SEGMENTED RESIDUAL-CENTERED DIFFUSION ANCHORS WITH DETERMINISTIC ROLLOUTS FOR ALLEN–CAHN DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

We propose a segmented diffusion–Euler method for conditional coarse-to-fine reconstruction of Allen–Cahn dynamics. Residual-centered diffusion reconstructs sparse high-resolution anchors from coarse observations, and deterministic fine-grid Euler steps update each segment interior. An Allen–Cahn grid residual provides physical guidance, with the ReMD multigrid correction kept as an optional refinement. For any fixed finite spatial dimension, we derive a rollout bound that separates discretization bias, anchor risk, and segment-local error amplification. Controlled experiments show that periodic re-anchoring substantially improves rollout accuracy over unsegmented original ReMD. The Allen–Cahn-aware variant achieves comparable field accuracy with lower mean energy mismatch, and independent training runs show low variability. These results identify anchor reconstruction as the main accuracy bottleneck and quantify the tradeoffs between learned reconstruction and deterministic dynamics.

1 INTRODUCTION

The Allen–Cahn equation is a prototypical semilinear reaction–diffusion equation in phase-field modeling. It was originally introduced to describe antiphase boundary motion and domain coarsening in binary alloys (Allen & Cahn, 1979), and it has since become a canonical model for diffuse-interface dynamics, phase separation, image processing, and geometric motion by mean curvature (Cahn & Hilliard, 1958; Cahn, 1961; Modica & Mortola, 1977; Bronsard & Kohn, 1991; Evans et al., 1992; Chen, 1992; de Mottoni & Schatzman, 1995; Ilmanen, 1993; Boettinger et al., 2002; Provatas & Elder, 2010). In the standard nondimensional form,

$$\partial_t u = \varepsilon^2 \Delta u + u - u^3, \tag{1}$$

the parameter $\varepsilon > 0$ controls the interfacial width. When ε is small, the solution develops thin transition layers and large gradients. Resolving these layers over long time intervals requires fine spatial grids and small time steps, especially for explicit or semi-explicit time integrators. Classical finite difference, finite element, spectral, IMEX, and energy-stable schemes provide reliable approximations with mature stability and convergence theory (Strikwerda, 2004; LeVeque, 2007; Morton & Mayers, 2005; Feng & Prohl, 2003; 2004; Eyre, 1998; Shen & Yang, 2010; Du et al., 2018). Nevertheless, repeatedly evolving the full high-resolution state at every physical time level remains expensive. Recent progress in scientific machine learning has opened an alternative route: instead of solving every instance of a PDE at the finest resolution, one can learn maps between solution spaces or recover high-resolution fields from low-resolution observations. Physics-informed neural networks (Raissi et al., 2019), DeepONet (Lu et al., 2021), Fourier neural operators (Li et al., 2021), and general neural operator frameworks (Kovachki et al., 2023) have demonstrated that neural networks can learn solution operators for families of PDEs. Benchmarks such as PDEBench (Takamoto et al., 2022) have further standardized the evaluation of learned solvers for time-dependent PDEs. Related work on fluid and climate super-resolution has shown that neural networks can reconstruct unresolved scales from coarse fields (Fukami et al., 2019; Liu et al., 2020; Fukami et al., 2021; Stengel et al., 2020). However, direct one-shot super-resolution or long neural rollouts may lose coarse-grid consistency, violate discrete physics constraints, and accumulate temporal errors. These issues are particularly relevant for Allen–Cahn dynamics, where the interface geometry and the balance between diffusion

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.

and reaction must be preserved. Classical numerical analysis often studies maps $u^0 \mapsto u(\cdot, t)$ without supervised HR trajectory data. We instead study *conditional* coarse-to-fine reconstruction at sparse anchor times,

$$\{\mathbf{u}_H^j\}_{j=0}^J \longmapsto \widehat{\mathbf{u}}_{h,j}^0,$$

followed by deterministic fine-grid Euler evolution inside each segment. The anchor generator uses residual-centered conditional diffusion; it may optionally include the residual–multigrid correction inherited from ReMD (Li et al., 2026). A grid-space Allen–Cahn residual guides training perturbations and sampling refinement. Only the current coarse snapshot conditions generation; all PDE residuals are evaluated by finite differences on generated grid fields. Sparse re-anchoring reduces HR label density and confines learned-error propagation to a segment, but still executes every fine Euler step and is not a drop-in acceleration of a fine-grid integrator. Section 2 gives the complete sampler and rollout definitions; diffusion background appears in (Ho et al., 2020; Song et al., 2021b;a).

Relation to ReMD and scope of novelty. Residual–Multigrid Diffusion (ReMD) provides the residual-centered conditional parameterization and posterior sampling machinery; we retain its multiscale correction only as an optional anchor refinement (Li et al., 2026). We do not claim these components as new. Our primary methodological contribution is sparse learned re-anchoring followed by deterministic fine-grid Euler evolution. We combine it with (i) a dimensionally scaled, grid-space Allen–Cahn vector-field residual and (ii) an error decomposition that converts measurable anchor risk into an end-to-end rollout bound with segment-length amplification. We do not claim a new diffusion sampler family; Appendix Table 3 itemizes inherited and new components.

Contributions.

- A segmented diffusion–Euler method that reconstructs sparse HR anchors and confines learned-error propagation to segment interiors.
- An Allen–Cahn grid residual that guides training perturbations and clean-state refinement during sampling.
- A modular error bound for any fixed finite dimension that translates anchor risk into rollout accuracy while exposing discretization bias and segment-length amplification.
- Controlled original-ReMD baselines and independent-seed replication that quantify the benefit of segmentation and reveal an accuracy–energy tradeoff.

Section 2 defines the method, Section 3 gives the error analysis, and Section 4 reports experiments.

Grid mass-weighted norms $\|\mathbf{v}_h\|_{M_h}$ and $\|\mathbf{v}_H\|_{M_H}$ are defined in Appendix A. Coarse conditions $Y \in \mathbb{R}^{N_H}$ use $\|Y\|_{\mathcal{Y}_H} := \|Y\|_{M_H}$. For random $\mathbf{Z}$, write $\|\mathbf{Z}\|_{\mathbb{E},M_h} := (\mathbb{E}\|\mathbf{Z}\|_{M_h}^2)^{1/2}$.

2 METHOD

2.1 THE ALLEN–CAHN EQUATION

Let $\Omega \subset \mathbb{R}^d$, where $d \in \mathbb{N}$ is any fixed finite spatial dimension, be a bounded Lipschitz domain and let $T > 0$. The analysis below is dimension-generic: the regularity, transfer-stability, and CFL constants may depend on d, and no estimate is claimed to be uniform as $d \to \infty$. We consider the Allen–Cahn initial-boundary value problem

$$\partial_t u = \varepsilon^2 \Delta u + \psi(u), \qquad (\mathbf{x}, t) \in \Omega \times (0, T], \tag{2}$$

where

$$\psi(s) = s - s^3, \qquad \varepsilon > 0. \tag{3}$$

The initial condition is

$$u(\mathbf{x}, 0) = u^0(\mathbf{x}), \qquad \mathbf{x} \in \Omega. \tag{4}$$

For definiteness, we use homogeneous Neumann boundary conditions,

$$\nabla u(\mathbf{x}, t) \cdot \mathbf{n} = 0, \qquad (\mathbf{x}, t) \in \partial\Omega \times [0, T], \tag{5}$$

where $\mathbf{n}$ is the unit outward normal. Periodic or Dirichlet boundary conditions can be handled by modifying the corresponding finite-difference stencil and ghost-cell/periodic index rules at the boundary.

Equation (2) is the L^2-gradient flow of the Ginzburg–Landau energy

$$\mathcal{E}(v) = \int_\Omega \left[\frac{\varepsilon^2}{2} |\nabla v|^2 + \frac{1}{4}(1 - v^2)^2 \right] d\mathbf{x}, \tag{6}$$

since

$$\partial_t u = -\frac{\delta \mathcal{E}}{\delta u} = \varepsilon^2 \Delta u + u - u^3. \tag{7}$$

This gradient-flow structure motivates physics-consistent residuals; however, independently sampled anchors followed by explicit Euler generally *do not* enforce a monotone discrete Ginzburg–Landau energy law $E_h(\widehat{\mathbf{u}}_{h,j}^{m+1}) \le E_h(\widehat{\mathbf{u}}_{h,j}^{m})$. We treat energy drift as an empirical diagnostic (Section 4), optionally strengthened by penalties as in Appendix E.

2.2 Coarse–fine time grids and fine Euler rollout

Spatial semidiscretization uses fine mesh size h and coarse $H>h$; L_h is the standard finite-difference Laplacian on the fine grid (norms, stencils, and boundary ghosts in Appendix A). Fine and coarse fields are grid vectors $\mathbf{v}_h \in \mathbb{R}^{N_h}$ and $\mathbf{v}_H \in \mathbb{R}^{N_H}$. We use the convention that the first subscript denotes the target grid and the second the source grid: R_{Hh} is the *restriction* operator that maps a fine-grid field to the coarse grid, whereas P_{hH} is the *prolongation* (interpolation) operator that lifts a coarse-grid field to the fine grid. Thus

$$R_{Hh} : \mathbb{R}^{N_h} \to \mathbb{R}^{N_H}, \qquad P_{hH} : \mathbb{R}^{N_H} \to \mathbb{R}^{N_h}, \qquad R_{Hh} P_{hH} \approx I_{N_H}. \tag{8}$$

The relation $R_{Hh}P_{hH} \approx I_{N_H}$ means that prolonging a coarse field and restricting it back approximately reproduces the original coarse field; it is exact for compatible nested transfers. The experiments use tensor-product interpolation and conservative block restriction for the residual operators, as detailed in Appendix F. On a uniform dimension-d tensor grid, the mass norm uses $M_h = h^d I$ (and analogously $M_H = H^d I$); a positive diagonal quadrature mass matrix may be used on boundary-adjusted grids. In analysis we use either a stable approximation bound $\|R_{Hh}P_{hH} - I_{N_H}\|_{M_H} \le CH^q$ (Appendix D) or exact nested prolongation on structured grids; see (58). The coarse time ladder and segment geometry are

$$0 = T_0 < T_1 < \cdots < T_J = T, \qquad \Delta T_j := T_{j+1} - T_j, \qquad M_j = \frac{\Delta T_j}{\Delta t_h} \in \mathbb{N}, \qquad t_{j,m} = T_j + m\Delta t_h. \tag{9}$$

For a chosen coarse integrator $\mathbf{\Psi}_H$, the coarse trajectory satisfies

$$\mathbf{u}_H^{j+1} = \mathbf{\Psi}_H^{\Delta T_j}(\mathbf{u}_H^j), \tag{10}$$

with coarse condition (network input)

$$Y_j := \mathbf{y}_j := \mathbf{u}_H^j, \qquad \mathbf{y}_j^+ := \mathbf{\Psi}_H^{\Delta T_j}(\mathbf{y}_j) \tag{11}$$

used only for increment-form flux targets (19) when desired; default residuals use (18). Fine-grid rollouts use the explicit Euler map

$$\mathbf{\Phi}_h^{\Delta t_h}(\mathbf{v}) = \mathbf{v} + \Delta t_h \left[\varepsilon^2 L_h \mathbf{v} + \boldsymbol{\psi}(\mathbf{v}) \right], \qquad \boldsymbol{\psi}(\mathbf{v}) = \mathbf{v} - \mathbf{v}^{\circ 3}, \tag{12}$$

The reference segment satisfies $\mathbf{u}_{h,j}^{m+1} = \mathbf{\Phi}_h^{\Delta t_h}(\mathbf{u}_{h,j}^m)$ and the predicted segment analogously for $\widehat{\mathbf{u}}_{h,j}^m$, with anchors $\widehat{\mathbf{u}}_{h,j}^0 = \mathcal{G}_{\vartheta_j,K}(Y_j)$.

2.3 Clean-state anchor sampler with optional multigrid refinement

We describe the implementation used for the controlled P1 experiments. Given $Y_j = \mathbf{y}_j$ and $\mathbf{c}_j = P_{hH}\mathbf{y}_j$, the network $f_\theta(\mathbf{x}_k, k, \mathbf{c}_j)$ predicts the clean HR field directly (`xstart`). We index the increasing noise schedule by $0 = \eta_0 < \eta_1 < \cdots < \eta_K$; the code uses indices $0, \ldots, K-1$. Set $a_k = \eta_k/4$, $\kappa = 2$, and initialize $\mathbf{x}_K = \mathbf{c}_j + \kappa\sqrt{\eta_K}\boldsymbol{\xi}$, $\boldsymbol{\xi} \sim \mathcal{N}(0, I)$. The clean prediction and its corrected value are

$$\mathbf{v}_k = f_\theta(\mathbf{x}_k, k, \mathbf{c}_j), \qquad \widetilde{\mathbf{v}}_k = \mathbf{v}_k + a_k \big[\mathbf{r}_{j,k}(\mathbf{v}_k) + b\, \mathcal{S}_\varphi(\mathbf{v}_k, \mathbf{r}_{j,k}(\mathbf{v}_k), k) \big], \tag{13}$$

where $b = 1$ enables the inherited learned M2NO refinement and $b = 0$ disables it. The raw residual correction remains active when $b = 0$. Thus NO-M2NO is not a sampler with all residual correction removed. The uncorrected posterior mean is

$$\mathbf{T}_{\theta,k}(\mathbf{x}_k, Y_j) = \frac{\eta_{k-1}}{\eta_k}\mathbf{x}_k + \frac{\eta_k - \eta_{k-1}}{\eta_k}\mathbf{v}_k, \tag{14}$$

and the deployed reverse update is

$$\mathbf{x}_{k-1} = \frac{\eta_{k-1}}{\eta_k}\mathbf{x}_k + \frac{\eta_k - \eta_{k-1}}{\eta_k}\widetilde{\mathbf{v}}_k + \sigma_k \mathbf{z}_k, \qquad \sigma_k^2 = \kappa^2 \frac{\eta_{k-1}(\eta_k - \eta_{k-1})}{\eta_k}. \tag{15}$$

Here $\mathbf{z}_k \sim \mathcal{N}(0, I)$ independently and $\sigma_1 = 0$. The map $\mathcal{G}_{\vartheta,K}(Y_j)$ returns $\mathbf{x}_0$. P1 uses stochastic posterior sampling with fixed evaluation seeds and no clean-state clipping; it does not use deterministic DDIM updates. The same trained network is reused at every segment head.

2.4 Allen–Cahn residual and implemented training loss

Define

$$\mathbf{F}_h(\mathbf{v}) = \varepsilon^2 L_h \mathbf{v} + \mathbf{v} - \mathbf{v}^{\circ 3}, \tag{16}$$

and $\mathbf{F}_H$ analogously. The implemented residual is

$$\mathbf{r}_{j,k}(\mathbf{v}) = P_{hH}\big[\mathbf{y}_j - R_{Hh}\mathbf{v} + \lambda_k(\mathbf{d}_H^j - R_{Hh}\mathbf{F}_h(\mathbf{v}))\big]. \tag{17}$$

The coefficient λ_k absorbs the relative state/flux scaling: in the dimensionally explicit analytical notation, $\lambda_k = \omega_{\mathrm{AC},k}\Delta T_j$ and $\omega_{\mathrm{data},k} = 1$. The code does not multiply the flux term by an additional ΔT_j. P1 uses the vector-field proxy; the increment proxy and its discrepancy are

$$\mathbf{d}_H^j = \mathbf{F}_H(\mathbf{y}_j), \tag{18}$$

$$\mathbf{d}_H^j = (\mathbf{y}_j^+ - \mathbf{y}_j)/\Delta T_j, \tag{19}$$

$$\eta_{\mathrm{dyn},j} = \|(\mathbf{y}_j^+ - \mathbf{y}_j)/\Delta T_j - \mathbf{F}_H(\mathbf{y}_j)\|_{M_H}. \tag{20}$$

In P1, λ_k follows the implemented cosine schedule; the exact schedule and boundary-dependent grid spacings appear in Appendix F.

Training samples one noise index per example. With HR target $\mathbf{x}_j^\star$, define $\mathbf{r}_{j,k}^{\mathrm{tr}} = \mathbf{r}_{j,k}(\mathbf{c}_j)$ and $\mathbf{e}_{j,k}^{\mathrm{tr}} = \mathcal{S}_\varphi(\mathbf{x}_j^\star, \mathbf{r}_{j,k}^{\mathrm{tr}}, k)$. The AC-aware training perturbation is

$$\mathbf{x}_{j,k} = \mathbf{x}_j^\star + a_k(\mathbf{r}_{j,k}^{\mathrm{tr}} + b\,\mathbf{e}_{j,k}^{\mathrm{tr}}) + \kappa\sqrt{\eta_k}\boldsymbol{\xi}. \tag{21}$$

The coarse field is the residual evaluation state during training, whereas the predicted clean field is used at inference. The M2NO state input is the HR target during training and the clean prediction during inference. These state choices are part of the implementation, rather than an identity between training and reverse dynamics. Original ReMD uses its inherited residual routine with the corresponding argument ordering. The implemented objective is

$$\mathcal{L}(\vartheta) = \mathbb{E}\frac{\|f_\theta(\mathbf{x}_{j,k}, k, \mathbf{c}_j) - \mathbf{x}_j^\star\|_2}{\|\mathbf{x}_j^\star\|_2} + 10^{-3} b\,\mathbb{E}\ \mathrm{mean}[(\mathbf{e}_{j,k}^{\mathrm{tr}})^{\circ 2}]. \tag{22}$$

Although named `mse` in the training dictionary, the reconstruction term is relative L^2, not squared error. Gradients reach M2NO through the training perturbation and its output regularizer. P1 uses a single training stage; no unrolled sampler-level fine-tuning or separate flux penalty is optimized. Alternative analytical training objectives are distinguished from this deployed loss in Appendix C.1.

2.5 Segmented diffusion–Euler algorithm

Given the coarse trajectory, the solver proceeds segmentwise: **(i)** at segment head T_j, set $Y_j = \mathbf{u}_H^j$ (and compute $\mathbf{y}_j^+ = \mathbf{\Psi}_H^{\Delta T_j}(\mathbf{u}_H^j)$ only if increment residuals (19) are used); **(ii)** reconstruct the HR anchor

$$\widehat{\mathbf{u}}_{h,j}^0 = \mathcal{G}_{\vartheta_j,K}(Y_j) \tag{23}$$

using the residual-centered sampler (15), with the learned M2NO refinement enabled or disabled by b; **(iii)** roll out inside the segment by

$$\widehat{\mathbf{u}}_{h,j}^{m+1} = \mathbf{\Phi}_h^{\Delta t_h}(\widehat{\mathbf{u}}_{h,j}^m), \qquad m = 0, \dots, M_j - 1; \tag{24}$$

(iv) re-anchor independently at the next coarse time T_{j+1}. Pseudocode appears as Algorithms 1–2 in Appendix B.1.

