Growing Images: Spatial Scheduling in Diffusion Inpainting

Anonymous Authors[1]

Abstract

Diffusion inpainting implicitly treats the order in which masked pixels are denoised as irrelevant. We argue in this manuscript that it is not. Holding the model, prompt, and per-patch compute budget fixed in Stable Diffusion inpainting, we show that progressively filling a mask from its boundary measurably shifts reconstruction, perceptual, and distributional metrics relative to standard parallel denoising. We study this analytically by recasting inpainting as *finite-depth approximate inference*. Specifically, we consider a block-structured Blume–Capel spin system: a diffusion-like stochastic interpolant tilts the prior into a time-dependent Gibbs posterior with block-dependent random fields. In this picture, schedule-induced bias turns into a predictable consequence of finite-depth inference.

1. Introduction

Diffusion and score-based generative models have become a dominant paradigm for high-fidelity image synthesis and editing (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021b; Karras et al., 2022; Rombach et al., 2022). A central reason for their practical success is that sampling is implemented as a finite sequence of denoising updates, e.g., Denoising Diffusion Implicit Models (DDIM) / Ordinary Differential Equation (ODE) solvers and related fast samplers, which necessarily trades computation for approximation error (Song et al., 2021a; Lu et al., 2022). Despite this, many influential implementation choices in *conditional* diffusion pipelines remain only weakly understood theoretically.

This paper focuses on one such choice in diffusion-based inpainting: *spatial scheduling*. Given a masked region, should one denoise the entire unknown set of pixels simultaneously (a *parallel* schedule), or reveal and denoise the mask progressively via a sequence of spatial patches (a *progressive* schedule)? While diffusion inpainting is commonly implemented with a parallel schedule, treating masked pixels as exchangeable, conditional generation is inherently *spatially asymmetric*: pixels adjacent to the observed boundary are strongly constrained, while pixels deep inside the mask are constrained primarily through the learned prior and long-range consistency. This raises a basic inference question: *under finite-step sampling, when does a spatial schedule preserve the intended conditional distribution, and when does it introduce systematic bias?* Beyond sample quality, schedule-dependent bias may also interact with memorization/copying phenomena documented for modern diffusion models (Carlini & et al., 2023; Somepalli et al., 2023; Gu et al., 2025).

Empirically, diffusion inpainting has been studied via both model/training design (Rombach et al., 2022; Saharia et al., 2022) and inference-time procedures that modify the reverse-time updates (Lugmayr et al., 2022). In contrast, we isolate *inference-time spatial scheduling* under a fixed pretrained model, and develop a theoretical framework that makes schedule effects analyzable. Our starting point is the standard denoising-as-inference viewpoint: at each noise level, denoising can be interpreted as approximate Bayesian inference under a noise-dependent posterior (cf. Tweedie's formula and its modern score-based instantiations) (Miyasawa, 1961; Efron, 2011; Hyvärinen, 2005; Vincent, 2011; Song et al., 2021b). Crucially, practical samplers perform only a *truncated* number of local updates, so the resulting distribution can depend on the order in which information is revealed.

Our approach We formalize spatial scheduling as *finite-depth approximate inference* in a block-structured spin system that isolates (i) spatial locality, (ii) boundary conditioning, and (iii) schedule-dependent inference. Concretely, we introduce a block-structured Blume–Capel model (Blume, 1966; Capel, 1966) in which observed blocks act as boundary conditions and unobserved blocks must be reconstructed consistently. A diffusion-like stochastic interpolant induces an explicit, time-dependent *tilted* Gibbs posterior with schedule-dependent effective fields, enabling closed-form comparisons between schedules (Albergo et al., 2025).

---

[1] Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the FoGen Workshop at ICML 2026. Do not distribute.

Contributions

- **A tractable block model for conditional diffusion with spatial structure:** We introduce a block-structured Blume–Capel system that mirrors inpainting: observed blocks provide boundary constraints while hidden blocks must be inferred jointly.
- **Schedule-dependent posteriors from diffusion-like interpolation:** We show that a block-wise stochastic interpolant yields an explicit time-dependent tilted Gibbs posterior with block-dependent effective fields and crystal terms, making the role of spatial scheduling analytically transparent.
- **Scheduling as finite-depth inference (and its consequences):** We characterize Bayes-consistent local updates and relate scheduled local Markov dynamics to an unrolled computation with shared parameters, where the number of denoising iterations corresponds to inference depth, clarifying when schedules agree and when finite-depth effects induce bias.
- **Empirical motivation in real diffusion inpainting under a fixed model:** Using a fixed pretrained Stable Diffusion inpainting model (LDM backbone (Rombach et al., 2022); model card / checkpoint documentation (RunwayML / Stability AI, 2022); trained on LAION-family data (Schuhmann et al., 2022; Schuhmann & Beaumont, 2022)), we vary *only* the spatial schedule and measure reconstruction, perceptual, and distributional metrics.

Related work and why statistical physics Most prior work on diffusion inpainting targets model or inference design: training-time objectives (Rombach et al., 2022; Saharia et al., 2022) and sampling-time modifications that re-inject the observed region at each reverse step (Lugmayr et al., 2022). We take the model as fixed and isolate the effect of the spatial order in which masked sites are denoised. Our analysis tools come from a line of work that uses exactly-solvable spin systems to probe diffusion: Biroli et al. (Biroli et al., 2024) show how mean-field models make the regimes of diffusion sampling analytically explicit; Caltagirone et al. (Caltagirone et al., 2014) and Hassani et al. (Hamed Hassani et al., 2012) characterise Glauber dynamics on a spatially-coupled Curie–Weiss chain. Our setup inherits the block chain of the latter two but extends it from Ising to Blume–Capel so that the "no-signal" state (the initialisation of a diffusion trajectory) is a genuine phase of the prior, and uses it to compare *spatial update schedules* rather than a single uniform dynamics—an axis their analysis do not consider. The payoff of this physics framing is that schedule-induced bias, which is nearly invisible at the level of real-image metrics (Heusel et al., 2017; Zhang et al., 2018; Radford et al., 2021), becomes a closed-form prediction at the level of block magnetisations.

Organization Section 2 presents empirical motivation on real models. Section 3 introduces the block-structured model and diffusion-induced posterior. Section 4 analyzes scheduled local dynamics and finite-depth effects. Additional experimental details are deferred to the appendix.

2. Empirical Motivation: Spatial Scheduling in Diffusion Inpainting

We begin by testing whether spatial scheduling matters in a standard diffusion inpainting pipeline. Inpainting is often implemented by denoising all masked pixels jointly (a parallel schedule), implicitly treating unknown pixels as exchangeable. However, conditional constraints are spatially non-uniform: pixels near the known boundary are tightly constrained, while pixels deep inside the mask must be resolved through prior structure and long-range consistency. This suggests that *progressive* schedules, filling the mask sequentially from the boundary inward, could change outcomes under finite-step sampling.

Setup (fixed model; schedule varies) We fix a pretrained Stable Diffusion v1.5 inpainting checkpoint (latent diffusion backbone (Rombach et al., 2022); checkpoint/model card (RunwayML / Stability AI, 2022)) and modify *only* the spatial schedule used to fill a left-half mask. Specifically, we partition the masked region into K vertical stripe patches and fill them sequentially from the boundary inward (left-to-right), with $K \in \{1, 2, 3, 4, 6, 8, 10\}$. All other inference choices (prompting, guidance, noise schedule, and step budgeting) are held fixed and reported in Appendix C.

Metrics We measure (i) reconstruction fidelity via Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) (Wang et al., 2004); (ii) perceptual similarity via Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al., 2018) and Contrastive Language-Image Pre-training (CLIP)-based similarity (Radford et al., 2021); and (iii) dataset-level distributional quality via Fréchet Inception Distance (FID) (Heusel et al., 2017) and precision/recall-style coverage metrics (Kynkäänniemi et al., 2019; Sajjadi et al., 2018). Example images are drawn from Caltech-256 (Griffin et al., 2007); additional datasets and evaluation backbones are detailed in the appendix.

