IMAGE FIDELITY IS NOT FIELD FIDELITY: JOINT THERMODYNAMIC RECONSTRUCTION AND ERROR LOCALIZATION IN NEURAL TOMOGRAPHY

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural fields for scientific tomography are optimized from 2D images, but the actual quantity of interest is often a latent 3D physical field. Because the forward map is many-to-one, low 2D image error need not certify a correct 3D field. Moreover, the latent field is not directly supervised during training, and its error cannot be evaluated against truth at deployment. We develop CoroNeRF to jointly optimize 3D electron density and temperature fields directly from multiview, multiline intensities through a differentiable atomic-emission renderer. Using solar coronal tomography as a controlled testbed, we evaluate physical-field recovery and test whether cross-seed instability provides a ground-truth-free-at-inference indicator of local physical-field error. We underscore the following two observations. (i) Image fidelity is not field fidelity: spectral ablations show that limited-channel reconstructions can fit their available observations well while recovering substantially worse fields, whereas evaluation on a common richer probe exposes the discrepancy. (ii) Cross-seed instability ranks local physical-field error across tested matched-model conditions, supported by sparsification and physical signal-strength controls. Seed-deviation projections provide complementary directional validation, but shared forward-model mismatch can still produce incorrect cross-seed consensus. These results characterize joint thermodynamic recovery and the usefulness and limits of seed-based error localization in a controlled, single-scene solar tomography testbed.

1 INTRODUCTION

Scientific tomographic imaging problems aim to recover latent 3D physical fields where the data available for supervised learning are indirect, multiview 2D measurements, such as projections, line-of-sight (LOS) integrals, or other nonlinear measurements. Neural fields (Mildenhall et al., 2020) and NeRF-style frameworks (Gao et al., 2026) have been developed primarily for 3D scene representation and are a promising paradigm for scientific tomography. These frameworks represent a spatial field as a continuous coordinate-conditioned function and optimize it through a differentiable forward model. Under these frameworks, neural rendering problems are often evaluated as novel view synthesis problems, where success is determined by how well the model reproduces the training images and novel views. However, in scientific inverse problems, notably tomographic reconstruction, such a criterion is incomplete: a model can achieve low 2D image-space error while recovering an incorrect latent 3D physical field, especially when the forward map is non-injective and ill-conditioned, allowing substantially different fields to produce similar measurements. Moreover, in standard novel view synthesis pipelines, error in a physically interpreted latent field is typically not part of the task definition, let alone the training objective. **Thus, the latent field is not directly supervised during training and its error is unverifiable against truth during deployment,** leaving image agreement as a common but unreliable validation signal.

We investigate this limitation through joint reconstruction of the solar corona's 3D electron density and temperature. The solar corona is the outermost layer of the Sun, composed of hot plasma that is optically thin across a large wavelength range. Multiline 2D intensity observations are LOS integrals of 3D plasma emissivities (analogous to the RGBσ radiance proxies in standard NeRFs), and these emissivities are functions of the latent 3D plasma density and temperature fields. The goal is

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.

therefore not merely to reconstruct each line's emitting structure, but to recover a shared density-temperature state that explains all the observed channels. This inverse problem is highly ill-posed as distinct 3D density-temperature fields can produce nearly indistinguishable measurements. There is a standard geometric LOS degeneracy: since each pixel is an integral along a ray, a redistribution of the emissivities along that ray can still produce the same image. In addition, without sufficient spectral coverage, there is a plasma thermodynamic degeneracy: many density-temperature realizations can produce similar multiline emissivity values because the line-response functions are not jointly injective.

We thus introduce CoroNeRF, a differentiable multiline neural tomography framework for the direct joint recovery of 3D electron density and temperature from multiview spectral line-intensity images. Instead of learning the RGBσ radiance proxy of the scene, CoroNeRF parameterizes a neural plasma field that maps 3D position $\mathbf{x}$ to electron density $\log_{10} n_{\mathrm{e}}(\mathbf{x})$ and temperature $\log_{10} T_{\mathrm{e}}(\mathbf{x})$. A differentiable line-emission renderer then maps these fields to multichannel spectral line intensities using precomputed emissivity tables on samples along each camera ray. All channels are synthesized from the same density and temperature fields, which are optimized jointly using only image supervision. Within this framework, we examine how observational information affects thermodynamic recovery and whether cross-seed instability can localize the physical-field error without ground truth at inference.

Our contributions are as follows:

1. **Direct joint thermodynamic neural tomography.** We develop CoroNeRF to optimize 3D electron density and temperature directly from multiview, forbidden-line intensities through a differentiable atomic-emission renderer. At each spatial location, all modeled lines share the same inferred electron density and temperature, coupling the channel reconstructions through their atomic-emission responses.
2. **Image fidelity is not field fidelity.** In a controlled synthetic scene, spectral ablations show that limited-channel reconstructions can fit their available observations well while recovering substantially worse physical fields. A common richer probe exposes these discrepancies, while noise-view experiments characterize reconstruction performance under reduced observational information.
3. **Empirical field-error localization and its limits.** Cross-seed instability ranks local physical-field error across tested matched-model conditions, with emissivity, radial, sparsification, and seed-stability controls. Seed-deviation projections provide complementary directional validation, while a specified abundance mismatch demonstrates that shared model error can produce incorrect consensus.

2 Related Work

In the past two decades, various tomography methods have been used for coronal reconstruction, such as recovering coronal density from LASCO-C2 white-light images via classical regularized-least-squares (Frazin & Janzen, 2002) and 3D coronal density and temperature through EUV tomography (Frazin et al., 2009). Vector coronal tomographic inversion theory has also been developed for magnetic field inversion from spectropolarimetric observations, validated on both simulations (Kramar et al., 2006; Kramar & Inhester, 2007; Kramar et al., 2013; Kramar & Lin, 2026) and real observations (Kramar et al., 2016). These approaches use classical inverse-problem formulations with explicit regularization or prior structure, rather than learned neural representations coupled with an end-to-end differentiable nonlinear emission model.

Implicit neural representations and related learned inverse models have increasingly been used for scientific tomography. In medical imaging (Wang et al., 2024; Kabika et al., 2026), uncertainty-aware null-space networks have been studied for MRI (Angermann et al., 2023). Neural reconstruction has also been used for heterogeneous macromolecular structures (Zhong et al., 2021). Related neural-field approaches have been studied for physical flow field tomography (Molnar & Grauer, 2022), seismic full-waveform inversion (Sun et al., 2023), black hole dynamic tomography (Levis et al., 2022; Feng et al., 2026), and thermal tomography in inverse heat conduction problems (Zhong et al., 2026). NeRFs have also been used for solar coronal reconstruction: Asensio Ramos (2023) reconstructs density from polarized-brightness observations, SuNeRF (Jarolim et al., 2024) mod-

els wavelength-specific EUV emission and absorption, while SuNeRF-CME (Jarolim et al., 2026) reconstructs time-dependent electron density through Thomson-scattering observations. Related efforts have reported density-temperature reconstruction from multithermal EUV imaging (Southwest Research Institute, 2024; Jarolim et al., 2025). To our knowledge, CoroNeRF is the first neural-field framework to directly and jointly reconstruct 3D coronal electron density and temperature from multiview forbidden infrared-line intensities through a differentiable atomic-emission model.

Methods for uncertainty quantification in NeRFs have also been developed over the past five years. Density-aware NeRF ensembles employ deep ensembles to identify regions unobserved during training (Sünderhauf et al., 2023); full probabilistic frameworks such as Stochastic NeRFs (Shen et al., 2021) and Conditional-Flow NeRFs (Shen et al., 2022) quantify uncertainty by learning the distribution over NeRFs. Prior post-hoc methods also study uncertainty in conventional radiance fields: Bayes' Rays (Goli et al., 2024) evaluates geometric depth uncertainty, while FisherRF (Jiang et al., 2024) uses Fisher information. Finally, concurrent work on sparse-view CT shows that strong global uncertainty-error association can deteriorate substantially within the reconstructed object, and that error shared across independently trained members can remain invisible to ensemble spread (Zhao et al., 2026). We study a complementary setting involving coupled density-temperature recovery through nonlinear multiline atomic-emission responses, characterizing when cross-seed instability localizes physical-field error and when shared forward-model mismatch causes it to fail.

3 Methodology: Differentiable Multiline Neural Tomography

CoroNeRF parameterizes a neural plasma field $f_\theta(\mathbf{x}) \mapsto (\log_{10} \hat{n}_{\mathrm{e},\theta}(\mathbf{x}), \log_{10} \hat{T}_{\mathrm{e},\theta}(\mathbf{x}))$. Rather than first inverting observed line ratios, it jointly fits the absolute intensities of all lines through their emissivity responses. Density and temperature diagnostic information therefore enters implicitly through the forward model, while multiview geometry constrains the 3D distribution.

The observations are multiview, multichannel line-intensity images. Let $v \in \mathcal{V}$ index viewpoints, $p \in \Omega_v$ index valid image pixels under the view mask, and $c \in \{1, \ldots, C\}$ index the spectral line channels. We write $q = (v, p)$ for the ray associated with pixel p in view v. Each ray is then parameterized as $\mathbf{x}_q(t) = \mathbf{o}_q + t\mathbf{d}_q,\ t \in [t_s(q), t_f(q)]$, where $\mathbf{o}_q$ is the ray origin and $\mathbf{d}_q$ is the unit direction. The observed multichannel intensity vector for this ray is $\mathbf{y}_q = (y_{q,1}, \ldots, y_{q,C}) \in \mathbb{R}^C$. The training ray dataset is then given by $\mathcal{D}_{\mathrm{ray}} = \{(\mathbf{x}_q(\cdot), \mathbf{y}_q) : q = (v, p),\ v \in \mathcal{V}_{\mathrm{train}},\ p \in \Omega_v\}$.

Given a neural plasma field, the differentiable line-emission renderer maps physical variables to predicted line intensities. For channel c, the renderer first evaluates a channel-dependent emissivity $\epsilon_c(n_{\mathrm{e}}, T_{\mathrm{e}}, r)$, where r is the heliocentric radius, using a precomputed atomic-physics lookup table from the CHIANTI database (Dere, K. P. et al. (1997); Dere et al. (2023), Appendix A.1). It then integrates the emissivities along each ray to produce the predicted line intensity:

$$\hat{I}_{\theta,q,c} = \int_{t_s(q)}^{t_f(q)} \epsilon_c\Big(\hat{n}_{\mathrm{e},\theta}(\mathbf{x}_q(t)), \hat{T}_{\mathrm{e},\theta}(\mathbf{x}_q(t)), r(\mathbf{x}_q(t))\Big)\, dt. \tag{1}$$

In practice, we compute LOS integrals using numerical quadrature at the fixed sampling resolution described in Appendix A.2. The inverse problem is then latent physical-field recovery of $n_{\mathrm{e}}(\mathbf{x})$ and $T_{\mathrm{e}}(\mathbf{x})$: we optimize the neural plasma field parameters θ using image supervision only:

$$\theta^\star = \arg\min_\theta \frac{1}{|\mathcal{D}_{\mathrm{ray}}|} \sum_{(x_q(\cdot), y_q) \in \mathcal{D}_{\mathrm{ray}}} \frac{1}{C} \sum_{c=1}^{C} \rho(\hat{I}_{\theta,q,c}, y_{q,c}). \tag{2}$$

For most experiments, we use a fixed-scale asinh image loss $\rho(\hat{I}, y) = |\operatorname{asinh}(\hat{I}/s_c) - \operatorname{asinh}(y/s_c)|$, where s_c is a channel-dependent fixed characteristic intensity scale, chosen during development so as to not allow bright channels to dominate. While our images have heteroscedastic Gaussian noise, image-space fidelity is not the ultimate objective: we thus use the asinh-L1 image loss to compress the intensity dynamic range. Among the tested objectives, it yields lower field error than the inverse-variance-weighted squared-residual objective (Appendix D.3).

4 Experimental Setup and Evaluation

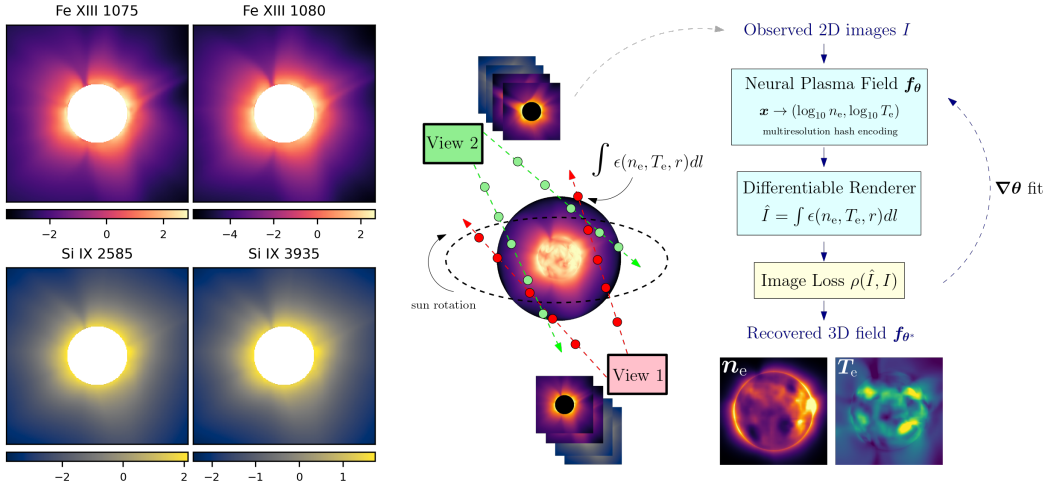

Figure 1: CoroNeRF overview. The left 2×2 grid displays multiline coronal observations of four spectral channels (Fe XIII $1075/1080$ nm, Si IX $2585/3935$ nm) forward-rendered from a PSI cube. The observations are used to fit a multiresolution hash-grid neural plasma field through the same differentiable renderer by minimizing an image loss (right panel). The fitted neural field yields jointly recovered 3D density and temperature fields.

We construct synthetic observations (left panel of Figure 1) of four spectral channels (Fe XIII $1075/1080$ nm, Si IX $2585/3935$ nm) from one ground truth PSI MAS (MHD) coronal cube (Mikić et al., 2007; Lionello et al., 2008) based on Carrington Rotation 2283 (April-May 2024). The GT field has $(299, 142, 154)$ longitude-colatitude-radius voxels spanning the full sphere and ranging from $[1, 30]$ $R_\odot$. The master dataset consists of 3000 views evenly spaced in longitude at $0°$ latitude: we mimic observational coverage expected from real Earth measurements, but the viewing geometry is a static-rotational tomography idealization. Each view is 256 by 256 pixels with a FOV of $[-3, +3]$ $R_\odot$, and each pixel corresponds to a particular ray through the coronal field. Because the corona is optically thin and faint ($\approx 10^{-6}$ of the central photospheric disk brightness), it is often observed by occulting the bright solar disk, as done in a coronagraph. We mirror this procedure by masking pixels within 1 $R_\odot$, which additionally confines the LOS integral to the $r \geq 1$ $R_\odot$ region where the emissivity model is well-defined. Consequently, all observation-space training and evaluation images apply the occulting mask. The canonical training set contains 300 evenly-spaced views (19.66M rays, 78.64M channel measurements, before the occulting mask), excluding a $30°$ evaluation arc and a $3°$ guard region around it: evaluation uses the arc, while the guard region is excluded from both sets. We chose this holdout protocol because it explicitly leaves a section of the corona views unseen by the model, which is a harder interpolation test rather than an interleaved-view held-out set.

We summarize each observational condition, or experiment family, by the tuple (η, V, Λ, a). We add heteroscedastic Gaussian noise (Equation 6) to our images, with the shot-noise variance multiplier η (e.g. $\eta = 9$ is $\times 3\sigma$). V is the number of views, Λ is the set of spectral channel indices, and $a \neq 1$ is a forward-model elemental abundance mismatch scale (Appendix A.4). We define the canonical noisy condition as $(\eta = 9, V = 300, \Lambda = \text{all } 4, a = 1)$, used later in Section 5.

Our joint-reconstruction model uses a multiresolution hash-grid encoder (Instant NGP-style, Müller et al. (2022)), with separate decoder heads for density and temperature predictions (Appendix A.5). We use the AdamW optimizer for 60k steps with batches of 1024 rays, optimized using a fixed-scale asinh loss (Appendices D.3, D.4). We evaluate performance using field-space mean absolute error on the inner coronal radial band $[1.1, 2.0]$ $R_\odot$, as well as held-out image-space loss (Appendix B). We show a visualization of the renderer and the training process in the right panel of Figure 1. Training a single reconstruction takes $\approx$ 3 GPU-hours on an NVIDIA A100, and rendering a 300-view dataset takes $\approx$ 10-100 minutes on an NVIDIA RTX 4090 depending on the resolution and integration step size (Appendix D.4).

We stress that this is a controlled study using one static coronal scene. Thus, we are evaluating joint thermodynamic reconstruction and within-scene error localization, not a probabilistic calibration or generalization across multiple coronal states. The objective family and architecture were selected based on physical-field recovery on this scene; for noise-level experiments, the asinh scale s_c is fixed at its baseline for $\eta \leq 1$ and scaled by $\sqrt{\eta}$ for $\eta > 1$ (Appendix D.6). All synthetic generation and inversion experiments also use the same discretized renderer (Appendix A.2), so the study evaluates reconstruction under a matched forward model except for abundance perturbations described in Appendix A.4.

5 Results

5.1 Joint Thermodynamic Reconstruction and Measurement Limitations

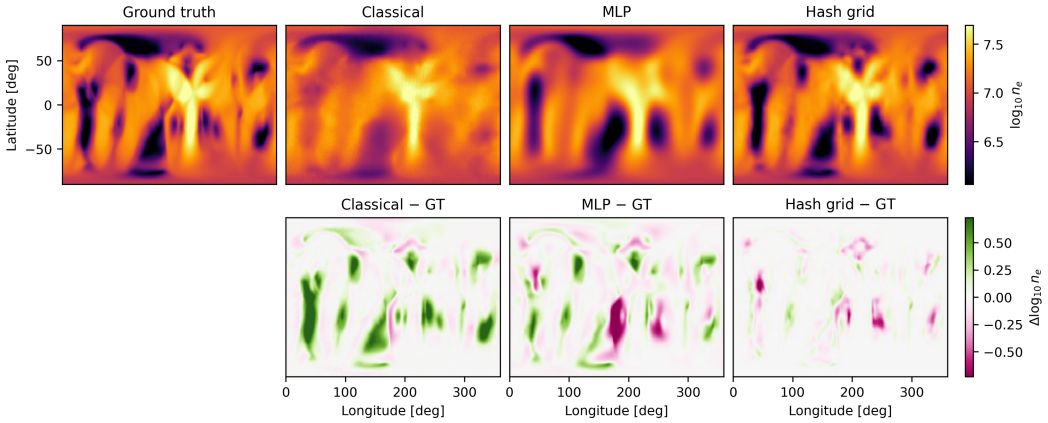

Figure 2: Density-field reconstruction (given GT temperature) at $1.51\ \mathrm{R}_\odot$. Each panel is a shell unraveled on a latitude-longitude plot. The top row shows the ground truth and reconstructions using a classical grid with Tikhonov regularization, positional encoder + MLP, and hash grid + decoder MLP representations. The bottom row shows corresponding residuals $\Delta \log_{10} n_e = \log_{10} n_{e,\mathrm{pred}} - \log_{10} n_{e,\mathrm{GT}}$. The hash grid is able to reproduce sharper small-scale features, substantially reducing inner-band MAE relative to the MLP by more than $2\times$ and classical grid methods by about $4\times$.

We first isolate the spatial representation by reconstructing density while supplying GT temperature and then select the hash-grid representation for the joint density-temperature inversion. Figure 2 shows density-field reconstruction results at $1.51\ \mathrm{R}_\odot$, trained using two matched-model, noiseless intensity channels (Fe XIII $1075/1080$) while given the ground truth temperature field. Latitude-longitude plots (flattened spherical shells) of the reconstructed density are shown on the top, while the signed residuals are shown on the bottom. We use $\mathrm{MAE}_{\mathrm{inner}}(\log_{10} n_e)$, the mean absolute error of log-density in the inner radial band $[1.1, 2.0]\ \mathrm{R}_\odot$, as our main metric. Under the tested grid resolution, architectures, and optimization settings, the multiresolution hash grid + decoder head attains the lowest reconstruction error, performing over twice as well (0.042 dex MAE) as a standard positional encoder + MLP (0.097 dex MAE), and about four times as well as a classical voxelized representation + Tikhonov regularizer (0.168 dex MAE). These comparisons are intended to isolate the effect of spatial representation under a common emission renderer, rather than to benchmark complete solar-tomography systems, which are generally tailored to different observations and inversion settings. See Appendix D.2 for detailed results.

