NEURAL FIELDS FOR NV-CENTER INVERSE SENSING

Anonymous authors
Paper under double-blind review

ABSTRACT

Inverse problems in scientific sensing are often solved with either hand-designed regularizers or supervised networks trained on simulated labels, yet both can fail when the forward model is nonlinear, spectrally coupled, and physically delicate. We study this issue for noise sensing based on nitrogen-vacancy (NV) centers in diamond, where a quantum sensor measures magnetic-noise spectra generated by sparse spin sources. We show that replacing a common scalar/coherent forward approximation with a tensor power-summed dipolar operator changes the inverse landscape and exposes a center-collapse failure mode in free-density optimization. We propose NeTMY, an amortization-free coordinate neural field coupled to the differentiable NV forward model, with annealed positional encoding, multi-scale optimization, sparsity/gating, and spectrum-fidelity losses. On a 512-sample benchmark generated by a direct source-side simulator and inverted under the tensor operator, NeTMY achieves the best localization and distributional metrics. Mechanism experiments show that NeTMY's parameterization filters the raw density-space gradient and avoids the center collapse reached by free-density solvers. These results position NV quantum sensing as a testbed for physics-faithful neural inverse problems.

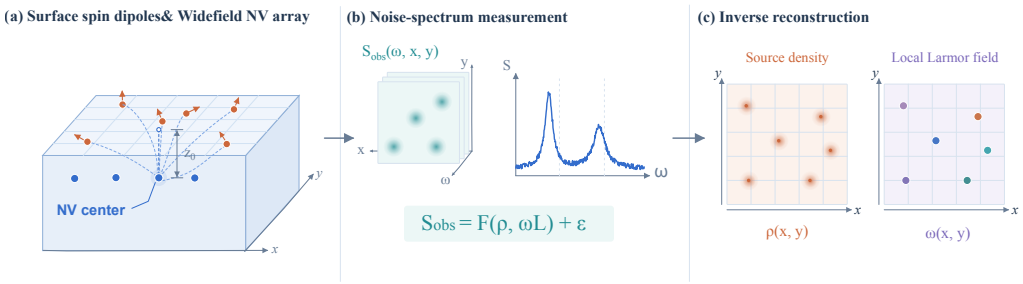


Figure 1: NV relaxometry maps sparse fluctuating spins to a noisy frequency-domain measurement. The inverse problem aims to recover the spin density and local Larmor response from the spectrum.

1 INTRODUCTION

Scientific machine learning is increasingly being used in settings where the quantity of interest is never observed directly. Instead, one observes the output of an instrument: a low-dimensional, noisy, spatially blurred, and often spectrally transformed measurement generated by a known but imperfect physical process (Arridge et al., 2019; Karniadakis et al., 2021). This setting differs from standard supervised prediction. Paired labels are expensive or impossible to obtain, synthetic data may not match experimental conditions, and seemingly minor approximations in the forward model can change not only the measurement distribution, but also the geometry of the inverse objective. As a result, a method that performs well under a convenient simulator may fail when the physical operator is made more faithful (Kaipio & Somersalo, 2007). This interaction between forward-model fidelity, representation, and optimization is a central challenge for scientific ML.

We study this issue in nitrogen-vacancy (NV) center noise sensing (Doherty et al., 2013; Degen et al., 2017). This problem is closely related to the better-established task of physical reconstruction from static magnetic-field images, where NV magnetic maps have been inverted to recover underlying

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.

current or magnetization distributions (Midha et al., 2024). Here, however, the target is not a static field map but a sparse distribution of fluctuating spin sources and their local Larmor response inferred from frequency-dependent magnetic-noise spectra. As illustrated in Figure 1, in this setting, NV centers measure spectra induced by nearby fluctuating spins, while the scientific target is the spatial and spectral structure of the spin sources (Casola et al., 2018; Mzyk et al., 2022; Tetienne et al., 2013). Recent non-ML NV noise/correlation measurements further motivate this reconstruction setting by showing that multiplexed NV platforms can access noise spectra and magnetic-field correlations at multiple spatial locations (Cheng et al., 2025). The resulting inverse problem is highly ill-conditioned (Fuchs & Shmueli, 2023): the dipolar response is low-pass and rapidly decaying (Mzyk et al., 2022), nearby sources can merge below the effective point-spread width (Grech et al., 2008), finite-window effects create center bias, and spectrum normalization introduces scale ambiguity and nonlocal gradients (Zhang et al., 2018). Thus, even with a differentiable forward model, direct reconstruction from magnetic noise spectra can have misleading descent directions and physically implausible local optima.

Classical sparse inverse methods such as Tikhonov regularization (Hoerl & Kennard, 1970b;a) or ADMM (Neal et al., 2011) optimize the density field directly, but in our setting, this free-density parameterization can follow pathological gradients and collapse to central artifacts under the physical objective. Supervised image-to-image predictors, such as U-Nets (Ronneberger et al., 2015) or GANs (Goodfellow et al., 2014; Radford et al., 2015), can be strong when paired training data match the test distribution, but paired spin-density labels are scarce, and synthetic dense training distributions may not generalize to sparse scenes. Generic untrained neural priors (Heckel & Hand, 2018; Ulyanov et al., 2018) avoid supervision, but do not by themselves address the tensor relaxometry operator, Larmor-spectrum coupling, or scale ambiguity of this sensing problem.

We propose NeTMY, an amortization-free neural-field solver for NV relaxometry inversion. NeTMY represents the unknown density and Larmor fields with a coordinate MLP and optimizes its parameters separately for each measurement through a differentiable tensor power-summed forward operator. The method uses annealed positional encoding (Tancik et al., 2020) and multiscale optimization to recover sparse high-frequency structure while maintaining stable descent, together with spectrum consistency, sparsity, TV regularization (Rudin et al., 1992), density gating, and measurement-derived scale correction. The forward operator sums tensor-channel magnetic-noise powers rather than squaring a coherently summed scalar field, avoiding nonphysical cross terms that appear in the simplified solver (Guan et al., 2024; Midha et al., 2024). This produces a more faithful inverse objective and exposes failure modes that are hidden or distorted under the scalar approximation.

Our mechanism experiments link NeTMY's advantage to optimization geometry. Direct density optimization follows the raw density-space gradient, whereas a neural-field update induces an image-space step filtered by the representation Jacobian. Our contributions are summarized as follows:

- We formulate NV noise sensing inversion as a physics-faithful differentiable inverse problem and introduce a tensor power-summed forward operator that avoids nonphysical cross terms in simplified scalar solvers.
- We propose NeTMY, an amortization-free coordinate neural-field solver that reconstructs sparse density and Larmor fields from individual measurements without paired density labels.
- We show empirically and mechanistically that forward-model fidelity changes inverse-problem landscapes and solver rankings, trace free-density center collapse to its first-step gradient, isolate the contributing NeTMY components by ablation, and check both operators against real α-$RuCl_3$ relaxometry data.

2 Related Work

NV centers have become a standard tool for nanoscale magnetic sensing, mapping local magnetic-noise into optical / T_1 and T_2 measurement with single-defect resolution (Casola et al., 2018; Degen et al., 2017; Doherty et al., 2013; Tetienne et al., 2013; Van der Sar et al., 2015). Recent applications include spin-bath sensing (Mzyk et al., 2022; Aharon et al., 2019; Homrighausen et al., 2023), nanoscale sensing for condensed matter physics and materials science (Rovny et al., 2024), and high-resolution nanoscale measurements (Liu et al., 2026; Mamin et al., 2013; Ávila Jiménez et al., 2025; Smooha et al., 2026; Tsukamoto et al., 2022; Midha et al., 2024). These works establish the sensing physics and forward models but do not study learning-centric inverse algorithms or

operator-fidelity-induced optimization-geometry changes, while we treat NV noise sensing as a differentiable inverse-problem benchmark.

Sparse optimization algorithms in classical inverse problems, such as Tikhonov and total-variation regularization (Rudin et al., 1992; Bunks et al., 1995), ADMM (Neal et al., 2011), and proximal methods (Chambolle & Pock, 2011; Chi et al., 2019) are standard for ill-posed reconstruction (Arridge et al., 2019; Kaipio & Somersalo, 2007; Virieux & Operto, 2009), with a long line of compressed-sensing (Candès et al., 2006; Donoho, 2006) and sparse-recovery results (Ahmed et al., 2013; Jagatap & Hegde, 2019; Bora et al., 2017; Ling & Strohmer, 2015; Lee et al., 2016). These tools have strong guarantees in convex / well-conditioned linear inverse settings, but the NV inverse objective is nonlinear, max-normalized, finite-windowed, tensorial, and spectrally coupled. Standard convexity-based stability analysis does not transfer, and free-density local methods can be attracted to center-biased gradients.

Data-driven inverse solvers, such as supervised image-to-image networks (Ronneberger et al., 2015; McCann et al., 2017), learned unrolling and primal-dual networks (Monga et al., 2021; Hammernik et al., 2018; Yang et al., 2018; Gupta et al., 2018), and GAN-based reconstruction (Ongie et al., 2020; Gilton et al., 2021) provide strong baselines when paired training data matches test data. Plug-and-play (PnP) and regularization-by-denoising (RED) priors (Venkatakrishnan et al., 2013; Romano et al., 2017; Mataev et al., 2019; Zhang et al., 2017; 2022) instead insert a denoiser, trained without paired inverse-problem data, into an iterative solver; we compare against both in Appendix F.5. The field has separately documented reconstruction instabilities under perturbation and model shift (Antun et al., 2020), robustness gaps (Henderson et al., 2018; Recht et al., 2019), and benchmarking pitfalls (Liao et al., 2021; Raji et al., 2021; Dehghani et al., 2021). In NV relaxometry, paired spin-density labels are scarce (Mzyk et al., 2022). We document train-on-dense / test-on-sparse failure modes and use per-instance physics-driven optimization to side-step them.

Untrained neural priors, neural fields, and physics-informed differentiable scientific ML. On the parameterization-as-prior side, Deep Image Prior (Ulyanov et al., 2018), Deep Decoder (Heckel & Hand, 2018; Heckel & Soltanolkotabi, 2020), Fourier features / SIREN (Tancik et al., 2020; Sitzmann et al., 2020), NeRF (Mildenhall et al., 2021), and INR-based reconstruction (Park et al., 2021; Martel et al., 2021; Molaei et al., 2023; Shi et al., 2024; Reed et al., 2021; Shen et al., 2022; Alkhouri et al., 2025; Habring & Holler, 2024; Zhao et al., 2024) demonstrate that the structure of a coordinate or generator network itself biases reconstruction. On the physics-in-loop side, physics-informed neural networks (Raissi et al., 2019; Karniadakis et al., 2021; Cuomo et al., 2022; Inda et al., 2022) and differentiable simulators (Zhong et al., 2026; Zhao et al., 2025) incorporate known physical laws as losses or as differentiable forward models. Generic neural priors do not encode NV tensor physics, Larmor-spectrum coupling, or scale ambiguity, and the physics-informed thread has rarely studied how forward-operator fidelity reshapes optimization geometry, which we measure here.

Adjacent inverse problems. Recovering sources from indirect, spatially blurred measurements recurs in PALM/STORM single-molecule localization microscopy (Betzig et al., 2006; Rust et al., 2006), whose computational methods range from compressed sensing to learned dense-emitter inference (Zhu et al., 2012; Speiser et al., 2021); sparse-spike super-resolution theory (Candès & Fernandez-Granda, 2014); biomagnetic dipole-source inversion (Hämäläinen et al., 1993); radio-interferometric deconvolution (Högbom, 1974; Wiaux et al., 2009); and static-field NV magnetometry (Broadway et al., 2020). The scalar operator $\mathcal{F}_1$ observes only the squared magnitude of a linear map of ρ, as in phase retrieval (Fienup, 1982; Shechtman et al., 2015). The NV pipeline differs in that density is coupled to an unknown Larmor field through tensor power-summed dipolar kernels and a Lorentzian response, and the fidelity applies max-normalization and a log transform; the optimization behavior here concerns this pipeline. What may transfer is the template (a structured per-instance parameterization, a differentiable instrument model, and benchmarks that generate and invert with different operators), instantiated by swapping the forward operator. We offer this as motivation rather than a result tested.

3 Preliminaries and Problem Formulation

This section fixes the measurement model, the inverse objective, and the sources of ill-posedness that Section 4 addresses; derivations and diagnostics are in Appendices B, C, and D.

3.1 MEASUREMENT MODEL

Let $\Omega \subset \mathbb{R}^2$ be a finite spatial grid and let $\rho : \Omega \to \mathbb{R}_{\geq 0}$ denote the unknown spin-source density. A widefield NV array sits at standoff height z_0 above Ω and reads out a frequency-dependent noise spectrum $S_{\mathrm{obs}}(\omega, \mathbf{r}) = \mathcal{F}(\rho, \omega_L)(\omega, \mathbf{r}) + \varepsilon$ on a discrete grid $W \subset \mathbb{R}_+$, where $\omega_L : \Omega \to \mathbb{R}_+$ is a per-pixel Larmor field encoding local detuning, $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is sensor noise, and $\mathcal{F}$ is a differentiable forward operator. The dipolar Green tensor coupling sample spins to the NV z-axis is

$$G_{ia}(\mathbf{R}) \;=\; \frac{\mu_0}{4\pi}\,\frac{3R_iR_a - \|\mathbf{R}\|^2\,\delta_{ia}}{\|\mathbf{R}\|^5}, \qquad i, a \in \{x, y, z\}, \tag{1}$$

where $\mathbf{R}(\mathbf{r}, \mathbf{r}_{\mathrm{src}}) = (\mathbf{r} - \mathbf{r}_{\mathrm{src}}, z_0)$, μ_0 is the vacuum permeability and δ_{ia} the Kronecker delta. The spectral response of a dipole resonant at ω_L with linewidth γ is the Lorentzian $L(\omega; \omega_L) = \gamma^2/((\omega - \omega_L)^2 + \gamma^2)$ (Tetienne et al., 2013; Doherty et al., 2013; Casola et al., 2018; Mzyk et al., 2022).

We compare two FFT-based forward operators that differ in how transverse-field noise is aggregated across the channels of G and across distinct sources:

$$\begin{aligned}
\mathcal{F}_1(\rho, \omega_L)(\omega, \mathbf{r}) \;&=\; \big|(G_{\mathrm{nv}} * \rho)(\mathbf{r})\big|^2\, L\big(\omega; \omega_L(\mathbf{r})\big), \\
\mathcal{F}_2(\rho, \omega_L)(\omega, \mathbf{r}) \;&=\; \Big[\sum_{a \in \{x,y,z\}} \big(|G_{az}|^2 * \rho\big)(\mathbf{r})\Big]\, L\big(\omega; \omega_L(\mathbf{r})\big).
\end{aligned} \tag{2}$$

where $*$ denotes 2D spatial convolution. The scalar/coherent operator $\mathcal{F}_1$ squares a single NV-axis projection $G_{\mathrm{nv}} = \sum_a n_a G_{az}$ ($\mathbf{n} = (0, 0, 1)$ here) after spatial superposition, while the tensor/incoherent operator $\mathcal{F}_2$ sums per-channel power kernels and treats distinct sources as independent thermal fluctuators. Their pointwise difference is a kernel-product cross-term coupling pairs of distinct sources, which is the algebraic origin of the inverse-landscape difference exploited in Section 5. We treat $\mathcal{F}_2$ as the physically appropriate inversion operator; because it factorizes the Lorentzian response outside the kernel convolution, it is exact for locally constant ω_L and controlled when $\chi \sim \Delta\omega_L^{(\mathrm{kernel})}/\gamma \ll 1$ (Appendix B.8). A slower direct simulator $\mathcal{F}_3$ evaluates the source-side superposition without FFT factorization (Appendix B.4). To avoid the inverse crime of reusing the inversion operator for data generation (Kaipio & Somersalo, 2007), the main benchmark generates spectra with $\mathcal{F}_3$ and inverts under $\mathcal{F}_1$ or $\mathcal{F}_2$; matched-operator results are a complementary view. Derivations of $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ and their regimes of applicability are in Appendix B.

