LEARNING PDE-FEASIBLE DESIGN MANIFOLDS FOR SOLVER-AMORTIZED OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

PDE-constrained optimization repeatedly solves a design-dependent state equation. For a class of separable elliptic PDEs, we show that one reference solution induces coefficient–state pairs that preserve compatible source and boundary data and satisfy the residual-inheritance identity $R_\theta = c_\theta R_0$. An exact reference yields a family of exact states, while an approximate reference transfers an explicit residual certificate. This enables hard-constrained design updates from cached fields without a new state solve. Because a fixed family spans only restricted directions, a full-space check at restricted stationarity constructs a bound- and volume-constrained Wendland-RBF candidate from the inaccessible gradient component; one candidate solve and an acceptance gate decide whether to update the reference. Across thermal experiments, analytical updates take $0.12 - 0.14$ s, versus $165 - 493$ s for independent PINN re-solves. In end-to-end physics-informed optimization, our method trains 3 state PINNs rather than 18 and reaches a 51.7% lower FVM-audited resistance. The framework therefore amortizes expensive state solves through residual-certified reuse.

1 INTRODUCTION

PDE-constrained optimization is a fundamental tool for designing physical systems, but its cost is often dominated by repeatedly solving state and adjoint equations as the design-dependent PDE operator changes. This bottleneck is especially severe in topology and material optimization, where hundreds or thousands of updates can each induce a new coefficient field (Hinze et al., 2009; Bendsøe & Sigmund, 2004). The central question here is therefore not merely how to approximate one PDE solution faster, but whether one reference solution–learned or numerical–can support many design updates while retaining a quantitative certificate of PDE feasibility.

Physics-informed neural networks embed differential equations into the training objective and can represent solutions without paired simulation data (Raissi et al., 2019). Neural operators instead learn maps between function spaces and amortize prediction over families of PDE instances (Lu et al., 2021a; Li et al., 2021); physics-informed operator learning additionally uses equation-based supervision (Li et al., 2024). These models, as well as learning-based PDE optimization schemes (Hao et al., 2023; Cheng et al., 2025), can reduce the cost of obtaining approximate states. Yet a standard optimization loop still evaluates a state model whenever the design changes, while freezing a previously computed state under a new coefficient generally violates the governing equation. Fast state prediction therefore does not by itself certify that a sequence of design iterates remains physically admissible.

We instead use a reference potential as a coordinate for an analytically related family of design–state pairs. Let K_0 be a reference potential for a design ρ_0 and set $a_0 = a(\rho_0)$. A potential-conditioned map defines $q_\theta = a(\rho_\theta) = a_0 r_\theta(K_0)$ and pairs it with a normalized state transformation $K_\theta = F_\theta(K_0)$. Under compatible source and boundary data, the construction preserves the prescribed Dirichlet and homogeneous-Neumann data and yields

$$q_\theta \nabla K_\theta = c_\theta a_0 \nabla K_0, \qquad R_\theta = c_\theta R_0.$$

An exact reference consequently induces a nontrivial family of exact solutions; an approximate reference transfers a directly quantifiable residual certificate. Traversing this family uses only cached fields and inexpensive transformations, with no new state solve.

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

A fixed reference family cannot represent every design direction. We therefore optimize many cheap updates on the induced family, then perform a full-space check only at restricted stationarity. The gradient component inaccessible to the current family defines a bound and volume-constrained Wendland-RBF candidate through a strictly convex subproblem. One candidate state solve and physics-based acceptance gates determine whether that candidate becomes the next reference; rejection retains the previous family. The procedure is solver-agnostic, while its amortization is especially valuable when the reference state is supplied by an expensive PINN or physics-informed neural operator.

Our main contributions are:

- We identify a potential-conditioned coefficient–state transformation for separable elliptic PDEs under compatible source and boundary data. An exact reference generates exact solutions, while an approximate reference carries the certificate $R_\theta = c_\theta R_0$.
- We turn this structure into a solver-amortized optimizer that performs many hard-constrained updates from cached fields. A full-space diagnostic detects missing directions, and a constrained Wendland-RBF candidate, one state solve, and an acceptance gate update the reference only when warranted.
- We verify the analytical identities against independent FVM and PINN solutions, evaluate four end-to-end thermal topology benchmarks, and study a fully physics-informed optimization case. Analytical updates require $0.12 - 0.14$ s rather than $165 - 493$ s for PINN re-solves; the physics-informed study uses 3 instead of 18 state-PINN trainings and attains a 51.7% lower FVM-audited resistance.

2 Related Work

Learning PDE solutions and solution operators. Physics-informed neural networks (PINNs) approximate PDE solutions by incorporating governing equations and boundary or observational conditions into the training objective (Raissi et al., 2019; Karniadakis et al., 2021; Lu et al., 2021b). Subsequent work has examined optimization difficulties caused by imbalanced physics-informed losses (Wang et al., 2021). Neural operators instead learn mappings between function spaces, amortizing prediction over families of PDE instances. Representative approaches include DeepONet (Lu et al., 2021a), the Fourier neural operator (Li et al., 2021), and physics-informed neural operators (Li et al., 2024). Recent architectures combine spectral and physical representations or explicitly encode local derivative structure (Yue et al., 2025; Liu et al., 2026). Parameterized PINNs similarly share one learned model across different PDE parameters (Cho et al., 2024). These methods evaluate a trained network to predict the state associated with each new PDE instance.

State-side acceleration of PDE-constrained optimization. Learning-based PDE optimization methods use neural networks to approximate the state, solution operator, or state derivatives required by the optimizer. Hard-constrained PINNs jointly optimize physics-informed states and design variables using penalty or augmented-Lagrangian strategies (Lu et al., 2021c), while bi-level PINNs separate the design objective from PDE-constrained network training and approximate the resulting hypergradients (Hao et al., 2023). Cheng et al. train neural operators and their derivatives on optimization trajectories and combine their predictions with numerical PDE solves (Cheng et al., 2025). Derivative-informed neural operators approximate both states and optimization-relevant derivatives under high-dimensional uncertainty (Luo et al., 2025), whereas BiLO repeatedly trains a local solution operator around the current PDE parameters (Zhang et al., 2026). Reduced-basis methods similarly evaluate a low-dimensional state surrogate during optimization and enrich it according to error estimates or trust-region tests (Hesthaven et al., 2016; Qian et al., 2017; Keil et al., 2021; Kartmann et al., 2024). In these methods, each design update requires evaluating an approximate state or solution operator.

Design-side learning in topology optimization. A separate line of work uses neural networks to represent design variables or predict optimized material layouts (Ramu et al., 2022; Shin et al., 2023). TOuNN uses a neural network to parameterize the material-density field while retaining finite-element state analysis (Chandrasekhar & Suresh, 2021). Prediction-based methods learn mappings from boundary conditions or initial physical fields to near-optimal layouts using convolutional, encoder–decoder, or generative networks (Yu et al., 2019; Nie et al., 2021). PINNTO replaces finite-

element state analysis with a physics-informed network while updating the topology through an explicit optimization procedure (Jeong et al., 2023). Recent variants extend this paradigm to complex nonlinear problems, accelerate state training with Fourier features, couple displacement and density networks, or parameterize topology boundaries directly (Jeong et al., 2025b;a; Singh et al., 2026; Zhou et al., 2026). Other studies analyze how neural density parameterizations alter the optimization landscape (Sanu et al., 2025), use discrete physics-informed networks to solve the state equation in geometrically nonlinear topology optimization (Yin et al., 2025), or learn derivative information online to skip selected finite-element and sensitivity evaluations (Xing & Tong, 2025). Parametric autoencoders have also been used to predict near-optimal topologies before high-fidelity correction (Giacomini & Huerta, 2026). These methods learn a material-field representation, a topology predictor, or a replacement for finite-element analysis. Their acceleration is therefore limited either by the expressiveness and optimization bias of the neural design parameterization or by the accuracy and out-of-distribution generalization of the learned topology or state predictor.

3 METHOD

Given one admissible reference design–state pair (ρ_0, K_0), our key observation is that it induces a parameterized family of new pairs whose compatible source, boundary data, and PDE feasibility are inherited analytically. Moving inside this family therefore requires no repeated state solve. The construction is independent of how K_0 is obtained, but its amortization is especially consequential for physics-informed neural solvers, whose repeated retraining can dominate design cost. We subsequently use the family as a restricted search space and update its reference only when a significant descent direction is not represented. Figure 1 depicts the overall method workflow.

3.1 REFERENCE PROBLEM AND ANALYTICAL STRUCTURE

Let $\rho : \Omega \to [0, 1]$ be a material field. We consider

$$-\nabla \cdot \big(k(\rho, u)\nabla u\big) = f, \qquad k(\rho, u) = a(\rho)b(u), \tag{1}$$

where $a : [0, 1] \to [a_{\min}, a_{\max}]$ is the monotone design-to-coefficient map specified in Eq. (5), and $b(u) > 0$ is the state-dependent factor. The classical Kirchhoff potential

$$K = B(u) := \int_{u_*}^{u} b(s)\,\mathrm{d}s, \qquad \nabla K = b(u)\nabla u, \tag{2}$$

where u_* is a fixed reference state, transforms Eq. (1) into

$$-\nabla \cdot \big(a(\rho)\nabla K\big) = f, \qquad R(\rho, K; f) := -\nabla \cdot \big(a(\rho)\nabla K\big) - f. \tag{3}$$

Let $\partial\Omega = \overline{\Gamma_D \cup \Gamma_N}$, with disjoint Dirichlet and Neumann parts, and write the boundary data abstractly as

$$K = g_D \text{ on } \Gamma_D, \qquad a(\rho)\nabla K \cdot n = g_N \text{ on } \Gamma_N, \tag{4}$$

where n is the outward unit normal. This notation includes multiple constant-temperature terminals, a common sink potential, and homogeneous or nonzero prescribed fluxes. For a monotone interpolation, such as

$$a(\rho) = a_{\min} + (a_{\max} - a_{\min})\rho^p, \qquad p \geq 1, \quad 0 < a_{\min} < a_{\max}, \tag{5}$$

let a^{-1} denote its inverse. On N design points with diagonal quadrature matrix W, hard box and volume feasibility is

$$\mathcal{D} := \{\rho \in [0, 1]^N : \mathbf{1}^\top W\rho = V_0\}. \tag{6}$$

Here V_0 is the prescribed discrete material volume and $\mathbf{1}$ is the vector of ones; norms below are discrete W-weighted $L^2(\Omega)$ norms.