Segment interfaces. Independence of (15) across j generally yields a *piecewise* HR trajectory. The *tail-to-head* interface jump compares the fine field at the *end* of segment j with the freshly reconstructed anchor at the start of segment $j+1$:

$$\mathbf{J}_{j+1} := \widehat{\mathbf{u}}_{h,j}^{M_j} - \widehat{\mathbf{u}}_{h,j+1}^0, \tag{25}$$

since $\widehat{\mathbf{u}}_{h,j}^{M_j} = \mathbf{\Phi}_h^{M_j \Delta t_h}(\widehat{\mathbf{u}}_{h,j}^0)$ advances the hybrid state through the segment interior. This must *not* be confused with the *anchor-to-anchor* gap $\widehat{\mathbf{u}}_{h,j}^0 - \widehat{\mathbf{u}}_{h,j+1}^0$, which measures only successive sampler outputs at coarse times. The global bound in Section 3 controls errors relative to the reference trajectory within segments but does not bound $\|\mathbf{J}_{j+1}\|_{M_h}$ unless additional stitching is imposed. P1 interface statistics are reported in Appendix G.1. Optional mitigations include adding a continuity penalty that directly targets (25), e.g. $\mathbb{E}\|\mathbf{\Phi}_h^{M_j \Delta t_h}(\mathcal{G}_{\vartheta_j,K}(Y_j)) - \mathcal{G}_{\vartheta_{j+1},K}(Y_{j+1})\|_{M_h}^2$ (Appendix E).

3 Error Analysis

All results in this section hold for every fixed finite $d \geq 1$, provided the dimension-d finite-difference operator and transfer maps satisfy the stated consistency and stability assumptions. In particular, the constants below may depend on d through solution regularity, boundary treatment, $\lambda_{\max}(-L_h)$, and the norms of R_{Hh} and P_{hH}; the result is not a dimension-uniform bound.

Let $\mathcal{S}_h u(t_{j,m})$ sample the exact solution on the fine grid and split

$$\mathcal{S}_h u(t_{j,m}) - \widehat{\mathbf{u}}_{h,j}^m = \big(\mathcal{S}_h u(t_{j,m}) - \mathbf{u}_{h,j}^m\big) + \big(\mathbf{u}_{h,j}^m - \widehat{\mathbf{u}}_{h,j}^m\big). \tag{26}$$

Coarse conditions compare algorithmic snapshots Y_j to ideal restrictions $Y_j^\star$:

$$Y_j = \mathbf{u}_H^j, \qquad Y_j^\star = R_{Hh}\mathbf{u}_{h,j}^0 = R_{Hh}\mathbf{X}_j. \tag{27}$$

Probability model. Fix $(\Omega, \mathcal{F}, \mathbb{P})$ and a data distribution $\mathcal{D}$ over scenarios ω (e.g. random initial phases) inducing a reference Allen–Cahn trajectory $u(\cdot, \cdot; \omega)$. Let $\mathcal{S}_h$ denote sampling onto Ω_h, set $\mathbf{X}_j(\omega) := \mathcal{S}_h u(T_j; \omega)$, and define ideal coarse conditions $Y_j^\star(\omega) = R_{Hh}\mathbf{X}_j(\omega)$ as in (27). Algorithmic coarse states $Y_j(\omega) = \mathbf{u}_H^j(\omega)$ arise from the chosen coarse integrator and may differ from $Y_j^\star(\omega)$. All expectations $\mathbb{E}[\cdot]$ are under $\omega \sim \mathcal{D}$ unless noted; ambiguity terms such as $\epsilon_{\mathrm{id},j}$ are thus distribution-dependent.

3.1 Anchor reconstruction risk

Definition 1 (Bayes reconstructor) *Let $\mathbf{X}_j$ be the random fine-grid anchor at T_j and let the ideal coarse restriction be $Y_j^\star = R_{Hh}\mathbf{X}_j$ as in* (27)*. Under squared-error loss, the conditional mean is the Bayes estimator and the L^2 projection onto the information carried by $Y_j^\star$ (Berger, 1985; Kallenberg, 2021). We therefore define the Bayes reconstructor by*

$$\mathcal{G}_j^\star(Y) := \mathbb{E}[\mathbf{X}_j \mid Y_j^\star = Y]. \tag{28}$$

Irreducible ambiguity is

$$\epsilon_{\mathrm{id},j}^2 := \mathbb{E}\big[\|\mathbf{X}_j - \mathcal{G}_j^\star(Y_j^\star)\|_{M_h}^2\big]. \tag{29}$$

Proposition 3.1 (Bayes risk decomposition) *For any deterministic reconstructor $\mathcal{A}(Y_j^\star)$,*

$$\mathbb{E}\|\mathbf{X}_j - \mathcal{A}(Y_j^\star)\|_{M_h}^2 = \epsilon_{\mathrm{id},j}^2 + \mathbb{E}\|\mathcal{G}_j^\star(Y_j^\star) - \mathcal{A}(Y_j^\star)\|_{M_h}^2. \tag{30}$$

Thus $\epsilon_{\mathrm{id},j}$ captures irreducible ambiguity rather than a modeling slack against a deterministic inverse.

Proof. This is the standard orthogonal decomposition of conditional expectation in L^2 (Kallenberg, 2021); see Lemma D.7 for the mass-norm form used here.

Coarse-condition mismatch is captured by

$$\epsilon_{\mathrm{cond},j} := \left(\mathbb{E}\big[\|Y_j - Y_j^\star\|_{\mathcal{Y}_H}^2\big]\right)^{1/2}, \tag{31}$$

with $\|\cdot\|_{\mathcal{Y}_H} = \|\cdot\|_{M_H}$ on $\mathbb{R}^{N_H}$ (single coarse snapshot). Thus $\epsilon_{\mathrm{cond},j}$ is an RMS mismatch in the *same* mass norm carried by Assumption 5; the factor $L_{G,j}\epsilon_{\mathrm{cond},j}$ in Proposition 3.2 is therefore taken with respect to pairwise differences $Y - \widetilde{Y} \in \mathbb{R}^{N_H}$ measured by $\|\cdot\|_{M_H}$, not a product norm from an obsolete two-component condition. Under standard coarse-solver consistency (Lemma D.2),

$$\epsilon_{\mathrm{cond},j} \lesssim H^{q_H} + (\Delta T_j)^{r_H} + h^2 + \Delta t_h. \tag{32}$$

Let $\widehat{\mathbf{X}}_j = \mathcal{G}_{\vartheta_j,K}(Y_j)$ and define the anchor RMS risk

$$\delta_j := \left(\mathbb{E}\big[\|\mathbf{X}_j - \widehat{\mathbf{X}}_j\|_{M_h}^2\big]\right)^{1/2}. \tag{33}$$

Proposition 3.2 (Modular anchor bound) *Assume Assumption 5 holds on a compact set containing the supports of $Y_j^\star$ and Y_j. Define the* algorithmic gap

$$\epsilon_{\mathrm{alg},j} := \left\|\mathcal{G}_j^\star(Y_j) - \mathcal{G}_{\vartheta_j,K}(Y_j)\right\|_{\mathbb{E},M_h}, \tag{34}$$

where $\mathcal{G}_j^\star(Y_j)$ denotes $\mathbb{E}[\mathbf{X}_j \mid Y_j^\star = Y]$ evaluated at $Y = Y_j$ under the same *scenario ω as $Y_j^\star$, i.e. the regular conditional expectation composed with the measurable map $\omega \mapsto Y_j(\omega)$ (equivalently, $\mathcal{G}_j^\star(Y_j)$ is well-defined on the joint support of $(Y_j, Y_j^\star)$ via coupling* (27)*; see Appendix D). Equivalently one may write $\mathbb{E}[\mathbf{X}_j \mid Y_j^\star]\big|_{Y_j^\star = Y_j}$ when $(Y_j, Y_j^\star)$ are synchronized along ω. Then*

$$\delta_j \le \epsilon_{\mathrm{id},j} + L_{G,j}\epsilon_{\mathrm{cond},j} + \epsilon_{\mathrm{alg},j}. \tag{35}$$

Proof sketch. Insert $\pm\mathcal{G}_j^\star(Y_j^\star)$ and $\pm\mathcal{G}_j^\star(Y_j)$ and apply Lipschitz continuity (88).

The core rollout bound uses anchor risk δ_j, which can be estimated from held-out HR anchors, without assuming multigrid improvement. The Bayes algorithmic gap $\epsilon_{\mathrm{alg},j}$ is generally unknown; the following optional refinement characterizes it under additional sampler assumptions.

Lemma 3.3 (Optional sampler-specific refinement) *Under Lemma D.8, Remark 13, Assumption 4, and the base diffusion error scale* (77)*, the algorithmic gap admits the sufficient bound*

$$\epsilon_{\mathrm{alg},j} \le C_{\mathrm{dyn},j}\,\eta_{\mathrm{dyn},j} + \rho_{\mathrm{mg},j}\,\epsilon_{\mathrm{D},j}(K) + \epsilon_{\mathrm{mg},j}. \tag{36}$$

Proof sketch. Insert the base diffusion map $\mathcal{D}_{\theta_j,K}$ between $\mathcal{G}_{\vartheta_j,K}$ and $\mathcal{G}_j^\star$, then apply Lemma D.8 and Assumption 4; see Appendix D.

Corollary 3.4 (Optional sampler-specific anchor bound) *Combining* (35)–(36)*,*

$$\delta_j \le \epsilon_{\mathrm{id},j} + L_{G,j}\epsilon_{\mathrm{cond},j} + C_{\mathrm{dyn},j}\,\eta_{\mathrm{dyn},j} + \rho_{\mathrm{mg},j}\,\epsilon_{\mathrm{D},j}(K) + \epsilon_{\mathrm{mg},j}, \tag{37}$$

with $\epsilon_{\mathrm{D},j}(K)$ from (77)*. Assumption 4 is an explicit* conditional diagnostic*: it encodes empirically checkable multigrid lift rather than a first-principles contraction theorem.*

Proposition 3.2 separates statistical ambiguity $\epsilon_{\mathrm{id},j}$, coarse mismatch $L_{G,j}\epsilon_{\mathrm{cond},j}$, and the remaining algorithmic gap $\epsilon_{\mathrm{alg},j}$. Lemma 3.3 is optional: it refines $\epsilon_{\mathrm{alg},j}$ using dynamic-proxy slack, base diffusion reconstruction error, and the ReMD improvement inequality (86). Proposition D.6 concerns an alternative score-trained model under additional assumptions; it is not a guarantee for the P1 sampler.

Remark 1 (Distributional scope of Lipschitz regularity) *Assumption 5 is imposed only on* $\mathrm{supp}(\mathcal{D}_j)$ *and may fail under distribution shift; near topological transitions of interfaces, coarse-to-fine conditioning can be stiff. Local moduli such as $L_{G,j}^{\mathrm{loc}}(r) := \sup\{\|\mathcal{G}_j^\star(Y) - \mathcal{G}_j^\star(\widetilde{Y})\|_{M_h}/\|Y - \widetilde{Y}\|_{\mathcal{Y}_H}\}$ over $\|Y - \widetilde{Y}\|_{\mathcal{Y}_H} \le r$ within the training/test support formalize this limitation (Appendix D).*

Lemma 3.5 (Segmentwise propagation) *For any fixed finite $d \geq 1$, fix j and $0 \leq m \leq M_j$. Under Assumption 1,*

$$\left\|\mathbf{u}_{h,j}^m - \widehat{\mathbf{u}}_{h,j}^m\right\|_{\mathbb{E},M_h} \leq \exp\left(L_B\, m\Delta t_h\right) \delta_j. \tag{38}$$

Proof. This follows from the local Lipschitz stability of $\mathbf{\Phi}_h^{\Delta t_h}$ under the CFL condition; see Theorem D.9.

3.2 Global error bound

Theorem 3.6 (Global error, modular form) *Fix any finite $d \geq 1$. Assume Theorem D.1, Lemma D.2, Assumption 1, and Lemma 3.5. Let δ_j be the anchor RMS risk* (33)*. Then*

$$\begin{aligned}&\max_{j,m}\left(\mathbb{E}\left[\|\mathcal{S}_h u(t_{j,m}) - \widehat{\mathbf{u}}_{h,j}^m\|_{M_h}^2\right]\right)^{1/2}\\ &\quad\leq C_{\mathrm{FD}}(h^2 + \Delta t_h) + \max_j \exp(L_B \Delta T_j)\, \delta_j.\end{aligned} \tag{39}$$

Proof sketch. Combine (26) with Theorem D.1 and Theorem D.9; see Appendix D.8. The value of Theorem 3.6 is thus a modular interface that converts the measurable anchor RMS risk δ_j into an end-to-end rollout guarantee while keeping discretization bias and segment-local amplification explicit.

Remark 2 (Segmentation amplification and Explicit δ_j constituents) *Re-anchoring limits propagation of each anchor error to ΔT_j, yielding $\exp(L_B \Delta T_j)$ instead of the full-horizon factor $\exp(L_B T)$. This exposes a tradeoff: shorter segments reduce the propagation factor, but need not improve accuracy if their anchor risks increase. Corollary 3.4 supplies a fully explicit upper bound on δ_j under the optional sampler assumptions; inserting it into* (39) *recovers the expanded global inequality from earlier drafts. Segment-interface jumps* (25) *(tail-to-head, not anchor-to-anchor) are not included in Theorem 3.6; P1 reports their empirical magnitude separately in Appendix G.1.*

4 Experiments

4.1 Setup and baselines

We use two explicitly separated evaluation suites because the earlier ablation artifact and the new controlled comparison employ different boundary protocols and sampling units. The first suite simulates (2) on $\Omega = [0,1]^2$ with $\varepsilon = 0.01$, homogeneous Neumann data, and HR grid 256×256 (h=1/255) versus LR 64×64 (H=1/63). The default segment stride is M_j=20, hence ΔT_j=0.20 at Δt_h=10^{-2}. Reverse diffusion uses K=15 reverse steps with the exponential schedule recorded in Appendix F; an inference-only sweep tests $K \in \{2, 5, 10, 15\}$. Held-out evaluation uses 200 segments except for stride 40, for which the complete split contains 100 segments. High-resolution labels enter training only at segment heads; within a default segment this is one labeled head among 21 reported time levels. The second suite uses periodic boundaries and a trajectory-level split and is described separately in Section 4.3; results from the two suites are not pooled or treated as paired observations.

We compare coarse bicubic prolongation followed by the same fine Euler rollout (Interp); the correction-enabled instantiation (Full); variants without the Allen–Cahn residual (No-AC), with the original generic ReMD residual (Old-ρ), and without M2NO correction (No-M2NO); two-frame conditioning, increment-proxy, and denoising-only variants; and an oracle numerical baseline initialized from the true HR anchor (GT-anchor+Euler). The denoising-only run also disables M2NO and is therefore a compound ablation. No-M2NO retains the raw residual update while disabling learned multigrid refinement. The P1 unsegmented comparator reuses the original-ReMD checkpoint, reconstructs only the initial anchor, and advances the full trajectory with Euler integration. Appendix G reports component diagnostics, segment-length and sampling-step sweeps, and a $32^2 \to 128^2$ resolution variant. The pipeline still executes every fine Euler interior step, so we make no automatic speedup claim.

4.2 MAIN RESULTS, ABLATIONS AND DIAGNOSTICS

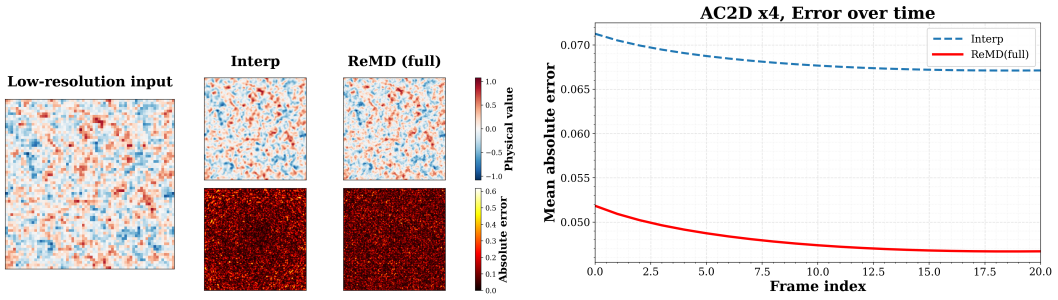
Table 1 summarizes accuracy on the fixed AC2D validation split. Because the available artifact contains pooled point estimates rather than independent training replicates, we state this scope explicitly and do not present the entries as mean±standard deviation.

Table 1: **AC2D results at segment stride 20** (200 held-out segments). Lower is better in every metric. Bold marks the best learned variant in each column. Entries are pooled point estimates from one trained checkpoint per configuration, not mean±standard deviation over independent training runs. INTERP artifacts contain only anchor and rollout-mean RMSE under this exact metric definition. †Denoising-only also disables M2NO.

Method	Anchor RMSE	Final RMSE	Final rel. L^2	Rollout RMSE	E_{spec}	$E_{\mathcal{E}}$
INTERP	0.091759	–	–	0.087710	–	–
FULL	0.066113	0.059634	0.189007	0.060986	177.415	0.002201
NO-AC	0.067059	0.060089	0.190707	0.061678	179.364	0.002079
OLD-ρ	0.066188	0.062553	0.197834	0.062845	219.716	0.004254
NO-M2NO	0.063747	0.060300	0.190329	0.060501	202.325	0.003786
Two-frame	0.064964	0.058465	0.185236	0.059811	163.677	0.002099
Increment proxy	0.064487	0.058363	0.184873	0.059598	170.246	0.002708
Denoising-only†	0.065039	0.058602	0.185699	0.059927	176.056	0.002629
GT-ANCHOR+EULER	0	0.012722	0.039221	0.007418	180.300	0.000765

Table 1 supports three conclusions. First, FULL lowers anchor and rollout-mean RMSE relative to INTERP by 28.0% and 30.5%. The oracle GT-ANCHOR+EULER rollout RMSE is only 0.007418, showing that learned anchor reconstruction, rather than the fine Euler interior, dominates the remaining error. Second, the increment proxy provides the best final-step and rollout RMSE among learned variants (2.1% and 2.3% below FULL), while two-frame conditioning gives the lowest spectral error. Third, the correction ablation does not support a strong empirical multigrid-contraction claim: NO-M2NO has slightly lower anchor and rollout RMSE than FULL, and the pooled $\widehat{\rho}_{\text{mix}}$ values are approximately one (Appendix G). The Allen–Cahn residual gives a modest rollout benefit over NO-AC, but different diagnostics favor different variants. We therefore treat the correction and regularity assumptions as conditional analytical devices rather than conclusions established by these experiments.

