JOINT 3D GRAVITY AND MAGNETIC INVERSION VIA LATENT RECTIFIED-FLOW POSTERIOR SAMPLING

Anonymous authors
Paper under double-blind review

ABSTRACT

Joint gravity and magnetic inversion, which recovers 3D subsurface density and magnetic susceptibility from surface measurements, is among the cheapest and most widely used tools in mineral exploration, but it is severely non-unique: many subsurface models explain the same data. Classical inversions resolve this ambiguity with hand-crafted regularization, and learning-based inversions are trained on simplified synthetics. However, both return a single model. We instead sample from the posterior under a learned geological prior. We train a 3D variational autoencoder and a latent rectified-flow prior on paired density–susceptibility volumes from Noddyverse, a collection of one million simulated 3D geological models, and condition the prior on both fields with flow-based posterior sampling at the full 2×200^3 resolution. Evaluated against the 3D ground truth of held-out models from largely unseen geological histories, our method reduces susceptibility RMSE by 31–58% relative to twelve established inversions. The learned prior alone already outperforms hand-crafted regularization in susceptibility error. Our samples also counteract the shallow bias of classical inversion: on average, they place bodies within 100 m of their true depth, while the baselines place them 171–961 m too shallow.

1 INTRODUCTION

Gravity and magnetic field measurements are among the most widely used geophysical data for regional mapping and mineral exploration (Blakely, 1995; Hinze et al., 2013), and they are becoming more important as shallow mineral resources are depleted and exploration moves to deeper targets. Inverting these measurements into 3D subsurface property models is ill-posed and non-unique: many distinct subsurface models produce indistinguishable fields (Tarantola, 2005).

Classical methods recover a single regularized estimate by balancing data misfit with hand-crafted priors (Li & Oldenburg, 1996; 1998; Portniaguine & Zhdanov, 1999). Such an estimate hides the ambiguity that matters most in practice: a model can reproduce the surface data almost perfectly while placing the source at the wrong depth. Monte Carlo sampling could expose this ambiguity but is prohibitive at millions of unknowns (Mosegaard & Tarantola, 1995).

Learned inversions are fast but mostly return point estimates and are trained on simplified synthetic bodies (Huang et al., 2021; Zhou et al., 2024). Noddyverse (Jessell et al., 2022), one million simulated 3D geological models, makes it possible to learn a geologically realistic prior instead.

Diffusion and flow models have emerged as powerful generative models for high-dimensional data such as images. They learn to transform Gaussian noise into samples from the data distribution, either by reversing a gradual noising process (Ho et al., 2020; Song et al., 2021) or by following a learned velocity field (Liu et al., 2023). Combined with a likelihood, such a learned prior gives a posterior sampler that draws many data-consistent solutions of an inverse problem (Song et al., 2022; Chung et al., 2023; Kim et al., 2025), and training generative models in a compressed latent space makes them tractable at scale (Rombach et al., 2022). This leads to the following question: *Can a learned geological prior make posterior sampling practical for joint 3D gravity and magnetic inversion, and how do its samples compare with established inversions against the 3D ground truth?*

We answer this by training a latent generative prior over paired density–susceptibility volumes on Noddyverse and sampling its posterior given both fields with FlowDPS (Kim et al., 2025) (Figure 1).

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.

We also tried adding a physics-inspired prior at inference time: a Ginzburg–Landau phase-separation energy, evaluated on the decoded susceptibility, that favors compact ore bodies with sharp contacts. It lowered this energy in the samples but did not measurably improve the reconstructions, so we report it in Appendix E. We evaluate against ground truth on held-out models and against a broad suite of established deterministic inversions. Our main contributions are:

- **The first (to our knowledge) generative framework for joint 3D gravity–magnetic inversion.** Sampling in an 8×40^3 latent space lets us operate at full Noddyverse resolution ($2 \times 200^3 \approx 1.6 \times 10^7$ unknowns) and draw paired density–susceptibility posterior samples, four in 4.5–5.7 minutes on a single GPU.
- **A ground-truth benchmark on held-out geology.** We compare against twelve established inversions, including Tikhonov, IRLS focusing, total variation, and cross-gradient inversion, on models from geological histories largely unseen in training, using 3D metrics, paired tests, and depth and center-of-mass localization.
- **Learned versus hand-crafted priors.** Our method reduces susceptibility RMSE by 31–58% relative to every baseline and improves on each for at least 83% of held-out models. Much of this gain comes from the learned prior, which alone already outperforms hand-crafted regularization in susceptibility error, and our samples reduce the shallow bias of classical inversion. Conversely, the baselines fit the data far below the noise level but reconstruct the subsurface poorly, and cross-gradient coupling does not improve.
- **Open assets for the community.** We release a 3D VAE and a rectified-flow prior for Noddyverse density–susceptibility volumes, together with an efficient processing pipeline for the full dataset.[1]

2 Inverse Problems

Inverse problems aim to recover hidden quantities from indirect measurements. In the simplest setting, observations are generated by

$$y = \mathcal{A}(x) + n, \tag{1}$$

where x is the unknown model, y is the observed data, $\mathcal{A}$ is the forward operator, and n denotes measurement noise. The inverse problem attempts to infer x from y.

This inference is difficult because many inverse problems are ill-posed: a solution may fail to exist, may not be unique, or may depend unstably on small perturbations in the data (Tikhonov & Arsenin, 1977; Tarantola, 2005). Potential-field inversion is a canonical example. Gravity and magnetic fields measured at the surface are spatially smoothed responses of subsurface density and susceptibility, so high-frequency or deeply buried structures can be weakly constrained by the data. As a result, distinct subsurface models can produce nearly identical observations (Saltus & Blakely, 2011).

Classical inversion addresses this by regularization. For a linear forward model with Gaussian noise, a common formulation is

$$\hat{x} = \arg\min_x \frac{1}{2\sigma_y^2}\|y - Ax\|^2 + \lambda R(x), \tag{2}$$

where the first term enforces data consistency and $R(x)$ encodes prior assumptions such as smoothness, sparsity, compactness, or structural coherence. However, the result can be sensitive to the selected regularizer and hyperparameters, and it does not by itself characterize the range of plausible models consistent with the same data.

A probabilistic view instead treats x as a random variable and targets the posterior distribution

$$p(x \mid y) \propto p(y \mid x)p(x), \tag{3}$$

where $p(y \mid x)$ is determined by the forward model and noise statistics, and $p(x)$ represents prior knowledge about plausible subsurface structure. Sampling from this posterior is especially valuable in geophysical settings because uncertainty and ambiguity are often as important as a single recon-

[1] Code: `https://anonymous.4open.science/r/maggrav-joint-diffusion-D23E/`.

struction. The challenge is that realistic 3D models are high-dimensional, and traditional sampling methods can be computationally prohibitive at this scale (Mosegaard & Tarantola, 1995).

Gravity and magnetic forward model. In this work, the unknown model is the joint density–susceptibility volume $m = (\rho, \chi)$, and the observations are gravity and magnetic surface fields. The surface gravity field g_z, in mGal, responds to subsurface density ρ: a dense ore body produces a local positive anomaly. The magnetic field, observed as the total magnetic intensity (TMI) anomaly ΔT in nT, responds to magnetic susceptibility χ, which is high in rocks rich in minerals such as magnetite. Both fields are linear superpositions over the volume: each subsurface element contributes an inverse-square gravitational pull, or the field of a small dipole induced by Earth's field (we neglect remanent magnetization, as Noddyverse does).

On the voxel grid, this gives the joint linear model

$$\boldsymbol{y} = \boldsymbol{G}\boldsymbol{m} + \boldsymbol{n}, \qquad \boldsymbol{G} = \mathrm{diag}(\boldsymbol{G}_\rho, \boldsymbol{G}_\chi), \tag{4}$$

where $\boldsymbol{m} = [\boldsymbol{\rho}^\top, \boldsymbol{\chi}^\top]^\top \in \mathbb{R}^{2N}$ stacks density and susceptibility over N voxels, $\boldsymbol{y}$ stacks the M_g gravity and M_m magnetic measurements, and $\boldsymbol{n}$ is noise. $\boldsymbol{G}$ is block-diagonal because gravity responds only to density and magnetics only to susceptibility (Appendix C gives the governing integrals and our implementation).

Assumption 1 (Noise Model)**.** *The measurement noise is Gaussian,* $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma^2_{\mathrm{grav}} \boldsymbol{I}_{M_g}, \sigma^2_{\mathrm{mag}} \boldsymbol{I}_{M_m})$.

This is the standard model in potential-field inversion and recovers the weighted least-squares misfit of the classical methods we compare against (Li & Oldenburg, 1996; Tarantola, 2005).

Both responses decay with depth, so a weak shallow source and a strong deep source can produce nearly the same surface data. Minimum-norm inversions therefore concentrate structure near the surface (Li & Oldenburg, 1996), a bias that we quantify in Section 6.2. Our goal is to sample subsurface models $\boldsymbol{m}$ from the posterior $p(\boldsymbol{m} \mid \boldsymbol{y})$ under a learned prior over realistic geology.

3 Related Work

Potential-field inversion: Voxel-based 3D inversion with Tikhonov-style regularization and depth weighting is the standard approach for gravity and magnetics (Li & Oldenburg, 1996; 1998). Matrix-free Krylov solvers such as CGLS and LSQR (Hestenes & Stiefel, 1952; Paige & Saunders, 1982), with the regularization weight chosen by data-tolerance or L-curve criteria (Hansen, 1992), make such inversions routine at scale. Compactness, focusing, and minimum-gradient-support stabilizers solved by iteratively reweighted least squares (IRLS) (Last & Kubik, 1983; Portniaguine & Zhdanov, 1999; Fournier & Oldenburg, 2019) and total variation (Rudin et al., 1992) recover compact bodies with sharp boundaries.

Joint gravity–magnetic inversion: Joint inversion reduces non-uniqueness by exploiting the complementary sensitivities of the two fields. Early approaches used layered or low-dimensional parameterizations (Gallardo-Delgado et al., 2003; Pilkington, 2006). 3D joint inversion couples density and susceptibility through hand-crafted terms such as cross-gradient (Gallardo & Meju, 2003; Fregoso & Gallardo, 2009), Gramian (Zhdanov et al., 2012; Lin & Zhdanov, 2018; Vatankhah et al., 2023), or joint sparsity and vectorial total variation (Blomgren & Chan, 1998; Utsugi, 2025) constraints. Our prior instead learns this coupling from paired geological models.

Stochastic and uncertainty-aware methods: Probabilistic inversion characterizes uncertainty, for example by Monte Carlo sampling of gravity and magnetic posteriors (Mosegaard & Tarantola, 1995; Sambridge & Mosegaard, 2002; Bosch et al., 2006). To stay tractable, these methods rely on low-dimensional parameterizations, Gaussian approximations, or simple analytic priors.

Machine learning-based potential-field inversion: Supervised networks map potential-field data to subsurface models for gravity (Huang et al., 2021) and joint gravity and magnetic data (Bai et al., 2024), or to the structural parameters of Noddy models from magnetic data (Guo et al., 2021). They are typically trained on smaller grids and return point estimates without posterior uncertainty.

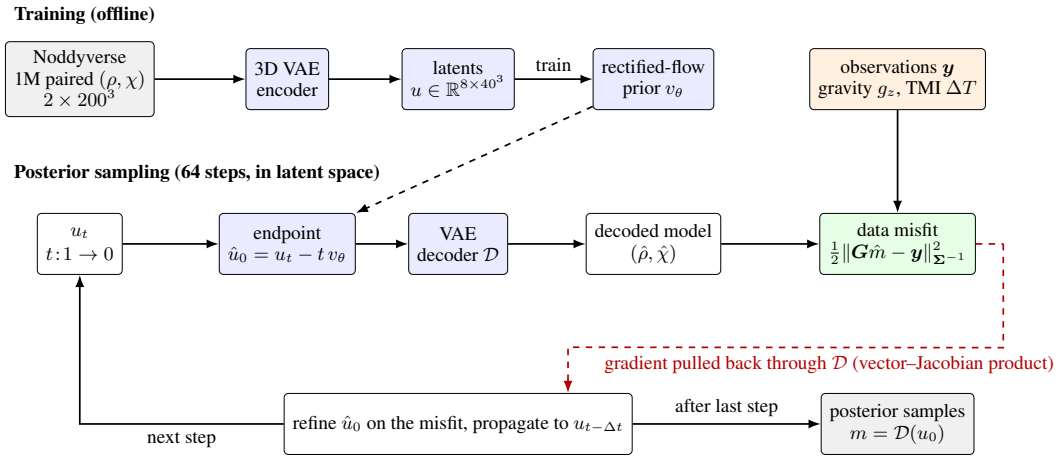

Figure 1: End-to-end pipeline. **Top:** a 3D VAE compresses paired density–susceptibility volumes from Noddyverse into a latent grid, on which a rectified-flow prior is trained. **Bottom:** at each FlowDPS step, the clean endpoint estimate is decoded to a full model, and the gradient of the joint data misfit is pulled back through the decoder to refine the latent.