3.2 INVERSE PROBLEM

Following standard practice in NV sensing (Kumar et al., 2024; Midha et al., 2024; Casola et al., 2018), the spectrum is summed across W and max-normalized before being compared on a logarithmic scale. Writing $N(\mathbf{r}; S) = \sum_{\omega \in W} S(\omega, \mathbf{r})$ and $\widehat{N}(\mathbf{r}; S) = N(\mathbf{r}; S)/\max_{\mathbf{r}'} N(\mathbf{r}'; S)$, the data-fidelity term $\mathcal{D}(\mathcal{F}(\rho, \omega_L), S_{\mathrm{obs}})$ is the pixelwise log-MSE between $\widehat{N}(\,\cdot\,; \mathcal{F}(\rho, \omega_L))$ and $\widehat{N}(\,\cdot\,; S_{\mathrm{obs}})$ (full form in Appendix C). The canonical regularized inverse problem is

$$\min_{\rho \geq 0,\ \omega_L} \;\mathcal{D}\big(\mathcal{F}(\rho, \omega_L), S_{\mathrm{obs}}\big) \;+\; \lambda_1 \|\rho\|_1 \;+\; \lambda_{\mathrm{TV}}\,\mathrm{TV}(\rho), \tag{3}$$

where $\|\rho\|_1 = \sum_{\mathbf{r} \in \Omega} |\rho(\mathbf{r})|$ is the ℓ_1 norm of the discretized density, $\mathrm{TV}(\rho) = \sum_{\mathbf{r} \in \Omega} \big(|\rho(\mathbf{r} + \mathbf{e}_x) - \rho(\mathbf{r})| + |\rho(\mathbf{r} + \mathbf{e}_y) - \rho(\mathbf{r})|\big)$ is the anisotropic total variation, $\mathbf{e}_x, \mathbf{e}_y$ are unit grid steps, and $\lambda_1, \lambda_{\mathrm{TV}} \geq 0$ are scalar weights. Every method in our benchmark shares Eq. 3; the classical baselines (Tikhonov, ADMM) and the proposed neural-field method differ in how ρ is parameterized and in which method-specific physics losses they add (Section 4).

Because of the max-normalization in $\widehat{N}$, the linearity $\mathcal{F}_2(c\rho, \omega_L) = c\,\mathcal{F}_2(\rho, \omega_L)$ leaves $\widehat{N}$ invariant for every $c > 0$, so the pre-normalization scale of ρ is unidentifiable from $\widehat{N}$ alone (Ahmed et al., 2013; Ling & Strohmer, 2015). Section 4 resolves the scale after optimization (Appendix C).

3.3 ILL-POSEDNESS

Even with $\mathcal{F}_2$ in place, four structural properties make Eq. 3 severely ill-conditioned (formal statements and diagnostics in Appendix D).

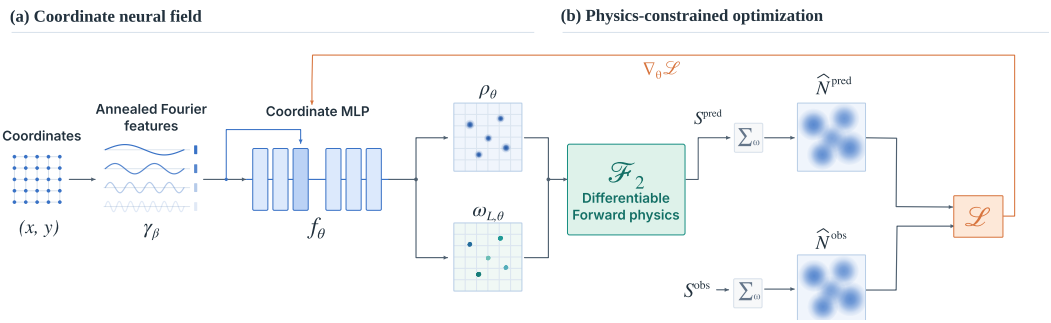


Figure 2: NeTMY pipeline. Coordinates pass through annealed Fourier features and a coordinate MLP f_θ to a density ρ and Larmor field ω_L; a differentiable forward operator $\mathcal{F}$ predicts a spectrum compared with the observation under loss $\mathcal{L}$. The same MLP is trained per instance across two resolutions in sequence; gradients flow back through the operator, with no paired training data.

Exponential frequency suppression. Each tensor channel $|G_{az}|^2$ in $\mathcal{F}_2$ has 2D Fourier transform with exponential envelope e^{-kz_0} up to algebraic factors in k, where $k = \|\mathbf{k}\|$ is the magnitude of the spatial-frequency vector $\mathbf{k}$ (Van der Sar et al., 2015), so high-spatial-frequency density features carry vanishingly little forward signal (Lemma 1).

Finite-window center bias. The convolutional footprint of a source pixel is more fully observable at the window center than at the boundary, so the per-pixel forward sensitivity $\|\partial\mathcal{F}/\partial\rho(\mathbf{r}_0)\|$ is larger near the center, producing a centered descent direction even from a uniform initialization.

Max-normalization peak coupling. The derivative of the max-normalized noise map contains a nonlocal term concentrated at the current peak pixel, sharpening any incipient peak and amplifying the center bias once a peak appears.

Source merging and Larmor ambiguity. Source pairs separated by less than $\mathcal{O}(z_0)$ merge and cannot be disambiguated, and the Larmor field ω_L is identifiable only on the support of ρ.

These pathologies make the parameterization of ρ matter more than regularizer tuning; Section 4 addresses frequency suppression, center bias, and peak coupling through the parameterization and its curriculum, and Larmor ambiguity by restricting Larmor updates to the predicted support.

4 Method

NeTMY solves the inverse problem of Eq. 3 per measurement: rather than learning an amortized inverse map from a paired dataset, it represents the unknown density ρ and Larmor field ω_L as the output of a coordinate MLP and optimizes the MLP parameters against a single observed spectrum through a differentiable forward operator $\mathcal{F} \in \{\mathcal{F}_1, \mathcal{F}_2\}$. The field components and schedule below each target one pathology of Section 3.3 through the parameterization and its optimization rather than through additional regularizer tuning; Appendix D.5 summarizes the pathology-to-component map. Figure 2 gives an overview; Appendix E gives full details and derivations.

4.1 Coordinate Neural Field

We parameterize the unknowns as an MLP of grid coordinates,

$$f_\theta : (x, y) \;\mapsto\; \big(\rho_\theta(x, y),\ \omega_{L,\theta}(x, y)\big), \qquad (x, y) \in \Omega, \tag{4}$$

with the network input augmented by annealed Fourier positional features $\gamma_\beta(x, y)$ whose high-frequency bands are gated on through a scalar β (Tancik et al., 2020; Park et al., 2021; Sitzmann et al., 2020). Annealing addresses frequency suppression by fitting low-$\|\mathbf{k}\|$ structure of ρ before high-$\|\mathbf{k}\|$ detail. The MLP is a backbone with a skip connection re-injecting the encoded input partway through.

The two output heads use bounded transforms. The density head is a gated softplus,

$$\rho_\theta(x, y) \;=\; \mathrm{softplus}\big(h_\rho(x, y)\big)\, \sigma\big(g(x, y)\big), \tag{5}$$

where h_ρ is the raw density logit, g is a separate gate logit, and σ is the sigmoid. The softplus enforces $\rho_\theta \geq 0$, while the gate factor $\sigma(g) \in (0, 1)$ lets the network drive pixel regions to near zero

without saturating, which targets peak coupling. The Larmor head sigmoid-maps a logit into the device-determined band $[\omega_{L,\min}, \omega_{L,\max}]$ and is multiplied by a stop-gradient mask of the predicted support of ρ (Appendix E.1), so that ω_L receives gradients only where the data constrains it; this addresses Larmor ambiguity.

4.2 Multiscale Curriculum

Optimization proceeds in two stages on the same MLP parameters, with the coordinate grid resolution increasing between stages. Stage 1 fits coarse spatial structure under the coarsened forward operator at the base learning rate; Stage 2 refines high-frequency detail at the higher resolution at a reduced learning rate. By the bandlimit argument of Lemma 1, low-$\|\mathbf{k}\|$ modes carry the bulk of the forward signal and are well-conditioned, so fitting them first delivers a coarse minimum closer to the true support before the high-frequency curriculum activates. Per-stage settings are in Table 6.

4.3 Per-Measurement Objective

NeTMY minimizes an extension of Eq. 3 by gradient descent on θ alone, with no oracle access to $\rho_\star$ or $\omega_{L,\star}$, adding two physics losses to the canonical fidelity $\mathcal{D}$,

$$\mathcal{L}(\theta) \;=\; \mathcal{D}\big(\mathcal{F}(\rho_\theta, \omega_{L,\theta}), S_{\text{obs}}\big) \;+\; \lambda_1 \,\|\rho_\theta\|_1 \;+\; \lambda_{\text{TV}}\,\text{TV}(\rho_\theta) \;+\; \lambda_N\,\mathcal{R}_{\text{nm}} \;+\; \lambda_\rho\,\mathcal{R}_{\text{ds}}, \quad (6)$$

where $\mathcal{R}_{\text{nm}}$, weighted by $\lambda_N \geq 0$, is a mean-normalized noise-map loss, a scale-stable companion to $\mathcal{D}$ that anchors gradients on the support, and $\mathcal{R}_{\text{ds}}$, weighted by $\lambda_\rho \geq 0$, is a direct-density loss matching ρ_θ^2 to the mean-normalized observed noise map, an oracle-free support proxy. The loss weights are fixed across all runs; the balance between $\mathcal{D}$ and $\mathcal{R}_{\text{nm}}$ shifts from support discovery in Stage 1 to refinement in Stage 2 (Appendix E.4, Table 7).

Energy-anchored scale correction. The pre-normalization scale of ρ is unidentifiable from $\widehat{N}$ (Section 3.2; Proposition 1). After optimization, NeTMY rescales the predicted density once, $\widehat{\rho} \leftarrow \alpha\widehat{\rho}$ with $\alpha = E_{\text{obs}}/E_{\text{pred}}$ for $\mathcal{F}_2$, where $E_{\text{obs}} = \sum_{\omega,\mathbf{r}} S_{\text{obs}}$ and $E_{\text{pred}} = \sum_{\omega,\mathbf{r}} \mathcal{F}(\widehat{\rho}, \omega_{L,\theta})$; the main-text tables use the operator-matched form, which for the quadratic $\mathcal{F}_1$ is $\sqrt{E_{\text{obs}}/E_{\text{pred}}}$ (Appendix C, Eq. 21). The correction mainly affects density MSE; the primary localization metrics are insensitive to absolute scale. Applying it after optimization rather than as a penalty in Eq. 6 leaves the data-fidelity gradient unchanged and gives identical post-processing across baselines (Appendix E.5).

4.4 Optimization Geometry: a Filtering View of the Update

Free-density solvers (Tikhonov, ADMM) minimize Eq. 3 by stepping ρ directly, $\rho_{t+1} = \rho_t - \eta\,\nabla_\rho\mathcal{L}(\rho_t)$, and therefore execute the raw density-space gradient, which inherits center bias and peak coupling. NeTMY instead steps the parameters θ, with chain rule

$$\Delta\rho \;\approx\; J_\theta\,\Delta\theta \;=\; -\eta\,J_\theta J_\theta^\top\,\nabla_\rho\mathcal{L} \;=\; -\eta\,G_\theta\,\nabla_\rho\mathcal{L}, \qquad J_\theta \;=\; \partial f_\theta/\partial\theta, \quad (7)$$

so the realized image-space update is the raw density gradient *filtered* by the positive semidefinite kernel $G_\theta = J_\theta J_\theta^\top \in \mathbb{R}^{|\Omega|\times|\Omega|}$ induced by the parameterization (Appendix E.6). For a coordinate MLP with annealed Fourier features, G_θ is a smoothing operator that couples pixels: in the unpreconditioned first-order limit, a sharp center spike in $\nabla_\rho\mathcal{L}$ is redistributed by G_θ rather than executed as a single-pixel update, so the centered iter-0 gradient caused by center bias need not be amplified by peak coupling. NeTMY therefore changes how the centered gradient is realized in image space, which Section 5.3 tests.

5 Experiments

We evaluate NeTMY and free-density baselines on a cross-fidelity synthetic benchmark and on a real NV-relaxometry dataset, inverting under both $\mathcal{F}_1$ and $\mathcal{F}_2$. Implementation details, per-sample results, hyperparameter sweeps, the supervised-baseline study, and runtime are in Appendix F.

5.1 Setup

Datasets. Two regimes (construction in Appendix F.1). A cross-fidelity benchmark of 512 synthetic samples in which spectra are generated by the source-side direct simulator $\mathcal{F}_3$ (Eq. 11) and methods invert under either $\mathcal{F}_1$ or $\mathcal{F}_2$, avoiding the inverse-crime risk of reusing the inversion operator as the data-generation operator (Kaipio & Somersalo, 2007). The samples cover nine scene classes

Table 1: Cross-fidelity benchmark on $\mathcal{F}_3$-generated samples; methods invert under $\mathcal{F}_1$ or $\mathcal{F}_2$. Best per metric in bold. We report the mean and 95% CI across 3 seeds.

	Inversion under $\mathcal{F}_1$ (scalar)				Inversion under $\mathcal{F}_2$ (tensor)			
Method	GMSD↓	F1↑	SWD↓	MSE↓	GMSD↓	F1↑	SWD↓	MSE↓
Tikhonov	**0.178±0.041**	0.465±0.436	0.119±0.043	0.0119±0.0155	0.202±0.043	0.762±0.318	0.100±0.041	0.0082±0.0118
ADMM	0.235±0.088	0.267±0.389	0.052±0.028	0.0183±0.0177	0.197±0.061	0.084±0.208	0.134±0.048	3.8364±4.9726
NeTMY (Ours)	0.189±0.116	**0.733±0.334**	**0.051±0.048**	**0.0066±0.0080**	**0.125±0.058**	**0.882±0.102**	**0.036±0.023**	**0.0078±0.0082**

Table 2: Matched-operator benchmarks. The matched operator regime is operator-dependent.

	$\mathcal{F}_1/\mathcal{F}_1$		$\mathcal{F}_2/\mathcal{F}_2$	
Method	MSE↓	NoiseMSE↓	MSE↓	NoiseMSE↓
Tikhonov	0.015	2.25	0.012	0.81
ADMM	**0.006**	**0.32**	0.256	22.9
HybridNeTMY	0.025	6.47	1.011	0.23
DeepDecoder	1.61	31.38	0.256	24.3
UntrainedGAN	0.026	43.0	0.026	24.2
NeTMY (Ours)	0.009	1.15	**0.004**	**0.21**

Table 3: Cumulative ablation. Each row removes one component cumulatively.

Configuration	MSE↓	SSIM↑
Full model	**.0039**	**.913**
−TV	.0042	.893
−annealed PE	.0086	.897
−PE	.0091	.895
−multiscale	.0103	.890
$-\ell_1$	.0104	.880
−gate	.0118	.876
$-\mathcal{R}_{\mathrm{ds}}$	.0116	.890

(few/medium/many sources × close/medium/far minimum separations relative to the standoff z_0). A *real-data cross-check* on α-$RuCl_3$ (8 NVs, Kumar et al. (2024)) with an independent SRIM depth prior 14 ± 5 nm. Two matched-operator benchmarks (generation and inversion by the same $\mathcal{F}_1$ or $\mathcal{F}_2$, 512 samples each, broader baseline coverage) are in Table 2 and Appendix F.4.

Baselines. Three reconstruction families: Tikhonov (free-density gradient descent with ℓ_2/TV) (Hoerl & Kennard, 1970b; Rudin et al., 1992); ADMM (variable-splitting with augmented Lagrangian and box constraints) (Neal et al., 2011); and NeTMY (this work). L-BFGS (Liu & Nocedal, 1989) serves as a quasi-Newton reference in the mechanism analysis. Untrained-prior (DeepDecoder (Heckel & Hand, 2018)) and supervised baselines (U-Net (Ronneberger et al., 2015), GAN (Goodfellow et al., 2014), HybridNeTMY) are reported in Appendix F.9. Learned-prior plug-and-play (PnP) and regularization-by-denoising (RED) solvers with a frozen pretrained denoiser (Venkatakrishnan et al., 2013; Romano et al., 2017; Zhang et al., 2022) are compared on a held-out subset in Appendix F.5, where neither leads any primary metric under either operator.

Metrics. Three primary metrics that emphasize sparse-localization and distributional fidelity over pixelwise mean-square error: GMSD (gradient-magnitude similarity deviation) (Xue et al., 2013); Hungarian F1 (one-to-one matched localization at fixed radius) (Kuhn, 1955); and Sliced Wasserstein distance (SWD) over 128 random projections (Bonneel et al., 2015). Pixelwise density MSE is a secondary metric included for comparability with prior NV-relaxometry work (Midha et al., 2024; Kumar et al., 2024); on sparse scenes, it is dominated by background and undersells localization quality. Definitions and matching tolerances are in Appendix F.3.

