0.03$ | $0.14 \pm 0.03$ |
| + PE, FA, $\sigma$ | $31.43 \pm 8.58$ | $9.19 \pm 0.94$ | $0.26 \pm 0.03$ | $0.18 \pm 0.04$ | $35.60 \pm 6.07$ | $9.27 \pm 0.82$ | $0.24 \pm 0.04$ | $0.14 \pm 0.04$ |
| + PE, FA, $\sigma$, HM | $21.01 \pm 5.67$ | $10.95 \pm 0.94$ | $0.36 \pm 0.03$ | $0.24 \pm 0.05$ | $23.09 \pm 4.77$ | $11.26 \pm 0.84$ | $0.33 \pm 0.05$ | $0.22 \pm 0.06$ |
| **Full (+ TV)** | $\mathbf{3.66 \pm 1.31}$ | $\mathbf{18.48 \pm 0.53}$ | $\mathbf{0.77 \pm 0.02}$ | $\mathbf{0.45 \pm 0.04}$ | $\mathbf{9.26 \pm 2.81}$ | $\mathbf{15.88 \pm 0.79}$ | $\mathbf{0.74 \pm 0.03}$ | $\mathbf{0.37 \pm 0.06}$ |

and layered configurations. The variance is highest at the Base and +PE configurations, where the optimizer regularly converges to qualitatively different fields across samples. Subsequent additions (FA, $\sigma$, HM, TV) progressively tighten the CI as they constrain the optimization landscape. The single largest jump is the final TV regularizer, which more than halves the diffusivity MSE and nearly doubles IoU on both configurations by suppressing the residual high-frequency noise that the previous ablation rows still admit.

### G.4 Training-level efficiency benchmark

Table 10: **Training-level wall-clock and memory per specimen on a single NVIDIA RTX PRO 6000 (96 GB)**, the training-level companion to the solver-level Table 4 in the main body. Wall-clock includes the one-time `torch.compile` overhead (Grid Opt. $\sim$ 6 s, NeFTY $\sim$ 11 s). Iterations are reported at convergence or at the saturation plateau. Collocation columns apply only to the PINN family.

| Method | Params | Colloc. | Iters | ms/iter | Wall-clock | Peak GPU |
|---|---|---|---|---|---|---|
| PINN [51] | 4.91 M | 24,576 | 22,000 | 128 | 47.0 min | 10.4 GB |
| Causal-PINN [66] | 4.91 M | 24,576 | 22,000 | 144 | 52.7 min | 10.4 GB |
| DCGD [28] | 4.91 M | 24,576 | 22,000 | 159 | 58.3 min | 11.7 GB |
| SPINN [9] | 2.64 M | 24,576 | 22,000 | 49.5 | 18.2 min | 3.9 GB |
| Grid Opt. | 0.07 M | — | 10,000 | 47.8 | 8.0 min | 1.2 GB |
| **NeFTY (Ours)** | **2.44 M** | — | **10,000** | 57.5 | **9.6 min** | **4.3 GB** |

NeFTY converges in 9.6 minutes per specimen, comparable to Grid Opt. (8.0 minutes) and 1.9$\times$ to 6.1$\times$ faster than the four PINN variants (18.2-58.3 minutes), even though the PINN baselines were given 2.2$\times$ more iterations and 2$\times$ the parameter count. The peak GPU memory of NeFTY (4.3 GB) is dominated by the unrolled Jacobi iterations of the forward solve. The discrete adjoint of Section 4.3 keeps the solver working set of the backward sweep independent of $N_t$ and stores about one forward

field per frame, in contrast to the $\mathcal{O}(KN_gN_t)$ retained iterates of standard backpropagation through time (Table 4).

### G.5 Frequency-domain diagnostics: Edge F1 and radial power spectrum

To quantify whether each method recovers sharp defect boundaries (Section 5.2), we threshold the gradient magnitude $\|\nabla\alpha\|$ at $50\%$ of the ground-truth maximum, restrict to a dilated boundary mask, and compute Edge F1. Averaged across the $1{,}000$-sample homogeneous benchmark we obtain $0.470 \pm 0.137$ for NeFTY, $0.029 \pm 0.036$ for Grid Opt., and $0.000$ for vanilla PINN, SPINN, Causal-PINN, and DCGD. NeFTY is the only label-free method whose recovered field carries detectable defect-scale gradients, while the four PINN variants produce featureless interiors. Figure 9 plots the radial power spectrum, computed over a $16 \times 16 \times 8$ defect crop and radially averaged across the same $1{,}000$-sample benchmark. NeFTY tracks the ground-truth spectrum from the lowest wavenumbers up to the Nyquist limit ($\sim 8$ cycles per unit length in the $z$-direction), while all four PINN variants collapse to near-zero power across the spectrum and Grid Opt. falls off above the mid frequencies. The pattern is consistent with Corollary 1: the parameter-to-observation map attenuates high spatial frequencies, so a reconstruction recovers defect-scale structure only if it restores that content.

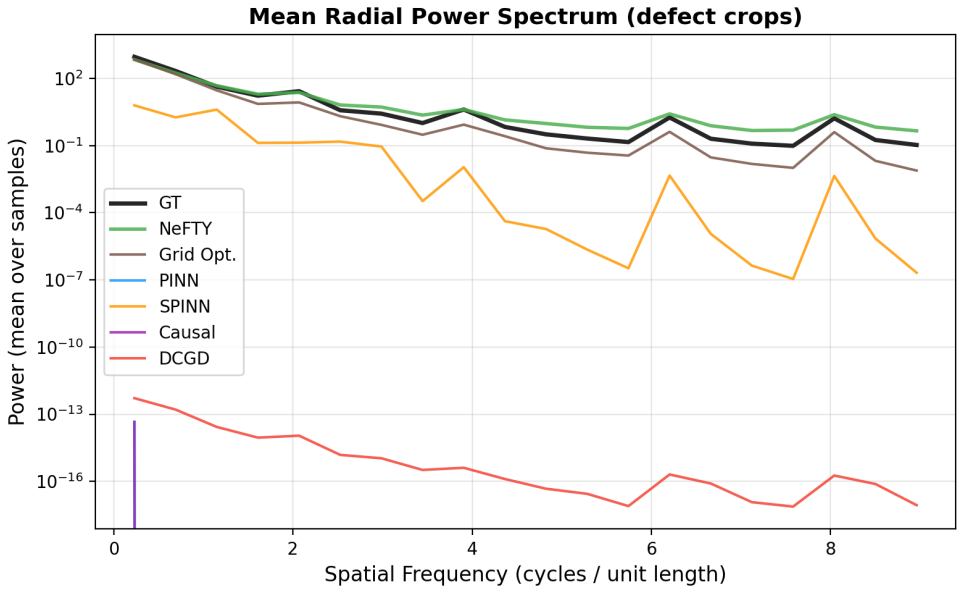



Figure 9: **Mean radial power spectrum** of the recovered diffusivity field, averaged across the $1{,}000$-sample homogeneous benchmark. NeFTY (green) tracks the ground-truth spectrum (black) across all spatial frequencies up to the Nyquist limit. Grid Opt. tracks at low frequencies but drops off above the mid range, and the four PINN variants collapse to near-zero power.