Generative models for inverse problems: Posterior samplers built on diffusion and flow priors differ in how they enforce measurement consistency: diffusion posterior sampling (DPS) guides the reverse process with likelihood gradients (Chung et al., 2023), DDRM exploits the SVD of a linear forward operator (Kawar et al., 2022), and FlowDPS extends posterior sampling to rectified flows (Kim et al., 2025). We apply FlowDPS in the latent space of a 3D VAE.

4 Method

Figure 1 summarizes the pipeline: a 3D VAE (Section 4.1), a latent flow prior (Section 4.2), and posterior sampling in the latent space (Section 4.3).

4.1 Latent Representation: 3D VAE

A 3D VAE maps paired density–susceptibility volumes to a latent grid, using residual convolutional down- and upsampling with attention at the bottleneck. Based on sweeps over latent size, capacity, and KL weight β (Appendix F), we use an 8×40^3 latent grid with $\beta = 10^{-3}$ (614,091 parameters).

4.2 Latent Generative Prior: Rectified Flow

We compared two 3D latent velocity networks: a timestep-conditioned 3D UNet (Çiçek et al., 2016) with residual blocks, FiLM time conditioning, skip connections, and low-resolution self-attention, and BiFlowNet (Wang et al., 2025), which merges features from patch-level DiT blocks into the corresponding blocks of a 3D UNet. BiFlowNet was more parameter-efficient and optimized faster. We also compared EDM diffusion (Karras et al., 2022) with rectified flow (Liu et al., 2023) and use the latter, which converged in fewer epochs.

Rectified flow interpolates linearly between a clean latent u_0 (the VAE encoding of a training volume) and Gaussian noise $u_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, $u_t = (1-t)u_0 + tu_1$ with $t \sim \mathcal{U}(0,1)$, and trains $v_\theta(u_t, t)$ to predict the velocity $u_1 - u_0$. Our final prior is a 20M-parameter BiFlowNet (Appendix G).

4.3 Posterior Sampling in Latent Space

Sampling runs in the VAE latent space. With $\mathcal{D}$ the decoder, the posterior over latents combines the joint likelihood with the learned prior:

$$p(u \mid \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \mathcal{D}(u))\, p_\theta(u). \tag{5}$$

The weighted misfit $\mathcal{L}_{\text{data}}(u) = \frac{1}{2}\|\boldsymbol{G}\mathcal{D}(u) - \boldsymbol{y}\|^2_{\boldsymbol{\Sigma}^{-1}}$ is the negative log-likelihood under Assumption 1.

FlowDPS (Kim et al., 2025) enforces data consistency at every step: it forms the clean endpoint estimate $\hat{u}_0 = u_t - t\, v_\theta(u_t, t)$, refines it with a few gradient steps on $\mathcal{L}_{\text{data}}$, and propagates the result to the next time step. The misfit is evaluated on the decoded volume, and its gradient is a vector–Jacobian product through $\mathcal{D}$. Algorithm 1 and all sampler settings are in Appendix D.

A hand-crafted physics prior. We also tested a physics-based prior for the two-phase structure of mineral systems, added to the refinement objective at sampling time without retraining the flow. We map the decoded susceptibility to a continuous phase field $\phi \in [-1, 1]$, with $\phi = -1$ for host rock and $\phi = +1$ for ore, and penalize its Ginzburg–Landau energy, the continuum relaxation of the Ising model. Its double-well potential $(\phi^2 - 1)^2/4\varepsilon^2$ drives each voxel toward a pure phase, and its gradient term $\frac{\kappa}{2}|\nabla\phi|^2$ penalizes interfaces, so the energy favors compact bodies with sharp contacts. Adding it to $\mathcal{L}_{\text{data}}$ more than halved the energy of the samples but did not measurably improve the held-out reconstructions, and it raised the magnetic misfit (Appendix E).

5 Experimental Setup

5.1 Dataset and Preprocessing

We use Noddyverse (Jessell et al., 2022), a synthetic benchmark of one million 3D geological models generated by kinematic simulation of varied deformation histories and petrophysical properties, including compact intrusive bodies with coherent susceptibility contrasts that serve as controlled ore-like targets. Each sample contains a 3D subsurface block of size $200 \times 200 \times 200$ voxels, where each voxel stores density ($\boldsymbol{\rho}$) and magnetic susceptibility ($\boldsymbol{\chi}$). The corresponding observations are two forward-simulated 2D surface fields on a 200×200 grid: the gravity anomaly ($\boldsymbol{g}_{\text{grav}}$) and the total magnetic intensity anomaly ($\boldsymbol{g}_{\text{mag}}$).

Direct training on the raw files is difficult because (i) the target dimensionality is very large, which makes inversion highly underdetermined and memory-intensive, and (ii) magnetic susceptibility is strongly heavy-tailed, which can destabilize optimization. We therefore model susceptibility in the log domain, $\chi_{\log} = \log_{10}(\chi + \delta)$ with $\delta = 10^{-6}$, and map back with $\chi = 10^{\chi_{\log}} - \delta$ before applying the magnetic forward operator.

Held-out split. Noddyverse is ordered by its 343 deformation-history classes, and we split it contiguously: the VAE is trained on the first 980,019 models (the next 5,000 are its validation set) and the flow prior on the first 950,019. We evaluate on 100 models drawn at random from the last 15,001, which neither model saw during training. They span 12 history classes, all involving an unconformity, and 11 of these classes do not occur in the training range, so the evaluation also tests generalization to unseen geological histories. All hyperparameters of the sampler and of the baselines were tuned on observations from the training range and then frozen.

5.2 Baselines

We compare against twelve established deterministic inversions that are tractable at this scale, all with depth weighting (Li & Oldenburg, 1996) and the same forward operator:

- **CGLS, LSQR, and Tikhonov-CG** (Hestenes & Stiefel, 1952; Paige & Saunders, 1982): matrix-free Krylov solvers for the unregularized and Tikhonov-regularized least-squares problems.
- **Projected CGLS**: nonnegative susceptibility by projected FISTA (Beck & Teboulle, 2009a), with density by CGLS.
- **IRLS focusing** (Last & Kubik, 1983; Portniaguine & Zhdanov, 1999; Fournier & Oldenburg, 2019): four inversions with sparse ℓ_p/ℓ_q model norms, $(p, q) \in \{(0.1, 1), (1, 1), (1, 2), (2, 1)\}$, in the spirit of classical compact and focusing inversion.
- **Isotropic TV** (Rudin et al., 1992) and **joint vectorial TV** (Blomgren & Chan, 1998; Bresson & Chan, 2008): edge-preserving regularization, applied per property or with

a shared gradient norm that couples density and susceptibility, solved with Chambolle–Pock (Chambolle & Pock, 2011).
- **Cross-gradient joint inversion** (Gallardo & Meju, 2003; Fregoso & Gallardo, 2009): the structural coupling penalty $\lambda_{\mathrm{cg}}\|\nabla\rho \times \nabla\chi\|_2^2$, which aligns density and susceptibility interfaces. We use $\lambda_{\mathrm{cg}} = 1000$ and report an otherwise identical control with $\lambda_{\mathrm{cg}} = 0$ to isolate the effect of explicit coupling.

Regularization weights were tuned once on ten training-range observations with criteria that do not use the ground truth, a data-tolerance rule or the L-curve corner (Hansen, 1992), and then frozen. Susceptibility is clipped to $[0, \chi_{0.999}]$ after each solve, where $\chi_{0.999} = 1.09$ is the 99.9th percentile of rock-unit susceptibilities in Noddyverse. Each baseline returns a single deterministic reconstruction. We do not include supervised neural inversions, which are designed for smaller grids and return point estimates. No sampling-based probabilistic method is tractable at this scale (Section 7).

5.3 Evaluation Metrics

All methods are evaluated on the same 100 held-out observations. FlowDPS draws four posterior samples per observation, and each baseline returns one reconstruction. We score the posterior mean of the four samples against the 3D ground truth, with density in g/cc and susceptibility in the $\log_{10}(\chi + 10^{-6})$ domain:

- **Voxel RMSE** between the reconstruction and the ground truth.
- **CRPS**: the fair ensemble continuous ranked probability score (Gneiting & Raftery, 2007; Ferro, 2014) of the four samples, averaged over voxels. It rewards both accuracy and appropriate spread. For a single reconstruction we use the standard CRPS, which reduces to the mean absolute error.
- **Rank-matched Dice** (Dice, 1945): the ground-truth body is the most anomalous 15% of voxels of each property. We label the same number of voxels in the reconstruction, breaking ties at random, and measure the overlap. This scores localization independently of amplitude.
- **Body localization**: the distance between the centers of mass of the matched bodies and their depth offset, negative when the reconstruction is too shallow. Dice and localization use volumes block-averaged by a factor of 2.
- **Data misfit**: the RMSE between forward-simulated and observed fields, ΔT in nT and g_z in mGal, averaged over the four samples, reported as the median over observations.

Paired comparisons report the mean per-observation difference with a 95% t interval and the number of observations on which each method is better, without correction for multiple comparisons.

Implementation details. For all posterior-sampling experiments we use 64 steps, $K_{\mathrm{ref}} = 16$ refinement steps with step size $\alpha = 0.1$, $\omega_t = 1$, $\sigma_{\mathrm{mag}} = 15$ nT, and $\sigma_{\mathrm{grav}} = 0.1$ mGal, without re-noising (Appendix D). Appendix D.1 ablates these settings. The magnetic response depends on exponentiated susceptibility, which makes trajectories highly sensitive to perturbations, so we clip the per-sample norm of the refinement gradient at $c = 50$. The initial noise is seeded per observation, so FlowDPS and the prior-only sampler start from the same noise. The networks run in BF16 and the loss and forward operator in FP32. The VAEs were trained on eight NVIDIA RTX 6000 Ada GPUs (48 GB) and the flow prior on two 3g.90gb MIG instances of an NVIDIA B200. Sampling ran on four RTX 6000 Ada GPUs and on one B200 MIG instance.

6 Results

6.1 Comparison with Established Inversions

Table 1 compares FlowDPS with the deterministic baselines and with samples from its own prior, drawn from the same initial noise without data.

A learned prior outperforms hand-crafted regularization. FlowDPS has the lowest susceptibility RMSE and CRPS of all methods, 31–58% and 52–72% below the baselines, and improves on every baseline in at least 83 of 100 observations. It also has the lowest density CRPS, better than every

Table 1: Reconstruction quality on held-out Noddyverse models. FlowDPS RMSE and Dice are scored on the posterior mean of four samples, and "prior only" runs the same sampler without data. Depth is the depth offset of the matched body (negative: too shallow). Voxel metrics and depth are means, and misfits are medians over observations. Susceptibility is in $\log_{10}(\chi + 10^{-6})$ and density in g/cc. Best per column in bold. Appendix A lists all twelve baselines. $^{\dagger}0.025$ nT before clipping susceptibility to its physical range.

	Susceptibility χ				Density ρ				Misfit ↓	
Method	RMSE ↓	CRPS ↓	Dice ↑	Depth [m]	RMSE ↓	CRPS ↓	Dice ↑	Depth [m]	ΔT [nT]	g_z [mGal]
FlowDPS (ours)	**0.897**	**0.535**	0.302	**+36**	0.153	**0.092**	0.328	**−98**	15.6	0.264
Prior only	0.943	0.570	0.180	+350	0.208	0.126	0.172	+258	419	17.7
CGLS / LSQR	2.141	1.877	0.350	−548	0.148	0.118	**0.370**	−961	124†	**0.0002**
Projected CGLS	1.348	1.105	0.348	−394	0.148	0.118	**0.370**	−960	0.021	0.0005
IRLS ($p = 2$, $q = 1$)	1.692	1.352	0.333	−309	**0.147**	0.118	0.369	−950	0.22	0.0013
Isotropic TV	1.693	1.342	0.330	−239	**0.147**	0.117	0.364	−677	8.12	0.065
Joint vectorial TV	1.651	1.320	**0.365**	−396	**0.147**	0.117	0.363	−687	5.26	0.063
Cross-gradient	1.320	1.147	0.304	−179	0.148	0.119	0.353	−850	0.019	0.0009
Cross-gradient control	1.309	1.135	0.301	−171	0.149	0.119	0.355	−875	**0.010**	0.0009

baseline in at least 97 observations. Surprisingly, much of the susceptibility gain comes from the learned prior alone: sampled without any data ("prior only" in Table 1), it already beats every baseline on RMSE and CRPS (0.943 and 0.570). This highlights the importance of a plausible prior: one learned from realistic geology regularizes susceptibility better than smoothness, sparsity, or structural penalties.

