AN ENERGY-AWARE SAMPLING METHOD FOR PHYSICS-INFORMED NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Adaptive sampling is a critical technique in Physics-Informed Neural Networks (PINNs), particularly for solving stiff and multiscale partial differential equations (PDEs). Existing adaptive sampling methods predominantly rely on the PDE residual. However, residuals measure local equation violations rather than true approximation errors, frequently neglecting regions with large actual errors, which leads to inefficient resource allocation and slow convergence. To overcome this limitation, we explore an energy-aware sampling method that leverages the intrinsic structure of energy-driven PDEs. Specifically, focusing on phase-field equations as a representative class, our theoretical analysis reveals that the potential not only identifies physically important regions, but also provides a rationale for denser sampling in high-potential regions under suitable assumptions. Motivated by these insights, we propose Potential-based Adaptive Sampling (PAS), which exploits the derivative-free nature of the potential to integrate global exploration with local refinement. Extensive comparisons demonstrate that PAS significantly outperforms residual-based baselines in accuracy, convergence speed, and training efficiency.

1 INTRODUCTION

Energy-driven partial differential equations (PDEs) provide a fundamental framework for modeling complex physical phenomena governed by energy dissipation. Representative examples in the context of interfacial dynamics include the Allen-Cahn and Cahn-Hilliard equations (Allen & Cahn, 1979; Cahn & Hilliard, 1958). These phase-field equations are usually driven by the Ginzburg–Landau free energy, whose gradient and potential terms become asymptotically equipartitioned near the sharp-interface limit (Ilmanen, 1993; Hutchinson & Tonegawa, 2000). Due to their versatility, these models are widely applied in crystal growth (Cheng & Warren, 2008; Yang, 2019), phase transitions (Chen, 2002), and computational biology (Shao et al., 2010).

Since analytic solutions for these nonlinear PDEs are generally unavailable, their simulation relies on numerical methods. However, traditional grid-based solvers (Hughes, 2012; Thomas, 2013) scale poorly in high dimensions (Powell, 2007) and struggle with costly mesh generation. Deep learning approaches (Sirignano & Spiliopoulos, 2018; Long et al., 2019), particularly physics-informed neural networks (PINNs) (Raissi et al., 2019), provide a powerful mesh-free alternative by embedding PDE constraints directly into the loss function. Despite their promise in solving phase-field equations (Wight & Zhao, 2021; Mattey & Ghosh, 2022; 2025; Geng et al., 2025), PINN optimization is sensitive to the spatial distribution of collocation points. This sensitivity is particularly pronounced for the Allen–Cahn and Cahn–Hilliard equations, where thin, moving interfaces and strong numerical stiffness make accurate approximation especially challenging.

Adaptive collocation sampling is widely employed to mitigate such training bottlenecks by dynamically allocating points to complex regions. Most existing methods rely on the PDE residual as an indicator, guiding sampling through residual-induced distributions (Tang et al., 2023; Wu et al., 2023; Jiao et al., 2024; Gao et al., 2025) or residual-driven local refinement (Lu et al., 2021; Luo et al., 2025; Wang & Hu, 2025; Mao & Meng, 2023; Gao et al., 2023). These strategies share a core premise: regions with large residuals dictate where the network struggles most. However, this premise can fail, particularly for energy-driven phase-field equations. The residual misaligns with the true approximation error, failing to pinpoint the physically critical interfaces governed by the potential, for which our theoretical analysis motivates denser sampling under suitable regularity

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.

assumptions. Furthermore, evaluating these residuals requires computing higher-order derivatives of the neural network, imposing computational overhead that significantly reduces training efficiency, especially when scaling to high-dimensional problems.

To overcome the limitations of residual-based approaches, we explore an energy-aware sampling method that directly leverages the intrinsic structure of the governing PDEs. Focusing specifically on phase-field equations, a challenging class of energy-driven PDEs, we use the potential of the energy functional as an alternative indicator to the PDE residual. Our theoretical analysis shows that, under suitable assumptions, high-potential regions admit larger VC-dimension lower bounds for the same approximation accuracy, providing a rationale for denser sampling in these regions.

Motivated by these theoretical analyses, we introduce a Potential-based Adaptive Sampling (PAS) method. By exploiting the derivative-free nature of the potential, PAS efficiently integrates global exploration with local refinement, circumventing the computational bottleneck of higher-order derivative evaluations during sampling.

Main Contributions.

- **Theoretical insight into residual failure**: We establish limitations of the interior PDE residual as an error indicator, demonstrating that under specific conditions, a small residual alone need not imply a small approximation error.
- **Potential as a superior alternative indicator**: We introduce the potential as a simpler replacement for the PDE residual. It captures physically important regions, while our analysis motivates higher sampling budgets in high-potential regions under suitable assumptions.
- **Energy-aware sampling method**: We propose Potential-based Adaptive Sampling (PAS) as a practical realization of this energy-aware method, designed specifically to exploit the derivative-free advantage of the potential indicator.
- **Empirical validation**: Extensive comparisons with state-of-the-art baselines demonstrate that PAS achieves superior accuracy and faster convergence, while avoiding the computational overhead of higher-order residual evaluations during sampling.

The remainder of this paper is organized as follows. Section 2 provides the preliminaries on PINNs, energy-driven PDEs, and residual-based sampling. Section 3 investigates the failure of the residual indicator and motivates the potential as a principled sampling signal. Section 4 presents the proposed PAS method and its theoretical foundation. Finally, Section 5 evaluates the effectiveness of our approach through extensive numerical experiments.

2 Preliminaries

2.1 Physics-informed Neural Networks

Let $\Omega \subseteq \mathbb{R}^d$ be a spatial domain. We consider a time-dependent PDE:

$$\begin{cases} \mathcal{N}u(\mathbf{x},t) = f(\mathbf{x},t), & (\mathbf{x},t) \in \Omega \times (0,T], \\ \mathcal{B}u(\mathbf{x},t) = g(\mathbf{x},t), & (\mathbf{x},t) \in \partial\Omega \times [0,T], \\ \mathcal{I}u(\mathbf{x},0) = h(\mathbf{x}), & \mathbf{x} \in \Omega, \end{cases} \tag{1}$$

where $\mathcal{N}$, $\mathcal{B}$, and $\mathcal{I}$ denote the differential, boundary, and initial operators, respectively. PINNs (Raissi et al., 2019) approximate the exact solution with a neural network $\hat{u}(\mathbf{x}, t; \Theta)$, optimized by minimizing the weighted empirical loss over sets of interior ($\mathcal{S}_f$), boundary ($\mathcal{S}_b$), and initial ($\mathcal{S}_i$) collocation points:

$$\mathcal{L}(\Theta) = \frac{\lambda_{pde}}{|\mathcal{S}_f|} \sum_{\mathcal{S}_f} \|\mathcal{N}\hat{u} - f\|^2 + \frac{\lambda_{bc}}{|\mathcal{S}_b|} \sum_{\mathcal{S}_b} \|\mathcal{B}\hat{u} - g\|^2 + \frac{\lambda_{ic}}{|\mathcal{S}_i|} \sum_{\mathcal{S}_i} \|\mathcal{I}\hat{u} - h\|^2. \tag{2}$$

Because the spatial-temporal distribution of these points dictates which regions are prioritized during training, adaptive sampling becomes critical for stiff or multiscale problems where approximation errors are highly localized.

2.2 Energy-driven PDEs and Phase-field Equations

A broad class of time-dependent PDEs is governed by energy dissipation principles, encompassing Dirichlet energy gradient flows like the heat equation (Evans, 2022) and various reaction-diffusion systems (Mielke, 2011). This energy-centric perspective is crucial for phase-field equations, whose evolution is governed by a free energy functional and involves numerically challenging interfacial dynamics. Crucially, the moving interfaces and sharp transition layers in these models inherently manifest as regions of concentrated energy density. Specifically, we focus on phase-field equations governed by the Ginzburg–Landau free energy:

$$E(u) = \int_\Omega \left(\frac{\varepsilon}{2} |\nabla u|^2 + \frac{F(u)}{\varepsilon} \right) \mathrm{d}\mathbf{x}, \qquad F(u) = \frac{1}{4}(u^2 - 1)^2. \tag{3}$$

The gradient term penalizes spatial variation, while the potential $F(u)$ drives phase separation by favoring the stable states $u = \pm 1$. Their competition forms diffuse interfaces of width $\mathcal{O}(\varepsilon)$. The Allen-Cahn and Cahn-Hilliard equations are derived as the L^2 and H^{-1} gradient flows of this energy, respectively:

$$\partial_t u = \Delta u - \frac{1}{\varepsilon^2} F'(u) \quad \text{(Allen-Cahn)} \quad \partial_t u = \Delta(-\varepsilon \Delta u + \frac{1}{\varepsilon} F'(u)) \quad \text{(Cahn-Hilliard)} \tag{4}$$

In the sharp-interface limit ($\varepsilon \to 0$), Allen-Cahn and Cahn-Hilliard follow curvature-driven and mass-conservative dynamics, respectively (Ilmanen, 1993; Soner, 1997; Hutchinson & Tonegawa, 2000). In addition, the presence of fourth-order spatial derivatives renders the Cahn-Hilliard equation significantly more challenging to solve than the second-order Allen-Cahn equation.

2.3 Adaptive Sampling Methods

Collocation point selection is a critical component of PINN optimization. While random sampling provides a non-adaptive baseline with uniform domain coverage, it fails to prioritize computationally difficult regions. To address this, most adaptive methods leverage the PDE residual as an indicator for difficult regions, which can be broadly divided into two main paradigms.

One primary line of work constructs global sampling probabilities based on residual magnitudes. For example, the residual-based adaptive distribution (RAD) method samples collocation points according to a distribution proportional to the residual (Wu et al., 2023). Similarly, DAS-PINNs models the residual as a probability density function approximated by a deep generative model (Tang et al., 2023), while the Gaussian mixture-based adaptive sampling method captures this distribution via Gaussian mixtures to draw more points from high-residual regions (Jiao et al., 2024).

The second main approach focuses on high-residual regions through local refinement. For instance, RSmote treats high-residual locations as hard examples, employing the SMOTE algorithm to perform local oversampling around these challenging regions (Luo et al., 2025). Other methods in this category include residual-based adaptive refinement (RAR), which iteratively adds points exhibiting the largest residuals (Lu et al., 2021), and R3, which refines the training set by retaining high-residual points while resampling the remainder (Daw et al., 2023).

3 Why Potential Indicators Better Guide PINNs than Residuals

Residual-based sampling methods rely on a core premise: regions with large residuals dictate where the network struggles most. However, this assumption fails under specific conditions, leading to a misalignment between the PDE residual and the actual approximation error. This section explains this fundamental mismatch and explores an alternative indicator that leverages the intrinsic physical structure of the governing PDEs. Using phase-field equations as an example, we motivate the potential as a simpler sampling indicator.

3.1 Empirical Motivation

To illustrate this mismatch, we consider two representative phase-field benchmarks: the 1D Allen–Cahn (107) and Cahn–Hilliard equations (111). Following the general PDE formulation (1) above,

for a PINN approximation $\hat{u}$, we define the pointwise interior PDE residual and the absolute approximation error as

$$R(\mathbf{x}, t) := |\mathcal{N}\hat{u}(\mathbf{x}, t) - f(\mathbf{x}, t)|, \qquad E(\mathbf{x}, t) := |u(\mathbf{x}, t) - \hat{u}(\mathbf{x}, t)|, \tag{5}$$

where $\mathcal{N}$ denotes the corresponding Allen–Cahn or Cahn–Hilliard operator.

As shown in Figure 1, the PDE residual and the actual approximation error exhibit markedly different spatiotemporal distributions for both equations. This discrepancy is also evident in representative time slices: locations with large residuals do not necessarily correspond to large approximation errors, while substantial errors can occur where the residual remains small. These observations demonstrate that the residual alone may provide a misleading signal for adaptive collocation and motivate the use of indicators that better reflect the physically important regions of phase-field dynamics.

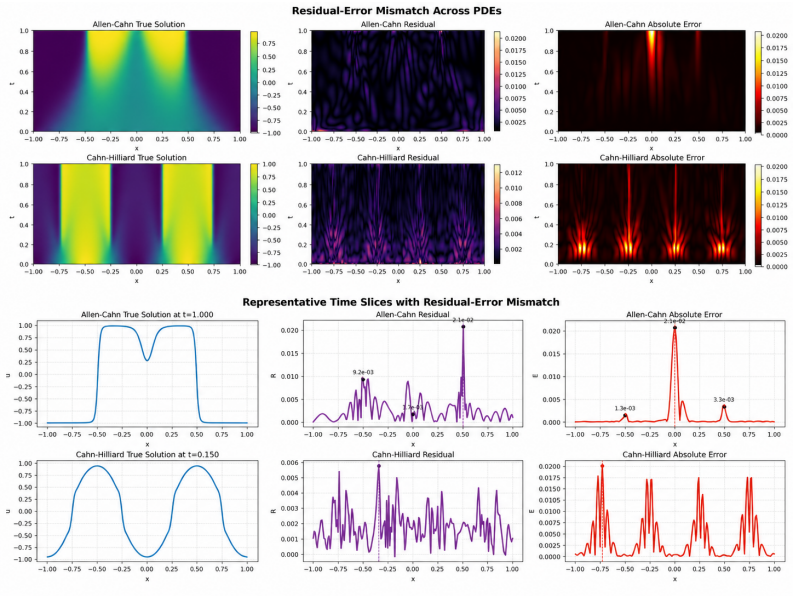


Figure 1: Global spatiotemporal and representative-time comparisons of the residual–error mismatch for the 1D Allen–Cahn and Cahn–Hilliard equations.