5.2 Operator Fidelity Reshapes the Reconstruction Benchmark

The inversion operator changes the ranking (Table 1). Moving from $\mathcal{F}_1$ to $\mathcal{F}_2$ inversion lifts Hungarian F1 and lowers SWD for two of three methods, with the largest gain for NeTMY (Hungarian F1: $0.733 \to 0.882$; SWD: $0.051 \to 0.036$), while ADMM's density MSE rises from 0.0183 to 3.8364 because the more faithful operator sharpens the centered collapse of Section 5.3. Under $\mathcal{F}_1$, NeTMY leads Hungarian F1, SWD, and MSE, and Tikhonov leads GMSD (0.178 vs 0.189, within the intervals); under $\mathcal{F}_2$, NeTMY tops every primary metric, with Hungarian F1 0.882 versus 0.762 and SWD 0.036 versus 0.100 for Tikhonov, the next-best method on both. Directional crosses between $\mathcal{F}_1$ and $\mathcal{F}_2$ (Appendix F.6) show that the ranking is operator-dependent in both directions: NeTMY leads GMSD and SWD when generating under $\mathcal{F}_1$ and inverting under $\mathcal{F}_2$, while the reverse direction favors the free-density solvers.

On the matched-operator benchmarks (Table 2), ADMM has the lowest MSE under $\mathcal{F}_1/\mathcal{F}_1$ (0.006; NeTMY second at 0.009) but degrades by two orders of magnitude under $\mathcal{F}_2/\mathcal{F}_2$, where NeTMY leads both columns. Table 1 is our primary evidence: its spectra come from the independent simulator

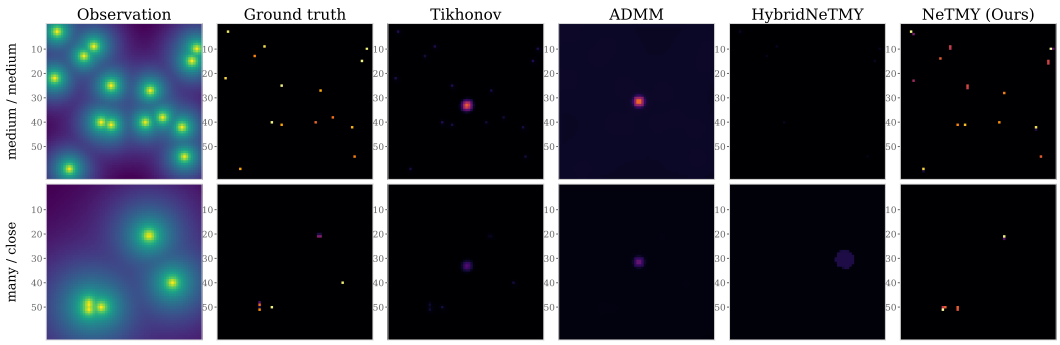


Figure 3: Qualitative reconstructions on two $\mathcal{F}_3$-generated samples inverted under $\mathcal{F}_2$. Free-density solvers center-collapse; HybridNeTMY and NeTMY preserve off-center structure.

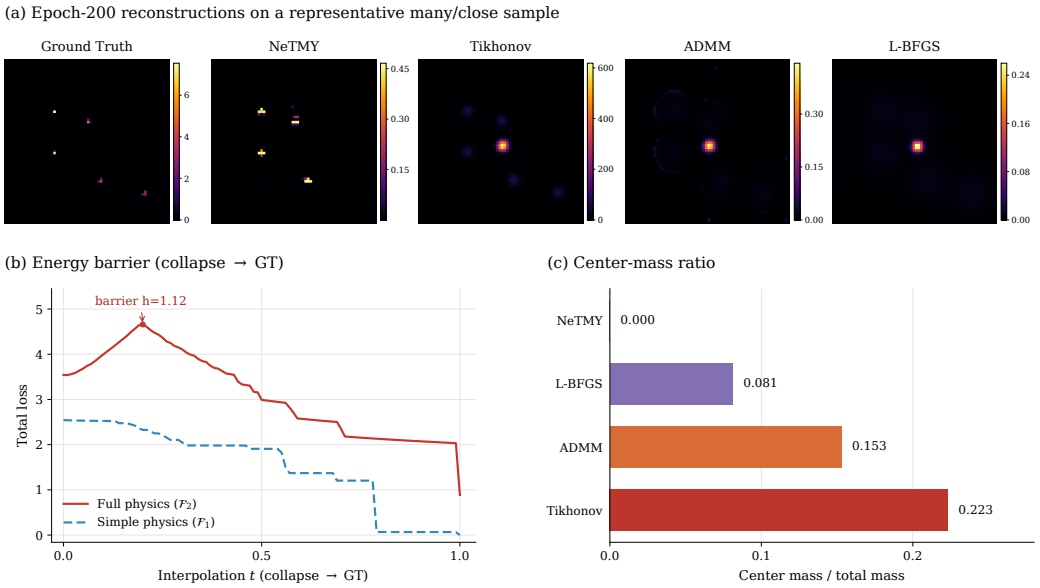


Figure 4: Optimization-geometry diagnostics. (a) Epoch-200 reconstructions. (b) Loss along $\rho(t) = (1-t)\rho_{\text{collapse}} + t\,\rho_\star$. (c) Center-mass ratio at epoch 200.

$\mathcal{F}_3$, whereas the matched-operator benchmarks generate and invert with the same operator (Kaipio & Somersalo, 2007). The real-data check (Section 5.5) uses no simulator.

Figure 3 shows the pattern: free-density solvers follow the centered iter-0 gradient to a centrally collapsed minimum, while NeTMY preserves off-center structure and detail (examples in Appendix F.13).

5.3 Optimization Geometry: Why Free-Density Solvers Center-Collapse at an Early Stage

Lemma 2 predicts that the centered iter-0 gradient produced by center bias and peak coupling is executed directly by gradient-descent free-density solvers and, in the unpreconditioned first-order limit, filtered by the parameterization Jacobian for NeTMY. We measure three quantities on representative many-source samples (Figure 4).

Iter-0 gradient. On a uniform initialization, the $\mathcal{F}_2$ data-fidelity gradient has a center-to-outer-ring magnitude ratio of $18.29\times$, with the peak at the grid center and not at any source: the center-bias signature predicted by Eq. 23.

Energy barrier. The path from a centered-collapse iterate to the ground truth (Figure 4b) crosses an $h \approx 1.12$ barrier under $\mathcal{F}_2$ at $t = 0.20$ but is monotonically decreasing under $\mathcal{F}_1$: a free-density solver in the centered minimum is hard to escape by gradient descent.

Fixed budget. At a 200-iteration budget (Figure 4c), center-mass ratios order as NeTMY (0.00) < L-BFGS (0.081) < ADMM (0.153) < Tikhonov (0.223). L-BFGS, a quasi-Newton free-density solver, still falls into the centered-collapse basin. The realized NeTMY iter-0 update has no singular center spike (App. F.7, Fig. 10); a 32-run ADMM sweep confirms the operator effect (mean center-mass ratio $0.55 \to 0.78$ from $\mathcal{F}_1$ to $\mathcal{F}_2$; App. F.7). In an operator-spectral diagnostic of the linearized measurement map (Appendix F.7.1), the parameterization suppresses energy in the unresolved residual subspace by a factor of 2–5 relative to the raw gradient under the full objective; under the data-fidelity term alone, the raw gradient already lies almost entirely in the resolved band.

5.4 Component Ablation

We ablate seven design choices on the full dataset (Table 3). Cumulatively removing {annealing schedule, PE, multiscale, ℓ_1, gate, $\mathcal{R}_{\mathrm{ds}}$} degrades density MSE by 2–3×. The ordering tracks the ill-posedness mechanisms that motivated each component: annealed PE, PE, and multiscale address frequency suppression, the gate addresses peak coupling, and the direct-density loss provides the missing-amplitude proxy. TV has a small aggregate effect (×1.1) but qualitatively controls cross-shaped artifacts on dense scenes. Per-class and hyperparameter sweeps are in Appendix F.8.

5.5 Real-Data Cross-Check on α-$RuCl_3$

We test the operator-fidelity gap on the α-$RuCl_3$ dataset of Kumar et al. (2024) (8 NVs, SRIM depth prior 14 ± 5 nm) on two axes, depth-amplitude and Hessian conditioning; the full four-step protocol, including an operator-neutral spectral calibration and a power-law slope axis, is in Appendix F.10.

Depth-amplitude. Solving $\Gamma_{\mathrm{Ru}}^{\mathrm{pred}}(d) = \Gamma_{\mathrm{Ru}}^{\mathrm{meas}}$ for the implied NV depth $d^\star$ at fixed crystallographic density and moment, the $\mathcal{F}_1$ inversion has no real solution for 7 of the 8 NVs and places $0/8$ in $[9, 20]$ nm, because the point amplitude ratio is $\Gamma^{\mathcal{F}_1}/\Gamma^{\mathcal{F}_2} \approx 210\times$ at $d = 14.5$ nm; $\mathcal{F}_2$ places $1/8$ NVs in the window.

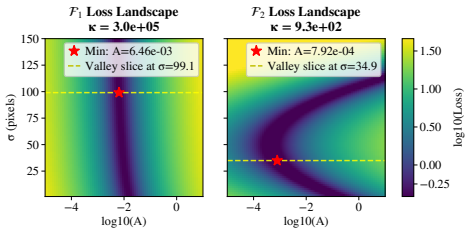


Figure 5: Loss landscape on α-$RuCl_3$: $\mathcal{F}_2$ (right) is well-conditioned, $\mathcal{F}_1$ (left) is a degenerate valley.

Hessian conditioning. On the Gaussian-density ansatz $\rho(\mathbf{r}; A, \sigma_g) = A \exp(-\|\mathbf{r}\|^2 / 2\sigma_g^2)$, where σ_g denotes the Gaussian width, over the 8 NVs (Figure 5), the log-MSE condition numbers are $\kappa_{\mathcal{F}_2} = 931$ and $\kappa_{\mathcal{F}_1} = 301{,}139$ (ratio $\sim 323\times$): the $\mathcal{F}_2$ landscape is a parabolic bowl while $\mathcal{F}_1$ is a degenerate valley along $A^2\sigma_g^2 \approx \mathrm{const}$, a direct algebraic consequence of $\Gamma \propto A^2$ in $\mathcal{F}_1$. Both axes disfavor $\mathcal{F}_1$. Appendix B.9 reproduces the dipolar $1/T_1 \propto \sigma_b/D^4$ scaling under the forward operator.

6 Conclusion

We framed NV-noise sensing inversion as an ill-posed inverse problem in which the choice of forward operator reshapes both the measurement distribution and the optimization geometry. Moving from a scalar/coherent approximation $\mathcal{F}_1$ to a tensor power-summed dipolar operator $\mathcal{F}_2$ exposes a center-collapse pathology in free-density solvers, traced to a centered iter-0 gradient and max-normalization peak coupling. NeTMY, an amortization-free coordinate neural-field solver whose parameterization filters the raw density-space gradient through a positive semidefinite kernel, attains the best localization and distributional metrics under $\mathcal{F}_2$ on the cross-fidelity benchmark, and the α-$RuCl_3$ check disfavors $\mathcal{F}_1$ on the depth-amplitude and conditioning axes.

Limitations and Broader Impact. Per-measurement optimization costs minutes per spectrum, which suits offline reconstruction when paired data are scarce; the matched-hardware cost accounting, including amortized training and compute crossovers, is in Appendix F.11. The real-data check is a multi-axis consistency test on 8 NVs under a Gaussian ansatz rather than a full validation. The analysis is scoped to the fluctuation-dominated dipolar regime; coherent-coupling regimes and other quantum-sensing modalities are left to future work. We do not anticipate direct dual-use harms; the main risk is misplaced confidence in reconstructions outside the operator's modeled regime, so downstream claims should report operator-fidelity caveats and consistency checks.

AI USE STATEMENT

We used generative AI tools to aid and polish the writing of this manuscript (language editing, restructuring of paragraphs, and LaTeX formatting). The authors reviewed and edited all AI-assisted text and take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work develops a reconstruction method evaluated on synthetic spectra and on a published NV relaxometry dataset. It involves no human subjects, personal data, or new data collection. We do not anticipate direct dual-use harm. The main risk is over-trusting reconstructions outside the regime the forward operator models, and we report operator-fidelity caveats and consistency checks alongside the results.

REPRODUCIBILITY STATEMENT

The forward operators and their regime of validity are derived in Appendix B, and the fidelity and scale-correction results are proved in Appendix C. Appendix E specifies the NeTMY architecture, training schedule, loss weights, and algorithm. Dataset construction, baseline settings, and metric definitions are given in Appendices F.1, F.2, and F.3, the real-data protocol in Appendix F.10, and runtime in Appendix F.11. The code for data generation, training, and evaluation is available in an anonymized repository (Appendix A).

REFERENCES

Nati Aharon, Amit Rotem, Liam P McGuinness, Fedor Jelezko, Alex Retzker, and Zohar Ringel. Nv center based nano-nmr enhanced by deep learning. *Scientific reports*, 9(1):17802, 2019.

Ali Ahmed, Benjamin Recht, and Justin Romberg. Blind deconvolution using convex programming. *IEEE Transactions on Information Theory*, 60(3):1711–1732, 2013.

Ismail Alkhouri, Evan Bell, Avrajit Ghosh, Shijun Liang, Rongrong Wang, and Saiprasad Ravishankar. Understanding untrained deep models for inverse problems: Algorithms and theory. *arXiv preprint arXiv:2502.18612*, 2025.

Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen. On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proceedings of the National Academy of Sciences*, 117(48):30088–30095, 2020.

Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta numerica*, 28:1–174, 2019.

Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, 2006.

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *International conference on machine learning*, pp. 537–546. PMLR, 2017.

David A Broadway, Scott E Lillie, Sam C Scholten, Dominik Rohner, Nikolai Dontschuk, Patrick Maletinsky, Jean-Philippe Tetienne, and Lloyd CL Hollenberg. Improved current density and magnetization reconstruction through vector magnetic field measurements. *Physical Review Applied*, 14(2):024076, 2020.

Carey Bunks, Fatimetou M Saleck, Stephane Zaleski, and Guy Chavent. Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473, 1995.

Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

Francesco Casola, Toeno Van Der Sar, and Amir Yacoby. Probing condensed matter physics with magnetometry based on nitrogen-vacancy centres in diamond. *Nature Reviews Materials*, 3(1): 17088, 2018.

Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, 40(1):120–145, 2011.

Kai-Hung Cheng, Zeeshawn Kazi, Jared Rovny, Bichen Zhang, Lila S Nassar, Jeff D Thompson, and Nathalie P De Leon. Massively multiplexed nanoscale magnetometry with diamond quantum sensors. *Physical Review X*, 15(3):031014, 2025.

Yuejie Chi, Yue M Lu, and Yuxin Chen. Nonconvex optimization meets low-rank matrix factorization: An overview. *IEEE Transactions on Signal Processing*, 67(20):5239–5269, 2019.

Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics-informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3):88, 2022.

Christian L Degen, Friedemann Reinhard, and Paola Cappellaro. Quantum sensing. *Reviews of modern physics*, 89(3):035002, 2017.

Mostafa Dehghani, Yi Tay, Alexey A Gritsenko, Zhe Zhao, Neil Houlsby, Fernando Diaz, Donald Metzler, and Oriol Vinyals. The benchmark lottery. *arXiv preprint arXiv:2107.07002*, 2021.

Marcus W Doherty, Neil B Manson, Paul Delaney, Fedor Jelezko, Jörg Wrachtrup, and Lloyd CL Hollenberg. The nitrogen-vacancy colour centre in diamond. *Physics Reports*, 528(1):1–45, 2013.

David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

James R Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, 1982.

Patrick Fuchs and Karin Shmueli. Incomplete spectrum qsm using support information. *Frontiers in Neuroscience*, 17:1130524, 2023.

Davis Gilton, Gregory Ongie, and Rebecca Willett. Deep equilibrium architectures for inverse problems in imaging. *IEEE Transactions on Computational Imaging*, 7:1123–1133, 2021.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Roberta Grech, Tracey Cassar, Joseph Muscat, Kenneth P Camilleri, Simon G Fabri, Michalis Zervakis, Petros Xanthopoulos, Vangelis Sakkalis, and Bart Vanrumste. Review on solving the inverse problem in eeg source analysis. *Journal of neuroengineering and rehabilitation*, 5(1):25, 2008.

Peimeng Guan, Naveed Iqbal, Mark A Davenport, and Mudassir Masood. Solving inverse problems with model mismatch using untrained neural networks within model-based architectures. *arXiv preprint arXiv:2403.04847*, 2024.

Harshit Gupta, Kyong Hwan Jin, Ha Q Nguyen, Michael T McCann, and Michael Unser. Cnn-based projected gradient descent for consistent ct image reconstruction. *IEEE transactions on medical imaging*, 37(6):1440–1453, 2018.