3.2 REFERENCE-INDUCED PDE-FEASIBLE FAMILY

Let $\rho_0 \in [0, 1]^N$ be a reference design, and let K_0 be one reference state for data $(f_0, g_{D,0}, g_{N,0})$. The feasibility result constrains the generated designs ρ_θ to $\mathcal{D}$ but does not require the inducing

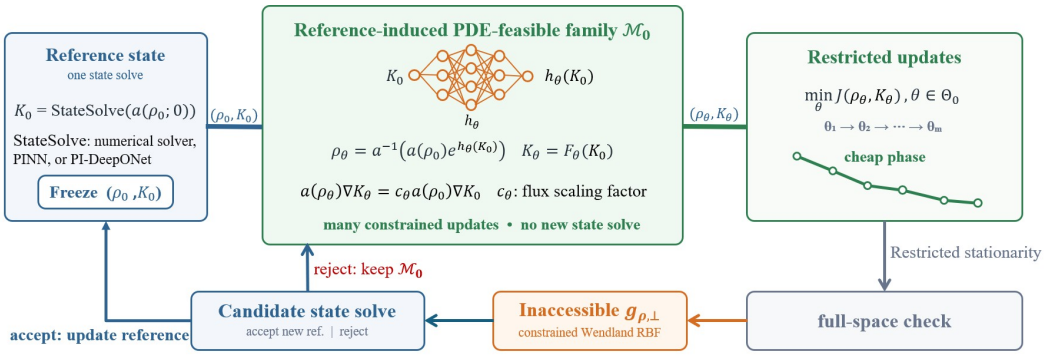

Figure 1: **Reference-induced solver amortization.** One reference induces $\mathcal{M}_0$ for many constrained updates without re-solving. The diagram depicts the source-free regime; the other compatible source and boundary regimes are covered by Theorem 1. At restricted stationarity, one constrained Wendland-RBF candidate is solved and either updates the reference or leaves $\mathcal{M}_0$ unchanged.

reference itself to have volume V_0; Algorithm 1 additionally assumes $\rho_0 \in \mathcal{D}$ so that every identity restart is feasible. Set $a_0 := a(\rho_0)$ and

$$R_0 := R(\rho_0, K_0; f_0). \tag{7}$$

A small network produces a positive scalar modulation

$$r_\theta(s) := \exp(h_\theta(s)) > 0. \tag{8}$$

For parameters whose coefficient remains in the range of a, define

$$q_\theta(\mathbf{x}) := a_0(\mathbf{x}) r_\theta(K_0(\mathbf{x})), \qquad \rho_\theta(\mathbf{x}) := a^{-1}(q_\theta(\mathbf{x})). \tag{9}$$

Thus $q_\theta = a(\rho_\theta)$ is the transformed coefficient, not a second material map. We restrict

$$\Theta_0 := \{\theta : a_{\min} \le q_\theta \le a_{\max},\ \mathbf{1}^\top W \rho_\theta = V_0\}, \tag{10}$$

so every generated design satisfies the hard constraints without post-hoc clipping. Choose $c_\theta > 0$ and a differentiable scalar map F_θ such that

$$F'_\theta(s) = \frac{c_\theta}{r_\theta(s)}, \qquad K_\theta(\mathbf{x}) := F_\theta(K_0(\mathbf{x})). \tag{11}$$

Then the pointwise flux identity is

$$a(\rho_\theta)\nabla K_\theta = c_\theta a(\rho_0)\nabla K_0. \tag{12}$$

Theorem 1 (Reference-induced PDE feasibility)**.** *Assume $r_\theta > 0$, $\theta \in \Theta_0$, and let the transformed problem data be*

$$f_\theta = c_\theta f_0, \qquad g_{D,\theta} = F_\theta(g_{D,0}), \qquad g_{N,\theta} = c_\theta g_{N,0}. \tag{13}$$

If (ρ_0, K_0) solves Eqs. (3)–(4) for $(f_0, g_{D,0}, g_{N,0})$, then (ρ_θ, K_θ) solves the same equations for $(f_\theta, g_{D,\theta}, g_{N,\theta})$. In particular, the original boundary data are preserved whenever

$$F_\theta(g_{D,0}) = g_{D,0} \text{ on } \Gamma_D, \qquad (c_\theta - 1) g_{N,0} = 0 \text{ on } \Gamma_N. \tag{14}$$

Proof. Equation (12) follows from Eq. (11); taking its divergence gives the source relation in Eq. (13), and its trace gives the two boundary relations.

The theorem yields three practically relevant source regimes: *source-free*, $f_0 = f_\theta = 0$; *source-scaled*, $f_\theta = c_\theta f_0$; and *fixed-source*, $f_\theta = f_0$ with $c_\theta = 1$ for nonzero f_0. The first is the zero-source specialization of the second, but is listed separately because it permits nontrivial flux normalization under fixed Dirichlet levels. For two constant levels $K_L < K_R$, boundary compatibility is obtained with

$$c_\theta := \frac{K_R - K_L}{\int_{K_L}^{K_R} r_\theta(s)^{-1}\,\mathrm{d}s}, \qquad F_\theta(s) := K_L + c_\theta \int_{K_L}^{s} r_\theta(t)^{-1}\,\mathrm{d}t. \tag{15}$$

For a common Dirichlet value K_D in the fixed-source regime, take $c_\theta = 1$ and $F_\theta(s) = K_D + \int_{K_D}^{s} r_\theta(t)^{-1}\,\mathrm{d}t$. Homogeneous Neumann data are compatible for every c_θ; fixed nonzero Neumann data require $c_\theta = 1$.

For an approximate reference and arbitrary requested source $\widetilde{f}_\theta$,

$$R(\rho_\theta, K_\theta; \widetilde{f}_\theta) = c_\theta R_0 + c_\theta f_0 - \widetilde{f}_\theta, \qquad R(\rho_\theta, K_\theta; f_\theta) = c_\theta R_0. \tag{16}$$

Thus every compatible regime inherits the residual exactly, while the first identity exposes source mismatch. The family

$$\mathcal{M}_0 := \{(\rho_\theta, K_\theta) : \theta \in \Theta_0\} \tag{17}$$

is design feasible for every reference, exactly PDE feasible for compatible data when $R_0 = 0$, and residual-certified by Eq. (16) otherwise. Under smoothness and constant-rank assumptions, it is locally a manifold. The certificate depends on the accuracy of (ρ_0, K_0), not on the solver type. In our primary instantiation K_0 is learned by a PINN or PI-DeepONet; a conventional discretized solver can also provide the reference.

3.3 Solver-Amortized Search on the Feasible Family

Let $K(\rho; f)$ denote the state returned by the chosen solver and define

$$\min_{\rho \in \mathcal{D}} j_f(\rho), \qquad j_f(\rho) := J(\rho, K(\rho; f)). \tag{18}$$

With (ρ_0, K_0) cached, we instead solve the restricted problem

$$\min_{\theta \in \Theta_0} L(\theta) := J(\rho_\theta, K_\theta). \tag{19}$$

During each restricted solve, the physical source and boundary data are fixed. Source-free problems may use the normalized transform in Eq. (15), whereas nonzero fixed-source problems use $c_\theta = 1$ and the common-Dirichlet construction above. The source-scaled branch of Theorem 1 instead describes a parameterized family of PDE instances and is used only when load scaling is part of the problem definition. Derivatives pass only through r_θ, a^{-1}, the one-dimensional state transform of the selected regime, and cached fields; no state solve or neural-state retraining occurs. We stop when the constrained first-order residual $\chi_{\mathrm{in}}(\theta) \le \varepsilon_{\mathrm{in}}$ or an iteration budget is reached. This establishes stationarity only within $\mathcal{M}_0$.

3.4 Full-Space Diagnosis and Reference Refresh

At the end of a cheap phase, a solver-specific sensitivity routine evaluates the projected full-density gradient mapping of the current reduced objective j_f. Here $\nabla_\rho j_f$ denotes the Riesz gradient under $\langle v, z\rangle_W := v^\top W z$; if $\nabla_{\rho,E} j_f$ is the Euclidean coordinate gradient, then $\nabla_\rho j_f = W^{-1}\nabla_{\rho,E} j_f$.

$$g_\rho := \frac{1}{\gamma_g}\left[\rho - \Pi_{\mathcal{D}}\big(\rho - \gamma_g \nabla_\rho j_f(\rho)\big)\right], \tag{20}$$

where $\gamma_g > 0$ is the diagnostic step size and $\Pi_{\mathcal{D}}$ is the W-orthogonal projection onto $\mathcal{D}$. The mapping uses the current feasible pair $(\rho, K) = (\rho_\theta, K_\theta)$ without a new primal state solve; non-self-adjoint objectives may still require an adjoint/sensitivity evaluation. Let $G_\theta := D_\theta \rho_\theta$ be the density Jacobian. Accounting for the volume constraint and active bounds, the first-order feasible directions represented by the current family form the closed convex cone

$$\mathcal{C}_\theta := \big\{d \in \mathrm{Range}(G_\theta) : \mathbf{1}^\top W d = 0,\ d_i \ge 0 \text{ if } \rho_i = 0,\ d_i \le 0 \text{ if } \rho_i = 1\big\}. \tag{21}$$

Because $-g_\rho$ is the projected feasible descent direction, let $\Pi^W_{\mathcal{C}_\theta}$ denote the W-metric projection onto $\mathcal{C}_\theta$ and define the corresponding gradient-like components

$$g_{\rho,\parallel} := -\Pi^W_{\mathcal{C}_\theta}(-g_\rho), \qquad g_{\rho,\perp} := g_\rho - g_{\rho,\parallel}. \tag{22}$$

Metric projection gives $\langle g_{\rho,\parallel}, g_{\rho,\perp}\rangle_W = 0$; the first component is accessible to the current family, whereas the second is not. After the check, one of three actions is taken. If $\|g_\rho\| \le \varepsilon_{\mathrm{stop}}$, terminate; if $\|g_{\rho,\parallel}\| > \varepsilon_\parallel$, continue the cheap phase. Otherwise the family is restricted-stationary: refresh when $\|g_{\rho,\perp}\| > \tau_{\mathrm{abs}} + \tau_{\mathrm{rel}}\|g_\rho\|$, and terminate when this inequality fails. For reporting only, we use $\eta_{\mathrm{miss}} := \|g_{\rho,\perp}\|/(\|g_\rho\| + \varepsilon_{\mathrm{num}})$ with $\varepsilon_{\mathrm{num}} > 0$; it is not a standalone trigger.