Explaining the metrics The eight metrics in Figure 1 fall into three groups with distinct operational meanings. *(i) Per-pixel reconstruction fidelity to the ground-truth image*: PSNR and SSIM, which directly compare inpainted pixels to the held-out reference. *(ii) Semantic/perceptual similarity to the ground truth*: CLIP similarity, which scores

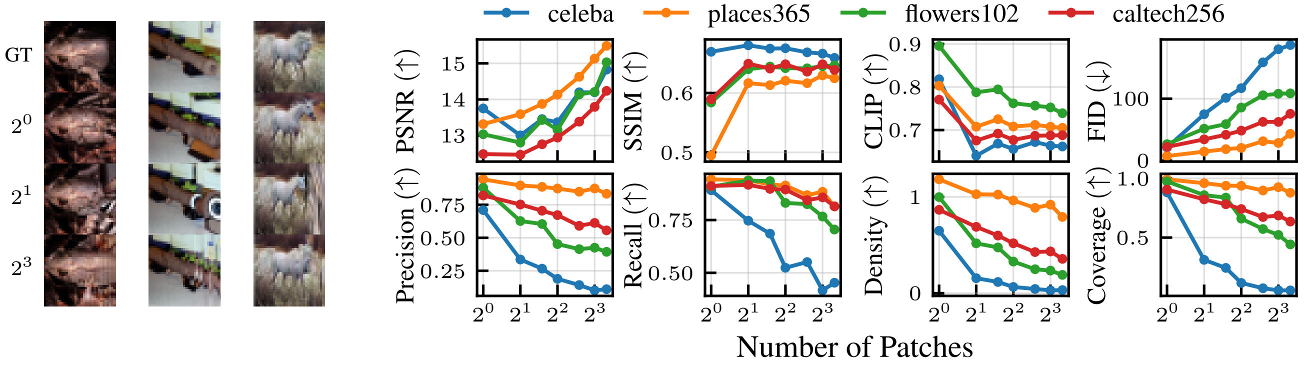

Figure 1. **Spatial scheduling affects inpainting quality metrics** *(Left)* Examples from Caltech-256 (Griffin et al., 2007) for different patch counts K. *(Right)* Reconstruction, perceptual, and distributional metrics as a function of the number of sequential patches K used in progressive inpainting (Stable Diffusion inpainting; left-half mask). Only the spatial schedule varies; the model, prompts, and inference hyperparameters are fixed. See Appendix C for full details.

high-level content agreement rather than pixel alignment. *(iii) Distributional quality of the generated set against real data*: FID, precision/recall (Kynkäänniemi et al., 2019), and density/coverage (Naeem et al., 2020), which probe on-manifold-ness and diversity of the sample distribution rather than per-image agreement with any single reference.

Empirical takeaway: a schedule-induced dichotomy Figure 1 reveals a pattern across all four datasets: as the number of schedule steps increase (larger K), *pixel-level* reconstruction metrics (PSNR, SSIM) *improve*, while *every* semantic and distributional metric (CLIP, FID, precision, recall, density, coverage) *degrades* — monotonically or nearly so. Concretely, PSNR rises by roughly $1{-}2\,\%$ from $K{=}1$ to $K{=}8$, whereas FID increases by $2{-}5\times$, and precision/density/coverage drop sharply (most severely on CelebA, where the distributional metrics collapse). The two families of metrics therefore do not move together: one improves precisely when the other deteriorates.

We interpret this as a local-versus-global trade-off induced by the schedule. Progressive filling conditions each new patch on the *already-resolved* neighboring pixels, which enforces strong local consistency with the known boundary and produces completions that look like smooth extensions of the observed region. This is exactly what pixel-level fidelity metrics reward. However, because each patch commits to a single realization before the rest of the mask is filled, the procedure behaves as a *greedy, finite-depth* approximation to the joint posterior: early choices cannot be revised in light of later ones, long-range consistency with the unseen half of the image is lost (CLIP ↓), samples drift off the learned data manifold (FID ↑, precision/density ↓), and the set of realized completions shrinks (recall/coverage ↓). The effect is strongest on CelebA, whose tight global face structure is most easily broken by locally-greedy completion, and weakest on Places365, whose scenes tolerate local reassembly.

This dichotomy is the phenomenon our theory is designed to explain: *when does a spatial schedule preserve the intended conditional distribution under finite-step inference, and when can it induce systematic bias through truncation and metastability?* The block-structured Blume–Capel model in Section 3 makes this trade-off analytically tractable, and Section 4 shows how finite-depth progressive updates lock in metastable configurations that are consistent with the boundary but biased away from the joint posterior, mirroring the empirical pattern above.

3. Theoretical Model

To isolate the effect of spatial scheduling in generative diffusion, we introduce a minimal, analytically tractable spin model. The design goal is a setting in which (i) the target distribution is explicit, (ii) the geometry of the mask is explicit, and (iii) schedule choices can be compared in closed form or by a cheap block-level recursion.

Block-structured geometry A configuration is a collection of spins $s = \{s_{\ell,i}\} \in \{-1, 0, +1\}^{LN}$ indexed by a block $\ell \in \{1, \dots, L\}$ and a site $i \in \{1, \dots, N\}$ within that block. Each block is a *Curie–Weiss cluster*: every pair of spins inside the block is coupled with strength J_{intra}/N. Consecutive blocks along the chain are coupled *all-to-all* between their spins with strength J_{inter}/N; non-adjacent blocks do not interact. This is the Blume–Capel extension of the spatially-coupled Curie–Weiss Ising chain of Caltagirone et al. (2014) and Hamed Hassani et al. (2012). The chain is open at $\ell = 1$ and $\ell = L$; these terminal blocks will act as boundary conditions for the inpainting analogy below.

Target distribution The target is the Gibbs measure

$$p(s) \;=\; \frac{1}{Z}\exp\bigl(-E(s)\bigr), \tag{1}$$

$$\begin{aligned} E(s) = &-\frac{1}{N}\sum_{\langle(\ell,i),(\ell',j)\rangle} J_{\ell\ell'}\, s_{\ell,i}\, s_{\ell',j} \\ &+ \Delta\sum_{\ell,i} s_{\ell,i}^2 - \sum_{\ell} h_\ell \sum_{i} s_{\ell,i}, \end{aligned} \tag{2}$$

where $\langle\cdot,\cdot\rangle$ runs over unordered pairs of spins $(\ell,i) \neq (\ell',j)$ sitting in the same block ($\ell = \ell'$) or in adjacent blocks ($|\ell-\ell'| = 1$); the block-level couplings are $J_{\ell\ell} = J_{\text{intra}}$, $J_{\ell,\ell\pm1} = J_{\text{inter}}$, and $J_{\ell\ell'} = 0$ otherwise. $\Delta > 0$ is the Blume–Capel crystal field penalizing the magnetized states $s = \pm1$; $h_\ell \in \mathbb{R}$ is an optional block-level field. For $L = 1$ (single block) and $h = 0$ this is the standard mean-field Blume–Capel model (Blume, 1966; Capel, 1966), with three locally stable phases ($m \in \{-1, 0, +1\}$) separated by first-order lines—the metastability we will exploit.

Convention. We absorb the inverse temperature β into $(J_{\text{intra}}, J_{\text{inter}}, \Delta, h_\ell)$, equivalently setting $\beta = 1$. The $1/N$ prefactor is the mean-field scaling that keeps intensive thermodynamic quantities finite as $N \to \infty$; as a consequence, the block-level recursion derived in Section 4 is *exact* in the large-N limit.

Why Blume–Capel rather than Ising? A binary Ising alphabet $\{-1,+1\}$ has no state that represents the *no-signal* regime of early diffusion: a pure-noise sample has magnetization $m \approx 0$, which in Ising is not a local minimum of the free energy but a saddle between the two magnetized phases. Blume–Capel adds $s = 0$ as a genuine third state, and the crystal field Δ makes it a locally stable phase whenever Δ is large enough. This matters for our question: an inpainting trajectory is initialized from noise (magnetization zero) and must decide, in finite depth, whether to roll into $m = -1$, $m = 0$, or $m = +1$ under boundary pressure. Ising collapses this ternary decision to a binary one and removes precisely the metastability a diffusion scheduler has to navigate.

Inpainting as conditional sampling Partition the block indices into an *observed* set $\mathcal{O} \subseteq \{1,\dots,L\}$—typically the two chain ends $\{1, L\}$—and a *hidden* set $\mathcal{H} = \{1,\dots,L\} \setminus \mathcal{O}$. Observed blocks are clamped (for instance, block 1 to $+1$ and block L to -1); the generative task is to draw $s^{\mathcal{H}} \sim p(\cdot \mid s^{\mathcal{O}})$. This is the block-structured analogue of inpainting: observed blocks play the role of boundary pixels, and hidden blocks must be reconstructed consistently with both the boundary and the coupling structure.