Body overlap. In susceptibility Dice, the posterior mean is on par with cross-gradient inversion (differences within ± 0.002). It trails the best baseline by 0.063 for susceptibility (joint TV) and 0.042 for density (CGLS). Individual samples do better: the best of the four reaches Dice 0.41 (χ) and 0.44 (ρ), above every baseline, so the posterior contains solutions that localize the bodies better than any of the deterministic inversions.

Qualitative comparison. Figure 2 shows four held-out models with deep or dipping bodies. The classical inversions place the anomalous body near the surface, even when the true body is deep (first and third rows) or dips to depth (second and fourth rows). FlowDPS places the bodies at depth and follows their layered and dipping shapes. The samples differ where the data are ambiguous, and the posterior mean averages them into smoother, partly fragmented bodies.

6.2 Depth of Recovered Bodies

Potential-field kernels decay with depth, so inversions tend to concentrate structure near the surface, and depth weighting is the standard remedy (Li & Oldenburg, 1996). Even with depth weighting, all twelve baselines place bodies too shallow on average (Tables 1 and 5). For density, the offset is 677–961 m for every baseline except Tikhonov-CG, whose density reconstruction is anomalous (RMSE 0.296, twice that of the other baselines). FlowDPS reduces it to 98 m, and its density bodies are closer to the true ones (center-of-mass distance 1290 m vs. 1415–1457 m). For susceptibility, the baselines are 171–548 m too shallow, while FlowDPS is 36 m too deep.

Figure 3 shows the mean depth profiles. The baselines concentrate susceptibility near the surface, while the FlowDPS profiles stay closer to the ground truth, apart from excess density in the top few hundred meters.

6.3 Computational Cost

Table 2 reports wall-clock times. Four FlowDPS samples take 5.7 minutes per observation (4.5 minutes on a B200 MIG instance). Refinement dominates the cost: sampling from the prior alone takes 22 s. The Krylov and IRLS baselines take 4–12 s per observation, and the TV and cross-gradient

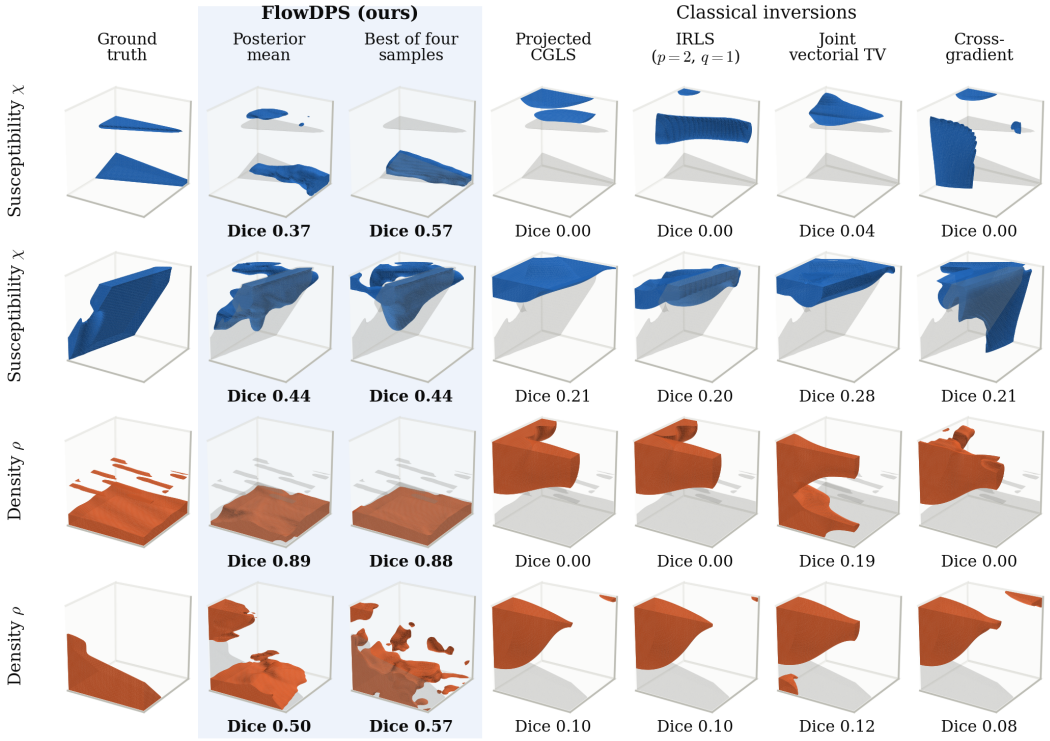

Figure 2: On deep bodies, FlowDPS does better than the classical inversions, which place the bodies near the surface. We illustrate this on four held-out models. Section 6.2 gives the statistics over all models. Each panel shows the 15% most anomalous voxels of a $4 \times 4 \times 4$ km volume, with depth increasing downward and the true body in gray. The best of four samples is selected using the ground truth. Values are rank-matched Dice.

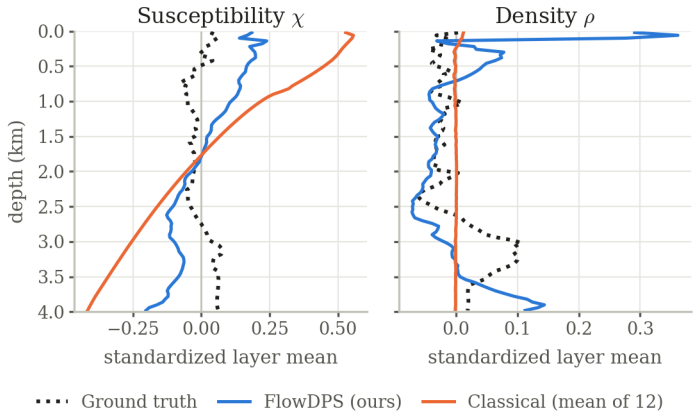

Figure 3: Mean depth profiles of 100 held-out models: horizontal layer means, standardized per model and averaged over models. “Classical” is the mean of the twelve baselines.

solvers 2.5–5 minutes. Peak memory for one FlowDPS sample is about 3.3 GB. Asymptotic costs are discussed in Appendix B.

Table 2: Median wall-clock time per held-out observation on an NVIDIA RTX 6000 Ada. FlowDPS times cover four samples. TV and cross-gradient solves were batched over up to 9 observations, and their time is the batch time divided by the batch size.

	FlowDPS	Prior only	CGLS / LSQR / Tikhonov-CG	IRLS	Projected CGLS	TV	Joint TV	Cross-gradient	Cross-grad. control
Time [s]	340	22	3.8–3.9	7.4–7.5	12.4	289	287	169	155

7 LIMITATIONS

Synthetic, noise-free evaluation. We evaluate on Noddyverse because it provides paired 3D ground truth at scale, which field data cannot. Performance on real measurements, with topography, irregular observation grids, correlated noise, and forward-model mismatch, remains to be established.

Induced magnetization. Our magnetic forward model assumes induced magnetization. The learned prior never evaluates the forward model, so remanence would misspecify only the inference-time likelihood. If the total-magnetization direction is known and spatially uniform, the kernel can simply be recomputed for that direction, and the inferred quantity becomes an effective magnetization amplitude. Unknown or spatially varying remanence would require vector-magnetization variables and a revised prior.

Posterior calibration. An exact reference posterior is unavailable at this dimensionality, and we do not compare against a sampling-based probabilistic inversion. Generic voxel-space MCMC is impractical, and a chain in the flow's latent coordinates (for example with a prior-preserving proposal (Cotter et al., 2013)) would require a flow transformation, decoding, and forward evaluation per proposal, with untested mixing. With four samples per observation we can score the posterior with CRPS, and the best of four samples overlaps the true body better than any baseline, but we cannot assess whether the spread of the samples matches the true posterior uncertainty.

Decoder backpropagation. Each sampling step backpropagates through the 3D decoder. This is practical at the present resolution, but its cost grows with resolution. A latent surrogate would need to approximate the gradient of the misfit through the decoder.

8 CONCLUSION

We introduced a flow-based framework for joint 3D gravity and magnetic inversion that samples paired density and susceptibility volumes at full Noddyverse resolution from a posterior defined by a learned geological prior and both fields. On held-out Noddyverse models from largely unseen geological histories, our method reduces susceptibility RMSE by 31–58% relative to twelve established inversions and counteracts their shallow bias. Much of this gain comes from the learned prior, which shows that a prior learned from realistic geology can regularize potential-field inversion better than hand-crafted penalties. The benchmark also shows that near-exact data fit does not imply an accurate subsurface. A hand-crafted physics prior, a Ginzburg–Landau phase-separation energy, did not measurably help (Appendix E), so priors learned from data matter all the more. FlowDPS is only one way to condition such a prior on the data. We also release the first VAE and rectified-flow prior for Noddyverse. Promising next steps include:

- **Better posterior samplers.** Score-based data assimilation (Rozet & Louppe, 2023) and the family of sequential Monte Carlo samplers (Cardoso et al., 2024; Dou & Song, 2024), such as the twisted diffusion sampler (Wu et al., 2023), show promise for improving on FlowDPS.
- **Scaling the VAE and flow prior.** Our sweeps show that larger latents and more parameters substantially reduce reconstruction and flow-matching losses.
- **Toward field data.** Single-field ablations, noisy and misspecified observations, and remanent magnetization.

All code is publicly available at
`https://anonymous.4open.science/r/maggrav-joint-diffusion-D23E/`.

REFERENCES

Samuel M. Allen and John W. Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metallurgica*, 27(6):1085–1095, 1979. doi: 10.1016/0001-6160(79)90196-2.

Zhijing Bai, Yanfei Wang, Chenzhang Wang, Caixia Yu, Dmitry Lukyanenko, Inna Stepanova, and Anatoly G. Yagola. Joint gravity and magnetic inversion using CNNs' deep learning. *Remote Sensing*, 16(7):1115, 2024. doi: 10.3390/rs16071115.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009a. doi: 10.1137/080716542.

Amir Beck and Marc Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009b. doi: 10.1109/TIP.2009.2028250.

Richard J. Blakely. *Potential Theory in Gravity and Magnetic Applications*. Cambridge University Press, Cambridge, 1995. doi: 10.1017/CBO9780511549816.

P. Blomgren and T. F. Chan. Color TV: Total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, 1998. doi: 10.1109/83.661180.

Miguel Bosch, Ronny Meza, Rosa Jiménez, and Alfredo Hönig. Joint gravity and magnetic inversion in 3D using Monte Carlo methods. *Geophysics*, 71(4):G153–G156, 2006. doi: 10.1190/1.2209952.

Xavier Bresson and Tony F. Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Problems and Imaging*, 2(4):455–484, 2008. doi: 10.3934/ipi.2008.2.455.

Gabriel Cardoso, Yazid Janati El Idrissi, Sylvain Le Corff, and Eric Moulines. Monte Carlo guided denoising diffusion models for Bayesian linear inverse problems. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=nHESwXvxWK.

Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. doi: 10.1007/s10851-010-0251-1.

Tony F. Chan and Luminita A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001. doi: 10.1109/83.902291.

Hyungjin Chung, Jeongsol Kim, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, volume 9901 of *Lecture Notes in Computer Science*, pp. 424–432. Springer, 2016. doi: 10.1007/978-3-319-46723-8_49.

S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science*, 28(3):424–446, 2013. doi: 10.1214/13-STS421.

Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3): 297–302, 1945. doi: 10.2307/1932409.

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=tplXNcHZs1.

Christopher A. T. Ferro. Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1917–1923, 2014. doi: 10.1002/qj.2270.

Dominique Fournier and Douglas W. Oldenburg. Inversion using spatially variable mixed ℓ_p norms. *Geophysical Journal International*, 218(1):268–282, 2019. doi: 10.1093/gji/ggz156.

Emilia Fregoso and Luis A. Gallardo. Cross-gradients joint 3D inversion with applications to gravity and magnetic data. *Geophysics*, 74(4):L31–L42, 2009. doi: 10.1190/1.3119263.

Luis A. Gallardo and Max A. Meju. Characterization of heterogeneous near-surface materials by joint 2D inversion of dc resistivity and seismic data. *Geophysical Research Letters*, 30(13):1658, 2003. doi: 10.1029/2003GL017370.

Luis A. Gallardo-Delgado, Marco Antonio Pérez-Flores, and Enrique Gómez-Treviño. A versatile algorithm for joint 3D inversion of gravity and magnetic data. *Geophysics*, 68(3):949–959, 2003. doi: 10.1190/1.1581067.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

Jiateng Guo, Yunqiang Li, Mark Walter Jessell, Jérémie Giraud, Chaoling Li, Lixin Wu, Fengdan Li, and Shanjun Liu. 3D geological structure inversion from Noddy-generated magnetic data using deep learning methods. *Computers & Geosciences*, 149:104701, 2021. doi: 10.1016/j.cageo.2021.104701.