Constrained RBF candidate. We may locally filter $g_{\rho,\perp}$ and place RBF centers where its filtered magnitude or interface uncertainty is high. Using the compactly supported Wendland basis $\psi_m(\mathbf{x}) = \varphi_{\mathrm{W}}(\|\mathbf{x}-\mathbf{z}_m\|_2/\sigma_m)$, where $\varphi_{\mathrm{W}}(r) = (1-r)_+^4(4r+1)$ and $(x)_+ := \max(x, 0)$, set $\Phi_{im} = \psi_m(\mathbf{x}_i)$ and denote the filtered component by $\widetilde{g}_{\rho,\perp}$. The correction solves

$$\begin{aligned} \boldsymbol{\beta}^\star := \arg\min_{\boldsymbol{\beta}} \quad & \widetilde{g}_{\rho,\perp}^\top W\Phi\boldsymbol{\beta} + \frac{\mu_{\mathrm{RBF}}}{2}\|\Phi\boldsymbol{\beta}\|_W^2 + \frac{\lambda_{\mathrm{RBF}}}{2}\|\boldsymbol{\beta}\|_2^2 \\ \text{s.t.} \quad & \mathbf{1}^\top W\Phi\boldsymbol{\beta} = 0, \quad -\rho \le \Phi\boldsymbol{\beta} \le \mathbf{1} - \rho. \end{aligned} \tag{23}$$

Here $\mu_{\mathrm{RBF}} \ge 0$ controls the correction size; $\lambda_{\mathrm{RBF}} > 0$ makes the feasible QP strongly convex and hence uniquely solvable (indeed, $\boldsymbol{\beta} = 0$ is feasible).

One candidate solve and acceptance. Let $\mathrm{StateSolve}(q; f)$ denote the chosen state solver for coefficient q and source f. Set $\rho^+ := \rho + \Phi\boldsymbol{\beta}^\star \in \mathcal{D}$, $a^+ := a(\rho^+)$, and compute the single candidate state $K^+ := \mathrm{StateSolve}(a^+; f)$. Projection continuation may be relaxed during candidate construction, but box and volume constraints remain hard. For a PINN or PI-DeepONet, the previous accepted parameters are warm-started, optionally distilled toward the analytical transformed state, and then physics-informed retrained. Accept (ρ^+, K^+) only if the hard constraints, objective guard, boundary error, residual, and held-out conservation diagnostics pass; otherwise retain the preceding reference and terminate. Every rejected candidate still counts as a state solve.

After acceptance, set $(\rho_0, K_0) \leftarrow (\rho^+, K^+)$ and identity-restart the design map with reinitialized parameters θ^+ by setting $h_{\theta^+} \equiv 0$, $r_{\theta^+} \equiv 1$, $c_{\theta^+} = 1$, and $F_{\theta^+} = \mathrm{Id}$; hence $(\rho_{\theta^+}, K_{\theta^+}) = (\rho_0, K_0)$ and the accepted design–state pair is inherited exactly by the next cheap phase.

3.5 Complete Algorithm

Algorithm 1 Solver-amortized optimization on reference-induced feasible families

Require: $\rho_0 \in \mathcal{D}$, compatible regime and state solver; tolerances, thresholds, N_{in}, P_{stall}, N_{cyc}

$K_0 \leftarrow \mathrm{StateSolve}(a(\rho_0); f)$; $(\rho, K) \leftarrow (\rho_0, K_0)$
Initialize $h_\theta \equiv 0$, $r_\theta \equiv 1$, $c_\theta \leftarrow 1$, and $F_\theta \leftarrow \mathrm{Id}$
for $s = 0, \ldots, N_{\mathrm{cyc}} - 1$ **do**
  Run the cheap optimizer until $\chi_{\mathrm{in}} \le \varepsilon_{\mathrm{in}}$, L stalls for P_{stall} steps, or N_{in} steps are reached
  $(\rho, K) \leftarrow (\rho_\theta, K_\theta)$
  Compute g_ρ, $g_{\rho,\|}$, and $g_{\rho,\perp}$ from the current feasible pair
  if $\|g_\rho\| \le \varepsilon_{\mathrm{stop}}$ **then**
    return ρ
  else if $\|g_{\rho,\|}\| > \varepsilon_{\|}$ **then**
    Continue cheap optimization on the current family without changing (ρ_0, K_0)
  else if $\|g_{\rho,\perp}\| > \tau_{\mathrm{abs}} + \tau_{\mathrm{rel}}\|g_\rho\|$ **then**
    Build ρ^+ by the filtered, adaptive RBF-QP in Eq. (23)
    $K^+ \leftarrow \mathrm{StateSolve}(a(\rho^+); f)$
    if the objective and state-accuracy guards pass **then**
      $(\rho_0, K_0) \leftarrow (\rho^+, K^+)$; $(\rho, K) \leftarrow (\rho^+, K^+)$
      Reset $h_\theta \equiv 0$, $r_\theta \equiv 1$, $c_\theta \leftarrow 1$, and $F_\theta \leftarrow \mathrm{Id}$
    else
      Discard (ρ^+, K^+) and retain (ρ_0, K_0)
      return ρ
    end if
  else
    return ρ
  end if
end for
return ρ

The full-space check reuses the current primal state. Thus the candidate state K^+ above is the only new primal state solve in a refresh; rejected candidates are counted. Optional volume-preserving interface regularization may be applied once to the final design, but is not part of Theorem 1.

4 EXPERIMENTS

The experiments follow the method's three-stage claim structure: analytical feasibility, optimization utility, and learned-solver amortization. Section 4.1 isolates the analytical construction and tests whether transformed design–state pairs retain accurate PDE states without re-solving. Section 4.2 evaluates the complete cheap/check/refresh procedure on thermal topology benchmarks, measuring independently audited design quality and reductions in full state solves, while Section 4.3 replaces the numerical state solver with a PINN and tests whether the same mechanism reduces repeated network retraining without sacrificing audited performance.

4.1 PDE-FEASIBLE STATES WITHOUT RE-SOLVING

Setup. We test the source-free branch of Theorem 1 on a square GH4169 domain with $T = 800\,\mathrm{K}$ on the left, $T = 300\,\mathrm{K}$ on the bottom, and insulated remaining boundaries. A reference PINN trained once at $\rho_0 \equiv 1$ provides K_0; only transformed designs satisfy $\bar{\rho}_\theta = 0.60$. Linear, geometric, and harmonic paths share contrast $C_\alpha = (1+\alpha)/(1-\alpha)$ but differ in state sensitivity. Independent 256^2 FVM solutions are used only for evaluation. Reference and re-solved PINNs share architecture, boundary parameterization, loss, and optimizer family; the reference is trained more accurately because its one-time cost is amortized.

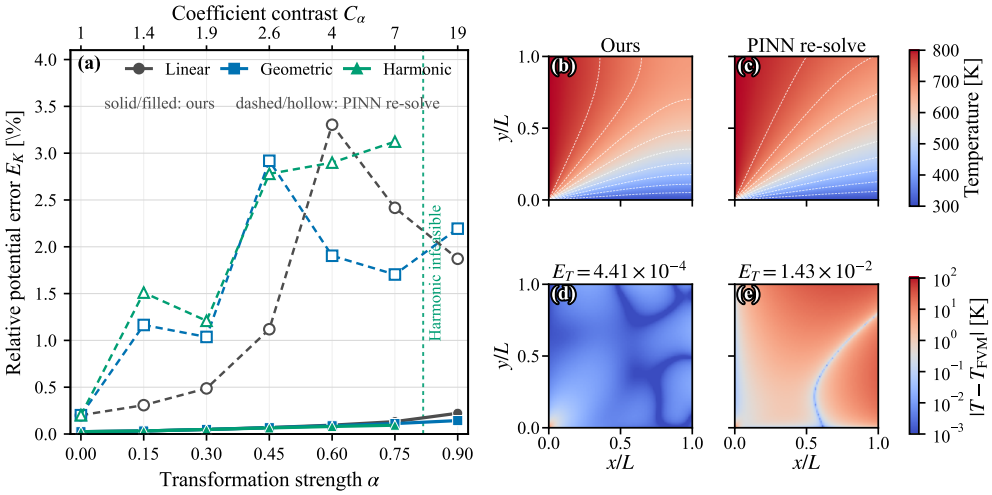

Figure 2: **Accurate PDE states without re-solving.** (a) Relative potential error under increasingly strong linear, geometric, and harmonic coefficient transformations. Solid curves with filled markers denote the analytical transformation; dashed curves with hollow markers connect independent PINN re-solves using the same preassigned seed. The dotted vertical line marks the harmonic feasibility limit under $\bar{\rho} = 0.60$. (b–c) Temperature fields produced by our analytical transformation and an independent PINN re-solve for the harmonic path at $\alpha = 0.75$, shown on a shared temperature scale. (d–e) Their corresponding absolute errors against the finite-volume reference, shown on a shared logarithmic scale.