Different update orders over $\mathcal{H}$ define different *spatial schedules*. A *parallel* schedule resamples every hidden site with equal probability, mirroring standard diffusion denoising. A *progressive* schedule—e.g., a left-to-right sweep that equilibrates one hidden block before moving on—corresponds to boundary-to-interior inpainting, where information propagates sequentially along the chain. Section 4 makes this precise and shows (Figure 2) that, at matched iteration budget, progressive schedules lock the chain into schedule-dependent metastable profiles while the parallel schedule converges to the Bayes-optimal posterior mean.

3.1. Diffusion as Stochastic Interpolation

We now connect diffusion to the block-structured system above. A stochastic-interpolant corruption process induces, at each time t, a *tilted* Gibbs distribution over the clean spins; denoising is then Bayesian inference in an effective random-field Blume–Capel model whose parameters depend explicitly on the diffusion schedule.

Block-wise stochastic interpolant Write s^0 for a clean configuration on the chain; we corrupt each block at its own pace via

$$y_{\ell,i}(t) = \alpha_\ell(t)\, s^0_{\ell,i} + \sigma_\ell(t)\,\xi_{\ell,i}, \qquad \xi_{\ell,i} \sim \mathcal{N}(0,1), \tag{3}$$

with $\sigma_\ell(t) > 0$. Define the per-block signal-to-noise ratio $\gamma_\ell(t) := \alpha_\ell(t)/\sigma_\ell(t)$. A uniform noise schedule sets $\gamma_\ell(t) \equiv \gamma(t)$ for all blocks; a *spatial* schedule lets $\gamma_\ell(t)$ differ across blocks, e.g., already-denoised blocks have $\gamma \to \infty$ and still-noisy blocks have $\gamma \to 0$.

Proposition 3.1 (Posterior induced by diffusion schedules)**.** *Let $y(t) = \{y_{\ell,i}(t)\}$ be generated from Equation (3) and let the prior be Equations (1) and (2). Define the effective crystal field and effective site field*

$$\widetilde{\Delta}_\ell(t) \;:=\; \Delta + \tfrac{1}{2}\gamma_\ell^2(t), \tag{4}$$

$$\widetilde{h}_{\ell,i}(t) \;:=\; h_\ell + \gamma_\ell(t)\,\xi_{\ell,i} + \gamma_\ell^2(t)\, s^0_{\ell,i}. \tag{5}$$

Then the posterior over $s \in \{-1,0,+1\}^{LN}$ satisfies

$$\begin{aligned} P(s \mid y(t)) \;\propto\; \exp\Bigl[\, &\tfrac{1}{N}\sum_{\langle(\ell,i),(\ell',j)\rangle} J_{\ell\ell'}\, s_{\ell,i}\, s_{\ell',j} \\ &- \sum_{\ell,i}\widetilde{\Delta}_\ell(t)\, s_{\ell,i}^2 \\ &+ \sum_{\ell,i}\widetilde{h}_{\ell,i}(t)\, s_{\ell,i}\,\Bigr]. \end{aligned} \tag{6}$$

The proof is a direct application of Bayes' rule under the Gaussian observation model Equation (3) followed by a completion of squares; see Appendix A.

Proposition 3.1 identifies denoising at diffusion time t with inference in a Blume–Capel model of the same coupling structure as the prior, but with two ingredients reshaped by the per-block signal-to-noise ratio $\gamma_\ell(t)$:

1. **A signal-dependent site field $\widetilde{h}_{\ell,i}(t)$.** It has a score-like pull $\gamma_\ell^2(t)\, s^0_{\ell,i}$ toward the clean sample and a noise term $\gamma_\ell(t)\, \xi_{\ell,i}$ inherited from the corruption Equation (3). At high SNR ($\gamma_\ell \to \infty$) the pull dominates and the posterior mass concentrates on s^0; at low SNR the noise term dominates and the posterior reverts to the prior.
2. **A signal-dependent "vacancy" penalty $\widetilde{\Delta}_\ell(t)$.** As $\gamma_\ell(t)$ grows, the effective crystal field stiffens, *ceteris paribus* favoring the no-signal state $s = 0$. Whether $s = 0$ actually wins depends on the competition with the alignment term above: on sites where $s^0 \neq 0$ the γ_ℓ^2 coefficient of the alignment (1) beats the $\frac{1}{2}\gamma_\ell^2$ of the stiffening (2), so concentration on s^0 wins in the high-SNR limit.

Both mechanisms depend on $\gamma_\ell(t)$ block-wise, so a spatial schedule (already-denoised blocks have γ_ℓ large and still-noisy blocks have γ_ℓ small) corresponds to traversing different slices of this $\gamma(t)$-parameterised family of posteriors in parallel. Section 4 studies finite-depth inference inside this family.

Regimes of the single-block posterior The single-block mean-field free energy of Equation (6) (one block, magnetization m, averaged over the planted noise ξ and the clean marginal s^0) exhibits up to three distinct phases in the (h, Δ) plane. Figure 3 in Appendix B tracks how these phases migrate as γ grows: (i) at low γ, the crystal field dominates and the landscape has a single "no-signal" minimum at $m \approx 0$; (ii) at intermediate γ, the alignment term opens a second minimum near $m \approx s^0$, producing the coexistence region in which the schedule determines which basin is reached; (iii) at large γ the aligned minimum is globally lowest and the posterior concentrates on s^0. The finite-depth bias analysed in Section 4 lives in regime (ii), where multiple basins coexist and the spatial schedule picks between them.

4. Scheduled Dynamics and Finite-Depth Effects

Proposition 3.1 identified denoising at diffusion time t with inference in the posterior $P(s \mid y(t))$. This section studies how a *finite* number of local updates approximates that posterior, shows that the resulting computation is a tied-weight 1D CNN, and quantifies the schedule-induced bias that remains at finite depth.

4.1. Bayes-Consistent Local Updates and the CNN Picture

Proposition 4.1 (Bayes-consistent local update). *Fix a target distribution $q(s)$ on a discrete product space and consider updating a single variable s_v with $s_{\setminus v}$ held fixed. (i) Any Markov kernel that only rewrites s_v (leaves $s_{\setminus v}$ fixed) and admits q as an invariant distribution must resample $s_v \sim q(\cdot \mid s_{\setminus v})$; this is the unique q-invariant single-site randomised update. (ii) Among deterministic updates measurable with respect to $s_{\setminus v}$, the conditional mean $\hat{s}_v = \mathbb{E}_q[s_v \mid s_{\setminus v}]$ minimises the mean squared error $\mathbb{E}_q[(s_v - f(s_{\setminus v}))^2]$.*

Part (i) is the standard characterisation of the Gibbs sampler; part (ii) is the elementary MMSE property of the conditional expectation. The proof is collected in Appendix A.

Applied to the posterior in Equation (6) and grouped at block level, this says the Bayes-consistent update of a block magnetization is its conditional expectation given the current neighboring blocks. Within each block the large-N mean-field collapses the conditional expectation to a function of the current block magnetizations, giving a deterministic block-level map

$$m^{(k+1)} \;=\; \Phi(m^{(k)}), \tag{7}$$

with Φ depending only on the block couplings $J_{\ell\ell'}$, the fields h_ℓ, and the Blume–Capel response function. When $J_{\ell\ell'}$ is nearest-neighbor along the chain, each iteration of Φ is a 1D convolution followed by a pointwise nonlinearity, so iterating Φ for T steps is exactly a tied-weight 1D CNN of depth T. Fixing the number of denoising iterations, therefore, amounts to truncating an otherwise infinite-depth Bayes-consistent computation, and the schedule chooses *which* blocks receive how much depth.

4.2. Block-Glauber Dynamics and the Mean-Field Recursion

We make the picture concrete with a discrete-time local Markov dynamics (Glauber / heat-bath) that targets $p(s)$ and admits an explicit block-wise description.

Block magnetization For each block $\ell \in \{1, \dots, L\}$,

$$m_\ell(s) \;:=\; \frac{1}{N}\sum_{i=1}^{N} s_{\ell,i} \in [-1, 1], \tag{8}$$

and its expectation at discrete time t,

$$m_\ell^{(t)} \;:=\; \mathbb{E}\Big[m_\ell(s^{(t)})\Big], \tag{9}$$

with $m^{(t)} = (m_1^{(t)}, \dots, m_L^{(t)})$.