Per Christian Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34(4):561–580, 1992. doi: 10.1137/1034115.

M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, 1952. doi: 10.6028/jres.049.044.

William J. Hinze, Ralph R. B. von Frese, and Afif H. Saad. *Gravity and Magnetic Exploration: Principles, Practices, and Applications*. Cambridge University Press, Cambridge, 2013. doi: 10.1017/CBO9780511843129.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Rui Huang, Shuang Liu, Rui Qi, and Yujie Zhang. Deep learning 3D sparse inversion of gravity data. *Journal of Geophysical Research: Solid Earth*, 126(11):e2021JB022476, 2021. doi: 10.1029/2021JB022476.

Mark Jessell, Jiateng Guo, Yunqiang Li, Mark Lindsay, Richard Scalzo, Jérémie Giraud, Guillaume Pirot, Ed Cripps, and Vitaliy Ogarko. Into the Noddyverse: A massive data store of 3D geological models for machine learning and inversion applications. *Earth System Science Data*, 14(1):381–392, 2022. doi: 10.5194/essd-14-381-2022.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 26565–26577. Curran Associates, Inc., 2022. doi: 10.52202/068431-1926.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 23593–23606. Curran Associates, Inc., 2022. doi: 10.52202/068431-1714.

Jeongsol Kim, Bryan Sangwoo Kim, and Jong Chul Ye. FlowDPS: Flow-driven posterior sampling for inverse problems. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12328–12337. IEEE, 2025. doi: 10.1109/ICCV51701.2025.01146.

B. J. Last and K. Kubik. Compact gravity inversion. *Geophysics*, 48(6):713–721, 1983. doi: 10.1190/1.1441501.

Yaoguo Li and Douglas W. Oldenburg. 3-D inversion of magnetic data. *Geophysics*, 61(2):394–408, 1996. doi: 10.1190/1.1443968.

Yaoguo Li and Douglas W. Oldenburg. 3-D inversion of gravity data. *Geophysics*, 63(1):109–119, 1998. doi: 10.1190/1.1444302.

Wei Lin and Michael S. Zhdanov. Joint multinary inversion of gravity and magnetic data using Gramian constraints. *Geophysical Journal International*, 215(3):1540–1557, 2018. doi: 10.1093/gji/ggy351.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Luciano Modica. The gradient theory of phase transitions and the minimal interface criterion. *Archive for Rational Mechanics and Analysis*, 98(2):123–142, 1987. doi: 10.1007/BF00251230.

Luciano Modica and Stefano Mortola. Un esempio di Γ^--convergenza. *Bollettino della Unione Matematica Italiana B (5)*, 14:285–299, 1977.

Klaus Mosegaard and Albert Tarantola. Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth*, 100(B7):12431–12447, 1995. doi: 10.1029/94JB03097.

Christopher C. Paige and Michael A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, 8(1):43–71, 1982. doi: 10.1145/355984.355989.

Mark Pilkington. Joint inversion of gravity and magnetic data for two-layer models. *Geophysics*, 71 (3):L35–L42, 2006. doi: 10.1190/1.2194514.

Oleg Portniaguine and Michael S. Zhdanov. Focusing geophysical inversion images. *Geophysics*, 64 (3):874–887, 1999. doi: 10.1190/1.1444596.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10674–10685. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01042.

François Rozet and Gilles Louppe. Score-based data assimilation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 40521–40541. Curran Associates, Inc., 2023. doi: 10.52202/075280-1763.

Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1–4):259–268, 1992. doi: 10.1016/0167-2789(92)90242-F.

Richard W. Saltus and Richard J. Blakely. Unique geologic insights from "non-unique" gravity and magnetic interpretation. *GSA Today*, 21(12):4–10, 2011. doi: 10.1130/G136A.1.

Malcolm Sambridge and Klaus Mosegaard. Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40(3):1009, 2002. doi: 10.1029/2000RG000089.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/forum?id=vaRCHVj0uGI.

Albert Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Philadelphia, 2005. doi: 10.1137/1.9780898717921.

Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of Ill-Posed Problems*. Scripta Series in Mathematics. V. H. Winston & Sons, Washington, D.C., 1977.

Mitsuru Utsugi. Joint inversion of magnetic and gravity data using group lasso regularization. *Earth, Planets and Space*, 77:146, 2025. doi: 10.1186/s40623-025-02270-1.

Saeed Vatankhah, Xingguo Huang, Rosemary A. Renaut, Kevin Mickus, and Jun Lin. Joint Gramian inversion of gravity, magnetic and gravity gradient tensor data for the modelling of Mesoproterozoic iron oxide deposits, southeast Missouri, USA. *Geophysical Journal International*, 235(2):1064–1085, 2023. doi: 10.1093/gji/ggad290.

Haoshen Wang, Zhentao Liu, Kaicong Sun, Xiaodong Wang, Dinggang Shen, and Zhiming Cui. 3D MedDiffusion: A 3D medical latent diffusion model for controllable and high-quality medical image generation. *IEEE Transactions on Medical Imaging*, 44(12):4960–4972, 2025. doi: 10.1109/TMI.2025.3585372.

Luhuan Wu, Brian L. Trippe, Christian A. Naesseth, David M. Blei, and John P. Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 31372–31403. Curran Associates, Inc., 2023. doi: 10.52202/075280-1363.

Michael S. Zhdanov, Alexander Gribenko, and Glenn Wilson. Generalized joint inversion of multimodal geophysical data using Gramian constraints. *Geophysical Research Letters*, 39(9):L09301, 2012. doi: 10.1029/2012GL051233.

Shuai Zhou, Yue Wei, Pengyu Lu, Guangrui Yu, Shuqi Wang, Jian Jiao, Ping Yu, and Jianwei Zhao. A deep learning gravity inversion method based on a self-constrained network and its application. *Remote Sensing*, 16(6):995, 2024. doi: 10.3390/rs16060995.

A Additional Baseline Results

Settings. Table 3 lists the frozen settings of the twelve baselines. All use depth weighting (Li & Oldenburg, 1996) with exponent 1 and floor 0.01, and the forward operator of Appendix C. Susceptibility is clipped to $[0, \chi_{0.999}]$ after each solve, which raises the CGLS, LSQR, and Tikhonov-CG magnetic misfit from 0.025 to 124 nT. Weights were tuned once on ten training-range observations with criteria that do not use the ground truth. Tikhonov-CG gives the same susceptibility as CGLS at this weight. Its density RMSE of 0.296 is twice that of the other solvers.

Table 3: Frozen settings and selection criteria of the baselines.

Method	Settings	Selection
CGLS, LSQR	100 iterations	–
Tikhonov-CG	$\lambda = 10^{-6}\sigma_{\max}^2$, 100 CG iterations	default
Projected CGLS	$\chi \geq 0$ by monotone FISTA (Beck & Teboulle, 2009b) (300 iterations), ρ by CGLS	–
IRLS (p, q)	$\lambda = 10^{-7}$ for $(0.1, 1)$, 10^{-6} for $(1, 1)$ and $(1, 2)$, 10^{-5} for $(2, 1)$, with 10 outer $\times$ 15 CG iterations	largest λ with misfit at most $2\times$ that at $\lambda = 10^{-8}$
TV, joint TV	$\lambda = 3.16 \times 10^{-3}$, 4,000 Chambolle–Pock iterations (Chambolle & Pock, 2011)	L-curve corner (Hansen, 1992)
Cross-gradient	$\lambda_{\mathrm{cg}} = 1000$, 3,000 iterations	smaller of the largest λ_{cg} with misfit at most $1.5\times$ that at $\lambda_{\mathrm{cg}} = 0$ and the L-curve corner
Cross-gradient control	$\lambda_{\mathrm{cg}} = 0$, otherwise identical	–

Full results. Table 4 extends Table 1 to all twelve baselines, Table 5 reports body localization, and Table 6 gives paired differences between FlowDPS and each reference.

B Computational Complexity of Inversion Methods

Table 7 summarizes the asymptotic cost of common inversion approaches for n model parameters and m observations. Here C_G and $C_{G^\top}$ denote the cost of applying the forward operator and its adjoint,

Table 4: Reconstruction quality of all methods on 100 held-out observations (columns as in Table 1).

	Susceptibility χ				Density ρ				Misfit ↓	
Method	RMSE ↓	CRPS ↓	Dice ↑	Depth [m]	RMSE ↓	CRPS ↓	Dice ↑	Depth [m]	ΔT [nT]	g_z [mGal]
FlowDPS (ours)	0.897	0.535	0.302	+36	0.153	0.092	0.328	−98	15.6	0.264
Prior only	0.943	0.570	0.180	+350	0.208	0.126	0.172	+258	419	17.7
CGLS	2.141	1.877	0.350	−548	0.148	0.118	0.370	−961	124	0.0002
LSQR	2.141	1.877	0.350	−548	0.148	0.118	0.370	−961	124	0.0002
Tikhonov-CG	2.141	1.877	0.350	−548	0.296	0.234	0.255	−258	124	0.0054
Projected CGLS	1.348	1.105	0.348	−394	0.148	0.118	0.370	−960	0.021	0.0005
IRLS ($p = 0.1, q = 1$)	1.720	1.376	0.334	−333	0.149	0.119	0.361	−913	0.79	0.0020
IRLS ($p = 1, q = 1$)	1.714	1.371	0.335	−327	0.148	0.118	0.363	−912	0.39	0.0010
IRLS ($p = 1, q = 2$)	1.717	1.374	0.335	−330	0.148	0.118	0.363	−911	0.38	0.0010
IRLS ($p = 2, q = 1$)	1.692	1.352	0.333	−309	0.147	0.118	0.369	−950	0.22	0.0013
Isotropic TV	1.693	1.342	0.330	−239	0.147	0.117	0.364	−677	8.12	0.065
Joint vectorial TV	1.651	1.320	0.365	−396	0.147	0.117	0.363	−687	5.26	0.063
Cross-gradient	1.320	1.147	0.304	−179	0.148	0.119	0.353	−850	0.019	0.0009
Cross-gradient control	1.309	1.135	0.301	−171	0.149	0.119	0.355	−875	0.010	0.0009

Table 5: Body localization on 100 held-out observations. Dice is that of the posterior mean, followed for the samplers by the best of the four samples, which requires the ground truth to select. CoM is the center-of-mass distance between the matched bodies, and depth the signed depth offset (negative: too shallow).

	Susceptibility χ			Density ρ		
Method	Dice	CoM [m]	Depth [m]	Dice	CoM [m]	Depth [m]
FlowDPS (ours)	0.302 / 0.408	1204	+36	0.328 / 0.436	1290	−98
Prior only	0.180 / 0.309	1582	+350	0.172 / 0.327	1635	+258
CGLS / LSQR	0.350	1232	−548	0.370	1457	−961
Tikhonov-CG	0.350	1232	−548	0.255	1151	−258
Projected CGLS	0.348	1404	−394	0.370	1457	−960
IRLS ($p = 0.1, q = 1$)	0.334	1196	−333	0.361	1455	−913
IRLS ($p = 1, q = 1$)	0.335	1203	−327	0.363	1443	−912
IRLS ($p = 1, q = 2$)	0.335	1201	−330	0.363	1443	−911
IRLS ($p = 2, q = 1$)	0.333	1220	−309	0.369	1452	−950
Isotropic TV	0.330	1219	−239	0.364	1429	−677
Joint vectorial TV	0.365	1287	−396	0.363	1434	−687
Cross-gradient	0.304	1218	−179	0.353	1415	−850
Cross-gradient control	0.301	1220	−171	0.355	1432	−875

C_L, C_D, and C_{cg} the cost of applying the regularization, difference, and cross-gradient operators, k the number of Krylov iterations, K and K_{out} outer iterations, N_λ the number of regularization weights tried, and S the number of MCMC samples. For our method, N is the number of sampling steps, K_{ref} the number of refinement steps, and C_v, $C_{\mathcal{D}}$, and $C_{\mathcal{D}^\top}$ the cost of the flow network, the decoder, and its vector–Jacobian product.

At our scale ($n = 2 \times 200^3 \approx 1.6 \times 10^7$), the explicit pseudoinverse and direct Hessian methods are infeasible, but all matrix-free methods are tractable, and we evaluate representatives of each in Section 6. MCMC is cheap per step but would require a very large number S of forward evaluations to mix in n dimensions. Our method is matrix-free and linear in n per step, with a fixed number of steps. Table 2 reports its measured cost.