Accuracy and cost. Figure 2(a) shows that the analytically transformed state remains accurate as the coefficient contrast increases. At $C_\alpha = 19$, the worst errors are 2.22×10^{-3} and 2.19×10^{-2} for ours and PINN re-solves, respectively; the latter reaches 3.30×10^{-2} over the full sweep. The transformed error is not expected to be constant: an approximate reference state carries the nonzero residual R_0. Writing $\widehat{K}_\alpha = F_\alpha(K_0)$ for the analytical state and $K^\star_\alpha$ for the exact transformed state, the map can amplify this fixed error according to

$$\|\widehat{K}_\alpha - K^\star_\alpha\|_{L^2} \le C_{\mathrm{stab}} \mathcal{A}_\alpha \|R_0\|_{H^{-1}}, \qquad \mathcal{A}_\alpha := \|F'_\alpha\|_{L^\infty}. \tag{24}$$

Here C_{stab} is a problem-dependent stability constant. This also explains the observed ordering at high contrast: the harmonic path has the smallest $\mathcal{A}_\alpha$, followed by the geometric and linear paths.

The reference PINN costs 472 s once; each analytical transformation then costs only 0.12–0.14 s, compared with 165–493 s for an independent PINN re-solve.

Field-level verification. Figure 2(b–e) compares temperature fields and absolute errors for the harmonic case at $\alpha = 0.75$. The transformed potential/temperature errors are $9.38 \times 10^{-4}/4.41 \times 10^{-4}$, versus $3.12 \times 10^{-2}/1.43 \times 10^{-2}$ for the re-solved PINN; the corresponding 99th-percentile temperature errors are 0.103 and 28.8 K. The harmonic path is infeasible beyond $\alpha = 0.818$ under the simultaneous density and volume constraints. Complete fields for all paths appear in Appendix A.

4.2 End-to-End Solver-Amortized Topology Optimization

Benchmarks and protocol. We evaluate four thermal benchmarks adapted from Ullah et al. (2021): Examples 1–2 use a central source with four cold edges or discrete corner sink cells, and Examples 3–4 use four interior sources with discrete corner or inset sink cells. Shared sink potential and insulated remainder place these fixed-load cases in the exact fixed-source regime of Theorem 1, with $c_\theta = 1$. All cases use $\bar{\rho} = 0.5$ and SIMP exponent $p = 3$. Within each case, both methods share the grid, initial design, hard constraints, and convergence-based stopping criteria. FVM supplies our initial reference and the candidate states evaluated at refreshes, whereas the baseline re-solves the FVM state after every design update. Solve counts include rejected candidates and final verification, and both final designs receive the same volume-preserving interface regularization.

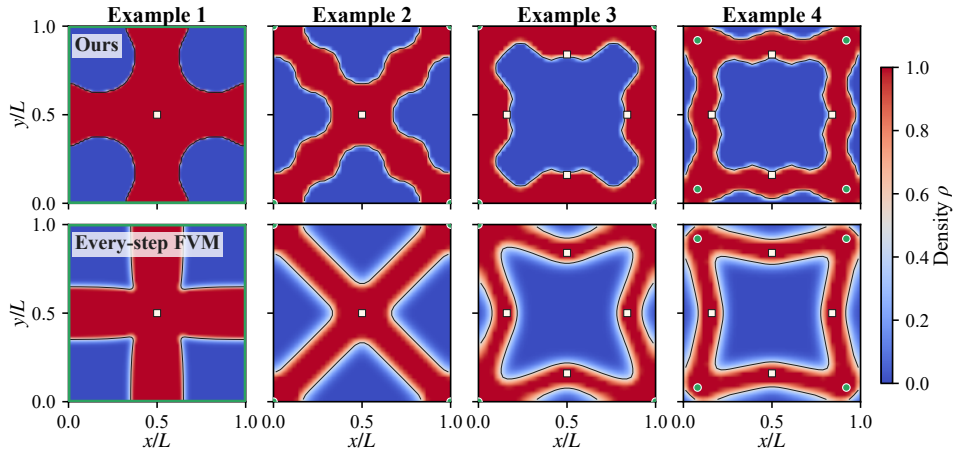

Figure 3: **Final designs across four thermal benchmarks.** Our method (top) and every-step FVM (bottom) recover the same dominant symmetries and source-to-sink routes under $\bar{\rho} = 0.5$. Squares mark sources, circles mark discrete sink cells, and the green boundary in Example 1 is cold.

Topology quality and cost. Across Figure 3, $J_{\text{ours}}/J_{\text{FVM}} = 0.941$–$0.992$ (mean 0.967), with 42 versus 155 full solves (72.9% fewer). Figure 4 separates cheap (blue) and refresh (orange) gains; terminal counts are $13, 9, 9, 11$ versus $27, 27, 54, 47$, with details in Appendix B.

4.3 End-to-End Optimization with a Physics-Informed Solver

Benchmark and protocol. Using Algorithm 1, we study a source-free GH4169 square with eight hot terminals around one cold center and insulated exterior, setting $(T_h, T_c, \bar{\rho}, p, a_{\min}/a_{\max}) = (700\,\text{K}, 300\,\text{K}, 0.40, 3, 0.05)$. Both methods start at $\rho_0 \equiv 0.40$ from the same pretrained PINN, which predicts the Kirchhoff potential and minimizes

$$J_E(\rho, K) = \left[\int_{\Omega_d} a(\rho) |\nabla K|^2 \, \mathrm{d}\mathbf{x} + \varepsilon_E \right]^{-1}, \tag{25}$$

Equation equation 25 is proportional to thermal resistance; nondimensional $\varepsilon_E = 10^{-12}$ stabilizes its inversion. Methods share architecture, optimizer, warm starts, and gates; the baseline retrains each update and ours at refreshes. Only initial/final designs receive FVM audits.

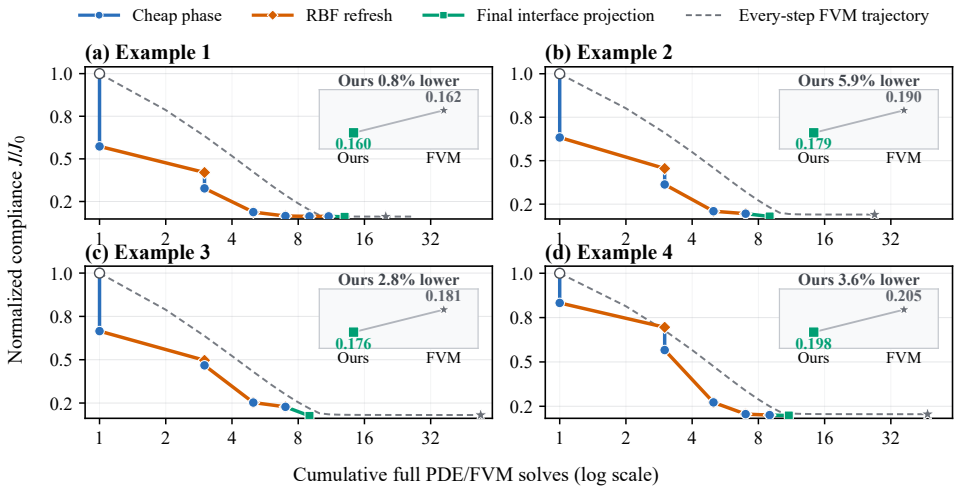

Figure 4: **Solver-amortized objective trajectories.** Colors identify cheap phases, refreshes, final projection, and every-step FVM; stars mark FVM minima.

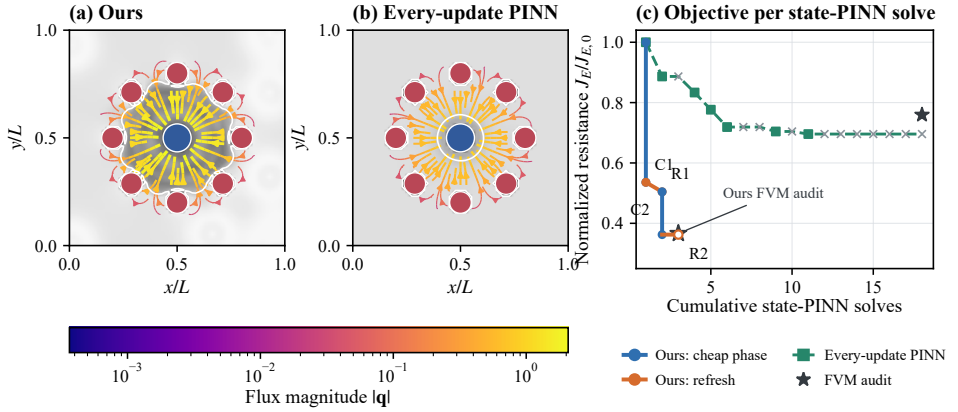

Figure 5: **Physics-informed amortization.** FVM heat-flow maps (a–b) and resistance versus PINN solves (c); markers show rejections and audits.

Dynamics and audit. Cheap phases reduce the PINN objective $2.019 \to 1.082$ and $1.019 \to 0.732$; R1/R2 are accepted/rejected. Ours trains 3 versus 18 PINNs; FVM audits give $J_E/J_{E,0} = 0.3665/0.7596$ (51.7% lower resistance; $2.07\times$ capacity). Relative objective-audit discrepancies are $\varepsilon_J = 3.6\%/10.7\%$ for ours/baseline; see Appendix C.

5 Conclusion

We present a reference-induced PDE-feasible family framework to amortize iterative state-solve overhead in PDE-constrained optimization. This transformation constructs many hard-constrained, residual-verifiable design–state pairs from one high-fidelity reference solution. Full-space diagnostics and constrained RBF refreshes recover feasible optimization directions beyond the current parameterized family's representation capacity. Numerical and physics-informed benchmarks show our approach substantially cuts redundant state evaluations while yielding optimized designs of comparable or improved physically audited quality. Our theoretical formulation, derived for scalar separable elliptic PDEs, lays a rigorous foundation for extensions to broader PDE classes and multi-physics coupled systems. Furthermore, optimized state transfer and refined retraining pipelines can reduce iterative refresh overhead and boost computational efficiency.