Scheduled heat-bath updates Fix a (possibly time-dependent) probability vector $\pi^{(t)} = (\pi_1^{(t)}, \dots, \pi_L^{(t)})$, $\pi_\ell^{(t)} \geq 0$, $\sum_\ell \pi_\ell^{(t)} = 1$; this is the *spatial schedule*. Observed/inpainting blocks are captured by setting $\pi_\ell^{(t)} \equiv 0$

there. At each step: (a) draw $\ell_t \sim \pi^{(t)}$; (b) draw $i_t \sim \mathrm{Unif}\{1,\dots,N\}$; (c) resample $s^{(t)}_{\ell_t,i_t}$ from the Gibbs conditional induced by $E(s)$ in Equation (2), leaving all other spins fixed.

Schedules condisered Let $\mathcal{H}$ be the set of blocks the scheduler updates. The *parallel* schedule is the diffusion-inpainting default: $\pi^{(t)}_\ell = 1/|\mathcal{H}|$ on every $\ell \in \mathcal{H}$ for every t. The two *progressive* schedules concentrate all mass on one block at a time and advance through $\mathcal{H}$ in order: progressive L→R places unit mass on the leftmost block for a phase budget T_{phase}, then the next-to-leftmost, and so on up to the rightmost; progressive R→L is the mirror sweep. The *interleaved* schedule is also concentrated on one block at a time but cycles through $\mathcal{H}$ round-robin with a shorter per-block budget T_{int}, revisiting each block many times before the horizon T is exhausted. We set $T_{\mathrm{phase}} = T/|\mathcal{H}|$ for the progressive variants and $T_{\mathrm{int}} = 20$ for interleaved; with these choices every block in $\mathcal{H}$ receives the same total number $T/|\mathcal{H}|$ of updates over the horizon, so differences between schedules reflect only *when* each block is visited rather than *how often*.

Proposition 4.2 (Magnetization evolution under scheduled block-Glauber)**.** *Let $\{s^{(t)}\}_{t\geq 0}$ be the scheduled heat-bath chain with target $p(s)$ from Equation (1), driven by any (deterministic or random, time-homogeneous or time-inhomogeneous) schedule $\{\pi^{(t)}\}_{t\geq 0}$. Then for each block ℓ,*

$$m^{(t+1)}_\ell = m^{(t)}_\ell + \frac{\pi^{(t)}_\ell}{N}\,\mathbb{E}\left[\mathbb{E}_{p(\,\cdot\,|s^{(t)}_{\setminus(\ell,I)})}[s_{\ell,I}] - s^{(t)}_{\ell,I}\right], \quad (10)$$

with $I \sim \mathrm{Unif}\{1,\dots,N\}$ independent of $s^{(t)}$.

The proof is in Appendix A. Proposition 4.2 identifies one source of schedule dependence, i.e., the per-step update rate $\pi^{(t)}_\ell/N$ in front of the drift, and leaves the drift itself schedule-independent. In particular, the drift vanishes exactly when $s^{(t)}_{\ell,I}$ already equals its Gibbs conditional mean for every I, and this zero-drift condition does not depend on π: all schedules share the same set of fixed points for the magnetization recursion. Differences at finite t therefore correspond to finite-depth truncation of the same underlying dynamics rather than to different targets.

Mean-field block map In the large-N limit the conditional mean in Equation (10) depends on $s^{(t)}$ only through the current block magnetizations,

$$\mathbb{E}_{p(\,\cdot\,|s_{\setminus(\ell,i)})}[s_{\ell,i}] \;\approx\; \mu_\Delta\Big(\textstyle\sum_{\ell'} J_{\ell\ell'}\, m^{(t)}_{\ell'} + h_\ell\Big), \quad (11)$$

with Blume–Capel response (at the $\beta = 1$ convention of Section 3)

$$\mu_\Delta(u) \;=\; \frac{2\sinh(u)}{e^{\Delta} + 2\cosh(u)}. \quad (12)$$

Substituting Equation (11) into Equation (10) gives the closed block-level recursion

$$m^{(t+1)}_\ell = m^{(t)}_\ell + \frac{\pi^{(t)}_\ell}{N}\Big[\,\mu_\Delta\Big(\textstyle\sum_{\ell'} J_{\ell\ell'}\, m^{(t)}_{\ell'} + h_\ell\Big) - m^{(t)}_\ell\Big], \quad (13)$$

which is exactly the map Φ of Equation (7): a local linear aggregation across blocks followed by the pointwise nonlinearity μ_Δ. Each iteration is one CNN layer; spatial scheduling corresponds to non-uniform per-block step sizes $\pi^{(t)}_\ell$, i.e. *different effective depths across space*. We take $m^\star$ to be the unique fixed point of the infinite-depth parallel recursion ($\pi^{(t)}_\ell \equiv 1/|\mathcal{H}|$, $T \to \infty$) from the noise initialization $m^{(0)} = 0$; it is the mean-field posterior mean, the *Bayes-optimal* reconstruction under the tilted Blume Capel posterior, and is shared across schedules by the invariance of Proposition 4.2.

4.3. Finite-Depth Bias: Why Schedule Matters

We iterate Equation (13) on an 8-block chain ($L = 8$) with couplings $J_{\mathrm{intra}} = 1$, $J_{\mathrm{inter}} = 0.6$, $\Delta = 0.5$; the results are insensitive to these exact values inside the ordered regime. The leftmost block is clamped to $+1$; the other seven ($|\mathcal{H}| = 7$) are the unclamped blocks that the scheduler operates on, initialised at $m^{(0)}_\ell \approx 0$ plus a small seed-dependent Gaussian perturbation (standard deviation 0.1). With the horizon $T = 21{,}000$ the per-phase budget is $T_{\mathrm{phase}} = T/|\mathcal{H}| = 3{,}000$ and the per-block interleaved budget is $T_{\mathrm{int}} = 20$; each concentrated schedule then spends exactly $T/|\mathcal{H}| = 3{,}000$ total updates on every unclamped block over the horizon, so differences in the resulting profile reflect *when* a block is updated rather than *how often*.

Figure 2 shows the block magnetization profile at three depths—$T = T_{\mathrm{phase}}$ (one progressive phase complete), $T = 4\,T_{\mathrm{phase}}$ (progressive past the halfway point), and $T = |\mathcal{H}|\,T_{\mathrm{phase}}$ (progressive sweep finished)—averaged over 24 seeds with ±1 standard-deviation error bars. The Bayes-optimal profile (dotted line) is a monotonic decay from $m = 1$ at the clamp to $m \approx 0.36$ at the far end: information from the clamped block propagates through the nearest-neighbour coupling but attenuates with distance. Parallel and interleaved schedules reach this profile within machine precision even at the earliest depth shown, because every unclamped block accumulates updates throughout the run. Progressive L→R locks into a *different* monotonic decay: the leftmost hidden block equilibrates against the clamp and a still-zero right neighbour, ending up below its Bayes level; the next block inherits that undershoot; the deficit accumulates along the chain. Progressive R→L fails catastrophically: its first phase updates the rightmost hidden block, which has no clamped neighbour, and the sweep never picks up a source of magnetization—the entire chain sits near zero except for the seed-dependent perturbation.

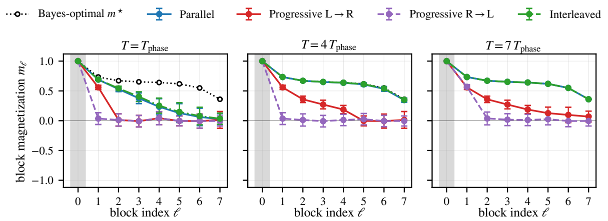

Figure 2. Block magnetization profile m_ℓ vs. block index for each schedule at three horizons, expressed in units of the progressive phase budget $T_{\text{phase}} = T/|\mathcal{H}| = 3{,}000$: after one phase (left), past the halfway point (middle), and after all $|\mathcal{H}| = 7$ progressive phases are complete (right). Block 1 (grey band) is clamped to $+1$; the remaining seven blocks are updated by the scheduler from a near-zero initial magnetization. Curves and error bars are the mean and ± 1 standard deviation across 24 seeds. Dotted black: the Bayes-optimal profile $m^\star$, which is a monotonic decay from the clamped boundary. Parallel (blue) and interleaved (green) recover $m^\star$ within error; progressive L→R (red) cascades into a magnetization deficit that accumulates along the chain; progressive R→L (purple) sweeps away from the clamp and never picks up the boundary signal.