### G.6 Additional qualitative reconstructions

To complement the headline qualitative reconstructions of Figures 4 and 5 in the main body, we provide a stratified qualitative comparison across defect density and layer count. Figures 10–12 show single-, two-, and four-defect reconstructions in the homogeneous setting, illustrating that NeFTY's defect-localization quality is preserved as the number of subsurface scatterers grows from one to four. In particular, NeFTY separates two laterally adjacent defects (Figure 11) without merging their thermal signatures, a regime in which the voxel grid of Grid Opt. blurs both defects together. Figure 13 extends the comparison to a four-layer composite where the bulk diffusivity steps with depth (visible as changing background intensity in the ground-truth row). NeFTY isolates the embedded defects against the layered background, confirming the layered-setting trend of Table 1 at the per-specimen level.

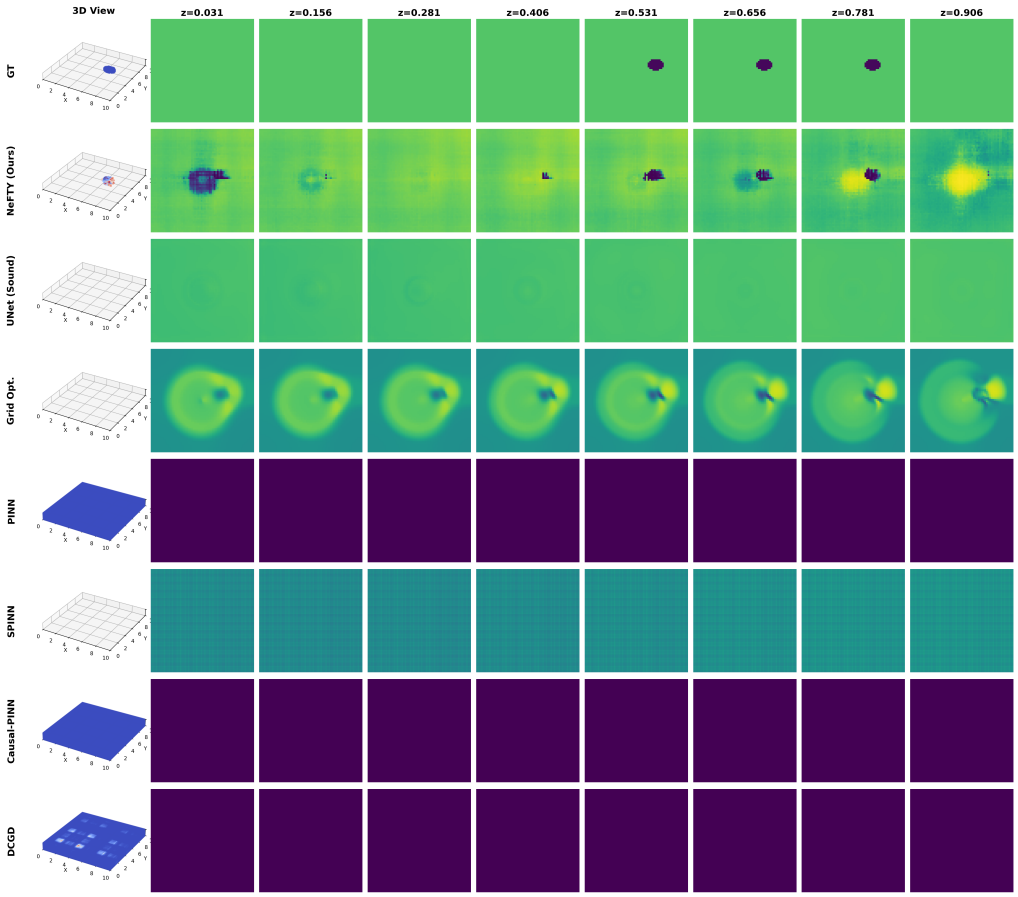



Figure 10: **Single-defect reconstruction (homogeneous bulk).** NeFTY recovers a clean defect outline at the correct depth, while Grid Opt. exhibits ringing artifacts and the four PINN variants converge to featureless fields.

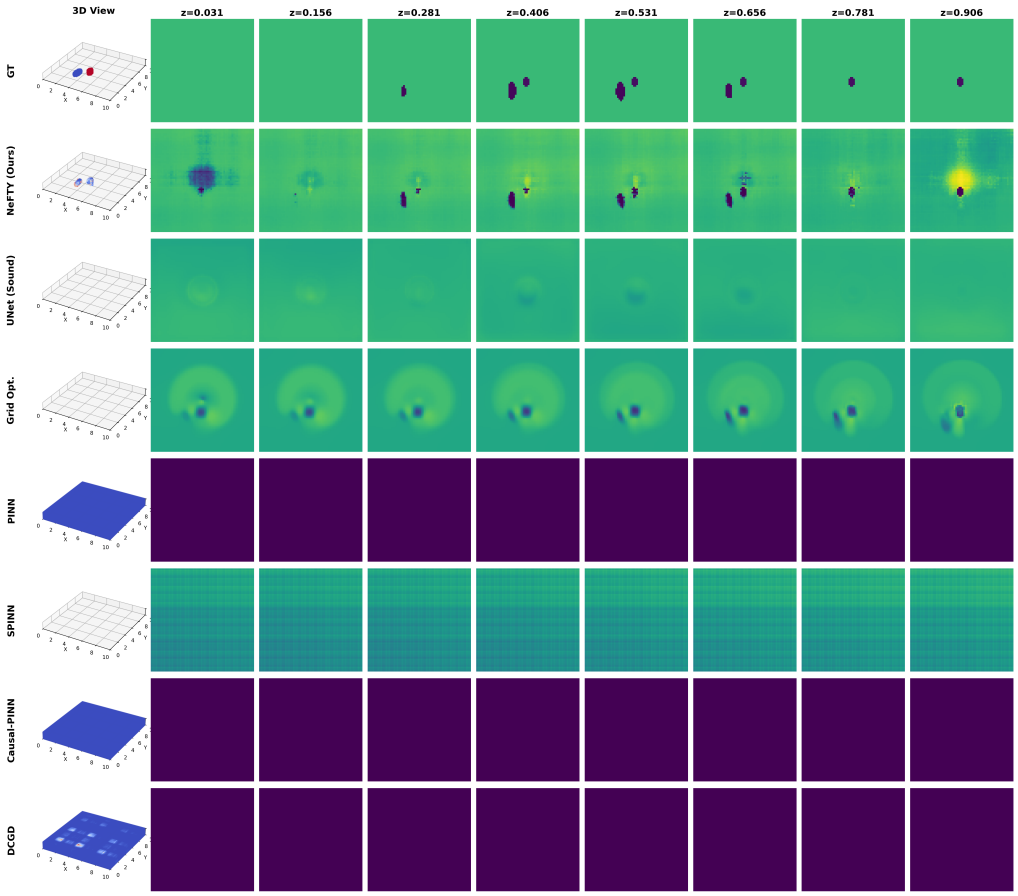



Figure 11: **Two-defect reconstruction.** NeFTY separates two laterally adjacent defects without merging their thermal signatures, while Grid Opt. blurs them together.