The segment-length sweep further shows that stride 40 is clearly worse in anchor error and energy mismatch than strides 10 or 20, while the inference step sweep is nearly flat from $K = 2$ to 15. These observations are consistent with segment-local error amplification but do not by themselves verify its worst-case exponential upper bound.



(a) Held-out anchor (`IBVP2_s180`, segment 0): LR input (left), INTERP/FULL HR anchors with shared color scales and paired absolute-error maps.

(b) Mean absolute error vs. fine index averaged over 200 segments (dashed INTERP, solid FULL; identical Euler rollout (12)).

Figure 1: Qualitative AC2D validation accompanying Table 1; supplementary diagnostics are reported in Appendix G.

4.3 CONTROLLED COMPARISON WITH ORIGINAL REMD

To separate the effect of segmentation from the change in residual parameterization, we additionally trained FULL and original ReMD under a common training protocol and evaluated five pipelines on 20 held-out complete trajectories. This independent experiment uses periodic boundaries, a trajectory-level $160/20/20$ train/validation/test split, stride 20, $K = 15$, and identical optimization budgets for the two learned models; it is therefore reported separately from the pooled-segment study in Table 1. Both original-ReMD rows reuse one checkpoint: the unsegmented run uses only an initial anchor, whereas segmentation uses ten coarse-conditioned anchors. Both perform 200 Euler steps. The comparison measures the benefit of repeated reconstruction with additional coarse observations and sampler calls, rather than an equal-observation or equal-compute advantage. Full numerical details are given in Appendix G.1.

Table 2: Controlled comparison on 20 complete held-out trajectories under the common periodic-boundary protocol. Lower is better; bold marks the numerically best accuracy or energy value among the two segmented learned methods and does not imply statistical significance. GT-ANCHOR+Euler is an oracle. Complete metrics and multi-seed results appear in Appendix G.1.

Method	Rollout RMSE	Rel. L^2	$E_{\mathcal{E}}$	Time (s/trajectory)
INTERP+Euler	0.088954	0.280147	0.002072	0.039
Original ReMD, unsegmented	0.080480	0.245475	0.007573	1.263
Original ReMD, segmented	**0.061956**	**0.195034**	0.001547	12.232
FULL ACReMD	0.062606	0.197309	**0.001396**	9.786
GT-ANCHOR+Euler	0.004780	0.012677	0.000602	0.039

Under this controlled protocol (Table 2), segmentation provides the largest observed field-error reduction: original ReMD with segmentation reduces rollout RMSE from 0.080480 to 0.061956. The proposed AC-aware residual does not improve every metric over that segmented control—FULL obtains rollout RMSE 0.062606 versus 0.061956—but it reduces mean energy mismatch from 0.001547 to 0.001396. The single timing run records 12.232 s and 9.786 s, respectively, but without repeated timing trials we treat these values as descriptive measurements rather than evidence of an intrinsic speed advantage. Three independent FULL trainings give rollout RMSE 0.063063 ± 0.000401 (between-seed standard deviation), indicating low training variability under this protocol. The original-ReMD controls use one training seed; statistical significance of the small differences between the segmented models has therefore not been established.

5 LIMITATIONS AND CONCLUSION

Evidence is limited to synthetic 2D Allen–Cahn data with one interface width and initial-condition family; robustness to other equations and 3D remains untested. The method requires supervised HR anchors and every fine Euler step, while interface jumps, maximum-principle violations, and distribution shift remain possible. Multigrid and Lipschitz assumptions are conditional.

The segmented diffusion–Euler hybrid combines sparse learned re-anchoring with deterministic interiors. Its dimension-generic analysis converts anchor risk into an end-to-end rollout bound. Experiments identify segmentation as the main field-error gain; the Allen–Cahn residual lowers energy mismatch, whereas optional multigrid correction is nearly neutral. We claim these measurable tradeoffs, not uniform superiority or acceleration.

AI USE STATEMENT

Generative AI tools assisted with language editing, manuscript organization, code–manuscript consistency checks, revisions to mathematical formulations and pseudocode. AI assistance also informed discussions of theoretical scope and possible experimental diagnostics. Reported numerical measurements originate from the experimental code and result files. The authors take responsibility for the final content, including the correctness of mathematical claims, implementation descriptions, citations, and numerical results.

Ethics statement

This work studies numerical and machine-learning methods on synthetic Allen–Cahn trajectories. It uses no human subjects, personal data, or sensitive attributes. The main risks are scientific rather than societal: learned surrogates may fail outside their training distribution and should not be used as a substitute for validated physical simulation in safety-critical settings.

Reproducibility statement

The main text states the discretization, conditioning variables, baselines, metrics, validation-set size, timing scope, and the distinct Neumann-boundary and periodic-boundary evaluation protocols. Appendix G specifies model and solver hyperparameters, data construction, pseudocode, and the protocol for the theory-linked diagnostics. Complete proofs are provided in Appendix D.

References

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.

James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, 2 edition, 1985. doi: 10.1007/978-1-4757-4286-2.

W. J. Boettinger, J. A. Warren, C. Beckermann, and A. Karma. Phase-field simulation of solidification. *Annual Review of Materials Research*, 32:163–194, 2002.

L. Bronsard and R. V. Kohn. Motion by mean curvature as the singular limit of Ginzburg–Landau dynamics. *Journal of Differential Equations*, 90(2):211–237, 1991.

J. W. Cahn. On spinodal decomposition. *Acta Metallurgica*, 9(9):795–801, 1961.

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.

X. Chen. Generation and propagation of interfaces for reaction-diffusion equations. *Journal of Differential Equations*, 96(1):116–141, 1992.

P. de Mottoni and M. Schatzman. Geometrical evolution of developed interfaces. *Transactions of the American Mathematical Society*, 347(5):1533–1589, 1995.

Q. Du, L. Ju, X. Li, and Z. Qiao. Stabilized linear semi-implicit schemes for the nonlocal Cahn–Hilliard equation. *Journal of Computational Physics*, 363:39–54, 2018.

L. C. Evans, H. M. Soner, and P. E. Souganidis. Phase transitions and generalized motion by mean curvature. *Communications on Pure and Applied Mathematics*, 45(9):1097–1123, 1992.

D. J. Eyre. Unconditionally gradient stable time marching the Cahn–Hilliard equation. In *Computational and Mathematical Models of Microstructural Evolution*, pp. 39–46. Materials Research Society, 1998.

X. Feng and A. Prohl. Numerical analysis of the Allen–Cahn equation and approximation for mean curvature flows. *Numerische Mathematik*, 94:33–65, 2003.

X. Feng and A. Prohl. Analysis of a fully discrete finite element method for the phase field model and approximation of its sharp interface limits. *Mathematics of Computation*, 73(246):541–567, 2004.

K. Fukami, K. Fukagata, and K. Taira. Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics*, 870:106–120, 2019.

K. Fukami, K. Fukagata, and K. Taira. Machine-learning-based spatio-temporal super resolution reconstruction of turbulent flows. *Journal of Fluid Mechanics*, 909:A9, 2021.

J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.

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

Olav Kallenberg. *Foundations of Modern Probability*. Springer, Cham, 3 edition, 2021. doi: 10.1007/978-3-030-61871-1.

N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

R. J. LeVeque. *Finite Difference Methods for Ordinary and Partial Differential Equations*. SIAM, 2007.

Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Zhihao Li, Shengwei Dong, Chuang Yi, Junxuan Gao, Zhilu Lai, Zhiqiang Liu, Wei Wang, and Guangtao Zhang. Physics-consistent diffusion for efficient fluid super-resolution via multiscale residual correction, 2026. URL `https://arxiv.org/abs/2603.00149`. arXiv preprint.

B. Liu, J. Tang, H. Huang, and X.-Y. Lu. Deep learning methods for super-resolution reconstruction of turbulent flows. *Physics of Fluids*, 32:025105, 2020.

L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3: 218–229, 2021.

L. Modica and S. Mortola. Un esempio di Γ-convergenza. *Bollettino dell'Unione Matematica Italiana*, 14:285–299, 1977.

K. W. Morton and D. F. Mayers. *Numerical Solution of Partial Differential Equations*. Cambridge University Press, 2 edition, 2005.

N. Provatas and K. Elder. *Phase-Field Methods in Materials Science and Engineering*. Wiley-VCH, 2010.

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.

J. Shen and X. Yang. Numerical methods for Allen–Cahn and Cahn–Hilliard equations. *Discrete and Continuous Dynamical Systems*, 28(4):1669–1691, 2010.

J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021a.

Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021b.

K. Stengel, A. Glaws, D. Hettinger, and R. N. King. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29):16805–16815, 2020.

J. C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. SIAM, 2 edition, 2004.

M. Takamoto, T. Praditia, R. Leiteritz, D. MacKinlay, F. Alesiani, D. Pflüger, and M. Niepert. PDEBench: An extensive benchmark for scientific machine learning. In *Advances in Neural Information Processing Systems*, 2022.

A Additional notation and discretization details

A.1 Initial data, boundary conditions, and energy

Initial data, Neumann boundary conditions, and the Ginzburg–Landau energy are recorded in (4)–(6) (Section 2.1); periodic or Dirichlet cases only require stencil modifications at boundaries.

A.2 Finite-difference Laplacian and mass norms

Grid functions $\mathbf{v}_\eta \in \mathbb{R}^{N_\eta}$ on meshes $\eta \in \{h, H\}$ use the discrete inner product

$$M_\eta := \eta^d I_{N_\eta}, \qquad \|\mathbf{v}_\eta\|_{M_\eta} := \left(\eta^d \sum_{i=1}^{N_\eta} |v_{\eta,i}|^2 \right)^{1/2}. \tag{40}$$

Away from boundaries, the standard second-order Laplacian stencil reads

$$(L_\eta \mathbf{v}_\eta)_{\mathbf{i}} = \frac{1}{\eta^2} \sum_{m=1}^{d} \left(v_{\mathbf{i}+\mathbf{e}_m} - 2v_{\mathbf{i}} + v_{\mathbf{i}-\mathbf{e}_m} \right). \tag{41}$$

Neumann boundaries $\nabla u \cdot \mathbf{n} = 0$,use ghost values set as $v_{i+\mathbf{e}_m} = v_{i-\mathbf{e}_m}$; periodic grids wrap indices.

A.3 Restriction/prolongation and coarse solvers

We assume approximate stability/consistency of (R_{Hh}, P_{hH}) beyond (8); see (Li et al., 2026) for operator-norm discussions used in residual scaling. These operators are assumed to satisfy the following stability bounds:

$$\|R_{Hh}\mathbf{v}_h\|_{M_H} \le C_R \|\mathbf{v}_h\|_{M_h}, \qquad \|P_{hH}\mathbf{v}_H\|_{M_h} \le C_P \|\mathbf{v}_H\|_{M_H},$$

with constants C_R, C_P independent of h and H The composition is assumed to approximate the identity on coarse-grid functions:

$$\|R_{Hh}P_{hH}\mathbf{v}_H - \mathbf{v}_H\|_{M_H} \le C_{\text{prol}} H^q \|\mathbf{v}_H\|_{M_H}$$

for some $q \ge 2$ (exact nested prolongation on structured grids achieves $q = 2$ for smooth functions).

Remark 3 (Coarse trajectories from FEM or external solvers) *If coarse states are produced by FEM, finite differences, or an external solver, sample or project onto the coarse grid at times $\{T_j\}$ (or $\{t_n\}$). It suffices that*

$$\max_{0 \le n \le N_t} \|\mathbf{u}_H^n - \mathcal{S}_H u(t_n)\|_{M_H} \le C \left(H^{q_H} + \Delta t \right), \tag{42}$$

with $\mathcal{S}_H$ sampling/averaging and spatial order q_H after grid representation; (42) *applies with $\{T_j\}$ in place of uniform coarse steps when segments drive the coarse ladder.*

Remark 4 (Optional one-step residual across many fine steps) *Instead of comparing vector fields at time T_j via $\mathbf{d}_H^j = \mathbf{F}_H(\mathbf{y}_j)$ in* (18)*, one may directly relate the coarse increment to the fine-grid rollout over the entire segment:*

$$\mathbf{d}_H^j = \frac{\mathbf{y}_j^+ - \mathbf{y}_j}{\Delta T_j} \quad \textit{with residual} \quad \mathbf{r}^{(roll)}(\mathbf{x}) = P_{hH} \left(\frac{\mathbf{y}_j^+ - \mathbf{y}_j}{\Delta T_j} - R_{Hh} \frac{\Phi_h^{M_j \Delta t_h}(\mathbf{x}) - \mathbf{x}}{M_j \Delta t_h} \right).$$

This replaces the instantaneous vector-field mismatch with an average over M_j fine Euler steps, at the cost of M_j fine-map evaluations per reverse step of the diffusion sampler.

B Implemented anchor sampler

The controlled P1 implementation is specified by (13)–(15). It predicts clean states and refines them before a stochastic posterior update. In the notation of the general error analysis, the clean residual and base drift are

$$\widehat{\mathbf{x}}_{0,\theta} = \mathbf{v}_k, \qquad \widehat{\mathbf{s}}_{0,\theta} = \mathbf{v}_k - \mathbf{c}_j, \qquad \mathbf{b}_{\theta,k} = \mathbf{T}_{\theta,k} - \mathbf{x}_k. \tag{43}$$

The data and full residuals are

$$\mathbf{r}_{\text{data}}(\mathbf{v}; \mathbf{y}_j) = P_{hH}(\mathbf{y}_j - R_{Hh}\mathbf{v}), \qquad \mathbf{r}_{j,k}(\mathbf{v}) = \mathbf{r}_{\text{data}}(\mathbf{v}; \mathbf{y}_j) + \lambda_k P_{hH}(\mathbf{d}_H^j - R_{Hh}\mathbf{F}_h(\mathbf{v})). \tag{44}$$

Learned and raw correction. The M2NO module $\mathcal{S}_\varphi$ takes a state, a residual, and a timestep as inputs. Its output is added to the raw residual:

$$\mathcal{M}_{b,\varphi,k}(\mathbf{v},\mathbf{r}) = \mathbf{r} + b\,\mathcal{S}_\varphi(\mathbf{v},\mathbf{r},k), \qquad \vartheta = (\theta,\varphi). \tag{45}$$

Here $b = 0$ removes only the learned refinement. An entirely uncorrected analytical comparator requires removing the raw residual term as well and is not the reported NO-M2NO ablation. Equivalently, the reverse step is

$$\mathbf{x}_{k-1} = \mathbf{x}_k + \mathbf{b}_{\theta,k} + \frac{\eta_k - \eta_{k-1}}{\eta_k} a_k \mathcal{M}_{b,\varphi,k}(\mathbf{v}_k, \mathbf{r}_{j,k}(\mathbf{v}_k)) + \sigma_k \mathbf{z}_k. \tag{46}$$

B.1 Pseudocode for anchor and segmented solvers

Algorithm 1 P1 clean-state anchor reconstruction at segment head T_j

Require: Coarse field $\mathbf{y}_j$, $\mathbf{c}_j = P_{hH}\mathbf{y}_j$; trained $f_\theta, \mathcal{S}_\varphi$; switch $b \in \{0,1\}$; $0 = \eta_0 < \cdots < \eta_K$, $\kappa = 2$.
Ensure: HR anchor $\widehat{\mathbf{u}}^0_{h,j}$.
Draw $\boldsymbol{\xi} \sim \mathcal{N}(0, I)$; set $\mathbf{x}_K \leftarrow \mathbf{c}_j + \kappa\sqrt{\eta_K}\boldsymbol{\xi}$.
for $k = K, \ldots, 1$ **do**
  $\mathbf{v}_k \leftarrow f_\theta(\mathbf{x}_k, k, \mathbf{c}_j)$.
  $\mathbf{r}_k \leftarrow \mathbf{r}_{j,k}(\mathbf{v}_k)$ using (17).
  $\mathbf{e}_k \leftarrow \mathbf{r}_k$.
  if $b = 1$ **then**
    $\mathbf{e}_k \leftarrow \mathbf{e}_k + \mathcal{S}_\varphi(\mathbf{v}_k, \mathbf{r}_k, k)$.
  $\widetilde{\mathbf{v}}_k \leftarrow \mathbf{v}_k + (\eta_k/4)\mathbf{e}_k$.
  $\boldsymbol{\mu}_k \leftarrow (\eta_{k-1}/\eta_k)\mathbf{x}_k + (1 - \eta_{k-1}/\eta_k)\widetilde{\mathbf{v}}_k$.
  $\sigma_k \leftarrow \kappa\sqrt{\eta_{k-1}(\eta_k - \eta_{k-1})/\eta_k}$.
  Draw $\mathbf{z}_k \sim \mathcal{N}(0, I)$; set $\mathbf{x}_{k-1} \leftarrow \boldsymbol{\mu}_k + \sigma_k \mathbf{z}_k$.
return $\widehat{\mathbf{u}}^0_{h,j} \leftarrow \mathbf{x}_0$.

Algorithm 2 P1 segmented diffusion–Euler reconstruction

Require: Observed coarse snapshots $\{\mathbf{y}_j\}_{j=0}^{J-1}$, trained anchor map $\mathcal{G}_{\vartheta,K}$, fine Euler map $\boldsymbol{\Phi}_h^{\Delta t_h}$, segment lengths M_j.
for $j = 0, \ldots, J-1$ **do**
  Reconstruct $\widehat{\mathbf{u}}^0_{h,j} \leftarrow \mathcal{G}_{\vartheta,K}(\mathbf{y}_j)$ using Algorithm 1.
  for $m = 0, \ldots, M_j - 1$ **do**
    $\widehat{\mathbf{u}}^{m+1}_{h,j} \leftarrow \boldsymbol{\Phi}_h^{\Delta t_h}(\widehat{\mathbf{u}}^m_{h,j})$.
Concatenate segments, retaining the new head at shared times and the final tail.