The corresponding ℓ_2 bias $\|m^{(T)} - m^\star\|_2$ confirms the same ordering (Figure 4, appendix): parallel and interleaved reach machine precision; progressive L→R plateaus near $\|m^{(T)} - m^\star\|_2 \approx 1.0$; progressive R→L plateaus near 1.4. Both plateaus persist up to $T = 4.2 \times 10^5$, 20× the horizon used here, so we treat them as extremely long-lived metastable basins rather than soft transients.

Reading the empirical dichotomy through the block map The progressive cascade in Figure 2 is the toy-model signature of the same mechanism we argued for in Section 2. Each time the progressive schedule advances to a new block, that block equilibrates against a *partial* view of the posterior, with respect only to its already-resolved neighbour, and commits to the magnetization that partial view favours. This matches what pixel-level fidelity metrics reward: strong local consistency with the observed boundary (PSNR and SSIM ↑ in Figure 1). But the committed profile is not the one the joint posterior would select: it decays more sharply than $m^\star$ as one moves into the interior, each block's undershoot is inherited by its successor, and the resulting chain sits in a narrow metastable basin. In the image experiments, the analogous object is a completion that is locally coherent with the observed half of the image but whose global image statistics are off: FID rises, precision, density, and coverage fall, and CLIP similarity drops because semantic consistency with the unseen half is lost. The schedule does not change the fixed points of the block recursion, as by Proposition 4.2 these are shared across schedules, but only which basin a finite number of local updates settles into.

5. Conclusions

Varying *only* the spatial schedule in a fixed pretrained Stable Diffusion inpainter produces a reproducible dichotomy: progressive boundary-to-interior filling improves pixel-level fidelity (PSNR, SSIM) while degrading every semantic and distributional metric we tested (CLIP, FID, precision, recall, density, coverage). The block-structured Blume–Capel model of Sections 3 and 4 explains the mechanism: the mean-field block recursion has the same fixed points under every schedule, but finite-depth progressive updates leave the chain on a long metastable plateau in which every hidden block aligns with the boundary it was swept from, compactly: pixel-locally consistent, globally off the joint posterior. A concrete follow-up is a *schedule-aware* inference rule that spends more denoising iterations on blocks farthest from any observed boundary, closing the effective-depth gap our theory predicts.

References

Albergo, M., Boffi, N. M., and Vanden-Eijnden, E. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26 (209):1–80, 2025.

Biroli, G., Bonnaire, T., de Bortoli, V., and Mézard, M. Dynamical regimes of diffusion models. *Nature Communications*, 15(1):9957, November 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-54281-3.

Blume, M. Theory of the first-order magnetic phase change in uo$_2$. *Physical Review*, 141(2):517–524, 1966.

Caltagirone, F., Franz, S., Morris, R. G., and Zdeborová, L. Dynamics and termination cost of spatially coupled mean-field models. *Phys. Rev. E*, 89:012102, Jan 2014. doi: 10.1103/PhysRevE.89.012102. URL `https://link.aps.org/doi/10.1103/PhysRevE.89.012102`.

Capel, H. W. On the possibility of first-order phase transitions in ising systems of triplet ions with zero-field splitting. *Physica*, 32(5):966–988, 1966.

Carlini, N. and et al. Extracting training data from diffusion models. In *USENIX Security*, 2023.

Efron, B. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 2011.

Griffin, G., Holub, A., and Perona, P. Caltech-256 object category dataset. Technical Report CNS-TR-2007-001, California Institute of Technology, 2007.

Gu, X., Du, C., Pang, T., Li, C., Lin, M., and Wang, Y. On memorization in diffusion models. *Transactions on Machine Learning Research*, 2025.

Hamed Hassani, S., Macris, N., and Urbanke, R. Threshold saturation in spatially coupled constraint satisfaction problems. *Journal of Statistical Physics*, 150(5): 807–850, December 2012. ISSN 1572-9613. doi: 10.1007/s10955-012-0664-x. URL `http://dx.doi.org/10.1007/s10955-012-0664-x`.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, 2017.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

Hyvärinen, A. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4):695–709, 2005.

Karras, T., Aittala, M., Aila, T., and Laine, S. Elucidating the design space of diffusion-based generative models. In *NeurIPS*, 2022.

Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., and Aila, T. Improved precision and recall metric for assessing generative models. In *NeurIPS*, 2019.

Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., and Zhu, J. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. In *NeurIPS*, 2022.

Lugmayr, A., Danelljan, M., Timofte, R., and Van Gool, L. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022.

Miyasawa, K. An empirical bayes estimator of the mean of a normal distribution, 1961. Tech. report / early reference commonly cited for Tweedie's formula.

Naeem, M. F., Oh, S. J., Uh, Y., Choi, Y., and Yoo, J. Reliable fidelity and diversity metrics for generative models. In *ICML*, 2020.

Radford, A., Kim, J. W., Hallacy, C., and et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022.

RunwayML / Stability AI. Stable diffusion inpainting model card / checkpoint documentation, 2022. Model card as released with the Stable Diffusion inpainting checkpoint.

Saharia, C., Chan, W., Chang, H., Lee, C. A., Ho, J., Salimans, T., Fleet, D. J., and Norouzi, M. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH*, 2022.

Sajjadi, M. S. M., Bachem, O., Lucic, M., Bousquet, O., and Gelly, S. Assessing generative models via precision and recall. *arXiv preprint arXiv:1806.00035*, 2018.

Schuhmann, C. and Beaumont, R. Laion-aesthetics, 2022. Dataset release by LAION.

Schuhmann, C., Beaumont, R., Vencu, R., and et al. Laion-5b: An open large-scale dataset for training next generation image-text models. In *NeurIPS Datasets and Benchmarks Track*, 2022.

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML*, 2015.

Somepalli, G., Singla, V., Goldblum, M., Geiping, J., and Goldstein, T. Understanding and mitigating copying in diffusion models. In *NeurIPS*, 2023.

Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. In *ICLR*, 2021a.

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

Vincent, P. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

A. Proofs

Proof of Proposition 3.1. By Bayes' rule,

$$P(s \mid y(t)) \;\propto\; p(y(t) \mid s^0 = s)\, p(s), \tag{14}$$

where the proportionality is in s (treating the observation $y(t)$ as fixed). Under the stochastic interpolant Equation (3), conditional on the clean configuration the observations are independent Gaussians,

$$p(y(t) \mid s^0 = s) \;=\; \prod_{\ell=1}^{L}\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_\ell(t)} \exp\!\left(-\frac{\big(y_{\ell,i}(t) - \alpha_\ell(t)\, s_{\ell,i}\big)^2}{2\,\sigma_\ell(t)^2}\right). \tag{15}$$

Expanding the square and collecting s-independent terms into a single constant $C(y,t)$,

$$\log p(y(t) \mid s) \;=\; C(y,t) \;+\; \sum_{\ell,i}\left[\frac{\alpha_\ell(t)\, y_{\ell,i}(t)}{\sigma_\ell(t)^2}\, s_{\ell,i} \;-\; \frac{\alpha_\ell(t)^2}{2\,\sigma_\ell(t)^2}\, s_{\ell,i}^2\right]. \tag{16}$$

Using $\gamma_\ell(t) = \alpha_\ell(t)/\sigma_\ell(t)$ and substituting the observation model $y_{\ell,i}(t)/\sigma_\ell(t) = \gamma_\ell(t)\, s^0_{\ell,i} + \xi_{\ell,i}$ into the linear-in-s term in Equation (16) gives

$$\frac{\alpha_\ell(t)\, y_{\ell,i}(t)}{\sigma_\ell(t)^2}\, s_{\ell,i} \;=\; \gamma_\ell(t)\, \frac{y_{\ell,i}(t)}{\sigma_\ell(t)}\, s_{\ell,i} \;=\; \gamma_\ell(t)^2\, s^0_{\ell,i}\, s_{\ell,i} \;+\; \gamma_\ell(t)\, \xi_{\ell,i}\, s_{\ell,i}. \tag{17}$$