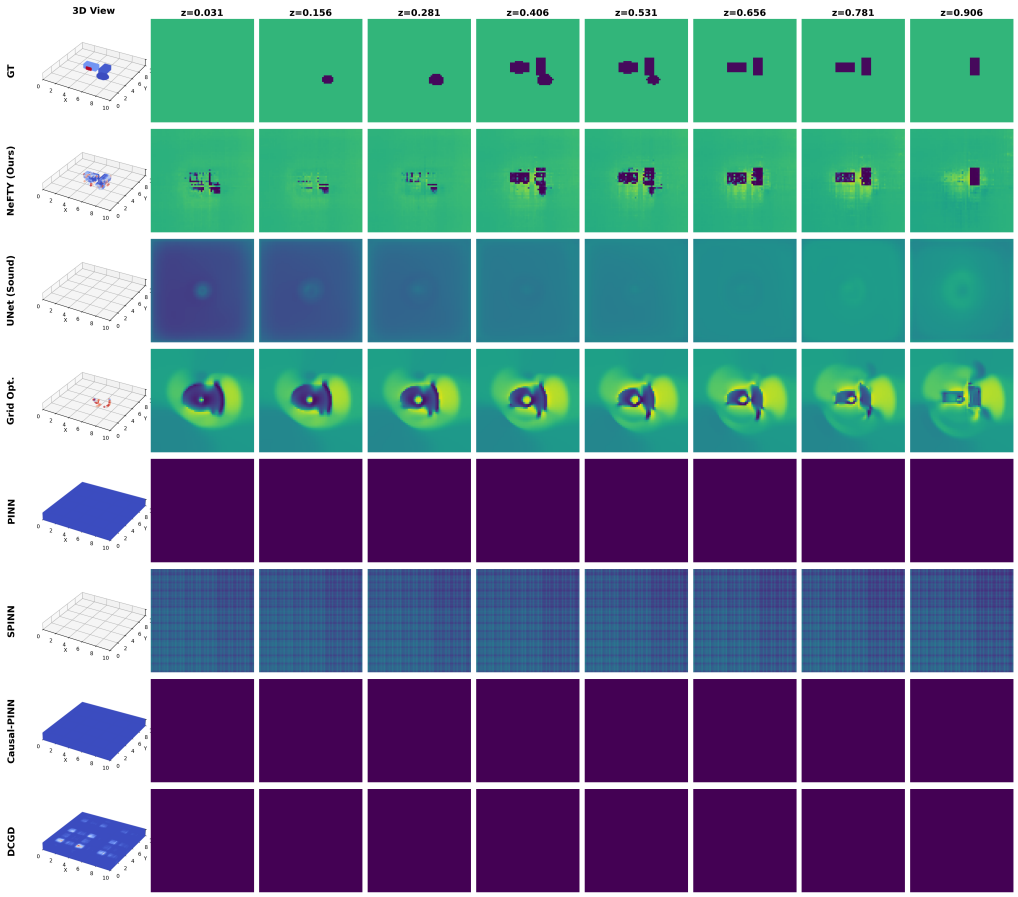



Figure 12: **Four-defect reconstruction.** NeFTY isolates each of four defects at the highest density of the synthetic benchmark. The Sound-Only U-Net ghosts all defects, in line with its zero IoU in Table 7.

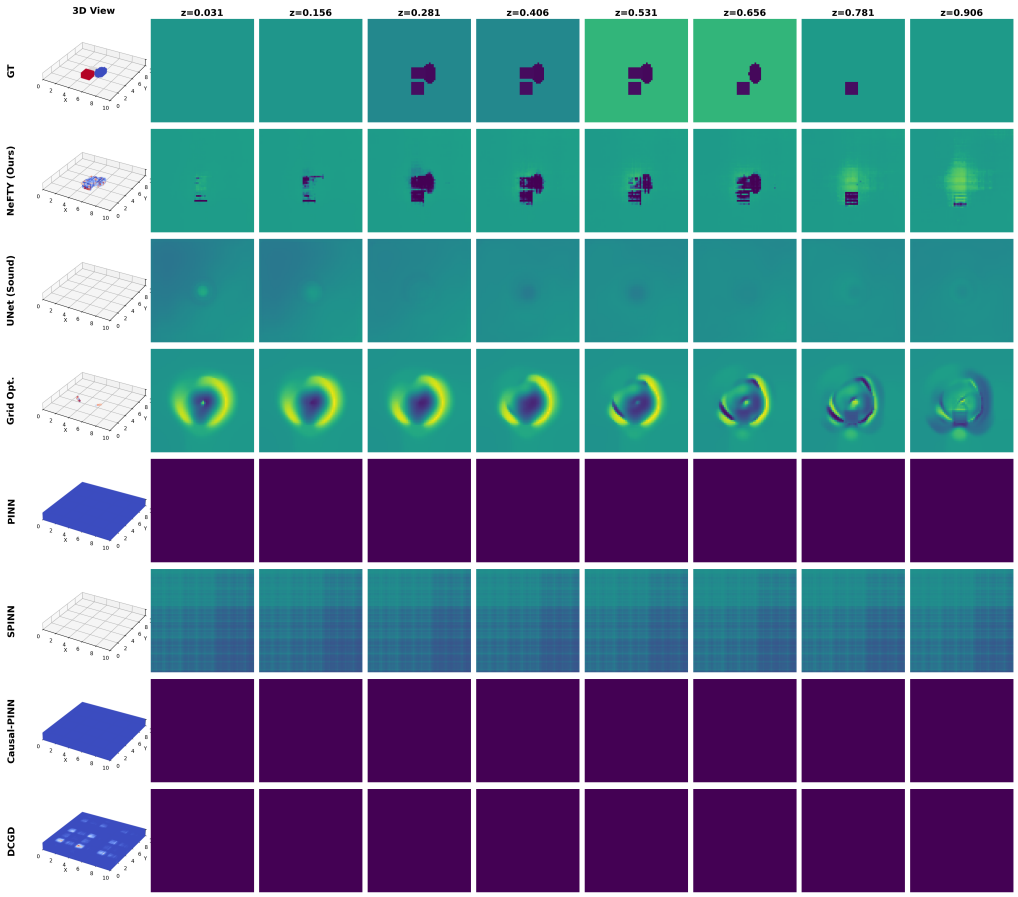



Figure 13: **Four-layer composite reconstruction.** The bulk diffusivity steps with depth (visible as changing background intensity in the ground-truth row). NeFTY correctly localizes the embedded defects against this stratified background.

### G.7 Failure mode

We identify one consistent failure mode of NeFTY: when defects are placed shallow and close to the laser-illuminated front surface, the surface-only data become weakly sensitive to the back-face diffusivity, and the optimizer occasionally introduces a spurious low-$\alpha$ artifact along the back boundary that is consistent with the surface decay but not with the ground truth (Figure 14). The defect itself is still recovered at the correct lateral position, so the artifact does not affect 2D segmentation, but it propagates into the depth metric of Table 2 for shallow defects. Three remedies target this mechanism directly: stronger back-face priors, a multi-side acquisition that adds back-face thermograms, and a depth-adaptive TV weight that suppresses high-curvature solutions far from observed pixels.

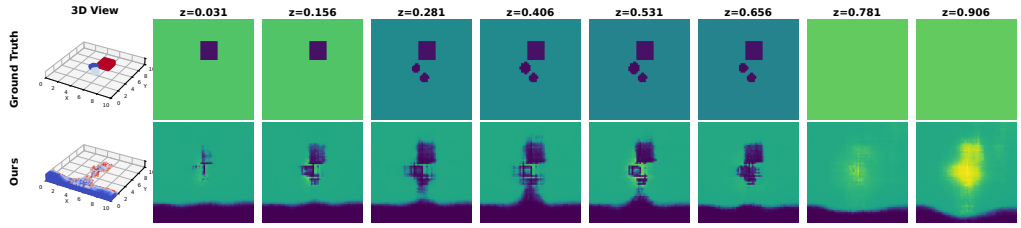



Figure 14: **Failure mode: shallow defects close to the heat source.** NeFTY correctly localizes the central defects but introduces a low-diffusivity artifact along the back-face boundary of the lateral plane.