Next, we jointly reconstruct density and temperature using all four spectral channels. The resulting inner-band MAEs are 0.039 dex for density (9.2% AbsRel) and 0.014 dex for temperature (3.2% AbsRel). Figure 3 visualizes the recovered fields through column-density and emissivity-weighted-temperature projections, and Figure 13 in Appendix D.5 compares reconstructed fields with ground truth at three radii. We find empirical evidence that increasing spectral coverage substantially reduces the density-temperature degeneracies: single-line models are underconstrained because the observed intensity depends jointly on density and temperature through a single emissivity response, allowing multiple density-temperature states to produce similar measurements. Because

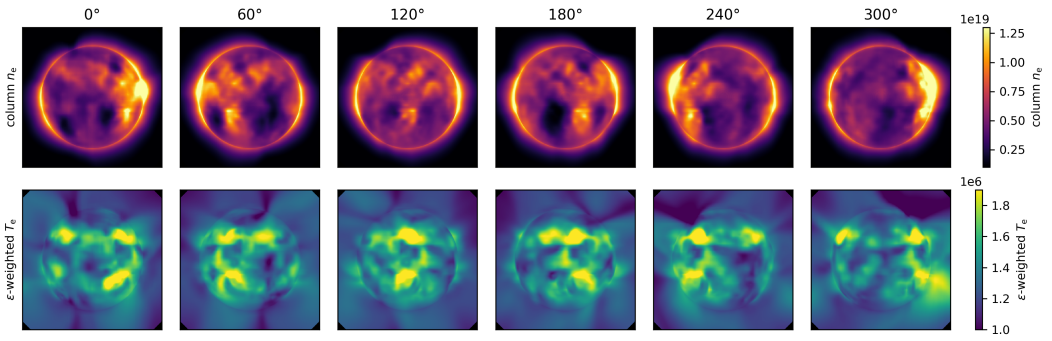

Figure 3: LOS projections of the jointly-reconstructed column density (top) and total-emissivity-weighted temperature (bottom), using a multiresolution hash-grid representation trained on all four spectral channels (Appendix A.3).

each line also has a finite temperature-response range, additional lines with different thermodynamic responses provide complementary constraints that substantially reduce these degeneracies: a pair of lines drops the n_e MAE by a factor of 5-10, and using four lines further halves the error. Finally, additional spectral lines also reduce the radial MAE at larger radii (Figure 14).

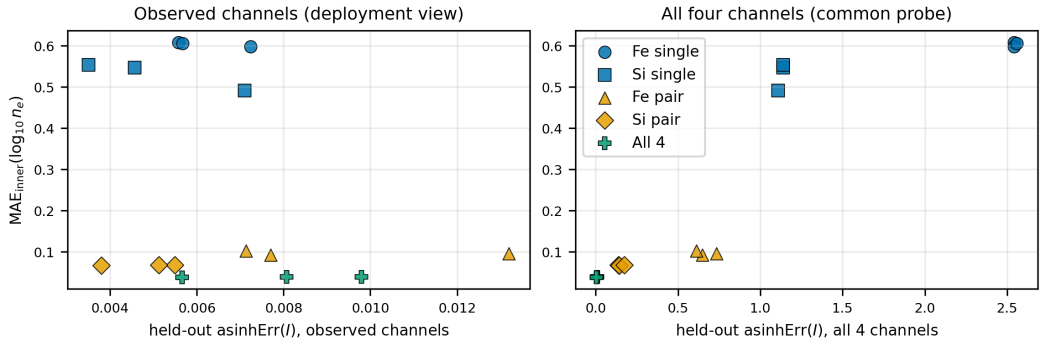

Figure 4: Image fidelity is not field fidelity. We show 3D density field-space MAE vs 2D held-out image-space loss. Each marker is a trained model (color = number of spectral lines, shape = the specific line set). The left plot uses held-out image error measured on each model's own channels, which is a GT-free image-space validation quantity available on held-out observed views. Single-channel models achieve lower observed-channel image error than the four-line model despite having substantially larger field error. Evaluation on a common four-line probe (right panel) exposes these failures and largely restores the density-error ranking, showing that the low observed-channel image error of these configurations reflects limited spectral coverage rather than correct field recovery.

We further show that **image fidelity is not field fidelity**: a low image-space error need not certify a low field-space error. Figure 4 shows a per-model scatter of field-space MAE vs 2D held-out validation image error (color = number of spectral lines, shape = the specific line set). The left panel uses the held-out image error measured on each model's own channels, which is a GT-free image-space validation quantity available on held-out observed views. We see that single-channel experiments fit their observed images better than the four-line models, yet they recover a far worse density field. The right panel recomputes the image-space evaluation on a common four-channel set across all models with the same fixed observational setup, showing that own-channel image error is not comparable as a certificate of field fidelity across spectral configurations. In the density comparison, the common probe largely restores the field-error ranking. Temperature reconstructions likewise exhibit the same discrepancy, although the common-probe image error does not perfectly order temperature error across line sets (Figure 15).

Finally, we assess the robustness under sparse-view and high-noise regimes. Figure 5 crosses four noise levels, $\eta \in \{0, 1, 9, 25\}$, with four view counts, $V \in \{5, 20, 100, 300\}$. For $V \leq 20$, density

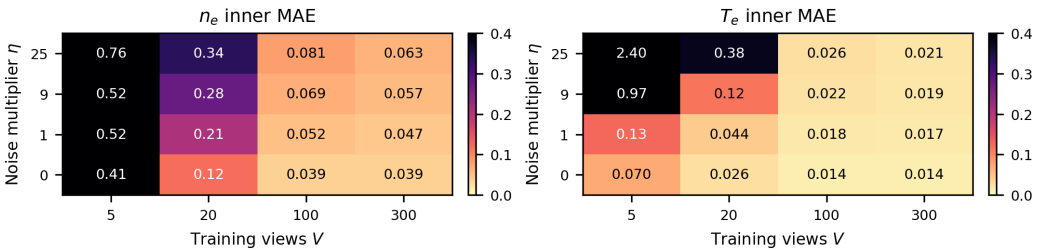

Figure 5: Noise-view sweep, fixed at 4 matched-model spectral channels, of n_{e} and T_{e} MAE. With at least 100 views, field recovery remains accurate across the tested noise levels under the noise-matched loss. Severe angular undersampling substantially increases error and makes the reconstruction much more vulnerable to noise.

MAE increases several-fold relative to $V \geq 100$, and severe angular undersampling substantially increases sensitivity to noise. However, for $V \geq 100$, increasing the noise scale from $1 \rightarrow 25$ only modestly increases the MAE (for example, $0.052 \rightarrow 0.081$ at 100 views). We conclude that within this scene, reconstructions with at least 100 views retain low field error across the tested noise levels. We additionally note that because s_c sets the scale of the asinh training loss, we scale it with the measurement-noise standard deviation; Figure 5 therefore compares reconstructions under a noise-matched loss.

5.2 Cross-Seed Instability as a GT-free-at-Inference Error-Ranking Proxy

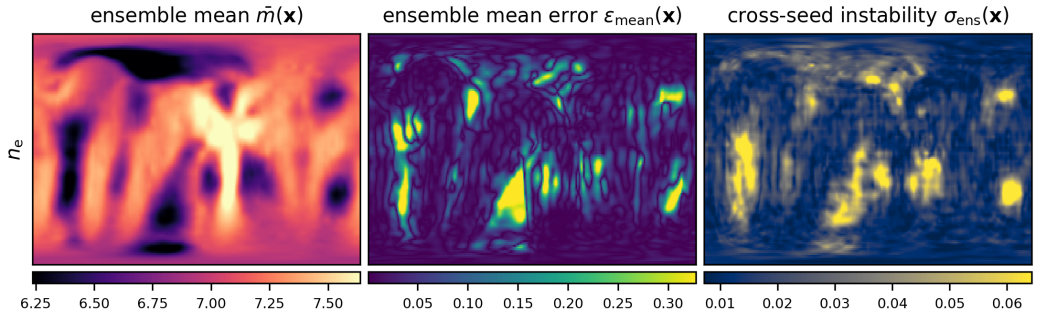

Figure 6: Ensemble panels of the 3D density field at $1.51\mathrm{R}_{\odot}$ using the canonical high-noise condition. We show the ensemble mean $\overline{m}(\mathbf{x})$, ensemble-mean error $\epsilon_{\mathrm{mean}}(\mathbf{x})$, and cross-seed instability $\sigma_{\mathrm{ens}}(\mathbf{x})$. The maps exhibit similar spatial structure, which we quantify using correlations below.

We evaluate whether standard seed ensembles provide useful local error rankings for the jointly reconstructed thermodynamic fields. We train $K = 10$ models for each condition, varying only the training seeds (Lakshminarayanan et al., 2017) while holding the model architecture, observational condition, and dataset fixed. The per-voxel standard deviation in the field predictions across the models is the cross-seed instability σ_{ens} (Appendix C.1). Figure 6 shows reconstructions at 1.51 $\mathrm{R}_{\odot}$ with the canonical high-noise regime for ensemble mean $\overline{m}(\mathbf{x})$, absolute ensemble-mean error $\epsilon_{\mathrm{mean}}(\mathbf{x})$, and cross-seed instability $\sigma_{\mathrm{ens}}(\mathbf{x})$. **Large σ_{ens} identifies regions where the recovered field is sensitive to training seeds under the fixed pipeline, while small σ_{ens} identifies regions of cross-seed consensus.** We note that consensus is not a correctness guarantee, because error shared by all ensemble members (such as a common-mode forward-model discrepancy) will go undetected.

We further examine whether rank association is explained by local signal strength. The left panel in Figure 7 shows the Spearman rank spatial correlation for the inner band density field between ϵ_{mean} and (i) σ_{ens} and (ii) an oracle inverse-emissivity signal-strength baseline $b_{\mathcal{E}}$. Error bars are 95% within-scene longitude-slice bootstrap intervals computed on the inner radial band using 2000 resamples. We find that σ_{ens} localizes error across three matched-model regimes: noiseless at 300 and 100 views (0.76-0.77), sparse-view at 20 and 5 views (0.71-0.78), and high-noise (0.55-0.57);

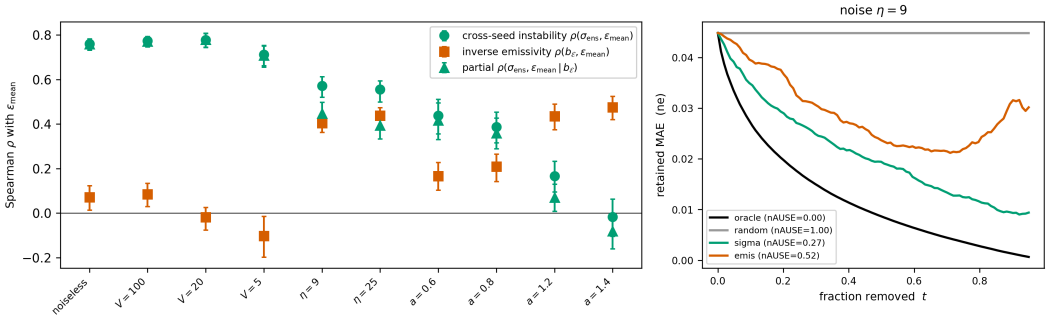

Figure 7: Left panel plots Spearman rank correlation between the true ensemble mean error ϵ_{mean} and σ_{ens} (green circle), inverse-emissivity $b_\mathcal{E}$ (orange), and partial correlation $\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}} \mid b_\mathcal{E})$ (green triangle), showing that the rank association persists after controlling for the oracle inverse-emissivity proxy. Error bars are 95% within-scene longitude-slice bootstrap intervals. We find that σ_{ens} localizes error in three matched-model regimes: noiseless, sparse-view, and high-noise, outperforming the emissivity proxy $b_\mathcal{E}$. The partial correlation shows that the rank association remains after conditioning on an oracle inverse-emissivity proxy. Finally, we note that the localization of σ_{ens} deteriorates substantially under positive Si over-scaling and approaches chance for density at $a = 1.4$, but under-scaling retains partial localization ability; while $b_\mathcal{E}$ can still track the high-error, low-emissivity regions in this regime, it requires the GT field and is thus used only as an oracle baseline. The right panel consists of sparsification curves for the canonical noisy condition, showing that σ_{ens} tracks the oracle more closely than does the baseline $b_\mathcal{E}$ (nAUSE of 0.27 compared to 0.52).

$b_\mathcal{E}$ has weak rank correlation with error in the noiseless and sparse-view regimes, while σ_{ens} localizes strongly. Moreover, the partial correlation $\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}} \mid b_\mathcal{E})$ remains positive across matched-model conditions, suggesting that σ_{ens} captures error structure beyond what is explained by signal strength alone. This association also remains positive within radial shells (Tables 4 and 5), indicating that it is not driven solely by radial stratification, although its magnitude varies strongly under five views and high noise. Finally, at the strongest tested Si over-scaling ($a = 1.4$), σ_{ens} approaches chance for density, showing that error shared across ensemble members can produce incorrect cross-seed consensus; although $b_\mathcal{E}$ can still localize low-signal, high-error regions in this regime, it requires the GT field and is used only as an oracle diagnostic baseline (Tables 2 and 3).

Beyond correlation, we use a normalized Area Under the Sparsification Error curve nAUSE (Ilg et al., 2018; Poggi et al., 2020) to quantify how well a ranking proxy approximates the oracle ϵ_{mean} (nAUSE is 0 for oracle-level sparsification and 1 for the expected random-reference performance). We find that σ_{ens} (nAUSE 0.11-0.30) is substantially closer to oracle sparsification than the tested baselines. In particular, our inverse-emissivity baseline $b_\mathcal{E}$ is consistently weaker than σ_{ens} in the matched-model conditions (nAUSE 0.47-1.25) and becomes near- or worse-than-random in sparse-view cases. Even in the high-noise regime where $b_\mathcal{E}$ can track the error, σ_{ens} performs better (nAUSE 0.27 vs 0.52 for $\times 9$, right panel in Figure 7). Finally, consistent with spatial correlations, under abundance mismatch σ_{ens} approaches chance (nAUSE $= 0.93$ for density at $a = 1.4$). See Appendix C.2 (Figure 10) for more details on sparsification analysis and temperature field results.

Next, we provide a complementary field-space consistency test by computing the directional per-field energy overlap E_{cap} (Equation 27), or fraction of the squared L^2 norm of the signed field error e (whose magnitude is ϵ_{mean}) that lies within the seed-deviation subspace. These percentages are computed on the full 1.05-3 $R_\odot$ control grid, as opposed to the inner radial band used by localization metrics. We also compute the overlap enrichment relative to the rank-matched isotropic expectation $E_{\text{cap}}^{\text{rand}} = r/M$. Across the three matched-model regimes, shown in Figure 8, the observed rank-9 deviation span (green) captures 4.8-32.1% of the density signed-error energy and 2.4-24.0% of the temperature signed-error energy, strongest under severe view sparsity, and modest under full coverage (all of which are substantially above the isotropic reference of 0.2% in dashed red). The corresponding enrichment ranges between 23-154$\times$ for density and 11-115$\times$ for temperature. We additionally note that even under forward model mismatch, 5-58$\times$ enrichment can persist above the rank-matched isotropic expectation.

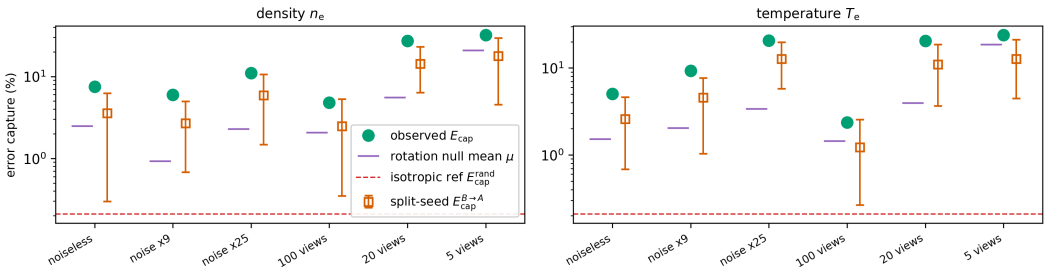

Figure 8: Per-field error-energy capture computed by projecting the error onto ensemble deviation subspaces. Rank-9 seed-deviation spans capture 2.4-32.1% of the signed field error energy (green), which is above the 0.2% isotropic reference (dashed red). Split-seed evaluation, using span and error from disjoint seed halves (orange), shows that the directional overlap persists when the span and target error are constructed from disjoint seed subsets. We additionally find that the original unrotated capture exceeds the longitude-rotation-null mean (purple) in all tested matched-model regimes and lies at the top of the tested rotation distribution in 11 of 12 combinations. Finally, we note that capture is largest under sparse-view and high-noise regimes.

While a high E_{cap} indicates directional alignment, it could be inflated by (i) reusing the same finite ensemble to construct e and the deviation span, or (ii) generic spatial structure. We provide two corresponding controls to test this. First, we split the $K = 10$ ensemble into two 5-seed groups A and B and compute the error from B captured by the field deviation subspace of A ($E_{\text{cap}}^{B\to A}$), repeated over all 252 splits. Figure 8 shows the mean and 2.5th and 97.5th percentile interval of the split-sensitivity distribution in orange: across matched-model conditions, a span estimated from one half of the seeds has a mean capture of the other half's error up to 18% versus 0.09% for a random subspace of equal rank. Thus, the observed alignment is not solely an artifact of using the same finite seed set to construct both quantities. Second, we cyclically rotate the field deviations relative to the signed field error in longitude, preserving latitudinal/radial structure while disrupting their longitudinal alignment. Shown in purple is the mean μ of the error capture computed after the rotations, where the unrotated capture exceeds the mean and lies at the top of the tested distribution in $11/12$ field-condition combinations. The reduced capture indicates that the original longitudinal alignment contributes to the observed overlap beyond generic spatial structure alone. We finally note that the captured energy from the seed-deviation span is a minority of the total error energy but substantially greater than the rank-matched reference expectations (Tables 9 and 10).

To summarize, CoroNeRF reports the ensemble mean and cross-seed instability without access to the true field; correlation, partial correlation, sparsification, and error-capture diagnostics require ground truth and are used only to validate this inference-time proxy. In addition, errors shared by ensemble members are absent from cross-seed instability, so consensus is not a correctness guarantee: we characterized this explicitly under a controlled Si abundance mismatch, where localization degrades under the tested range with a pronounced asymmetry between under and over-scaling. Notably, this relative abundance uncertainty can be reduced through co-elemental line selection, although absolute-intensity inversions remain sensitive to the common elemental scale. Finally, we note that the present line set was selected for informative coronal density and temperature response, and conditioning under arbitrary line selections remains future work.

6 Conclusion

CoroNeRF jointly reconstructs 3D coronal electron density and temperature through a differentiable multiline atomic-emission renderer. In a controlled synthetic testbed, we show that image fidelity is not a sufficient certificate of thermodynamic-field fidelity and evaluate cross-seed instability as a GT-free-at-inference error-ranking proxy. Additionally, increasing angular coverage substantially mitigates sparse-view geometric degeneracies, while increasing spectral coverage mitigates plasma-thermodynamic degeneracies. Future work will assess generalization and calibration across coronal states and noise realizations, extend the framework to time-dependent and spectropolarimetric observations, and evaluate it on DKIST (Rimmele et al., 2020; Schad et al., 2024) or UCoMP (Landi et al., 2016; Tomczyk et al., 2021) data.

REFERENCES

Christoph Angermann, Simon Göppel, and Markus Haltmeier. Uncertainty-aware null space networks for data-consistent image reconstruction, 2023. URL https://arxiv.org/abs/2304.06955.