C Alternative training objectives for analysis

C.1 Optional noise-prediction and sampler-level objectives

The objectives below describe an alternative training formulation retained for the conditional analysis. They were not used to produce the P1 results. P1 instead uses the single-stage relative-L^2 loss in (22). For the alternative noise-prediction formulation only, define

$$\mathbf{x}^{\mathrm{alt}}_{j,k} = \mathbf{c}_j + \sqrt{\bar{a}_k}(\mathbf{x}^\star_j - \mathbf{c}_j) + \sqrt{1 - \bar{a}_k}\boldsymbol{\xi}, \qquad \bar{a}_k = \prod_{i=1}^{k}(1 - \beta_i). \tag{47}$$

Any noise-prediction expressions below refer to this alternative perturbation, not the implemented training perturbation in (21).

Table 3: Attribution of inherited and new components. The proposed method is a segmented diffusion–Euler framework, not a new diffusion sampler family; the ReMD multigrid correction is optional.

Component	Source	Role in this paper
Conditional diffusion and posterior sampling	ReMD (Li et al., 2026); diffusion literature	Reused as the high-resolution anchor generator.
Optional residual–multigrid correction	ReMD (Li et al., 2026)	Inherited anchor refinement, evaluated by ablation rather than required by the core method.
Sparse re-anchoring plus fine Euler rollout	This work	Separates learned reconstruction from deterministic segment dynamics.
Bayes–anchor–Euler error decomposition	This work	Separates ambiguity, conditioning, algorithmic gap, finite-difference bias, and segment-local amplification.

$$\mathcal{D}_j = \left\{ \left(\mathbf{x}^\star_{p,j}, \mathbf{y}_{p,j} \right) \right\}_{p=1}^{N_{\mathrm{tr}}}, \qquad Y_{p,j} := \mathbf{y}_{p,j}, \qquad \mathbf{y}^+_{p,j} := \mathbf{\Psi}_H^{\Delta T_j}(\mathbf{y}_{p,j}), \tag{48}$$

where $\mathbf{x}^\star_{p,j}$ is the high-resolution anchor at coarse time T_j. The one-step-ahead coarse state $\mathbf{y}^+_{p,j}$ is computed from the coarse integrator whenever increment residuals (19) are used; it is *not* an extra conditioning argument to $\boldsymbol{\epsilon}_{\theta_j}$. Let

$$\mathbf{c}_{p,j} := P_{hH}\mathbf{y}_{p,j}. \tag{49}$$

For diffusion pretraining, sample

$$k \sim \mathrm{Unif}\{1, \dots, K\}, \qquad \boldsymbol{\xi} \sim \mathcal{N}(0, I_{N_h}),$$

and form the residual-centered noisy state

$$\mathbf{x}_{p,j,k} = \mathbf{c}_{p,j} + \sqrt{\bar{a}_k}(\mathbf{x}^\star_{p,j} - \mathbf{c}_{p,j}) + \sqrt{1-\bar{a}_k}\,\boldsymbol{\xi}. \tag{50}$$

The denoising loss is

$$\mathcal{L}_{\mathrm{den},j}(\theta_j) = \mathbb{E}_{p,k,\boldsymbol{\xi}} \left[\left\| \boldsymbol{\xi} - \boldsymbol{\epsilon}_{\theta_j}(\mathbf{x}_{p,j,k}, k, Y_{p,j}) \right\|_2^2 \right]. \tag{51}$$

An optional alternative is to supervise the output of an unrolled sampler. Let

$$\mathbf{x}^G_{p,j,0} = \mathcal{G}_{\vartheta_j, K'}(Y_{p,j}; \boldsymbol{\xi}) \tag{52}$$

be the output of the full ReMD sampler unrolled for $K' \le K$ reverse steps. The sampler-level loss is

$$\begin{aligned} \mathcal{L}_{\mathrm{roll},j}(\vartheta_j) = {} & \mathbb{E}_{p,\boldsymbol{\xi}} \left[\left\| \mathbf{x}^\star_{p,j} - \mathbf{x}^G_{p,j,0} \right\|_{M_h}^2 \right] \\ & + \gamma_{\mathrm{data}} \mathbb{E}_{p,\boldsymbol{\xi}} \left[\left\| R_{Hh}\mathbf{x}^G_{p,j,0} - \mathbf{y}_{p,j} \right\|_{M_H}^2 \right] \\ & + \gamma_{\mathrm{AC}} \mathbb{E}_{p,\boldsymbol{\xi}} \left[\left\| \Delta T_j \left(R_{Hh}\mathbf{F}_h(\mathbf{x}^G_{p,j,0}) - \mathbf{d}^j_{H,p} \right) \right\|_{M_H}^2 \right], \end{aligned} \tag{53}$$

with $\mathbf{F}_h$ in (16); the ΔT_j factor gives this alternative penalty state units. The default choice is $\mathbf{d}^j_{H,p} = \mathbf{F}_H(\mathbf{y}_{p,j})$, consistent with (18); the increment form $(\mathbf{y}^+_{p,j} - \mathbf{y}_{p,j})/\Delta T_j$ may be used when the coarse solver is explicit Euler (19). The total training objective is

$$\min_{\vartheta_j} \quad \mathcal{L}_{\mathrm{den},j}(\theta_j) + \lambda_{\mathrm{roll}}\mathcal{L}_{\mathrm{roll},j}(\vartheta_j). \tag{54}$$

In this alternative formulation, one first trains the noise predictor using (51), and then fine-tunes the full sampler with (53). This ensures that trainable parameters of $\mathcal{S}_\varphi$ receive gradients from the final reconstructed anchor.

This optional objective is not the objective used by the delivered P1 checkpoints.

D Proofs of the error estimates

Roadmap. Appendix D follows the main-text logic: finite-difference stability and classical error, coarse–fine conditioning, Bayes risk and anchor reconstruction, base diffusion error $\epsilon_{\mathrm{D},j}$, ReMD improvement, segment propagation, and finally assembly into Theorem 3.6.

D.1 CLASSICAL FINITE-DIFFERENCE DISCRETIZATION ERRORS

Assumption 1 (Regularity, boundedness, and finite-difference CFL condition) *Fix any finite $d \geq 1$. Let L_h be the dimension-d central-difference Laplacian on a Cartesian tensor-product grid, with a periodic, Dirichlet, or consistent ghost-cell Neumann boundary treatment, and let $M_h = h^d I$ (or the corresponding positive diagonal quadrature mass matrix). Assume that the exact solution is sufficiently smooth for the chosen finite-difference stencil. For the second-order central difference stencil, it is enough to assume, for example,*

$$u \in L^\infty(0,T;C^4(\overline{\Omega})), \qquad \partial_{tt} u \in L^\infty(0,T;L^2(\Omega)). \tag{55}$$

Assume also that all exact, reference, coarse, and ReMD-generated grid states remain uniformly bounded by B in the discrete ℓ^∞ norm (for generated anchors this is an unverified *stability requirement unless enforced, e.g. by projecting $\mathcal{G}_{\vartheta_j,K}(Y_j)$ into $[-B,B]$ or using a bounded head). Finally, assume*

$$\varepsilon^2 \Delta t_h \, \lambda_{\max}(-L_h) \leq 2 - c_0, \qquad c_0 \in (0,2), \tag{56}$$

and analogously for any explicit coarse integrator with segment lengths ΔT_j. Since $\lambda_{\max}(-L_h) \leq C_d h^{-2}$ for the standard dimension-d Laplacian (for the usual uniform tensor grid, one may take C_d proportional to $4d$), this is the dimension-dependent parabolic CFL condition $\Delta t_h \lesssim_d h^2/\varepsilon^2$.

Remark 5 (Dimension dependence) *The analysis is valid for every fixed finite d. Constants such as C_d, C_{FD}, C_R, C_P, and regularity bounds are allowed to depend on d; no estimate is asserted uniformly in the limit $d \to \infty$. The statistical Bayes-risk identities and segmentwise Gronwall argument are finite-dimensional norm arguments and introduce no additional restriction to $d = 2$ or 3.*

Remark 6 (Neumann boundary compatibility) *Second-order ghost-cell Neumann treatments typically require sufficient smoothness of u and compatibility conditions at $\partial\Omega$ (e.g. $\partial_{\mathbf{n}} u = 0$ together with $\partial_{\mathbf{n}} \Delta u = 0$ for the Allen–Cahn solution) to attain the stated boundary truncation order. The proof of Theorem D.1 therefore separates interior consistency from boundary modifications of the stencil.*

Theorem D.1 (Classical finite-difference Euler error) *For every fixed finite $d \geq 1$, under Assumption 1, the second-order finite-difference Euler reference solution satisfies*

$$\max_{\substack{0 \leq j \leq J-1 \\ 0 \leq m \leq M_j}} \|\mathcal{S}_h u(t_{j,m}) - \mathbf{u}^m_{h,j}\|_{M_h} \leq C_{\mathrm{FD}} \left(h^2 + \Delta t_h + \|\mathcal{S}_h u^0 - \mathbf{u}^0_{h,0}\|_{M_h} \right). \tag{57}$$

More generally, if the fine spatial stencil has order q_h, then h^2 is replaced by h^{q_h}.

Proof. Let $e^m_{h,j} := \mathcal{S}_h u(t_{j,m}) - \mathbf{u}^m_{h,j}$ denote the error between the exact projected solution and the finite-difference Euler solution at coarse time segment j and fine step m.

Step 1: Local truncation errors. For the central-difference Laplacian, the spatial truncation error satisfies

$$\|L_h \mathcal{S}_h u(t) - \Delta u(t)\|_{M_h} \leq C_1 h^2,$$

where $C_1 = C_1(d,u)$ depends on the dimension and on the fourth spatial derivatives of u. For the explicit Euler temporal discretization, the local time truncation error satisfies

$$\left\| \frac{\mathcal{S}_h u(t+\Delta t_h) - \mathcal{S}_h u(t)}{\Delta t_h} - \partial_t u(t) \right\|_{M_h} \leq C_2 \Delta t_h,$$

with $C_2 = C_2(d,u)$ controlled by $\|\partial_{tt} u\|_{L^\infty(0,T;L^2(\Omega))}$ under the mass norm.

Step 2: Reaction term Lipschitz bound. For the Allen–Cahn reaction $\psi(s) = s - s^3$, restricted to $[-B,B]$, the Lipschitz constant is

$$L_B = \sup_{s \in [-B,B]} |\psi'(s)| = 1 + 3B^2.$$

Step 3: Error recursion. The Euler scheme gives the update

$$\mathbf{u}^{m+1}_{h,j} = \mathbf{u}^m_{h,j} + \Delta t_h \left(\varepsilon^2 L_h \mathbf{u}^m_{h,j} + \psi(\mathbf{u}^m_{h,j}) \right).$$

Subtracting the projected exact solution increment

$$\mathcal{S}_h u(t_{j,m+1}) = \mathcal{S}_h u(t_{j,m}) + \Delta t_h \left(\varepsilon^2 L_h \mathcal{S}_h u(t_{j,m}) + \psi(\mathcal{S}_h u(t_{j,m}))\right) + \tau_{h,j}^m,$$

where $\tau_{h,j}^m = O(h^2 + \Delta t_h)$, we obtain the error recursion

$$e_{h,j}^{m+1} = e_{h,j}^m + \Delta t_h \left(\varepsilon^2 L_h e_{h,j}^m + \psi(\mathcal{S}_h u(t_{j,m})) - \psi(\mathbf{u}_{h,j}^m)\right) - \tau_{h,j}^m.$$

Step 4: Stability and discrete Gronwall. By the CFL condition (56), the operator $I + \varepsilon^2 \Delta t_h L_h$ is stable in the M_h-norm. Using the Lipschitz bound of ψ, we get

$$\|e_{h,j}^{m+1}\|_{M_h} \le (1 + \Delta t_h L_B)\|e_{h,j}^m\|_{M_h} + \|\tau_{h,j}^m\|_{M_h}.$$

Applying the discrete Gronwall inequality over $m = 0, \dots, M_j - 1$, we obtain

$$\|e_{h,j}^{M_j}\|_{M_h} \le e^{L_B \Delta T_j} \left(\|e_{h,j}^0\|_{M_h} + C(h^2 + \Delta t_h)\right),$$

where $C = C(d, u, T)$ depends on the fixed dimension and on the stated solution regularity. Taking the maximum over all j and m gives the stated bound (57). □

D.2 Coarse–fine consistency

Assumption 2 (Transfer stability and sampling compatibility) *The restriction and prolongation operators satisfy*

$$\|R_{Hh}\mathbf{v}_h\|_{M_H} \le C_R\|\mathbf{v}_h\|_{M_h}, \qquad \|P_{hH}\mathbf{v}_H\|_{M_h} \le C_P\|\mathbf{v}_H\|_{M_H}. \tag{58}$$

Moreover, the sampling/restriction operators are compatible in the sense that

$$\|\mathcal{S}_H u(t_n) - R_{Hh}\mathcal{S}_h u(t_n)\|_{M_H} \le C_R^\star(H^2 + h^2), \qquad 0 \le n \le N_t. \tag{59}$$

If the coarse trajectory is generated by an external solver on $\{T_j\}$, assume

$$\max_{0\le j\le J} \|\mathbf{u}_H^j - \mathcal{S}_H u(T_j)\|_{M_H} \le C_H\Big(H^{q_H} + \max_{0\le j\le J-1}(\Delta T_j)^{r_H} + \Delta t_h\Big), \tag{60}$$

where q_H is spatial order, r_H is temporal order of the coarse integrator, and the Δt_h term accounts for fine-grid reference mismatch when comparing to sampled exact trajectories.

Lemma D.2 (Coarse–fine consistency) *Under Assumption 2 and Theorem D.1,*

$$\max_{0\le j\le J-1} \|\mathbf{u}_H^j - R_{Hh}\mathbf{u}_{h,j}^0\|_{M_H} \le C_{\mathrm{cf}}\left(H^{q_H} + \max_{0\le j\le J-1}(\Delta T_j)^{r_H} + h^2 + \Delta t_h\right). \tag{61}$$

The same bound holds with (j, M_j) in place of $(j, 0)$ for the segment endpoint. Consequently,

$$\|Y_j - Y_j^\star\|_{\mathcal{Y}_H} \le C_{\mathrm{cf}}\left(H^{q_H} + \max_{0\le j\le J-1}(\Delta T_j)^{r_H} + h^2 + \Delta t_h\right). \tag{62}$$

Proof. Apply the triangle inequality at $t = T_j$:

$$\begin{aligned}\|\mathbf{u}_H^j - R_{Hh}\mathbf{u}_{h,j}^0\|_{M_H} \le{}& \|\mathbf{u}_H^j - \mathcal{S}_H u(T_j)\|_{M_H} + \|\mathcal{S}_H u(T_j) - R_{Hh}\mathcal{S}_h u(T_j)\|_{M_H} \\ &+ \|R_{Hh}(\mathcal{S}_h u(T_j) - \mathbf{u}_{h,j}^0)\|_{M_H}.\end{aligned} \tag{63}$$

The three terms are controlled by (60), (59), the stability of R_{Hh}, and Theorem D.1. Since $Y_j = \mathbf{u}_H^j$ and $Y_j^\star = R_{Hh}\mathbf{u}_{h,j}^0$, this is (62). □

D.3 Stability of the fine-grid finite-difference Euler rollout

Lemma D.3 (Local Lipschitz stability of $\mathbf{\Phi}_h$) *Under Assumption 1, for all bounded grid vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{N_h}$,*

$$\|\mathbf{\Phi}_h^{\Delta t_h}(\mathbf{v}) - \mathbf{\Phi}_h^{\Delta t_h}(\mathbf{w})\|_{M_h} \le (1 + L_B\Delta t_h)\|\mathbf{v} - \mathbf{w}\|_{M_h}, \qquad L_B = 1 + 3B^2. \tag{64}$$

Proof. Using (12),

$$\mathbf{\Phi}_h^{\Delta t_h}(\mathbf{v}) - \mathbf{\Phi}_h^{\Delta t_h}(\mathbf{w}) = (I + \varepsilon^2\Delta t_h L_h)(\mathbf{v} - \mathbf{w}) + \Delta t_h\left[\boldsymbol{\psi}(\mathbf{v}) - \boldsymbol{\psi}(\mathbf{w})\right]. \tag{65}$$

Because L_h is symmetric negative semidefinite and (56) holds, $I + \varepsilon^2\Delta t_h L_h$ is nonexpansive in the M_h norm. Since ψ is Lipschitz on $[-B, B]$ with constant L_B, the result follows. □

D.4 CONDITIONAL DIAGNOSTICS FOR AN ALTERNATIVE NOISE-PREDICTION MODEL

The score-training statements in this subsection concern the alternative Gaussian perturbation in (47); they are not validated by the P1 clean-state training loss. The modular rollout theorem uses anchor risk directly and does not depend on these statements.

We refine the abstract learning term used in the anchor estimate for the residual-centered forward law (47). Fix a segment j and set

$$X_j := \mathbf{u}^0_{h,j} \in \mathbb{R}^{N_h}, \qquad Y_j = \mathbf{y}_j = \mathbf{u}^j_H \text{ as in (27)}, \qquad \mathbf{c}_j = P_{hH}\mathbf{y}_j, \tag{66}$$

with predictor $\boldsymbol{\epsilon}_{\theta_j}(\mathbf{x}_k, k, Y_j)$ as in Appendix C.1.

Set $\tau_k = k/K$ and define the score surrogate

$$\mathbf{s}_{\theta_j}(\mathbf{x}_k, k, Y_j) := -\frac{1}{\sqrt{1-\bar{a}_k}}\boldsymbol{\epsilon}_{\theta_j}(\mathbf{x}_k, k, Y_j). \tag{67}$$

Let $\mathbf{s}^\star_j(\mathbf{x}, \tau, Y_j)$ denote a true conditional score for the residual-centered forward marginal (existence as in Assumption 3). The population segment-wise score error is

$$\epsilon^2_{\text{score},j} := \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}\left[\left\|\mathbf{s}_{\theta_j}(\mathbf{x}_{j,k}, k, Y_j) - \mathbf{s}^\star_j(\mathbf{x}_{j,k}, \tau_k, Y_j)\right\|_2^2\right], \tag{68}$$

where $\mathbf{x}_{j,k}$ follows (47) with $\mathbf{x}^\star_j = X_j$, and the expectation includes Y_j, $\mathbf{x}^\star_j$, Gaussian noise, and k.