Andreas Habring and Martin Holler. Neural-network-based regularization methods for inverse problems in imaging. *GAMM-Mitteilungen*, 47(4):e202470004, 2024.

Matti Hämäläinen, Riitta Hari, Risto J Ilmoniemi, Jukka Knuutila, and Olli V Lounasmaa. Magnetoencephalography: theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497, 1993.

Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine*, 79(6):3055–3071, 2018.

Reinhard Heckel and Paul Hand. Deep decoder: Concise image representations from untrained non-convolutional networks. *arXiv preprint arXiv:1810.03982*, 2018.

Reinhard Heckel and Mahdi Soltanolkotabi. Compressive sensing with un-trained neural networks: Gradient descent finds a smooth approximation. In *International conference on machine learning*, pp. 4149–4158. PMLR, 2020.

Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, Doina Precup, and David Meger. Deep reinforcement learning that matters. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Arthur E Hoerl and Robert W Kennard. Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12(1):69–82, 1970a.

Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970b.

Jan A Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement Series*, 15:417–426, 1974.

Jonas Homrighausen, Ludwig Horsthemke, Jens Pogorzelski, Sarah Trinschek, Peter Glösekötter, and Markus Gregor. Edge-machine-learning-assisted robust magnetometer based on randomly oriented nv-ensembles in diamond. *Sensors*, 23(3):1119, 2023.

Adan Jafet Garcia Inda, Shao Ying Huang, Nevrez İmamoğlu, Ruian Qin, Tianyi Yang, Tiao Chen, Zilong Yuan, and Wenwei Yu. Physics informed neural networks (pinn) for low snr magnetic resonance electrical properties tomography (mrept). *Diagnostics*, 12(11):2627, 2022.

Gauri Jagatap and Chinmay Hegde. Algorithmic guarantees for inverse imaging with untrained network priors. *Advances in neural information processing systems*, 32, 2019.

Jari Kaipio and Erkki Somersalo. Statistical inverse problems: discretization, model reduction and inverse crimes. *Journal of computational and applied mathematics*, 198(2):493–504, 2007.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

Jitender Kumar, Dan Yudilevich, Ariel Smooha, Inbar Zohar, Arnab K Pariari, Rainer Stöhr, Andrej Denisenko, Markus Hücker, and Amit Finkler. Room temperature relaxometry of single nitrogen vacancy centers in proximity to α-rucl3 nanoflakes. *Nano Letters*, 24(16):4793–4800, 2024.

Kiryung Lee, Yanjun Li, Marius Junge, and Yoram Bresler. Blind recovery of sparse signals from subsampled convolution. *IEEE Transactions on Information Theory*, 63(2):802–821, 2016.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Thomas Liao, Rohan Taori, Inioluwa Deborah Raji, and Ludwig Schmidt. Are we learning yet? a meta review of evaluation failures across machine learning. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.

Shuyang Ling and Thomas Strohmer. Self-calibration and biconvex compressive sensing. *Inverse Problems*, 31(11):115002, 2015.

Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989.

Keqing Liu, Jiazhao Tian, Bokun Duan, Hao Zhang, Kangze Li, Guofeng Zhang, Fedor Jelezko, Ressa S Said, Jianming Cai, and Liantuan Xiao. High-resolution wide-field magnetic imaging with sparse sampling using nitrogen-vacancy centers. *arXiv preprint arXiv:2602.00679*, 2026.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

HJ Mamin, M Kim, MH Sherwood, Charles T Rettner, K Ohno, DD Awschalom, and D Rugar. Nanoscale nuclear magnetic resonance with a nitrogen-vacancy spin sensor. *Science*, 339(6119): 557–560, 2013.

Julien NP Martel, David B Lindell, Connor Z Lin, Eric R Chan, Marco Monteiro, and Gordon Wetzstein. Acorn: Adaptive coordinate networks for neural scene representation. *arXiv preprint arXiv:2105.02788*, 2021.

Gary Mataev, Peyman Milanfar, and Michael Elad. Deepred: Deep image prior powered by red. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pp. 0–0, 2019.

Michael T McCann, Kyong Hwan Jin, and Michael Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95, 2017.

Siddhant Midha, Madhur Parashar, Anuj Bathla, David A Broadway, Jean-Philippe Tetienne, and Kasturi Saha. Optimized current-density reconstruction from wide-field quantum diamond magnetic field maps. *Physical Review Applied*, 22(1):014015, 2024.

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. *Communications of the ACM*, 65(1):99–106, 2021.

Amirali Molaei, Amirhossein Aminimehr, Armin Tavakoli, Amirhossein Kazerouni, Bobby Azad, Reza Azad, and Dorit Merhof. Implicit neural representation in medical imaging: A comparative survey. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2381–2391, 2023.

Vishal Monga, Yuelong Li, and Yonina C Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021.

Aldona Mzyk, Alina Sigaeva, and Romana Schirhagl. Relaxometry with nitrogen vacancy (nv) centers in diamond. *Accounts of chemical research*, 55(24):3572–3580, 2022.

Parikh Neal, Chu Eric, Peleato Borja, and Eckstein Jonathan. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Gregory Ongie, Ajil Jalal, Christopher A Metzler, Richard G Baraniuk, Alexandros G Dimakis, and Rebecca Willett. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020.

Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5865–5874, 2021.

Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Inioluwa Deborah Raji, Emily M Bender, Amandalynne Paullada, Emily Denton, and Alex Hanna. Ai and the everything in the whole wide world benchmark. *arXiv preprint arXiv:2111.15366*, 2021.

Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. Do imagenet classifiers generalize to imagenet? In *International conference on machine learning*, pp. 5389–5400. PMLR, 2019.

Albert W Reed, Hyojin Kim, Rushil Anirudh, K Aditya Mohan, Kyle Champley, Jingu Kang, and Suren Jayasuriya. Dynamic ct reconstruction from limited views with implicit neural representations and parametric motion fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2258–2268, 2021.

Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM journal on imaging sciences*, 10(4):1804–1844, 2017.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

Jared Rovny, Sarang Gopalakrishnan, Ania C Bleszynski Jayich, Patrick Maletinsky, Eugene Demler, and Nathalie P de Leon. Nanoscale diamond quantum sensors for many-body physics. *Nature Reviews Physics*, 6(12):753–768, 2024.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

Michael J Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3(10):793–796, 2006.

Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, 2015.

Liyue Shen, John Pauly, and Lei Xing. Nerp: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE transactions on neural networks and learning systems*, 35(1):770–782, 2022.

Kexuan Shi, Xingyu Zhou, and Shuhang Gu. Improved implicit neural representation with fourier reparameterized training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25985–25994, 2024.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

Ariel Smooha, Jitender Kumar, Dan Yudilevich, John W Rosenberg, Valentin Bayer, Rainer Stöhr, Andrej Denisenko, Tatyana Bendikov, Anna Kossoy, Iddo Pinkas, et al. Sensing single-molecule magnets with nitrogen-vacancy centers. *Nano Letters*, 2026.

Artur Speiser, Lucas-Raphael Müller, Philipp Hoess, Ulf Matti, Christopher J Obara, Wesley R Legant, Anna Kreshuk, Jakob H Macke, Jonas Ries, and Srinivas C Turaga. Deep learning enables fast and dense single-molecule localization with high accuracy. *Nature Methods*, 18(9):1082–1090, 2021.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.

J-P Tetienne, Thomas Hingant, Loïc Rondin, Adrien Cavaillès, Ludovic Mayer, Géraldine Dantelle, Thierry Gacoin, Jörg Wrachtrup, J-F Roch, and Vincent Jacques. Spin relaxometry of single nitrogen-vacancy defects in diamond nanocrystals for magnetic noise sensing. *arXiv preprint arXiv:1304.1197*, 2013.

Moeta Tsukamoto, Shuji Ito, Kensuke Ogawa, Yuto Ashida, Kento Sasaki, and Kensuke Kobayashi. Accurate magnetic field imaging using nanodiamond quantum sensors enhanced by machine learning. *Scientific reports*, 12(1):13942, 2022.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9446–9454, 2018.

Toeno Van der Sar, Francesco Casola, Ronald Walsworth, and Amir Yacoby. Nanometre-scale probing of spin waves using single electron spins. *Nature communications*, 6(1):7886, 2015.

Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE global conference on signal and information processing*, pp. 945–948. IEEE, 2013.

Jean Virieux and Stéphane Operto. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26, 2009.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Yves Wiaux, Laurent Jacques, Gilles Puy, Anna MM Scaife, and Pierre Vandergheynst. Compressed sensing imaging techniques for radio interferometry. *Monthly Notices of the Royal Astronomical Society*, 395(3):1733–1742, 2009.

Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE transactions on image processing*, 23(2): 684–695, 2013.

Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Admm-csnet: A deep learning approach for image compressive sensing. *IEEE transactions on pattern analysis and machine intelligence*, 42(3): 521–538, 2018.

Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3929–3938, 2017.

Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2022.

Yuqian Zhang, Han-wen Kuo, and John Wright. Structured local minima in sparse blind deconvolution. *Advances in neural information processing systems*, 31, 2018.

Brandon Zhao, Aviad Levis, Liam Connor, Pratul P Srinivasan, and Katherine L Bouman. Single view refractive index tomography with neural fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25358–25367, 2024.

Brandon Zhao, Aviad Levis, Liam Connor, Pratul P Srinivasan, and Katherine L Bouman. Revealing the 3d cosmic web through gravitationally constrained neural fields. *arXiv preprint arXiv:2504.15262*, 2025.

Tao Zhong, Yixun Hu, Dongzhe Zheng, Aditya Sood, and Christine Allen-Blanchette. Neural field thermal tomography: A differentiable physics framework for non-destructive evaluation. *arXiv preprint arXiv:2603.11045*, 2026.

Lei Zhu, Wei Zhang, Daniel Elnatan, and Bo Huang. Faster STORM using compressed sensing. *Nature Methods*, 9(7):721–723, 2012.

José Luis Ávila Jiménez, Ann-Sophie Bülter, Ludwig Horsthemke, Francisco J. Rodriguez-Lozano, Manuel Ortiz-Lopez, and Peter Glösekötter. Enhancing all-optical nv center magnetometry with machine learning: Model selection for efficient and deployable quantum sensors. *IEEE Sensors Journal*, 25(24):44473–44481, 2025. doi: 10.1109/JSEN.2025.3628661.

A Code Availability

To facilitate the review process, we provide the full implementation of our method in an anonymized repository: `https://anonymous.4open.science/r/NeTMY-18974/`. Instructions for setting up and reproducing the proposed method are provided in the repository `README`.

B Forward Operator Derivation

This appendix expands the measurement model of Section 3.1 from first principles, makes the difference between the scalar $\mathcal{F}_1$ and tensor $\mathcal{F}_2$ operators explicit at the dipole level, and identifies the regimes in which each form is physically appropriate.

B.1 Sample model and dipolar Green tensor

We model the sample as a (possibly continuous) distribution of independent magnetic fluctuators with nonnegative density $\rho : \Omega \to \mathbb{R}_{\geq 0}$ over a 2D plane $\Omega \subset \mathbb{R}^2$. A widefield NV array sits at standoff height z_0, so the relative position from a sample point $\mathbf{r}_{\mathrm{src}} \in \Omega \times \{0\}$ to an NV pixel $\mathbf{r} \in \Omega \times \{z_0\}$ is $\mathbf{R}(\mathbf{r}, \mathbf{r}_{\mathrm{src}}) = (\mathbf{r} - \mathbf{r}_{\mathrm{src}}, z_0)$. The dipolar Green tensor

$$G_{ia}(\mathbf{R}) \;=\; \frac{\mu_0}{4\pi} \frac{3R_i R_a - \|\mathbf{R}\|^2 \delta_{ia}}{\|\mathbf{R}\|^5}, \qquad i, a \in \{x, y, z\} \tag{8}$$

maps a unit dipole moment along axis a at $\mathbf{r}_{\mathrm{src}}$ to a magnetic-field component along axis i at $\mathbf{r}$. For an NV with quantization axis $\mathbf{n}$, the projected response is $G_{\mathrm{nv},a} = \sum_i n_i G_{ia}$; for a z-aligned NV (the convention used throughout this paper) we have $G_{\mathrm{nv},a} = G_{za} = G_{az}$ by symmetry of G in its two indices. Thermal magnetic fluctuations of a dipole at $\mathbf{r}_{\mathrm{src}}$ along axis a produce a spectral density $\rho(\mathbf{r}_{\mathrm{src}})\, L(\omega; \omega_L(\mathbf{r}_{\mathrm{src}}))$ at the source, with L the Lorentzian of Section 3.1 (Tetienne et al., 2013; Doherty et al., 2013; Casola et al., 2018).

We retain three structural assumptions that are standard in the NV-relaxometry literature: (i) sample dipoles are independent thermal fluctuators (no inter-source coherence), (ii) the NV array reads out the noise spectrum at the readout pixel $\mathbf{r}$ rather than at the source pixel $\mathbf{r}_{\mathrm{src}}$ (a controlled approximation when ω_L varies slowly over the kernel support), and (iii) the source-spin fluctuation components $a \in \{x, y, z\}$ are independent thermal channels with equal variance, i.e., isotropic in the lab frame, so each channel contributes $|G_{az}|^2$ to the NV-axis noise field at pixel $\mathbf{r}$. The physical regime in which these assumptions hold is discussed in Appendix B.6.

B.2 Pointwise derivation of $\mathcal{F}_2$ (incoherent thermal sum)

For independent thermal fluctuators, discretize the source plane into cells of area Δ centered at $\{\mathbf{r}_i\}$ and write $\rho_i := \rho(\mathbf{r}_i)\, \Delta$ for the dimensionless source weight in cell i. The noise spectrum measured at NV position $\mathbf{r}$ is the sum of squared field amplitudes from all sources, summed over independent fluctuation channels (Tetienne et al., 2013; Van der Sar et al., 2015):

$$\begin{aligned} \mathcal{F}_2(\rho, \omega_L)(\omega, \mathbf{r}) \;&=\; \sum_{a \in \{x,y,z\}} \sum_i |G_{az}(\mathbf{R}_i)|^2 \; \rho_i \, L\big(\omega; \omega_L(\mathbf{r}_i)\big) \\ &\approx \; \Big[\sum_{a \in \{x,y,z\}} \sum_i |G_{az}(\mathbf{R}_i)|^2 \; \rho_i \Big] \, L\big(\omega; \omega_L(\mathbf{r})\big), \end{aligned} \tag{9}$$

where $\mathbf{R}_i = \mathbf{R}(\mathbf{r}, \mathbf{r}_i)$. The first line is the source-side exact form ($\mathcal{F}_3$ in our internal notation, used as the simulator $\mathcal{F}_3$ for ground-truth data generation in Section 5); the second line factors the Lorentzian out of the spatial sum, which is exact when ω_L is locally constant over the kernel support and equivalent (up to grid-area factors) to $[\sum_a (|G_{az}|^2 * \rho)(\mathbf{r})]\, L(\omega; \omega_L(\mathbf{r}))$ used by the FFT-based reconstruction stack. The contraction over a is incoherent: distinct channels and distinct sources contribute independently, with no kernel-product cross-terms. Density enters Eq. 9 *linearly* in ρ_i, which is the property exploited by the scale-correction proof of Appendix C.

B.3 Pointwise derivation of $\mathcal{F}_1$ (scalar coherent square)

The simplified operator $\mathcal{F}_1$ retains only a single NV-axis projection and squares the convolved field as a final, post-superposition operation. Using the same discrete cell weights $\rho_i = \rho(\mathbf{r}_i)\Delta$, expanding

the modulus square gives

$$\begin{aligned}\mathcal{F}_1(\rho,\omega_L)(\omega,\mathbf{r}) &= \left|\sum_i G_{\mathrm{nv}}(\mathbf{R}(\mathbf{r},\mathbf{r}_i))\,\rho_i\right|^2 L(\omega;\omega_L(\mathbf{r})) \\ &= \Bigg[\underbrace{\sum_i \rho_i^2\, G_{\mathrm{nv}}(\mathbf{R}_i)^2}_{\text{diagonal}} + \underbrace{\sum_{i\neq j}\rho_i\rho_j\, G_{\mathrm{nv}}(\mathbf{R}_i)\, G_{\mathrm{nv}}(\mathbf{R}_j)}_{\text{cross term } \mathcal{C}(\mathbf{r})}\Bigg] L(\omega;\omega_L(\mathbf{r})), \end{aligned} \tag{10}$$

where $\mathbf{R}_i = \mathbf{R}(\mathbf{r},\mathbf{r}_i)$. The diagonal term reintroduces the squared cell weight ρ_i^2 instead of the linear cell weight ρ_i, so $\mathcal{F}_1(c\rho,\omega_L) = c^2\,\mathcal{F}_1(\rho,\omega_L)$; the cross-term $\mathcal{C}(\mathbf{r})$ couples pairs of distinct sources through the product of their NV-axis kernels, which has no counterpart in Eq. 9. When ρ is sparse and well-separated relative to the support of G_{nv}, $\mathcal{C}$ is small pointwise and $\mathcal{F}_1$ behaves as a (nonlinear-in-ρ) close cousin of $\mathcal{F}_2$; when sources are close, $\mathcal{C}$ dominates and the two operators differ in both magnitude and spatial structure.