3.2 Why Residual Indicators Can Mislead PINNs

We first use a finite-dimensional linear system as a general analogy to demonstrate a fundamental limitation: in ill-conditioned problems, a small residual can coexist with a large approximation error.

Theorem 1 (Residual–error gap, linear case)**.** *Let $A \in \mathbb{R}^{n\times n}$ be nonsingular with smallest singular value σ_n. For any $\epsilon > 0$ and any $M > 0$, whenever $\sigma_n \le \epsilon/M$, there exists an approximation $\tilde{\mathbf{x}}$ of the exact solution to $A\mathbf{x} = \mathbf{b}$ such that*

$$\|R\|_2 = \|\mathbf{b} - A\tilde{\mathbf{x}}\|_2 \le \epsilon \qquad \textit{and} \qquad \|E\|_2 = \|\mathbf{x} - \tilde{\mathbf{x}}\|_2 \ge M. \tag{6}$$

Extending this intuition to phase-field equations, this issue is fundamentally structural: the double-well nonlinearity inherently generates misleading residual values.

Proposition 1 (Residual failure for Allen-Cahn)**.** *Consider the Allen-Cahn equation*

$$\partial_t u = D\, u_{xx} + \lambda(u - u^3) \tag{7}$$

on $[-1, 1] \times [0, 1]$ with $D, \lambda > 0$, and let u^* be a non-constant exact solution. Define

$$\mathcal{N}_{\mathrm{AC}}[\hat{u}] := \partial_t \hat{u} - D\hat{u}_{xx} + \lambda(\hat{u}^3 - \hat{u}). \tag{8}$$

Then the constant functions $\hat{u} \equiv 0, \pm 1$ satisfy $\mathcal{N}_{\mathrm{AC}}[\hat{u}] \equiv 0$ everywhere; however, at any point (x, t), at most one of them can be within distance $1/2$ of the non-constant exact solution $u^*(x, t)$, so at least two zero-residual approximations have pointwise error no less than $1/2$.

Proposition 2 (Residual failure for Cahn-Hilliard)**.** *Define*

$$\mathcal{N}_{\mathrm{CH}}[\hat{u}] := \partial_t \hat{u} - \Delta\left[-\varepsilon\Delta\hat{u} + \frac{1}{\varepsilon}(\hat{u}^3 - \hat{u})\right]. \tag{9}$$

The residual failure described in Proposition 1 holds for $\mathcal{N}_{\mathrm{CH}}$, and manifests even more severely in the Cahn-Hilliard case because of the higher-order derivatives.

The detailed proofs of Theorem 1, Proposition 1 and Proposition 2 are deferred to Appendix A. Theorem 1 establishes that residual smallness alone does not guarantee an accurate approximation. As demonstrated in Propositions 1 and 2, these phase-field equations admit trivial zero-residual states with nonzero approximation errors, without relying on ill-conditioning. These examples concern the interior PDE residual alone, without imposing the prescribed initial and boundary conditions.

3.3 Potential as a Physical Indicator

The mismatch described above motivates a physics-aware indicator based on the double-well potential in Eq. (3), which vanishes at the stable phases $\hat{u} = \pm 1$ and becomes positive in diffuse interfacial regions. This is consistent with the Ginzburg–Landau energy underlying Allen–Cahn and Cahn–Hilliard dynamics: in the sharp-interface regime, the energy concentrates near phase boundaries, with the gradient and potential contributions becoming asymptotically equipartitioned (Ilmanen, 1993; Hutchinson & Tonegawa, 2000). Thus, the potential naturally reflects interfacial activity and distinguishes intermediate phase states from stable phases, providing richer phase-state information than the gradient alone.

A complementary motivation follows from Modica-type gradient estimates, which, under suitable assumptions, bound the spatial variation of the exact solution in terms of its potential (Banerjee & Garofalo, 2015), for example

$$\frac{\varepsilon^2}{2}|\nabla u|^2 \le F(u). \tag{10}$$

Guided by this connection, we use $F(\hat{u})$ as a physically meaningful, interface-sensitive indicator for allocating more collocation points to dynamically important regions.

As shown in Fig. 2, the high-potential regions closely track the diffuse interfaces and exhibit substantially stronger spatial correspondence with the dominant error regions than the PDE residual in the considered benchmarks. Quantitative statistics supporting these observations are provided in Appendix B.

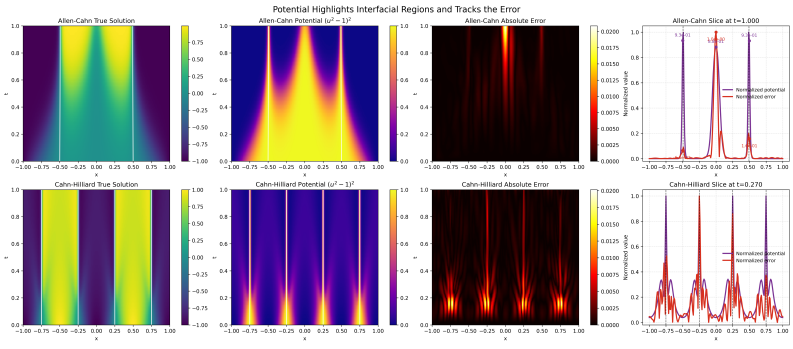


Figure 2: Representative time-slice comparisons among the reference solution, the double-well potential, and the approximation error for the benchmarks. The potential concentrates around the diffuse interfaces and qualitatively aligns with the dominant error regions.

4 Potential-based Adaptive Sampling

Motivated by the theoretical analysis in Section 3, we propose Potential-based Adaptive Sampling (PAS) as a practical realization of the energy-aware sampling method. Focusing primarily on phase-field equations, PAS shifts adaptive sampling from the PDE residual to the predicted double-well

potential. Specifically, PAS integrates this indicator into a probability-based global estimation framework. This substitution yields a significant computational advantage: unlike PDE residuals, which require expensive higher-order automatic differentiation over massive candidate pools, the potential is derivative-free and computed directly from the network's forward pass. Consequently, PAS provides an efficient, low-overhead sampling distribution that naturally biases collocation points toward physically critical regions—such as moving interfaces and rapid transition layers—without sacrificing global domain coverage. We also evaluate PAS on a semilinear elliptic inverse problem with a different potential to assess its applicability beyond the double-well setting.

4.1 ALGORITHM

PAS executes an iterative cycle of potential-guided global exploration and local densification. In each adaptive step, the method first constructs a potential-weighted global dataset from a massive uniform candidate pool. It then isolates high-potential points to serve as centers for localized isotropic perturbation. Finally, it reconstructs a fixed-size training set, integrating refinement of high-potential regions with comprehensive domain coverage.

Potential-based global sampling. In each iteration, we generate a uniform candidate pool $\mathcal{S}_{\text{pool}} = \{\mathbf{x}_i \sim \mathcal{U}(\Omega)\}_{i=1}^{N_{\text{pool}}}$ and compute the sampling weight for each $\mathbf{x} \in \mathcal{S}_{\text{pool}}$:

$$w(\mathbf{x}) = \frac{F(\hat{u}(\mathbf{x}))^k}{\frac{1}{|\mathcal{S}_{\text{pool}}|}\sum_{\mathbf{x}' \in \mathcal{S}_{\text{pool}}} F(\hat{u}(\mathbf{x}'))^k} + c. \tag{11}$$

Following the weighted-sampling formulation used in adaptive collocation methods (Wu et al., 2023), we set $k = 1$ and $c = 1$. We then sample N points without replacement, proportional to these weights, to form the global dataset $\mathcal{D}$. This step ensures comprehensive domain coverage while statistically prioritizing high-potential regions.

Partition. We partition the global dataset $\mathcal{D}$ into high- and low-potential subsets based on a fraction P_1:

$$\mathcal{S}_{\text{High}} = \{\mathbf{x} \in \mathcal{D} \mid F(\hat{u}(\mathbf{x})) \geq \zeta\}, \qquad \mathcal{S}_{\text{Low}} = \mathcal{D} \setminus \mathcal{S}_{\text{High}}, \tag{12}$$

where ζ is the potential value of the $(\lceil P_1 N \rceil)$-th point when all samples in $\mathcal{D}$ are sorted by their potential in descending order. The high-potential points in $\mathcal{S}_{\text{High}}$ are used for subsequent local densification, while $\mathcal{S}_{\text{Low}}$ is preserved as background points for the final dataset reconstruction.

Local perturbation. We allocate new collocation points to each high-potential point $\mathbf{x}_H \in \mathcal{S}_{\text{High}}$ proportionally to its relative potential. The number of assigned points is defined as $n_H = \lceil K F(\hat{u}(\mathbf{x}_H))/F_{\max} \rceil$, where $F_{\max} := \max_{\mathbf{x} \in \mathcal{D}} F(\hat{u}(\mathbf{x}))$ and K is a scaling multiplier. Consequently, the total number of newly generated points is bounded by $N_{\text{new}} \leq KNP_1$.

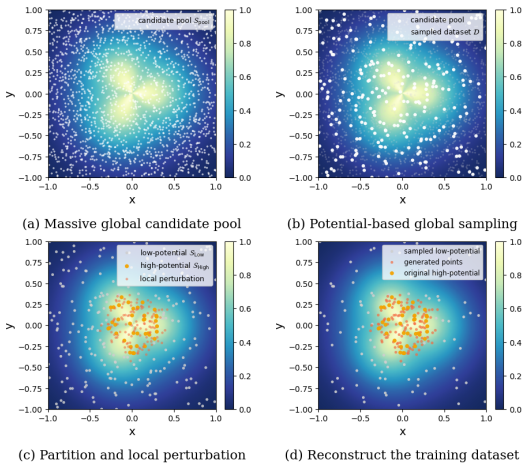
For each high-potential point $\mathbf{x}_H \in \mathcal{S}_{\text{High}}$, we generate n_H local samples via isotropic perturbation:

$$\mathbf{x}_{\text{new},j} = \mathbf{x}_H + j\,\nu\,\frac{\boldsymbol{\xi}_j}{\|\boldsymbol{\xi}_j\|_2}, \qquad \boldsymbol{\xi}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d), \qquad j = 1, \ldots, n_H, \tag{13}$$

where ν is a small fixed value ($\nu = 0.005$) controlling the radius of this slight local exploration. To ensure strict domain compliance, any candidate point generated outside the computational domain Ω is immediately clipped to the boundary.

Dataset reconstruction. To maintain a constant computational footprint, the updated training set of fixed size N is formed by uniting three components: (i) the high-potential points $\mathcal{S}_{\text{High}}$, (ii) the newly generated local perturbations, and (iii) a random subset drawn from $\mathcal{S}_{\text{Low}}$ to fulfill the remaining capacity. This composition ensures comprehensive domain coverage while densely refining high-potential regions.

The complete PAS procedure is summarized in Algorithm 1 and schematically illustrated in Figure 3.



(a) Massive global candidate pool (b) Potential-based global sampling

(c) Partition and local perturbation (d) Reconstruct the training dataset

Figure 3: Schematic overview of the PAS algorithm: (a) generating a uniform candidate pool $\mathcal{S}_{\text{pool}}$; (b) potential-based sampling to extract an intermediate global dataset $\mathcal{D}$; (c) partitioning of $\mathcal{D}$ and perturbing the high-potential points; (d) reconstruction of the final updated training set.

Algorithm 1 Potential-based Adaptive Sampling (PAS)

Require: Initial training set $\mathcal{D}$ of size N; model $\mathcal{M}$; parameters P_1, K, N_{pool}.

Train $\mathcal{M}$ on $\mathcal{D}$.
while not converged **do**
  Generate uniform candidate pool $\mathcal{S}_{\text{pool}}$ of size N_{pool}.
  Compute weights $w(\mathbf{x})$ based on potential via Eq. (11).
  Sample N points proportional to $w(\mathbf{x})$ to form global $\mathcal{D}$.
  Partition $\mathcal{D}$ into $\mathcal{S}_{\text{High}}$ (top P_1 fraction) and $\mathcal{S}_{\text{Low}}$.
  Allocate n_H points to each $\mathbf{x}_H \in \mathcal{S}_{\text{High}}$ proportional to potential.
  Generate local perturbations $\mathcal{S}_{\text{pert}}$ around $\mathcal{S}_{\text{High}}$.
  Reconstruct updated $\mathcal{D}$ of size N via $\mathcal{S}_{\text{High}} \cup \mathcal{S}_{\text{pert}} \cup \text{subset}(\mathcal{S}_{\text{Low}})$.
  Train $\mathcal{M}$ on the updated $\mathcal{D}$.
end while

Ensure: Trained model $\mathcal{M}$.