### G.8 Classical PDE-constrained inversion baseline

NeFTY belongs to the discretize-then-optimize, adjoint-state family of PDE-constrained inversion [2, 25, 30]: the heat equation is enforced by a numerical solver, and gradients reach the unknown field through the discrete adjoint. To separate the contributions of the hard constraint, of the optimizer and regularizer, and of the neural-field representation, we compare NeFTY with a classical member of this family.

**Protocol.** A bounded voxel diffusivity field is optimized with L-BFGS under total-variation regularization and, separately, $H^1$ (Tikhonov) regularization, through the same solver and discrete adjoint as NeFTY (Section 4). Regularization weights are tuned on a calibration set disjoint from the evaluation specimens. The iteration budget is set by convergence: the non-stalled runs reached stationarity by about 300 L-BFGS iterations (roughly 600 closure evaluations, since each iteration line-searches through multiple full forward and adjoint solves). The classical arm therefore received more wall-clock than NeFTY: in an eager-mode spot-check ($n = 5$), its median wall-clock was 20.5 min per specimen, more than double NeFTY's compiled 9.6 min (Table 10). Both L-BFGS arms are evaluated on 64 paired benchmark specimens.

Table 11: **Classical PDE-constrained inversion versus NeFTY** (volumetric IoU↑). The two L-BFGS rows are means over 64 paired benchmark specimens, while the Grid Opt. and NeFTY rows are the full-benchmark values of Table 1. The multiples quoted in the text are ratios of means computed on the same 64 specimens.

| METHOD | BASIS | IOU (HOMOG.) | IOU (LAYERED) |
|---|---|---|---|
| GRID OPT. | TABLE 1, FULL BENCHMARK | 0.04 | 0.03 |
| VOXEL FIELD, L-BFGS + TV | 64 PAIRED SPECIMENS | 0.129 | 0.112 |
| VOXEL FIELD, L-BFGS + $H^1$ | 64 PAIRED SPECIMENS | 0.125 | 0.127 |
| **NEFTY (OURS)** | TABLE 1, FULL BENCHMARK | **0.45** | **0.37** |

**Decomposition.** The hard constraint and adjoint alone (Grid Opt., with the same solver, adjoint, and TV regularizer, lacking only the neural prior; Appendix F.2) set the base. Computed on the same 64 specimens (ratio of means), a quasi-Newton optimizer with classical regularization multiplies this base by about 3.2×, and the neural-field prior adds a further factor of about 3.3× over the L-BFGS arms. Among the evaluated configurations, the representation carries the largest share of the gap,

without exhausting optimizer-representation interactions. The two regularizers recover essentially the same fields (a paired comparison detects no difference).

**Scope of the comparison.** We evaluate the TV/Tikhonov family because it can be tuned to a fair standard under a shared solver and adjoint, which isolates the optimizer and the regularizer. Gauss-Newton inversion [21] needs repeated linearized PDE solves per iteration, and level-set shape inversion [56] brings its own shape-parameterization machinery. Each merits a dedicated, fully tuned implementation, since an under-tuned one would understate it. Our comparative claims are scoped to the evaluated methods.

### G.9 Irregular defect geometry

The main benchmark uses ellipsoidal, cylindrical, and box-shaped defects in homogeneous or layered hosts (Appendix E.1), a setting aligned with the piecewise-constant structure that the TV regularizer promotes. To test dependence on regular shapes, we generate two nonregular categories, with 32 new specimens per category and difficulty parameters fixed before training: irregular blobs, whose boundaries are perturbed by up to $50\%$ of the base defect radius, and curved strata interfaces. All defects are fully embedded. We run the primary NeFTY configuration and the Grid Opt. baseline (voxel field through the same solver, adjoint, and TV regularizer, without the neural prior) on each category.

Table 12: **Irregular defect geometry.** Volumetric IoU↑ on two nonregular categories (32 new specimens per category). Brackets give the $95\%$ CI.

| CATEGORY | NEFTY 3D IOU [95% CI] | GRID OPT. 3D IOU |
|---|---|---|
| IRREGULAR BLOBS | $0.316\ [0.241, 0.392]$ | $0.000$ |
| CURVED STRATA INTERFACES | $0.222\ [0.076, 0.368]$ | $0.000$ |

NeFTY recovers volumetric defect structure in both nonregular categories, while Grid Opt. reaches IoU $0.000$ in both (Table 12). We primarily target thermal NDT of specimens whose defects (voids, delaminations, inclusions, flat-bottom holes) form discrete regions in a host material, where piecewise-constant diffusivity predominantly holds. Smoothly varying diffusivity, as in functionally graded materials, is more ill-posed: absent structural assumptions, many smooth interior fields map to nearly identical surface histories (Section 3.2). Extending NeFTY to this regime is important future work.

### G.10 Observation noise

We add Gaussian observation noise with standard deviation $\sigma_{\mathrm{n}} \in \{0.5, 1, 2, 5\}\%$ of the peak signal to 16 synthetic specimens and run the primary NeFTY configuration on each, pairing every noisy run with the same specimen's noise-free run. Table 13 reports IoU retention, the mean IoU under noise divided by the mean noise-free IoU on the same specimens (ratio of means).

Table 13: **Observation-noise sweep.** NeFTY IoU retention relative to the paired noise-free runs on the same 16 synthetic specimens (ratio of means).

| NOISE (% OF PEAK SIGNAL) | 0.5 | 1 | 2 | 5 |
|---|---|---|---|---|
| NEFTY IOU RETENTION (RATIO OF MEANS) | 96% | 82% | 64% | 37% |

Degradation is monotone in the noise level, with no collapse at low noise. The confidence interval of the paired IoU drop excludes zero only at $5\%$. The failure mode under noise is over-segmentation: across all runs of the sweep, including the noise-free anchors (16 specimens at five noise levels including $\sigma_{\mathrm{n}} = 0$), only $3.75\%$ return empty predictions. These are within-method curves against NeFTY's own noise-free anchor. Its noise levels are percentages of the peak post-flash signal, whereas the camera that recorded the real PVC sequences, a cooled FLIR SC5000 [69], has a manufacturer-specified NETD of about 20 mK, which is roughly $0.007\%$ of an absolute room-temperature (293 K) reading.

### G.11 Material contrast and solver diagnostic

We vary the defect-to-bulk diffusivity contrast over $1:5$, $1:10$, $1:20$, and $1:50$ with 5 seeds per ratio and run the primary NeFTY configuration at each ratio. Seed-paired IoU differences relative to $1:5$ all have confidence intervals straddling zero, so IoU shows no detectable change from $1:5$ to $1:50$.

**Solver diagnostic.** Production inversion uses a fixed budget of $K = 50$ Jacobi sweeps per implicit-Euler step, so the work per step does not depend on contrast. To check whether the linear solve itself becomes harder with contrast, a convergence diagnostic at ground-truth coefficients (first implicit-Euler step, fixed bulk diffusivity $\alpha_{\text{base}} = 0.15$) instead counts Jacobi sweeps until a fixed diagnostic tolerance is reached (Table 14).