B.4 Direct simulator $\mathcal{F}_3$ and the operator-fidelity check

The two operators $\mathcal{F}_1$ and $\mathcal{F}_2$ are FFT-factorized: each evaluates a 2D convolution between a per-channel kernel and the source density, then multiplies by a Lorentzian factor at the readout pixel. The factorization is exact for $\mathcal{F}_2$ when the Larmor field ω_L is locally constant over the kernel support (Eq. 9), and is only an approximation when ω_L varies across the support. The direct simulator $\mathcal{F}_3$ removes this approximation by evaluating the source-side superposition pixel-by-pixel without FFT factorization,

$$\begin{aligned}\mathcal{F}_3(\rho,\omega_L)(\omega,\mathbf{r}) &= \sum_{\mathbf{r}_{\mathrm{src}}\in\Omega} \rho(\mathbf{r}_{\mathrm{src}})\, P\big(\mathbf{R}(\mathbf{r},\mathbf{r}_{\mathrm{src}})\big)\, L\big(\omega;\omega_L(\mathbf{r}_{\mathrm{src}})\big), \\ P(\mathbf{R}) &= \sum_{a\in\{x,y,z\}} |G_{az}(\mathbf{R})|^2, \end{aligned} \tag{11}$$

where the Lorentzian is evaluated at each *source* pixel $\mathbf{r}_{\mathrm{src}}$ rather than factored out at the readout pixel, restoring the per-source spectral response of the underlying physics. $\mathcal{F}_3$ is the source-side form (the first line of Eq. 9) without the locally-constant-ω_L factorization, and is implemented in float64 with a direct source-loop evaluator.

Where $\mathcal{F}_3$ is used. The primary main-text benchmark in this paper is the cross-fidelity benchmark on a 512-sample dataset (Table 1, Section 5.2, Appendix F.1): spectra are produced by the source-side direct simulator $\mathcal{F}_3$ and inverted by $\mathcal{F}_1$ or $\mathcal{F}_2$, so the inversion operator is never reused as the data-generation operator and the inverse-crime risk (Kaipio & Somersalo, 2007) is avoided by construction. The matched-operator benchmarks ($\mathcal{F}_1/\mathcal{F}_1$ on 512 samples and $\mathcal{F}_2/\mathcal{F}_2$ on 512 samples; Table 2, Appendix F.4) are reported alongside as a complementary view that includes amortized and untrained-prior baselines for which the $\mathcal{F}_3$ form is not implemented; they are intentionally inverse-crime by construction and we do not use them as the central evidence for the operator-fidelity claim.

Exact convolutional form and learned surrogates. On the uniform grid used in this paper, Eq. 11 has an exactly equivalent convolutional form: defining $q_f(\mathbf{r}) = \rho(\mathbf{r})\, L(\omega_f;\omega_L(\mathbf{r}))$ at each frequency ω_f gives $S_f = P * q_f$, which a zero-padded batched FFT evaluates while retaining the source-side Lorentzian. $\mathcal{F}_3$ is used only for offline benchmark generation, while NeTMY differentiates through the FFT-based $\mathcal{F}_2$. On one NVIDIA L40 (Table 14), median generation time is 0.013664 s for $\mathcal{F}_3$ and 0.000542 s for $\mathcal{F}_2$, against 141.342 s for a NeTMY solve, so forward evaluation is not the bottleneck. For nonuniform or 3D geometries, neural-operator surrogates such as the Fourier neural operator (Li et al., 2021) and DeepONet (Lu et al., 2021) are plausible replacements for the direct evaluator. Such a surrogate should be validated on forward error and on downstream reconstruction against $\mathcal{F}_3$: Section 5 shows that substituting one forward operator for another changes the inverse landscape, so surrogate error is not innocuous. We leave surrogates for nonuniform and 3D geometries to future work.

B.5 Operator difference

Subtracting the discrete $\mathcal{F}_2$ approximation in Eq. 9 from Eq. 10 at fixed ω and $\mathbf{r}$,

$$\big(\mathcal{F}_1-\mathcal{F}_2\big)(\rho,\omega_L)(\omega,\mathbf{r}) \;=\; \Big[\mathcal{C}(\mathbf{r}) + \sum_i \big(\rho_i^2-\rho_i\big) G_{\mathrm{nv}}(\mathbf{R}_i)^2 - \sum_{a\neq z}\sum_i \rho_i \,\big|G_{az}(\mathbf{R}_i)\big|^2\Big]\, L(\omega;\omega_L(\mathbf{r})), \tag{12}$$

i.e., the gap is the cross-term $\mathcal{C}$, plus the signed diagonal discrepancy $\rho_i^2 - \rho_i$, minus the transverse-channel power $a \in \{x, y\}$ that is present in $\mathcal{F}_2$ and omitted by $\mathcal{F}_1$. These contributions are spatially structured across $\mathbf{r}$ and generally nonzero in multi-source scenes. Eq. 12 is the algebraic origin of the inverse-landscape and ranking changes documented in Section 5.

B.6 Physical regime of applicability

Both forms are legitimate approximations in different physical regimes. $\mathcal{F}_1$ is appropriate for coherent-emitter superposition (e.g., a small number of phase-locked classical dipoles driven by the same coherent source), where the field amplitudes literally add before being squared by the detector. $\mathcal{F}_2$ is appropriate for incoherent thermal ensembles, where distinct fluctuators have independent random phases and only second moments of the field add (Tetienne et al., 2013; Van der Sar et al., 2015; Casola et al., 2018). NV relaxometry of nuclear, electronic, or magnonic spin baths is in the second regime (Mzyk et al., 2022; Kumar et al., 2024; Midha et al., 2024), so $\mathcal{F}_2$ is the physically appropriate choice. We retain $\mathcal{F}_1$ as a benchmark because it is computationally cheaper, has historically appeared in some open-source simulators (Midha et al., 2024), and produces a meaningfully different inverse landscape that exposes the operator-fidelity sensitivity exploited in Section 5.

B.7 Lorentzian and Larmor mapping

The spectral response in Eq. 9 is the Lorentzian

$$L(\omega;\omega_L) \;=\; \frac{\gamma^2}{(\omega-\omega_L)^2+\gamma^2}, \tag{13}$$

with linewidth γ fixed by the experimental setup (we use $\gamma = 0.5\,\mathrm{GHz}$ throughout, matching the simulator). The Larmor map $\omega_L : \Omega \to \mathbb{R}_+$ encodes the per-pixel quasi-static detuning of the local resonance and is treated as a second unknown field in the inverse problem. In the FFT-based factorization used by all reconstruction methods, $L(\omega;\omega_L(\mathbf{r}))$ is evaluated at the readout pixel $\mathbf{r}$ rather than at the source pixel $\mathbf{r}_{\mathrm{src}}$, which is exact when ω_L is approximately constant over the kernel support and a controlled approximation otherwise. The direct simulator $\mathcal{F}_3$ used to generate ground-truth data evaluates L at $\mathbf{r}_{\mathrm{src}}$ and is therefore closer to the underlying physics; in the main-text cross-fidelity benchmark we use $\mathcal{F}_3$ for data generation and either $\mathcal{F}_1$ or $\mathcal{F}_2$ for reconstruction (Appendix B.4).

B.8 Robustness to spatially varying Larmor frequencies

We further validated the robustness of the FFT-factorized solver F_2 against spatial variation of the Larmor frequency within one dipolar kernel footprint. In the real-data setting, the Larmor frequency is primarily set by the externally applied magnetic field and is therefore nearly uniform across the reconstruction field of view, with only weak local perturbations. The relevant control parameter is the dimensionless ratio $\chi \sim \Delta\omega_L^{(\mathrm{kernel})}/\gamma$, where $\Delta\omega_L^{(\mathrm{kernel})}$ denotes the in-kernel variation of ω_L and γ is the effective linewidth. In a controlled stress test, the relative forward error of F_2 with respect to the direct source-side operator F_3 remained small when $\chi \ll 1$ (about 2.6×10^{-2} at $\chi \approx 0.3$), but increased substantially as χ approached unity (about 1.6×10^{-1} at $\chi \approx 1.08$). The error remains small in the broad-linewidth, slowly varying regime tested here.

B.9 Further physical variation of the forward operator

Figure 6 shows the spectrum noise generated by different components of the green kernel. The shape and scale of the spectrum noise stay consistent with the physical formula

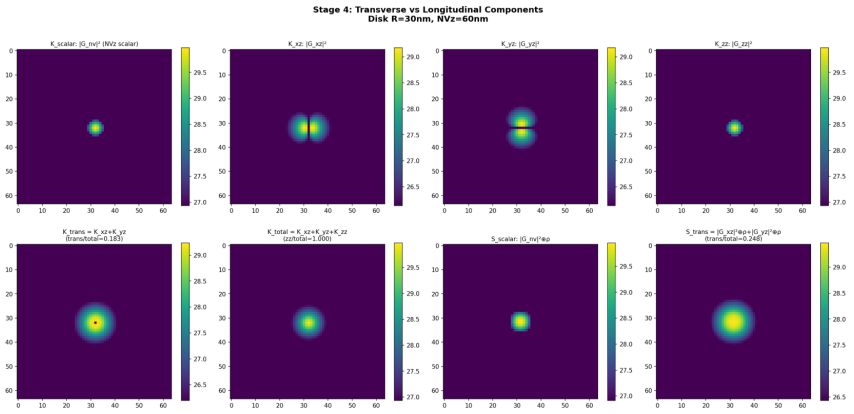


Figure 6: Verification of the Green kernel in physical solver

According to (Tetienne et al., 2013), the experiment detects the longitudinal relaxation time T_1 of the dipole spin bath (with bath surface density σ_b) on the nanodiamond with the NV center.

The cited work gives the formula

$$\frac{1}{T_1} = \frac{1}{T_1^{\text{bulk}}} + \left(\frac{48\mu_0^2\gamma_e^4\hbar^2 C_S}{\pi D^4}\right)\left(\frac{\sigma_b R(\sigma_b)}{\omega_0^2 + R(\sigma_b)^2}\right) \tag{14}$$

and

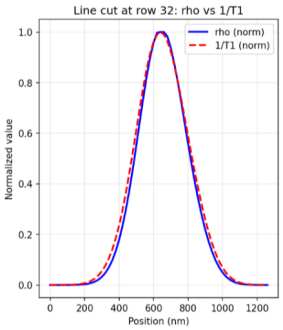
$$B_\perp^2 = \left(\frac{4\mu_0\gamma_e\hbar}{\pi}\right)^2 \pi C_S \frac{\sigma_b}{D^4} \tag{15}$$

where $C_S = \frac{1}{2S+1}\sum_{m=-S}^{S} m^2 = \frac{S(S+1)}{3}$ and $R(\sigma_b) = \frac{1}{\tau_c}$. The following relationship has been clearly stated in the research paper

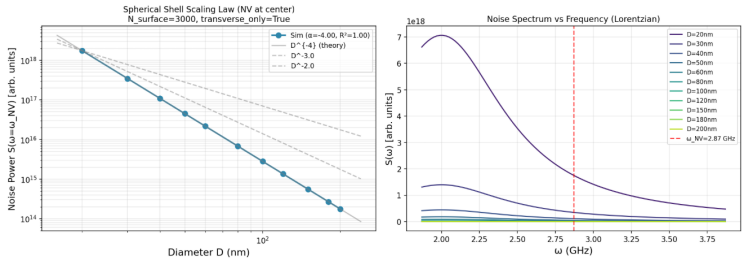
$$B_\perp^2 \propto \frac{\sigma_b}{D^4} \tag{16}$$

$$\frac{1}{T_1} \propto \frac{\sigma_b}{D^4} \tag{17}$$

To verify the physical correctness of the forward operator in the research, the experiment is reproduced. As shown in Figure 7, the noise power is inversely proportional to the fourth power of the diameter of the diamond D.



(a) Normalized line cut at row 32 comparing ρ and $1/T_1$ spatial profiles, demonstrating the qualitative agreement between the two distributions.



(b) Forward operator reproduction of the Tetienne *et al.* dipole bath model Tetienne et al. (2013). **Left:** Simulated noise power (dots) follows the theoretical D^{-4} scaling, and the Lorentzian spectrum $S(\omega)$ shows stronger noise for smaller D near $\omega_{\text{NV}} = 2.87\,\text{GHz}$. **Right:** The reconstructed $1/T_1(x, y)$ map matches the input $\rho(x, y)$ distribution (Pearson $r = 99.55\%$), confirming $1/T_1(x) \approx (|G|^2 * \rho)(x)$

Figure 7: The reproduction of the relationship between magnetic noise and $\frac{1}{T_1}$ using the forward operator and the relationship between magnetic noise and diameter using the forward operator

From Eq. 16 and Eq. 17, we can find out that the magnetic noise has the same order of magnitude, which means

$$\frac{1}{T_1(x)} \approx S = \left(|G|^2 * \rho\right)(x) \tag{18}$$

Thus for a stated dipole bath, the distribution of ρ and $\frac{1}{T_1(x)}$ should be similar, which is also followed by the forward operator.

C Data Fidelity, Scale Ambiguity, and Energy-Ratio Correction

This appendix gives the full form of the data-fidelity term $\mathcal{D}$ used in Eq. 3, records the formal proof that the pre-normalization scale of ρ is unidentifiable from $\mathcal{D}$ alone, and shows how an energy-ratio anchor restores identifiability. The proof is operator-agnostic and applies to any reconstruction method that minimizes Eq. 3; the specific scale-correction *post-processing step* adopted by NeTMY is described in Section 4 and Appendix E.5.

C.1 Pixelwise log-MSE form of $\mathcal{D}$

With $N(\mathbf{r}; S) = \sum_{\omega \in W} S(\omega, \mathbf{r})$, $\widehat{N}(\mathbf{r}; S) = N(\mathbf{r}; S) / \max_{\mathbf{r}'} N(\mathbf{r}'; S)$, and $\ell(\mathbf{r}; S) = \log_{10}(\widehat{N}(\mathbf{r}; S) + 10^{-10})$, the data-fidelity discrepancy in Eq. 3 is

$$\mathcal{D}\big(\mathcal{F}(\rho, \omega_L), S_{\text{obs}}\big) \;=\; \frac{1}{|\Omega|} \sum_{\mathbf{r} \in \Omega} \Big(\ell\big(\mathbf{r}; \mathcal{F}(\rho, \omega_L)\big) \;-\; \ell\big(\mathbf{r}; S_{\text{obs}}\big) \Big)^2, \tag{19}$$

where $|\Omega|$ denotes the number of grid pixels in Ω. The two preprocessing steps (frequency summation followed by spatial max-normalization) are standard in NV relaxometry to suppress per-sample amplitude variation and emphasize spatial-distribution mismatch (Kumar et al., 2024; Midha et al., 2024; Casola et al., 2018); the small additive constant 10^{-10} avoids numerical singularity on background pixels.

C.2 Setup: max-normalization induces a scale-flat ray

Let S_{obs} be the noise spectrum produced by an unknown ground-truth pair $(\rho_\star, \omega_{L,\star})$ under the linear operator $\mathcal{F}_2$, and let $\widehat{N}(\,\cdot\,; S)$ denote its frequency-summed, max-normalized noise map (Section 3.2). Linearity gives $\mathcal{F}_2(c\rho, \omega_L) = c\,\mathcal{F}_2(\rho, \omega_L)$ for any $c > 0$, so

$$\widehat{N}(\,\cdot\,; \mathcal{F}_2(c\rho, \omega_L)) \;=\; \widehat{N}(\,\cdot\,; \mathcal{F}_2(\rho, \omega_L)) \qquad \forall\, c > 0. \tag{20}$$

The fidelity term $\mathcal{D}$ in Eq. 19 is therefore exactly invariant along the ray $\{c\rho : c > 0\}$, and the data alone cannot identify the absolute density scale. This is a finite-dimensional instance of the bilinear-identifiability symptom analyzed in compressed-sensing and blind-deconvolution literature (Ahmed et al., 2013; Ling & Strohmer, 2015; Lee et al., 2016).