Remark 7 (Relation to the denoising loss) *For the residual-centered perturbation $\mathbf{x}_{j,k} = \mathbf{c}_j + \sqrt{\bar{a}_k}(X_j - \mathbf{c}_j) + \sqrt{1-\bar{a}_k}\,\boldsymbol{\xi}$, the conditional score of the Gaussian forward map satisfies*

$$\nabla_{\mathbf{x}_{j,k}} \log q(\mathbf{x}_{j,k} \mid X_j, Y_j) = -\frac{\boldsymbol{\xi}}{\sqrt{1-\bar{a}_k}}. \tag{69}$$

Thus the ϵ-prediction loss controls a denoising-score objective up to the same schedule factor as in DDPM:

$$\mathbb{E}\left[\left\|\mathbf{s}_{\theta_j}(\mathbf{x}_{j,k}, k, Y_j) + \frac{\boldsymbol{\xi}}{\sqrt{1-\bar{a}_k}}\right\|_2^2\right] = \frac{1}{1-\bar{a}_k}\mathbb{E}\left[\left\|\boldsymbol{\xi} - \boldsymbol{\epsilon}_{\theta_j}(\mathbf{x}_{j,k}, k, Y_j)\right\|_2^2\right]. \tag{70}$$

Consequently, if $c_a := \max_{1\le k\le K}(1-\bar{a}_k)^{-1}$, then

$$\mathcal{L}^{\text{pop}}_{\text{DSM},j}(\theta_j) \le c_a\, \mathcal{L}^{\text{pop}}_{\text{den},j}(\theta_j), \tag{71}$$

where $\mathcal{L}_{\text{den},j}$ is (51)*.*

Assumption 3 (Segment-wise score regularity and bounded support) *For each segment j, assume the following.*

1. *There are compact sets $\mathcal{K}_{h,j} \subset \mathbb{R}^{N_h}$ and $\mathcal{K}_{H,j} \subset \mathbb{R}^{N_H}$ such that the relevant forward-diffused high-resolution fields and coarse conditions lie in $\mathcal{K}_{h,j} \times \mathcal{K}_{H,j}$ with high probability. We work on*
$$D_j := \mathcal{K}_{h,j} \times \mathcal{K}_{H,j} \times [\tau_{\min}, 1], \qquad 0 < \tau_{\min} < 1. \tag{72}$$
2. *The joint density of $(\mathbf{x}_\tau, \mathbf{y}_j, \tau)$ on D_j is bounded by $\rho_{\max,j}$.*
3. *Each component of the true conditional score $\mathbf{s}^\star_j$ belongs to $H^m(D_j)$ for some $m \ge 3$ and satisfies*
$$\max_{1\le i\le N_h} \left|(s^\star_j)_i\right|_{H^m(D_j)} \le M_{m,j}. \tag{73}$$
4. *The reverse-time drift induced by either the true score or the learned score is globally Lipschitz on the considered compact set. Moreover, the support of the exact and learned reverse marginals at $\tau = 0$ is contained in an M_h-ball of radius R_j.*

Lemma D.4 (Symbolic score-error decomposition) *Under Assumption 3, decompose the trained score error* (68) *into approximation, optimization, and finite-sample components:*

$$\epsilon^2_{\mathrm{score},j} \le \epsilon^2_{\mathrm{app},j} + \epsilon^2_{\mathrm{opt},j} + \epsilon^2_{\mathrm{gen},j}. \tag{74}$$

We do not use pixel-dimension exponents; in high field dimension, $\epsilon_{\mathrm{app},j}$ is best interpreted under a low-dimensional data-manifold hypothesis (for example a rate in a manifold dimension $d_{\mathcal{M},j} \ll N_h$) or left as a qualitative small term. Equivalently, for ϵ-prediction training,

$$\epsilon^2_{\mathrm{score},j} \le C_a \left(\epsilon^2_{\mathrm{app},j} + \epsilon^2_{\mathrm{opt},j} + \epsilon^2_{\mathrm{gen},j}\right), \tag{75}$$

for a schedule-dependent constant C_a from (71)*.*

Proof. This is the standard empirical-risk decomposition for score matching combined with (70). □

We collect optimization, generalization, and network approximation errors into

$$\mathcal{E}^2_{\mathrm{train},j} := \epsilon^2_{\mathrm{app},j} + \epsilon^2_{\mathrm{opt},j} + \epsilon^2_{\mathrm{gen},j}, \tag{76}$$

where $\epsilon^2_{\mathrm{app},j}$ may be bounded explicitly as in Lemma D.4, but the main bounds below only use the aggregate $\mathcal{E}_{\mathrm{train},j}$.

Let $\mathcal{D}_{\theta_j,K}$ denote an abstract uncorrected comparator. For the deployed clean-state sampler it is defined by setting the correction step size a_k to zero in Algorithm 1, while retaining its noise schedule and network. The score-specific statements below require an alternative noise predictor and their additional assumptions. Neither comparator is identified with the reported raw-residual-only No-M2NO ablation.

Theorem D.5 (Base residual-centered diffusion error) *Let $\mathcal{G}^\star_j(Y) = \mathbb{E}[\mathbf{X}_j \mid Y^\star_j = Y]$ be the Bayes reconstructor (Definition 1). Define the base diffusion reconstruction error as RMS deviation from this conditional mean:*

$$\epsilon_{\mathrm{D},j}(K) := \left\|\mathcal{G}^\star_j(Y_j) - \mathcal{D}_{\theta_j,K}(Y_j)\right\|_{\mathbb{E},M_h}. \tag{77}$$

This quantity is logged directly on validation anchors (Section 4). For deterministic sampling ($\sigma_k = 0$), $\epsilon_{\mathrm{D},j}(K)$ is treated as an empirically measurable error rather than as the output of a path-space KL bound. If the sampler is stochastic, $\mathcal{D}_{\theta_j,K}(Y_j, \mathbf{Z})$ with noise $\mathbf{Z}$ conditionally independent of $\mathbf{X}_j$ given Y_j, then the expectation in (77) *also averages over $\mathbf{Z}$. Define the conditional sampler mean $\overline{\mathcal{D}}_{\theta_j,K}(Y_j) := \mathbb{E}_{\mathbf{Z}}[\mathcal{D}_{\theta_j,K}(Y_j, \mathbf{Z}) \mid Y_j]$ and the bias component*

$$\epsilon^{\mathrm{bias}}_{\mathrm{D},j}(K) := \left\|\mathcal{G}^\star_j(Y_j) - \overline{\mathcal{D}}_{\theta_j,K}(Y_j)\right\|_{\mathbb{E},M_h}. \tag{78}$$

For deterministic DDIM ($\sigma_k = 0$), $\overline{\mathcal{D}}_{\theta_j,K} = \mathcal{D}_{\theta_j,K}$ and $\epsilon^{\mathrm{bias}}_{\mathrm{D},j} = \epsilon_{\mathrm{D},j}$.

For optional stochastic analysis when $\sigma_k \ge \sigma_{\min} > 0$, standard decompositions bound $\epsilon_{\mathrm{D},j}(K)$ by score error and sampling discretization:

$$\mathbb{E}\left[\left\|\mathcal{D}_{\theta_j,K}(Y_j) - \mathcal{G}^\star_j(Y_j)\right\|^2_{M_h}\right] \le 2E_{\mathrm{sam},j} + 2E_{\mathrm{score},j}, \tag{79}$$

$$E_{\mathrm{sam},j} \le C_{\mathrm{sam},j} K^{-2\gamma}, \qquad \gamma \ge 1/2, \tag{80}$$

$$D_{\mathrm{KL}}\left(p^\theta_{0,j} \,\|\, p^\star_{0,j}\right) \le C_{\mathrm{KL},j}\epsilon^2_{\mathrm{score},j}, \tag{81}$$

$$E_{\mathrm{score},j} \le C_{\mathrm{TV},j}\epsilon_{\mathrm{score},j} + C'_{\mathrm{KL},j}\epsilon^2_{\mathrm{score},j}. \tag{82}$$

Proposition D.6 (Heuristic optional score-based bound under stochastic sampling) *Scope. This diagnostic requires the explicitly stated score-training and stochastic-kernel assumptions. Stochastic P1 sampling alone does not satisfy them; the P1 loss is relative clean-state error and this result is not claimed as a guarantee for the evaluated sampler. This proposition is a* diagnostic/heuristic *inequality chain (not a fully justified non-asymptotic theorem without additional transport/KL regularity). Under Assumption 3 and $\sigma_k \ge \sigma_{\min} > 0$, absorbing constants from* (79)–(82) *and Lemma D.4 yields*

$$\epsilon^2_{\mathrm{D},j}(K) \le C_{1,j}\mathcal{E}^2_{\mathrm{train},j} + C_{2,j}K^{-2\gamma} + C_{3,j}\mathcal{E}_{\mathrm{train},j}, \tag{83}$$

for constants $C_{i,j}$ depending on schedules and R_j.

Proof.[Proof sketch] Combine (79)–(82) with Lemma D.4. □

Remark 8 (When KL tools do not apply) *The estimates* (81)–(82) *require nondegenerate reverse noise. Deterministic DDIM ($\sigma_k = 0$) is outside this framework unless additional deterministic stability is assumed.*

Remark 9 (Local ϵ-prediction control of the residual) *From* (13)*,*

$$\widehat{\mathbf{s}}_{0,\theta_j} - \mathbf{s}_{j,0} = \frac{\sqrt{1-\bar{a}_k}}{\sqrt{\bar{a}_k}} \left(\boldsymbol{\xi} - \boldsymbol{\epsilon}_{\theta_j}(\mathbf{x}_{j,k}, k, Y_j)\right), \tag{84}$$

with $\mathbf{s}_{j,0} = X_j - \mathbf{c}_j$, hence

$$\left(\mathbb{E}\left[\left\|\widehat{\mathbf{s}}_{0,\theta_j} - \mathbf{s}_{j,0}\right\|_2^2\right]\right)^{1/2} \le \chi_{\max} \mathcal{L}_{\mathrm{den},j}^{\mathrm{pop}}(\theta_j)^{1/2}, \qquad \chi_{\max} := \max_{1\le k\le K} \left(\frac{1-\bar{a}_k}{\bar{a}_k}\right)^{1/2}. \tag{85}$$

D.5 ReMD correction relative to base diffusion

This is a conditional comparison with an abstract entirely uncorrected sampler. It is not a characterization of the No-M2NO switch, which retains the raw residual update. Unless the additive slack below is independently controlled, the inequality alone does not establish improvement.

The base diffusion error $\epsilon_{\mathrm{D},j}(K)$ in (77) measures how far the *uncorrected* residual-centered sampler $\mathcal{D}_{\theta_j,K}$ is from the Bayes reconstructor (conditional mean) on the given coarse data. The deployed sampler refines the clean prediction before the posterior update. The following abstract comparison imposes an additional sufficient inequality; it is not established by the reported ablation ratios.

Assumption 4 (ReMD correction improves the base sampler (conditional diagnostic)) *This inequality is treated as an* empirical or conditional *diagnostic relating corrected and uncorrected samplers, not as a consequence of a proved contraction principle. Let $\mathcal{D}_{\theta_j,K}$ be the unrolled residual-centered sampler (Algorithm 1) with $a_k \equiv 0$ (removing raw and learned corrections) and let $\mathcal{G}_{\vartheta_j,K}$ be the corrected sampler. There exist $\rho_{\mathrm{mg},j} \in [0,1)$ and $\epsilon_{\mathrm{mg},j} \ge 0$ such that*

$$\left\|\mathcal{G}_j^\star(Y_j) - \mathcal{G}_{\vartheta_j,K}(Y_j)\right\|_{\mathbb{E},M_h} \le \rho_{\mathrm{mg},j}\, \epsilon_{\mathrm{D},j}(K) + \epsilon_{\mathrm{mg},j}, \tag{86}$$

with $\epsilon_{\mathrm{D},j}(K)$ from (77)*.*

Remark 10 (Stepwise diagnostics) *Section 4 logs empirical ratios built from $\mathbf{T}_{\theta_j,k}(\mathbf{x}_k, Y_j)$ and $\mathbf{r}_{j,k}(\mathbf{x}_k)$ to check* (86) *at training resolution.*

D.6 ReMD anchor reconstruction error

We now combine local regularity of the Bayes reconstructor, coarse–fine consistency, the score/diffusion estimate, and the residual–multigrid estimate. Throughout, $\mathbf{X}_j$ denotes the random fine-grid anchor at time T_j induced by the data-generating distribution; along reference trajectories its realization is $\mathbf{u}_{h,j}^0$. Lemma D.7 records the standard orthogonal decomposition of squared Bayes risk into irreducible ambiguity and learnable excess. Definition 1 and (29) are stated in Section 3; equivalently,

$$\epsilon_{\mathrm{id},j}^2 = \mathbb{E}\left[\mathrm{tr}\left(M_h \,\mathrm{Cov}(\mathbf{X}_j \mid Y_j^\star)\right)\right]. \tag{87}$$

Remark 11 (Why conditional expectation is used) *The conditional mean formulation avoids assuming a deterministic inverse from LR to HR fields. Such an inverse generally does not exist because restriction removes high-frequency information. The Bayes reconstructor $\mathcal{G}_j^\star(Y) = \mathbb{E}[\mathbf{X}_j \mid Y_j^\star = Y]$ is instead the optimal target under the squared reconstruction loss. The ambiguity term $\epsilon_{\mathrm{id},j}$ represents irreducible conditional variance and is therefore part of the problem, not a failure of the algorithm. The local Lipschitz assumption is imposed only on the data support and can be checked empirically by nearest-neighbor slope diagnostics.*

Assumption 5 (Local regularity of the Bayes reconstructor) *Let $\mathcal{K}_j$ be a compact condition set containing the support of $Y_j^\star$ and the actual conditions used by the algorithm. We assume that the Bayes reconstructor is locally Lipschitz on $\mathcal{K}_j$:*

$$\left\|\mathcal{G}_j^\star(Y) - \mathcal{G}_j^\star(\widetilde{Y})\right\|_{M_h} \le L_{G,j} \left\|Y - \widetilde{Y}\right\|_{\mathcal{Y}_H}, \qquad Y, \widetilde{Y} \in \mathcal{K}_j. \tag{88}$$

This is a regularity assumption on the conditional mean over the data support, not a global inverse-stability assumption for arbitrary grid fields.

Lemma D.7 (Bayes risk decomposition) *Let $\mathcal{A}(Y_j^\star)$ be any deterministic reconstruction depending only on the ideal coarse restriction $Y_j^\star$. Then*

$$\mathbb{E}\left[\left\|\mathbf{X}_j - \mathcal{A}(Y_j^\star)\right\|_{M_h}^2\right] = \epsilon_{\mathrm{id},j}^2 + \mathbb{E}\left[\left\|\mathcal{G}_j^\star(Y_j^\star) - \mathcal{A}(Y_j^\star)\right\|_{M_h}^2\right]. \tag{89}$$

If the reconstruction is stochastic, $\widehat{\mathbf{X}}_j = \mathcal{A}(Y_j^\star, \mathbf{Z})$, with sampler noise $\mathbf{Z}$ conditionally independent of $\mathbf{X}_j$ given $Y_j^\star$, and

$$\overline{\mathcal{A}}(Y_j^\star) := \mathbb{E}_{\mathbf{Z}}[\mathcal{A}(Y_j^\star, \mathbf{Z}) \mid Y_j^\star],$$

then

$$\begin{aligned}\mathbb{E}\left[\left\|\mathbf{X}_j - \mathcal{A}(Y_j^\star, \mathbf{Z})\right\|_{M_h}^2\right] = \epsilon_{\mathrm{id},j}^2 &+ \mathbb{E}\left[\left\|\mathcal{G}_j^\star(Y_j^\star) - \overline{\mathcal{A}}(Y_j^\star)\right\|_{M_h}^2\right] \\ &+ \mathbb{E}\left[\left\|\mathcal{A}(Y_j^\star, \mathbf{Z}) - \overline{\mathcal{A}}(Y_j^\star)\right\|_{M_h}^2\right].\end{aligned} \tag{90}$$

Proof. The first identity follows from the orthogonality property of conditional expectation. Namely,

$$\mathbb{E}\left[\left(\mathbf{X}_j - \mathcal{G}_j^\star(Y_j^\star)\right)^\top M_h \left(\mathcal{G}_j^\star(Y_j^\star) - \mathcal{A}(Y_j^\star)\right)\right] = 0.$$

Expanding the squared norm gives (89). The stochastic version follows by conditioning on $Y_j^\star$ and applying the variance decomposition to the sampler output. □

Proof.[Proof of Theorem 3.4] Let $\|\cdot\|_{\mathbb{E},M_h} := \left(\mathbb{E}\|\cdot\|_{M_h}^2\right)^{1/2}$ denote the RMS norm. Inserting the Bayes reconstructor and applying the triangle inequality for this norm yields

$$\begin{aligned}\left\|\mathbf{X}_j - \mathcal{G}_{\vartheta_j,K}(Y_j)\right\|_{\mathbb{E},M_h} \le\ &\left\|\mathbf{X}_j - \mathcal{G}_j^\star(Y_j^\star)\right\|_{\mathbb{E},M_h} \\ &+ \left\|\mathcal{G}_j^\star(Y_j^\star) - \mathcal{G}_j^\star(Y_j)\right\|_{\mathbb{E},M_h} \\ &+ \left\|\mathcal{G}_j^\star(Y_j) - \mathcal{G}_{\vartheta_j,K}(Y_j)\right\|_{\mathbb{E},M_h}.\end{aligned} \tag{91}$$

The first term equals $\epsilon_{\mathrm{id},j}$ by (29). For the second term, pointwise Lipschitz continuity implies

$$\left\|\mathcal{G}_j^\star(Y_j^\star) - \mathcal{G}_j^\star(Y_j)\right\|_{M_h}^2 \le L_{G,j}^2 \left\|Y_j^\star - Y_j\right\|_{\mathcal{Y}_H}^2,$$

hence after taking expectations and square roots, $\|\mathcal{G}_j^\star(Y_j^\star) - \mathcal{G}_j^\star(Y_j)\|_{\mathbb{E},M_h} \le L_{G,j}\epsilon_{\mathrm{cond},j}$. The third term is bounded by (86). This establishes (36) (hence (37)) up to the $C_{\mathrm{dyn},j}\eta_{\mathrm{dyn},j}$ slack. When increment-based residuals interact with a non-compatible coarse integrator, Remark 13 supplies the advertised contribution with coefficient $C_{\mathrm{dyn},j}$; otherwise $\eta_{\mathrm{dyn},j} = 0$. □