Andrés Asensio Ramos. Tomographic reconstruction of the solar k-corona using neural fields. *Solar Physics*, 298(11), November 2023. ISSN 1573-093X. doi: 10.1007/s11207-023-02226-2. URL http://dx.doi.org/10.1007/s11207-023-02226-2.

M. Asplund, A. M. Amarsi, and N. Grevesse. The chemical make-up of the sun: A 2020 vision. *Astronomy & Astrophysics*, 653:A141, 2021. ISSN 1432-0746. doi: 10.1051/0004-6361/202140445. URL http://dx.doi.org/10.1051/0004-6361/202140445.

Kenneth P. Dere, G. Del Zanna, P. R. Young, and E. Landi. Chianti—an atomic database for emission lines. xvii. version 10.1: Revised ionization and recombination rates and other updates. *The Astrophysical Journal Supplement Series*, 268(2):52, 2023. ISSN 1538-4365. doi: 10.3847/1538-4365/acec79. URL http://dx.doi.org/10.3847/1538-4365/acec79.

Dere, K. P., Landi, E., Mason, H. E., Monsignori Fossi, B. C., and Young, P. R. Chianti - an atomic database for emission lines. *Astron. Astrophys. Suppl. Ser.*, 125(1):149–173, 1997. doi: 10.1051/aas:1997368. URL https://doi.org/10.1051/aas:1997368.

Berthy T. Feng, Andrew A. Chael, David Bromley, Aviad Levis, William T. Freeman, and Katherine L. Bouman. Dynamic black-hole emission tomography with physics-informed neural fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12511–12521, June 2026. URL https://arxiv.org/abs/2602.08029.

Richard A. Frazin and Paul Janzen. Tomography of the solar corona. ii. robust, regularized, positive estimation of the three-dimensional electron density distribution from lasco-c2 polarized white-light images. *The Astrophysical Journal*, 570(1):408–422, May 2002. ISSN 1538-4357. doi: 10.1086/339572. URL http://dx.doi.org/10.1086/339572.

Richard A. Frazin, Alberto M. Vásquez, and Farzad Kamalabadi. Quantitative, three-dimensional analysis of the global corona with multi-spacecraft differential emission measure tomography. *The Astrophysical Journal*, 701(1):547–560, 2009. ISSN 1538-4357. doi: 10.1088/0004-637x/701/1/547. URL http://dx.doi.org/10.1088/0004-637X/701/1/547.

Kyle Gao, Yina Gao, Hongjie He, Dening Lu, Linlin Xu, and Jonathan Li. Neural radiance fields in 3d vision: A comprehensive review. *Computational Visual Media*, pp. 1–55, 2026. doi: 10.26599/CVM.2026.9450535. URL http://dx.doi.org/10.26599/CVM.2026.9450535.

Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes' rays: Uncertainty quantification for neural radiance fields. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20061–20070. IEEE, 2024. doi: 10.1109/cvpr52733.2024.01896. URL http://dx.doi.org/10.1109/CVPR52733.2024.01896.

Eddy Ilg, Özgün Çiçek, Silvio Galesso, Aaron Klein, Osama Makansi, Frank Hutter, and Thomas Brox. *Uncertainty Estimates and Multi-hypotheses Networks for Optical Flow*, pp. 677–693. Springer International Publishing, 2018. ISBN 9783030012342. doi: 10.1007/978-3-030-01234-2_40. URL http://dx.doi.org/10.1007/978-3-030-01234-2_40.

Robert Jarolim, Benoit Tremblay, Andrés Muñoz-Jaramillo, Kyriaki-Margarita Bintsi, Anna Jungbluth, Miraflor Santos, Angelos Vourlidas, James P. Mason, Sairam Sundaresan, Cooper Downs, and Ronald M. Caplan. Sunerf: 3d reconstruction of the solar euv corona using neural radiance fields. *The Astrophysical Journal Letters*, 961(2):L31, jan 2024. doi: 10.3847/2041-8213/ad12d2. URL https://dx.doi.org/10.3847/2041-8213/ad12d2.

Robert Jarolim, Benoit Tremblay, Andrés Muñoz-Jaramillo, Kyriaki-Margarita Bintsi, Anna Jungbluth, Miraflor Santos, Angelos Vourlidas, James Mason, Sairam Sundaresan, Cooper Downs, and Ronald M. Caplan. 3D Reconstructions of the Solar EUV Corona with Neural Radiance Fields. *Bulletin of the American Astronomical Society*, 57(4), June 2025. URL https://baas.aas.org/pub/2025n4i309p02. AAS 246, Presentation 309.02, SPD IX: Unveiling the Solar Poles.

Robert Jarolim, Martin Sanner, Chia-Man Hung, Emma Stevenson, Hala Lamdouar, Josh Veitch-Michaelis, Ioanna Bouri, Anna Malanushenko, Elena Provornikova, Vít Růžička, and Carlos Urbina-Ortega. Sunerf-cme: Physics-informed neural radiance fields for tomographic reconstruction of coronal mass ejections. *The Astrophysical Journal*, 1004(2):168, 2026. ISSN 1538-4357. doi: 10.3847/1538-4357/ae6e39. URL http://dx.doi.org/10.3847/1538-4357/ae6e39.

Wen Jiang, Boshu Lei, and Kostas Daniilidis. *FisherRF: Active View Selection and Mapping with Radiance Fields Using Fisher Information*, pp. 422–440. Springer Nature Switzerland, October 2024. ISBN 9783031726248. doi: 10.1007/978-3-031-72624-8_24. URL http://dx.doi.org/10.1007/978-3-031-72624-8_24.

Twaha Kabika, Cai Hongsen, Ramadhan Said, Dong Jingxian, Bhahat Zimba, Omari Pinda, Zhang Siyuan, Li Yan, and Hou Wenguang. The application of generative neural radiance fields in medical imaging: A systematic review. *Medical Physics*, 53(2), January 2026. ISSN 2473-4209. doi: 10.1002/mp.70296. URL http://dx.doi.org/10.1002/mp.70296.

M. Kramar and B. Inhester. Inversion of coronal Zeeman and Hanle observations to reconstruct the coronal magnetic field. *Memorie della Società Astronomica Italiana*, 78:120, 2007. URL https://www.hao.ucar.edu/people/phil-judge/collaborators/Plowman/kramar2007.pdf.

M. Kramar, B. Inhester, and S. K. Solanki. Vector tomography for the coronal magnetic field. I. Longitudinal Zeeman effect measurements. *Astronomy & Astrophysics*, 456(2):665–673, September 2006. doi: 10.1051/0004-6361:20064865. URL https://doi.org/10.1051/0004-6361:20064865.

M. Kramar, B. Inhester, H. Lin, and J. Davila. Vector Tomography for the Coronal Magnetic Field. II. Hanle Effect Measurements. *The Astrophysical Journal*, 775(1):25, September 2013. doi: 10.1088/0004-637X/775/1/25. URL http://dx.doi.org/10.1088/0004-637X/775/1/25.

M. Kramar, H. Lin, and S. Tomczyk. Direct Observation of Solar Coronal Magnetic Fields by Vector Tomography of the Coronal Emission Line Polarizations. *Astrophysical Journal Letters*, 819(2): L36, March 2016. doi: 10.3847/2041-8205/819/2/L36. URL http://dx.doi.org/10.3847/2041-8205/819/2/L36.

Maxim Kramar and Haosheng Lin. Validation of linear polarization tomographic inversion for the 3d reconstruction of the solar coronal magnetic fields. *The Astrophysical Journal*, 997(1):69, January 2026. ISSN 1538-4357. doi: 10.3847/1538-4357/ae0f97. URL http://dx.doi.org/10.3847/1538-4357/ae0f97.

Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 6405–6416. Curran Associates Inc., 2017. ISBN 9781510860964. URL https://dl.acm.org/doi/10.5555/3295222.3295387.

E. Landi, S. R. Habbal, and S. Tomczyk. Coronal plasma diagnostics from ground-based observations. *Journal of Geophysical Research (Space Physics)*, 121(9):8237–8249, September 2016. doi: 10.1002/2016JA022598. URL https://doi.org/10.1002/2016JA022598.

Aviad Levis, Pratul P Srinivasan, Andrew A Chael, Ren Ng, and Katherine L Bouman. Gravitationally lensed black hole emission tomography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19841–19850, 2022. URL https://ieeexplore.ieee.org/document/9879282.

Roberto Lionello, Jon A. Linker, and Zoran Mikić. Multispectral emission of the sun during the first whole sun month: Magnetohydrodynamic simulations. *The Astrophysical Journal*, 690(1): 902–912, December 2008. ISSN 1538-4357. doi: 10.1088/0004-637x/690/1/902. URL http://dx.doi.org/10.1088/0004-637X/690/1/902.

Z. Mikić, J. A. Linker, R. Lionello, P. Riley, and V. Titov. Predicting the Structure of the Solar Corona for the Total Solar Eclipse of March 29, 2006. In O. Demircan, S. O. Selam, and B. Albayrak (eds.), *Solar and Stellar Physics Through Eclipses*, volume 370 of *Astronomical Society of the Pacific Conference Series*, pp. 299, May 2007. URL https://ui.adsabs.harvard.edu/scan/manifest/2007ASPC..370..299M.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

Joseph P Molnar and Samuel J Grauer. Flow field tomography with uncertainty quantification using a bayesian physics-informed neural network. *Measurement Science and Technology*, 33 (6):065305, March 2022. ISSN 1361-6501. doi: 10.1088/1361-6501/ac5437. URL http://dx.doi.org/10.1088/1361-6501/ac5437.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):1–15, 2022. ISSN 1557-7368. doi: 10.1145/3528223.3530127. URL http://dx.doi.org/10.1145/3528223.3530127.

Matteo Poggi, Filippo Aleotti, Fabio Tosi, and Stefano Mattoccia. On the uncertainty of self-supervised monocular depth estimation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3224–3234, 2020. URL https://api.semanticscholar.org/CorpusID:218613781.

Thomas R. Rimmele, Mark Warner, Stephen L. Keil, Philip R. Goode, Michael Knölker, Jeffrey R. Kuhn, Robert R. Rosner, Joseph P. McMullin, Roberto Casini, Haosheng Lin, Friedrich Wöger, Oskar von der Lühe, Alexandra Tritschler, Alisdair Davey, Alfred de Wijn, David F. Elmore, André Fehlmann, David M. Harrington, Sarah A. Jaeggli, Mark P. Rast, Thomas A. Schad, Wolfgang Schmidt, Mihalis Mathioudakis, Donald L. Mickey, Tetsu Anan, Christian Beck, Heather K. Marshall, Paul F. Jeffers, Jacobus M. Oschmann, Andrew Beard, David C. Berst, Bruce A. Cowan, Simon C. Craig, Eric Cross, Bryan K. Cummings, Colleen Donnelly, Jean-Benoit de Vanssay, Arthur D. Eigenbrot, Andrew Ferayorni, Christopher Foster, Chriselle Ann Galapon, Christopher Gedrites, Kerry Gonzales, Bret D. Goodrich, Brian S. Gregory, Stephanie S. Guzman, Stephen Guzzo, Steve Hegwer, Robert P. Hubbard, John R. Hubbard, Erik M. Johansson, Luke C. Johnson, Chen Liang, Mary Liang, Isaac McQuillen, Christopher Mayer, Karl Newman, Brialyn Onodera, LeEllen Phelps, Myles M. Puentes, Christopher Richards, Lukas M. Rimmele, Predrag Sekulic, Stephan R. Shimko, Brett E. Simison, Brett Smith, Erik Starman, Stacey R. Sueoka, Richard T. Summers, Aimee Szabo, Louis Szabo, Stephen B. Wampler, Timothy R. Williams, and Charles White. The daniel k. inouye solar telescope – observatory overview. *Solar Physics*, 295(12), December 2020. ISSN 1573-093X. doi: 10.1007/s11207-020-01736-7. URL http://dx.doi.org/10.1007/s11207-020-01736-7.

Thomas Schad and Gabriel Dima. Forward Synthesis of Polarized Emission in Target DKIST Coronal Lines Applied to 3D MURaM Coronal Simulations. *Solar Physics*, 295(7):98, July 2020. doi: 10.1007/s11207-020-01669-1. URL https://ui.adsabs.harvard.edu/abs/2020SoPh..295...98S.

Thomas A. Schad, Gordon J.D. Petrie, Jeffrey R. Kuhn, Andre Fehlmann, Thomas Rimmele, Alexandra Tritschler, Friedrich Woeger, Isabelle Scholl, Rebecca Williams, David Harrington, Alin R. Paraschiv, and Judit Szente. Mapping the sun's coronal magnetic field using the zeeman effect. *Science Advances*, 10(37):eadq1604, 2024. doi: 10.1126/sciadv.adq1604. URL https://www.science.org/doi/abs/10.1126/sciadv.adq1604.

Jianxiong Shen, Adria Ruiz, Antonio Agudo, and Francesc Moreno-Noguer. Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In *2021 International Conference on 3D Vision (3DV)*, pp. 972–981. IEEE, December 2021. doi: 10.1109/3dv53792.2021.00105. URL http://dx.doi.org/10.1109/3DV53792.2021.00105.

Jianxiong Shen, Antonio Agudo, Francesc Moreno-Noguer, and Adria Ruiz. *Conditional-Flow NeRF: Accurate 3D Modelling with Reliable Uncertainty Quantification*, pp. 540–557. Springer Nature Switzerland, 2022. ISBN 9783031200625. doi: 10.1007/978-3-031-20062-5_31. URL http://dx.doi.org/10.1007/978-3-031-20062-5_31.

Southwest Research Institute. 3D Reconstruction of Solar Coronal Temperature and Density Using Neural Radiance Fields, 19-R6509. 2024 Internal R&D Annual Report: Earth & Space, 2024. URL https://www.swri.org/what-we-do/internal-research-development/2024/earth-space/3d-reconstruction-of-solar-coronal-temperature-density-using-neural-radiance-fields-19-r6509. Principal Investigator: Andres Munoz.

Jian Sun, Kristopher Innanen, Tianze Zhang, and Daniel Trad. Implicit seismic full waveform inversion with deep neural representation. *Journal of Geophysical Research: Solid Earth*, 128(3), February 2023. ISSN 2169-9356. doi: 10.1029/2022jb025964. URL http://dx.doi.org/10.1029/2022JB025964.

Niko Sünderhauf, Jad Abou-Chakra, and Dimity Miller. Density-aware nerf ensembles: Quantifying predictive uncertainty in neural radiance fields. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9370–9376, 2023. doi: 10.1109/ICRA48891.2023.10161012. URL https://ieeexplore.ieee.org/document/10161012.

Steven Tomczyk, Enrico Landi, Ben Berkey, Joan Burkepile, Marc Cotter, Dennis Gallaher, Michael D. Galloy, Rob Graves, Philip Oakley, Lisa Perez-Gonzalez, Scott Sewell, Giuliana de Toma, and Patrick Zmarzly. First Images from the Upgraded Coronal Multi-channel Polarimeter (UCoMP). In *AGU Fall Meeting Abstracts*, volume 2021, pp. 2089, December 2021. URL https://ui.adsabs.harvard.edu/abs/2021AGUFMSH15G2089T/abstract.

Xin Wang, Yineng Chen, Shu Hu, Heng Fan, Hongtu Zhu, and Xin Li. Neural radiance fields in medical imaging: A survey, 2024. URL https://arxiv.org/abs/2402.17797.

Chulin Zhao, Yiran Xu, and Shu Liu. When variance is not an error map: Calibrated uncertainty for radiative gaussian splatting in sparse-view ct, 2026. URL https://arxiv.org/abs/2607.13682.

Ellen D. Zhong, Tristan Bepler, Bonnie Berger, and Joseph H. Davis. Cryodrgn: reconstruction of heterogeneous cryo-em structures using neural networks. *Nature Methods*, 18(2):176–185, February 2021. ISSN 1548-7105. doi: 10.1038/s41592-020-01049-4. URL http://dx.doi.org/10.1038/s41592-020-01049-4.

Tao Zhong, Yixun Hu, Dongzhe Zheng, Aditya Sood, and Christine Allen-Blanchette. Neural field thermal tomography: A differentiable physics framework for non-destructive evaluation, 2026. URL https://arxiv.org/abs/2603.11045.

AI use statement. The authors used large language models (ChatGPT and Claude) for finding related works, software development and debugging, checking mathematical rigor, feedback on experimental design and analysis of results, and for editing and proofreading the manuscript. However, all methods, experiments, results, and claims were designed, executed, and verified by the authors, who take full responsibility for the content of this paper.

Reproducibility statement. The authors conducted this study with reproducibility in mind. Various parts of the forward model (Appendix A), hyperparameters (Table 11), and benchmark/ablation specifications are detailed in the Appendix of this manuscript. All parameters associated with each experiment, as well as figures in the text, are documented explicitly in configuration code files. Additionally, the full codebase and experiment configuration files will be publicly available on GitHub.

A Forward Model and Rendering Details

A.1 Ground Truth Corona Cube and Emissivity Table Construction

Our ground truth coronal model is a MAS (Magnetohydrodynamic Algorithm outside a Sphere) cube from Predictive Science Inc. (Mikić et al., 2007; Lionello et al., 2008), based on Carrington Rotation 2283 that occurred around April-May 2024. The corona cube is in (longitude $[0, 2\pi]$, colatitude $[0, \pi]$, radius $[1, 30]$ $\mathrm{R}_\odot$) coordinates with resolution $(299, 142, 154)$. Density and temperature have ranges $\log_{10}(n_\mathrm{e}/\mathrm{cm}^{-3}) \in [2.16, 12.32]$ and $\log_{10}(T_\mathrm{e}/\mathrm{K}) \in [5.75, 6.35]$, respectively.

We use the CHIANTI 10.1 atomic database (Dere, K. P. et al., 1997; Dere et al., 2023) and a Python package `pyCELP` (Schad & Dima, 2020) to construct the emissivity tables for our forward model. Our spectral lines are two density-sensitive pairs of forbidden coronal infrared lines: Fe XIII $1075/1080$ nm and Si IX $2585/3935$ nm. While these are the nominal values used in scientific literature, the actual values used in the database are Fe XIII $1074.7/1079.8$ nm and Si IX $2584.6/3929.3$ nm, respectively. When synthesizing these lines, we assume the elemental abundance described in the 2021 Asplund paper (Asplund et al., 2021).

The emissivity LUT is a function of density n_e, temperature T_e, and heliocentric radius r, and its outputs are in units of erg s^{-1} cm^{-3} sr^{-1}. We retain only the scalar Stokes-I emissivity, neglecting the quadrupole atomic-alignment/depolarization terms associated with the magnetic-field geometry. The LUT axes are chosen from the PSI cube statistics and atomic-response ranges. Queries outside the tabulated $(\log_{10} n_\mathrm{e}, \log_{10} T_\mathrm{e}, \log_{10} r)$ domain are clamped to the nearest LUT boundary before interpolation. In particular, the radial LUT spans approximately 1.001-10 $\mathrm{R}_\odot$, and LOS samples farther than 10 $\mathrm{R}_\odot$ use the 10 $\mathrm{R}_\odot$ radial boundary response. Our grid resolution is $128 \times 64 \times 32$ in $(\log_{10} n_\mathrm{e}, \log_{10} T_\mathrm{e}, \log_{10} r)$. We implement the table in log-space, and we use `PyTorch` to make it tensor-compatible, notably a tensor-product trilinear interpolator in LUT-coordinate space that uses `grid_sample` from `torch.nn.functional`. The trilinear interpolator additionally has a backstop that returns the nearest boundary even if it is given a value outside the bounds, rather than zeros or a linear extrapolation.

This emissivity LUT is used both in (i) synthetic observation generation using the GT density and temperature values from the PSI cube, and (ii) CoroNeRF for rendering the outputs from the neural plasma field. For (i), we trilinearly interpolate the PSI density and temperature values on a spherical grid using the same `grid_sample`, this time with the longitude dimension augmented to support periodic longitude wrapping.

A.2 Sampling Geometry and Numerical Quadrature

CoroNeRF contains objects that live in different coordinate systems. We place the original $[-30, 30]^3$ $\mathrm{R}_\odot$ PSI cube centered in a normalized axis-aligned bounding box (AABB) that has bounds $[-1, 1]^3$ by normalizing all coordinates by 30 $\mathrm{R}_\odot$. Our ray sampler works in this coordinate system.