C.3 Energy-ratio correction is exact under $\mathcal{F}_2$ in the noiseless limit

Proposition 1 (Scale correction under $\mathcal{F}_2$)**.** *Let $\mathcal{F}$ be linear in ρ and assume the noiseless observation $S_{\text{obs}} = \mathcal{F}(\rho_\star, \omega_{L,\star})$. Let $\widehat{\rho}$ be a candidate reconstruction with the same shape as $\rho_\star$ up to scale, $\widehat{\rho} = c^{-1}\rho_\star$ for some unknown $c > 0$. Define*

$$E_{\text{obs}} \;=\; \sum_{\omega, \mathbf{r}} S_{\text{obs}}(\omega, \mathbf{r}), \qquad E_{\text{pred}} \;=\; \sum_{\omega, \mathbf{r}} \mathcal{F}(\widehat{\rho}, \omega_{L,\star})(\omega, \mathbf{r}),$$

and $\alpha = E_{\text{obs}}/E_{\text{pred}}$. Then $\alpha\widehat{\rho} = \rho_\star$ exactly, and α depends only on the observation and the reconstruction.

Proof. Linearity and $\widehat{\rho} = c^{-1}\rho_\star$ give $\mathcal{F}(\widehat{\rho}, \omega_{L,\star}) = c^{-1}\mathcal{F}(\rho_\star, \omega_{L,\star})$ pointwise. Hence $E_{\text{pred}} = c^{-1}E_{\text{obs}}$, so $\alpha = E_{\text{obs}}/E_{\text{pred}} = c$ and $\alpha\widehat{\rho} = \rho_\star$. □

With additive sensor noise $S_{\text{obs}} = \mathcal{F}(\rho_\star, \omega_{L,\star}) + \varepsilon$, write $E_\varepsilon = \sum_{\omega, \mathbf{r}} \varepsilon(\omega, \mathbf{r})$ and $E_\star = \sum_{\omega, \mathbf{r}} \mathcal{F}(\rho_\star, \omega_{L,\star})$. Then $\alpha = c\,(1 + E_\varepsilon/E_\star)$, so no deterministic exactness is claimed; if $\mathbb{E}\varepsilon = 0$,

the correction is unbiased in expectation and its relative fluctuation is $O(\sigma_\varepsilon \sqrt{|W|\,|\Omega|}/E_\star)$. A scale-correction regularizer that penalizes deviation of α from 1 (or, equivalently, anchors ρ to the observed energy after every step) restores identifiability without any oracle quantity. Two further caveats are worth recording independently of method. (i) The proof requires linearity of $\mathcal{F}$ in ρ, which holds for $\mathcal{F}_2$ but not $\mathcal{F}_1$ (handled below). (ii) The shape assumption $\widehat{\rho} = c^{-1}\rho_\star$ is exact only when the reconstruction recovers the true support; in practice we observe graceful degradation and a residual scale bias proportional to the shape error.

C.4 Square-root correction under $\mathcal{F}_1$

For the quadratic operator $\mathcal{F}_1(c\rho, \omega_L) = c^2\mathcal{F}_1(\rho, \omega_L)$, the same noiseless calculation with $\widehat{\rho} = c^{-1}\rho_\star$ gives $E_{\mathrm{pred}} = c^{-2}E_{\mathrm{obs}}$, so the physically matched closed form is

$$c_1 \;=\; \sqrt{E_{\mathrm{obs}}/E_{\mathrm{pred}}}\,, \qquad c_2 \;=\; E_{\mathrm{obs}}/E_{\mathrm{pred}}, \tag{21}$$

i.e., the $\mathcal{F}_1$-form correction is the square root of the $\mathcal{F}_2$-form correction in the noiseless limit. The two forms agree only at the trivial fixed point $E_{\mathrm{obs}} = E_{\mathrm{pred}}$. We use the operator-matched convention in the main-body tables: c_1 for $\mathcal{F}_1$ and c_2 for $\mathcal{F}_2$; numerically, the asymmetry is small for the localization-oriented metrics (GMSD, Hungarian F1, SWD) reported in Section 5 because none of them depend on the absolute density scale, but it does affect density MSE.

D Ill-posedness Diagnostics

This appendix gives the formal counterparts of the four ill-posedness statements of Section 3.3 (frequency suppression, center bias, peak coupling, and source merging and Larmor ambiguity), and quantifies the resulting forward-Jacobian and gradient structure that motivates the Method of Section 4.

D.1 Exponential frequency suppression

Lemma 1 (Frequency decay of the tensor channel power)**.** *For $z_0 > 0$ and $a \in \{x, y, z\}$, the 2D Fourier transform of the source-plane Green-tensor channel evaluated at standoff z_0 satisfies $|\widehat{G_{az}}(\mathbf{k}; z_0)| = O(k\, e^{-kz_0})$ for $k = \|\mathbf{k}\|$. Consequently, for fixed z_0 there exist constants C_{a,z_0} and $m < \infty$ such that*

$$\left|\, \widehat{|G_{az}|^2}\,(\mathbf{k}; z_0)\right| \;\le\; C_{a,z_0}\,(1+k)^m\, e^{-kz_0}, \tag{22}$$

so the linear forward map $\rho \mapsto \sum_a |G_{az}|^2 * \rho$ suppresses spatial mode $\mathbf{k}$ with exponential envelope e^{-kz_0} up to algebraic factors in k.

Proof. Let $\Phi(\mathbf{r}, z) = (4\pi)^{-1}(\|\mathbf{r}\|^2 + z^2)^{-1/2}$ be the half-space magnetostatic scalar potential. Its 2D Fourier transform in $\mathbf{r}$ is, up to convention constants, $\widehat{\Phi}(\mathbf{k}, z) = k^{-1}e^{-kz}$. Each component G_{ia} is a second derivative of Φ in real space; Fourier differentiation contributes factors ik_i or $-k$, hence $|\widehat{G_{az}}(\mathbf{k}; z_0)| \le C_a\, k\, e^{-kz_0}$ for a constant C_a. Since $|G_{az}|^2$ is the pointwise square of a real-valued function, its 2D Fourier transform is the autoconvolution $\widehat{|G_{az}|^2} = (2\pi)^{-2}\widehat{G_{az}} * \widehat{G_{az}}$. Therefore

$$\left|\widehat{|G_{az}|^2}(\mathbf{k})\right| \;\le\; C\int_{\mathbb{R}^2} \|\mathbf{q}\|\;\|\mathbf{k}-\mathbf{q}\|\; e^{-z_0(\|\mathbf{q}\|+\|\mathbf{k}-\mathbf{q}\|)}\, d\mathbf{q} \;\le\; C_{z_0}\,(1+k)^m\, e^{-kz_0},$$

where the second inequality uses the triangle bound $\|\mathbf{q}\| + \|\mathbf{k}-\mathbf{q}\| \ge k$ and bounds the residual integral by an algebraic factor in k for fixed z_0. □

The implication for optimization is that the eigenvalues of the linearized forward map drop off geometrically across spatial scales, making the inverse map exponentially ill-conditioned at high k. Accordingly, NeTMY fits low-k modes before high-k modes, and reversing that order degrades reconstruction (Section 4).

D.2 Finite-window center bias

On a bounded sensing window Ω, the convolutional footprint of a single source pixel at $\mathbf{r}_0$ under $\mathcal{F}_2$ is the translated channel-summed power kernel $\sum_a |G_{az}(\,\cdot\, - \mathbf{r}_0)|^2$, whose energy on Ω is maximized

when $\mathbf{r}_0$ is at the window center and decreases monotonically as $\mathbf{r}_0$ approaches the boundary, because the translated footprint is increasingly truncated by $\partial\Omega$. The forward Jacobian column for source pixel $\mathbf{r}_0$ has squared norm

$$\left\| \frac{\partial \mathcal{F}_2}{\partial \rho(\mathbf{r}_0)} \right\|^2 = \int_W \int_\Omega \Big(\sum_{a \in \{x,y,z\}} |G_{az}(\mathbf{r} - \mathbf{r}_0)|^2 \Big)^2 L(\omega; \omega_L(\mathbf{r}))^2 \, d\mathbf{r} \, d\omega, \tag{23}$$

which expands into auto-channel terms $\sum_a |G_{az}|^4$ plus positive cross-channel terms $\sum_{a \neq a'} |G_{az}|^2 |G_{a'z}|^2$. Since the integrand is a nonnegative localized footprint, truncation by $\partial\Omega$ preserves the center-vs-edge asymmetry. Even from a uniform initialization $\rho_0 \equiv c$, the iter-0 gradient of the data-fidelity term therefore points more strongly toward central pixels than toward boundary pixels, before any structure has formed. We measure this directly in Section 5: the iter-0 physics-loss gradient under $\mathcal{F}_2$ has a center-to-outer ratio of $18.29\times$ on a representative sample, with the peak at the grid center, even though the ground-truth density is off-center. The center bias is therefore a property of the operator and the window, not a bug of any one solver.

D.3 Max-normalization peak coupling

The data-fidelity term in Eq. 19 compares max-normalized noise maps $\widehat{N}_i = N_i / M$, $M = \max_j N_j$. Differentiating $\widehat{N}$ with respect to the unnormalized N:

$$\frac{\partial \widehat{N}_i}{\partial N_j} = \frac{\delta_{ij}}{M} - \frac{N_i}{M^2} \frac{\partial M}{\partial N_j}, \qquad \frac{\partial M}{\partial N_j} = \delta_{jk} \quad \text{where} \quad k = \arg\max_j N_j. \tag{24}$$

For a non-peak pixel ($j \neq k$) this reduces to δ_{ij}/M (purely local), but for the peak pixel ($j = k$) Eq. 24 contains an extra term $-N_i/M^2$ that couples the gradient at N_k to *every* pixel i where N_i is appreciable, with weight proportional to N_i. The chain-rule combination of Eq. 24 with the center bias amplifies the peak whenever the current peak is at the window center, which is exactly the iter-0 situation produced by the center bias on a uniform initialization. The qualitative consequence is that free-density methods (Tikhonov, ADMM) following $\nabla_\rho \mathcal{D}$ literally execute this center-amplifying step at iter 0, and once a centered peak forms it self-reinforces through Eq. 24. We document the resulting energy-barrier between the centered-collapse minimum and the ground-truth density in Section 5.

D.4 Source merging and joint (ρ, ω_L) ambiguity

The effective point-spread width w_{psf} of $\mathcal{F}_2$ is set by the standoff z_0 through the kernel decay of Lemma 1: each tensor channel $|G_{az}|^2$ has spatial extent $\sim z_0$ on the source side, so two sources separated by $\Delta r < w_{\text{psf}} \sim z_0$ produce nearly identical forward responses and cannot be disambiguated from the data alone. A simple lower-bound estimate (Grech et al., 2008) useful for our setup is

$$w_{\text{psf}} \approx 2 z_0 \sqrt{\log 2 \,/\, (k z_0 + 3)} \quad \text{at characteristic mode } k, \tag{25}$$

which we use as a working threshold to label "ill-conditioned" data regimes in Section 5. Beyond pairwise merging, in many-source scenes a continuous family of densities can fit the same forward measurement to within noise, producing axis-aligned cross-shaped artifacts when the optimizer settles on a low-cost, anisotropic explanation (Zhang et al., 2018).

A separate non-uniqueness arises from the joint role of ρ and ω_L in Eq. 2: at any pixel $\mathbf{r}$ where the source contribution $\sum_a |G_{az}|^2 * \rho(\mathbf{r})$ is small, the Lorentzian $L(\omega; \omega_L(\mathbf{r}))$ multiplies a near-zero coefficient and the fidelity term places no constraint on $\omega_L(\mathbf{r})$. The Larmor map is therefore identifiable only on the support of ρ. NeTMY enforces this scope at the architectural level (Appendix E.1, Eq. 26): the predicted Larmor field is multiplicatively masked by an indicator on the predicted support of ρ_θ, so gradients flow into $\omega_L(\mathbf{r})$ only where $\rho_\theta(\mathbf{r})$ is appreciable. Without such gating, any local minimum acquires arbitrary Larmor values on the background, contaminating the per-pixel loss landscape and slowing convergence on the support.

D.5 Summary: which pathology drives which experiment

Table 4 maps each ill-posedness pathology to the empirical signature it predicts and to the experimental section that reports the measurement. **Design rationale.** The components of NeTMY (Section 4)

map to these pathologies as follows. NeTMY is a per-instance coordinate field because matched paired labels are scarce: the ground-truth density is never directly observed, so the reconstruction is optimized against the forward operator for each measurement instead of being learned from pairs. The coordinate parameterization itself filters the raw density gradient (Lemma 2), which targets center bias. Positional encoding with frequency annealing targets frequency suppression: it fits the low-k modes that carry forward signal before it unlocks the suppressed high-k modes (Appendix E.2). The density gate targets peak coupling: writing density as a nonnegative amplitude times a sigmoid gate lets background pixels switch off without driving the amplitude logit into saturation (Appendix E.1). The masked Larmor head targets Larmor ambiguity by restricting gradients into ω_L to the predicted support of ρ_θ (Eq. 26). The softplus density head and the sigmoid Larmor head enforce feasibility, namely nonnegative density and Larmor frequencies inside the device band. Of the two auxiliary losses, $\mathcal{R}_{\mathrm{nm}}$ is a scale-stable companion to the max-normalized fidelity and $\mathcal{R}_{\mathrm{ds}}$ is a measurement-derived support proxy (Appendix E.4). Table 3 reports the cumulative ablation of these components.

Pathology	Predicted signature	Reported in	Mitigated by
Frequency suppression	Coarse-to-fine $>$ fine-to-coarse curriculum	Sec. 5, ablation	Sec. 4, annealed PE
Center bias	Iter-0 centered gradient	Sec. 5, mechanism	Sec. 4, parameterization
Peak coupling	Energy barrier at center collapse	Sec. 5, mechanism	Sec. 4, gating + PE
Source merging, Larmor ambiguity	Cross artifacts, background Larmor noise	Sec. 5, failure	Sec. 4, masked Larmor head

Table 4: Mapping from ill-posedness pathology to empirical signature, the experimental section that reports it, and the design choice in Section 4 that mitigates it.

E Method Details

This appendix gives the full functional and algorithmic details deferred from Section 4: the architecture of the coordinate MLP and density/Larmor heads, the frequency-annealing schedule, the per-stage training schedule, the explicit forms of all loss terms, the post-training scale correction, and the formal derivation of the Jacobian-induced filtering relation $\Delta\rho \approx -\eta\, G_\theta \nabla_\rho \mathcal{L}$.

E.1 Architecture and Output Heads

The coordinate field f_θ in Eq. 4 is realized by the network $f_\theta(\mathbf{x}) = \mathrm{Heads}\big(\mathrm{MLP}_\theta(\gamma_\beta(\mathbf{x}))\big)$, where $\mathbf{x} = (x, y) \in \Omega$, γ_β is the annealed Fourier encoder (Appendix E.2), MLP_θ is a 6-layer fully connected backbone, and Heads is the two-output transform of Eq. 5 and Eq. 26. Coordinates are normalized to the cube $[-1, 1]^2$ before encoding so that the Fourier basis frequencies are device-agnostic. Table 5 summarizes the architectural choices used in every reported run.

Component	Setting
Coordinate input	$(x, y) \in [-1, 1]^2$, normalized by half grid extent
Positional encoding	annealed Fourier features, $K = 12$ octaves, base 2
Encoded input dim	$2 + 2 \cdot 2 \cdot K = 50$
Backbone	6 fully connected layers, hidden 320, $\tanh$ activation
Skip connection	encoded input re-injected at layer 3
Output channels	3: density logit h_ρ, density gate g, Larmor logit h_ω
Density head	$\rho_\theta = \mathrm{softplus}(h_\rho)\, \sigma(g)$
Larmor head	$\omega_{L,\theta} = \omega_{L,\min} + (\omega_{L,\max} - \omega_{L,\min})\, \sigma(h_\omega)$, with mask in Eq. 26
Parameter count	$\approx 4.5 \times 10^5$
Optimizer	AdamW, weight decay 10^{-4}, gradient clip $\|\nabla\theta\| \le 1$
Learning-rate schedule	cosine annealing per stage to $0.01 \cdot \eta_{\mathrm{base}}$

Table 5: NeTMY architectural and optimization settings. All reported main-body and ablation runs use these defaults except where the ablation explicitly toggles a component.

The density-head transform of Eq. 5 is a gated softplus: the softplus enforces nonnegativity, and the sigmoid gate $\sigma(g) \in (0, 1)$ acts as a pixelwise on/off switch that the optimizer can drive close to

zero without saturating the softplus. Empirically, removing the gate (i.e., $\rho_\theta = \mathrm{softplus}(h_\rho)$) makes background suppression substantially harder and weakens sparse-localization metrics (Section 5, ablation).