Table 14: **Solver convergence diagnostic across material contrast**: Jacobi sweeps until a fixed diagnostic tolerance, at ground-truth coefficients (first implicit-Euler step, fixed $\alpha_{\text{base}} = 0.15$). Production inversion instead runs a fixed budget of $K = 50$ sweeps per step.

| DEFECT:BULK CONTRAST | 1:5 | 1:10 | 1:20 | 1:50 |
|---|---|---|---|---|
| SWEEPS TO TOLERANCE (MEAN) | 55.4 | 55.6 | 55.6 | 55.6 |

Across all specimens the sweep count spans 53 to 57, every solve converged, and the mean is essentially unchanged with contrast (55.4 to 55.6), so solver work is insensitive to contrast in this range. Beyond roughly $1:1000$ the linear system becomes severely ill-conditioned and the solver stalls (Appendix E.1). Preconditioned or multigrid solvers are the path to that regime (Appendix J).

### G.12 Runtime and memory scaling

To measure how cost grows with problem size, we record NeFTY's runtime and peak memory on the same hardware used for Tables 4 and 10, on grids up to $128 \times 128 \times 32$ and sequences up to 200 frames.

**Memory.** The adjoint's solver-level working set is $\mathcal{O}(N_g)$, independent of the frame count (Appendix D.3; 21.9 MB versus 18.63 GB for autograd in Table 4). The fitted peak-memory exponents are 0.985 in voxels and 0.694 in frames ($R^2 = 0.9998$): measured total memory is a frame-independent solver term plus about one stored field per frame, versus the $K = 50$ iterates per frame that backpropagation through the solver would retain.

**Runtime.** Runtime follows an affine model in the frame count whose base term scales as $N_g^{0.909}$, with a per-frame marginal cost that stays near-flat across a $64\times$ increase in voxels (Table 15). Held-out extrapolation of the fitted model is within $11.8\%$.

Table 15: **Per-frame marginal cost across grid sizes**: the fitted per-frame term of the affine runtime model, measured on the hardware of Tables 4 and 10.

| GRID | $32 \times 32 \times 8$ | $64 \times 64 \times 16$ | $96 \times 96 \times 24$ | $128 \times 128 \times 32$ |
|---|---|---|---|---|
| PER-FRAME MARGINAL COST (MS/FRAME) | 0.49 | 0.50 | 0.59 | 0.68 |

## H Real PVC Data Details

### H.1 Datasets

The PVC-Infrared [69] and PVC-Depth [68] benchmarks consist of 19 and 38 pulsed-thermography sequences respectively on $100 \times 100 \times 5$ mm PVC specimens with cylindrical flat-bottom holes of varying diameter and depth, recorded with a cooled mid-wave-infrared camera (FLIR SC5000 [69]) at second-scale dynamics under flash-lamp excitation. PVC-Infrared annotates each specimen with a 2D binary defect mask derived from CAD drawings and bottom-view inspection, and PVC-Depth additionally annotates discrete defect depths. Neither benchmark provides a 3D volumetric diffusivity ground truth: X-ray CT provides volumetric defect geometry [19], with reported scan times of 5 to 96 hours per specimen [52] on systems priced from roughly \$100K to over \$1M [11], and volumetric diffusivity labels additionally require material characterization or destructive sectioning. The

dataset authors themselves describe labeling the 3D spatial location of defects as a substantial challenge [68]. Thermal sequences are temporally resampled and amplitude-normalized as documented in Appendix F.5 for compatibility across baselines.

### H.2 Physical adaptations

The synthetic benchmark of Appendix E.1 is set on micron-scale specimens at microsecond timescales with negligible convective dissipation, while the PVC specimens are centimeter-scale at second-scale dynamics. Three physical adaptations of Appendix A.5 are required for NeFTY to apply without re-training the ground-truth simulator.

**Initial condition.** The Gaussian post-flash profile is replaced by a near-uniform front-face flash $T_0(\mathbf{x}) = T_{\mathrm{amb}} + A\exp(-z^2/(2w_z^2))$, with $w_z$ the absorption depth and $A$ calibrated against the first measured frame. This matches the broad flash-lamp excitation used by Wei et al. [68, 69].

**Convective Robin boundary.** On the back face we replace the synthetic adiabatic Neumann boundary by a Robin condition $\mathbf{n}\cdot(\alpha\nabla T) = -h\,(T - T_{\mathrm{amb}})$, with $h > 0$ a calibrated convective coefficient. The unitless $h$ is set to match the asymptotic Biot regime of natural convection on PVC over the relevant second-scale window. After the ambient shift of Section 4.2, this becomes $\mathbf{n}\cdot(\alpha\nabla T) = -h\,T$, so the discrete implicit-Euler step remains affine-free and the diagonal-dominance argument of Appendix D.2 carries over without change.

**Dimensional rescaling.** The unitless solver evolves in $(t/t_{\mathrm{total}}, \mathbf{x}/L_0)$ coordinates with effective diffusivity $\alpha_{\mathrm{sim}} = \alpha_{\mathrm{phys}}\, t_{\mathrm{total}}/L_0^2$ (Appendix A.5). For PVC we set $L_0 = 100\,\mathrm{mm}$ (lateral specimen size) and $t_{\mathrm{total}}$ to the recording window of the benchmark, so the unitless diffusivity ranges of the synthetic benchmark map onto the physical PVC diffusivity range. Lateral boundaries remain periodic when the camera footprint is far from the specimen edge. On cropped specimens we replace lateral periodicity by adiabatic Neumann boundaries.

### H.3 Projection from 3D diffusivity to 2D and 2.5D labels

The reconstructed 3D field $\hat{\alpha}_\theta(x, y, z)$ is reduced to the labels supplied by the benchmarks through two complementary projections.

**2D defect mask.** For each lateral pixel $(x, y)$, we compute the depth-averaged diffusivity $\bar{\alpha}(x, y) = N_z^{-1}\sum_z \hat{\alpha}_\theta(x, y, z)$ and the normalized one-sided diffusivity deficit $d(x, y) = [1 - \bar{\alpha}(x, y)/\alpha_{\mathrm{base}}]_+$. Because sound pixels should have $\bar{\alpha}(x, y) \approx \alpha_{\mathrm{base}}$, we estimate the reconstruction noise floor by $\hat{\sigma} = 1.4826\,\mathrm{MAD}\{d(x, y)\}_{(x,y)}$ and define the mask to be $\mathbf{1}\{d(x, y) > 2\hat{\sigma}\}$. The same fixed post-processing rule is applied to NeFTY, Grid Opt., Swin-UNETR, and SPINN, while PPT/TSR follow their publication-default segmentation rules.

**2.5D defect depth.** For each pixel $p$ inside the predicted 2D mask we extract the through-thickness depth $\hat{d}_p$ as the spatial median of the depths of the $z$-voxels whose $\hat{\alpha}_\theta$ falls below the bulk by more than the noise floor of the defect-free region. PPT and TSR use their published depth formulas with the calibrated PVC diffusivity. Swin-UNETR uses the same spatial-median rule applied to its supervised prediction.

### H.4 Real PVC depth, full metric breakdown

We provide the full table and an additional qualitative example in Table 16 and Figure 15.