Combining with the prior $\log p(s) = -E(s) - \log Z$ (the inverse temperature $\beta = 1$ is already absorbed into (J, Δ, h) by the convention of Section 3; $E(s)$ in Equation (2) contains the field term with a leading $-$ sign, so $-E(s)$ contributes $+\sum_\ell h_\ell \sum_i s_{\ell,i}$ to the log-posterior), and absorbing the remaining s-independent normalisation into a single constant $\widetilde{C}(y,t)$,

$$\begin{aligned}\log P(s \mid y(t)) \;=\; \widetilde{C}(y,t) + \frac{1}{N} \sum_{\langle(\ell,i),(\ell',j)\rangle} J_{\ell\ell'}\, s_{\ell,i}\, s_{\ell',j} \;-\; \sum_{\ell,i}\big(\Delta + \tfrac{1}{2}\gamma_\ell(t)^2\big)\, s_{\ell,i}^2 \\ +\; \sum_{\ell,i}\big(h_\ell \;+\; \gamma_\ell(t)\, \xi_{\ell,i} \;+\; \gamma_\ell(t)^2\, s^0_{\ell,i}\big)\, s_{\ell,i}.\end{aligned} \tag{18}$$

Recognising the coefficient of $s_{\ell,i}^2$ as $\widetilde{\Delta}_\ell(t)$ and the coefficient of $s_{\ell,i}$ as $\widetilde{h}_{\ell,i}(t)$ in Equation (5), and exponentiating, yields Equation (6). □

Proof of Proposition 4.2. Fix a block $\ell \in \{1, \dots, L\}$ and a site index $i \in \{1, \dots, N\}$, and condition on $s^{(t)}$. Under the scheduled heat-bath rule at time t, the site (ℓ, i) is selected as the update target with probability $\pi_\ell^{(t)} \cdot (1/N)$ (block index drawn from $\pi^{(t)}$, then a site uniformly within the block). If (ℓ, i) is not selected, then $s_{\ell,i}^{(t+1)} = s_{\ell,i}^{(t)}$. If it is selected, then $s_{\ell,i}^{(t+1)}$ is resampled from the Gibbs conditional $p(\cdot \mid s^{(t)}_{\setminus(\ell,i)})$, while all other spins are unchanged. Taking expectations conditional on $s^{(t)}$,

$$\mathbb{E}\Big[s_{\ell,i}^{(t+1)} - s_{\ell,i}^{(t)} \;\Big|\; s^{(t)}\Big] \;=\; \frac{\pi_\ell^{(t)}}{N}\Big(\mathbb{E}_{p(\,\cdot\,\mid s^{(t)}_{\setminus(\ell,i)})}[s_{\ell,i}] \;-\; s_{\ell,i}^{(t)}\Big). \tag{19}$$

Summing Equation (19) over i and dividing by N,

$$\mathbb{E}\Big[m_\ell(s^{(t+1)}) - m_\ell(s^{(t)}) \;\Big|\; s^{(t)}\Big] \;=\; \frac{\pi_\ell^{(t)}}{N^2} \sum_{i=1}^{N}\Big(\mathbb{E}_{p(\,\cdot\,\mid s^{(t)}_{\setminus(\ell,i)})}[s_{\ell,i}] \;-\; s_{\ell,i}^{(t)}\Big). \tag{20}$$

Introducing $I \sim \mathrm{Unif}(\{1, \dots, N\})$ independent of $s^{(t)}$ rewrites the empirical average as an expectation, $\frac{1}{N}\sum_i f(i) = \mathbb{E}_I[f(I)]$, so that

$$\mathbb{E}\Big[m_\ell(s^{(t+1)}) - m_\ell(s^{(t)}) \;\Big|\; s^{(t)}\Big] \;=\; \frac{\pi_\ell^{(t)}}{N}\, \mathbb{E}_I\Big[\mathbb{E}_{p(\,\cdot\,\mid s^{(t)}_{\setminus(\ell,I)})}[s_{\ell,I}] \;-\; s_{\ell,I}^{(t)}\Big]. \tag{21}$$

Taking the outer expectation over $s^{(t)}$ and using the definition $m_\ell^{(t)} := \mathbb{E}[m_\ell(s^{(t)})]$ from Equation (9) yields Equation (10). The derivation is entirely path-wise and does not constrain the joint distribution of $s^{(t)}$, so no assumption on the initial law is required: the recursion tracks the unconditional mean of the block magnetization under any initialisation and any choice of schedule $\{\pi^{(t)}\}_{t \geq 0}$. □

Proof of Proposition 4.1. We prove the two parts separately.

Part (i). Let K be any Markov kernel that only rewrites s_v: for every s, $K(s, \cdot)$ is supported on configurations that agree with s on $s_{\setminus v}$. Write $K(s, s'_v) := K(s, (s'_v, s_{\setminus v}))$ for the induced conditional density on the updated coordinate. Assume K leaves q invariant. Summing the invariance equation $\sum_s q(s)K(s, s') = q(s')$ over only s_v (so $s'_{\setminus v} = s_{\setminus v}$ is held fixed) and dividing by the marginal $q(s_{\setminus v}) > 0$,

$$\sum_{s_v} \underbrace{\frac{q(s_v, s_{\setminus v})}{q(s_{\setminus v})}}_{= q(s_v \mid s_{\setminus v})} K((s_v, s_{\setminus v}),\, s'_v) \;=\; q(s'_v \mid s_{\setminus v}). \tag{22}$$

The left-hand side is the expectation of $K(\,\cdot\,, s'_v)$ under $q(\,\cdot \mid s_{\setminus v})$. The only way for this to equal $q(s'_v \mid s_{\setminus v})$ for every choice of s'_v and every $s_{\setminus v}$ is for the kernel not to depend on the input s_v at all and to output $s'_v \sim q(\,\cdot \mid s_{\setminus v})$, i.e. $K(s, s'_v) = q(s'_v \mid s_{\setminus v})$. This is the Gibbs (heat-bath) update and it is the unique q-invariant single-site kernel.

Part (ii). For any deterministic $f : s_{\setminus v} \mapsto \hat{s}_v$,

$$\mathbb{E}_q[(s_v - f(s_{\setminus v}))^2] \;=\; \mathbb{E}_q\left[\mathbb{E}_q[(s_v - f(s_{\setminus v}))^2 \mid s_{\setminus v}]\right], \tag{23}$$

and for each fixed $s_{\setminus v}$ the inner conditional expectation is minimised (by the elementary property of conditional mean as L^2-projection) at $f(s_{\setminus v}) = \mathbb{E}_q[s_v \mid s_{\setminus v}]$. Taking the outer expectation preserves the minimiser. □

B. Properties of the Posterior

The posterior of Proposition 3.1 is a random-field Blume–Capel model with a planted alignment term $\gamma_\ell^2(t)\, s^0_{\ell,i}\, s_{\ell,i}$ and an effective crystal field $\Delta + \frac{1}{2}\gamma_\ell^2(t)$ that stiffens with the diffusion schedule. Its single-site mean-field free energy $f(m)$ (single-block, magnetization m, planted spin s^0 drawn from the clean-prior marginal at (J, Δ, h)) organises the single-block equilibria. Figure 3 tracks the number and ranking of minima of $f(m)$ across (h, Δ) at twelve values of the per-block signal-to-noise ratio γ; the boundaries separate regions with (1) a single minimum, (2) two minima with the lower-m minimum lowest, (3) two minima with the higher-m one lowest, (4) three minima with the central one lowest, and (5) three minima with the high-m one lowest.

B.1. ℓ_2 Bias vs. Iteration

Figure 4 reports the ℓ_2 bias $\|m^{(T)} - m^\star\|_2$ of the same single-boundary run as in Figure 2, log–log in T, seed-averaged over six initial-condition perturbations with ±1 sd bands, and extended to $T = 4.2 \times 10^5$—20× the horizon of the main-text figure. Parallel and interleaved reach machine precision; progressive L→R and R→L both sit on plateaus (of order 1.0 and 1.4, respectively) with no visible relaxation throughout the horizon.