C Forward Model Details

Governing integrals. Let $\boldsymbol{x} = (x, y, z)$ be an observation point, with z positive upward, and let $\boldsymbol{x}'$ be a source point in the subsurface volume V. A density contrast $\rho(\boldsymbol{x}')$ generates the gravitational potential $U(\boldsymbol{x}) = \gamma \int_V \rho(\boldsymbol{x}') \, |\boldsymbol{x} - \boldsymbol{x}'|^{-1} \, d^3\boldsymbol{x}'$, where γ is the gravitational constant, and the gravity

Table 6: FlowDPS minus each reference on 100 held-out observations: mean paired difference and, in parentheses, the number of observations on which FlowDPS is better. Lower is better for RMSE and CRPS, higher for Dice. *: the 95% t interval excludes zero (no multiple-comparison correction).

	Susceptibility χ			Density ρ		
Reference	RMSE	CRPS	Dice	RMSE	CRPS	Dice
Prior only	-0.046^* (64)	-0.035^* (60)	$+0.123^*$ (73)	-0.055^* (96)	-0.034^* (75)	$+0.156^*$ (72)
CGLS / LSQR	-1.244^* (96)	-1.342^* (98)	-0.048^* (44)	$+0.005^*$ (33)	-0.026^* (97)	-0.042^* (42)
Tikhonov-CG	-1.244^* (96)	-1.342^* (98)	-0.048^* (44)	-0.143^* (100)	-0.142^* (100)	$+0.073^*$ (67)
Projected CGLS	-0.451^* (83)	-0.571^* (92)	-0.046 (56)	$+0.005^*$ (33)	-0.026^* (97)	-0.042^* (42)
IRLS (0.1, 1)	-0.823^* (95)	-0.842^* (98)	-0.032 (48)	$+0.004^*$ (38)	-0.027^* (97)	-0.033 (40)
IRLS (1, 1)	-0.817^* (95)	-0.836^* (98)	-0.033 (46)	$+0.005^*$ (30)	-0.026^* (97)	-0.035 (41)
IRLS (1, 2)	-0.819^* (96)	-0.839^* (98)	-0.033 (46)	$+0.005^*$ (31)	-0.026^* (97)	-0.035 (41)
IRLS (2, 1)	-0.795^* (94)	-0.818^* (97)	-0.031 (47)	$+0.006^*$ (31)	-0.026^* (97)	-0.042^* (44)
Isotropic TV	-0.795^* (94)	-0.807^* (98)	-0.028 (45)	$+0.006^*$ (28)	-0.025^* (98)	-0.036^* (42)
Joint vectorial TV	-0.754^* (93)	-0.786^* (97)	-0.063^* (45)	$+0.006^*$ (27)	-0.025^* (98)	-0.036 (42)
Cross-gradient	-0.423^* (88)	-0.612^* (94)	-0.002 (52)	$+0.004^*$ (33)	-0.027^* (97)	-0.026 (45)
Cross-grad. control	-0.411^* (88)	-0.601^* (94)	$+0.002$ (54)	$+0.004^*$ (34)	-0.027^* (97)	-0.028 (46)

Table 7: Asymptotic memory and time per inversion.

Method	**Memory**	**Time per inversion**
Pseudoinverse / SVD	$\mathcal{O}(mn)$ to $\mathcal{O}(n^2)$	$\mathcal{O}(\min(mn^2,\, m^2n))$
Direct Tikhonov	$\mathcal{O}(n^2)$	$\mathcal{O}(n^3)$
CGLS / LSQR	$\mathcal{O}(n+m)$, matrix-free	$\mathcal{O}\big(k(C_G + C_{G^\top})\big)$
Tikhonov–CG	$\mathcal{O}(n+m)$, matrix-free	$\mathcal{O}\big(k(C_G + C_{G^\top} + C_L)\big)$
Tikhonov sweep	same as CG	$\mathcal{O}\big(N_\lambda\, k(C_G + C_{G^\top})\big)$
IRLS / focusing	as CG, plus weights	$\mathcal{O}\big(K_{\text{out}}\, k(C_G + C_{G^\top})\big)$
TV / ADMM / primal–dual	$\mathcal{O}(n+m)$	$\mathcal{O}\big(K(C_G + C_{G^\top} + C_D)\big)$
Cross-gradient	$\mathcal{O}(2n+m)$	$\mathcal{O}\big(K_{\text{out}}\, k(C_G + C_{G^\top} + C_{\text{cg}})\big)$
MCMC	$\mathcal{O}(n)$ per chain	$\mathcal{O}(S\, C_G)$
Ours (FlowDPS)	$\mathcal{O}(n+m)$ + weights	$\mathcal{O}\big(N[C_v + (K_{\text{ref}}+1)(C_{\mathcal{D}} + C_{\mathcal{D}^\top}) + K_{\text{ref}}(C_G + C_{G^\top})]\big)$

anomaly is its downward derivative,

$$g_z(\boldsymbol{x}) = -\frac{\partial U}{\partial z} = \gamma \int_V \rho(\boldsymbol{x}')\, \frac{z - z'}{|\boldsymbol{x} - \boldsymbol{x}'|^3}\, d^3\boldsymbol{x}'. \tag{6}$$

In Earth's field $\boldsymbol{H}_0 = H_0 \hat{\boldsymbol{h}}_0$, a susceptibility $\chi(\boldsymbol{x}')$ acquires the induced magnetization $\boldsymbol{M}(\boldsymbol{x}') = \chi(\boldsymbol{x}')\boldsymbol{H}_0$. Like the Noddyverse generator, we neglect remanent magnetization (Section 7). The TMI anomaly is the anomalous field projected onto $\hat{\boldsymbol{h}}_0$,

$$\Delta T(\boldsymbol{x}) = -\frac{\mu_0}{4\pi}\, \hat{\boldsymbol{h}}_0 \cdot \nabla \int_V \boldsymbol{M}(\boldsymbol{x}') \cdot \nabla' \frac{1}{|\boldsymbol{x} - \boldsymbol{x}'|}\, d^3\boldsymbol{x}', \tag{7}$$

where μ_0 is the vacuum permeability and ∇' acts on $\boldsymbol{x}'$.

Spectral evaluation. Noddy computes both fields with a spectral method (Jessell et al., 2022), and all methods in this paper use our differentiable PyTorch reimplementation of it. The volume is a stack of $D = 200$ horizontal layers of cubic cells with side $\Delta = 20\,\text{m}$, observed on the 200×200 cell grid at a sensor altitude of $100\,\text{m}$. Below, lengths are measured in cells. For layer $\ell = 1, \ldots, D$, counted from the top, let $\hat{\rho}_\ell(\boldsymbol{f})$ and $\hat{\chi}_\ell(\boldsymbol{f})$ be the 2D discrete Fourier transforms of its density (in $\text{kg}\,\text{m}^{-3}$) and susceptibility, where $\boldsymbol{f}$ is the horizontal spatial frequency in cycles per cell, and let $d_\ell = (100\,\text{m})/\Delta + \ell - \frac{1}{2}$ be the distance from the sensor plane to the layer center. For the vertical inducing field of Noddyverse (inclination $90°$, declination $0°$, intensity $F = 50{,}000\,\text{nT}$), each layer

contributes

$$\hat{g}_\ell(\boldsymbol{f}) = 2\pi\,\gamma\,\Delta\,\hat{\rho}_\ell(\boldsymbol{f})\,e^{-2\pi|\boldsymbol{f}|d_\ell},$$
$$\widehat{\Delta T}_\ell(\boldsymbol{f}) = 4\pi^2 F\,|\boldsymbol{f}|\,\hat{\chi}_\ell(\boldsymbol{f})\,e^{-2\pi|\boldsymbol{f}|d_\ell}.$$

For other field directions, the magnetic kernel takes the standard projected form, which the implementation also supports. The exponential is the upward-continuation factor: it damps fine-scale structure more strongly the deeper the layer. The extra factor $|\boldsymbol{f}|$ makes the magnetic response emphasize shallower and sharper features than gravity. The layer spectra are summed and inverted once, $g_z = 10^5\,\mathcal{F}^{-1}\big[\sum_\ell \hat{g}_\ell\big]$ in mGal and $\Delta T = \mathcal{F}^{-1}\big[\sum_\ell \widehat{\Delta T}_\ell\big]$ in nT, and cropped to the observation grid. Before the transform, each layer is extended to twice its width and length by mirror reflection, as in Noddy, to suppress wrap-around from the periodic transform, and the continuation factor is floored at e^{-12}. Remanent magnetization and demagnetization are neglected. Applied to the ground-truth volumes, the implementation reproduces the released Noddyverse fields to within 0.02 nT and 10^{-4} mGal RMSE. One evaluation costs $\mathcal{O}(D\,n\log n)$ for the $n = 400^2$ padded grid, and gradients follow by automatic differentiation.

D FlowDPS Update Details

Algorithm 1 gives the full sampler. Following FlowDPS (Kim et al., 2025), each step forms the linear-flow endpoint estimates $\hat{u}_0$ and $\hat{u}_1$, refines $\hat{u}_0$ by K_{ref} gradient steps on the refinement objective $\mathcal{L}_t$, blends the result with $\hat{u}_0$ using the FlowDPS weight $\omega_t \in [0,1]$, and propagates to the next time $t' < t$. FlowDPS also mixes fresh Gaussian noise into $\hat{u}_1$. We disable this, so each sample is a deterministic function of its initial noise. Without GL guidance, $\mathcal{L}_t = \mathcal{L}_{\text{data}}$, whose gradient is a vector–Jacobian product through the decoder,

$$\nabla_u \mathcal{L}_{\text{data}}(u) = J_{\mathcal{D}}(u)^\top \boldsymbol{G}^\top \boldsymbol{\Sigma}^{-1}\big(\boldsymbol{G}\mathcal{D}(u) - \boldsymbol{y}\big),$$

which we compute with one decode and one backward pass. We clip its per-sample norm at c for stability. Appendix E.2 adds the GL term.

Settings. All held-out results use $S = 64$ steps, $K_{\text{ref}} = 16$ refinement steps with constant step size $\alpha = 0.1$, $\omega_t = 1$, $c = 50$, $\sigma_{\text{mag}} = 15$ nT, and $\sigma_{\text{grav}} = 0.1$ mGal. Appendix D.1 describes how these values were chosen, on observations from the training range (Section 5.1).

Algorithm 1 Latent FlowDPS. The objective $\mathcal{L}_t$ is $\mathcal{L}_{\text{data}}$, or Eq. (11) with GL guidance.

Require: observations $\boldsymbol{y}$, flow prior v_θ, decoder $\mathcal{D}$, steps S, refinement steps K_{ref} and step size α, blend weight ω_t, clip threshold c

$u \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ ▷ $t = 1$
for $k = S, \dots, 1$ **do**
  $t \leftarrow k/S, \quad t' \leftarrow (k-1)/S$
  $v \leftarrow v_\theta(u, t); \quad \hat{u}_0 \leftarrow u - t\,v; \quad \hat{u}_1 \leftarrow u + (1-t)\,v$ ▷ endpoint estimates
  $u_0^{\text{ref}} \leftarrow \hat{u}_0$
  for $j = 1, \dots, K_{\text{ref}}$ **do**
    $u_0^{\text{ref}} \leftarrow u_0^{\text{ref}} - \alpha\,\text{clip}\big(\nabla_u \mathcal{L}_t(u_0^{\text{ref}}),\, c\big)$ ▷ Section 4.3
  end for
  $\tilde{u}_0 \leftarrow (1-\omega_t)\,\hat{u}_0 + \omega_t\,u_0^{\text{ref}}$ ▷ $\omega_t = 1$ in our experiments
  $u \leftarrow (1-t')\,\tilde{u}_0 + t'\,\hat{u}_1$ ▷ propagate to t', no re-noising
end for
return $\boldsymbol{m} = \mathcal{D}(u)$

D.1 Sampler Ablation

Our initial sampler (V0 in Table 8) averaged the misfit over the 8×10^4 map values, used the FlowDPS weight $\omega_t = t$, which vanishes at the clean end, and took $K_{\text{ref}} = 8$ refinement steps. Its samples were almost identical to prior samples drawn from the same initial noise, with voxel correlation 0.96 for susceptibility and 0.98 for density. Averaging divides the log-likelihood by 8×10^4, which is

equivalent to inflating the noise variance by the same factor. Table 8 changes one setting at a time on ten tuning observations from the training range.

Table 8: Sampler ablation on ten training-range observations with four samples each. Values are medians. "Twin r" is the voxel correlation with the prior-only sample drawn from the same initial noise, and "truth r" the correlation with the ground truth. Time is the median sampling time per observation (four samples, RTX 6000 Ada). A no-data surface slab reaches Dice 0.029 (χ) and 0.142 (ρ) on these observations.