4.2 Theoretical Analysis of PAS

To motivate allocating a larger sampling budget to high-potential regions, we analyze the regional complexity under a potential-based domain partition, providing a conditional rationale for localized densification.

Without loss of generality, let $\Omega \subset [0,1]^d$. We divide the domain into high- and low-potential regions Ω_H and Ω_L according to a potential threshold. This partition naturally separates the diffuse interfacial regions from the low-potential bulk phases.

To establish this connection, recall that the Ginzburg–Landau energy (3) comprises two components: the potential energy density and the gradient energy density. We respectively define these as:

$$P_\varepsilon(\mathbf{x}) := \frac{1}{\varepsilon} F(u(\mathbf{x})), \qquad G_\varepsilon(\mathbf{x}) := \frac{\varepsilon}{2} |\nabla u(\mathbf{x})|^2. \tag{14}$$

For the Allen-Cahn equation, the discrepancy condition (Ilmanen, 1993),

$$\xi_\varepsilon(\mathbf{x}) := G_\varepsilon(\mathbf{x}) - P_\varepsilon(\mathbf{x}) \le 0, \tag{15}$$

dictates that gradient energy is bounded by potential energy. For the Cahn-Hilliard equation, an analogous relationship emerges asymptotically via energy equipartition near the sharp-interface limit (Hutchinson & Tonegawa, 2000). Furthermore, such potential–gradient control extends beyond classical phase-field models; Modica-type estimates have been established for broad classes of semilinear elliptic and reaction–diffusion equations (Modica, 1987; 1985; Banerjee & Garofalo, 2015). Motivated by this connection, we introduce the following regional approximation assumption for the potential-induced partition.

Assumption 1 (Regional approximation)**.** *For each $i \in \{H, L\}$, let*

$$\mathcal{F}_i = \{f \in W^{s_i,\infty}(\Omega_i) : \|f\|_{W^{s_i,\infty}(\Omega_i)} \le B_i\}, \qquad s_i > 2,$$

where Ω_i contains a cube of side length $\ell > 0$. For a target accuracy $\delta > 0$, assume

$$\sup_{f \in \mathcal{F}_i} \inf_{\phi \in \mathcal{H}_i(\delta)} \|\phi - f\|_{W^{2,\infty}(\Omega_i)} \le \delta. \tag{16}$$

The potential determines the partition, while (s_i, B_i) characterize the regional approximation classes; their ordering is not implied by the potential alone.

Theorem 2 (Regional complexity under potential-based partitioning)**.** *Under Assumption 1, the regional VC complexity satisfies*

$$V_i(\delta) \geq \underline{V}_i(\delta) := c_{s_i,d}\ell^d \left(\frac{B_i}{\delta}\right)^{d/(s_i-2)}, \qquad i \in \{H, L\}, \tag{17}$$

for sufficiently small $\delta > 0$. In particular, if $s_H = s_L = s > 2$ and $B_H > B_L$, then

$$\frac{\underline{V}_H(\delta)}{\underline{V}_L(\delta)} = \left(\frac{B_H}{B_L}\right)^{d/(s-2)} > 1. \tag{18}$$

Under the VC-type generalization certificate specified in Appendix C, this further induces

$$\underline{M}_i(\delta) = C_i B_i^{d/(s_i-2)} \delta^{-d/(s_i-2)-4}, \tag{19}$$

as a certificate-induced complexity scale rather than a necessary statistical sample-complexity lower bound.

The proof is deferred to Appendix C. The result provides a conditional complexity rationale for preferential refinement within the potential-induced partition. A complementary phase-field analysis relating the potential to a sufficient local spatial-resolution scale is provided in Appendix D.

5 Experiments

We comprehensively evaluate the effectiveness of the proposed PAS method, focusing on its accuracy, computational efficiency, and robustness. Performance is quantitatively measured using the mean and standard deviation of the relative L^2 error across five independent runs, alongside GPU memory consumption and total training time. For the inverse problem, the corresponding accuracy measure is the relative error of the recovered coefficient. To demonstrate its superiority, we benchmark PAS against representative sampling strategies, including Random-R, RAR (Lu et al., 2021), RAD (Wu et al., 2023), R3 (Daw et al., 2023) and RSmote (Luo et al., 2025), across a diverse suite of challenging PDEs: 1D/2D Allen-Cahn, 1D Cahn-Hilliard, and 10D/20D reaction-diffusion equations, together with a 1D semilinear elliptic *inverse* problem. Comprehensive experimental settings including specific PDE formulations, potential definitions, and network configurations are detailed in Appendix E.

5.1 Main results

Table 1 and Fig. 4 summarize the overall performance across all benchmark problems. PAS consistently achieves the lowest errors across all six test cases, demonstrating clear superiority over baseline methods. In phase-field equations, high-potential regions naturally align with sharp transition layers, where local refinement is particularly beneficial. This observation is consistent with our theoretical motivation in Section 3 and the region-complexity analysis in Theorem 2. For instance, on the 1D Allen-Cahn equation, PAS reduces the error to 6.81×10^{-4}, significantly outperforming other methods. Furthermore, the convergence trajectories illustrate that PAS not only attains a lower approximation error but also exhibits faster and remarkably stable convergence.

This superiority is pronounced in high-dimensional problems. Unlike RAD, which necessitates expensive residual evaluations over massive candidate pools, PAS leverages the derivative-free potential as a lightweight indicator, thereby circumventing the computational bottleneck of global residual estimation. Consequently, PAS drastically curtails peak GPU memory consumption during global exploration, reducing it from 2446 MB to 906 MB in the 10D case and from 7510 MB to 4388 MB in the 20D case, while simultaneously achieving the highest accuracy and the shortest training time on five of the six benchmarks. These results establish that PAS simultaneously enhances predictive performance and computational efficiency.

5.2 Ablation and Sensitivity Studies

To elucidate the fundamental drivers behind the superiority of PAS, our ablation study isolates the contribution of the sampling indicator from the overall adaptive strategy. We conduct matched

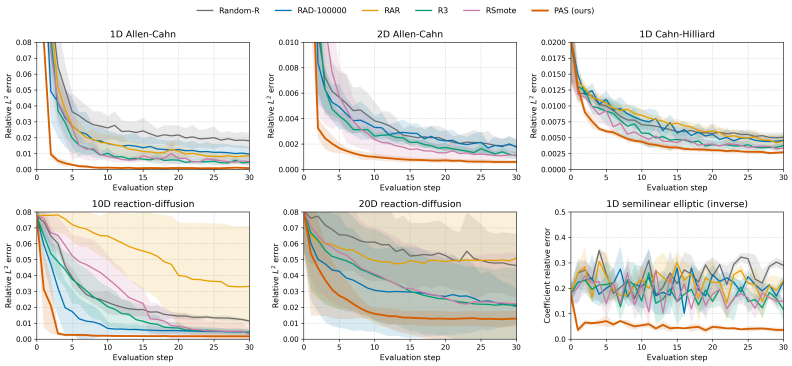


Figure 4: Convergence of adaptive sampling strategies on six benchmarks. Solid curves show five-seed means of relative L^2 errors for forward problems and relative coefficient errors for the inverse problem.

Table 1: Relative errors, GPU memory, and training time on all benchmarks. Best results are in bold. Results are averaged over 5 random seeds.

Problem	Metric	Random-R	RAD-100000	RAR	R3	RSmote	PAS (ours)
1D AC	L^2 error	$1.80\times10^{-2}\pm4.86\times10^{-3}$	$9.62\times10^{-3}\pm4.52\times10^{-3}$	$1.10\times10^{-2}\pm2.28\times10^{-3}$	$4.06\times10^{-3}\pm1.00\times10^{-3}$	$4.52\times10^{-3}\pm1.59\times10^{-3}$	$\mathbf{6.81\times10^{-4}\pm1.35\times10^{-4}}$
	Memory (MB)	710	1458	1458	710	710	**710**
	Time (s)	8036.21	8515.08	10119.64	9212.29	9141.98	**5688.44**
2D AC	L^2 error	$1.80\times10^{-3}\pm5.61\times10^{-4}$	$1.74\times10^{-3}\pm6.27\times10^{-4}$	$1.90\times10^{-1}\pm6.20\times10^{-2}$	$1.07\times10^{-3}\pm2.70\times10^{-4}$	$1.11\times10^{-3}\pm2.81\times10^{-4}$	$\mathbf{5.82\times10^{-4}\pm1.49\times10^{-4}}$
	Memory (MB)	786	2050	2050	786	786	**786**
	Time (s)	17474.02	18188.46	21280.53	17751.78	19319.24	**16426.68**
1D CH	L^2 error	$4.98\times10^{-3}\pm7.92\times10^{-4}$	$4.38\times10^{-3}\pm8.48\times10^{-4}$	$4.18\times10^{-3}\pm1.09\times10^{-3}$	$3.40\times10^{-3}\pm5.86\times10^{-3}$	$3.25\times10^{-3}\pm7.14\times10^{-4}$	$\mathbf{2.57\times10^{-3}\pm2.45\times10^{-4}}$
	Memory (MB)	778	2184	2184	778	778	**778**
	Time (s)	**15479.83**	17662.20	16330.26	17858.29	17220.57	17799.43
10D RD	L^2 error	$1.14\times10^{-2}\pm1.01\times10^{-3}$	$4.40\times10^{-3}\pm1.84\times10^{-3}$	$3.29\times10^{-2}\pm3.79\times10^{-2}$	$4.21\times10^{-3}\pm1.65\times10^{-3}$	$3.15\times10^{-3}\pm2.76\times10^{-3}$	$\mathbf{1.78\times10^{-3}\pm4.48\times10^{-4}}$
	Memory (MB)	906	2446	2446	906	906	**906**
	Time (s)	10001.16	8017.02	14212.79	10268.64	12008.12	**7369.43**
20D RD	L^2 error	$4.63\times10^{-2}\pm1.97\times10^{-2}$	$2.06\times10^{-2}\pm1.09\times10^{-2}$	$4.74\times10^{-2}\pm4.18\times10^{-2}$	$2.13\times10^{-2}\pm2.33\times10^{-2}$	$2.21\times10^{-2}\pm3.18\times10^{-2}$	$\mathbf{1.25\times10^{-2}\pm5.11\times10^{-3}}$
	Memory (MB)	4388	7510	7510	4388	4388	**4388**
	Time (s)	18814.42	17722.53	19015.50	18036.47	20264.33	**17319.03**
1D SE (inverse)	Coefficient rel. error	$2.90\times10^{-1}\pm1.90\times10^{-2}$	$2.27\times10^{-1}\pm2.80\times10^{-2}$	$2.30\times10^{-1}\pm5.60\times10^{-3}$	$1.16\times10^{-1}\pm4.30\times10^{-2}$	$1.50\times10^{-1}\pm1.20\times10^{-2}$	$\mathbf{3.63\times10^{-2}\pm8.40\times10^{-3}}$
	Memory (MB)	696	1400	1400	696	696	**696**
	Time (s)	3953	4023	5037	5043	4550	**3602**

comparisons in which the sampling procedure is held fixed while the indicator is varied among the PDE residual, the gradient, and the potential. As detailed in Appendix F, the potential-based indicator consistently improves accuracy and computational efficiency over residual-based variants, while PAS also substantially outperforms the gradient-based baseline. These results suggest that the gains of PAS cannot be attributed solely to generic gradient-based interface detection.

Furthermore, we investigate the sensitivity of PAS to essential hyperparameters: the partition ratio P_1, the local point generation multiplier K, the global candidate pool size N_{pool}, and the number of collocation points N, as shown in Appendix F. Empirical evaluations reveal that PAS exhibits remarkable stability across a broad spectrum of parameter configurations, consistently outperforming the baseline methods, which validates that PAS is highly robust.

6 Conclusion

This paper investigates an energy-aware sampling method for physics-informed neural networks (PINNs) to tackle a class of challenging PDEs characterized by extreme numerical stiffness and sharp transitions. Our theoretical analysis reveals limitations of the interior PDE residual as an error indicator and provides a rationale for denser sampling in high-potential regions of phase-field equations. Motivated by these insights, we propose PAS. Leveraging the derivative-free nature of the potential, PAS efficiently integrates global exploration with local refinement. Experimental results demonstrate that PAS significantly outperforms state-of-the-art baselines in accuracy, convergence speed, and memory efficiency.

AI USE STATEMENT

In this work, we used generative AI tools to assist with language editing, writing refinement, and LaTeX formatting. We have not used generative AI tools to generate experimental results or independently establish the theoretical claims of this work. Additionally, we used generative AI tools to improve the readability and presentation of the manuscript. We have reviewed and verified all AI-assisted content for correctness and consistency with the original research. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We provide detailed experimental settings, network architectures, hyperparameters, and benchmark configurations in the appendix. All reported results are averaged over five random seeds unless otherwise specified. Complete assumptions and proofs of the theoretical results are also provided in the appendix. These materials are intended to facilitate reproduction of both the experimental and theoretical results. The implementation code will be provided in the supplementary material to further facilitate reproducibility.

REFERENCES

Y. Abu-Mostafa. The vapnik-chervonenkis dimension: Information versus complexity in learning. *Neural Computation*, 1(3):312–317, 1989.