REPRODUCIBILITY STATEMENT

Section 3 states the assumptions, transformations, refresh rules, and complete algorithm. Appendices A–D report the problem definitions, material laws, transformation paths, discretizations, essential network and optimizer hyperparameters, stopping and acceptance thresholds, deterministic seeds, timing decomposition, complete solve counts, and independent refined-grid/FVM audits used for the reported claims.

AI USE STATEMENT

The authors conceived the original ideas, designed the study and experiments, developed the methodology, derived the mathematical formulations, interpreted the results, and prepared the manuscript. Generative AI tools were used as assistive tools for literature organization, code debugging, scientific figures, and language refinement. All research decisions, implementations, results, and conclusions were independently verified and finalized by the authors, who retain full responsibility for the work.

ACKNOWLEDGMENTS

The authors wish to acknowledge the project supported by Young Scientists Fund (A Class, Grant No. 52525603) and the financial support of Excellence Research Group Program (ERGP, the former Basic Science Center Program) (Grant No.52488101).

REFERENCES

Martin Philip Bendsøe and Ole Sigmund. *Topology Optimization: Theory, Methods, and Applications*. Springer, Berlin, Heidelberg, 2004. doi: 10.1007/978-3-662-05086-6.

Aaditya Chandrasekhar and Krishnan Suresh. TOuNN: Topology optimization using neural networks. *Structural and Multidisciplinary Optimization*, 63:1135–1149, 2021. doi: 10.1007/s00158-020-02748-4.

Ze Cheng, Zhuoyu Li, Xiaoqiang Wang, Jianing Huang, Zhizhou Zhang, Zhongkai Hao, and Hang Su. Accelerating PDE-constrained optimization by the derivative of neural operators. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 10090–10105, 2025. URL https://proceedings.mlr.press/v267/cheng25f.html.

Woojin Cho, Minju Jo, Haksoo Lim, Kookjin Lee, Dongeun Lee, Sanghyun Hong, and Noseong Park. Parameterized physics-informed neural networks for parameterized PDEs. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 8510–8533, 2024. URL https://proceedings.mlr.press/v235/cho24b.html.

Matteo Giacomini and Antonio Huerta. A surrogate model for topology optimisation of elastic structures via parametric autoencoders. *Computer Methods in Applied Mechanics and Engineering*, 448:118503, 2026. doi: 10.1016/j.cma.2025.118503.

Zhongkai Hao, Chengyang Ying, Hang Su, Jun Zhu, Jian Song, and Ze Cheng. Bi-level physics-informed neural networks for PDE constrained optimization using broyden's hypergradients. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=kkpL4zUXtiw.

Jan S. Hesthaven, Gianluigi Rozza, and Benjamin Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. Springer, 2016. doi: 10.1007/978-3-319-22470-1.

Michael Hinze, René Pinnau, Michael Ulbrich, and Stefan Ulbrich. *Optimization with PDE Constraints*. Springer, Dordrecht, 2009. doi: 10.1007/978-1-4020-8839-1.

Hyogu Jeong, Jinshuai Bai, Chamila Batuwatta-Gamage, Chathurika Rathnayaka, Ying Zhou, and Yajing Gu. A physics-informed neural network-based topology optimization (PINNTO) framework for structural optimization. *Engineering Structures*, 278:115484, 2023. doi: 10.1016/j.engstruct.2022.115484.

Hyogu Jeong, Jinshuai Bai, Chanaka Batuwatta-Gamage, Zachary J. Wegert, Connor N. Mallon, Vivien J. Challis, Yilin Gui, and YuanTong Gu. Fourier feature embedded physics-informed neural network-based topology optimization (FF-PINNTO) framework for geometrically nonlinear structures. *Computer Methods in Applied Mechanics and Engineering*, 446:118244, 2025a. doi: 10.1016/j.cma.2025.118244.

Hyogu Jeong, Chanaka P. Batuwatta-Gamage, Jinshuai Bai, Charith Rathnayaka, Ying Zhou, and Yuantong Gu. An advanced physics-informed neural network-based framework for nonlinear and complex topology optimization. *Engineering Structures*, 322:119194, 2025b. doi: 10.1016/j.engstruct.2024.119194.

George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3:422–440, 2021. doi: 10.1038/s42254-021-00314-5.

Michael Kartmann, Tim Keil, Mario Ohlberger, Stefan Volkwein, and Barbara Kaltenbacher. Adaptive reduced basis trust region methods for parameter identification problems. *Computational Science and Engineering*, 1:3, 2024. doi: 10.1007/s44207-024-00002-z.

Tim Keil, Luca Mechelli, Mario Ohlberger, Felix Schindler, and Stefan Volkwein. A non-conforming dual approach for adaptive trust-region reduced basis approximation of PDE-constrained parameter optimization. *ESAIM: Mathematical Modelling and Numerical Analysis*, 55(3):1239–1269, 2021. doi: 10.1051/m2an/2021019.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Z. Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3), 2024. doi: 10.1145/3648506.

Shouyi Liu, Xiaokang Yang, and Yuntian Chen. Riesz neural operator for solving partial differential equations. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ed3fea9033a80fea1376299fa7863f4a-Abstract-Conference.html.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021a. doi: 10.1038/s42256-021-00302-5.

Lu Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis. DeepXDE: A deep learning library for solving differential equations. *SIAM Review*, 63(1):208–228, 2021b. doi: 10.1137/19M1274067.

Lu Lu, Raphael Pestourie, Wenjie Yao, Zhicheng Wang, Francesc Verdugo, and Steven G. Johnson. Physics-informed neural networks with hard constraints for inverse design. *SIAM Journal on Scientific Computing*, 43(6):B1105–B1132, 2021c. doi: 10.1137/21M1397908.

Dingcheng Luo, Thomas O'Leary-Roseberry, Peng Chen, and Omar Ghattas. Efficient PDE-constrained optimization under high-dimensional uncertainty using derivative-informed neural operators. *SIAM Journal on Scientific Computing*, 47(4), 2025. doi: 10.1137/23M157956X.

Zhenguo Nie, Tong Lin, Haoliang Jiang, and Levent Burak Kara. TopologyGAN: Topology optimization using generative adversarial networks based on physical fields over the initial domain. *Journal of Mechanical Design*, 143(3):031715, 2021. doi: 10.1115/1.4049533.

Elizabeth Qian, Martin A. Grepl, Karen Veroy, and Karen Willcox. A certified trust region reduced basis approach to PDE-constrained optimization. *SIAM Journal on Scientific Computing*, 39(5): S434–S460, 2017. doi: 10.1137/16M1081981.

Maziar Raissi, Paris Perdikaris, and George Em 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. doi: 10.1016/j.jcp.2018.10.045.

Palaniappan Ramu, Pugazhenthi Thananjayan, Erdem Acar, Gamze Bayrak, Jeong Woo Park, and Ikjin Lee. A survey of machine learning techniques in structural and multidisciplinary optimization. *Structural and Multidisciplinary Optimization*, 65(9):266, 2022. doi: 10.1007/s00158-022-03369-9.

Suryanarayanan Manoj Sanu, Alejandro M. Aragon, and Miguel A. Bessa. Neural topology optimization: The good, the bad, and the ugly. *Structural and Multidisciplinary Optimization*, 68(10): 213, 2025. doi: 10.1007/s00158-025-04135-3.

Seungyeon Shin, Dongju Shin, and Namwoo Kang. Topology optimization via machine learning and deep learning: A review. *Journal of Computational Design and Engineering*, 10(4):1736–1766, 2023. doi: 10.1093/jcde/qwad072.

Ajendra Singh, Souvik Chakraborty, and Rajib Chowdhury. A dual physics-informed neural network for topology optimization. *Journal of Computational Physics*, 551:114666, 2026. doi: 10.1016/j.jcp.2026.114666.

Baseer Ullah, Siraj ul Islam, Zahur Ullah, and Wajid Khan. A parametrized level set based topology optimization method for analysing thermal problems. *Computers & Mathematics with Applications*, 99:99–112, 2021. doi: 10.1016/j.camwa.2021.07.018.

Sifan Wang, Yujun Teng, and Paris Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021. doi: 10.1137/20M1318043.

Yi Xing and Liyong Tong. A stochastic gradient online learning and prediction method for accelerating structural topology optimization using recurrent neural network. *Engineering Structures*, 338:120507, 2025. doi: 10.1016/j.engstruct.2025.120507.

Jichao Yin, Shuhao Li, Yaya Zhang, and Hu Wang. An efficient discrete physics-informed neural networks for geometrically nonlinear topology optimization. *Computer Methods in Applied Mechanics and Engineering*, 442:118043, 2025. doi: 10.1016/j.cma.2025.118043.

Yonggyun Yu, Taeil Hur, Jaeho Jung, and In Gwun Jang. Deep learning for determining a near-optimal topological design without any iteration. *Structural and Multidisciplinary Optimization*, 59(3):787–799, 2019. doi: 10.1007/s00158-018-2101-5.

Xihang Yue, Yi Yang, and Linchao Zhu. Holistic physics solver: Learning PDEs in a unified spectral-physical space. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 73877–73896, 2025. URL https://proceedings.mlr.press/v267/yue25e.html.

Ray Zirui Zhang, Christopher E. Miles, Xiaohui Xie, and John S. Lowengrub. BiLO: Bilevel local operator learning for PDE inverse problems. *Journal of Computational Physics*, 551:114679, 2026. doi: 10.1016/j.jcp.2026.114679.

Yuanye Zhou, Zhaokun Wang, Kai Zhou, Hui Tang, and Xiaofan Li. LT-PINN: Lagrangian topology-conscious physics-informed neural network for boundary-focused engineering optimization. *Computer Methods in Applied Mechanics and Engineering*, 448:118453, 2026. doi: 10.1016/j.cma.2025.118453.