		Misfit					
Run	Change	ΔT [nT]	g_z [mGal]	Twin r χ/ρ	Truth r χ/ρ	Dice χ/ρ	Time [s]
Prior only	$K_{\mathrm{ref}} = 0$	1272	15.5	–	$+0.009\,/\,-0.035$	$0.107\,/\,0.211$	17
V0	mean misfit, ω_t=t, K_{ref}=8	551	2.08	$0.957\,/\,0.982$	$+0.025\,/\,+0.036$	$0.187\,/\,0.202$	–
V1	V0 + summed misfit	166	0.43	$0.766\,/\,0.850$	$+0.055\,/\,+0.022$	$0.165\,/\,0.207$	178
V2	V1 + $\omega_t = 1$	107	0.37	$0.730\,/\,0.833$	$+0.064\,/\,+0.026$	$0.171\,/\,0.206$	181
V3	V2 + $\alpha = 0.3$	141	0.73	$0.661\,/\,0.768$	$+0.073\,/\,+0.045$	$0.194\,/\,0.203$	179
V4 (final)	V2 + $K_{\mathrm{ref}} = 16$	59	0.32	$0.693\,/\,0.794$	$+0.066\,/\,+0.037$	$0.172\,/\,0.206$	335
V5	V2 + $K_{\mathrm{ref}} = 32$	50	0.25	$0.66\,/\,0.75$	$+0.068\,/\,+0.046$	$0.182\,/\,0.205$	666

Summing the misfit is the largest single change: the median magnetic misfit falls from 551 to 166 nT and the twin correlation from 0.96 to 0.77. Setting $\omega_t = 1$ and doubling K_{ref} lower the misfit further, to 59 nT. A larger step size raises it, and $K_{\mathrm{ref}} = 32$ lowers it only to 50 nT at twice the cost, so we use V4. The misfit on these observations is bimodal: the median over four samples is 2–7 nT for five of them and 150–2369 nT for the other five, which have high-amplitude fields. Dice changes little between the guided variants on these ten observations, and none exceeds the prior-only density Dice.

E Ginzburg–Landau Guidance

Mineral systems often contain compact ore bodies with sharp contacts against the host rock. We tested whether a physics-inspired prior that encodes this two-phase structure improves posterior sampling when it is added to the learned prior at inference time. Appendix E.1 derives the prior, Appendix E.2 adds it to FlowDPS, Appendix E.3 reports its effect on held-out data, and Appendix E.4 describes how its hyperparameters were chosen. The remaining subsections give supporting derivations.

E.1 From the Ising Model to Ginzburg–Landau Energy

We build this prior from the physics of phase separation.

The Ising model is the simplest description of two-phase structure: each voxel i carries a label $s_i \in \{-1, +1\}$, host or ore, with energy $H_{\mathrm{Ising}} = -J \sum_{\langle i,j \rangle} s_i s_j$ over pairs of face-adjacent voxels, where $J > 0$ rewards neighbors for sharing a phase. Discrete labels are not differentiable, so we relax them to a continuous phase field $\phi \in [-1, 1]$, with $\phi = -1$ for host rock and $\phi = +1$ for ore:

$$\phi(\chi) = \mathrm{clip}\left(2\,\frac{\chi - \chi^{\mathrm{host}}}{\chi^{\mathrm{ore}} - \chi^{\mathrm{host}}} - 1,\ -1,\ 1 \right), \tag{8}$$

where χ^{host} and χ^{ore} are the susceptibilities of pure host and pure ore. In practice, we apply this map to $\log_{10} \chi$, since susceptibility spans several orders of magnitude (Appendix E.5).

The Ginzburg–Landau (GL) energy of ϕ on the volume V is

$$\mathcal{E}_{\mathrm{GL}}[\phi] = \int_V \left[\frac{\kappa}{2} |\nabla \phi|^2 + W(\phi) \right] d\boldsymbol{x}, \qquad W(\phi) = \frac{1}{4\varepsilon^2} (\phi^2 - 1)^2. \tag{9}$$

The first term, weighted by $\kappa > 0$, charges for interfaces between phases. The double well W vanishes at $\phi = \pm 1$ and pushes each voxel toward pure host or pure ore, and $\varepsilon > 0$ sets the width of the transition between them. As $\varepsilon \to 0$, the GL energy Γ-converges to a perimeter-penalized two-phase model, the continuum analogue of the Ising energy (Modica–Mortola, Appendix E.8). On a voxel grid with spacing h, the interface term becomes a sum of squared differences between

neighbors (Appendix E.7), so a voxel that disagrees with all six neighbors costs an extra $12\kappa/h^2$, and the energy favors coherent ore bodies.

We turn the energy into a prior on susceptibility,

$$p_{\mathrm{GL}}(\boldsymbol{\chi}) \propto \exp\big(-\lambda_{\mathrm{GL}}\,\mathcal{E}_{\mathrm{GL}}[\phi(\boldsymbol{\chi})]\big), \tag{10}$$

with strength $\lambda_{\mathrm{GL}} > 0$, which Appendix E.2 combines with the learned prior. The same energy induces Allen–Cahn phase-separation dynamics (Appendix E.11).

E.2 GL-Guided FlowDPS

With $\mathcal{D}_\chi$ the susceptibility channel of the decoder, adding the GL prior (10) to the latent posterior (5) gives

$$p(u \mid \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \mathcal{D}(u))\, p_\theta(u)\, p_{\mathrm{GL}}(\mathcal{D}_\chi(u)).$$

We add the GL energy of the decoded susceptibility to the FlowDPS refinement objective:

$$\mathcal{L}_t(u) = \mathcal{L}_{\mathrm{data}}(u) + \lambda_t\,\mathcal{E}_{\mathrm{GL}}\big[\phi(\mathcal{D}_\chi(u))\big], \qquad \lambda_t = \lambda_0(1-t)^\nu. \tag{11}$$

Refinement therefore minimizes the negative log of the likelihood times p_{GL}. The energy is evaluated on the decoded volume, since latent coordinates do not preserve the geometry or phase meaning of susceptibility. Both terms share one decode and one backward pass through the decoder,

$$\nabla_u \mathcal{L}_t(u) = J_{\mathcal{D}}(u)^\top \boldsymbol{G}^\top \boldsymbol{\Sigma}^{-1}\big(\boldsymbol{G}\mathcal{D}(u) - \boldsymbol{y}\big) + \lambda_t\, J_{\mathcal{D}_\chi}(u)^\top \nabla_{\boldsymbol{\chi}} \mathcal{E}_{\mathrm{GL}}[\phi(\boldsymbol{\chi})]\big|_{\boldsymbol{\chi}=\mathcal{D}_\chi(u)},$$

and the per-sample clip of Algorithm 1 applies to this total gradient. The weight λ_t rises from 0 at $t = 1$ to λ_0 at $t = 0$, so early high-noise steps explore freely and GL sharpens structure once it has formed. Setting $\lambda_0 = 0$ recovers unguided FlowDPS exactly, and GL needs no retraining.

We use $\kappa = 1$, $h = 1$, $\varepsilon = 0.5$ voxels, $\nu = 1$, and $\lambda_0 = 10^6$, with per-sample anchors (Appendix E.5). The GL energy is a per-voxel mean (Appendix E.7), whereas $\mathcal{L}_{\mathrm{data}}$ sums 8×10^4 squared residuals, so λ_0 is large in absolute terms. All other settings are those of Appendix D. GL adds 4–7% to the sampling time (364 vs. 340 s per observation on an RTX 6000 Ada, and 280 vs. 268 s on a B200 MIG instance).

E.3 Effect on Held-Out Data

Table 9: GL-guided minus unguided FlowDPS on the 78 held-out observations with both runs, which share initial noise. Entries are mean paired differences, with 95% intervals for Dice, and the number of observations on which GL is better.

Metric	Susceptibility χ	Density ρ
RMSE ↓	-0.003 (39)	-0.000 (41)
CRPS ↓	-0.007 (41)	-0.001 (50)
Dice ↑	$+0.008$ $[-0.015, +0.031]$ (30)	$+0.004$ $[-0.009, +0.017]$ (38)

GL-guided and unguided FlowDPS ($\lambda_0 = 0$) start from the same initial noise, so their paired difference isolates the effect of GL guidance. The GL runs cover 78 of the 100 held-out observations: a class-stratified subset of 25, scored first, and 53 of the remaining 75. GL guidance lowers the energy it targets: the mean GL energy of the samples falls from 0.217 to 0.091. This does not carry over to the reconstructions (Table 9). No reconstruction metric changes significantly. The change in susceptibility Dice, $+0.008$ $[-0.015, +0.031]$, was $+0.056$ $[-0.003, +0.116]$ on the first 25 observations but did not replicate on the other 53 (-0.015 $[-0.033, +0.002]$). GL raises the median magnetic misfit from 16.2 to 23.9 nT and slightly lowers the gravity misfit, from 0.255 to 0.238 mGal. On the tuning observations, stronger guidance ($\lambda_0 \geq 3 \times 10^6$) raised susceptibility Dice by up to 0.033 but increased the median magnetic misfit by at least 54% (Appendix E.4). Figure 4 shows GL-guided samples for four held-out models.

Two properties of the setup may limit the effect. The data and GL gradients are nearly orthogonal (median cosine -0.002 on the tuning observations), so GL mostly acts in directions that the data do

not constrain. With per-sample anchors, the energy is also scale-free: it sharpens two-phase structure relative to each sample's own quantiles, but it cannot correct absolute susceptibility levels or the depth of bodies (Section 6.2). Phase separation is, moreover, an assumption about the geology, and gradual or multiphase structure is poorly served by a two-phase prior.

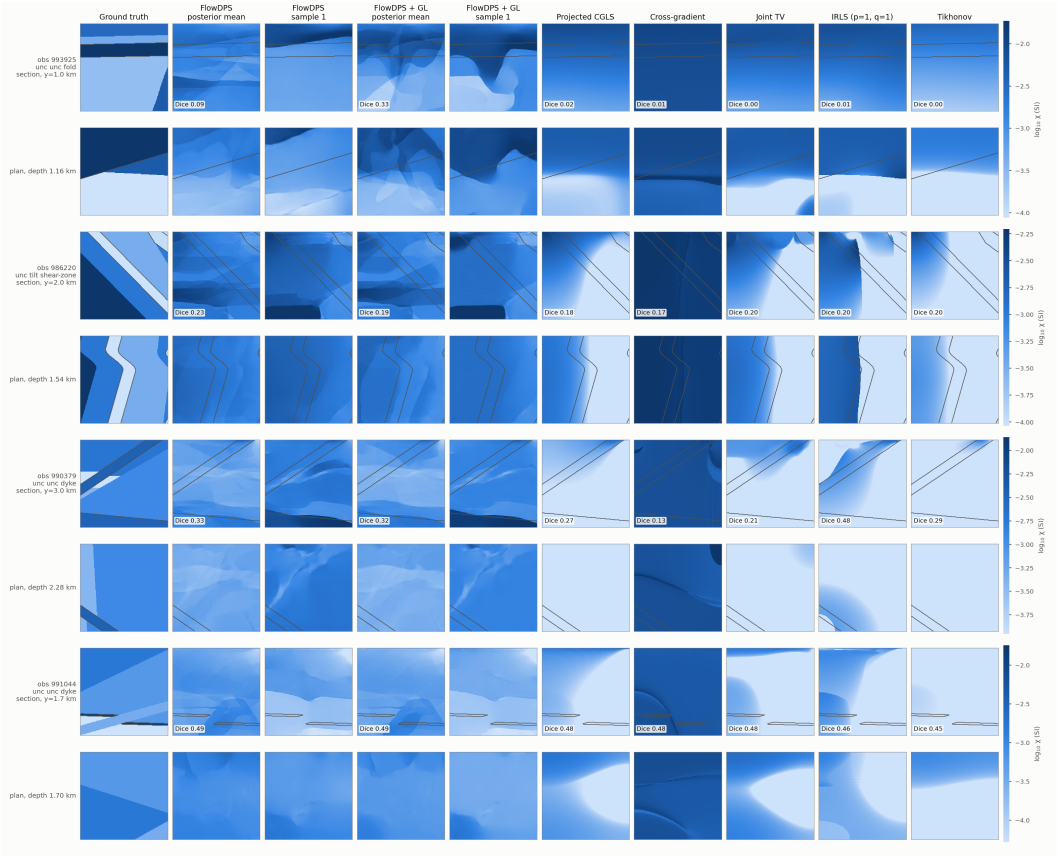

Figure 4: Susceptibility ($\log_{10} \chi$) of four held-out models at the 20th, 40th, 60th and 80th percentiles of FlowDPS Dice, each shown as a vertical section and a horizontal slice. Columns show the ground truth, the posterior mean and one sample for FlowDPS and for GL-guided FlowDPS, and five baselines. Gray lines trace ground-truth unit boundaries. Inset values are rank-matched Dice.