S. M. Allen and J. W. Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metallurgica*, 27(6):1085–1095, 1979.

A. Banerjee and N. Garofalo. Modica type gradient estimates for reaction-diffusion equations. In *Geometric Methods in PDE's*, pp. 215–242. Springer International Publishing, Cham, 2015.

J. P. Boyd. *Chebyshev and Fourier Spectral Methods*. Dover Publications, 2001.

Kevin Buck and Woojeong Kim. Auto-adaptive pinns with applications to phase transitions. *Numerical Mathematics: Theory, Methods and Applications*, 19(2):510–534, 2026.

J. W. Cahn and J. E. Hilliard. Free Energy of a Nonuniform System. I. Interfacial Free Energy. *The Journal of Chemical Physics*, 28(2):258–267, 1958.

L.-Q. Chen. Phase-field models for microstructure evolution. *Annual Review of Materials Research*, 32:113–140, 2002.

M. Cheng and J. A. Warren. An efficient algorithm for solving the phase field crystal model. *Journal of Computational Physics*, 227(12):6241–6248, 2008.

A. Daw, J. Bu, S. Wang, P. Perdikaris, and A. Karpatne. Mitigating propagation failures in physics-informed neural networks using retain-resample-release (R3) sampling. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 2023.

L. Evans. *Partial differential equations*, volume 19. American Mathematical Society, 2022.

L. Gao, D. Zhang, Y. Chen, X. Li, C. Li, and Y. Peng. Global–local adaptive resampling strategy for enhancing the performance of Physics-informed neural networks. *Engineering Applications of Artificial Intelligence*, 158:111283, 2025.

Z. Gao, L. Yan, and T. Zhou. Failure-informed adaptive sampling for PINNs. *SIAM Journal on Scientific Computing*, 45(4):A1971–A1994, 2023.

Y. Geng, O. Burkovska, L. Ju, G. Zhang, and M. Gunzburger. An end-to-end deep learning method for solving nonlocal Allen–Cahn and Cahn–Hilliard phase-field models. *Computer Methods in Applied Mechanics and Engineering*, 436:117721, 2025.

B. Hao, U. Braga-Neto, C. Liu, L. Wang, and M. Zhong. Stability in training PINNs for stiff PDEs: Why initial conditions matter. *arXiv preprint arXiv:2404.16189*, 2025.

T. J. R. Hughes. *The finite element method: linear static and dynamic finite element analysis*. Dover Publications, 2012.

J. E. Hutchinson and Y. Tonegawa. Convergence of phase interfaces in the van der Waals-Cahn-Hilliard theory. *Calculus of Variations and Partial Differential Equations*, 10(1):49–84, 2000.

T. Ilmanen. Convergence of the Allen-Cahn equation to Brakke's motion by mean curvature. *Journal of Differential Geometry*, 38(2):417–461, 1993.

Y. Jiao, D. Li, X. Lu, J. Z. Yang, and C. Yuan. A gaussian mixture distribution-based adaptive sampling method for physics-informed neural networks. *Engineering Applications of Artificial Intelligence*, 135:108770, 2024.

A. K. Kassam and L. N. Trefethen. Fourth-order time-stepping for stiff PDEs. *SIAM Journal on Scientific Computing*, 26(4):1214–1233, 2005.

Chunyan Li, Wenkai Yu, and Qi Wang. Energy dissipation rate guided adaptive sampling for physics-informed neural networks: Resolving surface-bulk dynamics in allen-cahn systems. *International Journal of Numerical Analysis and Modeling*, 23(4):576–604, 2026.

Z. Long, Y. Lu, and B. Dong. PDE-Net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network. *Journal of Computational Physics*, 399:108925, 2019.

L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis. DeepXDE: A deep learning library for solving differential equations. *SIAM Review*, 63(1):208–228, 2021.

J. Luo, Y. Yang, Y. Yuan, S. Xu, and W. Hao. An imbalanced learning-based sampling method for physics-informed neural networks. *Journal of Computational Physics*, 534:114010, 2025.

Z. Mao and X. Meng. Physics-informed neural networks with residual/gradient-based adaptive sampling methods for solving partial differential equations with sharp solutions. *Applied Mathematics and Mechanics*, 44(7):1069–1084, 2023.

R. Mattey and S. Ghosh. A novel sequential method to train physics informed neural networks for Allen Cahn and Cahn Hilliard equations. *Computer Methods in Applied Mechanics and Engineering*, 390:114474, 2022.

R. Mattey and S. Ghosh. Gradient flow based phase-field modeling using separable neural networks. *Computer Methods in Applied Mechanics and Engineering*, 439:117897, 2025.

A. Mielke. A gradient structure for reaction–diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24(4):1329–1346, 2011.

Luciano Modica. A gradient bound and a liouville theorem for nonlinear poisson equations. *Communications on Pure and Applied Mathematics*, 38(5):679–684, 1985.

Luciano Modica. The gradient theory of phase transitions and the minimal interface criterion. *Archive for Rational Mechanics and Analysis*, 98(2):123–142, 1987.

H. T. Nguyen and S. Wang. Quantization of the energy for the inhomogeneous Allen–Cahn mean curvature. *Mathematische Annalen*, 390(4):6399–6457, 2024.

W. B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2007.

M. Raissi, P. Perdikaris, and G. 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.

D. Shao, W.-J. Rappel, and H. Levine. Computational model for cell morphodynamics. *Physical Review Letters*, 105:108104, 2010.

J. Sirignano and K. Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339–1364, 2018.

H. M. Soner. Ginzburg-Landau equation and motion by mean curvature, II: Development of the initial interface. *The Journal of Geometric Analysis*, 7(3):477–491, 1997.

K. Tang, X. Wan, and C. Yang. DAS-PINNs: A deep adaptive sampling method for solving high-dimensional partial differential equations. *Journal of Computational Physics*, 476:111868, 2023.

J. W. Thomas. *Numerical Partial Differential Equations: Finite Difference Methods*, volume 22. Springer Science & Business Media, 2013.

R. Wang and D. Hu. Non-uniform random walk for adaptive sampling. *Journal of Computational Physics*, 538:114160, 2025.

C. L. Wight and J. Zhao. Solving Allen-Cahn and Cahn-Hilliard Equations using the Adaptive Physics Informed Neural Networks. *Communications in Computational Physics*, 29(3):930–954, 2021.

C. Wu, M. Zhu, Q. Tan, Y. Kartha, and L. Lu. A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 403:115671, 2023.

X. Yang. Efficient linear, stabilized, second-order time marching schemes for an anisotropic phase field dendritic crystal growth model. *Computer Methods in Applied Mechanics and Engineering*, 347:316–339, 2019.

Y. Yang, Y. Wu, H. Yang, and Y. Xiang. Nearly Optimal Approximation Rates for Deep Super ReLU Networks on Sobolev Spaces. *arXiv preprint arXiv:2310.10766*, 2025.

APPENDIX

A ANALYSIS OF THE RESIDUAL AND ERROR

We prove Theorem 1 and Propositions 1–2. The proof first establishes a generic linear residual–error gap, and then verifies explicit zero-residual failure modes for the Allen-Cahn and Cahn-Hilliard equations.

Proof of Theorem 1. Let

$$A = U\Sigma V^\top \tag{20}$$

be a singular value decomposition of A, where the singular values satisfy

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0, \tag{21}$$

and let v_n denote the right singular vector associated with the smallest singular value σ_n.

Let $\mathbf{x}$ be the exact solution of $A\mathbf{x} = \mathbf{b}$, and define the approximation

$$\tilde{\mathbf{x}} := \mathbf{x} - Mv_n. \tag{22}$$

Then the approximation error is

$$E = \mathbf{x} - \tilde{\mathbf{x}} = Mv_n. \tag{23}$$

Since v_n is a unit vector, it follows that

$$\|E\|_2 = M. \tag{24}$$

Next, since $A\mathbf{x} = \mathbf{b}$, the residual satisfies

$$\begin{aligned} R &= \mathbf{b} - A\tilde{\mathbf{x}} && (25)\\ &= A\mathbf{x} - A\tilde{\mathbf{x}} && (26)\\ &= AMv_n. && (27) \end{aligned}$$

Using $Av_n = \sigma_n u_n$, where u_n is the corresponding left singular vector, we obtain

$$R = M\sigma_n u_n. \tag{28}$$

Therefore,

$$\|R\|_2 = M\sigma_n. \tag{29}$$

By the assumption $\sigma_n \leq \epsilon/M$, we conclude that

$$\|R\|_2 \leq M \cdot \frac{\epsilon}{M} = \epsilon. \tag{30}$$

Thus $\tilde{\mathbf{x}}$ satisfies

$$\|R\|_2 \leq \epsilon \qquad \text{and} \qquad \|E\|_2 \geq M, \tag{31}$$

which proves the claim. □

Proof of Proposition 1. We prove the two statements separately.

(i) Characterization of constant zero-residual states. Let $\hat{u} \equiv c$ be a constant function. Then

$$\partial_t \hat{u} = 0, \qquad \partial_{xx}\hat{u} = 0, \tag{32}$$

and hence

$$\mathcal{N}_{\mathrm{AC}}[\hat{u}] = \partial_t \hat{u} - D\hat{u}_{xx} + \lambda(\hat{u}^3 - \hat{u}) = \lambda(c^3 - c). \tag{33}$$

Therefore,

$$\mathcal{N}_{\mathrm{AC}}[\hat{u}] \equiv 0 \iff c^3 - c = 0 \iff c(c-1)(c+1) = 0. \tag{34}$$

Thus the only constant zero-residual states are

$$c \in \{0, 1, -1\}. \tag{35}$$

(ii) Pointwise error lower bound among the three trivial states. Fix any point (x, t), and write

$$v := u^*(x, t). \tag{36}$$

Suppose, for contradiction, that two distinct constants $\hat{u}_i, \hat{u}_j \in \{0, 1, -1\}$ both satisfy

$$|v - \hat{u}_i| < \frac{1}{2} \qquad \text{and} \qquad |v - \hat{u}_j| < \frac{1}{2}. \tag{37}$$

Then, by the triangle inequality,

$$|\hat{u}_i - \hat{u}_j| \leq |\hat{u}_i - v| + |v - \hat{u}_j| < 1 \tag{38}$$

However, any two distinct elements of $\{0, 1, -1\}$ are separated by distance at least 1, which is a contradiction. Hence, at most one of the three constants can lie within distance $1/2$ of $u^*(x, t)$.

It follows that at least two of the three trivial zero-residual states must satisfy

$$|\hat{u}_i - u^*(x, t)| \geq \frac{1}{2}. \tag{39}$$

Since (x, t) was arbitrary, the conclusion holds pointwise throughout the domain. □

Proof of Proposition 2. Let $\hat{u} \equiv c$ be *any* constant function. Then $\partial_t \hat{u} = 0$ and $\Delta \hat{u} = 0$. Then

$$\mu = -\varepsilon \Delta \hat{u} + \frac{1}{\varepsilon}(\hat{u}^3 - \hat{u}) = \frac{1}{\varepsilon}(c^3 - c). \tag{40}$$

Because c is a constant, μ is also a spatial constant. Consequently, the outer Laplacian vanishes:

$$\Delta \mu = 0. \tag{41}$$

Therefore, we obtain

$$\mathcal{N}_{\mathrm{CH}}[\hat{u}] = \partial_t \hat{u} - \Delta \mu = 0. \tag{42}$$

Thus, unlike the Allen-Cahn equation where the residual vanishes only for $c \in \{0, \pm 1\}$, *every* constant $c \in \mathbb{R}$ is an exact zero-residual state for the Cahn-Hilliard operator. The residual completely fails to bound the approximation error, and can incur an arbitrarily large pointwise error. □

Remark 1. *This structural divergence explains why residual failure is more severe for Cahn-Hilliard. While Allen-Cahn's pathology relies on the discrete roots of the double-well polynomial, Cahn-Hilliard's outer Laplacian annihilates any constant chemical potential. This renders the PDE residual completely blind to arbitrary global constant shifts.*

B Quantitative statistics of potential

We further provide multi-seed statistical evidence to quantify the relationships between the potential, PDE residual, and pointwise approximation error for the Allen–Cahn and Cahn–Hilliard equations.

We use two complementary metrics. The Spearman rank correlation measures the global monotonic association between each indicator and the pointwise approximation error. The area under the ROC curve (AUC) measures the ability of an indicator to distinguish points belonging to the top 1% of approximation errors from the remaining points within each time slice. The reported values are aggregated over multiple time instances and random seeds and are presented as mean $\pm$ standard deviation.

As shown in Table 2, the potential consistently exhibits a stronger rank association with the pointwise approximation error than the PDE residual for both equations. It also achieves substantially higher AUC values, indicating a greater ability to identify extreme high-error regions. These results provide quantitative empirical evidence that, for the phase-field benchmarks considered here, the potential offers a more informative indicator than the residual for prioritizing regions that are both physically significant and numerically challenging.