A Additional Details for the PDE-Feasibility Experiment

A.1 Benchmark and material model

The physical domain is $\Omega = [0, L]^2$ with $L = 0.04\,\mathrm{m}$. We use the temperature-dependent thermal conductivity of GH4169,

$$\lambda(T) = 5.291 + 0.0152T + 1.382 \times 10^{-6}T^2 \quad [\mathrm{W\,m^{-1}\,K^{-1}}], \tag{26}$$

together with the SIMP coefficient

$$a(\rho) = 10^{-3} + (1 - 10^{-3})\rho^3. \tag{27}$$

Thus this benchmark instantiates Eq. (1) with $u = T$, $b(T) = \lambda(T)$, and the same design map $a(\rho)$ as in Eq. (5). The PINN outputs the Kirchhoff potential

$$K = B(T) := \int_{300\,\mathrm{K}}^{T} \lambda(\tau)\,\mathrm{d}\tau, \tag{28}$$

and is trained with the transformed source-free equation $-\nabla \cdot (a(\rho)\nabla K) = 0$. We impose $T = 800\,\mathrm{K}$ on the left edge, $T = 300\,\mathrm{K}$ on the bottom edge, and homogeneous Neumann conditions on the remaining edges. The incompatible Dirichlet values meet at the lower-left corner, creating a localized corner singularity. Field-error figures therefore use a logarithmic color scale so that a few corner cells do not hide the interior accuracy.

A.2 Transformation families

Let $\kappa_0 = (K_0 - K_L)/(K_R - K_L) \in [0, 1]$. For $0 \le \alpha < 1$, we compare

$$\widehat{r}_\alpha^{\mathrm{lin}}(\kappa) = 1 + \alpha(2\kappa - 1), \tag{29}$$

$$\widehat{r}_\alpha^{\mathrm{geo}}(\kappa) = (1 - \alpha)^{1-\kappa}(1 + \alpha)^{\kappa}, \tag{30}$$

$$\widehat{r}_\alpha^{\mathrm{har}}(\kappa) = \frac{1 - \alpha^2}{1 + \alpha - 2\alpha\kappa}. \tag{31}$$

All three satisfy $\widehat{r}_\alpha(0) = 1-\alpha$ and $\widehat{r}_\alpha(1) = 1+\alpha$, hence they share the contrast $C_\alpha = (1+\alpha)/(1-\alpha)$. A scalar $\gamma_\alpha > 0$ is found by a one-dimensional root solve so that

$$r_\alpha = \gamma_\alpha \widehat{r}_\alpha, \qquad \rho_\alpha = a^{-1}(r_\alpha a(\rho_0)), \qquad |\Omega|^{-1} \int_\Omega \rho_\alpha\,\mathrm{d}\mathbf{x} = 0.60, \tag{32}$$

while preserving $0 \le \rho_\alpha \le 1$. The global factor γ_α enforces the hard volume constraint but cancels from the normalized state map,

$$\kappa_\alpha = F_\alpha(\kappa_0), \qquad F'_\alpha(s) = \frac{c_\alpha}{r_\alpha(s)}, \qquad c_\alpha = \left(\int_0^1 r_\alpha(s)^{-1}\,\mathrm{d}s \right)^{-1}. \tag{33}$$

The three paths isolate transformation sensitivity at an identical endpoint contrast. Their maximum sensitivities are

$$\mathcal{A}_\alpha^{\mathrm{lin}} = \frac{2\alpha}{(1-\alpha)\log C_\alpha}, \qquad \mathcal{A}_\alpha^{\mathrm{geo}} = \frac{(1+\alpha)\log C_\alpha}{2\alpha}, \qquad \mathcal{A}_\alpha^{\mathrm{har}} = 1 + \alpha, \tag{34}$$

with the continuous value 1 at $\alpha = 0$. Thus the harmonic path provides bounded amplification, whereas the linear path becomes increasingly sensitive as $\alpha \to 1$.

A.3 Numerical protocol

The reference PINN and all re-solved PINNs use the same Cartesian-input MLP with four hidden layers of width 64, the same boundary construction, the same residual terms, and the same optimizer family. The one-time reference solve uses 8192 interior points, 6000 Adam steps, and 1000 L-BFGS steps, yielding a relative potential error of 2.48×10^{-4} and a held-out residual RMS of 8.65×10^{-4}. Each independent re-solve uses 4096 interior points, 512 points per boundary, 3000 Adam steps, and 300 L-BFGS steps. This asymmetric budget is deliberate: the more accurate reference is amortized

over every transformed design, whereas a re-solved PINN incurs its cost separately for each design. To avoid post-hoc seed selection, every re-solve in Figure 2 uses the preassigned seed 20260821 exactly once.

All errors are computed against an independent cell-centered 256^2 finite-volume solution with harmonic face interpolation. Refinement over 64^2, 128^2, and 256^2 confirms that the 256^2 field is adequate as the evaluation reference. No finite-volume field is used in PINN training or in the analytical construction of K_α. Relative norms use equal cell-area quadrature weights.

A.4 Complete field comparisons

What the fields verify. Figure 6 complements the error curves in Section 4.1 with complete two-dimensional fields at the strongest reported setting, $\alpha = 0.75$. For every transformation, the finite-volume solution is computed for the transformed density ρ_α; the analytical result is obtained solely from K_0 through F_α; and the re-solved result is produced by an independently trained PINN. Identical temperature limits within each column prevent visual rescaling from concealing discrepancies.

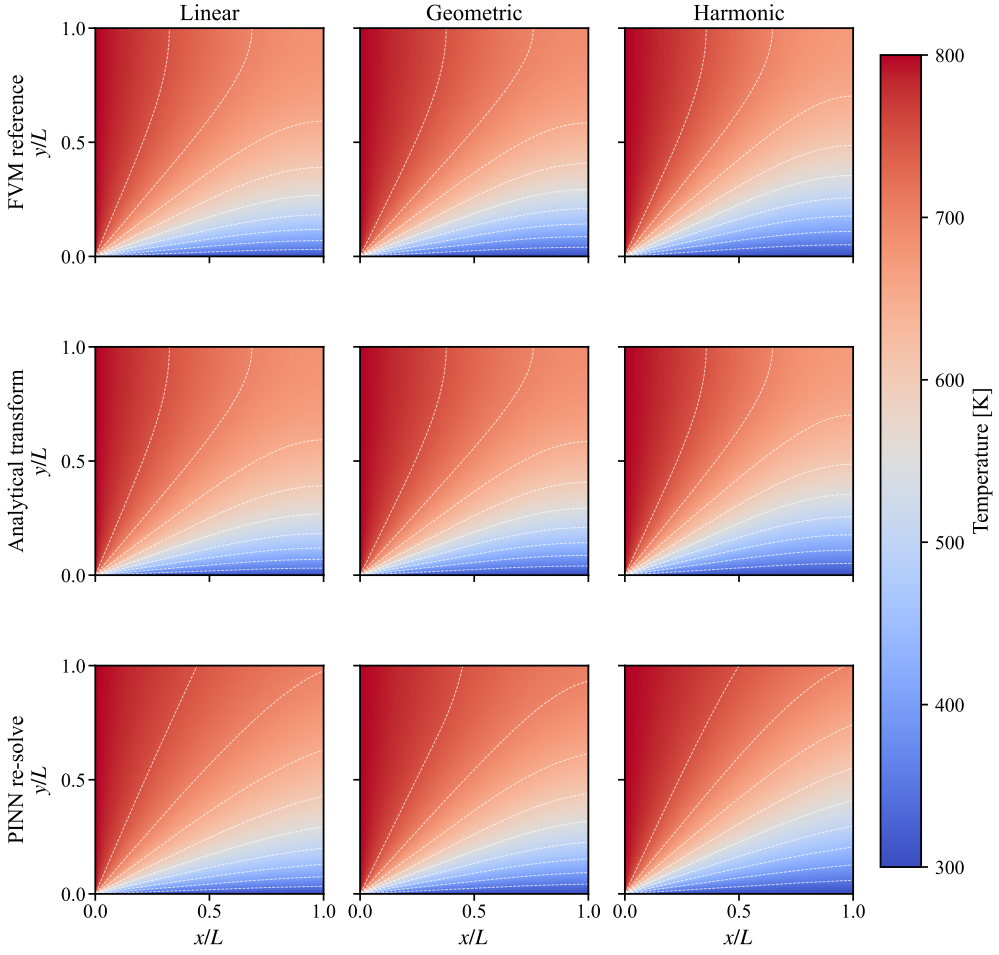

Figure 6: **Complete temperature fields at $\alpha = 0.75$.** Columns correspond to the linear, geometric, and harmonic paths; rows show the independent finite-volume reference, the analytical transformation, and the PINN re-solve. The analytical fields reproduce the reference structure without retraining a state network.

Where the remaining error occurs. Figure 7 reports absolute temperature error on a shared logarithmic scale. The analytical relative errors are 6.70×10^{-4}, 5.41×10^{-4}, and 4.41×10^{-4} for the linear, geometric, and harmonic paths, respectively, compared with 1.09×10^{-2}, 7.87×10^{-3},

and 1.43×10^{-2} for the corresponding PINN re-solves. The analytical error is therefore one to two orders of magnitude smaller in this test. Its mild growth with transformation strength is consistent with propagation of the nonzero reference-PINN error through F_α; it is not an additional PDE-discretization or optimization error. The largest local errors remain concentrated near the incompatible Dirichlet corner and high-sensitivity contours.