Remark 12 (Discrete coarse–fine scaling) *If the coarse–fine mismatch satisfies* (62)*, then*

$$\epsilon_{\mathrm{cond},j} \le C_{\mathrm{cf}}\left(H^{q_H} + \max_{0\le i\le J-1}(\Delta T_i)^{r_H} + h^2 + \Delta t_h\right). \tag{92}$$

Substituting into (37) *recovers the fully explicit display*

$$\begin{aligned}\delta_j \le\ &\epsilon_{\mathrm{id},j} + L_{G,j}C_{\mathrm{cf}}\left(H^{q_H} + \max_{0\le i\le J-1}(\Delta T_i)^{r_H} + h^2 + \Delta t_h\right) \\ &+ C_{\mathrm{dyn},j}\,\eta_{\mathrm{dyn},j} + \rho_{\mathrm{mg},j}\,\epsilon_{\mathrm{D},j}(K) + \epsilon_{\mathrm{mg},j},\end{aligned} \tag{93}$$

with $\epsilon_{\mathrm{D},j}(K)$ from (77)*. In the observed-LR protocol, if $Y_j = Y_j^\star$ and R_{Hh} matches the data downsampling operator, then $\epsilon_{\mathrm{cond},j} = 0$.*

Lemma D.8 (Increment versus vector-field flux targets) *Fix k, j and $\mathbf{x} \in \mathbb{R}^{N_h}$. Let $\mathbf{r}_{j,k}^{\mathrm{inc}}(\mathbf{x})$ and $\mathbf{r}_{j,k}^{\mathrm{vf}}(\mathbf{x})$ denote* (17) *with the same $(\omega_{\mathrm{data},k}, \omega_{\mathrm{AC},k})$ but with $\mathbf{d}_H^j = (\mathbf{y}_j^+ - \mathbf{y}_j)/\Delta T_j$ versus $\mathbf{d}_H^j = \mathbf{F}_H(\mathbf{y}_j)$, respectively. Then*

$$\left\|\mathbf{r}_{j,k}^{\mathrm{inc}}(\mathbf{x}) - \mathbf{r}_{j,k}^{\mathrm{vf}}(\mathbf{x})\right\|_{M_h} \le \omega_{\mathrm{AC},k}\Delta T_j\, C_P\, \eta_{\mathrm{dyn},j}, \tag{94}$$

where C_P is the $M_H \to M_h$ stability constant for P_{hH} in (58)*.*

Proof. Only the flux blocks differ; subtract and apply (58) to $\mathbf{d}^j_{H,\mathrm{inc}} - \mathbf{d}^j_{H,\mathrm{vf}} = (\mathbf{y}^+_j - \mathbf{y}_j)/\Delta T_j - \mathbf{F}_H(\mathbf{y}_j)$. □

Remark 13 (Scaling of $C_{\mathrm{dyn},j}$) *Write the physics block of* (17) *schematically as* $\omega_{\mathrm{AC},k}\Delta T_j\, P_{hH}(\mathbf{d}^j_H - R_{Hh}\mathbf{F}_h(\mathbf{x}))$. *If one trains or analyzes with* $\mathbf{d}^j_H = \mathbf{F}_H(\mathbf{y}_j)$ *but evaluates with the increment proxy* (19)*, or vice versa, the induced change in the coarse-grid target is bounded by* $\eta_{\mathrm{dyn},j}$ *in* $\|\cdot\|_{M_H}$. *Passing through* P_{hH} *contributes an operator factor comparable to the stability constant* C_P *in* (58)*; composing with the reverse-chain Lipschitz/stability assumptions underlying Assumption 4 and with* $L_{G,j}$ *from* (88) *when comparing anchors produced under perturbed residuals yields a contribution of the form* $C_{\mathrm{dyn},j}\eta_{\mathrm{dyn},j}$ *with* $C_{\mathrm{dyn},j}$ *depending on* C_P, $\max_k(\omega_{\mathrm{AC},k}\Delta T_j)$*, those sampler moduli, and* $L_{G,j}$. *When* (18) *is enforced throughout,* $\eta_{\mathrm{dyn},j} = 0$ *and this slack disappears.*

Remark 14 (Residual penalties and inverse stability) *The sampler-level penalties in* (53) *can be connected to the anchor error through the observation map*

$$\mathcal{O}_j(\mathbf{x}) := (R_{Hh}\mathbf{x}, R_{Hh}\mathbf{F}_h(\mathbf{x}))\,, \tag{95}$$

If $\mathcal{O}_j$ *is inverse-stable on the learned solution manifold,*

$$\|\mathbf{v} - \mathbf{w}\|_{M_h} \le C_{\mathrm{obs},j}\,\|\mathcal{O}_j(\mathbf{v}) - \mathcal{O}_j(\mathbf{w})\|_{\mathcal{Y}_H} + \epsilon_{\mathrm{id},j}, \tag{96}$$

then the rollout penalties in (53) *admit analogous validation bounds in terms of* $\mathcal{L}_{\mathrm{roll},j}(\vartheta_j)^{1/2}$ *(up to generalization error). This provides a residual-based interpretation of* $\epsilon_{\mathrm{mg},j}$ *and the irreducible ambiguity scale* $\epsilon_{\mathrm{id},j}$. *Here* $\|\cdot\|_{\mathcal{Y}_H}$ *denotes the Euclidean combination norm on the two coarse snapshots* $(R_{Hh}\mathbf{v}, R_{Hh}\mathbf{F}_h(\mathbf{v}))$*, unrelated to the single-vector coarse condition* $Y_j = \mathbf{u}^j_H$ *in* (27)*. We do not require a global strong-monotonicity estimate for the nonlinear Allen–Cahn residual; the weaker manifold inverse-stability condition is enough for the proposed super-resolution setting.*

D.7 Propagation inside each segment

Theorem D.9 (Segmentwise error propagation) *For any fixed finite* $d \ge 1$*, fix* j *and let* $0 \le m \le M_j$. *Under Assumption 1,*

$$\left\|\mathbf{u}^m_{h,j} - \widehat{\mathbf{u}}^m_{h,j}\right\|_{\mathbb{E},M_h} \le \exp\left(L_B\, m\Delta t_h\right)\delta_j. \tag{97}$$

Consequently,

$$\max_{0\le m\le M_j}\left\|\mathbf{u}^m_{h,j} - \widehat{\mathbf{u}}^m_{h,j}\right\|_{\mathbb{E},M_h} \le \exp(L_B\Delta T_j)\delta_j. \tag{98}$$

Proof. For $0 \le m \le M_j - 1$, both the fine reference solution and the segmented approximation are advanced by the same fine-grid Euler map:

$$\mathbf{u}^{m+1}_{h,j} - \widehat{\mathbf{u}}^{m+1}_{h,j} = \mathbf{\Phi}^{\Delta t_h}_h(\mathbf{u}^m_{h,j}) - \mathbf{\Phi}^{\Delta t_h}_h(\widehat{\mathbf{u}}^m_{h,j}). \tag{99}$$

Using Lemma D.3,

$$\left\|\mathbf{u}^{m+1}_{h,j} - \widehat{\mathbf{u}}^{m+1}_{h,j}\right\|_{\mathbb{E},M_h} \le (1 + L_B\Delta t_h)\left\|\mathbf{u}^m_{h,j} - \widehat{\mathbf{u}}^m_{h,j}\right\|_{\mathbb{E},M_h}. \tag{100}$$

Iteration gives $(1 + L_B\Delta t_h)^m \le \exp(L_B\, m\Delta t_h) \le \exp(L_B\Delta T_j)$ for $m \le M_j$, which proves the claim. □

D.8 Proof of Theorem 3.6

Proof. For fixed (j, m), decompose

$$\|\mathcal{S}_h u(t_{j,m}) - \widehat{\mathbf{u}}^m_{h,j}\|_{M_h} \le \|\mathcal{S}_h u(t_{j,m}) - \mathbf{u}^m_{h,j}\|_{M_h} + \|\mathbf{u}^m_{h,j} - \widehat{\mathbf{u}}^m_{h,j}\|_{M_h}. \tag{101}$$

The first term is bounded by Theorem D.1. The second term is bounded by Theorem D.9 in terms of the anchor RMS δ_j in (33). Taking expectations and the maximum over j, m yields (39). □

E NUMERICAL DIAGNOSTICS AND EVALUATION PROTOCOLS

Scope of the diagnostic protocols. This section defines candidate diagnostics for the analysis. It does not assert that every quantity has been computed or that empirical diagnostics verify uniform theoretical assumptions. Reported measurements and their aggregation conventions are specified in Appendix G; P1 interface jumps are reported in Section G.1. Noise-prediction and score diagnostics below concern the alternative model in Appendix C.1, not the P1 training loss.

E.1 CODE PATHS, TENSORS, AND BASELINES

We study the two-dimensional Allen–Cahn problem (2) on $\Omega = (0,1)^2$ with Neumann data (5). Coarse and fine Cartesian grids with spacings H and h carry segmentwise states $\{\mathbf{u}^j_{H,s}\}$, $\{\mathbf{u}^m_{h,s,j}\}$, and hybrid outputs $\{\widehat{\mathbf{u}}^m_{h,s,j}\}$ (Section 2.2). Operators R_{Hh}, P_{hH}, $\mathbf{\Phi}_h^{\Delta t_h}$, L_h, and $M_h = h^d I$ follow Section 2. Segment conditions are $Y_{s,j} = (\mathbf{u}^j_{H,s}, \mathbf{u}^{j+1}_{H,s})$; anchors are $\mathcal{G}_{\vartheta_j,K}(Y_{s,j})$ from Algorithm 1. Training uses (54); inference follows Algorithms 1–2. Report ΔT_j, Δt_h, and $M_j = \Delta T_j/\Delta t_h$ for each segment; verify $M_j \in \mathbb{N}$. **Baselines to ship in the same codebase:** (i) full fine FD reference (12); (ii) $P_{hH}\mathbf{u}^j_H$ or coarse-only upsampling; (iii) ablations removing $\mathcal{M}_{\varphi_j,k}$, removing the Allen–Cahn residual in (17), or conditioning on a single coarse frame instead of Y_j. Coarse data may come from (10) or from FEM/FD sampled onto the grid (Remark 3).

E.2 VALIDATION CHECKLIST TIED TO SECTION 3

For each item, log batch aggregates over validation scenarios s and keep per-step tensors for debugging (saved as `.npz`/`.pt` as appropriate). Scenario aggregation is not a substitute for uncertainty across independently trained random seeds; the latter should be reported separately whenever multiple training runs are available.

CFL and boundedness (Assumption 1). Implement power-iteration or dense eigen-solve for $\lambda_{\max}(-L_h)$ with the same boundary stencil as production:

$$\theta_h := \varepsilon^2 \Delta t_h \, \lambda_{\max}(-L_h). \tag{102}$$

Pass criterion: $\theta_h \le 2 - c_0$ with fixed $c_0 \in (0,2)$. **Track** ℓ^∞ monitors on references and hybrids:

$$B_{\mathrm{val}} := \max_{s,j,m} \left\{ \|\mathbf{u}^m_{h,s,j}\|_{\ell^\infty}, \|\mathbf{u}^j_{H,s}\|_{\ell^\infty}, \|\widehat{\mathbf{u}}^m_{h,s,j}\|_{\ell^\infty} \right\}. \tag{103}$$

Optional safeguards: clipping heads or bounded activations on anchors if B_{val} drifts outside the training range.

Transfer stability / sampling compatibility (Assumption 2). The discrete operators used in code must satisfy stability budgets analogous to

$$\|R_{Hh}\mathbf{v}_h\|_{M_H} \le C_R \|\mathbf{v}_h\|_{M_h}, \qquad \|P_{hH}\mathbf{v}_H\|_{M_h} \le C_P \|\mathbf{v}_H\|_{M_H}. \tag{104}$$

Compute the empirical scaled mismatch on validation references:

$$\eta_{\mathrm{comp}}(h,H) := \max_{s,n} \frac{\|\mathcal{S}_H u_s(t_n) - R_{Hh}\mathcal{S}_h u_s(t_n)\|_{M_H}}{H^{q_H} + h^2}. \tag{105}$$

Expectation: η_{comp} stays uniformly bounded as grids refine, matching Lemma D.2.

Score training / diffusion error (Assumption 3). Population scores need not be available; instead log the train-vs-validation split of the denoising objective. The score-energy bookkeeping used in theory is

$$\epsilon^2_{\mathrm{score},j} \lesssim \epsilon^2_{\mathrm{app},j} + \epsilon^2_{\mathrm{opt},j} + \epsilon^2_{\mathrm{gen},j}. \tag{106}$$

Implement the validation denoising loss

$$\mathcal{L}^{\mathrm{val}}_{\mathrm{den},j} := \frac{1}{N_{\mathrm{val}}} \sum_{s,k} \left\| \boldsymbol{\xi}_{s,k} - \boldsymbol{\epsilon}_{\theta_j}(\mathbf{x}_{s,j,k}, k, Y_{s,j}) \right\|_2^2, \tag{107}$$

and the anchor RMS error (same quantity as E_{anc} later):

$$E^{\mathrm{val}}_{\mathrm{anchor},j}(K) := \left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| \mathbf{u}^0_{h,s,j} - \mathcal{G}_{\vartheta_j,K}(Y_{s,j}) \right\|^2_{M_h} \right)^{1/2}. \tag{108}$$

Also implement base versus corrected anchors (cf. (77)**):**

$$E^{\mathrm{obs}}_{\mathrm{D}}(j;K) := \left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| \mathbf{u}^0_{h,s,j} - \mathcal{D}_{\theta_j,K}(Y_{s,j}) \right\|^2_{M_h} \right)^{1/2}, \tag{109}$$

$$E^{\mathrm{obs}}_{\mathrm{G}}(j;K) := \left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| \mathbf{u}^0_{h,s,j} - \mathcal{G}_{\vartheta_j,K}(Y_{s,j}) \right\|^2_{M_h} \right)^{1/2} = E^{\mathrm{val}}_{\mathrm{anchor},j}(K). \tag{110}$$

Expectation: $E^{\mathrm{obs}}_{\mathrm{G}}(j;K) < E^{\mathrm{obs}}_{\mathrm{D}}(j;K)$ when multigrid correction helps. Lowering $\mathcal{L}^{\mathrm{val}}_{\mathrm{den},j}$ correlates with lower $E^{\mathrm{obs}}_{\mathrm{D}}$, $E^{\mathrm{obs}}_{\mathrm{G}}$ as K grows.

Deterministic DDIM / few-step sampling. If $\sigma_k = 0$, treat $\epsilon_{\mathrm{D},j}(K)$ as the empirical base diffusion error. **Sweep** $K \in \{2, 5, 10, 20, \ldots\}$ and fit

$$\left(E^{\mathrm{obs}}_{\mathrm{D}}(j;K)\right)^2 \approx a_j + b_j K^{-2\gamma_j}, \tag{111}$$

as an empirical counterpart to Proposition D.6.

Restriction-only coarse conditioning vs. intrinsic ambiguity (Definition 1). Record the production condition (network input):

$$Y_{s,j} = \mathbf{u}^j_{H,s}. \tag{112}$$

Implement the nearest-neighbor ambiguity probe

$$s^\star(s) := \arg\min_{r \neq s} \|Y^\star_{s,j} - Y^\star_{r,j}\|_{\mathcal{Y}_H}, \tag{113}$$

$$A^{\mathrm{val}}_{\mathrm{id},j} := \left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| \mathbf{u}^0_{h,s,j} - \mathbf{u}^0_{h,s^\star(s),j} \right\|^2_{M_h} \right)^{1/2}. \tag{114}$$

Large $A^{\mathrm{val}}_{\mathrm{id},j}$ flags ill-posed coarse-to-fine pairs even on the dataset.

Intrinsic ambiguity (kNN conditional mean). For each validation condition $Y_{s,j}$, let $\mathcal{N}_q(s)$ be its q nearest neighbors in the condition norm $\|\cdot\|_{\mathcal{Y}_H}$, excluding s. With HR anchors $\mathbf{u}^0_{h,s,j}$ playing the role of realizations of $\mathbf{X}_j$, approximate the conditional mean by

$$\widehat{\mu}_{j,q}(Y_{s,j}) := \frac{1}{q} \sum_{r \in \mathcal{N}_q(s)} \mathbf{u}^0_{h,r,j}. \tag{115}$$

The empirical ambiguity proxy is

$$\widehat{\epsilon}^2_{\mathrm{id},j}(q) := \frac{1}{N_{\mathrm{val}}} \sum_{s=1}^{N_{\mathrm{val}}} \left\| \mathbf{u}^0_{h,s,j} - \widehat{\mu}_{j,q}(Y_{s,j}) \right\|^2_{M_h}. \tag{116}$$

A small value means that nearby LR conditions determine similar HR anchors; a large value indicates intrinsic coarse-to-fine ambiguity.

Local Lipschitz diagnostic. Using the same nearest-neighbor graph, estimate local slopes by

$$S^{(j)}_{s,r} := \frac{\|\widehat{\mu}_{j,q}(Y_{s,j}) - \widehat{\mu}_{j,q}(Y_{r,j})\|_{M_h}}{\|Y_{s,j} - Y_{r,j}\|_{\mathcal{Y}_H} + \delta_{\mathrm{num}}}, \qquad r \in \mathcal{N}_q(s), \tag{117}$$

where $\delta_{\mathrm{num}} > 0$ is a small numerical floor. Report

$$\widehat{L}^{(p)}_{G,j} := \mathrm{Quantile}_p\left(\{S^{(j)}_{s,r}\}_{s,r} \right), \qquad p \in \{0.90, 0.95\}. \tag{118}$$

A bounded high-quantile slope supports Assumption 5 on the observed data support.