Table 16: **Real PVC 2.5D depth recovery, full metric breakdown.** Companion to Table 2 in the main body, reporting RMSE in millimetres and the full $\delta < 1.25^k$ threshold accuracies for $k \in \{1, 2, 3\}$ alongside Abs Rel. NeFTY is best label-free on every metric.

| METHOD | ABS REL ↓ | RMSE (MM) ↓ | $\delta < 1.25$ ↑ | $\delta < 1.25^2$ ↑ | $\delta < 1.25^3$ ↑ |
|---|---|---|---|---|---|
| SWIN-UNETR [24] | 3.131 | 1.983 | 0.052 | 0.126 | 0.190 |
| PPT [38] | 3.416 | 2.075 | 0.005 | 0.006 | 0.073 |
| TSR [59] | 13.366 | 7.655 | 0.002 | 0.002 | 0.002 |
| SPINN [9] | 6.737 | 3.913 | 0.002 | 0.002 | 0.004 |
| GRID OPT. | 0.941 | 0.983 | 0.104 | 0.399 | 0.614 |
| **NEFTY (OURS)** | **0.465** | **0.541** | **0.385** | **0.621** | **0.743** |

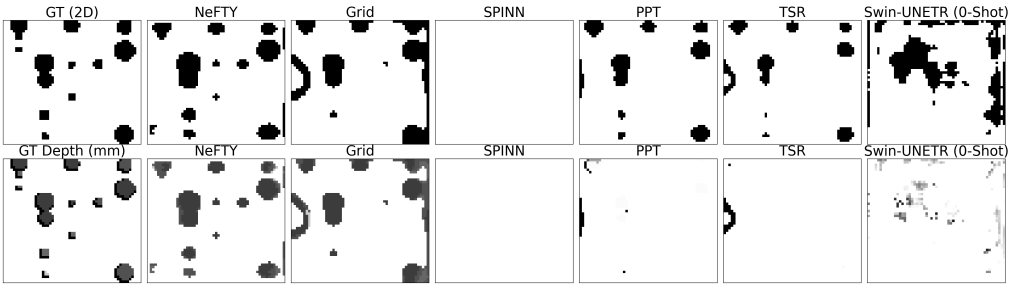



Figure 15: **Real PVC reconstruction on a second held-out specimen** (companion to Figure 6). Row 1: defect masks; Row 2: predicted depth (greyscale, darker is deeper). NeFTY recovers both the defect layout and the depth ordering from the surface IR sequence, in contrast to PPT/TSR which produce coarse depth and SPINN which collapses to a near-uniform field.

# I Validation of the Differentiable Heat Diffusion Simulator

We validate the correctness of our differentiable heat diffusion simulator, implemented using an implicit Euler time discretization solved via a Jacobi iterative scheme, through both analytical consistency checks and numerical experiments.

## I.1 Governing Equation and Analytical Behavior

To validate our simulator, we fix $\alpha$ as a constant, set the source term $Q$ in Eq. (1) to zero, and prescribe a Gaussian initial temperature distribution. Under these settings the simulator reduces to

$$\frac{\partial T(\mathbf{x}, t)}{\partial t} = \alpha \nabla^2 T(\mathbf{x}, t), \tag{28}$$

where $T(\mathbf{x}, t)$ denotes temperature and $\alpha$ is a constant, isotropic thermal diffusivity.

For an initial Gaussian temperature distribution

$$T(\mathbf{x}, 0) = A \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_0\|^2}{2\sigma_0^2}\right), \tag{29}$$

where $A > 0$ is the peak amplitude, $\mathbf{x}_0 \in \mathbb{R}^3$ the source center, and $\sigma_0 > 0$ the initial standard deviation along each axis. The analytical solution of Eq. (28) remains Gaussian for all $t > 0$. In particular, the variance along each spatial dimension evolves as

$$\sigma^2(t) = \sigma_0^2 + 2\alpha t, \tag{30}$$

which implies a linear growth rate

$$\frac{d\sigma^2}{dt} = 2\alpha. \tag{31}$$

This property provides a quantitative criterion for validating the physical fidelity of a numerical diffusion solver.

### I.2 Numerical Setup

We discretize the spatial domain using a uniform Cartesian grid and advance Eq. (28) in time using an implicit Euler scheme. The resulting linear system at each time step is solved using a fixed number of Jacobi iterations, yielding a fully differentiable simulation pipeline.

Periodic boundary conditions are used in the $x$ and $y$ directions, and zero-flux (Neumann) boundary conditions are applied along the $z$ axis. Temperature observations are taken from the top surface of the domain to match the sensing configuration used in thermal imaging.

### I.3 Gaussian Diffusion Rate Verification

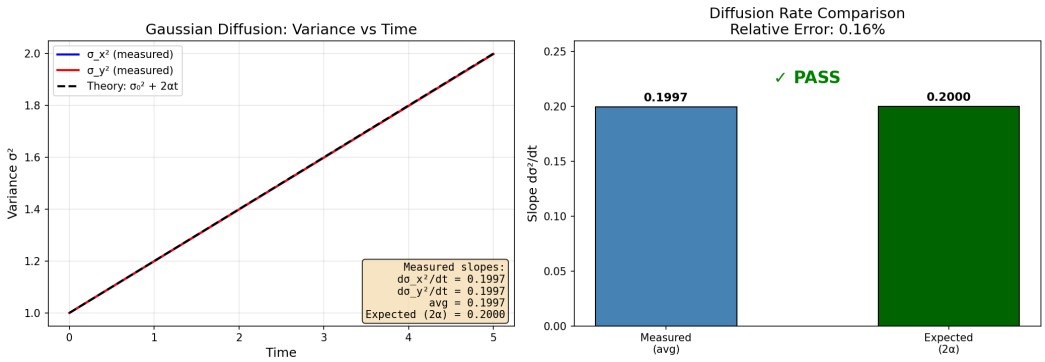



Figure 16: **Gaussian Diffusion Rate Validation.** (*Left*) Temporal evolution of the temperature variance $\sigma^2$ along the $x$ and $y$ directions, measured on the surface for a diffusing Gaussian heat source with constant diffusivity $\alpha = 0.1$. Both $\sigma_x^2(t)$ and $\sigma_y^2(t)$ grow linearly over time and closely follow the analytical prediction $\sigma^2(t) = \sigma_0^2 + 2\alpha t$. (*Right*) Comparison between the measured average slope $\mathrm{d}\sigma^2/\mathrm{d}t$ and the theoretical value $2\alpha$, showing a relative error of $0.16\%$. This result quantitatively confirms that the proposed simulator reproduces the correct diffusion rate of the heat equation.

To verify the analytical variance growth in Eq. (30), we simulate the diffusion of a 3D Gaussian heat source with constant diffusivity $\alpha = 0.1$ in a large domain. At each time step, we compute the temperature-weighted second moments on the surface,

$$\sigma_x^2(t) = \frac{\sum (x - \bar{x})^2 T(x, y, t)}{\sum T(x, y, t)}, \quad \sigma_y^2(t) = \frac{\sum (y - \bar{y})^2 T(x, y, t)}{\sum T(x, y, t)}, \tag{32}$$

with $\bar{x} = \sum x\, T(x, y, t) / \sum T(x, y, t)$ and $\bar{y}$ defined analogously, the temperature-weighted lateral centroids on the observed top surface.

Linear regression is performed on $\sigma_x^2(t)$ and $\sigma_y^2(t)$ after an initial transient period. As shown in Fig. 16, both measured variances exhibit a linear increase over time, with an average slope of $0.1997$, compared to the theoretical value $2\alpha = 0.2000$. The resulting relative error with respect to the analytical solution is $0.16\%$.

### I.4 Qualitative Validation: Constant and Variable Diffusivity

We further validate the simulator through qualitative visualization of the temperature evolution.