The Larmor head is sigmoid-mapped into the device-determined band and then multiplied by a hard predicted-support mask,

$$\omega_{L,\theta}^{\mathrm{mask}}(\mathbf{r}) \;=\; m_\theta(\mathbf{r})\,\omega_{L,\theta}^{\mathrm{band}}(\mathbf{r}), \qquad m_\theta(\mathbf{r}) \;=\; \mathbf{1}\big\{\rho_\theta(\mathbf{r}) > \tau \max_{\mathbf{r}'} \rho_\theta(\mathbf{r}')\big\}, \qquad \tau = 0.3, \tag{26}$$

with m_θ treated as a stop-gradient constant during back-prop. The nondifferentiable indicator is therefore not differentiated, and the Larmor logit receives gradients only where $m_\theta = 1$. Pixels with $m_\theta = 0$ are assigned the fixed value 0, which lies outside $[\omega_{L,\min}, \omega_{L,\max}]$ but is not interpreted as a physical Larmor frequency: since $\rho_\theta \simeq 0$ there, support-weighted ℓ_2 losses and metrics on ω_L ignore those pixels. This is the operational counterpart of the Larmor-ambiguity identifiability fact (Appendix D.4) that the data constrains ω_L only on the support of ρ: without the mask, the gradient through ω_L on the background is a near-zero coefficient that nonetheless steers the optimizer toward arbitrary Larmor values, contaminating the per-pixel loss landscape.

E.2 Annealed Fourier Features

Following the Nerfies frequency-annealing schedule (Park et al., 2021; Tancik et al., 2020), the encoded input is

$$\gamma_\beta(\mathbf{x}) \;=\; \Big(\mathbf{x},\; \big\{\, w_k(\beta)\,\sin(2^k\pi\,\mathbf{x}),\; w_k(\beta)\,\cos(2^k\pi\,\mathbf{x})\,\big\}_{k=0}^{K-1}\Big), \tag{27}$$

$$w_k(\beta) \;=\; \frac{1-\cos\big(\pi\,\mathrm{clip}(\beta-k,0,1)\big)}{2},$$

where $K = 12$ and $\beta \in [0, K]$ is a per-stage progress scalar. We schedule β linearly within each stage, $\beta(t) = K \cdot t/T$, with t the within-stage epoch and T the stage length. At $\beta = 0$ only the linear coordinate input contributes; at $\beta = K$ all K octaves are fully active and Eq. 27 reduces to the standard Fourier feature encoding of (Tancik et al., 2020). The cosine envelope $w_k(\beta)$ activates each band smoothly, avoiding the gradient-shock that abrupt high-frequency activation would induce.

The annealing schedule is the operational counterpart of frequency suppression (Lemma 1): low spatial frequencies of ρ carry $\Theta(1)$ forward signal, while high frequencies are suppressed with envelope e^{-kz_0} up to algebraic factors. Fitting low frequencies first delivers a coarse minimum that is closer to the true support before the high-frequency capacity of the MLP is unlocked. We measure this as a coarse-to-fine vs. fine-to-coarse curriculum ablation in Section 5.

E.3 Multiscale Training Schedule

Optimization runs in two stages on a single MLP (parameters carried across stages); only the coordinate grid resolution and the resolution at which $\mathcal{F}_2$ is evaluated change between stages. Table 6 summarizes the schedule used in every reported NeTMY run.

Stage	Resolution	Epochs (default)	Learning rate	Notes
1	32×32	3000	$\eta_{\mathrm{base}} = 10^{-3}$	coarse low-frequency support recovery
2	64×64	7000	$0.5 \cdot \eta_{\mathrm{base}}$	fine high-frequency refinement

Table 6: NeTMY two-stage multiscale schedule. Total 10,000 epochs per measurement. Within each stage, learning rate decays under cosine annealing to $0.01 \cdot \eta_{\mathrm{base}}$. Frequency annealing β is reset to 0 at the start of each stage so high-frequency PE bands are re-introduced from coarse to fine within each resolution.

The same MLP weights are carried across stages with no re-initialization, so the fitted parameters at the end of Stage 1 act as a structured initialization for Stage 2. Coordinates are recreated at the new resolution, and the differentiable forward operator $\mathcal{F}_2$ is evaluated at that resolution using a precomputed FFT kernel cache. Per-instance wall-clock on a single A6000 GPU ranges from 273 s (median) to $\sim$780 s (untrimmed mean) over the main 512-sample benchmark; runtime is reported in Appendix F.11.

E.4 Loss Terms

The full per-measurement objective of Eq. 6 is the sum of a canonical fidelity, two standard regularizers, and two NeTMY-specific physics losses. Below we give the explicit form and the role of each term; default loss weights are listed in Table 7.

- **Spectrum fidelity** $\mathcal{D}$: the pixelwise log-MSE on max-normalized noise maps as defined in Eq. 19. This is the sole data-fidelity term.
- **Sparsity** $\|\rho_\theta\|_1 = \sum_{\mathbf{r}} |\rho_\theta(\mathbf{r})|$: encourages point-like sources.
- **Anisotropic total variation** $\mathrm{TV}(\rho_\theta)$ as defined after Eq. 3: smooths small-scale noise without penalizing sparse peaks.
- **Mean-normalized noise-map loss** $\mathcal{R}_{\mathrm{nm}} = \mathrm{MSE}\big(N_{\mathrm{pred}}/\overline{N}_{\mathrm{pred}},\, N_{\mathrm{obs}}/\overline{N}_{\mathrm{obs}}\big)$, where $\overline{N}$ denotes the spatial mean of the frequency-summed noise map. This is a scale-stable companion to $\mathcal{D}$ that anchors gradients on the support without the max-normalization peak coupling (Appendix D.3). In the 64×64 stage we replace the log-MSE form of $\mathcal{D}$ with $\mathcal{R}_{\mathrm{nm}}$ as the primary fidelity term, since the log-domain comparison is most useful for early-stage support discovery.
- **Direct-density loss** $\mathcal{R}_{\mathrm{ds}} = \mathrm{MSE}\big(\rho_\theta^2/\overline{\rho_\theta^2},\, N_{\mathrm{obs}}/\overline{N}_{\mathrm{obs}}\big)$: a soft, oracle-free support regularizer. The interpretation "spectrum integral $\propto \rho^2$" is exact only for the coherent point-source operator $\mathcal{F}_1$; under the thermal operator $\mathcal{F}_2$ used in this paper, the spectrum is linear in ρ, so $\mathcal{R}_{\mathrm{ds}}$ is a heuristic that nudges large predicted density values toward bright observed noise pixels rather than an exact functional identity. We retain it because the ablation shows that it accelerates support discovery without harming $\mathcal{F}_2$ inversion. No ground-truth density labels enter this term.

Term	Weight	Role
$\mathcal{D}$	2.0	spectrum fidelity, log-MSE
$\mathcal{R}_{\mathrm{nm}}$	0.5	noise-map fidelity, mean-norm
$\mathcal{R}_{\mathrm{ds}}$	0.1	direct-density proxy, mean-norm
$\|\rho_\theta\|_1$	$10^{-2} + 10^{-3}$	sparsity (combined: support + L1)
$\mathrm{TV}(\rho_\theta)$	10^{-3}	smoothness on small scales

Table 7: Default loss weights for NeTMY. The two ℓ_1 entries are listed separately because the codebase implements two L1 contributions with different weights (one as a "sparsity" prior across both stages, one as an explicit L1 regularizer that can be ablated independently).

E.5 Energy-Anchored Scale Correction

After the per-measurement optimization terminates, the predicted density is rescaled by the energy-ratio anchor of Proposition 1. For a given inversion operator $\mathcal{F} \in \{\mathcal{F}_1, \mathcal{F}_2\}$, we forward-evaluate

$$\widehat{S} = \mathcal{F}(\widehat{\rho}, \omega_{L,\theta}) \tag{28}$$

on the final 64×64 grid and compute the pre-normalization energies

$$E_{\mathrm{pred}} = \sum_{\omega,\mathbf{r}} \widehat{S}(\omega, \mathbf{r}), \qquad E_{\mathrm{obs}} = \sum_{\omega,\mathbf{r}} S_{\mathrm{obs}}(\omega, \mathbf{r}). \tag{29}$$

We then apply the operator-matched scale correction

$$\widehat{\rho}_{\mathrm{scaled}} = \alpha_{\mathcal{F}} \widehat{\rho}, \qquad \alpha_{\mathcal{F}} = \begin{cases} \sqrt{E_{\mathrm{obs}}/E_{\mathrm{pred}}}, & \mathcal{F} = \mathcal{F}_1, \\ E_{\mathrm{obs}}/E_{\mathrm{pred}}, & \mathcal{F} = \mathcal{F}_2. \end{cases} \tag{30}$$

The square-root correction for $\mathcal{F}_1$ follows from the quadratic homogeneity $\mathcal{F}_1(c\rho, \omega_L) = c^2\mathcal{F}_1(\rho, \omega_L)$, whereas the linear correction for $\mathcal{F}_2$ follows from $\mathcal{F}_2(c\rho, \omega_L) = c\mathcal{F}_2(\rho, \omega_L)$; the derivation is given in Appendix C.4. This correction mainly affects density-scale-dependent quantities such as density MSE. The primary localization-oriented metrics reported in Section 5, including GMSD, Hungarian F1, and SWD, are insensitive or only weakly sensitive to the absolute density scale. Density MSE is therefore reported as a secondary metric. We apply the correction as a one-shot post-processing step rather than as a soft penalty inside Eq. 6. This avoids reweighting the $\mathcal{D}$-gradient

by a scalar that depends on the current iterate, which would otherwise couple the support-discovery dynamics to the absolute amplitude estimate. Within each fixed operator setting, all methods are evaluated with the same operator-matched scale convention, so comparisons across parameterizations are not affected by method-specific scale-correction choices.

E.6 Filtering View of the Update

We give the formal version of the filtering identity Eq. 7 that motivates the optimization-geometry interpretation in Section 4.4.

Lemma 2 (Jacobian-induced filtering of the density update)**.** *Let $f_\theta : \Omega \to \mathbb{R}_{\geq 0}$ be a coordinate field parameterized by $\theta \in \mathbb{R}^P$, write $\rho_\theta = f_\theta$ as a vector in $\mathbb{R}^{|\Omega|}$, and let $\mathcal{L}(\rho)$ be a differentiable scalar loss on density. A vanilla first-order step on θ at learning rate $\eta > 0$, $\theta_{t+1} = \theta_t - \eta \nabla_\theta \mathcal{L}(\rho_{\theta_t})$, induces*

$$\Delta\rho \;=\; \rho_{\theta_{t+1}} - \rho_{\theta_t} \;=\; -\eta\, J_\theta J_\theta^\top \nabla_\rho \mathcal{L}(\rho_{\theta_t}) \;+\; O\big(\eta^2 \,\|J_\theta\|_{\mathrm{op}}^2 \,\|\nabla_\rho \mathcal{L}\|^2 \,\big\|\partial_\theta^2 f_\theta\big\|_{\mathrm{op}}\big), \tag{31}$$

where $J_\theta = \partial f_\theta / \partial\theta \in \mathbb{R}^{|\Omega| \times P}$ is the parameterization Jacobian and $\big\|\partial_\theta^2 f_\theta\big\|_{\mathrm{op}}$ is the bilinear operator norm of the second-derivative tensor of f_θ. The first-order term is the raw density gradient $\nabla_\rho \mathcal{L}$ filtered by the positive semidefinite kernel $G_\theta = J_\theta J_\theta^\top \in \mathbb{R}^{|\Omega| \times |\Omega|}$.

Proof. Let $g = \nabla_\rho \mathcal{L}(\rho_{\theta_t})$. Chain rule gives $\nabla_\theta \mathcal{L} = J_\theta^\top g$, hence $\Delta\theta = -\eta\, J_\theta^\top g$ and $\|\Delta\theta\| \leq \eta\, \|J_\theta\|_{\mathrm{op}}\, \|g\|$. Taylor expansion of f_θ in θ yields $\rho_{\theta_t + \Delta\theta} = \rho_{\theta_t} + J_\theta \Delta\theta + \frac{1}{2}\, \partial_\theta^2 f_\theta[\Delta\theta, \Delta\theta] + O(\|\Delta\theta\|^3)$, which combined with the bound on $\|\Delta\theta\|^2$ gives Eq. 31. □

For any parameter-space preconditioner P_t, including AdamW's, the first-order update $-\eta\, J_\theta P_t J_\theta^\top \nabla_\rho \mathcal{L}$ lies in the range of J_θ, so the subspace statements below hold for the optimizer NeTMY uses.

Consequence for ill-posedness. The filtering kernel

$$G_\theta = J_\theta J_\theta^\top \tag{32}$$

is positive semidefinite by construction. It acts as a structurally coupled filtering operator across pixels. Writing the Jacobian column functions as

$$\psi_p(\mathbf{r}) = \frac{\partial f_\theta(\mathbf{r})}{\partial \theta_p}, \tag{33}$$

we have, for any raw density gradient $g(\mathbf{r})$,

$$(G_\theta g)(\mathbf{r}) = \sum_{p=1}^{P} \psi_p(\mathbf{r}) \langle \psi_p, g \rangle \in \mathrm{span}\{\psi_1, \ldots, \psi_P\}. \tag{34}$$

Hence G_θ maps the raw pixel-wise gradient into the structured function space allowed by the coordinate parameterization, rather than allowing an arbitrary free pixel-wise update. Equivalently,

$$\mathrm{rank}(G_\theta) = \mathrm{rank}(J_\theta J_\theta^\top) \leq \min(|\Omega|, P), \tag{35}$$

so the induced density update is constrained to a structured low-dimensional subspace of the full pixel space.

For a coordinate MLP with annealed Fourier features, the functions $\psi_p(\mathbf{r})$ are smooth in the early stages because high-frequency Fourier bands are suppressed in the encoding γ_β. If

$$w_k(\beta) \approx 0, \qquad k > k_\beta, \tag{36}$$

then the effective bandwidth of the encoding is approximately

$$B_\beta \sim 2^{k_\beta} \pi. \tag{37}$$

Consequently, the Jacobian columns are dominated by modes below B_β, and

$$\widehat{\psi_p}(\boldsymbol{\xi}) \approx 0, \qquad |\boldsymbol{\xi}| > B_\beta. \tag{38}$$

It follows that

$$\widehat{G_\theta g}(\boldsymbol{\xi}) = \sum_{p=1}^{P} \langle \psi_p, g \rangle \widehat{\psi_p}(\boldsymbol{\xi}) \approx 0, \qquad |\boldsymbol{\xi}| > B_\beta. \tag{39}$$

Thus G_θ initially behaves as a low-pass smoothing operator, consistent with the exponential frequency suppression of Lemma 1. As the annealing parameter β increases, higher Fourier bands are gradually activated, the effective bandwidth of the Jacobian columns grows, and the spectrum of G_θ becomes progressively richer. In this sense, the coordinate parameterization redistributes a single-pixel density perturbation into a smooth, spatially coupled update during early training, and only later allows sharper, higher-frequency corrections. A sharp center spike in $\nabla_\rho \mathcal{L}$ (induced by center bias and peak coupling on a uniform initialization) is therefore mapped to a spatially distributed image-space update whose center-to-outer ratio is bounded by the corresponding ratio of G_θ rather than by $\nabla_\rho \mathcal{L}$ itself. We measure this directly in Section 5: the realized iter-0 NeTMY update $|\Delta\rho|$ is spatially distributed, with a center-to-outer ratio of approximately $1.6\times$ against $18.29\times$ for the iter-0 free-density gradient under the same forward operator and uniform initialization.

Comparison to free-density and quasi-Newton solvers. For Tikhonov and ADMM, ρ is the optimization variable and the parameter-space update is a (sub-)gradient step, so the effective parameterization Jacobian is the identity ($J = I$, $G_\theta = I$) and Eq. 31 reduces to $\Delta\rho = -\eta \nabla_\rho \mathcal{L}$: in the unpreconditioned first-order limit the raw density gradient is executed verbatim, including the center bias and the self-reinforcing peak coupling. L-BFGS also optimizes ρ directly, but replaces the gradient direction with $-H_t^{-1}\nabla_\rho \mathcal{L}$ via a low-rank approximate-inverse-Hessian H_t^{-1} assembled from past gradient differences; this is a parameter-space preconditioner orthogonal to the parameterization-Jacobian filter analyzed here, and provides an independent route to escape the centered-collapse basin (consistent with the L-BFGS center-mass ratio of 0.081 in Section 5). The comparison in Section 5 is consistent with this filtering view: NeTMY, whose update is filtered by the G_θ of a fully connected MLP across the grid, is less attracted to the center collapse driven by center bias and peak coupling than Tikhonov and ADMM, which have no parameterization. The filtering view accounts for the observed iter-0 update structure and for the direction of the NeTMY versus free-density comparison; the gate and the curriculum also contribute (Table 3).