B.2. Per-Block Gibbs-Evolution Time Traces

Figure 5 shows the block-magnetization time traces $m_\ell(t)$ underlying Figure 2: one panel per unclamped block, one curve per schedule, seed-averaged with ±1 standard-deviation bands across 24 seeds. Parallel and interleaved schedules drive every block monotonically toward its Bayes-optimal level $m_\ell^\star$ (dotted line). Progressive L→R updates blocks in sequence: each block's trace is flat near zero until its phase begins (at $t = (\ell - 2)\, T_{\text{phase}}$ for block ℓ), jumps to an equilibrium value below $m_\ell^\star$, and stays there for the rest of the horizon. Progressive R→L does the same in reverse order but starting from the rightmost block, whose only neighbour is also unclamped; no block picks up a large magnetization, so the traces stay near zero throughout. The across-seed variance is small compared with the schedule gaps, so the effect is a property of the update order rather than of the initial perturbation.

B.3. Two-Boundary Ablation

The main-text experiment uses a single clamped boundary, matching the left-half mask of our empirical setup. If we instead clamp *both* chain ends to $+1$ and -1, the Bayes-optimal profile is antisymmetric and crosses zero in the middle, and the two progressive variants become mirror images of each other. Figure 6 shows this ablation with the same depth axes as Figure 2, now with $|\mathcal{H}| = 6$ and $T_{\text{phase}} = T/6 = 3500$. Parallel and interleaved recover the antisymmetric profile; progressive L→R

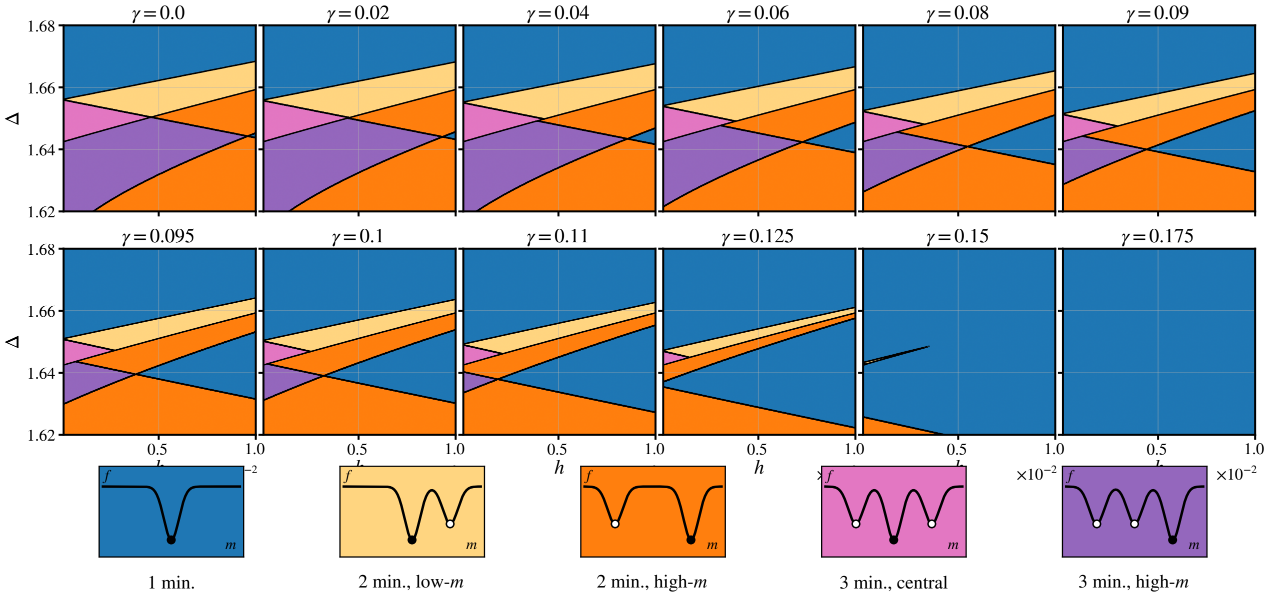

Figure 3. Phase diagram of the single-block posterior in the (h, Δ) plane as the per-block signal-to-noise ratio γ varies. Colour encodes the number of mean-field minima and which minimum is lowest; legend pictograms show the stylised $f(m)$ landscape for each class. Increasing γ stiffens the crystal field (effective $\Delta + \frac{1}{2}\gamma^2$) while the alignment term $\gamma^2 s^0$ biases the winning minimum toward the clean magnetization, reshaping the phase boundaries.

locks the chain into a positive basin (equilibrating each block against the already-resolved left side before the -1 side ever propagates to it); progressive R→L produces the mirror image. The qualitative lesson is the same as in the single-boundary case—finite-depth progressive updates sit in a metastable basin—but the failure is now symmetric in L→R vs. R→L: both progressive directions land on a plateau of comparable bias (≈ 0.84 at the horizon shown).

C. Experimental Setup

Model and inference. All experiments use the publicly available `runwayml/stable-diffusion-inpainting` checkpoint (Stable Diffusion v1.5, fine-tuned for inpainting) loaded via the HuggingFace `diffusers` library. Inference is performed in float16 precision with classifier-free guidance scale $w = 7.5$ and $T = 100$ DDPM denoising steps per patch. We use the default text prompt *"high quality, detailed, natural"* with negative prompt *"blurry, low quality, artifacts"* throughout; prompt embeddings are computed once and cached.

Masking and scheduling. Every test image is masked with a binary left-half mask (the right 50% of the image is unknown). We compare *parallel* inpainting ($K{=}1$ patch, entire mask denoised at once) against *progressive* schedules with $K \in \{1, 2, 3, 4, 6, 8, 10\}$ vertical stripe patches, filled sequentially from left to right. Each stripe is padded to a width divisible by 8 for VAE compatibility. To reduce boundary artifacts, a cosine ramp of width 4 pixels is applied at the diffusion-mask boundary of each stripe. The total number of network evaluations scales as $K \times T$; in this study we deliberately do *not* redistribute the step budget and instead hold T fixed per patch, so that improved quality (if any) comes purely from the spatial schedule rather than from additional compute.

Dataset. We evaluate on $5{,}000$ images from the Places365 validation set, center-cropped and resized to 512×512.

Metrics. Figure 1 reports three families of metrics, summarised below. We compute each metric on (a) the full 512×512 image, (b) the left (observed) half and the right (inpainted) half independently, and (c) where applicable, a *left-vs-right* variant that compares the two halves of the same image to quantify internal coherence across the inpainting seam.

Reconstruction metrics (higher is better, except LPIPS). *PSNR* (peak signal-to-noise ratio) is the log-ratio of the peak pixel intensity to the root-mean-squared error between the inpainted and ground-truth images, measured in decibels; it

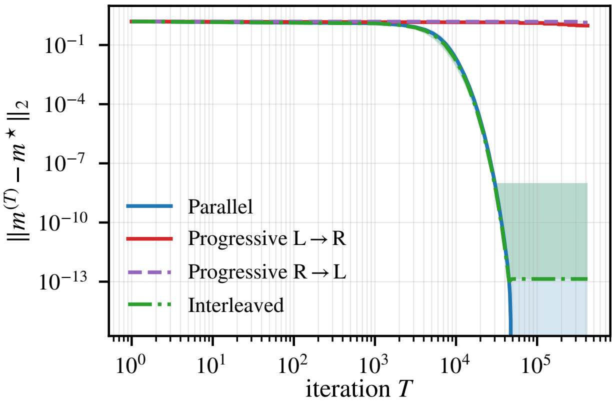

Figure 4. ℓ_2 bias $\|m^{(T)} - m^\star\|_2$ vs. iteration T (log–log) for the single-boundary chain, seed-averaged with ±1 sd bands. Parallel and interleaved converge to machine precision; both progressive variants remain on a finite plateau throughout $T = 4.2 \times 10^5$ iterations.

captures pixel-level fidelity but is insensitive to perceptually plausible but non-identical completions. *SSIM* (Wang et al., 2004) (structural similarity) compares local luminance, contrast, and structure statistics in sliding windows; it correlates better with perceived distortion than PSNR but is still sensitive to exact alignment.

Perceptual metrics. *LPIPS* (Zhang et al., 2018) (learned perceptual image patch similarity, lower is better) computes an ℓ_2 distance between deep-feature activations of a pretrained network, calibrated on human similarity judgments. We report three backbones (AlexNet, VGG, SqueezeNet) because the ranking between methods can depend on the feature extractor. *CLIP cosine similarity* (Radford et al., 2021) (higher is better) embeds both images with the ViT-B/32 CLIP image encoder and reports the cosine of the embedding angle; it captures high-level semantic agreement. The *left-vs-right* variant embeds each half separately and measures the semantic mismatch across the seam—independent of the ground truth—and is therefore the most direct measure of coherence under progressive scheduling.