**Constant diffusivity.** Figure 17 shows the surface temperature evolution under uniform diffusivity ($\alpha = 0.1$ and $\alpha = 1.0$). The initially localized heat source spreads isotropically over time, preserving Gaussian symmetry as expected from Eq. (28).

**Variable diffusivity with defect.** To test spatially varying material properties, we introduce a low-diffusivity spherical defect embedded in a homogeneous background. As illustrated in Fig. 18 (bottom row), the heat propagation is locally impeded near the defect region, resulting in a clearly asymmetric temperature field. This behavior is consistent with the physical interpretation of reduced thermal diffusivity.

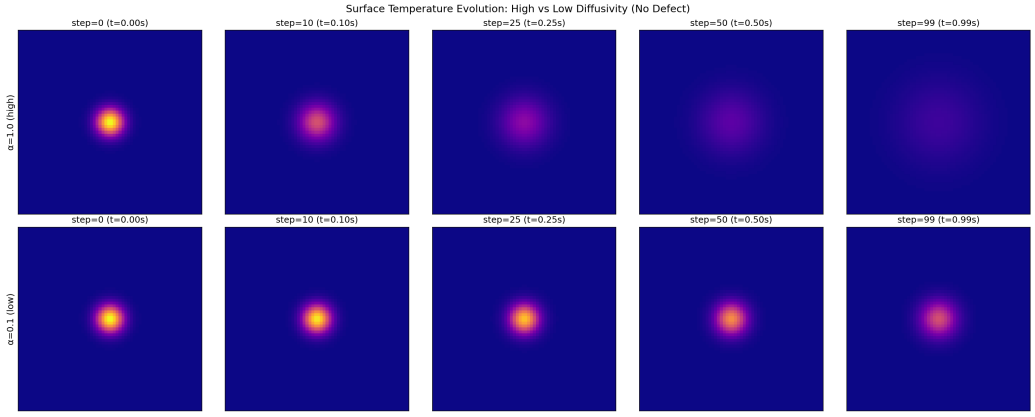



Figure 17: **Effect of Diffusivity Magnitude on Surface Heat Diffusion.** Surface temperature evolution for a defect-free domain under high diffusivity ($\alpha = 1.0$, top row) and low diffusivity ($\alpha = 0.1$, bottom row), shown at matched time steps. With identical spatial and temporal discretization, the high-diffusivity case exhibits substantially faster spatial spreading and a more rapid decay of peak temperature, while the low-diffusivity case retains a localized, high-contrast heat profile. These results qualitatively illustrate the expected dependence of diffusion dynamics on the thermal diffusivity parameter $\alpha$.

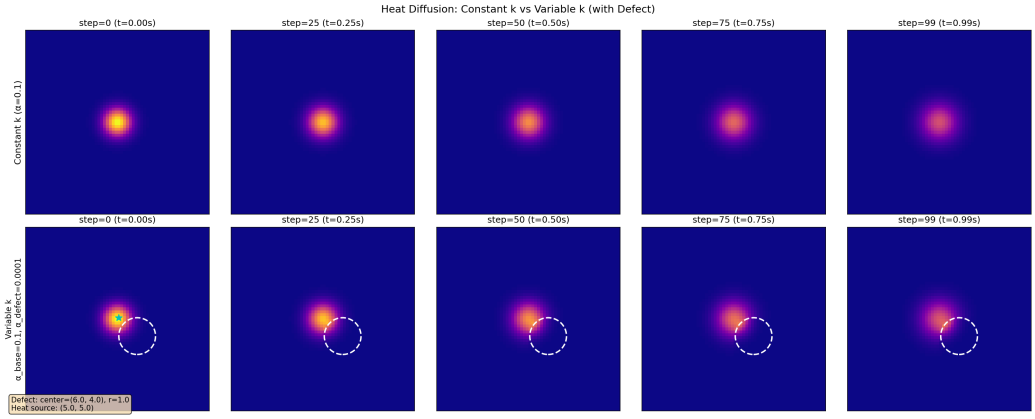



Figure 18: **Surface Temperature Evolution under Constant and Spatially Varying Diffusivity.** Surface temperature maps at selected time steps for heat diffusion with constant diffusivity ($\alpha = 0.1$, top row) and spatially varying diffusivity with an embedded low-diffusivity defect (bottom row). In the homogeneous case, the heat source spreads isotropically and preserves Gaussian symmetry over time. In contrast, the presence of a defect locally impedes heat propagation, leading to asymmetric temperature distributions and delayed diffusion in the defect region (outlined by the dashed circle). These results demonstrate that the simulator correctly captures both uniform and heterogeneous diffusion behavior.

### I.5 Effect of Diffusivity Magnitude

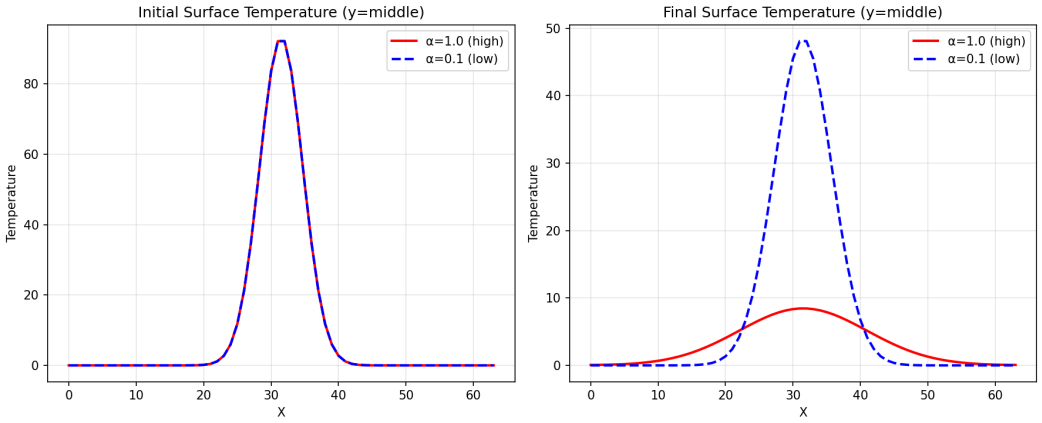



Figure 19: **Surface Temperature Profiles under Different Diffusivities.** One-dimensional cross-sections of the surface temperature along the $x$ direction at the midline ($y = \text{middle}$), comparing high diffusivity ($\alpha = 1.0$) and low diffusivity ($\alpha = 0.1$). (*Left*) At the initial time step, both cases exhibit identical Gaussian profiles, confirming consistent initialization. (*Right*) At the final time step, the high-diffusivity case shows a significantly broader and lower-amplitude profile, reflecting faster spatial spreading of heat, while the low-diffusivity case retains a sharper peak. These results qualitatively validate the expected dependence of diffusion dynamics on the thermal diffusivity parameter $\alpha$.

We also compare diffusion dynamics under different diffusivity values in a defect-free setting. Figure 19 contrasts $\alpha = 1.0$ (high diffusivity) with $\alpha = 0.1$ (low diffusivity) using identical spatial and temporal resolutions. As expected, higher diffusivity leads to faster spreading and lower peak temperatures, while preserving the overall Gaussian structure.