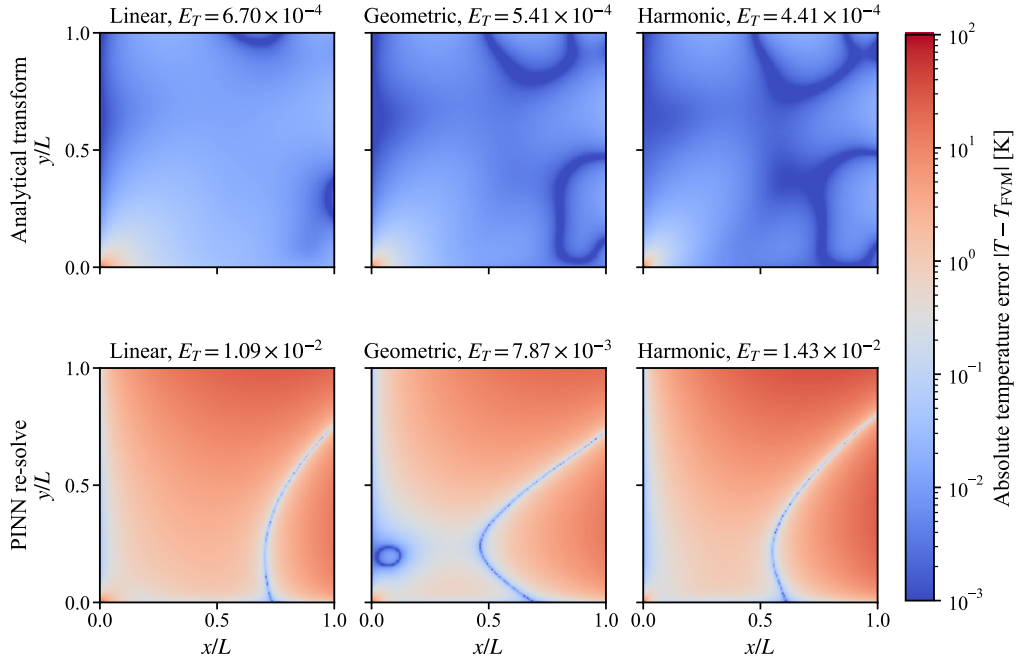

Figure 7: **Absolute temperature errors at $\alpha = 0.75$.** The analytical transformation (top) and PINN re-solve (bottom) are compared with their corresponding finite-volume references. All panels use the same logarithmic error scale.

B Additional Results for the Topology Benchmarks

Complete benchmark definition. For every case, the design domain is the unit square, and the implemented benchmark is the following cell-centred finite-volume problem:

$$\begin{aligned} \min_{\boldsymbol{\rho},\mathbf{K}} \quad & J(\boldsymbol{\rho},\mathbf{K}) = \mathbf{f}^{\mathsf{T}}\mathbf{K}, \\ \text{s.t.} \quad & \mathbf{A}(\boldsymbol{\rho})\mathbf{K} = \mathbf{f}, \qquad K_i = 0 \quad (i \in \mathcal{C}), \\ & 0 \le \rho_i \le 1, \qquad \frac{\mathbf{1}^{\mathsf{T}}\mathbf{M}\boldsymbol{\rho}}{\mathbf{1}^{\mathsf{T}}\mathbf{M}\mathbf{1}} = 0.5. \end{aligned} \tag{35}$$

Here $\mathbf{A}(\boldsymbol{\rho})$ is the finite-volume diffusion matrix with harmonic face conductivities, $\mathbf{f}$ is the fixed integrated source vector, $\mathcal{C}$ is the set of prescribed zero-potential cells or boundary contacts, and $\mathbf{M}$ contains cell areas. The coefficient is $a(\rho) = a_{\min} + (a_{\max} - a_{\min})\rho^p$, with $a_{\max} = 1$, $a_{\min} = 0.01$, and $p = 3$. Thus corner and inset sinks are discrete benchmark contacts, not zero-measure Dirichlet sets in a continuum PDE. Because $\mathbf{f}$ is fixed, this is the discrete analogue of the fixed-source branch of Theorem 1, with $c_\theta = 1$. All design updates are restricted to the D_4-symmetric subspace.

Discretization and common protocol. Each unit integrated source is distributed bilinearly to neighbouring cell centres; every listed discrete contact fixes the nearest cell (or symmetry-tied cells) to $K = 0$, and every unlisted boundary face is insulated. The point locations in Examples 2–4 are our reproducible normalized discretization of the graphical layouts in Ullah et al. (2021), rather than coordinates reported numerically by those authors. Both methods start from $\rho_0 \equiv 0.5$. Each cheap

Table 1: **Complete geometry and loading for the four Section 4.2 benchmarks.** Coordinates are normalized by the unit-square side length.

Case	Grid	Discrete source coordinates	Zero-potential contacts $\mathcal{C}$
Example 1	80×80	$(0.50, 0.50)$	all four boundary edges
Example 2	40×40	$(0.50, 0.50)$	cells nearest $(0,0), (1,0), (1,1), (0,1)$
Example 3	40×40	$(0.50, 0.84), (0.84, 0.50), (0.50, 0.16), (0.16, 0.50)$	corner contacts $(0,0), (1,0), (1,1), (0,1)$
Example 4	40×40	$(0.50, 0.84), (0.84, 0.50), (0.50, 0.16), (0.16, 0.50)$	cells nearest $(0.08, 0.08), (0.92, 0.08), (0.92, 0.92), (0.08, 0.92)$

phase uses the same two-hidden-layer, width-12 ratio network, Adam with initial learning rate 10^{-2} for at most 300 steps, and early stopping after at least 60 steps with patience 30, relative-gradient tolerance 10^{-4}, and relative-objective tolerance 10^{-7}. All four reported runs use seed 20260824. A refresh uses 16 compactly supported Wendland centres at support radii 0.09, 0.17, and 0.30; the full-space sensitivity is filtered with a 1.5-cell Gaussian before the volume- and bound-constrained QP. The every-step FVM baseline applies a density-based optimality-criteria update with move limit 0.055 after each state solve, using a 1.5-cell Gaussian sensitivity filter, D_4 symmetrization, and OC bisection to enforce the box and volume constraints. It stops after at least 25 updates when $\|\rho^{k+1} - \rho^k\|_\infty < 7 \times 10^{-4}$, or at the 80-update budget, followed by one final FVM solve. For the reported comparison, both terminal designs receive the same volume-preserving signed-distance interface regularization (filter radius 0.05 of the side length and 0.5-cell transition width) and are then re-audited by FVM.

Purpose of the comparison. Section 4.2 reports the optimization trajectories and final material layouts. Table 2 provides the complementary numerical audit: final objective, objective ratio, and complete state-solve count for each of the four benchmarks. Objectives are comparable within a row but not across rows because the benchmark geometries and load normalizations differ. Both methods use the same grid, initial density, volume constraint, and final interface regularization; rejected refreshes and the final verification solve are included in our count.

Table 2: **Final objective and complete state-solver cost.** A ratio below one favors our method; the solve count includes rejected candidates and final verification.

Case	J_{ours}	J_{FVM}	$J_{\text{ours}}/J_{\text{FVM}}$	Solves (ours)	Solves (FVM)	Reduction
Example 1	0.7815	0.7876	0.992	13	27	51.9%
Example 2	1.3024	1.3834	0.941	9	27	66.7%
Example 3	12.6273	12.9908	0.972	9	54	83.3%
Example 4	8.6642	8.9922	0.964	11	47	76.6%
Aggregate	–	–	0.967	42	155	72.9%

Result. Our method uses 42 complete state solves rather than 155, a 72.9% aggregate reduction. Its final objective is no larger than the every-step FVM baseline in any benchmark: the row-wise ratios range from 0.941 to 0.992, with mean 0.967. These values support the intended claim—solver calls are amortized without sacrificing the final audited objective—rather than claiming that the restricted manifold must dominate a full-space method on every problem. The trajectories in Figure 4 further separate objective decrease inside cheap phases from the decrease obtained at accepted refreshes.

C Additional Details for the Physics-Informed Optimization Study

C.1 Benchmark and objective

The design grid is 64^2. Eight identical hot circular terminals are evenly distributed on a ring around a central cold terminal, and the outer square boundary is adiabatic. The temperature-dependent GH4169 conductivity and Kirchhoff map are those in Appendix A, now with $T_c = 300\,\mathrm{K}$ and

$T_h = 700\,\mathrm{K}$. The density coefficient and hard volume constraint are

$$a(\rho) = 0.05 + 0.95\rho^3, \qquad |\Omega_d|^{-1} \int_{\Omega_d} \rho \, \mathrm{d}\mathbf{x} = 0.40. \tag{36}$$

For the fixed Dirichlet potential span ΔK, the source-free solution satisfies

$$E(\rho, K) := \int_{\Omega_d} a(\rho) |\nabla K|^2 \, \mathrm{d}\mathbf{x} = Q_{\mathrm{tot}} \Delta K. \tag{37}$$

Consequently $J_E = (E + \varepsilon_E)^{-1}$ is proportional to effective thermal resistance, and the normalized heat-transfer capacity is $(J_E/J_{E,0})^{-1}$. The objective avoids estimating a boundary-normal derivative and can be evaluated from the interior potential gradient used by the PINN loss.

C.2 PINN RETRAINING AND COMPARISON PROTOCOL

The initial and every-update state solvers use the same network, exact Dirichlet construction, residual terms, optimizer family, warm starts, and held-out acceptance diagnostics. After an accepted refresh, the state network is initialized from the preceding accepted PINN, analytically distilled toward the transformed state, and then physics-informed retrained for the new reference density. The design network is identity-restarted as described in Section 3; hence the first density produced in the next cheap phase is exactly the accepted refreshed density. Cheap phases use cached reference fields and do not train or evaluate another state network.

The state PINN has four tanh hidden layers of width 64. The initial reference uses 4096 interior points, 3000 Adam steps at learning rate 10^{-3}, and 150 L-BFGS steps; each candidate retraining uses 2048 interior points, at most 1200 Adam steps at the same learning rate, and 60 L-BFGS steps, with collocation resampling every 250 steps. The ratio network has two hidden layers of width 12 and uses Adam at 8×10^{-3} for at most 240 cheap steps, with patience 45 and relative-objective tolerance 10^{-8}. The refresh objective guard permits a 2% relative increase, and the held-out state gate uses fixed PDE, heat-balance, energy–flux, and Neumann thresholds of 8.0, 0.08, 0.08, and 0.12, respectively; these are acceptance settings rather than additional reported final-error metrics.