E.4 Choice of GL Hyperparameters

Interface width and anchors. A structural prior should assign lower energy to the ground truth than to the reconstructions it is meant to correct. We tested this on 20 training-range observations, pairing each ground truth with four unguided samples (80 pairs), and report the fraction of pairs in which the ground truth has the lower GL energy (Table 10). With per-sample anchors, the ground truth ranks lower in 73–75% of pairs for $\varepsilon \leq 0.5$, and this fraction falls quickly for wider interfaces: ground-truth volumes have sharp one-voxel boundaries, which a wide interface penalizes. Global anchors rank the ground truth lower in at most 38% of pairs at every ε. We use per-sample anchors and $\varepsilon = 0.5$, the widest value on the plateau.

Guidance strength. We swept λ_0 with the final sampler on the ten tuning observations (Table 11). The grid spans the point where the data and GL gradients have equal median norm, $\lambda_0 \approx 2.9 \times 10^8$. Before the sweep, we fixed a selection rule: the largest λ_0 whose median magnetic and gravity misfits stay within 5% of the $\lambda_0 = 0$ control, whose susceptibility Dice is at least the control's minus $\max(0.005, 2\,\mathrm{SE})$, and whose susceptibility CRPS stays within 2%. No value passed, since every λ_0 raised the median magnetic misfit by more than 5%. The misfit is bimodal (Appendix D.1), so its median lies between two clusters. We therefore adopted a criterion after the sweep: every observation that the control fits within σ_{mag} must remain within σ_{mag}. The largest value that passes is $\lambda_0 = 10^6$.

Table 10: Fraction of 80 pairs in which a volume has lower GL energy than an unguided posterior sample of the same observation. "VAE recon." is the ground truth passed through the VAE.

ε [voxels]	0.25	0.35	0.5	0.7	1.0	1.4	2.0
Per-sample anchors, ground truth	0.750	0.738	0.725	0.613	0.463	0.213	0.138
Per-sample anchors, VAE recon.	0.688	0.675	0.650	0.575	0.425	0.300	0.150
Global anchors, ground truth	0.375	0.350	0.350	0.350	0.275	0.200	0.175
Global anchors, VAE recon.	0.350	0.338	0.325	0.325	0.325	0.225	0.163

For the five such observations, the median misfit over four samples rises from 2–7 nT at $\lambda_0 = 0$ to 5–14 nT at 10^6 and 15–18 nT at 3×10^6. This choice was made before any held-out GL run, but it was not pre-registered.

Table 11: GL strength sweep on ten training-range observations (final sampler). Misfits are medians, and Dice and CRPS are means over observations. ΔDice is the mean paired change in susceptibility Dice relative to $\lambda_0 = 0$ ($\pm$ one standard error), and $n_\uparrow$ counts improved observations.

λ_0	ΔT [nT]	g_z [mGal]	GL energy	Dice χ / ρ	CRPS χ	ΔDice χ	$n_\uparrow$
0	58.7	0.324	0.205	0.172 / 0.206	0.573	–	–
10^6	81.5	0.335	0.089	0.186 / 0.207	0.558	$+0.014 \pm 0.011$	6
3×10^6	90.3	0.322	0.072	0.197 / 0.208	0.546	$+0.024 \pm 0.019$	6
10^7	97.1	0.344	0.066	0.205 / 0.206	0.541	$+0.033 \pm 0.027$	2
3×10^7	147.0	0.297	0.061	0.202 / 0.205	0.540	$+0.029 \pm 0.024$	5
10^8	181.4	0.290	0.038	0.196 / 0.207	0.540	$+0.024 \pm 0.028$	6
3×10^8	253.7	0.279	0.028	0.180 / 0.208	0.559	$+0.008 \pm 0.035$	4
10^9	442.6	0.309	0.020	0.187 / 0.184	0.629	$+0.015 \pm 0.050$	6

E.5 Phase-Field Map

Why a log domain. Magnetic susceptibility is heavy-tailed. Over the voxels of 500 training models, its median is 7.5×10^{-4} and its 95th percentile is 0.044, while the upper bound used by the baselines is $\chi_{0.999} = 1.09$ and the raw maximum is 1106.89 (Section 5.2). A map linear in χ, as in (8), must span this whole range, for example with $\chi^{\text{host}} = 0$ and $\chi^{\text{ore}} = 1.09$. It then sends the median voxel to $\phi = -0.9986$ and the 95th-percentile voxel to $\phi = -0.92$. Nearly the whole volume sits in the host well, and the GL energy and its gradient nearly vanish. The contrast between ore and host is a ratio rather than a difference, so we apply the map to log-susceptibility,

$$\phi(\chi) = \text{clip}\left(2\,\frac{\chi_{\log} - \chi_{\log}^{\text{host}}}{\chi_{\log}^{\text{ore}} - \chi_{\log}^{\text{host}}} - 1,\ -1,\ 1 \right), \qquad \chi_{\log} = \log_{10}(\chi + \delta), \quad \delta = 10^{-6},$$

where $\chi_{\log}^{\text{host}}$ and $\chi_{\log}^{\text{ore}}$ are the anchors in log units. The offset δ keeps the 0.17% of voxels with $\chi = 0$ finite. This is the transform the VAE already uses (Section 5.1), so between the clipping points ϕ is an affine function of the decoder output, and its gradient needs no exponentiation.

Choosing the anchors. We consider two ways to set $\chi_{\log}^{\text{host}}$ and $\chi_{\log}^{\text{ore}}$, both using the median of $\chi_{\log}$ for host and its 95th percentile for ore.

- **Global anchors** are fixed quantiles over the voxels of 500 training models: $\chi_{\log}^{\text{host}} = -3.124$ and $\chi_{\log}^{\text{ore}} = -1.356$, that is, $\chi \approx 7.5 \times 10^{-4}$ and 0.044, 1.77 decades apart.
- **Per-sample anchors** are the same quantiles of the sample being evaluated. They are re-estimated at every evaluation and held constant under differentiation, as the region means are in Chan–Vese segmentation (Chan & Vese, 2001). The ore anchor is kept at least 0.1 decades above the host anchor.

Global anchors ignore that background levels differ between models. Over 60 random training models, the median of $\chi_{\log}$ ranges from -4.66 to -0.91, with 10th, 50th and 90th percentiles -4.13, -3.24 and -1.74. Under global anchors, the background of a model at the 90th percentile maps to $\phi \approx 0.57$, closer to the ore well than to the host well, so the double-well term penalizes the host

rock itself. At the 10th percentile, the background and all contrasts below $\chi_{\log} = -3.12$ are clipped to $\phi = -1$, where the prior cannot see them. Per-sample anchors place every model's background at $\phi = -1$ by construction. They are also closer together: the median gap between a model's 95th percentile and median is 0.82 decades, against 1.77 for the global anchors. In 12 of the 60 models the two quantiles coincide, because one unit fills at least 45% of the volume, and the minimum gap of 0.1 decades keeps the map defined there. With per-sample anchors, (10) is no longer a fixed density over $\boldsymbol{\chi}$: the guidance gradient is that of the GL energy with the anchors frozen at their current values. The experiments in Appendix E.3 use per-sample anchors, which rank the ground truth below unguided samples far more often than global anchors do (Appendix E.4).

E.6 Graph Laplacian

The standard graph Laplacian has elements

$$L_{ij} = \begin{cases} |\mathcal{N}(i)| & \text{if } i = j, \\ -1 & \text{if } j \in \mathcal{N}(i), \\ 0 & \text{otherwise.} \end{cases}$$

E.7 Discrete GL Energy

Discretizing (9) on a grid with N cells yields

$$\mathcal{E}_{\mathrm{GL}}[\boldsymbol{\phi}] = \sum_{i=1}^{N} \Delta V \left[\frac{\kappa}{4} \sum_{j \in \mathcal{N}(i)} \frac{(\phi_i - \phi_j)^2}{h^2} + \frac{1}{4\varepsilon^2}(\phi_i^2 - 1)^2 \right],$$

where $\mathcal{N}(i)$ denotes the neighbors of cell i, h is the grid spacing, and $\Delta V = h^3$ is the cell volume. Since $\sum_i \sum_{j \in \mathcal{N}(i)} (\phi_i - \phi_j)^2 = 2 \sum_{\langle i,j \rangle} (\phi_i - \phi_j)^2 = 2\, \boldsymbol{\phi}^\top \boldsymbol{L} \boldsymbol{\phi}$ for the graph Laplacian in Appendix E.6, this equals ΔV times the matrix form

$$\mathcal{E}_{\mathrm{GL}}[\boldsymbol{\phi}] = \frac{\kappa}{2h^2} \boldsymbol{\phi}^\top \boldsymbol{L} \boldsymbol{\phi} + \frac{1}{4\varepsilon^2} \sum_{i=1}^{N} (\phi_i^2 - 1)^2, \tag{12}$$

The cell volume ΔV is constant, so we absorb it into the prior weight λ_{GL} in (10). We evaluate (12) divided by N, a per-voxel mean, with the neighbor sum over face-adjacent pairs inside the volume (no-flux boundary). A single voxel whose phase disagrees with all six neighbors creates six interfaces, each with $(\phi_i - \phi_j)^2 = 4$, and so costs an additional $12\kappa/h^2$: such configurations are allowed but strongly penalized. The gradient of (12) has a closed form (Appendix E.9).

E.8 Ising–Ginzburg–Landau Correspondence

Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Consider

$$\mathcal{E}_\varepsilon(\phi) = \int_\Omega \left(\frac{\varepsilon}{2} |\nabla \phi|^2 + \frac{1}{\varepsilon} W(\phi) \right) dx,$$

where $W \in C^2(\mathbb{R})$ satisfies $W \geq 0$, $W(\pm 1) = 0$, $W(s) > 0$ for $s \neq \pm 1$, and $W''(\pm 1) > 0$.

With $c_0 = \int_{-1}^{1} \sqrt{2W(s)}\, ds$, define

$$\mathcal{E}_0(u) = \begin{cases} c_0 \operatorname{Per}(E; \Omega), & u = \mathbb{1}_E - \mathbb{1}_{\Omega \setminus E}, \\ +\infty, & \text{otherwise,} \end{cases}$$

for sets $E \subset \Omega$ of finite perimeter, where $\mathbb{1}_E$ is the indicator of E. We assume in addition that W grows at least quadratically at infinity, as the double well $W(s) = \frac{1}{4}(s^2 - 1)^2$ does. Our GL energy (9) with $\kappa = 1$ and this double well equals $\varepsilon^{-1} \mathcal{E}_\varepsilon$, so $\varepsilon\, \mathcal{E}_{\mathrm{GL}}$ converges to the perimeter functional below.

Theorem 1 (Modica–Mortola (Modica & Mortola, 1977; Modica, 1987))**.** *As $\varepsilon \to 0$, $\mathcal{E}_\varepsilon$ Γ-converges to $\mathcal{E}_0$ in $L^1(\Omega)$.*

Proof. We prove compactness, the liminf inequality, and a recovery sequence (limsup inequality).

Compactness. Assume $\sup_\varepsilon \mathcal{E}_\varepsilon(\phi_\varepsilon) < \infty$ and define

$$\Phi(s) = \int_0^s \sqrt{2W(r)}\, dr.$$

By the Modica inequality ($a^2 + b^2 \geq 2ab$),

$$\frac{\varepsilon}{2}|\nabla\phi|^2 + \frac{1}{\varepsilon}W(\phi) \geq \sqrt{2W(\phi)}\,|\nabla\phi| = |\nabla\Phi(\phi)| \quad \Longrightarrow \quad \int_\Omega |\nabla\Phi(\phi_\varepsilon)|\, dx \leq C.$$

The growth of W also bounds $\Phi(\phi_\varepsilon)$ in $L^1(\Omega)$, so $\Phi(\phi_\varepsilon)$ is bounded in $BV(\Omega)$. By the compact embedding $BV(\Omega) \hookrightarrow L^1(\Omega)$, a subsequence converges in L^1 and almost everywhere. Since Φ is continuous and strictly increasing, ϕ_ε converges almost everywhere to some u, and in $L^1(\Omega)$ by equi-integrability, which follows from the growth of W. Finally, $\int_\Omega W(\phi_\varepsilon)\, dx \leq C\varepsilon \to 0$ and Fatou's lemma give $\int_\Omega W(u)\, dx = 0$, so $u(x) \in \{\pm 1\}$ almost everywhere, and $u \in BV(\Omega; \{\pm 1\})$ because $\Phi(u) \in BV(\Omega)$.

Liminf inequality. Let $\phi_\varepsilon \to u$ in L^1. By the Modica inequality and lower semicontinuity of the total variation,

$$\liminf_{\varepsilon\to 0} \mathcal{E}_\varepsilon(\phi_\varepsilon) \geq \liminf_{\varepsilon\to 0} \int_\Omega |\nabla\Phi(\phi_\varepsilon)|\, dx \geq |D\Phi(u)|(\Omega).$$

If $u = \mathbb{1}_E - \mathbb{1}_{\Omega\setminus E}$, then $\Phi(u) = \Phi(-1) + c_0 \mathbb{1}_E$, so $|D\Phi(u)|(\Omega) = c_0\, |D\mathbb{1}_E|(\Omega) = c_0 \operatorname{Per}(E;\Omega)$. Otherwise, the left-hand side is infinite by the compactness step.