### I.6 Error Sources: Discretization, Solver, and Inversion

NeFTY's temperature satisfies the discretized heat equation up to the chosen spatial and temporal discretization and the iterative linear-solver tolerance (Section 4.2), and the adjoint gradient is exact up to the same solver tolerance (Appendix D.2). Isolating continuous-model error requires an exact solution of the continuous PDE, which exists only for special configurations. We therefore quantify the combined numerical error where an analytic reference exists: the Gaussian diffusion-rate test above gives a relative error of $0.16\%$, and Table 4 reports the simulation error of the solver. The benchmark metrics of Section 5 measure inversion error against the ground-truth diffusivity.

## J Limitations and Future Work

**Per-specimen optimization.** NeFTY targets offline inspection: it recovers one specimen in $\sim 10$ minutes on one workstation GPU ($9.6$ min and $4.3$ GB of peak memory in Table 10), from the same one-sided, non-contact acquisition of a single surface temperature sequence that standard thermographic methods (PPT/TSR) already use, and without volumetric training labels. For reference, X-ray CT provides volumetric defect geometry [19], with reported scan times of $5$ to $96$ hours per specimen [52] on systems priced from roughly \$100K to over \$1M [11], and volumetric diffusivity labels additionally require material characterization or destructive sectioning. Supervised feedforward inversion (e.g., U-Net) runs in milliseconds once trained but requires volumetric labels. Online or high-throughput screening calls for lower per-specimen latency. Amortized initialization, for example meta-learning or hypernetworks that predict a good initialization for the neural field, is an important future direction. The measured memory and runtime scaling (Appendix G.12) supports growing the problem size.

**Quantitative diffusivity under high contrast.** Our benchmark uses a defect-to-bulk diffusivity contrast of $\sim 1:20$ (Appendix E.1), and IoU shows no detectable change from $1:5$ to $1:50$

(Appendix G.11). At realistic air-to-solid ratios ($> 1:1000$), the linear system $\mathbf{A}(\alpha_\theta)$ becomes severely ill-conditioned and the iterative solver stalls. NeFTY recovers defect *geometry* in the evaluated regime (Section 5.2, Section 5.3), while the precise quantitative magnitude of $\alpha$ inside high-contrast voids is weakly determined by the surface data, because the surface-temperature signature saturates as the diffusion time $L^2/\alpha$ inside an insulator diverges for $\alpha \to 0$. Preconditioned or multigrid solvers within the differentiable loop are the path to the regime of stiff air voids.

**Material and domain model.** The solver assumes a rectilinear slab with isotropic, temperature-independent effective diffusivity. Non-Cartesian geometries, anisotropy, and temperature-dependent coefficients are natural extensions for deployment beyond this setting. We primarily target specimens whose defects form discrete regions in a host material, where piecewise-constant diffusivity predominantly holds, and NeFTY also recovers irregular blobs and curved strata interfaces (Appendix G.9). Smoothly varying diffusivity, as in functionally graded materials, is more ill-posed and is important future work.

**Volumetric real labels.** The real PVC datasets used in Section 5.3 provide 2D defect masks and discrete defect depths from CAD. No public NDT dataset provides volumetric diffusivity ground truth, because X-ray CT supplies volumetric defect geometry at the time and cost above, and diffusivity labels additionally need material characterization or destructive sectioning (Appendix H). The full volumetric claim therefore rests on the 1,000-sample synthetic benchmark with exact 3D ground truth, generated by a finite-volume engine independent of NeFTY's solver (Appendix E.1), while the real experiments validate 2D segmentation and 2.5D depth recovery under real acquisition. A benchmark that pairs surface thermograms with volumetric diffusivity ground truth would be a standalone materials-science contribution and would enable direct volumetric validation of label-free methods.

## NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The abstract and introduction state our three contributions (hard-constrained neural-field formulation, theoretical account of soft-PINN failure, comprehensive evaluation), each delivered in Sections 4–5 with the headline numbers $0.45/0.37$ synthetic IoU and $0.431$ real-PVC IoU appearing in Tables 1 and 2.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: The conclusion has a dedicated bold-header "Limitation and Broader Impact" paragraph, and Appendix J discusses test-time optimization cost, quantitative recovery under high air-to-solid contrast, and the absence of 3D real ground truth in detail.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.

- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

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

   Answer: [Yes]

   Justification: Proposition 1, Proposition 2, and Corollary 1 state assumptions in their hypotheses; full proofs are in Appendices A, B, and C.

   Guidelines:

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

4. **Experimental result reproducibility**

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

   Answer: [Yes]

   Justification: Appendix D fully specifies the architecture, solver, and adjoint; Appendix D.4 tabulates all hyperparameters; Appendix E documents data generation and metrics; Appendix F gives baseline implementations.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.

- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

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

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

   Answer: [Yes]

   Justification: The code is released through the project page `https://cab-lab-princeton.github.io/nefty/`. The real PVC datasets are public benchmarks [68, 69]. The generated synthetic dataset is hosted at `https://huggingface.co/datasets/n3il666/NeFTY` and can also be regenerated by the released code with PhiFlow [26].

   Guidelines:

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

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

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

   Answer: [Yes]

   Justification: Section 5.1 summarizes the protocol; Appendix D.4 (Table 6) lists every hyperparameter; Appendix E gives data splits and metric definitions; Appendix F details each baseline configuration.

   Guidelines:

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

7. **Experiment statistical significance**

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

   Answer: [Yes]

   Justification: Table 1 and Appendix G (Table 9) report mean $\pm$ 95% confidence intervals across the 1,000-sample synthetic benchmark; Edge F1 is reported as mean $\pm$ standard deviation in Appendix G.

   Guidelines:

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

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: Appendix E.3 documents the hardware (NVIDIA RTX PRO 6000 96 GB and L40 cluster); Table 4 and Table 10 report per-step solver and per-specimen training-level wall-clock and peak memory.

   Guidelines:

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

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: The work involves no human subjects, no scraped personal data, and uses public materials-science benchmarks; it conforms to the NeurIPS Code of Ethics in every respect.

   Guidelines:

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

10. **Broader impacts**

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

    Answer: [Yes]

    Justification: The conclusion discusses the positive impact (safer infrastructure inspection without ionizing X-ray exposure) and the mitigated negative consideration (decision-support framing for safety-critical inspection, with uncertainty in low-sensitivity regions); see also Appendix J.

    Guidelines:

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

11. **Safeguards**

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

    Answer: [N/A]

    Justification: NeFTY is a per-specimen test-time inverse-problem solver, not a generative model or pre-trained model with dual-use risk; no safeguards are required.

    Guidelines:

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

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: PhiFlow [26], the PVC-Infrared and PVC-Depth benchmarks [68, 69], all baseline implementations [9, 24, 28, 38, 53, 59, 66], and PyTorch [46] are cited; their licenses are respected as documented in their original publications.

    Guidelines:

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

13. **New assets**

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

    Answer: [Yes]

    Justification: The released NeFTY codebase (`https://cab-lab-princeton.github.io/nefty/`) ships with documentation covering installation, training, evaluation, and a script that regenerates the synthetic benchmark from PhiFlow. The generated synthetic dataset is hosted at `https://huggingface.co/datasets/n3il666/NeFTY`.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.

- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

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

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

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing or human subjects.

    Guidelines:

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

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

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

    Answer: [N/A]

    Justification: The paper does not involve human subjects, so no IRB approval is required.

    Guidelines:

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

16. **Declaration of LLM usage**

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

    Answer: [N/A]

    Justification: LLMs were not used as part of the core methodology of this research.

    Guidelines:

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