A view is constructed as follows. Given a selected longitude angle, we place a camera at a heliocentric distance of ≈ 215 $\mathrm{R}_\odot$ (Earth-Sun distance), 0° latitude, pointed towards the origin. We use a perspective camera model: our rays start from the observer and span a $[-3, 3]$ $\mathrm{R}_\odot$ FOV, and each ray forms one pixel on the image plane. At an observer distance of ≈ 215 $\mathrm{R}_\odot$, perspective rays are nearly parallel across the $6\mathrm{R}_\odot$ FOV. We then mask sightlines with impact parameter $< 1R_\odot$ to mimic a coronagraph, leaving only a set of coronal pixels to render. Given a pixel to render, we shoot a ray through the AABB, compute the near and far bounds as intersections with the AABB, and compute emissivities at uniform samples along the ray with step size $1/256$ (the number of samples depends on the ray's intersection length with the AABB). To do so at a given sample (in AABB coordinates), we query the neural plasma field f_θ to get density and temperature predictions $(\log_{10} n_\mathrm{e}, \log_{10} T_\mathrm{e})$ and use our LUT to compute the emissivity at that point. Note that the LUT takes in the heliocentric radius $r = ||\mathbf{x}||$, computed from physical Cartesian coordinates $\mathbf{x}$. Next, we compute the differential length segment corresponding to each sample point, by scaling our coordinates from AABB to $R_\odot$ to cm. We additionally convert the raw LUT emissivity from erg to photon counts, and from steradian^{-1} to arcsec^{-2}, so that the integrand ϵ_c has units photons s^{-1} cm^{-3} arcsec^{-2}, and our final intensity observations after integrating along the ray are in units of photons s^{-1} cm^{-2} arcsec^{-2}. The LOS integration uses a midpoint Riemann sum.

step size	Δs ($R_\odot$)	rel-L2 vs finest	asinhErr vs finest
$1/128$	0.234	20.6%	1.0×10^{-2}
$\mathbf{1/256}$ (production)	0.117	7.5%	1.7×10^{-3}
$1/512$	0.059	2.3%	5.7×10^{-4}
$1/1024$	0.029	— (ref)	— (ref)

Table 1: Convergence table for selected renderer step sizes. For each step size, we render eight views using the GT cube and compute relative L2 and asinhErr compared to the finest resolution.

We assess the renderer convergence by re-rendering the GT forward model at integration steps $\Delta s \in \{1/128, 1/256, 1/512, 1/1024\}$ (AABB coordinates) on eight views and comparing to the finest resolution (Table 1). The images converge at roughly first-to-second order, and at production step ($1/256 \approx 0.117$ $R_\odot$) they differ from a $4\times$-finer render by 7.5% in relative L2 but only 1.7×10^{-3} in the asinh image metric, so production-resolution renders are close to the finer reference under the optimization metric. However, we do not claim discretization-independent field recovery. Finally, for synthetic dataset generation, the ray sampling and integral procedures are identical, except we use the oracle PSI density and temperature fields instead of the neural plasma field. These fields live on a spherical coordinate system, so we transform AABB to spherical coordinates before sampling for density and temperature values using spherical trilinear field interpolators.

A.3 LOS VISUAL PROJECTIONS

For visualizing the field, we use rendered field-space projections for intuition, and reserve latitude-longitude plots at fixed shell radii for metric evaluation.

First, our LOS integration support for visualization is $[1.1, 2.5]$ $R_\odot$. We reuse the ray geometry notation q and define the column density and emissivity-weighted temperature with LOS integrals analogous to Equation 1:

$$N_{\mathrm{e}}(q) = \int_{\mathcal{S}(q)} n_{\mathrm{e}}\big(\mathbf{x}_q(t)\big)\, dt, \tag{3}$$

$$\langle T_{\mathrm{e}} \rangle_{\epsilon}(q) = \frac{\displaystyle\int_{\mathcal{S}(q)} T_{\mathrm{e}}\big(\mathbf{x}_q(t)\big)\, \mathcal{E}\big(\mathbf{x}_q(t)\big)\, dt}{\displaystyle\int_{\mathcal{S}(q)} \mathcal{E}\big(\mathbf{x}_q(t)\big)\, dt}. \tag{4}$$

Here, we define a total summed emissivity $\mathcal{E}(\mathbf{x})$ that accounts for all the observed spectral channels

$$\mathcal{E}(\mathbf{x}) = \sum_{c=1}^{C} \epsilon_c\big(n_{\mathrm{e}}(\mathbf{x}),\, T_{\mathrm{e}}(\mathbf{x}),\, r(\mathbf{x})\big). \tag{5}$$

We choose the column density because it is a real physical quantity: the total electrons accumulated per cm^2. On the other hand, temperature is not additive, and thus we choose to compute a weighted representative temperature based on the total emission along the LOS.

A.4 OBSERVATIONAL NOISE AND FORWARD MODEL MISMATCH

We additionally add heteroscedastic Gaussian noise to our images: for a given channel c and pixel i, the observed intensity is drawn from

$$I_{i,c}^{\mathrm{obs}} \sim \mathcal{N}(I_{i,c}^{\mathrm{true}}, \eta\alpha_c I_{i,c}^{\mathrm{true}} + \beta_c^2), \tag{6}$$

where α_c is a channel-dependent coefficient for the signal-dependent shot-like variance term, and β_c is the background/read-noise floor. We scale the noise by multiplying α_c by η, a variance multiplier ($\eta = 9$ means a $\times 3$ increase in standard deviation). Next, to introduce forward model mismatch, we

define the truth $I_{i,c}^{\text{true}}$ to be a scaled version of the raw forward model output $I_{i,c}^{\text{raw}}$:

$$I_{i,c}^{\text{true}} = \begin{cases} aI_{i,c}^{\text{raw}}, & c \in \{\text{Si IX channels}\}, \\ I_{i,c}^{\text{raw}}, & c \in \{\text{Fe XIII channels}\}. \end{cases} \tag{7}$$

where a is an abundance-mismatch scale applied to the Si IX channels before noise is added. This mismatch models a controlled family of multiplicative channel-wise abundance discrepancies, in particular a relative Fe-to-Si abundance scaling, where $a = 1$ means no forward model error (matched-model). Our implementation scales both Si IX channels by the same value while the Fe XIII channels remain at one, so we refer to a specifically as a Si scale.

In practice, we select α_c such that radial bands across spectral channels achieve approximately the same SNR, reflecting slightly different per-channel noise levels as a real instrument would have. We set the background standard deviation floor to $\beta_c = 0$. For $\eta = 0$ specifically, we bypass this noise generation process entirely. Finally, we note that observation noise is applied to training views only, and held-out image-space evaluation uses corresponding noise-free synthetic views.

A.5 Multiresolution Hash Grid Model

Our multiresolution hash-grid representation is based on the encoding described by (Müller et al., 2022), where we encode spatial information as trainable grid features, and then separate decoder MLP heads combine the features into density and temperature estimates. The grids themselves do not store features directly at each spatial point, as this does not scale with resolution. The grid is indexed by hashing the spatial coordinate of the point, where we define the size as 2^G by F: G is the hashmap table size in $\log_2$, and F is the number of features. The full hash grid contains many such levels, each associated with a physical spatial resolution. Given a sample position, the features at that point are computed by trilinearly interpolating the features at the eight corners of the cube surrounding that point. The positions of those eight corners are computed relative to the fixed spatial resolution of that level. The coordinates are then hashed and used in the trainable hashmap to retrieve the features.

For a multiresolution hash grid with L levels, we define the coarsest (base) resolution as N_0 and the finest resolution as N_{L-1}. We then spread the resolution geometrically, where the resolution at level ℓ is given by

$$N_\ell = N_0 b^\ell, \quad b = \exp\left(\frac{\ln(N_{L-1}/N_0)}{L-1}\right). \tag{8}$$

Our hash function is a bitwise XOR hash. While hashing integer coordinates may introduce many collisions, we use multiple levels, each with a different physical spatial resolution. Thus, the multiresolution nature of our grid not only encodes information at different spatial levels, but also mitigates the effect of collisions once we concatenate the features at all levels.

In our implementation, each level has a hash table with 2^{20} entries and $F = 2$ features per entry, with a total of $L = 20$ levels that span a spatial resolution going from $N_0 = 16$ (coarsest) to $N_{L-1} = 512$ (finest). Our decoder heads are each two hidden layers of width 64.

During training (per minibatch of rays), because our forward model is differentiable with respect to the neural plasma field (and thus the model parameters), the gradients first backpropagate through the differentiable line-emission renderer, and then through the emissivity lookup table to the predicted $\log_{10} n_{\text{e}}$ and $\log_{10} T_{\text{e}}$ fields, and finally to the multiresolution hash grid. The lookup table itself and ray geometry and sampling are fixed during optimization.

B METRICS

Our evaluations are conducted in two spaces. *Image-space metrics* compare rendered line-intensity images to held-out observations. *Field-space metrics* compare the reconstructed density and temperature fields to their ground-truth counterparts on spherical shells. Field-space metrics are the primary metrics in this work as the goal is a latent joint physical-field recovery rather than image synthesis.

Notation. For any finite index set $\mathcal{A}$, we write

$$\langle f(a)\rangle_{a\in\mathcal{A}} = \frac{1}{|\mathcal{A}|}\sum_{a\in\mathcal{A}} f(a) \tag{9}$$

for the empirical average. We also use the same indexing as in Section 3 of the main text: $v \in \mathcal{V} = \{1,\dots,V\}$ indexes viewpoints, $p \in \Omega_v$ indexes valid pixels in view v, and $c \in \mathcal{C} = \{1,\dots,C\}$ indexes spectral line channels.

B.1 IMAGE-SPACE METRICS

Let $I_{v,p,c}$ and $\hat{I}_{v,p,c}$ denote the observed and predicted line intensities for view v, pixel p, and channel c. The per-view, per-channel mean-squared error is

$$\mathrm{MSE}_{v,c}(I) = \left\langle (\hat{I}_{v,p,c} - I_{v,p,c})^2 \right\rangle_{p\in\Omega_v}. \tag{10}$$

The per-view, per-channel PSNR is

$$\mathrm{PSNR}_{v,c}(I) = 10\log_{10}\left(\frac{R_{v,c}^2}{\mathrm{MSE}_{v,c}(I)}\right), \tag{11}$$

where $R_{v,c}$ is the 1st-to-99th-percentile target-intensity range for that view and channel, with max-minus-min as a fallback if the percentile range is degenerate. Dataset-level image-space metrics are obtained by averaging over views and channels:

$$\mathrm{PSNR}(I) = \langle \mathrm{PSNR}_{v,c}(I)\rangle_{v\in\mathcal{V},c\in\mathcal{C}}. \tag{12}$$

Next, our fixed-scale asinh image error is defined as follows:

$$\mathrm{asinhErr}(I) = \left\langle \left| \operatorname{asinh}\left(\frac{\hat{I}_{v,p,c}}{s_c}\right) - \operatorname{asinh}\left(\frac{I_{v,p,c}}{s_c}\right)\right| \right\rangle_{p\in\Omega_v,\ v\in\mathcal{V},\ c\in\mathcal{C}}, \tag{13}$$

where s_c is the channel-dependent intensity scale specified by the evaluation protocol (in the noise-view study, held-out asinhErr uses the common base per-channel scales). This metric is well-defined for signed intensities and behaves logarithmically for large intensity magnitudes.

B.2 FIELD-SPACE METRICS

Field metrics are computed on spherical shells. Let $r \in \mathcal{R}$ index evaluation shell radii, and let $s \in \mathcal{S}_r$ index valid samples on the shell at radius r. We use m to denote a density or temperature log-field

$$m \in \{\log_{10} n_{\mathrm{e}},\ \log_{10} T_{\mathrm{e}}\}. \tag{14}$$

Let $m_{r,s}$ and $\hat{m}_{\theta,r,s}$ be the ground-truth and predicted log-field values. For a radial band $\mathcal{R}_{\mathrm{band}} \subseteq \mathcal{R}$, we define the mean absolute log error

$$\mathrm{MAE}_{\mathrm{band}}(m) = \langle |\hat{m}_{\theta,r,s} - m_{r,s}|\rangle_{s\in\mathcal{S}_r,\ r\in\mathcal{R}_{\mathrm{band}}}. \tag{15}$$

We also report the absolute relative error

$$\text{AbsRel}_{\text{band}}(m) = \left\langle \frac{|10^{\hat{m}_{\theta,r,s}} - 10^{m_{r,s}}|}{\max(|10^{m_{r,s}}|, \epsilon_0)} \right\rangle_{s\in\mathcal{S}_r,\ r\in\mathcal{R}_{\text{band}}}, \tag{16}$$

where $\epsilon_0 = 10^{-30}$ is a small numerical floor.

Finally, for model-mismatch studies, the sign of the error is also important, so we additionally report signed shell mean error:

$$\text{ME}_{\text{band}}(m) = \langle \hat{m}_{\theta,r,s} - m_{r,s} \rangle_{s\in\mathcal{S}_r,\ r\in\mathcal{R}_{\text{band}}}. \tag{17}$$

Unlike absolute error metrics, this signed metric indicates whether a reconstructed field systematically overestimates or underestimates the ground truth field.

C Ensemble Uncertainty Quantification

C.1 Cross-Seed Instability and Correlation Analysis

We denote a plasma quantity as $m \in \{\log_{10} n_e, \log_{10} T_e\}$. Every definition below is per-m for readability (fields are in dex), and voxels are indexed by position $\mathbf{x}$. Let us reuse the shell and radial notation from Appendix B: $\mathcal{S}_r = \{\mathbf{x} : r(\mathbf{x}) = r\}$ as a radial shell, and $\mathcal{R}_{[r_1,r_2]} = \{\mathbf{x} : r_1 \le r(\mathbf{x}) \le r_2\}$ as a radial band from r_1 to r_2. Finally, let $m^*(\mathbf{x})$ be the ground truth discretized field.

We train K models, with fixed model architecture, observational protocol (η, V, Λ, a), and training dataset, only varying the seeds $k = 1 \dots K$ (Lakshminarayanan et al., 2017). In particular, seed k gives a predicted field $\hat{m}_k(\mathbf{x})$. Define the ensemble mean and cross-seed instability across the seeds, respectively, as follows:

$$\overline{m}(\mathbf{x}) = \frac{1}{K}\sum_k \hat{m}_k(\mathbf{x}), \tag{18}$$

$$\sigma_{\text{ens}}(\mathbf{x}) = \left[\frac{1}{K-1}\sum_k (\hat{m}_k(\mathbf{x}) - \overline{m}(\mathbf{x}))^2\right]^{1/2}. \tag{19}$$

σ_{ens} is a GT-free-at-inference empirical metric that is a field-error localization proxy. Large σ_{ens} identifies regions where the recovered field is sensitive to training seeds under the fixed pipeline, while small σ_{ens} identifies regions of cross-seed consensus. On the other hand, the absolute ensemble mean error, which can never be known without access to the GT field m^*, quantifies the error of the ensemble-mean field:

$$\epsilon_{\text{mean}}(\mathbf{x}) = |\overline{m}(\mathbf{x}) - m^*(\mathbf{x})|. \tag{20}$$

Note that ϵ_{mean} is not a statistical bias, which would require an expectation over multiple noise realizations.

Finally, to quantitatively express localization correlation, let us use error target ϵ_{mean} and spatial domain $\Omega \in \{\mathcal{S}_r, \mathcal{R}_{[r_1,r_2]}\}$. Define the Spearman rank correlation coefficient between the cross-seed instability and the error over Ω as

$$\rho_\epsilon = \text{Spearman}_{x\in\Omega}(\sigma_{\text{ens}}, \epsilon_{\text{mean}}) = \text{corr}(\text{rank}\,\sigma_{\text{ens}}, \text{rank}\,\epsilon_{\text{mean}}). \tag{21}$$

We note that our ensemble-localization study uses 300 evenly-spaced training views with a disjoint 30-view interleaved evenly-spaced holdout. All members within each condition share the same split, and for noisy conditions they share the same fixed noise realization. Figure 9 shows sample ensemble mean $\overline{m}(\mathbf{x})$, ensemble-mean error $\epsilon_{\text{mean}}(\mathbf{x})$, and cross-seed instability $\sigma_{\text{ens}}(\mathbf{x})$ panels for our canonical noisy condition.

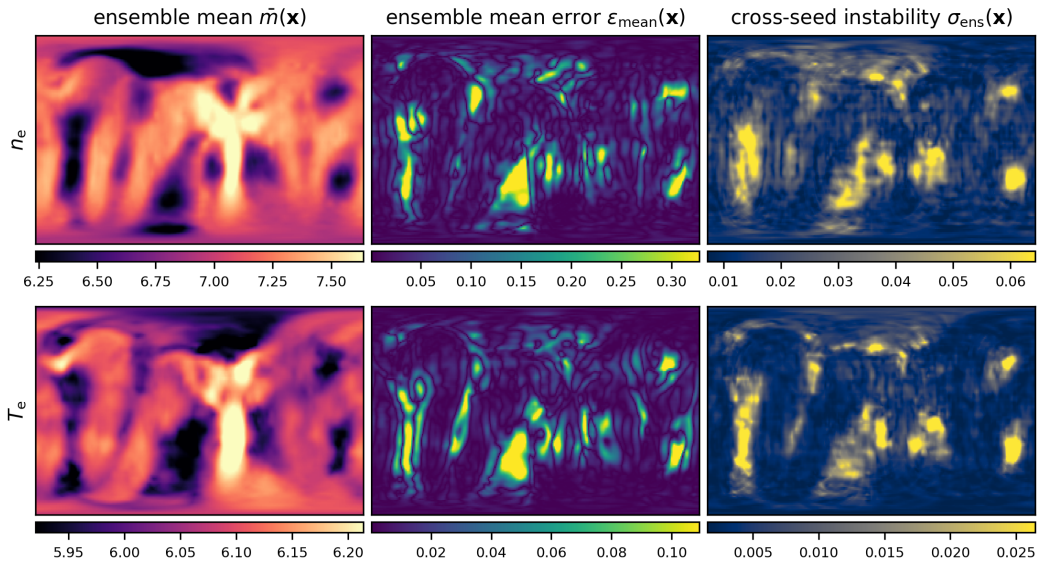

Figure 9: Ensemble reconstruction panels of the 3D density (top) and temperature (bottom) fields at $1.51\text{R}_\odot$ in the high-noise regime ($\eta = 9, V = 300$, all four lines, $a = 1$). We show the ensemble mean $\overline{m}(\mathbf{x})$ (Equation 18), ensemble-mean error $\epsilon_{\text{mean}}(\mathbf{x})$ (unobservable, Equation 20), and cross-seed instability $\sigma_{\text{ens}}(\mathbf{x})$ (Equation 19).

Next, our baseline signal-strength proxy is the inverse-emissivity proxy, given by $b_{\mathcal{E}} = 1/\mathcal{E}$, where $\mathcal{E} > 0$ is the total local emissivity summed over all spectral channels (Equation 5). We use this as a baseline signal-strength confound: low-emissivity regions are a plausible confound because they are more weakly constrained and may exhibit larger reconstruction error. We note that in mismatch conditions we intentionally retain the nominal, unscaled GT emissivity, so $b_{\mathcal{E}}$ remains an oracle probe of the underlying physical signal rather than an observation-consistent proxy.

We then compute per-condition pairwise correlations between the cross-seed instability σ_{ens}, the ensemble mean error ϵ_{mean}, and our emissivity proxy $b_{\mathcal{E}}$, to answer the following question: does the rank association between σ_{ens} and ϵ_{mean} remain after controlling for the oracle inverse-emissivity proxy? To do so, we control for $b_{\mathcal{E}}$ by rank-transforming each variable and computing the standard partial Pearson correlation on those ranks:

$$\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}} \,|\, b_{\mathcal{E}}) = \frac{\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}}) - \rho(\sigma_{\text{ens}}, b_{\mathcal{E}})\rho(b_{\mathcal{E}}, \epsilon_{\text{mean}})}{\sqrt{(1-\rho^2(\sigma_{\text{ens}}, b_{\mathcal{E}}))(1-\rho^2(b_{\mathcal{E}}, \epsilon_{\text{mean}}))}}. \tag{22}$$

A strong positive partial correlation indicates a nontrivial linear association between the rank-transformed σ_{ens} and ϵ_{mean}, after linear adjustment for the rank-transformed $b_{\mathcal{E}}$ proxy (in other words, suggesting that σ_{ens} captures error structure beyond what is explained by signal strength alone).