Excess risk after subtracting ambiguity. Let $E_{\mathrm{D}}^{\mathrm{obs}}(j;K)$ and $E_{\mathrm{G}}^{\mathrm{obs}}(j;K)$ be as in (109)–(110). Define ambiguity-corrected excess errors by

$$E_{\mathrm{ex},D}^2(j;K) := \left[\left(E_{\mathrm{D}}^{\mathrm{obs}}(j;K)\right)^2 - \widehat{\epsilon}_{\mathrm{id},j}^2(q)\right]_+, \tag{119}$$

$$E_{\mathrm{ex},G}^2(j;K) := \left[\left(E_{\mathrm{G}}^{\mathrm{obs}}(j;K)\right)^2 - \widehat{\epsilon}_{\mathrm{id},j}^2(q)\right]_+, \tag{120}$$

with $[a]_+ := \max\{a, 0\}$. The ambiguity-corrected improvement ratio is

$$\widehat{\rho}_{\mathrm{ex}}(j;K) := \frac{E_{\mathrm{ex},G}(j;K)}{E_{\mathrm{ex},D}(j;K) + \delta_{\mathrm{num}}}. \tag{121}$$

The expected behavior is $\widehat{\rho}_{\mathrm{ex}}(j;K) < 1$, indicating that ReMD reduces the learnable part of the anchor error beyond the irreducible ambiguity.

Multigrid improvement ratio (Assumption 4). Using (109)–(110), define the empirical mixture ratio

$$\widehat{\rho}_{\mathrm{mix}}(j;K) := \frac{E_{\mathrm{G}}^{\mathrm{obs}}(j;K)}{E_{\mathrm{D}}^{\mathrm{obs}}(j;K) + \mathrm{tol}}. \tag{122}$$

(Numerator: ReMD-corrected anchor RMS $E_{\mathrm{G}}^{\mathrm{obs}}$; denominator: base diffusion anchor RMS $E_{\mathrm{D}}^{\mathrm{obs}}$ plus tolerance.) We use the same quantity under the dashboard alias $\widehat{\rho}_{\mathrm{mg}}(j;K) := \widehat{\rho}_{\mathrm{mix}}(j;K)$. It should *not* be identified with the theoretical contraction factor $\rho_{\mathrm{mg},j}$ in (86), because (122) also mixes intrinsic ambiguity and coarse-mismatch effects; see Section E.3. **Expectation:** $\widehat{\rho}_{\mathrm{mix}}(j;K) < 1$ when ReMD improves over the base residual-centered sampler; MG-off runs hit $E_{\mathrm{G}}^{\mathrm{obs}} = E_{\mathrm{D}}^{\mathrm{obs}}$. Optionally log stepwise diagnostics comparing $\mathbf{T}_{\theta_j,k}$ with and without $(1 - \eta_{k-1}/\eta_k)a_k\mathcal{M}_{b,\varphi,k}(\mathbf{v}_k, \mathbf{r}_{j,k}(\mathbf{v}_k))$.

Segmentwise rollout amplification (Theorem 3.6). Monitor intra-segment RMS drift

$$E_{\mathrm{rollout}}(j) := \max_{0\le m\le M_j} \left(\frac{1}{N_{\mathrm{val}}}\sum_s \left\|\mathbf{u}_{h,s,j}^m - \widehat{\mathbf{u}}_{h,s,j}^m\right\|_{M_h}^2\right)^{1/2}. \tag{123}$$

Expectation: shrinking $\tau_{\max}$ lowers amplification consistent with $\exp(L_B\tau_{\max})$ (Remark 2).

The following subsections spell out convergence plots, aggregated error metrics, residual energies, spectra, and cost accounting. Implement them using the tensors and hooks above.

E.3 Observed and oracle diagnostics for ReMD improvement

The Bayes reconstructor $\mathcal{G}_j^\star(Y) = \mathbb{E}[\mathbf{X}_j \mid Y_j^\star = Y]$ is not available in finite validation sets. Therefore, theoretical quantities involving $\mathcal{G}_j^\star(Y_j)$ cannot be measured directly. Candidate observable proxies include (114) and the kNN ambiguity estimator above; their definitions alone do not constitute measured results. Let $\mathcal{D}_{\theta_j,K}$ be the base residual-centered diffusion sampler without multigrid correction, and let $\mathcal{G}_{\vartheta_j,K}$ be the full ReMD sampler. The observed base and corrected anchor errors are

$$E_{\mathrm{D}}^{\mathrm{obs}}(j;K) := \left(\frac{1}{N_{\mathrm{val}}}\sum_s \left\|\mathbf{u}_{h,s,j}^0 - \mathcal{D}_{\theta_j,K}(Y_{s,j})\right\|_{M_h}^2\right)^{1/2}, \tag{124}$$

$$E_{\mathrm{G}}^{\mathrm{obs}}(j;K) := \left(\frac{1}{N_{\mathrm{val}}}\sum_s \left\|\mathbf{u}_{h,s,j}^0 - \mathcal{G}_{\vartheta_j,K}(Y_{s,j})\right\|_{M_h}^2\right)^{1/2}, \tag{125}$$

consistent with (109)–(110). Their ratio (122) is an empirical improvement ratio and should not be interpreted as the exact theoretical contraction factor $\rho_{\mathrm{mg},j}$ because it also contains intrinsic ambiguity and coarse-mismatch effects.

To reduce the coarse-mismatch component, evaluate an oracle-condition diagnostic. Define

$$Y_{s,j}^\star := \left(R_{Hh}\mathbf{u}_{h,s,j}^0, R_{Hh}\mathbf{u}_{h,s,j}^{M_j}\right). \tag{126}$$

Then

$$E_{\mathrm{D}}^{\star}(j;K) := \left(\frac{1}{N_{\mathrm{val}}} \sum_{s} \left\| \mathbf{u}_{h,s,j}^{0} - \mathcal{D}_{\theta_j,K}(Y_{s,j}^{\star}) \right\|_{M_h}^{2} \right)^{1/2}, \tag{127}$$

$$E_{\mathrm{G}}^{\star}(j;K) := \left(\frac{1}{N_{\mathrm{val}}} \sum_{s} \left\| \mathbf{u}_{h,s,j}^{0} - \mathcal{G}_{\vartheta_j,K}(Y_{s,j}^{\star}) \right\|_{M_h}^{2} \right)^{1/2}. \tag{128}$$

The ratio $E_{\mathrm{G}}^{\star}(j;K)/E_{\mathrm{D}}^{\star}(j;K)$ measures the benefit of multigrid correction with computed coarse-trajectory error largely removed.

Finally, report the conservative upper proxy

$$\widehat{\rho}_{\mathrm{mg}}^{\mathrm{ub}}(j;K) = \frac{E_{\mathrm{G}}^{\mathrm{obs}}(j;K) + A_{\mathrm{id},j}^{\mathrm{val}} + L_{G,j} E_{\mathrm{cf}}^{\mathrm{val}}(j)}{\left[E_{\mathrm{D}}^{\mathrm{obs}}(j;K) - A_{\mathrm{id},j}^{\mathrm{val}} - L_{G,j} E_{\mathrm{cf}}^{\mathrm{val}}(j) \right]_{+}}, \tag{129}$$

where $[a]_+ := \max(a, 0)$ and $E_{\mathrm{cf}}^{\mathrm{val}}(j)$ denotes the validation coarse–fine mismatch $E_{\mathrm{cf}}(H,h)$ from Section 4 restricted to segment j (for example the maximum over coarse times T_j, T_{j+1}). The expected empirical pattern is

$$E_{\mathrm{G}}^{\mathrm{obs}}(j;K) < E_{\mathrm{D}}^{\mathrm{obs}}(j;K), \qquad E_{\mathrm{G}}^{\star}(j;K) < E_{\mathrm{D}}^{\star}(j;K). \tag{130}$$

When $A_{\mathrm{id},j}^{\mathrm{val}}$ and $E_{\mathrm{cf}}^{\mathrm{val}}(j)$ are small, (122) is a more faithful proxy for the theoretical contraction behavior.

E.4 Suggested verification protocol

The following checklist specifies possible tests of the analysis. It does not assert that every diagnostic has been evaluated. Statements labeled "Expected behavior" below describe hypotheses to test; measured results and their evaluation scope are reported in Appendix G.

- *Time indexing:* report ΔT_j, Δt_h, and $M_j = \Delta T_j / \Delta t_h$; verify $M_j \in \mathbb{N}$.
- *Multigrid versus base diffusion:* report $E_{\mathrm{D}}^{\mathrm{obs}}$, $E_{\mathrm{G}}^{\mathrm{obs}}$, and $\widehat{\rho}_{\mathrm{mix}}$.
- *Training ablations:* compare denoising-only, denoising+clean residual, denoising+sampler loss (53), and full ReMD with multigrid and Allen–Cahn residual.
- *Coarse dynamic defect:* if increment residuals are used with non-explicit coarse integrators, report $\eta_{\mathrm{dyn},j}$ from (20).
- *Intrinsic ambiguity / coarse mismatch:* report $A_{\mathrm{id},j}^{\mathrm{val}}$, $\widehat{\epsilon}_{\mathrm{id},j}(q)$, $\widehat{L}_{G,j}^{(p)}$, $\widehat{\rho}_{\mathrm{ex}}$, $E_{\mathrm{cf}}^{\mathrm{val}}(j)$, $E_{\mathrm{D}}^{\star}$, and $E_{\mathrm{G}}^{\star}$.

E.5 Primary error metrics versus Section 3

Let $\mathbf{u}_{h,s}^{n}$ denote the fine-grid reference on validation trajectory s and $\widehat{\mathbf{u}}_{h,s}^{n}$ the hybrid output (same indexing as Theorem 3.6). Define snapshot errors in the mass norm,

$$E_{\mathrm{snap}}(s,n) := \left\| \mathbf{u}_{h,s}^{n} - \widehat{\mathbf{u}}_{h,s}^{n} \right\|_{M_h}, \qquad E_{\mathrm{snap}}(n) := \left(\frac{1}{N_{\mathrm{val}}} \sum_{s} E_{\mathrm{snap}}(s,n)^2 \right)^{1/2}, \tag{131}$$

together with a relative counterpart (fixed scaling of the reference norm),

$$E_{\mathrm{rel}}(n) := \frac{E_{\mathrm{snap}}(n)}{\left(\frac{1}{N_{\mathrm{val}}} \sum_{s} \|\mathbf{u}_{h,s}^{n}\|_{M_h}^2 \right)^{1/2} + \mathrm{tol}}. \tag{132}$$

Global summaries are

$$E_{\max} := \max_{0 \le n \le N_t} E_{\mathrm{snap}}(n), \qquad E_{\mathrm{avg}} := \left(\frac{1}{N_t + 1} \sum_{n=0}^{N_t} E_{\mathrm{snap}}(n)^2 \right)^{1/2}. \tag{133}$$

Expected behavior: $E_{\max}$ and E_{avg} track the total error bound in Theorem 3.6; improvements from better training or smaller $\tau_{\max}$ should lower both.

E.6 Finite-difference discretization error

Isolate the FD–Euler contribution by comparing $\mathbf{u}_h^n$ to a highly resolved reference (mesh refinement or manufactured solution). Define

$$E_{\mathrm{FD}}(h, \Delta t) := \max_{0 \le n \le N_t} \left\| \mathbf{u}_h^n - \mathbf{u}_{h,\mathrm{ref}}^n \right\|_{M_h}, \tag{134}$$

and empirical convergence rates

$$r_h := \frac{\log\left(E_{\mathrm{FD}}(h_1, \Delta t)/E_{\mathrm{FD}}(h_2, \Delta t)\right)}{\log(h_1/h_2)}, \qquad r_t := \frac{\log\left(E_{\mathrm{FD}}(h, \Delta t_1)/E_{\mathrm{FD}}(h, \Delta t_2)\right)}{\log(\Delta t_1/\Delta t_2)}. \tag{135}$$

Expected behavior: for the second-order Laplacian stencil and Euler in time, $r_h \approx 2$ and $r_t \approx 1$ in the asymptotic regime, consistent with Theorem D.1.

E.7 CFL and boundedness

At each training/inference checkpoint, re-log θ_h from (102) and B_{val} from (103); these scalars should appear on dashboards next to the losses. *Expected behavior:* $\theta_h \le 2 - c_0$ (Section 3); solutions remain bounded and B_{val} stays aligned with the Allen–Cahn range used in analysis.

E.8 Coarse–fine consistency

Measure coarse restriction mismatch between reference trajectories:

$$E_{\mathrm{cf}}(H, h) := \max_{0 \le n \le N_t} \left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| \mathbf{u}_{H,s}^n - R_{Hh} \mathbf{u}_{h,s}^n \right\|_{M_H}^2 \right)^{1/2}. \tag{136}$$

Optionally compare with the scaled diagnostic $\eta_{\mathrm{comp}}(h, H)$ from (105). *Expected behavior:* $E_{\mathrm{cf}}(H, h) = O(H^{q_H} + h^2 + \Delta t)$ up to constants, matching Lemma D.2.

E.9 ReMD data and Allen–Cahn residuals

At anchors, evaluate population-level residual energies on validation fields:

$$E_{\mathrm{data}}(j) := \left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| P_{hH} \left(\mathbf{u}_{H,s}^j - R_{Hh} \widehat{\mathbf{x}}_{0,s,j} \right) \right\|_{M_h}^2 \right)^{1/2}, \tag{137}$$

$$E_{\mathrm{AC}}(j) := \left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| \mathbf{r}_{\mathrm{AC}} \left(\widehat{\mathbf{x}}_{0,s,j}; \mathbf{u}_{H,s}^{j+1} \right) \right\|_{M_h}^2 \right)^{1/2}, \tag{138}$$

where $\widehat{\mathbf{x}}_{0,s,j}$ is the predicted clean field at the final reverse step (or the chain output when using residual–drift sampling), and the time part of (17) coincides with the increment form (19) when $\mathbf{d}_H^j$ is chosen accordingly. *Expected behavior:* the full model with both residuals active attains lower $E_{\mathrm{data}}, E_{\mathrm{AC}}$ than ablations that remove either the data residual or the Allen–Cahn residual, echoing ReMD's emphasis on physics- aware correction.

E.10 Diffusion training and sampling errors

For the alternative noise-prediction formulation only, consider validation denoising loss $\mathcal{L}_{\mathrm{den},j}^{\mathrm{val}}$ from (107). Note $E_{\mathrm{anc}}(j; K)$ below coincides with $E_{\mathrm{anchor},j}^{\mathrm{val}}(K)$ in (108). Define an empirical score error proxy (schedule-weighted, optional) by

$$\widehat{\epsilon}_{\mathrm{score},j} := \left(\frac{1}{N_{\mathrm{val}} K} \sum_{s,k} \left\| \mathbf{s}_{\theta_j}(\mathbf{x}_{s,j,k}, k, \mathbf{y}_{s,j}) - \mathbf{s}_j^\star(\mathbf{x}_{s,j,k}, \tau_k, \mathbf{y}_{s,j}) \right\|_2^2 \right)^{1/2}, \tag{139}$$

when a score surrogate is available, or omit if only ϵ-prediction is trained. Clean-field error at a fixed reverse step and anchor reconstruction error are

$$E_{\mathrm{clean}}(j) := \left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| \mathbf{u}^0_{h,s,j} - \widehat{\mathbf{x}}_{0,s,j} \right\|^2_{M_h} \right)^{1/2}, \tag{140}$$

$$E_{\mathrm{anc}}(j;K) := \left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| \mathbf{u}^0_{h,s,j} - \mathcal{G}_{\vartheta_j,K}(Y_{s,j}) \right\|^2_{M_h} \right)^{1/2}. \tag{141}$$

Expected behavior: $E_{\mathrm{anc}}(j;K)^2$ stabilizes as K increases; fitting $E_{\mathrm{anc}}(j;K)^2 \approx a_j + b_j K^{-2\gamma_j}$ provides a descriptive diagnostic of sampling-step sensitivity. Such a fit does not verify Proposition D.6 or its assumptions, even when $\sigma_k > 0$.

E.11 Multigrid residual correction diagnostics

Let $\mathbf{r}_{s,j,k}$ be the total residual (17) at reverse step k. Define post-correction residuals $\mathbf{r}^{\mathrm{post}}_{s,j,k}$ after the multiscale update (same chain state as used for reporting). Set

$$Q_{\mathrm{res}}(j,k) := \frac{\left(\sum_s \left\| \mathbf{r}^{\mathrm{post}}_{s,j,k} \right\|^2_{M_h} \right)^{1/2}}{\left(\sum_s \left\| \mathbf{r}_{s,j,k} \right\|^2_{M_h} \right)^{1/2} + \mathrm{tol}}, \tag{142}$$

The anchor-level ratio $\widehat{\rho}_{\mathrm{mg}}(j;K)$ coincides with (122). *Expected behavior:* on most steps, $Q_{\mathrm{res}}(j,k) < 1$ and $\widehat{\rho}_{\mathrm{mg}}(j;K) < 1$, indicating non-expansive residual reduction by the learned multiscale corrector rather than a classical SPD multigrid theorem.

E.12 Discrete energy and phase-field quantities

Using finite-difference gradients consistent with (41), define the discrete Ginzburg–Landau energy

$$\mathcal{E}_h(\mathbf{v}) := h^d \sum_{\mathbf{i}} \left[\frac{\varepsilon^2}{2} \left| \nabla_h \mathbf{v}_{\mathbf{i}} \right|^2 + \frac{1}{4} \left(1 - \mathbf{v}^2_{\mathbf{i}} \right)^2 \right], \tag{143}$$

where $\left| \nabla_h \mathbf{v}_{\mathbf{i}} \right|^2$ is the standard square sum of centered differences at $\mathbf{i}$ with the chosen boundary treatment. Track

$$E_{\mathcal{E}}(n) := \left| \frac{1}{N_{\mathrm{val}}} \sum_s \mathcal{E}_h(\mathbf{u}^n_{h,s}) - \frac{1}{N_{\mathrm{val}}} \sum_s \mathcal{E}_h(\widehat{\mathbf{u}}^n_{h,s}) \right|, \tag{144}$$

the per-step drift of discrete energy

$$D^+_{\mathcal{E}}(n) := \frac{1}{N_{\mathrm{val}}} \sum_s \left[\mathcal{E}_h(\widehat{\mathbf{u}}^{n+1}_{h,s}) - \mathcal{E}_h(\widehat{\mathbf{u}}^n_{h,s}) \right], \tag{145}$$

boundary-interface flux proxies $J_{\mathrm{bdry}}(j)$ (e.g. discrete normal flux through $\partial\Omega$ averaged near anchors), and a phase indicator error $E_{\mathrm{phase}}(n)$ comparing zero-level-set surrogates between reference and hybrid fields. *Expected behavior:* $E_{\mathcal{E}}$ stays small when the hybrid track preserves the energetic structure of the FD reference; spikes flag inconsistent anchors.