C Theoretical Analysis of PAS Sampling

We provide the detailed analysis underlying Theorem 2. We first establish a VC-dimension lower bound for uniform approximation of Sobolev balls and extend it to the potential-induced regional domains. We then prove the main theorem and derive, separately, the complexity scale induced by an applicable VC-based generalization certificate.

Table 2: Statistical comparison between the PDE residual R and double-well potential F as indicators of pointwise approximation error. Higher values indicate stronger correspondence with the error distribution.

PDE	Indicator	Spearman	AUC
Allen–Cahn	Residual (R)	0.336 ± 0.080	0.639 ± 0.037
	Potential (F)	$\mathbf{0.720 \pm 0.083}$	$\mathbf{0.837 \pm 0.058}$
Cahn–Hilliard	Residual (R)	0.339 ± 0.113	0.574 ± 0.059
	Potential (F)	$\mathbf{0.792 \pm 0.022}$	$\mathbf{0.783 \pm 0.015}$

C.1 VC-DIMENSION LOWER BOUND ON SOBOLEV BALLS

We provide the detailed proof of Theorem 2. We first establish a VC-dimension lower bound for uniform approximation of Sobolev balls, extend it to the regional domains induced by the potential partition, and then derive the associated VC-based certificate scale.

Definition 1 (VC dimension; (Abu-Mostafa, 1989))**.** *For a family $\mathcal{H}$ of $\{0,1\}$-valued functions on a set $\mathcal{X}$, the VC dimension $\mathrm{VCdim}(\mathcal{H})$ is the largest integer m such that there exist points $x_1, \dots, x_m \in \mathcal{X}$ shattered by $\mathcal{H}$, i.e., such that $\mathcal{H}$ realizes all 2^m possible labelings on these points. For a real-valued family $\mathcal{G}$, we define*

$$\mathrm{VCdim}(\mathcal{G}) := \mathrm{VCdim}(\mathrm{sgn}(\mathcal{G})). \tag{43}$$

Lemma 1 (Unit-ball VC lower bound; (Luo et al., 2025, Proposition 1))**.** *Let $s, d \in \mathbb{N}_+$ with $s > 2$. Then there exist a small constant $\bar{C}_{s,d} > 0$ and*

$$C_{s,d} = (2\bar{C}_{s,d})^{-d/(s-2)} - 1 \tag{44}$$

such that the following holds. Let $\mathcal{H}$ be a function family on $[0,1]^d$ with $\mathrm{VCdim}(\mathcal{H}) \geq 1$. If

$$\inf_{\phi \in \mathcal{H}} \|\phi - f\|_{W^{2,\infty}([0,1]^d)} \leq \delta, \qquad \forall f \text{ with } \|f\|_{W^{s,\infty}([0,1]^d)} \leq 1, \tag{45}$$

then, for all sufficiently small δ,

$$\mathrm{VCdim}(\mathcal{H}_2) \geq C_{s,d}\, \delta^{-d/(s-2)}, \tag{46}$$

where

$$\mathcal{H}_2 := \{\partial^2_{x_i x_j} h : h \in \mathcal{H},\ 1 \leq i, j \leq d\}. \tag{47}$$

We now derive the corresponding bound for Sobolev balls of radius B.

Lemma 2 (Radius-B VC lower bound)**.** *Let $s, d \in \mathbb{N}_+$ with $s > 2$, let $B > 0$, and define*

$$\mathcal{F}_{s,B} := \left\{ f : \|f\|_{W^{s,\infty}([0,1]^d)} \leq B \right\}. \tag{48}$$

Suppose that a hypothesis class $\mathcal{H}$ satisfies

$$\inf_{\phi \in \mathcal{H}} \|\phi - f\|_{W^{2,\infty}([0,1]^d)} \leq \delta, \qquad \forall f \in \mathcal{F}_{s,B}. \tag{49}$$

Then, for sufficiently small δ,

$$\mathrm{VCdim}(\mathcal{H}_2) \geq C_{s,d}\, B^{d/(s-2)}\, \delta^{-d/(s-2)}. \tag{50}$$

where $C_{s,d}$ is the same constant as defined in Lemma 1.

Proof. **Step 1: construction of a localized bump.** Choose $\tilde{g} \in C^\infty([0,1]^d)$ such that

$$\partial^2_{x_1} \tilde{g}(\mathbf{0}) = 1, \qquad \tilde{g}(\mathbf{x}) = 0 \quad \text{whenever } \|\mathbf{x}\|_2 \geq \frac{1}{3}. \tag{51}$$

We can find a constant $\bar{C}_{s,d} > 0$ such that if we set

$$g := \frac{\tilde{g}}{\bar{C}_{s,d}}, \tag{52}$$

it satisfies

$$\|g\|_{W^{s,\infty}([0,1]^d)} \le 1, \qquad \partial_{x_1}^2 g(\mathbf{0}) = \frac{1}{\bar{C}_{s,d}}. \tag{53}$$

Step 2: scaled copies on a grid. Fix an integer $M \ge 1$. Partition $[0,1]^d$ into M^d cubes of side length $1/M$, and let x_θ denote the center of the cube indexed by $\theta \in \{1,\dots,M\}^d$. Define

$$g_\theta(\mathbf{x}) := g\big(M(\mathbf{x} - x_\theta)\big). \tag{54}$$

Because g is supported in the Euclidean ball of radius $1/3$, each g_θ is supported in a ball of radius $1/(3M)$ centered at x_θ. Since neighboring cube centers are separated by distance at least $1/M$, these supports are pairwise disjoint.

Step 3: a family of test functions with prescribed signs. For each sign pattern

$$\beta : \{1,\dots,M\}^d \to \{-1,+1\}, \tag{55}$$

define

$$h_\beta^B(\mathbf{x}) := B \sum_\theta M^{-s} \beta(\theta)\, g_\theta(\mathbf{x}). \tag{56}$$

Because the supports are disjoint, at each point $\mathbf{x}$ only one summand is active. Moreover, by the chain rule,

$$D^{\boldsymbol{\alpha}} g_\theta(\mathbf{x}) = M^{|\boldsymbol{\alpha}|} \big(D^{\boldsymbol{\alpha}} g\big)\big(M(\mathbf{x} - x_\theta)\big). \tag{57}$$

Hence, for every multi-index $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \le s$,

$$|D^{\boldsymbol{\alpha}} h_\beta^B(\mathbf{x})| \le B\, M^{-s+|\boldsymbol{\alpha}|} \|g\|_{W^{s,\infty}} \le B. \tag{58}$$

Therefore,

$$\|h_\beta^B\|_{W^{s,\infty}([0,1]^d)} \le B, \tag{59}$$

so $h_\beta^B \in \mathcal{F}_{s,B}$.

Step 4: signal size at the grid centers. Fix one center x_{θ_0}. Since the supports are disjoint, all terms vanish at x_{θ_0} except the one indexed by θ_0. Therefore,

$$\partial_{x_1}^2 h_\beta^B(x_{\theta_0}) = B\, M^{-s} \beta(\theta_0)\, \partial_{x_1}^2 g_{\theta_0}(x_{\theta_0}). \tag{60}$$

Using the scaling relation,

$$\partial_{x_1}^2 g_{\theta_0}(x_{\theta_0}) = M^2\, \partial_{x_1}^2 g(\mathbf{0}), \tag{61}$$

we obtain

$$\partial_{x_1}^2 h_\beta^B(x_{\theta_0}) = B\, M^{-s+2} \beta(\theta_0) \cdot \frac{1}{\bar{C}_{s,d}}. \tag{62}$$

Thus the second derivative at each center has the prescribed sign $\beta(\theta_0)$, with magnitude

$$\big|\partial_{x_1}^2 h_\beta^B(x_{\theta_0})\big| = \frac{B\, M^{-s+2}}{\bar{C}_{s,d}}. \tag{63}$$

Step 5: approximation by the hypothesis class. By assumption, for each $h_\beta^B \in \mathcal{F}_{s,B}$, there exists $\phi_\beta \in \mathcal{H}$ such that

$$\|\phi_\beta - h_\beta^B\|_{W^{2,\infty}([0,1]^d)} \le \delta. \tag{64}$$

Define

$$p_\beta := \partial_{x_1}^2 \phi_\beta \in \mathcal{H}_2. \tag{65}$$

Then

$$|p_\beta(\mathbf{x}) - \partial_{x_1}^2 h_\beta^B(\mathbf{x})| \le \delta, \qquad \forall\, \mathbf{x} \in [0,1]^d. \tag{66}$$

Step 6: preservation of signs. To guarantee that p_β has the same sign as $\partial_{x_1}^2 h_\beta^B$ at the grid centers, it is enough to ensure

$$\frac{B\, M^{-s+2}}{\bar{C}_{s,d}} \ge 2\delta. \tag{67}$$

Equivalently,

$$M \le \left(\frac{B}{2\bar{C}_{s,d}\delta}\right)^{1/(s-2)}. \tag{68}$$

We therefore choose

$$M := \left\lfloor \left(\frac{B}{2\bar{C}_{s,d}\delta}\right)^{1/(s-2)} \right\rfloor. \tag{69}$$

With this choice, for every center x_θ,

$$\operatorname{sgn}\big(p_\beta(x_\theta)\big) = \operatorname{sgn}\big(\partial_{x_1}^2 h_\beta^B(x_\theta)\big) = \beta(\theta). \tag{70}$$

Step 7: shattering. The previous step shows that the family $\mathcal{H}_2$ realizes all 2^{M^d} sign patterns on the M^d centers $\{x_\theta\}$. Hence these points are shattered by $\mathcal{H}_2$, and therefore

$$\operatorname{VCdim}(\mathcal{H}_2) \ge M^d. \tag{71}$$

Substituting the above choice of M yields

$$\operatorname{VCdim}(\mathcal{H}_2) \ge C_{s,d}\, B^{d/(s-2)}\, \delta^{-d/(s-2)} \tag{72}$$

for all sufficiently small δ, which proves the claim. □

Lemma 3 (Regional VC lower bound)**.** *Let $U \subset (0,1)^d$ contain an axis-aligned cube of side length $\ell > 0$, and let*

$$\mathcal{F}_{s,B}(U) := \{f \in W^{s,\infty}(U) : \|f\|_{W^{s,\infty}(U)} \le B\}, \qquad s > 2.$$

If a hypothesis class $\mathcal{H}$ satisfies

$$\sup_{f \in \mathcal{F}_{s,B}(U)} \inf_{\phi \in \mathcal{H}} \|\phi - f\|_{W^{2,\infty}(U)} \le \delta, \tag{73}$$

then, for sufficiently small $\delta > 0$,

$$\operatorname{VCdim}(\mathcal{H}_2) \ge c_{s,d}\ell^d \left(\frac{B}{\delta}\right)^{d/(s-2)}, \tag{74}$$

where $c_{s,d} > 0$ depends only on s and d.

Proof. Let $Q = a + [0,\ell]^d \subset U$. Using the compactly supported bump construction from Lemma 2, define its rescaled version on Q by

$$h_\beta^{B,\ell}(x) := B\ell^s h_\beta\left(\frac{x-a}{\ell}\right), \tag{75}$$

and extend it by zero outside Q. Since the underlying bumps are compactly supported in the interior of the cube, this extension belongs to $W^{s,\infty}(U)$ and satisfies

$$\|h_\beta^{B,\ell}\|_{W^{s,\infty}(U)} \le B. \tag{76}$$

At the corresponding rescaled grid points x_θ,

$$\partial_{x_1}^2 h_\beta^{B,\ell}(x_\theta) = a_{s,d} B\ell^{s-2} M^{-(s-2)} \beta(\theta), \tag{77}$$

for a constant $a_{s,d} > 0$ depending only on s and d.

Choose M such that

$$a_{s,d} B\ell^{s-2} M^{-(s-2)} > 2\delta. \tag{78}$$

The uniform approximation assumption then implies that, for every sign pattern β, there exists $\phi_\beta \in \mathcal{H}$ whose second derivatives realize the prescribed signs at all M^d grid points. Hence,

$$\operatorname{VCdim}(\mathcal{H}_2) \ge M^d \ge c_{s,d}\ell^d \left(\frac{B}{\delta}\right)^{d/(s-2)} \tag{79}$$

for sufficiently small $\delta > 0$. □

C.2 Proof of the main theorem and induced complexity scale

We now prove Theorem 2. For each $i \in \{H, L\}$, define

$$V_i(\delta) := \mathrm{VCdim}((\mathcal{H}_i(\delta))_2)\,, \tag{80}$$

where $(\mathcal{H}_i(\delta))_2$ denotes the class of second-order partial derivatives of functions in $\mathcal{H}_i(\delta)$.

Proof of Theorem 2. Fix $i \in \{H, L\}$. By Assumption 1, $\mathcal{H}_i(\delta)$ uniformly approximates the Sobolev ball $\mathcal{F}_i$ in $W^{2,\infty}(\Omega_i)$ with accuracy δ. Since Ω_i contains an axis-aligned cube of side length ℓ, the regional VC lower bound established in Lemma 3 gives

$$V_i(\delta) \geq \underline{V}_i(\delta) := c_{s_i,d}\ell^d \left(\frac{B_i}{\delta}\right)^{d/(s_i-2)} \tag{81}$$

for sufficiently small $\delta > 0$.