Tables 2 (density) and 3 (temperature) show pairwise correlations between σ_{ens}, ϵ_{mean}, and $b_{\mathcal{E}}$, as well as the partial correlation $\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}} \,|\, b_{\mathcal{E}})$, in four different regimes: noiseless (300v, 100v), sparse-view (20v, 5v), high-noise ($\times 9$, $\times 25$), and forward model abundance mismatch (Si $\times 0.6$, $\times 0.8$, $\times 1.2$, $\times 1.4$).

We evaluate the diagnostics on a $30 \times 18 \times 8$ longitude-latitude-radius grid, with the radial coordinates spanning 1.05-3 $\text{R}_\odot$. To assess within-scene resampling variability, we treat each longitude slice, containing its associated latitude-radius samples, as one cluster. Resampling entire slices preserves their internal spatial dependence but does not explicitly model dependence between neighboring longitude slices. For each of 2000 bootstrap replicates, we draw 30 longitude indices independently with replacement from the 30 available indices and concatenate the corresponding slices. The same sampled indices are applied to σ_{ens}, ϵ_{mean}, and $b_{\mathcal{E}}$. We recompute the correlation, partial corre-

Condition	$\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}}) \uparrow$	$\rho(b_{\mathcal{E}}, \epsilon_{\text{mean}})$	partial $\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}} \mid b_{\mathcal{E}}) \uparrow$	$\rho(\sigma_{\text{ens}}, b_{\mathcal{E}})$
noiseless (300v)	$0.761^{+0.022}_{-0.029}$	$0.072^{+0.052}_{-0.058}$	$0.759^{+0.022}_{-0.027}$	0.081
100 views	$0.773^{+0.021}_{-0.027}$	$0.085^{+0.050}_{-0.056}$	$0.772^{+0.020}_{-0.025}$	0.086
20 views	$0.778^{+0.029}_{-0.034}$	$-0.019^{+0.045}_{-0.056}$	$0.780^{+0.029}_{-0.035}$	0.038
5 views	$0.712^{+0.042}_{-0.050}$	$-0.103^{+0.088}_{-0.094}$	$0.708^{+0.043}_{-0.051}$	-0.161
Noise $\times 9$	$0.572^{+0.041}_{-0.052}$	$0.405^{+0.038}_{-0.042}$	$0.447^{+0.052}_{-0.059}$	0.621
Noise $\times 25$	$0.555^{+0.039}_{-0.055}$	$0.438^{+0.036}_{-0.044}$	$0.393^{+0.052}_{-0.059}$	0.663
Si $\times 0.6$	$0.438^{+0.073}_{-0.082}$	$0.167^{+0.060}_{-0.063}$	$0.417^{+0.079}_{-0.087}$	0.229
Si $\times 0.8$	$0.387^{+0.067}_{-0.071}$	$0.209^{+0.056}_{-0.068}$	$0.359^{+0.069}_{-0.070}$	0.207
Si $\times 1.2$	$0.166^{+0.067}_{-0.071}$	$0.435^{+0.055}_{-0.061}$	$0.070^{+0.061}_{-0.061}$	0.243
Si $\times 1.4$	$-0.017^{+0.081}_{-0.087}$	$0.476^{+0.049}_{-0.055}$	$-0.082^{+0.070}_{-0.079}$	0.115

Table 2: (Density) Pairwise Spearman correlations between cross-seed instability σ_{ens}, ensemble mean error ϵ_{mean}, and emissivity proxy $b_{\mathcal{E}}$.

Condition	$\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}}) \uparrow$	$\rho(b_{\mathcal{E}}, \epsilon_{\text{mean}})$	partial $\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}} \mid b_{\mathcal{E}}) \uparrow$	$\rho(\sigma_{\text{ens}}, b_{\mathcal{E}})$
noiseless (300v)	$0.726^{+0.025}_{-0.027}$	$0.001^{+0.067}_{-0.066}$	$0.732^{+0.024}_{-0.027}$	0.122
100 views	$0.732^{+0.019}_{-0.023}$	$0.014^{+0.063}_{-0.068}$	$0.735^{+0.019}_{-0.022}$	0.115
20 views	$0.717^{+0.031}_{-0.034}$	$0.092^{+0.053}_{-0.055}$	$0.716^{+0.029}_{-0.032}$	0.061
5 views	$0.611^{+0.034}_{-0.041}$	$0.015^{+0.049}_{-0.057}$	$0.626^{+0.032}_{-0.040}$	-0.189
Noise $\times 9$	$0.590^{+0.044}_{-0.046}$	$0.325^{+0.046}_{-0.052}$	$0.521^{+0.046}_{-0.049}$	0.579
Noise $\times 25$	$0.601^{+0.042}_{-0.049}$	$0.382^{+0.045}_{-0.051}$	$0.503^{+0.047}_{-0.058}$	0.624
Si $\times 0.6$	$0.424^{+0.068}_{-0.073}$	$0.317^{+0.058}_{-0.067}$	$0.361^{+0.075}_{-0.072}$	0.312
Si $\times 0.8$	$0.448^{+0.062}_{-0.065}$	$0.368^{+0.046}_{-0.052}$	$0.365^{+0.065}_{-0.064}$	0.356
Si $\times 1.2$	$0.250^{+0.068}_{-0.071}$	$0.228^{+0.069}_{-0.069}$	$0.226^{+0.067}_{-0.065}$	0.141
Si $\times 1.4$	$0.096^{+0.069}_{-0.063}$	$0.138^{+0.066}_{-0.066}$	$0.101^{+0.063}_{-0.057}$	-0.027

Table 3: (Temperature) Pairwise Spearman correlations between cross-seed instability σ_{ens}, ensemble mean error ϵ_{mean}, and emissivity proxy $b_{\mathcal{E}}$.

lation, and nAUSE (Appendix C.2) on each resample and report the 2.5th and 97.5th percentiles. Finally, correlation and sparsification calculations retain only samples in the inner radial band, and all statistics use unweighted samples rather than spherical area or volume weighting.

In Tables 4 (density) and 5 (temperature), we show within-shell Spearman correlation $\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}})$ between cross-seed instability and absolute ensemble mean error, with the partial correlation $\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}} \mid b_{\mathcal{E}})$ in brackets. We compute these correlations for 3 shells (columns) at R= $[1.33, 1.61, 1.89]$ $\text{R}_{\odot}$, each with 540 voxels on a controlled grid. Within-shell values track the pooled correlation for the selected conditions, showing that the correlation is not solely explained by a radial trend.

Condition	pooled	$r = 1.33\ \mathrm{R}_\odot$	$r = 1.61\ \mathrm{R}_\odot$	$r = 1.89\ \mathrm{R}_\odot$
noiseless (300v)	0.76 [0.76]	0.74 [0.74]	0.77 [0.75]	0.76 [0.75]
100 views	0.77 [0.77]	0.75 [0.75]	0.78 [0.77]	0.77 [0.75]
20 views	0.78 [0.78]	0.78 [0.78]	0.78 [0.78]	0.76 [0.73]
5 views	0.71 [0.71]	0.68 [0.68]	0.75 [0.75]	0.61 [0.60]
Noise $\times 9$	0.57 [0.45]	0.59 [0.58]	0.54 [0.44]	0.41 [0.20]
Noise $\times 25$	0.56 [0.39]	0.53 [0.51]	0.54 [0.41]	0.38 [0.15]
Si $\times 0.6$	0.44 [0.42]	0.47 [0.50]	0.46 [0.44]	0.37 [0.33]
Si $\times 0.8$	0.39 [0.36]	0.39 [0.43]	0.40 [0.37]	0.37 [0.28]
Si $\times 1.2$	0.17 [0.07]	0.12 [0.14]	0.13 [0.09]	0.17 [0.08]
Si $\times 1.4$	-0.02 [-0.08]	-0.09 [-0.02]	-0.04 [-0.01]	-0.04 [-0.05]

Table 4: (Density) Per-shell correlation $\rho(\sigma_{\mathrm{ens}}, \epsilon_{\mathrm{mean}})$ (partial $\rho(\sigma_{\mathrm{ens}}, \epsilon_{\mathrm{mean}} \,|\, b_{\mathcal{E}})$ in brackets), for selected conditions.

Condition	pooled	$r = 1.33\ \mathrm{R}_\odot$	$r = 1.61\ \mathrm{R}_\odot$	$r = 1.89\ \mathrm{R}_\odot$
noiseless (300v)	0.73 [0.73]	0.70 [0.70]	0.72 [0.70]	0.77 [0.75]
100 views	0.73 [0.74]	0.73 [0.73]	0.70 [0.68]	0.76 [0.75]
20 views	0.72 [0.72]	0.72 [0.73]	0.67 [0.65]	0.75 [0.72]
5 views	0.61 [0.63]	0.70 [0.72]	0.50 [0.50]	0.53 [0.51]
Noise $\times 9$	0.59 [0.52]	0.66 [0.66]	0.54 [0.45]	0.49 [0.30]
Noise $\times 25$	0.60 [0.50]	0.61 [0.61]	0.55 [0.44]	0.50 [0.29]
Si $\times 0.6$	0.42 [0.36]	0.30 [0.31]	0.38 [0.30]	0.50 [0.43]
Si $\times 0.8$	0.45 [0.36]	0.35 [0.35]	0.43 [0.30]	0.52 [0.38]
Si $\times 1.2$	0.25 [0.23]	0.30 [0.33]	0.18 [0.19]	0.24 [0.21]
Si $\times 1.4$	0.10 [0.10]	0.14 [0.20]	0.08 [0.12]	0.06 [0.07]

Table 5: (Temperature) Per-shell correlation $\rho(\sigma_{\mathrm{ens}}, \epsilon_{\mathrm{mean}})$ (partial $\rho(\sigma_{\mathrm{ens}}, \epsilon_{\mathrm{mean}} \,|\, b_{\mathcal{E}})$ in brackets), for selected conditions.

C.2 Sparsification Analysis

A useful error-ranking proxy should allow the average error to improve by systematically discarding voxels it flags as least trustworthy (sparsification process). A metric we use is the Area Under the Sparsification Error curve (AUSE, Ilg et al. (2018); Poggi et al. (2020)), a standard protocol used to assess neural-field uncertainty quantification (Shen et al., 2022; Sünderhauf et al., 2023; Goli et al., 2024). To construct this, we pick a ranking scalar z (such as σ_{ens}, $b_{\mathcal{E}}$, ϵ_{mean} (oracle), random), and sort all voxels by decreasing z. We note that the random baseline is the exact expected uniform-retention curve (the global mean error), not a finite permutation average. For a total of N voxels, we then parameterize the process of removing voxels via $K_z(t)$, the subset retaining approximately the

lowest-ranked $(1-t)$ fraction of voxels. We then define the risk as the mean error over the kept set

$$\mathrm{Risk}_z(t) = \frac{1}{|K_z(t)|} \sum_{i \in K_z(t)} \epsilon_{\mathrm{mean,i}}. \tag{23}$$

Plotting $\mathrm{Risk}_z(t)$ against t gives the sparsification curve. At $t = 0$, all curves retain all voxels. The oracle $z = \epsilon_{\mathrm{mean}}$ decreases the fastest as it removes the largest errors first, while a random ranking keeps the MAE relatively flat. A useful ranking proxy should approach the oracle, while arbitrary proxies may perform similar to or worse than the random baseline.

To summarize each curve with a single metric, we first compute the sparsification error

$$\mathrm{SE}_z(t) = \mathrm{Risk}_z(t) - \mathrm{Risk}_{\mathrm{oracle}}(t) \geq 0, \tag{24}$$

which is the gap to the oracle at each level, and then compute the area under the curve

$$\mathrm{AUSE}_z = \int_0^{t_{\max}} \mathrm{SE}_z(t) dt, \tag{25}$$

where we integrate up to $t_{\max} = 0.95$ to avoid a noisy terminal spike when few voxels remain. In practice, this is done using a discretized t-grid. A lower AUSE indicates a ranking closer to that of the oracle ($\mathrm{AUSE} = 0$). Finally, to compare across conditions, we normalize the AUSE by a random ranking, such that 0 gives perfect, while 1 gives no-better-than-random, and > 1 gives worse-than-random:

$$\mathrm{nAUSE}_z = \frac{\mathrm{AUSE_z}}{\mathrm{AUSE}_{\mathrm{random}}}. \tag{26}$$

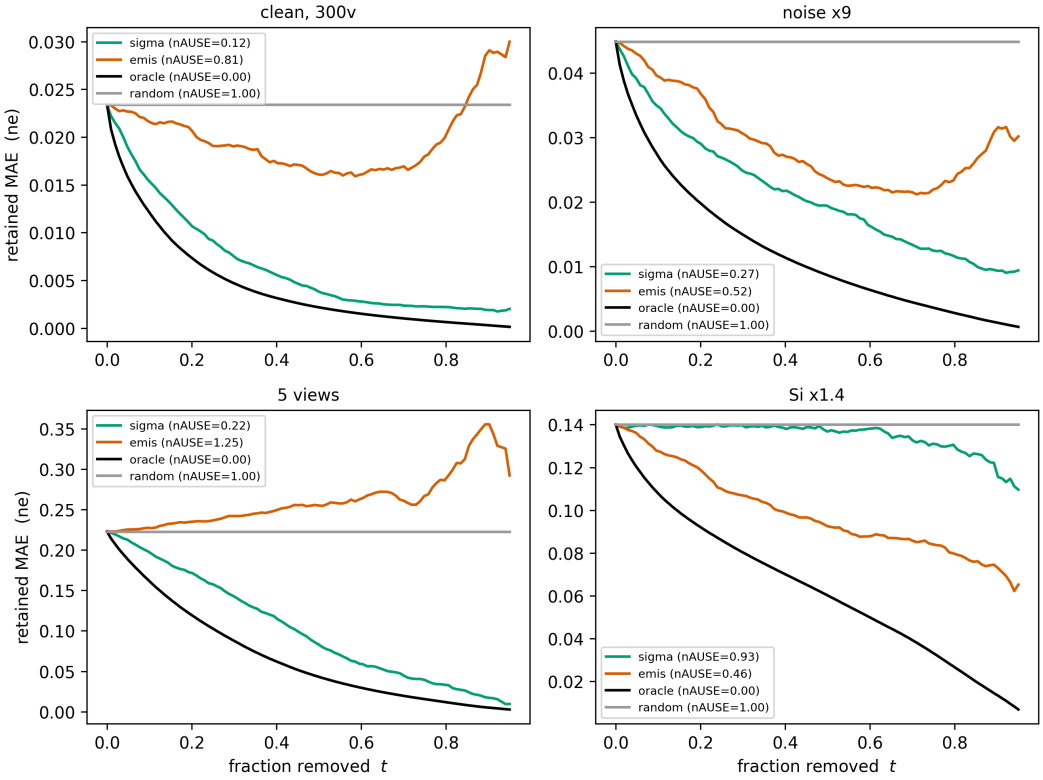

Figure 10: Density-error sparsification curves for the noiseless, high-noise, sparse-view, and Si-mismatch regimes. Lower remaining MAE is better, and the random reference is the exact expected uniform-retention curve.

Figure 10 shows the sparsification curves for four main regimes: noiseless, high-noise, sparse-view, and forward-model mismatch. In each plot, we show curves for various ranking scalars z: the cross-seed instability (green), emissivity proxy (orange), oracle (black), and random (gray). We find that across the matched-model noiseless, high-noise, and sparse-view regimes, the cross-seed instability tracks the oracle closely and much better than our physical emissivity proxy, and degrades towards random ranking under the strongest tested forward-model mismatch. Tables 6 (density) and 7 (temperature) tabulate the values in the curves, along with the raw AUSE values. The nAUSE intervals are 95% within-scene longitude-slice bootstrap intervals using the same 2000-resample procedure. Note that raw AUSE combines the scale of reconstruction error with ranking quality, so we therefore use nAUSE for comparisons of relative ranking performance across conditions and report raw AUSE for reference.

Condition	AUSE_σ	AUSE_b	nAUSE_σ	nAUSE_b
noiseless (300v)	0.0021	0.0145	$0.115^{+0.019}_{-0.012}$	$0.809^{+0.120}_{-0.096}$
100 views	0.0021	0.0147	$0.113^{+0.016}_{-0.012}$	$0.794^{+0.115}_{-0.094}$
20 views	0.0062	0.0657	$0.117^{+0.026}_{-0.020}$	$1.239^{+0.101}_{-0.102}$
5 views	0.0327	0.1834	$0.222^{+0.043}_{-0.035}$	$1.245^{+0.139}_{-0.135}$
Noise $\times 9$	0.0085	0.0161	$0.274^{+0.047}_{-0.036}$	$0.520^{+0.075}_{-0.063}$
Noise $\times 25$	0.0108	0.0168	$0.304^{+0.056}_{-0.044}$	$0.472^{+0.067}_{-0.055}$
Si $\times 0.6$	0.0676	0.1383	$0.408^{+0.074}_{-0.062}$	$0.834^{+0.102}_{-0.097}$
Si $\times 0.8$	0.0312	0.0579	$0.416^{+0.060}_{-0.050}$	$0.772^{+0.107}_{-0.096}$
Si $\times 1.2$	0.0302	0.0227	$0.610^{+0.072}_{-0.060}$	$0.459^{+0.063}_{-0.047}$
Si $\times 1.4$	0.0676	0.0335	$0.932^{+0.098}_{-0.084}$	$0.462^{+0.060}_{-0.046}$

Table 6: (Density) Sparsification summary for selected ranking metrics across selected conditions.

Condition	AUSE_σ	AUSE_b	nAUSE_σ	nAUSE_b
noiseless (300v)	0.0008	0.0052	$0.140^{+0.028}_{-0.025}$	$0.855^{+0.190}_{-0.156}$
100 views	0.0009	0.0051	$0.142^{+0.029}_{-0.026}$	$0.838^{+0.184}_{-0.152}$
20 views	0.0017	0.0098	$0.161^{+0.033}_{-0.026}$	$0.910^{+0.144}_{-0.123}$
5 views	0.0065	0.0240	$0.281^{+0.037}_{-0.031}$	$1.037^{+0.092}_{-0.085}$
Noise $\times 9$	0.0025	0.0060	$0.251^{+0.052}_{-0.045}$	$0.599^{+0.130}_{-0.107}$
Noise $\times 25$	0.0030	0.0061	$0.266^{+0.056}_{-0.042}$	$0.550^{+0.121}_{-0.095}$
Si $\times 0.6$	0.0155	0.0287	$0.349^{+0.090}_{-0.076}$	$0.644^{+0.096}_{-0.090}$
Si $\times 0.8$	0.0067	0.0098	$0.344^{+0.069}_{-0.060}$	$0.507^{+0.057}_{-0.051}$
Si $\times 1.2$	0.0064	0.0076	$0.587^{+0.067}_{-0.062}$	$0.706^{+0.102}_{-0.089}$
Si $\times 1.4$	0.0114	0.0112	$0.827^{+0.077}_{-0.078}$	$0.811^{+0.093}_{-0.086}$

Table 7: (Temperature) Sparsification summary for selected ranking metrics across selected conditions.

Finally, to confirm that no single training run drives the localization signal, we recompute the full correlation $\rho(\sigma_{\mathrm{ens}}, \epsilon_{\mathrm{mean}})$ and nAUSE_σ on each of the ten leave-one-seed-out sub-ensembles (Table 8). We find that the LOO values stay within a comparably tight band, so the reported correlation values are not artifacts of any individual seed.