Recovery sequence. Let $u = \mathbb{1}_E - \mathbb{1}_{\Omega\setminus E}$, where by density we may assume that ∂E is smooth. Let d_E be the signed distance to ∂E, positive in E and negative outside, and define

$$\phi_\varepsilon(x) = q\left(\frac{d_E(x)}{\varepsilon}\right),$$

where q solves the heteroclinic ODE

$$q' = \sqrt{2W(q)}, \quad q(-\infty) = -1, \quad q(+\infty) = 1.$$

Then $\phi_\varepsilon \to +1$ in E and -1 outside, so $\phi_\varepsilon \to u$ in L^1. Along the profile, equipartition $\frac{\varepsilon}{2}|\nabla\phi_\varepsilon|^2 = \frac{1}{\varepsilon}W(\phi_\varepsilon)$ holds, so the coarea formula and the one-dimensional energy $\int_{\mathbb{R}} \sqrt{2W(q)}\, q'\, ds = c_0$ give

$$\lim_{\varepsilon\to 0} \mathcal{E}_\varepsilon(\phi_\varepsilon) = c_0 \operatorname{Per}(E;\Omega).$$

Thus, the limsup inequality holds.

The Γ-convergence follows. □

E.9 Ginzburg–Landau Energy Gradient

The gradient of the discrete GL energy (12) with respect to $\boldsymbol{\phi}$ is

$$\nabla_{\boldsymbol{\phi}} \mathcal{E}_{\text{GL}} = \frac{\kappa}{h^2}\boldsymbol{L}\boldsymbol{\phi} + \frac{1}{\varepsilon^2}\,\boldsymbol{\phi} \odot (\boldsymbol{\phi} \odot \boldsymbol{\phi} - \mathbf{1}), \tag{13}$$

where $\odot$ denotes element-wise multiplication.

Proof. For the gradient term:

$$\frac{\partial}{\partial \phi_k}\left[\frac{\kappa}{2h^2}\boldsymbol{\phi}^\top \boldsymbol{L}\boldsymbol{\phi}\right] = \frac{\kappa}{h^2}[\boldsymbol{L}\boldsymbol{\phi}]_k$$

since $\boldsymbol{L}$ is symmetric.

For the double-well term:

$$\frac{\partial}{\partial \phi_k}\left[\frac{1}{4\varepsilon^2}(\phi_k^2-1)^2\right] = \frac{1}{\varepsilon^2}\phi_k(\phi_k^2-1)$$

□

E.10 Ginzburg–Landau Energy Hessian

The Hessian of the GL energy is

$$\boldsymbol{H}_{\mathrm{GL}} = \frac{\kappa}{h^2}\boldsymbol{L} + \frac{1}{\varepsilon^2}\mathrm{diag}(3\boldsymbol{\phi}\odot\boldsymbol{\phi}-\mathbf{1}).$$

E.11 Allen–Cahn and Gibbs Interpretation

The GL energy induces the Allen–Cahn gradient flow

$$\frac{\partial \phi}{\partial t} = -\frac{\delta \mathcal{E}_{\mathrm{GL}}}{\delta \phi} = \kappa\nabla^2\phi - W'(\phi) = \kappa\nabla^2\phi - \frac{1}{\varepsilon^2}\phi(\phi^2-1).$$

Lemma 1 (Energy Dissipation)**.** *The Allen–Cahn equation monotonically decreases the GL energy:*

$$\frac{d\mathcal{E}_{GL}}{dt} = -\int_\Omega \left|\frac{\partial \phi}{\partial t}\right|^2 d\boldsymbol{x} \le 0.$$

Proof.

$$\begin{aligned}\frac{d\mathcal{E}_{\mathrm{GL}}}{dt} &= \int_\Omega \frac{\delta \mathcal{E}_{\mathrm{GL}}}{\delta \phi}\frac{\partial \phi}{\partial t}\, d\boldsymbol{x} \\ &= \int_\Omega \frac{\delta \mathcal{E}_{\mathrm{GL}}}{\delta \phi}\left(-\frac{\delta \mathcal{E}_{\mathrm{GL}}}{\delta \phi}\right) d\boldsymbol{x} \\ &= -\int_\Omega \left|\frac{\delta \mathcal{E}_{\mathrm{GL}}}{\delta \phi}\right|^2 d\boldsymbol{x} \le 0.\end{aligned}$$

□

Let $\Omega \subset \mathbb{R}^d$ be bounded with Neumann boundary conditions. Consider the Allen–Cahn equation (Allen & Cahn, 1979) with additive noise

$$d\phi_t = (\kappa\Delta\phi_t - W'(\phi_t))\, dt + \sqrt{2T}\, dW_t, \tag{14}$$

where W_t is a cylindrical Wiener process in $L^2(\Omega)$ (not to be confused with the double-well potential W) and $T > 0$ is a temperature.

Theorem 2 (Invariant Gibbs Measure)**.** *Assume suitable coercivity of W and well-posedness of* (14)*. Then the measure*

$$\mu(d\phi) = Z^{-1}\exp\left(-\frac{\mathcal{E}_{GL}[\phi]}{T}\right) d\phi$$

is invariant for the stochastic Allen–Cahn equation.

Proof. The variational derivative of the GL energy is $\frac{\delta \mathcal{E}_{\mathrm{GL}}}{\delta \phi} = -\kappa\Delta\phi + W'(\phi)$, and the drift of (14) is its negative. Thus, the SPDE can be written as $d\phi_t = -\frac{\delta \mathcal{E}_{\mathrm{GL}}}{\delta \phi}dt + \sqrt{2T}\, dW_t$. The formal generator is $\mathcal{L}F(\phi) = \langle -\delta\mathcal{E}_{\mathrm{GL}}/\delta\phi, DF(\phi)\rangle + T\mathrm{Tr}(D^2F(\phi))$. By integration by parts in infinite dimensions, the adjoint operator satisfies $\mathcal{L}^*\mu = 0$. Thus, μ is invariant. □

Table 12: VAE sweep configurations. The latent shape is reported as channels $\times$ spatial size^3.

Sweep	KL weight	Latent shape	Base ch.	Bottleneck	Res. blocks	Attn. blocks	Params
1	10^{-5}	8×40^3	8	48	1	1	614,091
1	10^{-4}	8×40^3	8	48	1	1	614,091
1	10^{-3}	8×40^3	8	48	1	1	614,091
2	10^{-3}	4×40^3	4	32	1	1	235,135
2	5×10^{-3}	4×40^3	4	32	1	1	235,135
2	10^{-2}	4×40^3	4	32	1	1	235,135
3	10^{-3}	2×24^3	4	32	2	2	385,578
3	5×10^{-3}	2×24^3	4	32	2	2	385,578
3	10^{-3}	4×24^3	4	32	2	2	385,736
3	5×10^{-3}	4×24^3	4	32	2	2	385,736

Table 13: Best validation metrics for all completed VAE sweep runs. We report the checkpoint achieving the best validation reconstruction loss, along with the corresponding validation KL and total loss.

Sweep	KL weight	Latent shape	Best val recon	Val KL	Val total
1	10^{-5}	8×40^3	7.3×10^{-3}	7.2	7.4×10^{-3}
1	10^{-4}	8×40^3	7.3×10^{-3}	3.7	7.7×10^{-3}
1	10^{-3}	8×40^3	8.0×10^{-3}	1.2	9.2×10^{-3}
2	10^{-3}	4×40^3	2.0×10^{-2}	2.5	2.2×10^{-2}
2	5×10^{-3}	4×40^3	2.4×10^{-2}	0.98	2.9×10^{-2}
2	10^{-2}	4×40^3	2.3×10^{-2}	0.57	2.9×10^{-2}
3	10^{-3}	2×24^3	3.9×10^{-2}	4.0	4.3×10^{-2}
3	5×10^{-3}	2×24^3	3.7×10^{-2}	1.6	4.5×10^{-2}
3	10^{-3}	4×24^3	2.9×10^{-2}	3.0	3.2×10^{-2}
3	5×10^{-3}	4×24^3	3.5×10^{-2}	1.3	4.2×10^{-2}

Remark 1 (GL Weight and Temperature)**.** *The Gibbs measure $p_{eq}[\phi] \propto \exp\left(-\mathcal{E}_{GL}[\phi]/T\right)$ coincides with a GL-regularized prior $p_{GL}(\phi) \propto \exp(-\lambda_{GL}\mathcal{E}_{GL}[\phi])$ under the identification $\lambda_{GL} = 1/T$. This identification applies to the stationary Gibbs measure of the SPDE. The inference-time weight $\lambda_{GL}(t)$ is instead a continuation schedule for reverse diffusion, not a physical temperature evolution.*

F VAE Training Details and Sweeps

All VAE experiments were trained on joint density–susceptibility volumes using AdamW with learning rate 10^{-4}. Across all sweeps, we used bf16 mixed precision, gradient clipping of 1.0, random seed 42, and evaluated every $100{,}000$ processed samples. Training ran for up to 10^6 processed samples. The first two sweeps used a latent spatial size of 40^3 and a shallower architecture, while the third sweep used a deeper attention VAE with a smaller latent spatial size of 24^3, obtained by projecting the 25^3 bottleneck grid.

Sweep design. Sweep 1 varied only the KL weight while using a relatively large model. Sweep 2 again varied the KL weight, but reduced the model width and parameter count. Sweep 3 changed both the latent channel count and KL weight, while also switching to a deeper architecture with two residual blocks per stage and two bottleneck attention blocks.

Table 12 shows that the experiments span a clear reconstruction–regularization–compute tradeoff. Sweeps 1 and 2 use larger latent spatial grids (40^3), whereas Sweep 3 reduces the latent spatial resolution to 24^3 in order to make downstream latent generative modeling substantially more tractable.

The validation KL is important beyond the total loss alone, since the latent prior must be regular enough for a downstream diffusion or flow model to learn a useful distribution. Very small KL

Table 14: BiFlowNet backbone comparison. Both models were trained on latents encoded with the Sweep 3 VAE checkpoint (4×24^3, KL weight 10^{-3}).

Model	Model dim	Mid DiT blocks	Attn. heads	Latent shape	Params	Final val loss
S	32	1	4	4×24^3	8,397,188	0.215393
L	48	3	12	4×24^3	19,957,636	0.181192

weights improve reconstruction, but they also leave a much less regularized latent space. For example, in Sweep 1, moving from KL weight 10^{-3} to 10^{-5} improves validation reconstruction from 0.007987 to 0.007340, but increases the validation KL from 1.229854 to 7.163707. Conversely, increasing the KL weight reduces the latent KL substantially, but comes at the cost of worse reconstruction.

For the main experiments, we select the Sweep 1 model with latent shape 8×40^3 and KL weight 10^{-3} ($614{,}091$ parameters). It has the lowest validation KL of the Sweep 1 runs, its reconstruction loss is within 9% of the best run, and that loss is less than a third of the best Sweep 3 model's. The sweeps also show that larger latents and more parameters lower the reconstruction loss, so further scaling of the VAE is a direct route to improving the inversion pipeline.

G BiFlowNet Backbone Comparison

Before selecting the final VAE, we compared two BiFlowNet backbones for latent rectified-flow training on latents from the Sweep 3 VAE. Both models were trained on the same latent dataset with latent shape 4×24^3, batch size 256, learning rate 5×10^{-5}, warmup over 150,000 samples, and a total training budget of 5×10^6 processed samples. The two runs differ only in model capacity.

The corresponding final training losses were 0.208852 for the smaller model and 0.176247 for the larger model. The larger BiFlowNet improves both optimization and generalization, achieving a lower training loss as well as a lower validation loss. This indicates that the larger backbone is not merely overfitting, but is learning a better latent generative prior. Note that these values are latent flow-matching losses, not physical reconstruction metrics, so they have no absolute "acceptable" threshold. They are meaningful for comparing models trained under the same setup. Inversion quality is measured against the 3D ground truth in Section 6.

The BiFlowNet comparison mirrors the VAE sweep behavior: increasing model capacity improves performance when compute allows it. Accordingly, the final prior uses the larger backbone, trained on the Sweep 1 latents (8×40^3). With 8 latent channels instead of 4, it has $19{,}963{,}208$ parameters. It uses the same learning rate and warmup, with a batch size of 246–297. We use the checkpoint with the lowest validation loss, 0.1350 after 3.5×10^6 processed samples (about 3.7 passes over the training latents). This loss is not comparable to Table 14, which uses different latents. At the same time, the gap between the two models suggests that additional scaling of the latent generative model remains a promising direction for improving inversion quality further.