If $s_H = s_L = s > 2$ and $B_H > B_L$, the constants $c_{s_H,d}$ and $c_{s_L,d}$ coincide. Hence,

$$\frac{\underline{V}_H(\delta)}{\underline{V}_L(\delta)} = \left(\frac{B_H}{B_L}\right)^{d/(s-2)} > 1, \tag{82}$$

which proves the stated regional complexity comparison. □

Certificate-induced complexity scale. We next relate the above approximation-complexity bound to a VC-based generalization certificate. Let $\mathcal{R}_{D,i}$ and $\mathcal{R}_{S,i}$ denote the continuous and empirical regional losses, respectively, and let M_i denote the number of samples in region Ω_i.

Lemma 4 (Generalization certificate; (Yang et al., 2025))**.** *For a fully connected deep-network hypothesis class $\mathcal{H}_i(\delta)$, suppose the applicable regional generalization bound takes the form*

$$\mathbb{E}\mathcal{R}_{D,i}(\Theta_{S,i}) - \mathbb{E}\mathcal{R}_{S,i}(\Theta_{S,i}) \leq A_i\sqrt{\frac{V_i(\delta)}{M_i}}, \tag{83}$$

up to logarithmic factors, where $A_i > 0$ collects constants independent of M_i.

To certify a regional generalization gap no larger than δ^2, it is sufficient that

$$M_i \geq \frac{A_i^2 V_i(\delta)}{\delta^4}. \tag{84}$$

Accordingly, define the certified threshold

$$M_i^{\mathrm{cert}}(\delta) := \frac{A_i^2 V_i(\delta)}{\delta^4}. \tag{85}$$

Since Theorem 2 gives $V_i(\delta) \geq \underline{V}_i(\delta)$, we obtain

$$\begin{aligned} M_i^{\mathrm{cert}}(\delta) &\geq \frac{A_i^2 \underline{V}_i(\delta)}{\delta^4} \\ &= A_i^2 c_{s_i,d}\ell^d B_i^{d/(s_i-2)}\delta^{-d/(s_i-2)-4}. \end{aligned} \tag{86}$$

Defining

$$\underline{M}_i(\delta) := A_i^2 c_{s_i,d}\ell^d B_i^{d/(s_i-2)}\delta^{-d/(s_i-2)-4}, \tag{87}$$

we therefore have

$$M_i^{\mathrm{cert}}(\delta) \geq \underline{M}_i(\delta). \tag{88}$$

Remark 2 (Interpretation)**.** *The quantity $\underline{M}_i(\delta)$ is a lower bound on the threshold induced by the adopted VC-based generalization certificate, rather than a necessary statistical sample-complexity lower bound. It describes how the regional approximation parameters (s_i, B_i) propagate through this certificate and therefore provides a conditional complexity rationale for preferential refinement.*

D Potential Energy and Local Spatial Resolution

We provide a complementary interpretation of potential-guided refinement for phase-field energies. This analysis is separate from Theorem 2: it motivates the potential as a local spatial-resolution indicator, rather than establishing convergence or optimality of PAS.

Consider the Ginzburg–Landau energy density

$$e_\epsilon(u) = \frac{\epsilon}{2}|\nabla u|^2 + \frac{F(u)}{\epsilon}, \tag{89}$$

and define the associated discrepancy

$$\xi_\epsilon := \frac{\epsilon}{2}|\nabla u|^2 - \frac{F(u)}{\epsilon}. \tag{90}$$

Suppose that, on a region of interest,

$$\xi_\epsilon \le C. \tag{91}$$

Rearranging the above inequality gives

$$\frac{\epsilon^2}{2}|\nabla u|^2 \le F(u) + C\epsilon, \tag{92}$$

and hence

$$|\nabla u| \le \frac{\sqrt{2}}{\epsilon}\sqrt{F(u) + C\epsilon}. \tag{93}$$

Proposition 3 (Potential-dependent sufficient resolution)**.** *Let $D \subset \Omega$ be a compact and convex region, and let $u \in C^1(D)$. Denote its diameter by*

$$h_D := \operatorname{diam}(D),$$

and define

$$F_D := \sup_{x\in D} F(u(x)), \qquad V_D(u) := \sup_{x,y\in D} |u(x) - u(y)|.$$

If $\xi_\epsilon \le C$ on D, then

$$V_D(u) \le \frac{\sqrt{2}h_D}{\epsilon}\sqrt{F_D + C\epsilon}. \tag{94}$$

Consequently, for any prescribed tolerance $\tau > 0$, the condition

$$h_D \le \frac{\epsilon\tau}{\sqrt{2(F_D + C\epsilon)}} \tag{95}$$

is sufficient to guarantee $V_D(u) \le \tau$.

Proof. Since $\xi_\epsilon \le C$ on D, for every $z \in D$ we have

$$\frac{\epsilon}{2}|\nabla u(z)|^2 - \frac{F(u(z))}{\epsilon} \le C. \tag{96}$$

Multiplying by $\epsilon > 0$ and rearranging yields

$$|\nabla u(z)| \le \frac{\sqrt{2}}{\epsilon}\sqrt{F(u(z)) + C\epsilon} \le \frac{\sqrt{2}}{\epsilon}\sqrt{F_D + C\epsilon}. \tag{97}$$

Now take arbitrary $x, y \in D$. Since D is convex, the line segment $\gamma(t) = x + t(y - x)$, $t \in [0, 1]$, is entirely contained in D. Applying the fundamental theorem of calculus along this segment gives

$$u(y) - u(x) = \int_0^1 \nabla u(\gamma(t)) \cdot (y - x)\, dt. \tag{98}$$

Therefore, by the Cauchy–Schwarz inequality and Eq. (97),

$$|u(y) - u(x)| \le |y - x| \int_0^1 |\nabla u(\gamma(t))|\, dt \tag{99}$$

$$\le |y - x| \sup_{z \in D} |\nabla u(z)| \tag{100}$$

$$\le \frac{\sqrt{2}|y - x|}{\epsilon} \sqrt{F_D + C\epsilon} \tag{101}$$

$$\le \frac{\sqrt{2}h_D}{\epsilon} \sqrt{F_D + C\epsilon}. \tag{102}$$

Since $x, y \in D$ are arbitrary, taking the supremum over all such pairs gives

$$V_D(u) = \sup_{x,y \in D} |u(x) - u(y)| \le \frac{\sqrt{2}h_D}{\epsilon} \sqrt{F_D + C\epsilon}, \tag{103}$$

which proves Eq. (94).

Finally, requiring the right-hand side to be no larger than τ gives

$$h_D \le \frac{\epsilon\tau}{\sqrt{2(F_D + C\epsilon)}}. \tag{104}$$

Hence $V_D(u) \le \tau$, completing the proof. □

Remark 3. *Proposition 3 provides a one-sided, sufficient resolution criterion. In particular, a larger value of F_D decreases the sufficient upper bound on the local spatial scale h_D, thereby providing a conservative rationale for finer resolution in high-potential regions. The result does not imply that a larger potential necessarily yields a larger actual gradient or oscillation, nor does it assert that PAS enforces this spatial scale or realizes an optimal sampling law.*

E Experimental details

E.1 Common setup

Baselines and hyperparameters. For all experiments, we use the $\tanh$ activation function and run each method for 30 sampling iterations. In each iteration, the network is first trained with Adam for 1000 steps and then refined with L-BFGS for another 1000 steps. For PAS, we set $K = 3$, $P_1 = 0.2$, and $N_{\text{pool}} = 100000$. RAD uses an additional residual evaluation set containing $100{,}000$ points, while RSmote adopts an imbalance ratio of 0.45. Random sampling with resampling (Random-R) is included as the baseline random sampling strategy, which essentially corresponds to uniform sampling; specifically, random points are generated by uniform sampling in low-dimensional problems, whereas pseudo-random sampling is used in high-dimensional problems for computational efficiency. For R3, points with residuals larger than the mean residual are retained, while the remaining points are discarded and replaced with uniformly sampled points so that the total number of training points remains unchanged. For RAR, $100{,}000$ candidate points are uniformly generated at each sampling iteration, and the 50 points with the largest residuals are added to the training set as new anchor points. All experiments are conducted on an NVIDIA RTX 5090 GPU and averaged over 5 random seeds.

Network architectures and loss weights. Table 3 summarizes configurations. We enforce initial and boundary conditions as hard constraints in some low-dimensional experiments to ensure stability for stiff dynamics (Hao et al., 2025).

Table 3: Experimental configuration for all benchmarks.

Problem	Depth	Width	Input dim.	Training points	λ_{pde}	$(\lambda_{ic}, \lambda_{bc}, \lambda_{data})$
1D Allen-Cahn	3	64	2	10000	–	(–, –, –)
2D Allen-Cahn	3	64	3	10000	–	(–, –, –)
1D Cahn-Hilliard	3	64	2	10000	–	(–, –, –)
10D reaction-diffusion (Allen–Cahn type)	3	20	11	10000	1	(1, 1, –)
20D reaction-diffusion (Allen–Cahn type)	3	40	21	50000	1	(1, 1, –)
1D semilinear elliptic (inverse)	3	64	1	2000 (+200 obs.)	1	(–, 1, 20)

E.2 Other Details

The relative L_2 error reported in this paper is calculated as follows:

$$\text{Relative } L_2 \text{ error} = \frac{\|u - \hat{u}\|_2}{\|u\|_2}, \tag{105}$$

where $\hat{u}$ denotes the predicted solution obtained by the PINN, and u denotes the reference solution. In our experiments, u is either known analytically or approximated by a high-fidelity numerical solution. Apart from prescribed initial and boundary data and the observations used in the inverse problem, this reference solution is used only for evaluation. For the inverse problem, the accuracy of the identified coefficient is measured by the relative coefficient error

$$\text{Relative coefficient error} = \frac{|\lambda^\star - \hat{\lambda}|}{\lambda^\star}, \tag{106}$$

where $\hat{\lambda}$ is the coefficient recovered by the PINN and $\lambda^\star = 1$ is the value used to generate the reference solution. Both metrics are evaluated at every adaptive round.

E.3 1D Allen-Cahn equation

We solve

$$\begin{cases} \partial_t u = 0.001\, u_{xx} - 5(u^3 - u), & (x,t) \in [-1,1] \times [0,1], \\ u(x,0) = x^2 \cos(\pi x), \end{cases} \tag{107}$$

with Dirichlet boundary conditions.

$$u(-1,t) = -1, \quad u(1,t) = -1, \quad t \in [0,1]. \tag{108}$$

Figure 5 visualizes the sampling point distributions of PAS, RAD, RAR, R3 and RSmote for the one-dimensional Allen-Cahn equation. The background shows the reference solution, and the brown dots indicate the selected sampling points. PAS places more samples near the interfaces, whereas RAD and RSmote distribute samples more broadly over the space-time domain.

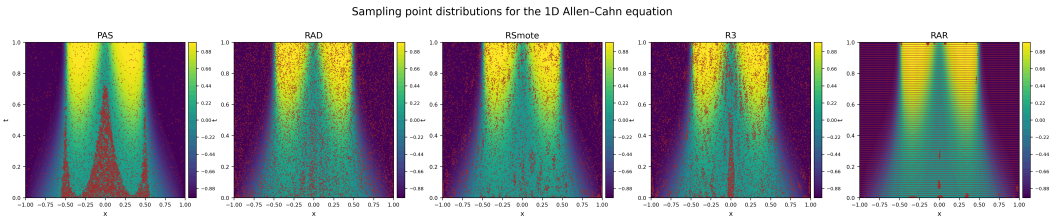


Figure 5: Sampling point distributions for the 1D Allen-Cahn equation. The figure compares the sampling patterns of PAS, RAD, RAR, R3, and RSmote over the space-time domain.

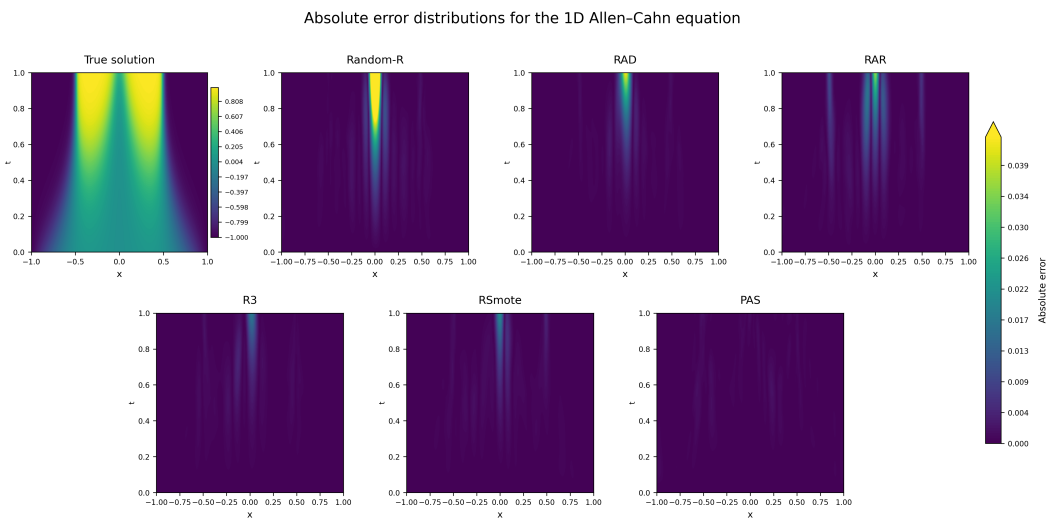


Figure 6: Reference solution and absolute error distributions for the 1D Allen-Cahn equation using Random-R, RAD, RAR, R3, RSmote, and PAS. The first panel shows the reference solution, and the remaining panels show the pointwise absolute error for each method.