Condition	n_e full ρ	n_e LOO ρ	n_e LOO nAUSE$_\sigma$	T_e full ρ	T_e LOO ρ	T_e LOO nAUSE$_\sigma$
300 views	0.761	$[0.755, 0.767]$	$[0.115, 0.120]$	0.726	$[0.717, 0.735]$	$[0.138, 0.144]$
100 views	0.773	$[0.765, 0.775]$	$[0.112, 0.128]$	0.732	$[0.726, 0.735]$	$[0.141, 0.145]$
20 views	0.778	$[0.773, 0.783]$	$[0.114, 0.124]$	0.717	$[0.708, 0.722]$	$[0.160, 0.166]$
5 views	0.712	$[0.695, 0.731]$	$[0.205, 0.240]$	0.611	$[0.599, 0.619]$	$[0.276, 0.293]$
Noise $\times 9$	0.572	$[0.559, 0.568]$	$[0.280, 0.287]$	0.590	$[0.582, 0.590]$	$[0.251, 0.262]$
Noise $\times 25$	0.555	$[0.540, 0.552]$	$[0.307, 0.324]$	0.601	$[0.585, 0.603]$	$[0.263, 0.286]$
Si $\times 0.6$	0.438	$[0.436, 0.442]$	$[0.405, 0.414]$	0.424	$[0.408, 0.440]$	$[0.337, 0.356]$
Si $\times 0.8$	0.387	$[0.380, 0.403]$	$[0.409, 0.422]$	0.448	$[0.437, 0.450]$	$[0.339, 0.352]$
Si $\times 1.2$	0.166	$[0.158, 0.174]$	$[0.605, 0.625]$	0.250	$[0.241, 0.256]$	$[0.585, 0.604]$
Si $\times 1.4$	-0.017	$[-0.023, -0.012]$	$[0.926, 0.951]$	0.096	$[0.088, 0.107]$	$[0.812, 0.835]$

Table 8: Leave-one-out (LOO) ranges of $\rho(\sigma_{\text{ens}}, \epsilon_{\text{mean}})$ and nAUSE$_\sigma$ for both density and temperature.

C.3 Seed-Deviation Error Capture

As a complementary field-space consistency test, we project the signed ensemble error onto the vectors spanned by the ensemble-deviations. Define the ensemble-deviation subspace as $S_{\text{ens}} = \text{span}\{\delta_1, \ldots, \delta_K\}$, where $\delta_k = \hat{m}_k - \overline{m}$ are the per-seed field deviations. Define the rank of S_{ens} as $r \leq K - 1$, since $\sum_k \delta_k = 0$ removes one degree of freedom. Then, we can define the orthogonal projection on this subspace by letting $Q \in \mathbb{R}^{M \times r}$ be an orthonormal basis of S_{ens}, and $P_{\text{ens}} = QQ^T$ as the orthogonal projection operator. This allows us to define an error-capture energy fraction

$$E_{\text{cap}} = \frac{||P_{\text{ens}} e||^2}{||e||^2} = \frac{||Q^T e||^2}{||e||^2} \quad \in [0, 1], \tag{27}$$

which is the fraction of squared L^2 norm of the signed reconstruction error that lies within the seed-deviation subspace. We additionally report enrichment relative to the rank-matched isotropic expectation $E_{\text{cap}}^{\text{rand}} = r/M$ for both density and temperature. Note that in our experiment, we use $K = 10$ seeds and all reported full-ensemble deviation matrices have numerical rank $r = K - 1 = 9$ under a relative tolerance of 10^{-8} on the singular values. This is computed using the same control grid as above ($M = 4320$ per-field), so our isotropic baseline sits at around $E_{\text{cap}}^{\text{rand}} \approx 0.002$. Additionally, E_{cap} depends on the chosen control-grid discretization and Euclidean inner product.

A high E_{cap} indicates directional alignment, but such alignment could be inflated by (i) reusing the same finite ensemble to construct both the signed error $e = \overline{m} - m^*$ and the deviation span, or (ii) generic spatial structure, where the signed error e and the seed-deviation fields share smooth, low-frequency modes even without scene-specific alignment.

We test (i) by computing a split-seed validation test, where we split the $K = 10$ seeds into two groups A and B. We construct the span from the per-seed field deviations in A, $S_A = \text{span}\{\hat{m}_k - \overline{m}_A : k \in A\}$, and separately construct the signed ensemble mean error $e_B = \overline{m}_B - m^*$ from B. We then compute the error from ensemble B captured by the span from ensemble A:

$$E_{\text{cap}}^{B \to A} = \frac{||P_A e_B||^2}{||e_B||^2}, \tag{28}$$

repeated over all balanced splits, and report the mean and the 2.5th and 97.5th percentile interval of the finite split-sensitivity distribution. This explicitly tests whether a subspace learned from one subset of seeds captures the error of a disjoint seed subset, rather than evaluating the subspace and error from exactly the same finite ensemble. For our $K = 10$ ensemble, we enumerate all

$\binom{10}{5} = 252$ ordered choices of A (126 complementary partitions evaluated in both directions), with B as its complement, and report the exact finite split-sensitivity distribution, rather than a Monte Carlo approximation. We find that across matched-model conditions, our $E_{\text{cap}}^{B\to A}$ values are substantially larger than the rank-matched isotropic expectation. We note that our five-member group has a numerical rank of four under the same tolerance, so its rank-matched isotropic baseline is $4/4320 \approx 0.09\%$, as opposed to the 0.2% rank-9 baseline for the original E_{cap}.

We test (ii) by using a stronger structured null: we cyclically rotate the ensemble-deviation fields δ_k relative to the signed error field e through all 29 nonzero longitude shifts, preserving spatial structure radially and latitudinally while disrupting the longitudinal alignment. More specifically, we choose a cyclic longitude shift, apply that same shift to all deviations δ_k, keeping e fixed. We then compute E_{cap} for the shifted deviation span, and repeat for all 29 nonzero cyclic longitude shifts. In implementation, we equivalently keep the deviation span fixed and cyclically shift the signed field error. Finally, we compute the mean μ and compare with the original E_{cap} with no longitudinal rotations. Again across matched-model conditions, we find that the non-rotated E_{cap} exceeds the rotation-null mean and is at the top of the distribution in all but the 100-view temperature reconstruction. The reduced capture indicates that the original longitudinal alignment contributes to the observed overlap beyond generic spatial structure alone.

Condition	E_{cap} %	$E_{\text{cap}}/E_{\text{cap}}^{\text{rand}}$ (enrichment)	$E_{\text{cap}}^{B\to A}$ %	rot-null μ %	E_{cap} rot %
noiseless (300v)	7.5	36×	$3.6^{+2.7}_{-3.3}$	2.5	100
100 views	4.8	23×	$2.5^{+2.7}_{-2.2}$	2.1	100
20 views	27.2	130×	$14.4^{+8.6}_{-8.0}$	5.5	100
5 views	32.1	154×	$17.9^{+11.6}_{-13.3}$	20.9	100
Noise ×9	6.0	29×	$2.7^{+2.3}_{-2.0}$	0.9	100
Noise ×25	11.0	53×	$5.9^{+4.7}_{-4.4}$	2.3	100
Si ×0.6	3.4	16×	$1.6^{+1.7}_{-1.1}$	2.1	97
Si ×0.8	8.0	38×	$4.3^{+3.8}_{-3.6}$	2.0	100
Si ×1.2	4.8	23×	$2.1^{+1.5}_{-1.5}$	1.7	100
Si ×1.4	1.0	5×	$0.6^{+0.6}_{-0.4}$	0.8	72

Table 9: (Density) Error-capture diagnostics as an additional field-space consistency metric.

Condition	E_{cap} %	$E_{\text{cap}}/E_{\text{cap}}^{\text{rand}}$ (enrichment)	$E_{\text{cap}}^{B\to A}$ %	rot-null μ %	E_{cap} rot %
noiseless (300v)	5.0	24×	$2.6^{+2.0}_{-1.9}$	1.5	100
100 views	2.4	11×	$1.2^{+1.4}_{-0.9}$	1.4	79
20 views	20.5	98×	$11.0^{+7.7}_{-7.3}$	4.0	100
5 views	24.0	115×	$12.8^{+8.3}_{-8.3}$	18.7	100
Noise ×9	9.3	45×	$4.6^{+3.1}_{-3.6}$	2.1	100
Noise ×25	20.6	99×	$12.8^{+7.0}_{-7.0}$	3.4	100
Si ×0.6	12.1	58×	$6.3^{+4.4}_{-3.7}$	1.0	100
Si ×0.8	6.0	29×	$3.5^{+3.7}_{-2.8}$	0.8	100
Si ×1.2	2.5	12×	$1.4^{+1.1}_{-1.3}$	1.1	93
Si ×1.4	1.9	9×	$1.2^{+1.2}_{-1.0}$	0.5	97

Table 10: (Temperature) Error-capture diagnostics as an additional field-space consistency metric.

Tables 9 and 10 summarize our error capture diagnostics. For selected conditions, we report the E_{cap} % as the fraction of the squared signed-error energy lying in the rank-9 span, and the enrichment relative to an equal-rank isotropic subspace. We also report the split-seed capture percentage $E_{\text{cap}}^{B\to A}$ % with its 2.5th-97.5th percentile interval, the mean capture percentage μ of the longitude-rolled deviations, and the percentile rot % of the unrotated E_{cap} relative to the rotation-null distribution.

D BENCHMARK STUDIES AND ABLATIONS

D.1 REPRODUCIBILITY

Table 11 provides a list of the main canonical hyperparameters used in our experiments. Note that control configurations are rerun independently within each ablation study. Because GPU hash-grid optimization uses nondeterministic operations, fixing the nominal RNG seed does not guarantee bitwise-identical training trajectories. Additionally, different studies may use different RNG streams, so numerically identical configurations can differ slightly across tables. We therefore interpret within-study mean $\pm$ standard deviation comparisons, and do not treat small cross-table differences between nominally identical controls as scientific effects.

Unless otherwise stated, we fix the model architecture to be a multiresolution hash grid with a hashmap size of 2^{20} $\times 2$ features, and 20 levels spanning resolutions from 16 (coarsest) to 512 (finest). Separate decoder heads, each containing two hidden layers of width 64, are used to predict log-density and log-temperature, and their biases are initialized to the midpoints of their corresponding clamp ranges, which lie within the respective LUT domains. We train with 300 evenly-spaced views outside a 30° held-out arc with a 3° buffer. We use the AdamW optimizer with our asinh loss, and joint reconstructions optimize using batches of 1024 rays for 60k steps, while our density-only representations (Appendix D.2) use 30k steps. No observation noise or abundance mismatch is applied. Our evaluation metrics are field-space errors and held-out image-space errors.

Group	Name	Description
Rendering	AABB	$x_{\mathrm{AABB}} \in [-1, 1]^3$
	sample step size	$dx_{\mathrm{AABB}} = 1/256$
	LOS integration	midpoint Riemann sum
Architecture	hash grid	2^{20} hash map, $\times 2$ features
	levels	20 levels, $16 \rightarrow 512$ coarse-to-fine resolutions
	MLP head	per-field, 2 hidden layers, width 64
Training	Optimizer	AdamW
	learning rate	5×10^{-4}
	weight decay	10^{-5}
	ray batch size	$b = 1024$
	steps	$s = 60$k
	asinh scale	$s_c = [1.3 \times 10^{-2}, 4.5 \times 10^{-4}, 10^{-2}, 5.8 \times 10^{-2}]$
	ensemble seeds	0xC0FFEE, 0xDECAF, 0xBADD1E, 0xD1CE 0x8888, 0xB0BA, 0xFACADE, 0xCAFE 0xF00D, 0xBADA55
Forward Model	shot coefficients	$\alpha_c = [10^{-2}, 5 \times 10^{-4}, 10^{-2}, 5 \times 10^{-2}]$
	noise multiplier	$\eta \in \{0, 1, 9, 25\}$
	mismatch scale	$a \in \{0.6, 0.8, 1, 1.2, 1.4\}$

Table 11: Table of canonical hyperparameters used in CoroNeRF. Note that the listed s_c values are the base scales used for $\eta \leq 1$. For $\eta > 1$, s_c is multiplied by $\sqrt{\eta}$ as described in Appendix D.6. The ordering of α_c and s_c values corresponds to Fe XIII 1075, Fe XIII 1080, Si IX 2585, and Si IX 3935, respectively.

D.2 Density Representation Baselines

Our density representational baselines below share the following properties. We only reconstruct density with the GT temperature field supplied. The observational dataset has 300 evenly-spaced views with a 30° arc holdout and a 3° buffer on both ends (the arc holdout is used as validation). Each view consists of a $[-3, 3]$ $R_\odot$ FOV of two Fe XIII channels ($1075/1080$ nm), with no observational noise or abundance scale mismatch. All models take in as inputs a 3D Cartesian position mapped to the AABB. We use the AdamW optimizer, with learning rate 5×10^{-4}, weight decay 10^{-5}, and 1024 ray batches for 30k steps. Our loss is the fixed-scale asinh image loss. Each model is trained with 3 seeds (0xC0FFEE, 0xDECAF, 0xBADD1E).

The first baseline we test is a classical voxel grid, which we call the `Grid` model. The optimized variable is a dense trainable log-density field with no encoder or MLP. The grid dimensions match those of the native PSI cube, with colatitude converted to latitude: $299 \times 142 \times 154$ longitude-latitude-radius, where the longitude $\in [0, 2\pi]$ is periodic (wrap-padded), latitude $\in [-\pi/2, \pi/2]$, and radius $\in [1, 30]$ $R_\odot$. Retrieving values from this trainable grid is done using trilinear interpolation via PyTorch's `grid_sample` function, with a longitude wrap for periodicity. All grid nodes are initialized as the midpoint of the $\log n_\mathrm{e}$ clamp range. The number of trainable parameters in this model is the total number of voxels $6,538,532$ ($= 299 \cdot 142 \cdot 154$).

We additionally test the grid representation with a first-order (Tikhonov/gradient) smoothness regularizer on the learned density grid, which we will call the `Grid-Reg` model. For a grid g, the regularizer has the form $\mathcal{L}_R = \lambda_\mathrm{lon}\langle(\Delta_\mathrm{lon}\, g)^2\rangle + \lambda_\mathrm{lat}\langle(\Delta_\mathrm{lat}\, g)^2\rangle + \lambda_\mathrm{r}\langle(\Delta_\mathrm{r}\, g)^2\rangle$, where Δ is the first difference along each axis and $\langle\cdot\rangle$ is the mean over nodes. We set the directional weights $\lambda_\mathrm{lon} = \lambda_\mathrm{lat} = \lambda_\mathrm{r} = 1$. The coefficient we use for the overall total loss is $\lambda_\mathrm{smooth} = 10.0$, where the training objective is $\mathrm{asinhErr} + 10.0 \times \mathcal{L}_R$. We chose 10.0 as it gave us the best preliminary results compared to other regularization values. Finally, the regularizer does not change the amount of trainable parameters (6.5M).

The second baseline we use is a positional encoder + MLP, which we will call the `MLP` model. The encoding frequencies are axis-wise Fourier positional encoding with $L = 6$ bands with corresponding frequencies $\{2^0, \ldots, 2^5\}$ (features are $\sin(2^k \pi\, x)$ and $\cos(2^k \pi\, x)$ per axis). This gives a total of 2 features $\times$ 3 axes $\times$ 6 bands $=$ 36 dims. We additionally have a spherical-harmonic encoding of the unit position direction up to degree 2 ($(2+1)^2 = 9$ dims). We concatenate the raw position with the positional encoding and spherical-harmonic encoding for a total of 48 dims to the MLP input. Our MLP has 4 hidden layers each with width 256. SiLU activations connect the hidden layers, and the final layer is a linear output layer. The output is a single scalar $\log_{10} n_\mathrm{e}$, and the final-layer bias is initialized to the midpoint of the $\log n_\mathrm{e}$ clamp range. The total number of trainable parameters is $210,177$ ($= 48 \cdot 256 + 256 + 3 \cdot (256^2 + 256) + (256 + 1)$).

Our final and main representation that we eventually select for the main experiments is a multiresolution hash grid as described in Appendix A.5. The encoding hash grid has 20 levels, 2 features per level, with the resolutions from 16 (coarsest) to 512 (finest). Each level has a 2^G hashmap (we test $G \in \{17, 20\}$). Given a position, the hash grid produces a 40-dimensional feature vector by concatenating the 2 features across all 20 levels. This is then fed into a decoder head with 2 hidden layers each with width 64. Activations are SiLU, and the final layer is a linear output layer. The decoder head has 6849 trainable parameters. Our `Hash20` model has 20 levels $\times 2^{20}$ hashmap $\times 2$ features + $6849 = 41,949,889 \approx 41.9$M parameters. Our `Hash17` model has 20 levels $\times 2^{17}$ hashmap $\times 2$ features + $6849 = 5,249,729 \approx 5.2$M parameters.

Figure 11 shows a meridional (latitude-radius) slice of the reconstructed density at 0° longitude for various representational baselines, providing a complementary view of the reconstructed corona compared to a latitude-longitude shell shown in Figure 2 in the main text. Table 12 provides the field-space and image-space metrics for our density representations. We find that the two tested hash-table sizes, $G \in \{17, 20\}$, achieve nearly identical field error.

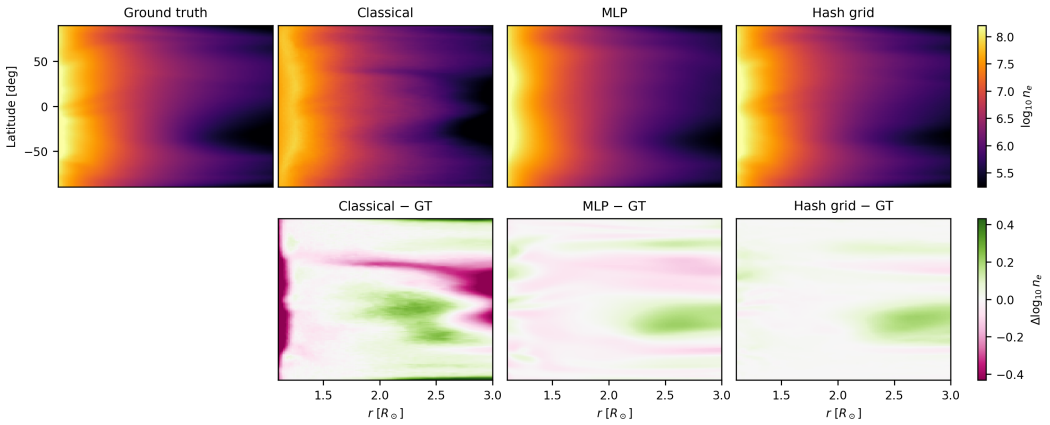

Figure 11: Density-field reconstruction (with GT temperature supplied) at 0° longitude, spanning $[1.1, 3.0]$ $\mathrm{R}_\odot$. Each panel is a meridional slice (latitude-radius plot). The top row shows the ground truth and reconstructions using classical spherical grid with Tikhonov regularization, positional-encoder + MLP, and hash grid + decoder head representations. The bottom row shows corresponding signed residuals $\Delta \log_{10} n_e = \log_{10} n_{e,\mathrm{pred}} - \log_{10} n_{e,\mathrm{GT}}$. The multiresolution hash grid is able to reconstruct consistent radial structures much farther out than the classical method, and achieves lower errors on the inner radial boundaries near the photosphere.

Model	$\mathrm{MAE}_{\mathrm{inner}}(\log_{10} n_e) \downarrow$	$\mathrm{AbsRel}_{\mathrm{inner}}(n_e) \downarrow$	$\mathrm{asinhErr}(I) \downarrow$	$\mathrm{PSNR}(I) \uparrow$
`Hash20`	$\mathbf{0.042} \pm < \mathbf{0.001}$	$\mathbf{0.094} \pm < \mathbf{0.001}$	$\mathbf{0.003} \pm < \mathbf{0.001}$	$\mathbf{39.59 \pm 0.67}$
`Hash17`	$\mathbf{0.042} \pm < \mathbf{0.001}$	$0.095 \pm < 0.001$	0.004 ± 0.001	38.77 ± 0.65
`MLP`	0.097 ± 0.001	0.254 ± 0.004	0.024 ± 0.003	24.65 ± 0.62
`Grid-Reg`	$0.168 \pm < 0.001$	$0.449 \pm < 0.001$	$0.053 \pm < 0.001$	$11.57 \pm < 0.01$
`Grid`	$0.217 \pm < 0.001$	$0.571 \pm < 0.001$	$0.080 \pm < 0.001$	$11.30 \pm < 0.01$

Table 12: Density representation benchmark. Each row indicates a model configuration, and metrics are reported as mean $\pm$ standard deviation over 3 seeds, with deviations below 5×10^{-4} reported as < 0.001. The first column indicates the experiment family, and the last four columns are the relevant metrics. Image metrics are averaged over views and channels, and density metrics are averaged over inner coronal radial shells 1.1-2 $\mathrm{R}_\odot$. The multiresolution hash grid outperforms the MLP by more than $2\times$, and the classical grid by about $4\times$.