E.13 Spectral error diagnostics

Let $\mathcal{F}_h$ denote the discrete Fourier transform on the fine periodic patch (or cosine transform for Neumann boxes). Define the band-wise error density

$$P_{\mathrm{err}}(q,n) := \frac{\sum_{\|\kappa\| \in \mathcal{B}_q} \left| \mathcal{F}_h(\mathbf{u}^n_{h,s} - \widehat{\mathbf{u}}^n_{h,s})(\kappa) \right|^2}{\sum_{\kappa} \left| \mathcal{F}_h(\mathbf{u}^n_{h,s})(\kappa) \right|^2 + \mathrm{tol}}, \tag{146}$$

with frequency bins $\mathcal{B}_q$ for low / mid / high wavenumbers, and the spectral error summary

$$E_{\mathrm{spec}}(n) := \left(\sum_q w_q \, P_{\mathrm{err}}(q,n)^2 \right)^{1/2}. \tag{147}$$

Expected behavior: ReMD + Allen–Cahn residual improves P_{err} across bands compared with pure interpolation or "no multigrid" ablations, consistent with multiscale correction.

E.14 Intrinsic ambiguity under restriction-only conditioning

Use the nearest-neighbor statistic $A^{\mathrm{val}}_{\mathrm{id},j}$ from (114) and the kNN ambiguity proxy $\widehat{\epsilon}_{\mathrm{id},j}(q)$; abbreviate $A_{\mathrm{id}}(j) := A^{\mathrm{val}}_{\mathrm{id},j}$. *Expected behavior:* small $A_{\mathrm{id}}(j)$ and $\widehat{\epsilon}_{\mathrm{id},j}(q)$ indicate that nearby coarse snapshots $Y_{s,j} = \mathbf{u}^j_{H,s}$ predict similar HR anchors on the dataset support (cf. $\epsilon_{\mathrm{id},j}$ in Definition 1).

E.15 Time segmentation and propagation

Let $\tau^{(m)}$ denote a schedule with segment count indexed by m. Define

$$E_{\mathrm{seg}}(j,m) := \max_{0 \le \ell \le M_j} \left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| \mathbf{u}^\ell_{h,s,j} - \widehat{\mathbf{u}}^\ell_{h,s,j} \right\|^2_{M_h} \right)^{1/2}, \tag{148}$$

and the segment-averaged amplification factor versus anchor RMS,

$$A_{\mathrm{seg}}(j) := \frac{E_{\mathrm{seg}}(j,m)}{\left(\frac{1}{N_{\mathrm{val}}} \sum_s \left\| \mathbf{u}^0_{h,s,j} - \widehat{\mathbf{u}}^0_{h,s,j} \right\|^2_{M_h} \right)^{1/2} + \mathrm{tol}}. \tag{149}$$

Expected behavior: shorter segments (smaller $\tau_{\max}$) reduce the growth captured by A_{seg}, consistent with $\exp(L_B \tau_{\max})$ in Theorem 3.6.

E.16 Training data budget and computational accounting

Let N_{anc} be the number of distinct anchor indices used for training. Define the high-resolution supervision fraction and inflation factor,

$$r_{\mathrm{HR}} := \frac{N_{\mathrm{anc}}}{N_t + 1}, \qquad S_{\mathrm{HR}} := \frac{N_t + 1}{N_{\mathrm{anc}}}. \tag{150}$$

Wall-clock cost is reported as an accounting identity rather than an automatic speedup claim:

$$T_{\mathrm{inf}} := T_H + N_{\mathrm{anc}} K T_{\mathrm{rev}} + N_t T_{\Phi_h}, \tag{151}$$

where T_H is coarse-rollout time, T_{rev} is mean reverse time per anchor (including multigrid passes), and T_{Φ_h} is one fine Euler step. The main savings targeted by segmentation are (i) fewer high-resolution labels ($r_{\mathrm{HR}} \ll 1$), (ii) parallel segment-wise training, and (iii) controlled amplification of learned errors via smaller $\tau_{\max}$. If the baseline is the *same* explicit FD fine rollout, one should not infer automatic inference speedup from (151); any runtime advantage must be measured against a declared reference solver (e.g. implicit fine scheme or ensemble fine runs), reported separately with identical hardware.

E.17 Reporting checklist

The qualitative checklist in E.2 is instantiated with pooled component diagnostics in Table 5 and with segment and sampler sweeps in Section G; symbol definitions appear in Appendix E. These diagnostics are descriptive and are not substituted for uniform theoretical constants.

F IMPLEMENTATION DETAILS

Scope and architecture. The earlier Neumann-boundary suite and the controlled periodic P1 suite have different training schedules and aggregation units. P1 is described below from the delivered code and resolved run manifests; its settings must not be inferred from an unused default configuration. The conditional Swin U-Net uses width 160, channel multipliers $(1, 2, 2, 4)$, two residual blocks per scale, attention resolutions $(64, 32, 16, 8)$, window size 8, and no dropout. M2NO uses width 64 and grid levels $(8, 8, 4, 2, 1)$.

P1 training and sampling. P1 trains for 100 epochs with AdamW, learning rate 4×10^{-5}, weight decay 10^{-3}, batch size four per GPU, and two-GPU DDP. It uses clean-state prediction, relative-L^2 reconstruction loss, and M2NO output regularization weighted by 10^{-3}. Training samples a single diffusion index; no sampler-level fine-tuning is performed. The implemented perturbation and reverse updates are given in (21) and (15). The noise schedule is exponential with power 0.3, terminal parameter $\eta_K = 0.99$, minimum noise level 0.04, $\kappa = 2$, and $K = 15$. Evaluation calls `p_sample_loop`: it draws noise at each reverse step except the final one and disables clean-state clipping. Fixed seeds make evaluation reproducible but do not make the kernel deterministic.

Residual weights and transfers. For code index $t = k - 1$ and $s = 1 - t/(K - 1)$, the AC weight is

$$\lambda_k = \tfrac{1}{2}(1 - \cos(\pi s)) \left[0.7 + 0.15 \left(1 + \cos\left(\pi t/(K - 1)\right)\right)\right].$$

The P1 weight range is $[0, 1]$; the field reaction coefficient is separately set to one. Native LR fields supply the data and vector-field targets. Restriction uses area averaging over 4×4 blocks, and prolongation uses bicubic interpolation. No additional segment-duration factor multiplies the implemented flux residual: λ_k corresponds to $\omega_{\mathrm{AC},k}\Delta T_j$ in the analytical scaling. Periodic AC operators use the five-point stencil with wrapped neighbors and mesh spacing L/N; Neumann operators use endpoint-inclusive spacing $L/(N - 1)$. The generic original-ReMD residual retains its inherited implementation.

Comparators and randomness. The two original-ReMD rows reuse one checkpoint. Unsegmented inference generates one initial anchor and advances 200 Euler steps. Segmented inference generates ten anchors from successive coarse observations and also advances 200 steps. P1 uses the same 20 test trajectories and evaluation base seed 42. For trajectory index i and segment index j, the seed is $42 + 1000i + j$; the unsegmented run uses $42 + 1000i$. Full training seeds are 42, 43, and 44. The reconstruction comparison therefore changes both the observation usage and number of sampler calls. It is not an equal-compute comparison. NO-M2NO disables $\mathcal{S}_\varphi$ while retaining the raw residual; it must not be identified with an entirely uncorrected diffusion sampler.

Earlier suite. The earlier Neumann-boundary artifact reports pooled segment statistics from one checkpoint per configuration. Its recorded training setup allows up to 500 epochs, validation every five epochs, and early-stopping patience 30. These settings do not describe P1. The K sweep resamples one trained checkpoint rather than training a model for each K. The correction-ratio diagnostics in that suite are descriptive ablation ratios, not estimates of a uniform multigrid contraction constant.

G VALIDATION SETUP AND ADDITIONAL NUMERICAL RESULTS

Physical and evaluation setup. All AC2D runs use $\Omega = [0, 1]^2$, homogeneous Neumann conditions, $\varepsilon = 0.01$, $\Delta t_h = 0.01$, and a $4\times$ spatial scale factor. The default pair is $64^2 \to 256^2$, the default segment stride is 20 ($\Delta T_j = 0.2$), and the default reverse chain has 15 steps. Table entries are pooled over held-out segments from a single trained checkpoint per configuration.

Component and objective ablations. NO-AC sets the Allen–Cahn residual weight to zero; OLD-ρ restores the original generic residual; and NO-M2NO disables the learned M2NO refinement while retaining the raw residual update and reconstruction training objective. Two-frame supplies both coarse segment endpoints. Increment proxy replaces the start-state vector-field proxy by the observed coarse increment divided by ΔT_j. Denoising-only both changes the training objective and disables M2NO, so it is a compound ablation.

Table 4: Resolved setup used by the reported AC2D experiments.

Category	Parameter	Value
Physics	domain; interface width	$[0,1]^2$; $\varepsilon = 0.01$
Grid	LR; HR; scale factor	64^2; 256^2; $4\times$
Rollout	Δt_h; default M_j; ΔT_j	0.01; 20; 0.20
Stability	$\theta_h = \Delta t_h \varepsilon^2 \rho(L_h)$	0.2601
Sampler	default reverse steps	$K = 15$
Evaluation	default number of segments	200
Training	split; seed	$0.8/0.1/0.1$; 42

Table 5: Supplementary component diagnostics for stride 20. Base RMSE is the anchor RMSE before the evaluated correction; consequently $\widehat{\rho}_{\mathrm{mix}} \approx 1$ indicates that the current correction changes anchor error very little. These pooled ratios do not verify the uniform contraction in Assumption 4.

Variant	Base RMSE	$\widehat{\rho}_{\mathrm{mix}}$	E_{avg}	E_{rollout}	E_{data}	E_{AC}	Q_{res}	B_{val}
FULL	0.066110	1.000048	0.003908	0.004317	0.005697	0.005714	0.184537	1.311943
NO-AC	0.067040	1.000336	0.003966	0.004419	0.005752	0.006380	0.102255	1.300985
OLD-ρ	0.064857	1.040807	0.004240	0.004423	0.006277	0.002966	0.899973	1.202089
NO-M2NO	0.063734	1.000786	0.003964	0.004152	0.005542	0.003162	0	1.245742
Two-frame	0.064963	1.000024	0.003771	0.004171	0.005599	0.005557	0.051207	1.326995
Increment proxy	0.064485	1.000022	0.003751	0.004113	0.005503	0.008718	0.111324	1.255029
Denoising-only	0.065040	1.000040	0.003784	0.004172	0.005502	0.005059	0	1.311751

Segment length. The same FULL design is evaluated with segment strides 10, 20, and 40. Strides 10 and 20 use 200 held-out segments; the complete stride-40 split contains 100. The change in sample count and segment distribution means this is a sensitivity study rather than a paired significance test.

Table 6: Segment-stride sensitivity. Bold marks the best value in each error column; A_{seg} is reported as a diagnostic rather than optimized.

M_j	N	ΔT_j	Anchor RMSE	E_{avg}	A_{seg}	E_{AC}	$E_{\mathcal{E}}$
10	200	0.10	**0.064897**	0.004327	0.976085	**0.004446**	**0.002030**
20	200	0.20	0.066113	**0.003908**	0.988518	0.005714	0.002201
40	100	0.40	0.074706	0.004341	1.088794	0.008637	0.004268

The GT-anchor oracle confirms the same trend in pure numerical propagation: its final-step RMSE increases from 0.007293 at stride 10 to 0.012722 at stride 20 and 0.023062 at stride 40. The corresponding rollout-mean RMSE values are 0.004414, 0.007418, and 0.012958. These are all substantially below the learned-model errors, identifying the reconstructed anchor as the dominant error source in the current pipeline.

Number of reverse steps. Table 7 evaluates one FULL checkpoint with four inference respacings. Differences are below 2×10^{-5} in anchor RMSE, so more reverse steps do not materially improve this checkpoint.

Resolution variant. A separately trained $32^2 \to 128^2$ model at stride 20 obtains anchor RMSE 0.219429, final-step RMSE 0.220531, rollout-mean RMSE 0.218912, and relative L^2 error 0.703807 on 200 segments, with $\widehat{\rho}_{\mathrm{mix}} = 1.028127$. This negative result shows that the reported architecture/checkpoint does not transfer its $64^2 \to 256^2$ accuracy automatically to another absolute resolution, even at the same $4\times$ scale factor. Because this is a separately trained model and the reference distributions differ, the result is a resolution-sensitivity check, not a controlled discretization-convergence estimate.

Reproducibility and statistical scope. Every scalar above is transcribed from the resolved evaluation summaries. They are pooled estimates over held-out segments, not uncertainty estimates over

Table 7: Inference-only reverse-step sweep using the same trained checkpoint.

K	Anchor RMSE	Base RMSE	$\widehat{\rho}_{\mathrm{mix}}$	E_{avg}	E_{spec}	$E_{\mathcal{E}}$
2	0.066103	0.066079	1.000310	0.003905	177.363	0.002189
5	0.066106	0.066101	1.000060	0.003908	177.394	0.002208
10	0.066111	0.066108	1.000041	0.003908	177.387	0.002203
15	0.066114	0.066106	1.000113	0.003908	177.408	0.002201

independently trained models. Independent multi-seed training remains necessary before making confidence or significance claims for these preceding configurations; the separate replication below quantifies this variability for FULL under the new periodic-boundary protocol.

G.1 P1 FAIR COMPARISON AND INDEPENDENT-SEED REPLICATION

This subsection reports a new, independent experiment suite and leaves all preceding results unchanged. Unlike the preceding Neumann-boundary, segment-pooled evaluation, this suite uses periodic boundaries and a trajectory-level split: 200 trajectories are divided into 160 training, 20 validation, and 20 test trajectories, with the test set `IBVP2_s180-s199`. Each trajectory contains 201 frames at $\Delta t_h = 0.01$; the spatial map is $64^2 \to 256^2$, the segment stride is 20, and $K = 15$ reverse steps are used. FULL and original ReMD are trained for 100 epochs with the same split, seed 42, optimizer (AdamW, learning rate 4×10^{-5} and weight decay 10^{-3}), two-GPU DDP, and batch size four per GPU. Their intended methodological difference is the residual definition; this affects both training perturbations and clean-state refinement. Evaluation uses FP32 and batch size one on one RTX 4090; timings include anchor generation and Euler propagation but exclude data loading and metric computation, with CUDA synchronization around timed regions.

Table 8: Controlled P1 comparison on 20 complete held-out trajectories. Lower is better except for Dice. Bold compares accuracy and physics metrics for the two segmented learned methods only; the oracle is not a learned competitor. "Time" is mean total time per trajectory under the common timing protocol and is left unranked because repeated timing trials are unavailable.

Method	Anchor	Rollout	Final	Rel. L^2	E_{AC}	$E_{\mathcal{E}}$	Hausdorff	Dice	Time (s)	Mem. (MiB)
INTERP+Euler	0.091759	0.088954	0.095027	0.280147	0.054365	0.002072	0.044926	0.914147	0.039	26.8
Original ReMD, unsegmented	0.145527	0.080480	0.107767	0.245475	9.30×10^{-7}	0.007573	0.045539	0.938574	1.263	1027.4
Original ReMD, segmented	**0.065085**	**0.061956**	**0.064039**	**0.195034**	**0.044638**	0.001547	0.045364	**0.938817**	12.232	1037.9
FULL ACReMD	0.065835	0.062606	0.064680	0.197309	0.046852	**0.001396**	**0.044960**	0.937837	9.786	1037.9
GT-ANCHOR+Euler	0	0.004780	0.009398	0.012677	0.037103	0.000602	0.006481	0.999498	0.039	26.8

The segmented original-ReMD control has slightly lower field errors than FULL (about 1.0% lower rollout RMSE), so these measurements do not support a blanket accuracy advantage for the AC-aware residual. In contrast, FULL lowers mean energy mismatch by 9.7% and slightly lowers the interface Hausdorff error. Its recorded total time is about 20.0% lower than that of segmented original ReMD in this run, but repeated timing trials are unavailable; we therefore report this as a descriptive measurement and do not infer an intrinsic speed advantage. Relative to interpolation, FULL lowers anchor and rollout RMSE by 28.2% and 29.6%, respectively. The nearly zero E_{AC} of unsegmented original ReMD is a consequence of evaluating a trajectory generated by repeated use of the same Euler update; it should not be interpreted as superior agreement with the reference trajectory, as its final RMSE and energy mismatch are substantially larger.

Table 9: Independent-training replication of FULL. Each entry is the mean over the same 20 test trajectories; the final row reports the mean and sample standard deviation across the three training seeds.

Full seed	Anchor	Rollout	Final	Rel. L^2	E_{AC}	$E_{\mathcal{E}}$	Dice	Time (s)
42	0.065835	0.062606	0.064680	0.197309	0.046852	0.001396	0.937837	9.786
43	0.066029	0.063231	0.064981	0.199080	0.046300	0.001443	0.937627	9.734
44	0.066524	0.063352	0.065305	0.199072	0.044461	0.001391	0.937734	9.814
Mean ± s.d.	0.066130±0.000355	0.063063±0.000401	0.064989±0.000313	0.198487±0.001020	0.045871±0.001252	0.001410±0.000029	0.937733±0.000105	9.778±0.040

The between-seed variation is small relative to the gap between interpolation and the learned methods. These repetitions quantify training variability for FULL; the original-ReMD controls use one training seed and hence do not provide paired multi-seed uncertainty. We therefore avoid significance claims for the small difference between the two segmented learned methods.

Observed segment-interface jumps. P1 stores the pixel RMSE between each previous segment tail and the new anchor at the same time. Across $20 \times 9 = 180$ internal interfaces, the mean and maximum are 0.009919 and 0.042303 for segmented original ReMD, 0.010411 and 0.036683 for Full, and 0.008245 and 0.009609 for GT-anchor+Euler. Full has a slightly larger mean but smaller maximum than segmented original ReMD. These values describe discontinuities; they do not establish continuity or a uniform improvement. The oracle also has a nonzero jump because each numerical tail is replaced by a reference-data head. Pixel RMSE equals $\|\mathbf{J}\|_2/\sqrt{N_h}$; on the P1 periodic unit-square grid, $M_h = h^2 I$ and $h = 1/256$, so it equals the mass norm as well.