Figure 6 shows the reference solution and compares the absolute error distributions obtained by different sampling strategies for the one-dimensional Allen-Cahn equation. The errors are mainly concentrated near the moving interfaces, where the solution changes rapidly. Random-R and RAD show relatively large errors in these regions, while RSmote and PAS reduce the error more effectively. Among all methods, PAS produces the smallest errors.

E.4 2D Allen-Cahn equation

We consider the following two-dimensional Allen-Cahn equation on the periodic domain $[-1,1]^2 \times [0,1]$:

$$u_t - 0.001(u_{xx} + u_{yy}) - 5(u - u^3) = 0, \qquad (x, y, t) \in [-1,1]^2 \times (0,1], \tag{109}$$

with periodic boundary conditions and initial condition

$$u(x, y, 0) = \sin(2\pi x)\cos(2\pi y). \tag{110}$$

The reference solution is computed by a Fourier spectral method (Boyd, 2001) with 256^2 modes and ETDRK4 time integration (Kassam & Trefethen, 2005) with time step $\Delta t = 10^{-5}$.

E.5 1D Cahn-Hilliard equation

We consider the one-dimensional Cahn-Hilliard system on $[-1,1] \times [0,1]$:

$$\begin{cases} \partial_t u = \gamma_2\, \partial_{xx}\mu, \\ \mu = u^3 - u - \gamma_1\, \partial_{xx} u, \end{cases} \tag{111}$$

with

$$\gamma_1 = 10^{-4}, \qquad \gamma_2 = 10^{-2}. \tag{112}$$

The initial condition is

$$u(x, 0) = -\cos(2\pi x), \qquad x \in [-1,1], \tag{113}$$

and periodic boundary conditions.

This problem is a Cahn-Hilliard-type phase-field equation. Different from the Allen-Cahn equation, the Cahn-Hilliard equation is an H^{-1}-gradient flow. Therefore, after eliminating μ, the equation contains the fourth-order term $-\gamma_1\gamma_2 u_{xxxx}$, which makes it essentially different from the second-order Allen-Cahn dynamics.

Figure 7 compares the sampling point distributions of PAS, RAD, RAR, R3, and RSmote for the one-dimensional Cahn-Hilliard equation. The background shows the reference solution, and the brown dots indicate the selected sampling points. PAS places more samples near the phase interfaces and early-time transient regions, whereas RAD and RSmote distribute samples more broadly over the space-time domain.

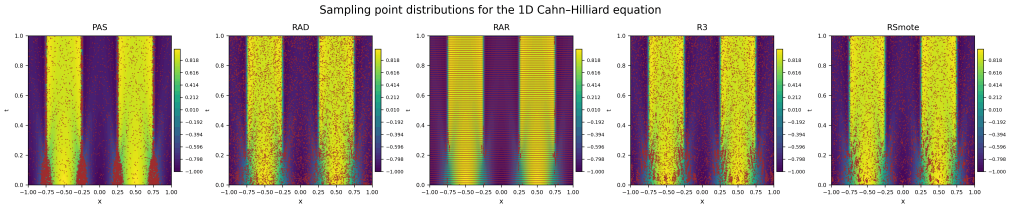


Figure 7: Sampling point distributions for the 1D Cahn-Hilliard equation. The figure compares the sampling patterns of PAS, RAD, RAR, R3 and RSmote over the space-time domain.

Figure 8 shows the reference solution and compares the absolute error distributions for the one-dimensional Cahn-Hilliard equation. All methods reproduce the main phase-separated structure, but their error levels differ significantly. Random-R has the largest errors, mainly concentrated near the early-time interfacial regions and along the vertical transition layers. RAD reduces the error magnitude compared with Random-R, but clear interface-related errors remain. RSmote further improves the prediction and produces smaller errors around most transition regions. PAS achieves the lowest overall error level among the tested methods, with errors more uniformly suppressed across the space-time domain. This indicates that PAS better captures the sharp phase-separation dynamics of the Cahn-Hilliard equation.

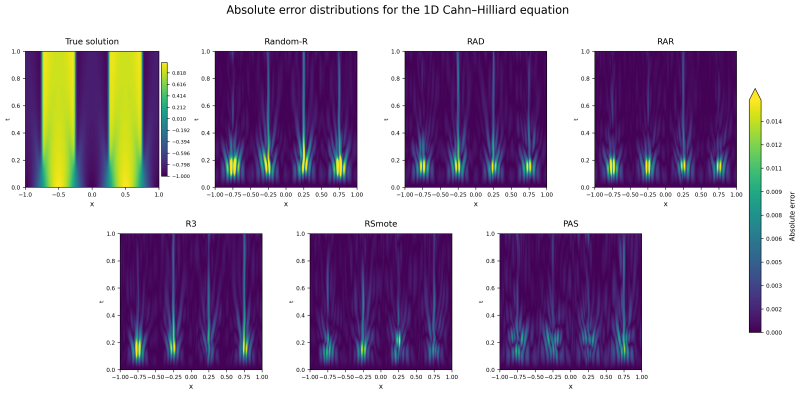


Figure 8: Reference solution and absolute error distributions for the 1D Cahn-Hilliard equation using Random-R, RAD, RAR, R3, RSmote, and PAS. The first panel shows the reference solution, and the remaining panels show the pointwise absolute error for each method.

E.6 High-dimensional problem

We consider the following high-dimensional problem on $[-1,1]^d \times [0,1]$:

$$\begin{cases} u_t - \varepsilon^2 \Delta u - u + u^3 = f(\mathbf{x}, t), & (\mathbf{x}, t) \in [-1,1]^d \times (0,1], \\ u(\mathbf{x}, 0) = u^*(\mathbf{x}, 0), & \mathbf{x} \in [-1,1]^d, \end{cases} \tag{114}$$

where

$$s(\mathbf{x}) = d^{-1/2} \sum_{i=1}^{d} x_i, \qquad d = 10, 20. \tag{115}$$

The source term is chosen as

$$f(\mathbf{x}, t) = -0.1 e^{-0.1t} \tanh\left(\frac{s(\mathbf{x})}{\sqrt{2}\varepsilon}\right) - \left(e^{-0.1t} - e^{-0.3t}\right) \tanh^3\left(\frac{s(\mathbf{x})}{\sqrt{2}\varepsilon}\right), \tag{116}$$

so that the exact solution is

$$u^*(\mathbf{x}, t) = e^{-0.1t} \tanh\left(\frac{s(\mathbf{x})}{\sqrt{2}\varepsilon}\right). \tag{117}$$

For general energy-driven reaction-diffusion equations $u_t - \Delta u + R'(u) = 0$ in a domain $\Omega \subseteq \mathbb{R}^d$, Modica-type gradient estimates (Banerjee & Garofalo, 2015) establish a link between the gradient energy and the reaction potential. Because their discrepancy is controlled and vanishes in appropriate limiting regimes, the potential characterizes regions with sharp spatial variations.

Within this general framework, the specific physical model is determined by the choice of $R(u)$; notably, specifying the double-well potential $R(u) = \frac{1}{4}(1-u^2)^2$ instantiates the classic Allen-Cahn equation, while Eq. (114) represents a forced Allen-Cahn equation. Although the source term $f(\mathbf{x}, t)$ affects the temporal evolution of the diffuse interfaces, the system fundamentally retains the core reaction mechanism governed by the double-well potential $R(u) = \frac{1}{4}(1-u^2)^2$. Furthermore, recent theoretical analyses (Nguyen & Wang, 2024) confirm that, under suitable forcing conditions, the gradient and potential energies remain asymptotically equipartitioned near the sharp-interface limit. That is, the energy discrepancy(15)$\xi_\varepsilon := G_\varepsilon - P_\varepsilon$ still vanishes asymptotically. This property rigorously justifies our use of the double-well potential as a reliable adaptive sampling indicator, even in the presence of external source terms.

E.7 1D SEMILINEAR ELLIPTIC INVERSE PROBLEM

To examine PAS beyond the phase-field equations considered above, we consider the semilinear elliptic equation

$$-\varepsilon^2 \Delta u + \lambda W'(u) = 0 \qquad \text{in } \Omega, \tag{118}$$

associated with the energy

$$\mathcal{E}_{\varepsilon,\lambda}[u] = \int_\Omega \left[\frac{\varepsilon^2}{2}|\nabla u|^2 + \lambda W(u)\right] \mathrm{d}x. \tag{119}$$

Here, $\lambda > 0$ controls the relative contribution of the potential energy.

We consider the one-dimensional inverse problem on $\Omega = (-1, 1)$,

$$-\varepsilon^2 u''(x) + \lambda W'(u(x)) = 0, \qquad u(-1) = u(1) = 1, \tag{120}$$

with

$$W(u) = \frac{1}{2}\big|u(1-u^2)\big|^{3/2}. \tag{121}$$

The coefficient λ is inferred from 200 interior observations together with the boundary conditions, with $\lambda^\star = 1$ used to generate the reference solution. The reference profile is constructed from the closed-form first integral of (120), sampled on 20001 uniform points.

Why potential-based sampling extends beyond the double-well case. The use of the potential as a sampling indicator is not specific to the double-well potential. For semilinear elliptic equations, the classical Modica gradient estimate relates spatial variation directly to a non-negative potential. In its standard setting, a bounded entire solution of

$$-\varepsilon^2 \Delta u + \lambda W'(u) = 0, \qquad W \geq 0, \tag{122}$$

satisfies, after rescaling,

$$\frac{\varepsilon^2}{2}|\nabla u|^2 \leq \lambda W(u). \tag{123}$$

Thus, the potential provides a direct energy-based scale for spatial variation. This observation motivates using $W(\hat{u})$ as a refinement indicator without requiring W to take the particular double-well form used in the Allen–Cahn and Cahn–Hilliard examples. More generally, whenever a non-negative potential admits an appropriate Modica-type gradient estimate, the same potential-guided sampling principle applies (Modica, 1987; Banerjee & Garofalo, 2015).

PAS obtains $\hat{\lambda} = 0.964 \pm 0.008$, corresponding to a relative coefficient error of $3.63 \times 10^{-2} \pm 8.4 \times 10^{-3}$, and reduces the relative field error to $2.37 \times 10^{-3} \pm 1.6 \times 10^{-4}$. These results are respectively

3.2× and 2.3× better than the strongest baseline (R3). Moreover, PAS places 92.4% of its samples near the transition regions, compared with 42.9–57.3% for the residual-based methods, consistent with the potential-guided refinement mechanism in (123).

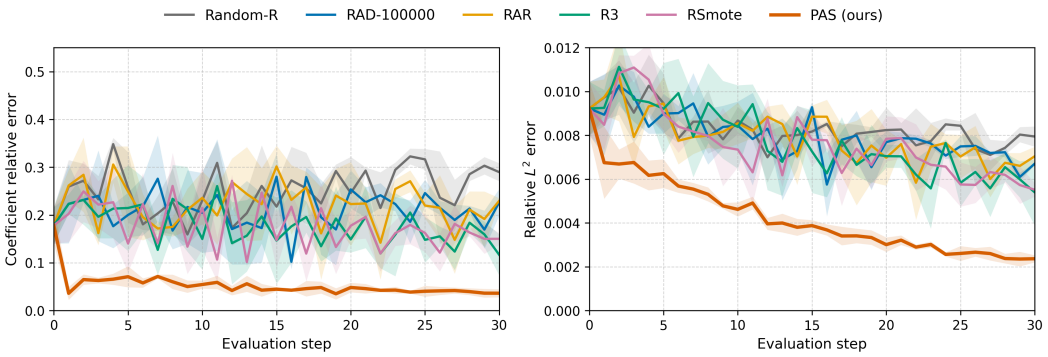


Figure 9: Convergence on the 1D semilinear elliptic inverse problem. Left: relative error of the inferred coefficient λ. Right: relative L^2 error of the recovered solution. All methods share the same initialization and training configuration and differ only in the collocation update.

F Ablation and Sensitivity Studies

We conduct a set of ablation and sensitivity studies to further examine the effectiveness and robustness of the proposed sampling strategy. Specifically, we study: (F.1) the effectiveness of the potential-based indicator in isolation; (F.2) the sensitivity of PAS to the hyperparameters (K, P_1); (F.3) the influence of the candidate pool size N_{pool}; and (F.4) the scaling behavior with respect to the number of collocation points N. These studies are performed on representative low-dimensional and high-dimensional problems, and we report the relative L^2 error, GPU memory usage, and total training time when applicable.