D.3 Loss Ablation Study

In our loss ablation study, we keep the same model architecture and observational setup as those described in Appendix D.1, except we introduce observational noise at $\eta = 1$ and alter the objective function. We use the following objective family:

$$\mathcal{L}_{\mathrm{train}} = \lambda_G \mathcal{L}_G + \lambda_A \mathcal{L}_A, \tag{29}$$

where $\mathcal{L}_G$ is the heteroscedastic Gaussian image loss, and $\mathcal{L}_A$ is the fixed-scale asinh image loss. More specifically, given ray q and channel c, the heteroscedastic Gaussian term is the inverse-variance-weighted squared image residual $((\hat{I}_{\theta,q,c} - y_{q,c})/\sigma_{q,c})^2$, which is equivalent for optimization to the Gaussian NLL up to a positive constant factor and additive terms independent of θ. The whitening-scale $\sigma_{q,c}$ is the target-σ, or the known per-pixel, per-channel observation noise standard deviation of our generative model (Equation 6): this is the exact heteroscedastic standard deviation used to add noise to the training images. By contrast, the asinh term uses $|\operatorname{asinh}(\hat{I}_{q,c}/s_c) - \operatorname{asinh}(y_{q,c}/s_c)|$, where s_c is a channel-dependent fixed characteristic intensity scale. Our experiment families vary λ_G and λ_A so as to test objectives that are pure Gaussian, mixed Gaussian and asinh, and pure asinh. Each experiment family is trained with 3 seeds, and replicate runs use different observation-noise realizations coupled to the run seed.

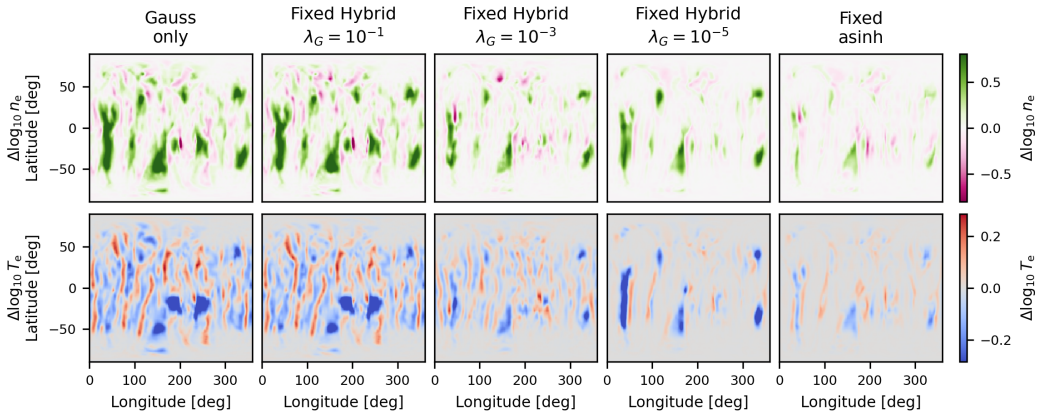

Figure 12: Density and temperature reconstructed residuals for different loss metric settings. We find that the Gaussian-only loss performs worst while the asinh-only loss performs best, with hybrid losses landing in between the two regimes.

Objective	$\text{MAE}_{\text{inner}}(\log_{10} n_e) \downarrow$	$\text{MAE}_{\text{inner}}(\log_{10} T_e) \downarrow$	$\text{asinhErr}(I) \downarrow$
Gaussian-only $\lambda_G = 1$	0.145 ± 0.010	0.075 ± 0.020	0.060 ± 0.006
Fixed hybrid $\lambda_G = 10^{-1}$	0.134 ± 0.009	0.064 ± 0.014	0.047 ± 0.005
Fixed hybrid $\lambda_G = 10^{-3}$	0.087 ± 0.001	0.036 ± 0.001	0.032 ± 0.002
Fixed hybrid $\lambda_G = 10^{-5}$	0.061 ± 0.002	0.022 ± 0.002	$\mathbf{0.021} \pm < \mathbf{0.001}$
Fixed asinh only $\lambda_G = 0$	$\mathbf{0.047 \pm 0.001}$	$0.017 \pm < 0.001$	$0.022 \pm < 0.001$
Target-σ asinh $\lambda_G = 0$	$0.047 \pm < 0.001$	$\mathbf{0.016} \pm < \mathbf{0.001}$	0.022 ± 0.001

Table 13: Loss ablation benchmark. For the Gaussian-only experiment, $\lambda_A = 0$, otherwise it is fixed at $\lambda_A = 1$. We find that objectives that favor the asinh term $\mathcal{L}_A$ generally perform better than those that favor the heteroscedastic Gaussian loss term.

Under the tested training protocol, the pure asinh objective yields lower log-field MAE than the Gaussian-only objective (Figure 12 and Table 13). We therefore use it as the task-aware objective for the subsequent reconstruction experiments. As an auxiliary oracle variant, the final row, our target-σ case, replaces s_c in $\mathcal{L}_A$ with the synthetic generator's per-pixel target σ. The corresponding loss term becomes $|\operatorname{asinh}(\hat{I}_{q,c}/\sigma_{q,c}) - \operatorname{asinh}(y_{q,c}/\sigma_{q,c})|$, and is thus included only as an oracle diagnostic ablation. Finally, we note that this ablation compares complete objective choices rather than separately isolating the effects of intensity transformation, residual norm, and weighting.

D.4 Compute and Runtime

Our multiresolution hash grid has around 40M parameters ($2M$ per hash table, 20 levels), and is trained in a cluster GPU setting: training a single reconstruction takes ≈ 3 GPU-hours on an 80 GB NVIDIA A100, and so a $K = 10$ ensemble takes ≈ 30 aggregate GPU-hours. The ten localization conditions reported in Tables 2-10 require approximately 300 aggregate GPU-hours. Rendering a 300-view set, as a dataset or post-training analysis, can take $\approx$ 10-100 minutes on a local NVIDIA RTX 4090 depending on the resolution and integration step size. Finally, diagnostics such as σ_{ens} are cheap as they only require K forward evaluations.

We performed pilot tests to pick batch size b and step count s from a compute-quality trade-off ablation study (Table 14). Note that these pilot runs use the earlier density-only configuration with no observational noise, and are not directly comparable to the noise-view sweep of the joint thermodynamic reconstruction in Figure 5. We find that field error varies modestly across the tested settings, and note that because these tests were done only for hyperparameter selection at an earlier stage of the pipeline, they should not be used for scientific conclusions.

Exp	b	s	V	D	**MAE**($\log_{10} n_e$) ↓	**PSNR**(I) ↑	**MSE**($\log_{10} I$) ↓
big	4096	60k	30	64	0.096	30.56	1.75×10^{-4}
big	4096	60k	1000	256	0.097	30.74	1.78×10^{-4}
med	4096	30k	1000	256	0.106	26.60	5.46×10^{-4}
med	4096	30k	30	64	0.107	27.47	4.81×10^{-4}
base	1024	30k	1000	256	0.109	24.88	6.40×10^{-4}
base	1024	30k	30	64	0.109	27.82	5.01×10^{-4}

Table 14: Compute-quality trade-off ablation study. We vary the batch size b, steps s, views V, and image dimension D. We find that the log-density MAE varies modestly across the tested settings, with slightly better performance at $b = 4096$, $s = 60$k.

D.5 Spectral Channel Ablation Study

This section provides details on our spectral channel ablation study. We ablate the number of spectral channels for a total of 5 experiment families: Fe XIII 1075, Si IX 3935, Fe XIII pair, Si IX pair, and All four. Each experiment family is trained with 3 seeds.

Channel Set	$\text{MAE}_{\text{inner}}(\log_{10} n_{\text{e}})$ ↓	$\text{MAE}_{\text{inner}}(\log_{10} T_{\text{e}})$ ↓	asinhErr(observed channels) ↓
all4	$\mathbf{0.039} \pm < \mathbf{0.001}$	$\mathbf{0.014} \pm < \mathbf{0.001}$	0.008 ± 0.002
Fe XIII pair	0.097 ± 0.004	0.023 ± 0.001	0.009 ± 0.003
Si IX pair	0.068 ± 0.001	0.058 ± 0.001	$\mathbf{0.004 \pm 0.001}$
Fe 1075	0.605 ± 0.004	$0.072 \pm < 0.001$	0.006 ± 0.001
Si 3935	0.530 ± 0.028	0.195 ± 0.006	0.005 ± 0.002

Table 15: Reconstruction metrics for our spectral ablation study. The four-line configuration achieves the lowest field errors and low across-seed variability in this three-seed ablation. We can also observe that a good image-space reconstruction does not imply a good physical-field reconstruction: while the Si IX pair has the lowest image-space validation loss, it still has higher physical-field errors.

Table 15 summarizes the field-space and image-space performance for all five experiment families, and Figure 13 shows a sample joint reconstruction at three selected radii and corresponding residual panels. We also provide radial profiles of the field-space MAEs (Figure 14). Finally, Figure 15 shows image-space versus field-space error for all five experiment families, for both density and temperature (as opposed to the density-only Figure 4 in the main text). We note that for visualizations of radial shells, we select the median-performing model (out of three seeds) using MAE of the inner band temperature field.

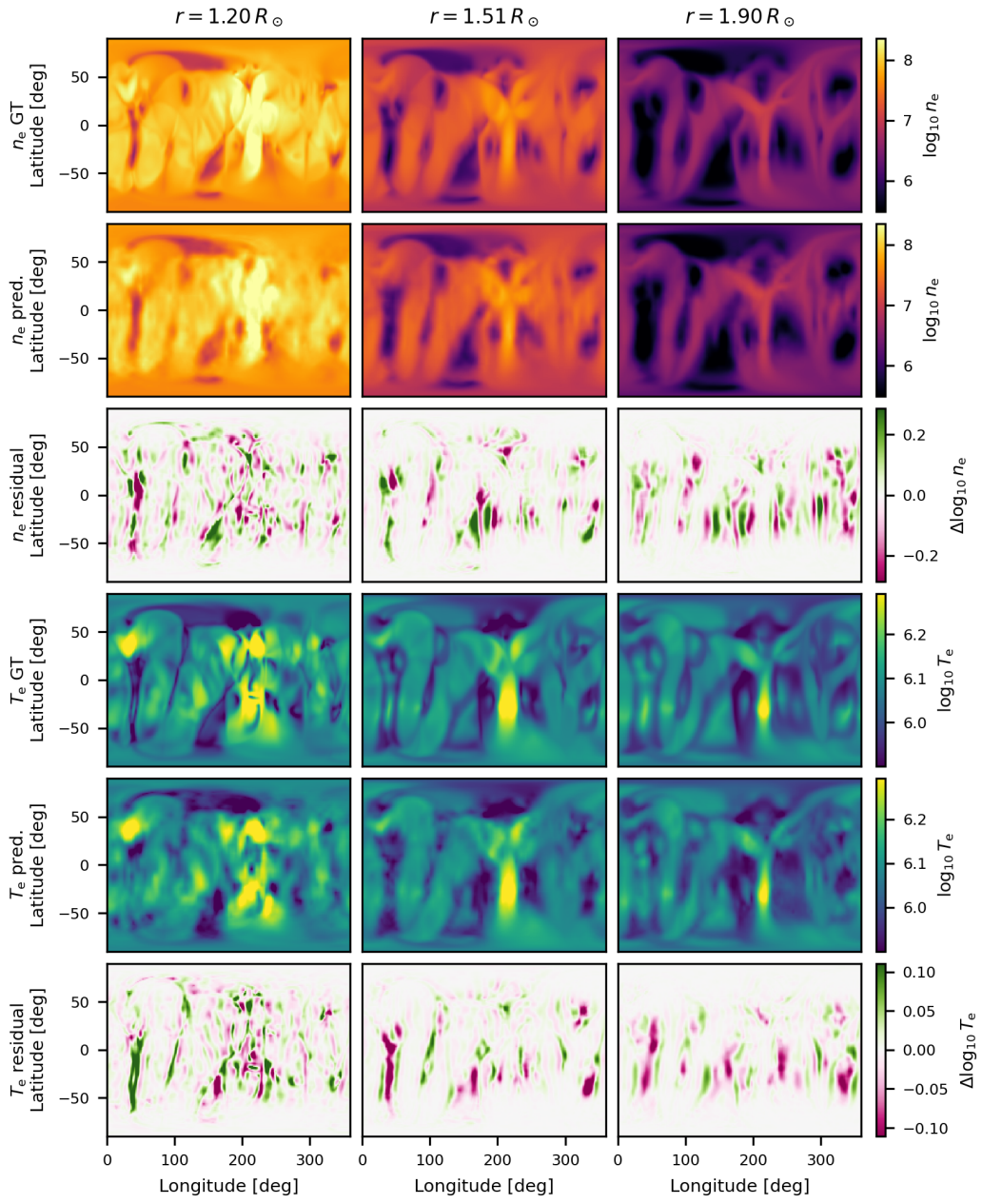

Figure 13: Field Reconstruction at three selected radii: $1.20, 1.51, 1.90\ \mathrm{R}_\odot$ (columns). We show the field-space GT, prediction, and signed residuals $\Delta m = \hat{m} - m^*$ for $m \in \{\log_{10} n_\mathrm{e}, \log_{10} T_\mathrm{e}\}$.

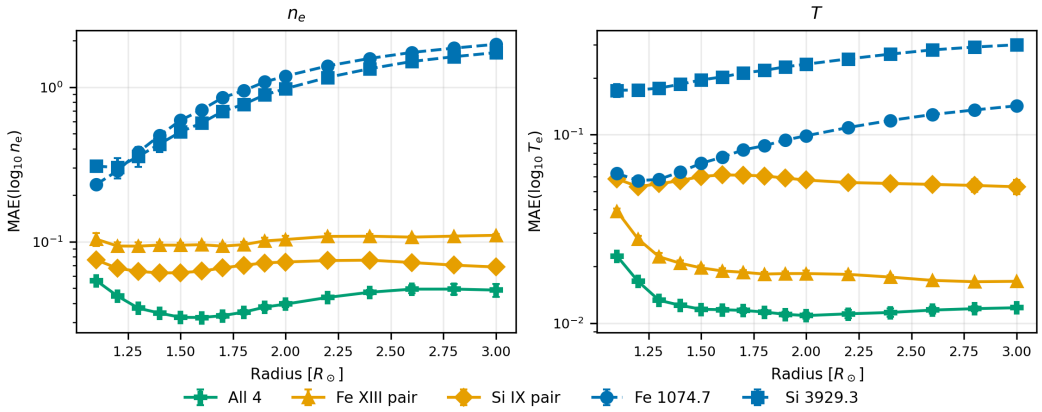

Figure 14: Field-space MAE radial dependence for the spectral ablation study. Color denotes the number of spectral lines, and marker shape denotes the line set. We find that having at least a pair of lines substantially reduces radial errors across 1.1-3 $R_\odot$.

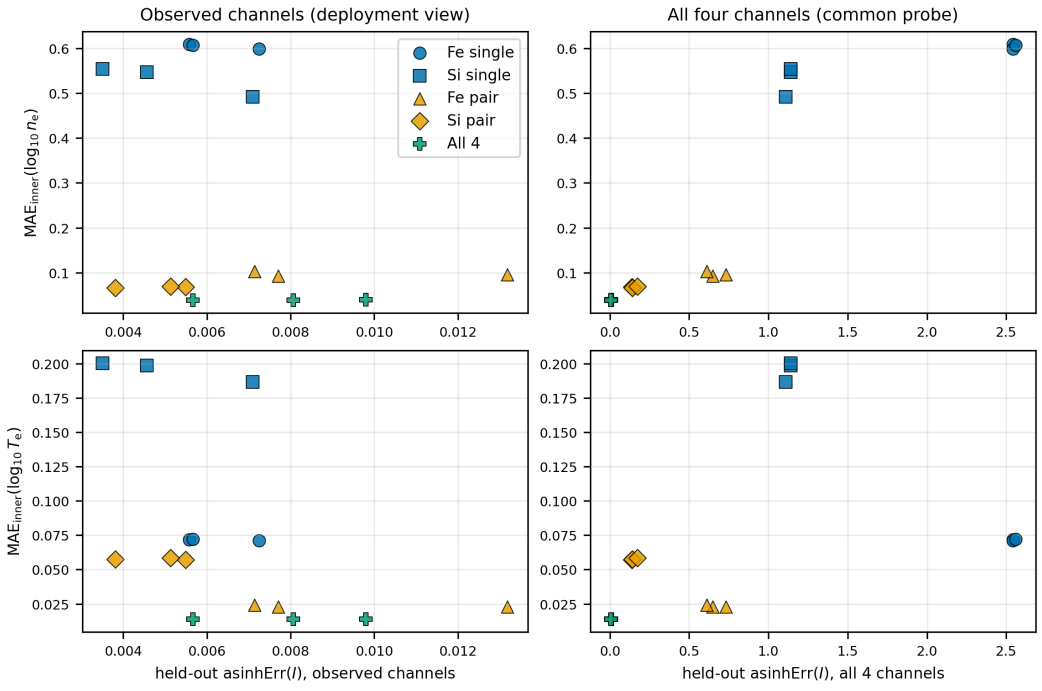

Figure 15: 3D field-space MAE vs 2D held-out image-space validation loss, showing that a good image-space reconstruction need not imply a good field-space reconstruction. Each point is one seed from the spectral ablation benchmark trained with different sets of spectral channels. We find that fewer-channel models can achieve lower observed-channel image error than the four-line model despite worse field recovery.

D.6 NOISE/VIEW ABLATION STUDY

This section details the noise-view stress ablation study described in Figure 5 in the main text. We keep the same model architecture and observational setup as described in Appendix D.1. However, we ablate the number of views V by selecting evenly sampled views from the original 300 views. Likewise, we also ablate the noise scale η. Additionally, we calibrate the per-channel asinh scale to the noise level: s_c is fixed at the baseline for $\eta \in \{0, 1\}$, and scaled by $\sqrt{\eta}$ for higher noise, so it tracks the noise standard deviation (variance scales as η, standard deviation as $\sqrt{\eta}$). The scale is constant within any single run.

Our ablation study has four noise levels ($\eta \in \{0, 1, 9, 25\}$) crossed with four view counts ($V \in \{5, 20, 100, 300\}$), for a total of 16 distinct combinations trained with three different seeds for 48 total models. For noisy conditions, the three replicate seeds also use distinct observation-noise realizations derived from the run seed.

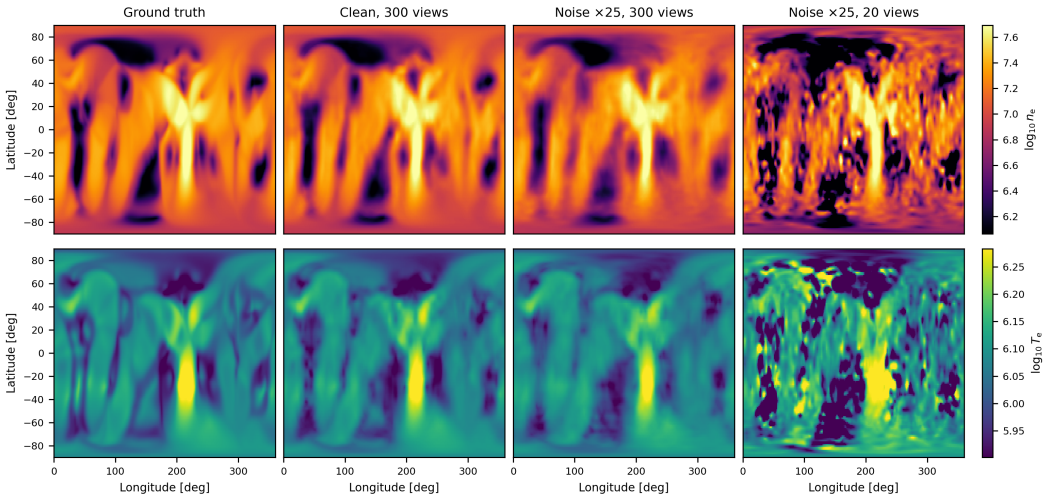

Figure 16: Robustness in physical field reconstruction. The columns show ground truth and the reconstructed density (top row) and temperature (bottom row) for different combinations of views and noise multipliers. We see that with 300 views, there is moderate degradation between a clean reconstruction with no noise and one with a noise multiplier of 25, while retaining low absolute error. However, once we reduce the number of views significantly to 20 views (last column), the combined effect of a large noise multiplier and sparse views is a failure mode for physical field reconstruction.