Distributional metrics (dataset-level, computed on Inception-v3 features). *FID* (Heusel et al., 2017) (Fréchet Inception Distance, lower is better) fits a Gaussian to the Inception-v3 features of real and generated images and reports the Fréchet distance between the two Gaussians; it summarises the distance between the two feature distributions in a single scalar. *Precision* and *recall* (Kynkäänniemi et al., 2019) (higher is better) estimate, respectively, the fraction of generated samples that lie in the support of the real manifold (quality) and the fraction of the real manifold covered by generated samples (diversity), using k-nearest-neighbour manifold approximations. *Density* and *coverage* (Naeem et al., 2020) (higher is better) are robust refinements of precision and recall: density counts how many real-sample neighbourhoods contain a given generated sample (less sensitive to outliers than precision), and coverage counts the fraction of real samples whose neighbourhood contains at least one generated sample (more stable than recall under mode dropping).

Aggregation. Per-sample metrics are gathered across processes and we report means, standard errors, medians, and the 10th/90th percentiles. When multiple generations per image are drawn, we additionally decompose the total variance into intra-image (noise-induced) and inter-image components.

Infrastructure. Experiments are distributed across two NVIDIA RTX 6000 Ada GPUs using HuggingFace `accelerate`. Each process operates on a disjoint shard of the dataset; per-sample metrics and Inception features are gathered via lightweight object collection (no gradient synchronization). A single seed schedule—`base_seed + generation_idx` $\times\ 10^6$ `+ image_idx`—ensures reproducible yet distinct noise realizations across images and generation rounds. All results, per-image CSVs, and representative examples are saved under a run directory uniquely identified by the hyperparameter configuration.

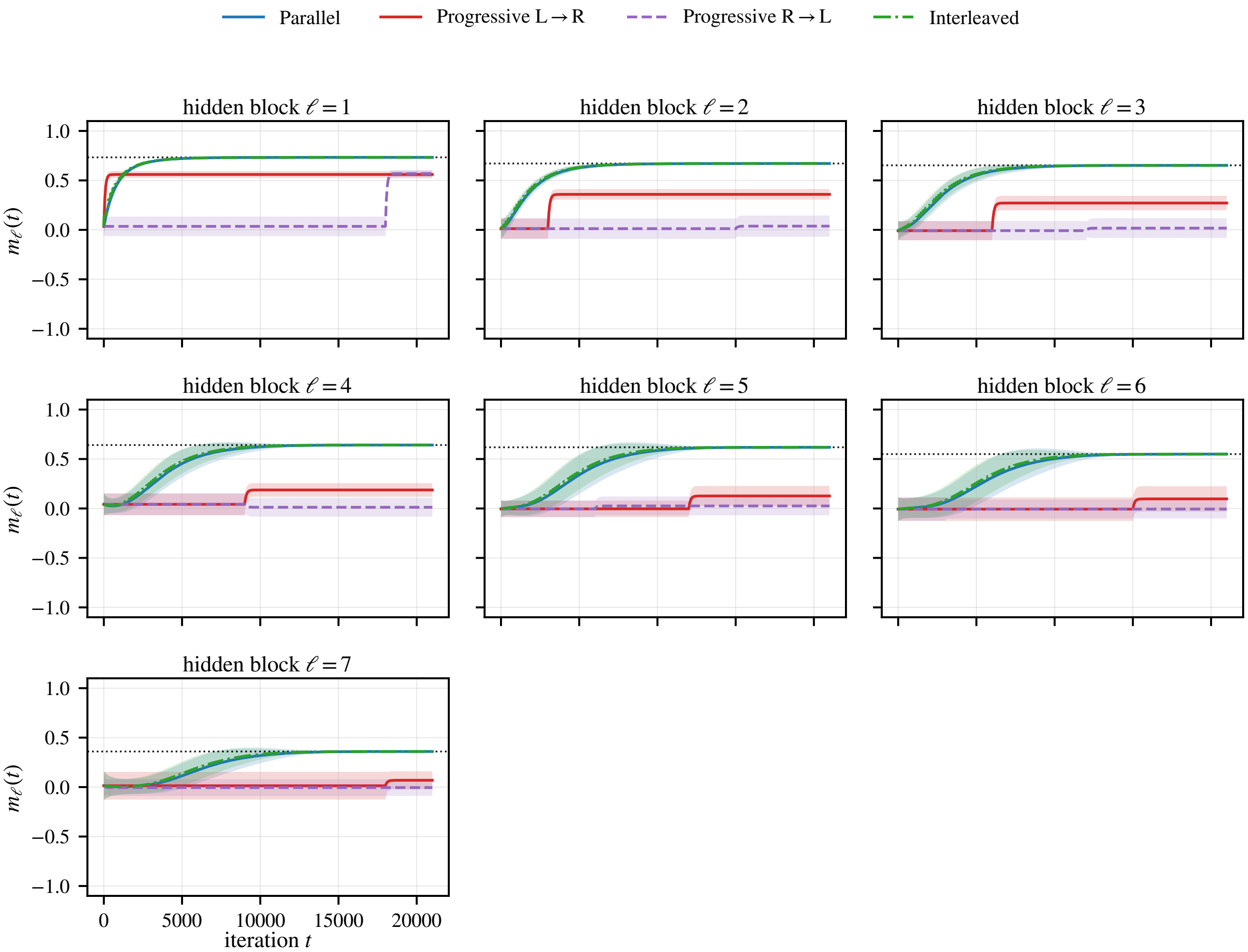

Figure 5. Per-block time traces $m_\ell(t)$ under each schedule, seed-averaged (mean $\pm$ 1 sd band across 24 seeds). Dotted line: Bayes-optimal $m^\star_\ell$. Parallel (blue) and interleaved (green) reach $m^\star$ on every block. Progressive L→R (red) updates one block per phase of length T_{phase}, and each block's equilibrium is below its Bayes level. Progressive R→L (purple) sweeps away from the clamped boundary and leaves all blocks near zero for the full horizon.

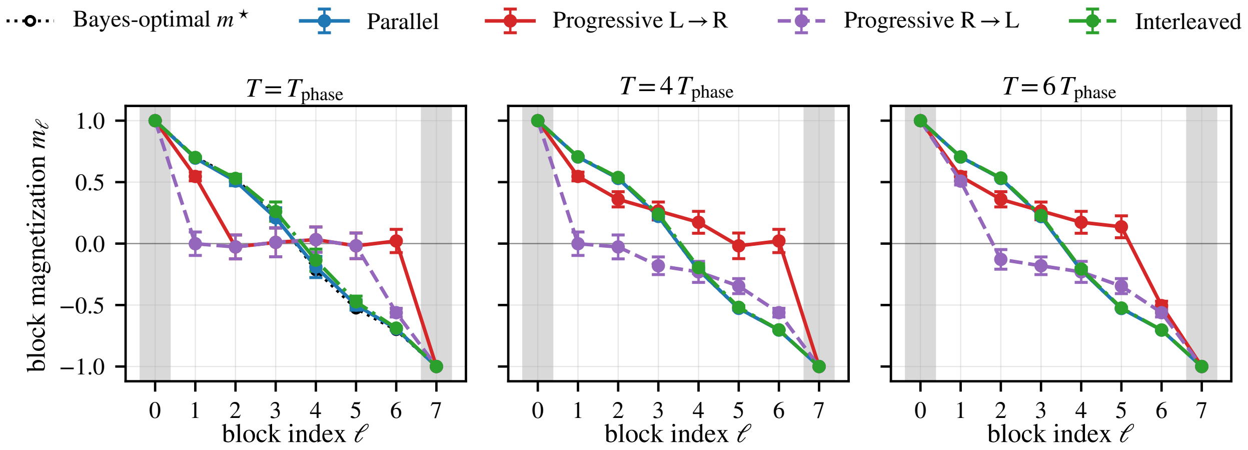

Figure 6. Two-boundary ablation: blocks 1 and L clamped to $+1$ and -1; $|\mathcal{H}| = 6$ and $T_{\text{phase}} = T/6$. Block magnetization m_ℓ with ± 1 sd error bars at three horizons (one, four, and six progressive phases), same colour code as Figure 2. Parallel and interleaved recover the antisymmetric Bayes target; progressive L→R locks the chain into the $+1$ basin; progressive R→L locks it into the -1 basin.