F.1 Effectiveness of the potential-based indicator

We first investigate whether the potential-based indicator alone is effective for identifying informative samples. To this end, we construct potential-based variants of existing adaptive sampling strategies by replacing the residual indicator in RAD and RSmote with the potential indicator, leading to PAD and PSmote, respectively. The comparison is performed on the one-dimensional Allen-Cahn equation.

Table 4: Effectiveness of the potential indicator on 1D Allen-Cahn.

Method	Indicator	L^2 relative error	Memory (MB)	Training time (s)
RAD	residual	$9.62 \times 10^{-3} \pm 4.52 \times 10^{-3}$	1458	8515.08
PAD	potential	$8.56 \times 10^{-3} \pm 3.19 \times 10^{-3}$	710	7395.50
RSmote	residual	$4.52 \times 10^{-3} \pm 1.59 \times 10^{-3}$	710	9141.98
PSmote	potential	$2.31 \times 10^{-3} \pm 9.51 \times 10^{-4}$	710	7274.16
PAS	potential	$\mathbf{6.81 \times 10^{-4} \pm 1.35 \times 10^{-4}}$	710	**5688.44**

As shown in Table 4 and Fig. 10, replacing the residual indicator with the potential indicator improves both RAD and RSmote. PAD reduces the relative L^2 error and memory usage compared with RAD, while PSmote achieves lower error and shorter training time than RSmote. PAS obtains the best accuracy and the shortest training time among the tested methods. These results suggest that the potential indicator provides an effective criterion for selecting informative samples in phase-field problems.

Remark 4. *Recent physics-aware sampling strategies have also exploited thermodynamic information. EDRAS (Li et al., 2026) uses the local energy dissipation rate density to guide adaptive*

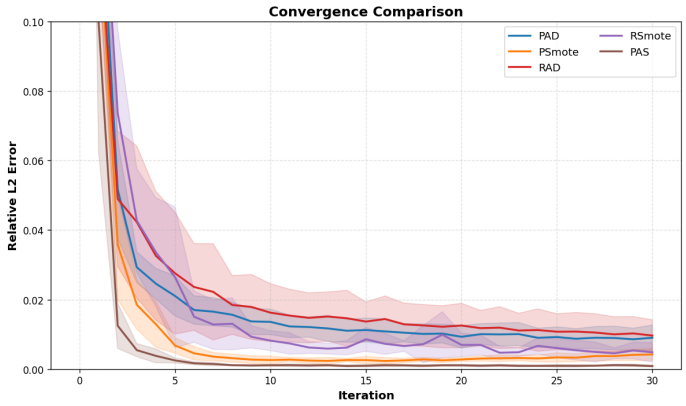


Figure 10: Convergence comparison between potential-based and residual-based adaptive sampling strategies for the 1D Allen-Cahn equation.

sampling, while Buck and Kim (Buck & Kim, 2026) develop an auto-adaptive framework and instantiate it with an energy-based sampling strategy for phase-transition problems. Importantly, however, the physics-aware instantiations developed and numerically validated in both works are centered on the Allen–Cahn equation. In particular, neither work provides or validates a corresponding sampling formulation across different governing equations such as Cahn–Hilliard, high-dimensional reaction–diffusion systems, or elliptic inverse problems.

This distinction is central to PAS. Rather than constructing a sampling criterion from an equation-specific energy dissipation rate or the full energy density, PAS uses a local potential of the predicted state as a derivative-free indicator. This construction is not tied to the differential operator of a particular governing equation and can be instantiated whenever an appropriate potential landscape is available. Consequently, the same potential-based sampling principle can be applied across substantially different PDE systems. In this work, we demonstrate this portability on Allen–Cahn and Cahn–Hilliard equations, high-dimensional reaction–diffusion problems, and an elliptic inverse problem.

For this reason, EDRAS and the energy-adaptive strategy of Buck and Kim are regarded as closely related physics-aware sampling methods rather than direct baselines: their existing formulations do not provide a common sampling method that can be applied across the heterogeneous benchmark suite considered here. Extending these methods to the non-Allen–Cahn problems studied in this work would require additional equation-specific formulations beyond those established and validated in the original works.

F.1.1 Comparison with gradient-based interface sampling

We further investigate whether the effectiveness of PAS can be explained solely by generic interface detection. To this end, we construct a gradient-based sampling baseline, denoted by PGS, which follows the same sampling procedure as PAS but replaces the potential-based score with the absolute spatial-gradient score

$$s_{\mathrm{grad}}(x,t) = \left| \frac{\partial \hat{u}(x,t)}{\partial x} \right|.$$

PAS and PGS use the same candidate pool, number of selected points, network architecture, optimizer, and training schedule. Therefore, the only difference between the two methods is the indicator used to rank candidate points.

As shown in Table 5, PAS substantially outperforms the gradient-based baseline under the same sampling and training configuration. PAS reduces the relative L^2 error from 2.80×10^{-2} to 6.81×10^{-4}, corresponding to approximately a $41.1\times$ reduction in error. It also reduces GPU memory consumption from 845 MB to 710 MB and total training time from 13198.20 s to 5688.44 s. These results indicate that the benefit of PAS cannot be attributed solely to concentrating samples near regions with large solution gradients.

Table 5: Comparison between potential-based and gradient-based sampling on the 1D Allen–Cahn equation under the same sampling and training configuration.

Method	Sampling indicator	Relative L^2 error	GPU memory (MB)	Total time (s)
PGS	Gradient	$2.80 \times 10^{-2} \pm 3.30 \times 10^{-3}$	845	13198.20
PAS	Potential	$\mathbf{6.81 \times 10^{-4} \pm 1.35 \times 10^{-4}}$	**710**	**5688.44**

For the potential, the stable bulk phases correspond to $u \approx \pm 1$, where $F(u) \approx 0$, while intermediate phase values have elevated potential energy. Consequently, the potential indicator can assign high importance to transition states even when the instantaneous solution gradient is not yet large. This distinction is particularly relevant for time-dependent phase-field dynamics, where inaccuracies in an evolving transition region may affect the subsequent solution. Moreover, evaluating $F(\hat{u})$ requires only the network prediction, whereas a gradient-based indicator additionally requires automatic differentiation with respect to the input coordinates. The accuracy and computational results in Table 5 therefore provide empirical evidence that the potential indicator captures useful phase-state information beyond generic gradient-based interface detection.

F.2 Sensitivity to hyperparameters

We next study the sensitivity of PAS to the parameters (K, P_1), where K is the local point generation multiplier and P_1 is the fraction of high-potential points selected for local densification. The results are reported in Table 6.

Table 6: Sensitivity of PAS to (K, P_1) settings on 1D Allen-Cahn.

K	P_1	1D AC L^2 error	Training Time (s)
1	0.30	$3.19 \times 10^{-3} \pm 1.84 \times 10^{-3}$	5875.27
	0.40	$2.13 \times 10^{-3} \pm 1.49 \times 10^{-3}$	**5628.16**
2	0.20	$2.22 \times 10^{-3} \pm 6.74 \times 10^{-4}$	5842.15
	0.30	$7.09 \times 10^{-4} \pm 1.23 \times 10^{-4}$	5830.72
3	0.20	$\mathbf{6.81 \times 10^{-4} \pm 1.35 \times 10^{-4}}$	5688.44
5	0.15	$9.90 \times 10^{-4} \pm 2.05 \times 10^{-4}$	5642.45

Overall, PAS is not overly sensitive to the choice of (K, P_1). All tested settings achieve errors on the order of 10^{-3}, and several settings further reduce the error below 10^{-3}. The best accuracy is obtained with $(K, P_1) = (3, 0.20)$, which gives a relative L^2 error of $6.81\text{e}{-4}$. The setting used in the main experiments, $(K, P_1) = (3, 0.20)$, achieves a comparable error of $6.81\text{e}{-4}$ with similar training cost. This shows that PAS is robust over a reasonable range of hyperparameters, and its performance does not rely on a highly tuned parameter choice.

F.3 Effect of candidate pool size

We further examine the effect of the candidate pool size N_{pool}, which determines the number of candidate points considered during adaptive sampling. Since RAD explicitly evaluates the residual over an additional candidate set, we use it as the main reference method in this study. The results are shown in Table 7.

For RAD, increasing N_{pool} leads to a clear increase in GPU memory usage, as the residual must be evaluated over a larger candidate set. In contrast, PAS keeps the memory usage nearly unchanged. On both the 1D Allen-Cahn problem and the 20D problem, PAS achieves lower errors than RAD while using less memory. In the 20D case, RAD runs out of memory when $N_{\text{pool}} = 500{,}000$, whereas PAS remains feasible. These results indicate that PAS scales more favorably with the candidate pool size.

Table 7: Effect of N_{pool} on 1D Allen-Cahn and 20D reaction-diffusion. Best results in each row are in bold. Memory denotes GPU memory usage.

Problem	N_{pool}	Method	Error	Memory (MB)
1D AC	100,000	RAD	$9.62 \times 10^{-3} \pm 4.52 \times 10^{-3}$	1458
		PAS (ours)	$\mathbf{6.81 \times 10^{-4} \pm 1.35 \times 10^{-4}}$	**710**
	200,000	RAD	$9.48 \times 10^{-3} \pm 3.21 \times 10^{-3}$	2148
		PAS (ours)	$\mathbf{6.69 \times 10^{-4} \pm 9.98 \times 10^{-5}}$	**710**
	500,000	RAD	$8.45 \times 10^{-3} \pm 2.66 \times 10^{-3}$	4182
		PAS (ours)	$\mathbf{7.84 \times 10^{-4} \pm 1.86 \times 10^{-4}}$	**710**
20D RD	100,000	RAD	$2.06 \times 10^{-2} \pm 1.09 \times 10^{-2}$	7510
		PAS (ours)	$\mathbf{1.25 \times 10^{-2} \pm 5.11 \times 10^{-3}}$	**4388**
	200,000	RAD	$1.90 \times 10^{-2} \pm 1.60 \times 10^{-2}$	15600
		PAS (ours)	$\mathbf{1.05 \times 10^{-2} \pm 5.62 \times 10^{-3}}$	**4388**
	500,000	RAD	–	OOM
		PAS (ours)	$\mathbf{7.18 \times 10^{-3} \pm 4.85 \times 10^{-3}}$	**4388**

F.4 Number of collocation points

Finally, we investigate the effect of the number of collocation points N on both the 1D Allen-Cahn problem and the 20D high-dimensional Allen-Cahn problem. The results are summarized in Table 8. For the 20D problem, we do not report the case $N = 2000$, since such a small number of collocation points is insufficient to cover the high-dimensional space and leads to an under-resolved training set. Therefore, we use $N = 10000$ as the starting point for the 20D experiments.

For the 1D problem, PAS achieves the lowest error for all tested values of N. RSmote improves as N increases, but PAS remains consistently more accurate. For the 20D problem, PAS also gives the lowest error at both $N = 10000$ and $N = 50000$. These results suggest that PAS makes effective use of limited collocation budgets and remains competitive as the number of training points increases.

Table 8: Effect of N (number of collocation points) on 1D Allen-Cahn and 20D reaction-diffusion. Best results in each row are in bold.

Problem	N	Method	Error	Mem. (MB)
1D AC	2000	RAD	$1.44 \times 10^{-2} \pm 2.09 \times 10^{-3}$	1398
		RSmote	$7.87 \times 10^{-2} \pm 1.37 \times 10^{-1}$	664
		PAS (ours)	$\mathbf{1.09 \times 10^{-3} \pm 7.93 \times 10^{-5}}$	664
	10000	RAD	$9.62 \times 10^{-3} \pm 4.52 \times 10^{-3}$	1458
		RSmote	$4.52 \times 10^{-3} \pm 1.59 \times 10^{-3}$	710
		PAS (ours)	$\mathbf{6.81 \times 10^{-4} \pm 1.35 \times 10^{-4}}$	710
	50000	RAD	$1.03 \times 10^{-2} \pm 2.45 \times 10^{-3}$	1824
		RSmote	$3.21 \times 10^{-3} \pm 7.61 \times 10^{-4}$	1088
		PAS (ours)	$\mathbf{9.53 \times 10^{-4} \pm 3.40 \times 10^{-4}}$	1088
20D RD	10000	RAD	$8.39 \times 10^{-2} \pm 5.03 \times 10^{-2}$	7118
		RSmote	$4.76 \times 10^{-2} \pm 2.53 \times 10^{-2}$	1706
		PAS (ours)	$\mathbf{2.39 \times 10^{-2} \pm 1.38 \times 10^{-2}}$	1706
	50000	RAD	$2.06 \times 10^{-2} \pm 1.09 \times 10^{-2}$	7510
		RSmote	$2.21 \times 10^{-2} \pm 3.18 \times 10^{-2}$	4388
		PAS (ours)	$\mathbf{1.25 \times 10^{-2} \pm 5.11 \times 10^{-3}}$	4388

G Limitations

Limitations and future work. The limitation of PAS is that its sampling indicator is problem-dependent. Extending the framework to broader classes of PDEs therefore requires identifying suitable physics- or energy-based indicators for the governing system. An important direction for future work is to generalize the energy-aware sampling paradigm to other equations, including heat-type and more general reaction–diffusion systems, by exploiting their corresponding energy functionals or other physically meaningful quantities.