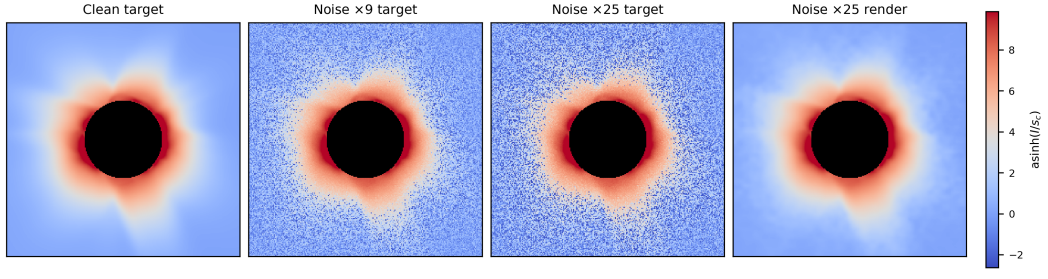

Figure 17: Demonstration of the accompanying denoising effect in the framework. We show the clean ground truth target, and sample noisy training observations. The final column shows a reconstructed render using a model trained on noise $\times 25$ images, where multiview consistency suppresses pixel-scale noise in the reconstructed render.

Noise Setting	$\mathrm{MAE}_{\mathrm{inner}}(\log_{10} n_{\mathrm{e}}) \downarrow$	$\mathrm{MAE}_{\mathrm{inner}}(\log_{10} T_{\mathrm{e}}) \downarrow$	$\mathrm{asinhErr}(I) \downarrow$
$\eta = 0$, 5 views	0.410 ± 0.020	0.070 ± 0.005	1.454 ± 0.041
$\eta = 0$, 20 views	0.122 ± 0.023	0.026 ± 0.003	0.458 ± 0.060
$\eta = 0$, 100 views	$0.039 \pm <0.001$	$0.014 \pm <0.001$	0.007 ± 0.001
$\eta = 0$, 300 views	$0.039 \pm <0.001$	$0.014 \pm <0.001$	0.007 ± 0.001
$\eta = 1$, 5 views	0.516 ± 0.026	0.126 ± 0.010	1.513 ± 0.046
$\eta = 1$, 20 views	0.209 ± 0.018	0.044 ± 0.003	0.823 ± 0.050
$\eta = 1$, 100 views	0.052 ± 0.002	$0.018 \pm <0.001$	$0.039 \pm <0.001$
$\eta = 1$, 300 views	0.047 ± 0.001	$0.017 \pm <0.001$	$0.022 \pm <0.001$
$\eta = 9$, 5 views	0.519 ± 0.021	0.970 ± 0.370	1.565 ± 0.050
$\eta = 9$, 20 views	0.275 ± 0.037	0.115 ± 0.030	1.239 ± 0.084
$\eta = 9$, 100 views	0.069 ± 0.003	$0.022 \pm <0.001$	0.092 ± 0.004
$\eta = 9$, 300 views	0.057 ± 0.002	$0.019 \pm <0.001$	0.041 ± 0.001
$\eta = 25$, 5 views	0.760 ± 0.458	2.404 ± 1.154	1.529 ± 0.059
$\eta = 25$, 20 views	0.336 ± 0.077	0.380 ± 0.138	1.418 ± 0.107
$\eta = 25$, 100 views	0.081 ± 0.005	0.026 ± 0.001	0.136 ± 0.008
$\eta = 25$, 300 views	0.063 ± 0.002	$0.021 \pm <0.001$	0.054 ± 0.001

Table 16: Noise-view stress benchmark results (also shown in Figure 5 in the main text). With sufficient angular coverage, the reconstruction remains robust even under strong heteroscedastic observation noise: increasing the noise multiplier to $\times 25$ at 300 training views only moderately increases the density and temperature field errors. In contrast, sparse-view regimes are substantially less stable, and the combination of high noise and sparse views produces a clear failure mode in latent physical-field recovery. The training objective uses the noise-matched s_c scaling described above, while held-out asinhErr is evaluated on noise-free held-out views using the same base per-channel scales across all η.

Figure 16 shows a sample reconstruction at $1.51\ \mathrm{R}_{\odot}$, and Table 16 shows detailed results of our ablation study. Additionally, we show an apparent denoising effect of training on noisy observations with sufficient views in Figure 17.

D.7 Abundance Mismatch Ablation Study

This section details an abundance scale ablation study that probes a controlled common-mode forward-model mismatch. We keep the same model architecture and observational setup as described in Appendix D.1. However, we ablate the abundance scale a of the Si IX channels to produce a forward model mismatch between the model used to generate synthetic observations and that used to train CoroNeRF. Because this ablation has no added observation noise, $I_c^{\text{obs}} = I_c^{\text{true}} = aI_c^{\text{raw}}$ for the Si IX channels with varying a while the Fe XIII channels remain unscaled; the inversion renderer continues to use the nominal $a = 1$ for all channels. Our study has 5 experiment families that vary $a \in \{0.6, 0.8, 1.0, 1.2, 1.4\}$, each trained with 3 seeds for a total of 15 runs. Figure 18 shows sample reconstructions at $1.51\ \mathrm{R}_\odot$, and Table 17 details the field-space and image-space errors.

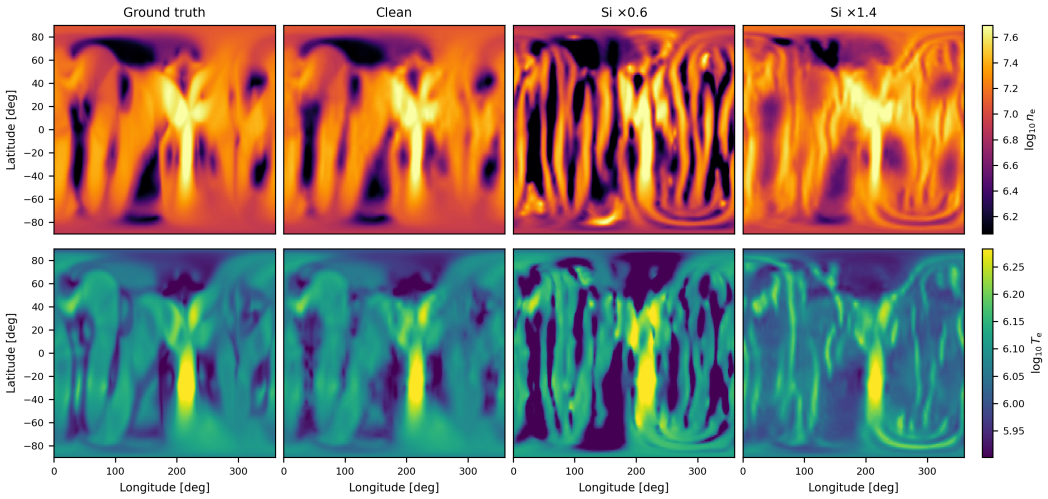

Figure 18: Errors in physical field reconstruction due to forward model mismatch. The columns show ground truth and the reconstructed density (top row) and temperature (bottom row) for different abundance mismatch scales. We see a clear failure mode when the Si IX intensities are scaled by $\pm 40\%$: the reconstructed field has spurious sharp features that do not exist in the ground truth fields.

a	asinhErr(I) ↓	n_e MAE ↓	n_e ME	T_e MAE ↓	T_e ME
0.6	$0.030 \pm < 0.001$	0.301 ± 0.006	-0.211 ± 0.004	0.076 ± 0.001	-0.030 ± 0.002
0.8	$0.017 \pm < 0.001$	0.146 ± 0.005	-0.076 ± 0.004	0.036 ± 0.001	-0.010 ± 0.001
1.0	$\mathbf{0.006 \pm < 0.001}$	$\mathbf{0.039 \pm < 0.001}$	-0.001 ± 0.001	$\mathbf{0.014 \pm < 0.001}$	$\mathbf{0.000 \pm < 0.001}$
1.2	$0.016 \pm < 0.001$	0.092 ± 0.001	0.062 ± 0.001	$0.023 \pm < 0.001$	$-0.002 \pm < 0.001$
1.4	$0.027 \pm < 0.001$	0.142 ± 0.001	0.107 ± 0.001	$0.031 \pm < 0.001$	$-0.006 \pm < 0.001$

Table 17: Abundance-mismatch benchmark. We scale the Si IX intensity channels in the synthetic observations by a while evaluating against the clean ground-truth physical fields. MAE reports inner-band absolute log-field error, while ME reports signed inner-band errors. The clean setting, corresponding to scale 1.0, gives the lowest image and field errors. Scaling the Si IX channels below or above the correct value produces systematic signed density errors, indicating that forward-model mismatch can be absorbed into incorrect latent physical fields.

D.8 Rendered Views

This section provides renders of the column density in Figure 19, total-emissivity weighted temperature in Figure 20, and total emissivity in Figure 21, for selected conditions and longitude view angles.

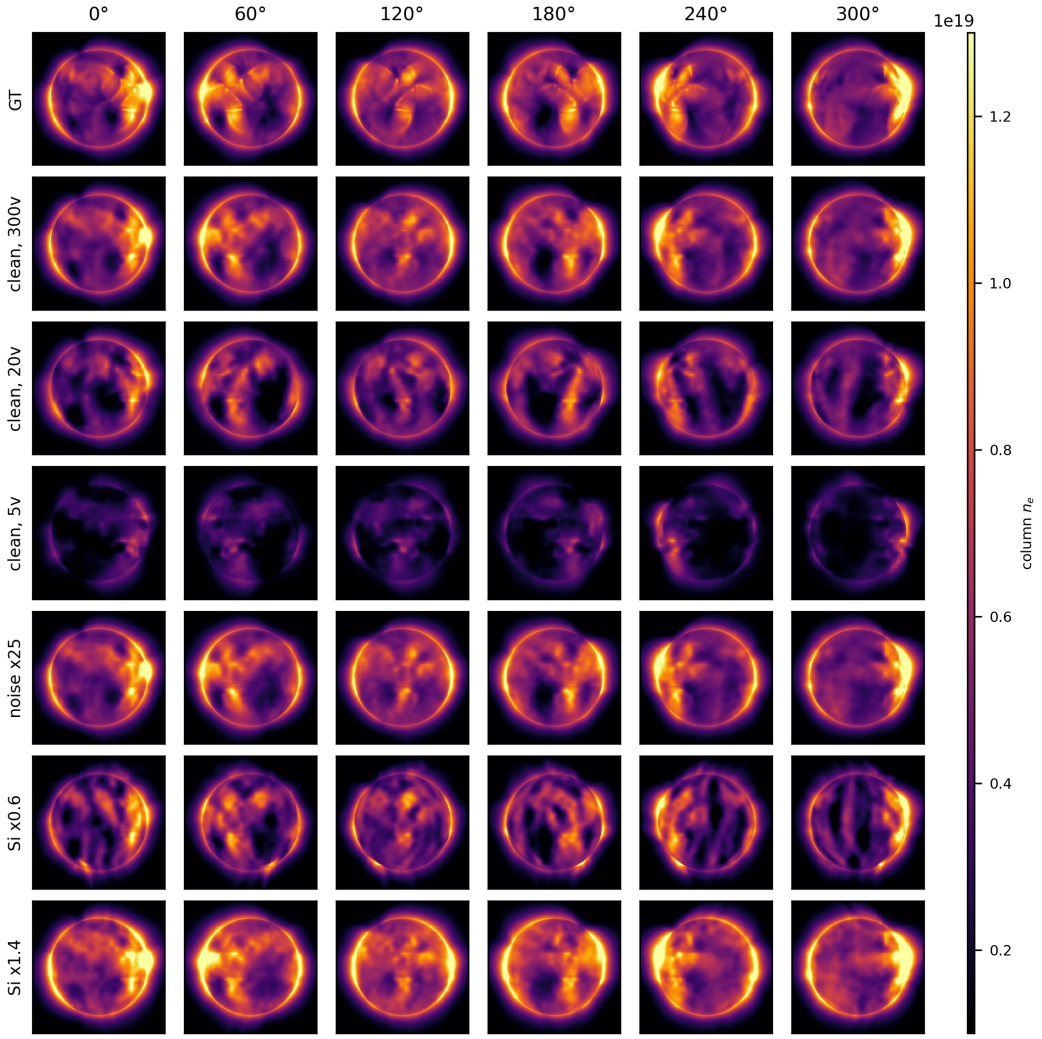

Figure 19: LOS projections of recovered column density for selected conditions and view angles.

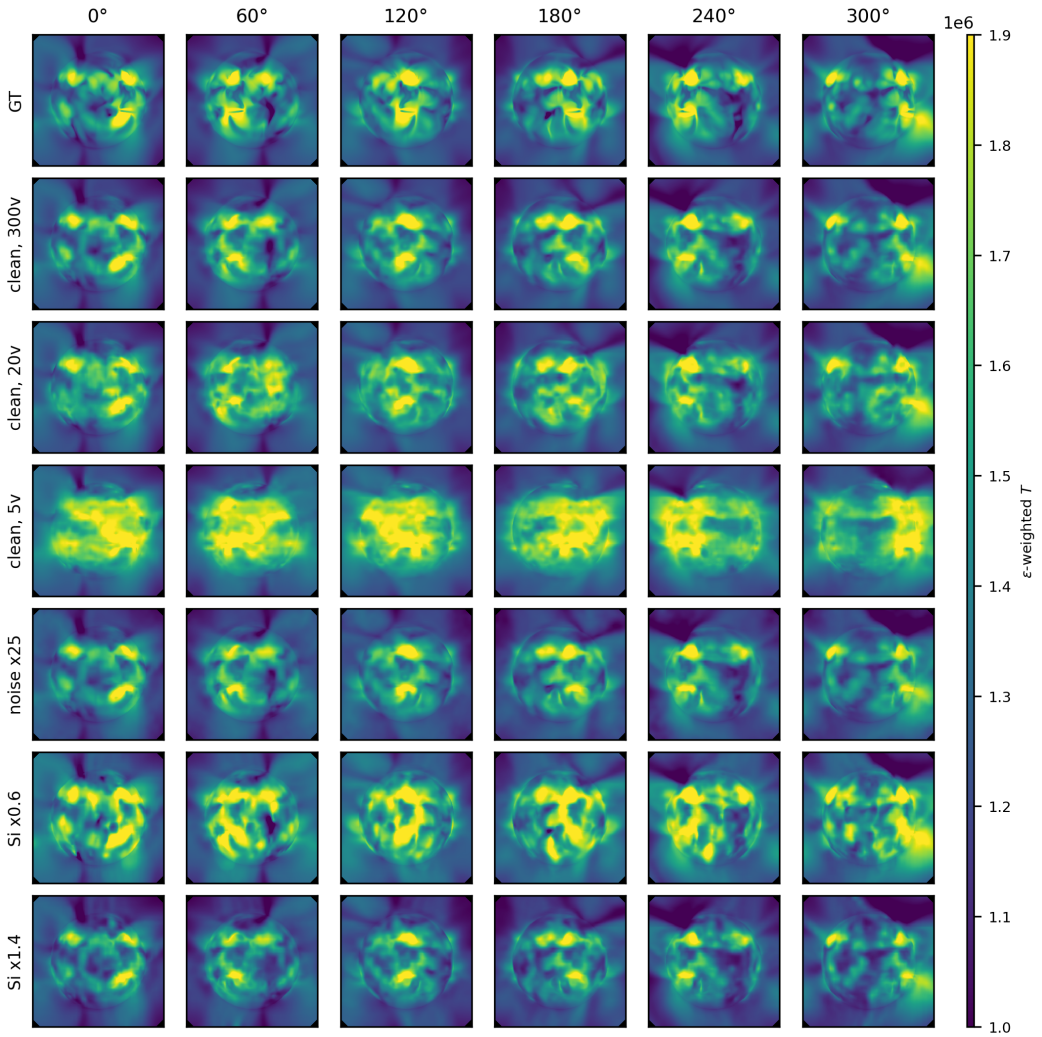

Figure 20: LOS projections of recovered emissivity-weighted temperature for selected conditions and view angles.

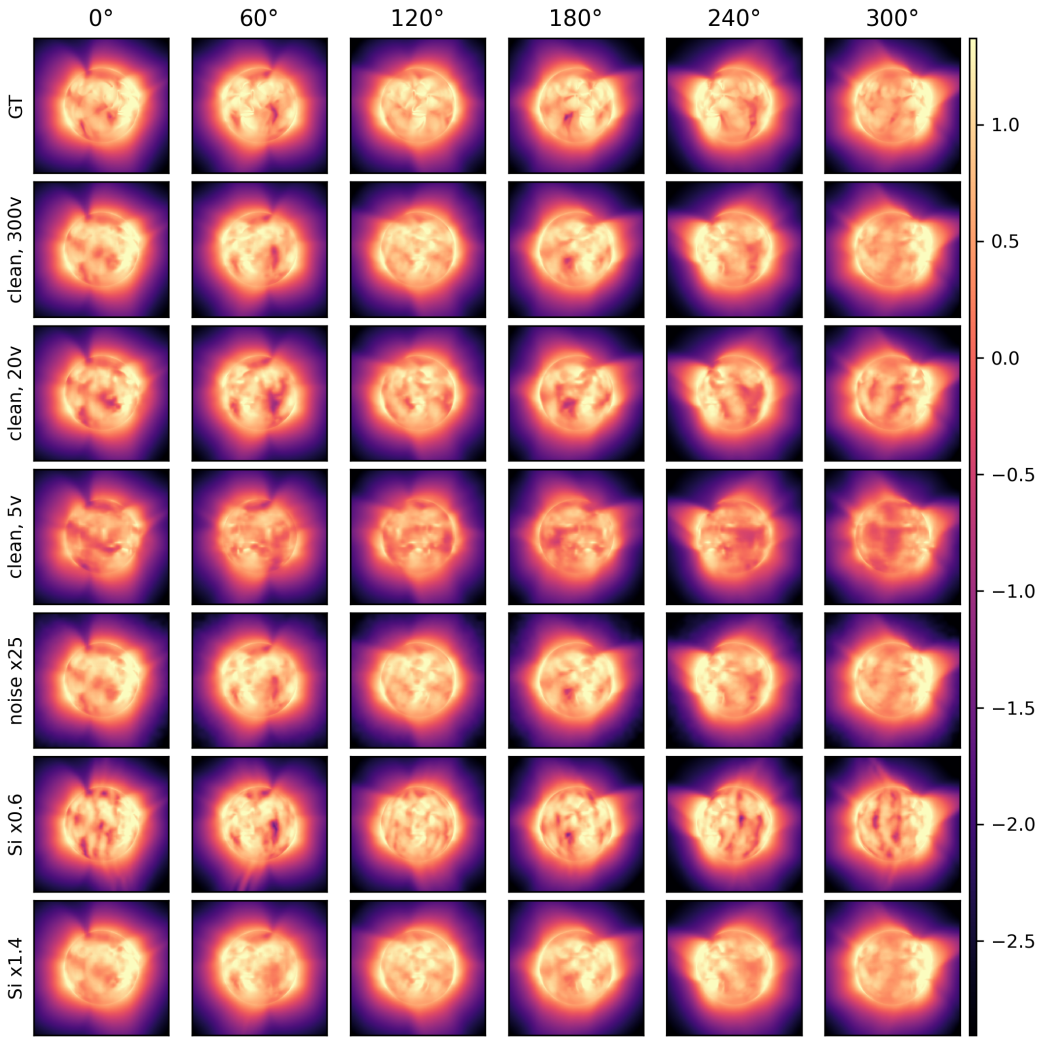

Figure 21: LOS projections of total-emissivity log-intensities for selected conditions and view angles.