The every-update baseline warm-starts the same state PINN after every candidate density step and applies the same objective and state-accuracy guards. It terminates when repeated rejections reduce the density step below its minimum threshold, not after a prescribed number of accepted updates. No FVM field is supplied to either optimizer or PINN. FVM is run only after both procedures terminate, together with one initial reference evaluation needed to normalize resistance. Every solve count includes the initial state and all rejected candidates.

C.3 REFRESH DECISIONS AND COMPLETE SOLVE COUNTS

Table 3 gives the complete two-cycle trajectory. C1 and C2 denote solver-free optimization. Each R event is a current-state sensitivity check followed by one candidate-state retraining. R1 passes the prespecified held-out state-accuracy gate and becomes the next reference. Although R2 has a lower predicted objective, it fails this gate and is rejected; the algorithm therefore retains the R1 reference rather than accepting a numerically attractive but insufficiently accurate state.

Table 3: **Complete event log for the physics-informed run.** Objectives are evaluated by the current PINN; the R2 candidate is not accepted as a new reference.

Event	J_E	$J_E/J_{E,0}$	Decision
Initial	2.0194	1.0000	reference
C1	1.0822	0.5359	cheap phase
R1	1.0187	0.5045	accepted
C2	0.7321	0.3625	cheap phase
R2 candidate	0.6834	0.3384	rejected by state gate

The every-update baseline attempts 17 density updates and trains 18 state PINNs including the initial state; six candidates are accepted and 11 are rejected. Our method trains three state PINNs including

the rejected R2 candidate, an 83.3% reduction in state-network trainings. Both final densities satisfy the hard volume constraint to numerical precision. Counting rejected states is essential: their training cost is paid even though they do not appear in the accepted optimization trajectory.

C.4 Post-hoc objective verification and timing

Table 4 separates independently audited design quality from implementation time using the same scalar metric as the main text. For each final design, we define

$$\varepsilon_J := \frac{\left|J_E^{\mathrm{PINN}} - J_E^{\mathrm{FVM}}\right|}{J_E^{\mathrm{FVM}}}, \tag{38}$$

where J_E^{PINN} is the objective used by the optimizer and J_E^{FVM} is the post-hoc FVM reevaluation. The resulting objective-audit discrepancies are 3.6% for our method and 10.7% for the every-update PINN baseline.

Table 4: **Independent objective audit and computational cost.** FVM is used only after optimization; ε_J compares the PINN objective with its post-hoc FVM reevaluation.

Method	State PINNs	FVM $J_E/J_{E,0}$	ε_J	Time [s]
Ours	3	0.3665	3.6%	562
Every-update PINN	18	0.7596	10.7%	373

All wall-clock times were measured on a single NVIDIA GeForce RTX 5060 Laptop GPU (8 GB) with an Intel Core Ultra 9 285H CPU; shared reference-PINN pretraining and post-hoc FVM audits are excluded. For our method, the measured $562\,\mathrm{s}$ comprise $12.9\,\mathrm{s}$ of cheap optimization, $0.03\,\mathrm{s}$ of full-space sensitivity checks, $57.3\,\mathrm{s}$ of RBF refresh construction, $84.3\,\mathrm{s}$ of analytical distillation, and $406.0\,\mathrm{s}$ of state retraining; the small remainder is data transfer and bookkeeping. Thus the cheap phases and checks are negligible, while reliable refreshed-PINN production is the dominant current cost. The shorter elapsed time of the every-update PINN should not be interpreted as faster successful convergence: 11 of its 17 candidate retrainings reached the 200-step cap without passing the held-out state-accuracy gate, and the run terminated when repeated rejections reduced the design step below its minimum rather than upon reaching design stationarity.

D Cross-Physics Validation on Pressure-Dependent Porous Flow

This appendix tests whether the proposed feasible-family/refresh mechanism is tied to heat conduction. We replace temperature by pressure and consider steady incompressible Darcy flow with pressure-dependent viscosity,

$$\mathbf{v} = -\frac{\chi(\rho)}{\mu(p)}\nabla p, \qquad \nabla\cdot\mathbf{v} = 0, \qquad \mu(p) = \mu_0\exp\big(\beta(p-p_{\mathrm{out}})\big), \tag{39}$$

where $\chi(\rho) = 0.02 + 0.98\rho^3$ is the dimensionless permeability factor. The Kirchhoff pseudo-pressure $K(p) = \int_{p_{\mathrm{out}}}^{p}\mu(s)^{-1}\,\mathrm{d}s$ converts Equation 39 to $-\nabla\cdot(\chi(\rho)\nabla K) = 0$. Thus the same analytical feasible-family construction applies, although the physical variable, constitutive law, geometry, and objective differ from the thermal studies.

Benchmark and protocol. A unit square contains one circular inlet centered at $(0.16, 0.50)$ and two circular outlets centered at $(0.84, 0.27)$ and $(0.84, 0.73)$; the exterior boundary is impermeable. We prescribe $p = p_{\mathrm{in}}$ at the inlet and $p = p_{\mathrm{out}}$ at both outlets, set $\beta(p_{\mathrm{in}} - p_{\mathrm{out}}) = 1.2$, and impose the hard design-volume constraint $|\Omega_d|^{-1}\int_{\Omega_d}\rho\,\mathrm{d}\mathbf{x} = 0.35$. With the normalized pseudo-pressure $\kappa = (K - K_{\mathrm{out}})/(K_{\mathrm{in}} - K_{\mathrm{out}})$, the objective is the dimensionless hydraulic resistance

$$J_{\mathrm{flow}} = \left[\int_{\Omega_d}\chi(\rho)|\nabla\kappa|^2\,\mathrm{d}\mathbf{x} + \varepsilon\right]^{-1}. \tag{40}$$

Both methods start from the same uniform design, use identical SIMP interpolation and volume enforcement, and finish with the same volume-preserving interface projection ($\beta_{\mathrm{proj}} = 48$) and

connectivity cleanup. The cleanup retains the component joining all three terminals, removes disconnected islands, and regrows the same material volume from that component in filtered-density order. Since the active-grid volume equals 1389.5 cells, each terminal design contains 1389 binary material cells and one half-density cell; its grayness is only 2.52×10^{-4}. The proposed method performs seven cheap/check/refresh cycles; the comparison method re-solves the finite-volume state after every design update. Optimization uses a 64^2 grid, while every reported terminal value is independently audited on a 128^2 grid.

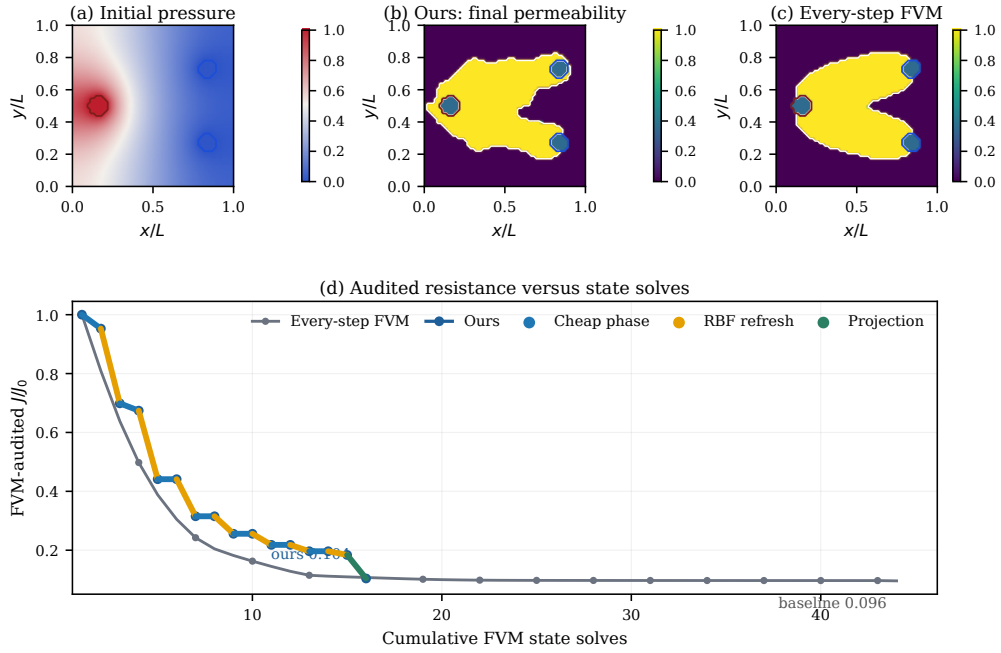

Figure 8: Non-thermal validation on pressure-dependent Darcy flow. (a) Initial physical pressure. (b–c) Near-binary final permeability factors for the proposed method and the every-step FVM baseline; outlined disks denote the inlet and outlets. (d) Independently audited resistance versus cumulative state solves.

Results. Figure 8 shows that the method constructs a connected inlet-to-outlet channel without re-solving the state inside cheap phases. The first two cheap phases reduce the objective before a refresh, while the later checks detect almost purely off-manifold descent directions ($\eta_{\text{miss}} = 0.98$–$1.00$); accepted RBF corrections then produce the visible objective drops. This is precisely the intended role separation between inexpensive restricted updates and occasional full-space corrections.

Table 5: Independent 128^2 audit of the porous-flow experiment. State-solve counts include the initial state, refresh candidates, and final projection verification.

Method	J_{flow}/J_0	State solves	Volume error
Ours	0.096	9	0.0
Every-step FVM	0.104	44	0.0

The proposed method therefore removes 79.5% of the state solves while remaining within 8.0% of the fully re-solved baseline in the independent audit. The material volume is exact, all terminals belong to one material component, and the mass-balance error is below 5×10^{-14} for both final states. We do not use this experiment to claim that a restricted family must dominate an every-step full-space optimizer; rather, it provides a deliberately different-physics check that the analytical state transport, tangent/normal diagnosis, hard constraint, and refresh logic remain operational beyond heat conduction.