E.7 Algorithm Summary

Algorithm 1 NeTMY per-measurement optimization (operator $\mathcal{F} \in \{\mathcal{F}_1, \mathcal{F}_2\}$)

Input: measured spectrum S_{obs}, inversion operator $\mathcal{F} \in \{\mathcal{F}_1, \mathcal{F}_2\}$, total epochs T_{tot}, base lr η_0
Initialize MLP parameters $\theta \sim$ default init
for stage $s \in \{1, 2\}$ with resolution $R_s \in \{32, 64\}$ and epochs T_s, lr η_s **do**
  Build coordinate grid $\{\mathbf{x}_{ij}\}$ at resolution R_s
  Precompute FFT kernels for $\mathcal{F}$ at R_s
  for $t = 1, \ldots, T_s$ **do**
    $\beta \leftarrow K \cdot t / T_s$
    $(\rho_\theta, \omega_{L,\theta}) \leftarrow f_\theta(\mathbf{x}_{ij}; \beta)$ ▷ Eq. 5, Eq. 26
    $\widehat{S} \leftarrow \mathcal{F}(\rho_\theta, \omega_{L,\theta})$ ▷ Eq. 2
    $\mathcal{L} \leftarrow$ Eq. 6 with weights from Table 7
    $\theta \leftarrow$ AdamW step with cosine-annealed lr, gradient clip
  end for
end for
$E_{\text{pred}} \leftarrow \sum_{\omega,\mathbf{r}} \mathcal{F}(\rho_\theta, \omega_{L,\theta})$; $E_{\text{obs}} \leftarrow \sum_{\omega,\mathbf{r}} S_{\text{obs}}$
$\alpha_{\mathcal{F}} \leftarrow E_{\text{obs}}/E_{\text{pred}}$ if $\mathcal{F} = \mathcal{F}_2$, else $\sqrt{E_{\text{obs}}/E_{\text{pred}}}$
$\widehat{\rho} \leftarrow \alpha_{\mathcal{F}}\, \rho_\theta$ ▷ Eq. 30
return $(\widehat{\rho}, \omega_{L,\theta})$

F Experimental Details

This appendix expands Section 5: dataset construction (Appendix F.1), baseline implementation (Appendix F.2), metric definitions (Appendix F.3), the full $\mathcal{F}_2/\mathcal{F}_2$ matched-operator results and per-class breakdowns (Appendix F.4), the learned-prior and directional-cross comparisons (Appendix F.5,

Appendix F.6), the mechanism aggregate (Appendix F.7), the full ablation sweep including hyperparameter sweeps (Appendix F.8), the supervised-baseline distribution-shift study (Appendix F.9), the four-step real-data protocol (Appendix F.10), runtime and total-cost accounting and failure modes (Appendix F.11, Appendix F.12). Implementation hyperparameters that are NeTMY-specific (architecture, optimizer, multiscale schedule, loss weights, scale correction) are in Appendix E.

F.1 Datasets

Cross-fidelity benchmark (Section 5.2, 512 samples). The cross-fidelity benchmark uses spectra generated by the source-side direct simulator $\mathcal{F}_3$ (Appendix B.4) on a 64×64 grid at standoff $z_0 = 20\,\text{nm}$ and grid spacing $20\,\text{nm}$, with sensor noise σ_ε^2 matched to per-sample dynamic range. Source counts and minimum separations are drawn from nine scene classes that cover the regimes most relevant to NV relaxometry: *few/close*, *few/medium*, *few/far*, *medium/close*, *medium/medium*, *medium/far*, *many/close*, *many/medium*, *many/far*. The Larmor field $\omega_L \in [1.5, 2.5]$ GHz is evaluated on a 50-frequency grid and is constant within each ground-truth source. Sample IDs follow the pattern $\langle\text{seq}\rangle_\langle\text{class}\rangle$ used internally for cross-referencing. Each method is run on the same 512 samples under both $\mathcal{F}_1$ and $\mathcal{F}_2$ for inversion (so the table reports $3 \times 2 = 6$ method-operator conditions).

Matched-operator benchmark (Section 5.2, 512 samples). The matched-operator benchmark uses spectra generated and inverted by $\mathcal{F}_2$, on the same physical setup as the cross-fidelity benchmark. The samples cover the same scene classes and are used as a broader-coverage view that includes amortized and untrained-prior baselines (DeepDecoder, HybridNeTMY) for which the $\mathcal{F}_3$ form is not implemented. The matched-operator setup is intentionally an inverse-crime configuration (Kaipio & Somersalo, 2007); we report it only as a complementary view, with the cross-fidelity benchmark as the primary evidence.

Real-data dataset (Section 5.5). The α-$RuCl_3$ dataset is the publicly released NV-relaxometry data from Kumar et al. (2024), comprising T_1-with-flake and T_1-without-flake measurements at 8 NVs (NV2, NV3, NV4, NV6, NV7, NV8, NV9, NV10) at the same magnetic-field setting, plus a magnetic-field-dependent dataset on a separate NV (used only for spectral calibration). The crystallographic spin density per layer is $\rho_0 = 6.45 \times 10^{18}\,\text{m}^{-2}$ and the effective magnetic moment is $\mu_{\text{eff}} = 2.31\,\mu_B$ (Kumar et al., 2024). The SRIM-implied NV-depth prior is $14 \pm 5\,\text{nm}$, derived from the reported ion implantation energy of $9.8\,\text{keV}$ and SRIM ion-range/straggle simulation.

F.2 Baseline Implementation

All baselines minimize Eq. 3 on the same coordinate grid as NeTMY, using the same forward operator and the same observed spectrum, but differ in how ρ is parameterized and which auxiliary regularizers are added. We give compact algorithmic settings for each.

Tikhonov. Free-density $\rho \in \mathbb{R}_{\geq 0}^{|\Omega|}$, optimized by Adam (Kingma & Ba, 2014) with learning rate 5×10^{-3}, weight decay 10^{-5}, and gradient clipping at 1 for 5000 epochs. Regularizers are ℓ_2 (weight 10^{-3}) and TV (weight 10^{-3}), matching the NeTMY weights up to the absent NeTMY-specific physics losses. The scale-correction post-processing of Eq. 30 is applied to all methods identically for paired comparison.

ADMM. Variable-splitting on $\rho = z$ with augmented-Lagrangian penalty $\mu = 10^{-3}$, box constraints $[0, \rho_{\max}]$ via projection, and proximal ℓ_1 shrinkage with weight 10^{-2}. The data-fidelity step uses Adam at learning rate 5×10^{-3} for 30 inner iterations per outer cycle, run for 200 outer cycles. Stopping criteria are $\|\rho - z\|_\infty < 10^{-3}$ or maximum iterations.

DeepDecoder. Untrained convolutional decoder prior of Heckel & Hand (2018) with 5 upsampling stages, channel widths $[128, 128, 128, 128, 1]$, and bilinear upsampling. Optimized for 5000 steps with Adam at learning rate 10^{-3}. The pixel output is multiplied by 1 to remain nonnegative; values < 0 are clamped at zero.

U-Net (and U-Net+physics). Four-level U-Net with channel widths $[32, 64, 128, 256]$, instance normalization, and a final pixelwise softplus, trained on 300 dense paired samples for 200 epochs with Adam at learning rate 10^{-3}. The U-Net+physics variant additionally optimizes a physics-residual loss $\|\mathcal{F}_2(\widehat{\rho}, \omega_L) - S_{\text{obs}}\|_2^2$ alongside the supervised pixel loss. We report only the supervised-trained model and evaluate on sparse test scenes (which are out-of-distribution by construction); see Appendix F.9.

GAN-SL+physics. Generative adversarial network with the same backbone as the U-Net for the generator, a PatchGAN discriminator, and a physics-loss term applied to the generator's output. Training and evaluation protocols mirror the U-Net.

HybridNeTMY. A coordinate-image hybrid in which the input to a U-Net is concatenated with a Fourier-feature encoding of the coordinate grid, mirroring the coordinate-feature input of NeTMY but retaining a CNN architecture and supervised training. HybridNeTMY tests whether coordinate features alone produce NeTMY-like behavior under supervised training; Appendix F.9 shows they do not.

L-BFGS. Limited-memory BFGS (Liu & Nocedal, 1989) on free-density ρ with history depth 20, line-search Wolfe conditions, and 100 outer iterations. Used only in the mechanism analysis (Section 5.3, Appendix F.7) as a reference quasi-Newton free-density solver.

F.3 Metric Definitions

Gradient Magnitude Similarity Deviation (GMSD). GMSD (Xue et al., 2013) measures structural fidelity through the standard deviation of a pointwise gradient-magnitude similarity map. Given the predicted density $\widehat{\rho}$ and ground-truth density $\rho_\star$, both rescaled to share the maximum value, define gradient magnitudes $G_X = \sqrt{(D_x * X)^2 + (D_y * X)^2}$ for $X \in \{\widehat{\rho}, \rho_\star\}$ with the Prewitt operators D_x, D_y. The pointwise GMS is $\mathrm{GMS}(\mathbf{r}) = (2G_{\widehat{\rho}}G_{\rho_\star} + c)/(G_{\widehat{\rho}}^2 + G_{\rho_\star}^2 + c)$ with stability constant c, and $\mathrm{GMSD} = \sqrt{\mathrm{Var}_{\mathbf{r}}\,\mathrm{GMS}(\mathbf{r})}$. Lower is better.

Hungarian $\mathbf{F_1}$. A localized peak-matching score: peaks are extracted from $\widehat{\rho}$ and $\rho_\star$ by local-maximum filtering above an absolute threshold of 5% of the per-image maximum, paired by the Hungarian algorithm (Kuhn, 1955) under a fixed matching radius of $r_{\mathrm{match}} = 2$ pixels (corresponding to one half of the Section 3.3 effective PSF width w_{psf} at the standoff and characteristic mode used in the cross-fidelity benchmark), and scored as $F_1 = 2\,\mathrm{TP}/(2\,\mathrm{TP} + \mathrm{FP} + \mathrm{FN})$. Higher is better. $\mathrm{F_1}$ is well-defined only when both reconstruction and ground truth contain at least one supra-threshold peak; we report $F_1 = 0$ for the degenerate case.

Sliced Wasserstein Distance (SWD). The sliced Wasserstein-2 distance (Bonneel et al., 2015), evaluated as the mean over 128 random projections drawn uniformly from the unit sphere of the 1-D Wasserstein-2 distance between the two projected mass distributions $P_{\theta\#}\widehat{\rho}/\left\|\widehat{\rho}\right\|_1$ and $P_{\theta\#}\rho_\star/\left\|\rho_\star\right\|_1$. Lower is better.

Density MSE. The pixelwise mean square error $\mathrm{MSE} = |\Omega|^{-1}\sum_{\mathbf{r}\in\Omega}(\widehat{\rho}(\mathbf{r}) - \rho_\star(\mathbf{r}))^2$ on the post-scale-corrected reconstruction. Reported as a secondary metric only because the background-dominated character of sparse density fields makes MSE insensitive to localization quality.

Masked SSIM. The structural similarity index (Wang et al., 2004) evaluated on the support of $\rho_\star$ (i.e., on pixels where $\rho_\star > 0$), reported in Section 5.4 alongside MSE.

F.4 Matched-Operator Density-MSE Leaderboard

Table 8 reports the per-method density-MSE leaderboard on the 512-sample $\mathcal{F}_2/\mathcal{F}_2$ matched-operator benchmark, with secondary metrics.

Table 8: Density-MSE leaderboard on the 512-sample $\mathcal{F}_2/\mathcal{F}_2$ matched-operator benchmark. Density MSE and noise-map MSE are means over 512 sparse samples. Train time is mean wall-clock on a single A6000 GPU.

Method	Density MSE↓	Noise MSE↓	Train time (s)	Notes
NeTMY (Ours)	**0.0037**	**0.21**	780	per-instance, no labels
Tikhonov	0.0117	0.81	1.5	free-density + ℓ_2/TV
ADMM	0.2555	22.94	1.1	free-density + ADMM
DeepDecoder	0.2558	24.25	259	untrained CNN prior
HybridNeTMY	1.011	0.23	399	supervised, dense-trained

The matched-operator ranking is consistent with the cross-fidelity benchmark of Table 1 on density-MSE, but, as discussed in Section 5.2, density-MSE alone is insensitive to centered-collapse failure modes that the cross-fidelity benchmark exposes through Hungarian $\mathrm{F_1}$ and SWD. The supervised

HybridNeTMY's poor MSE reflects the dense-to-sparse distribution-shift failure documented in Appendix F.9.

F.5 Learned-Prior Baselines: Plug-and-Play and RED

Plug-and-play (PnP) priors (Venkatakrishnan et al., 2013) and regularization by denoising (RED) (Romano et al., 2017) use a pretrained denoiser as the prior and need no paired inverse-problem data, so they are label-free comparators to NeTMY. We evaluate PnP with proximal gradient descent (PnP-PGD) and RED with gradient descent (RED-GD), both with the official frozen grayscale DRUNet checkpoint (Zhang et al., 2022). Hyperparameters were selected only on a disjoint 32-scene validation set. Spectrum input scaling is observation-only, and density is projected to be nonnegative. All five methods in Table 9 were freshly evaluated on an identical held-out 128-scene stratified subset of the $\mathcal{F}_3$-generated benchmark (Section 5.1); the entries for Tikhonov, ADMM, and NeTMY are therefore reruns on this subset and differ from Table 1. All five methods invert through the same $\mathcal{F}_1$/$\mathcal{F}_2$ operator definitions. For comparability with Table 1, ADMM retains its density-masked Larmor construction on $(0.0, 1.6)$, while the other four use the common derived-Larmor chain on $(1.4, 1.6)$. All methods run to convergence within fixed per-scene budgets: Tikhonov 3,000, ADMM 500, NeTMY 20,000, PnP/RED 200. Entries are untrimmed means. GMSD, Hungarian F_1, and SWD are primary and MSE is secondary (Appendix F.3); the MSE column is heavy-tailed (means exceed medians by 50–70%).

Table 9: Learned-prior comparison on a held-out 128-scene stratified subset of the $\mathcal{F}_3$-generated benchmark, inverted under $\mathcal{F}_1$ or $\mathcal{F}_2$. Untrimmed means. Tikhonov, ADMM, and NeTMY are reruns on this subset. Best per column within each operator block in bold.

Operator	Method	GMSD↓	Hungarian F_1 ↑	SWD↓	MSE↓
$\mathcal{F}_1$	Tikhonov	**0.1782**	0.4649	0.1196	0.01521
$\mathcal{F}_1$	ADMM	0.2354	0.4331	**0.05224**	0.01488
$\mathcal{F}_1$	NeTMY	0.2202	**0.6285**	0.06342	**0.01294**
$\mathcal{F}_1$	PnP-PGD	0.1992	0.1443	0.1379	0.01817
$\mathcal{F}_1$	RED-GD	0.1797	0.01418	0.08904	0.01763
$\mathcal{F}_2$	Tikhonov	0.2016	0.7615	0.09816	**1.1283**
$\mathcal{F}_2$	ADMM	0.1951	0.06825	0.1358	1.6271
$\mathcal{F}_2$	NeTMY	**0.1367**	**0.8477**	**0.04968**	4.4592
$\mathcal{F}_2$	PnP-PGD	0.2051	0.3772	0.1050	2.6609
$\mathcal{F}_2$	RED-GD	0.1737	0.04520	0.1792	6.8291

Neither learned-prior method leads any primary metric under either operator, and median Hungarian F_1 is 0.000 for PnP-PGD under $\mathcal{F}_1$ and for RED-GD under both operators. The validation record does not favor a larger budget: raising the PnP-PGD iteration cap from 200 to 500 was rejected on validation score under both operators and diverged in 7 of its 12 configurations under $\mathcal{F}_1$, returning a non-finite fidelity or predicted spectrum, while the same increase moved the data fidelity of RED-GD by under 6%. This behavior matches the scale and peak-coupling analysis of this paper. $\mathcal{F}_1$ is quadratic in density (Appendices B.3 and C.4) and the observations span tens of orders of magnitude, while the DRUNet interface requires per-iterate max normalization of the density iterate. That normalization is a global dependence on the current peak, analogous to the peak coupling of Appendix D.3, which the density gate of NeTMY mitigates (Section 4.1).

F.6 Directional Operator Crosses

The $\mathcal{F}_3$ benchmark tests mismatch against an independent generator. A literal cross isolates the two inversion operators: spectra are generated with one of $\mathcal{F}_1$ and $\mathcal{F}_2$ and inverted with the other. We run both directions on paired scenes per direction with three solvers (Tikhonov, ADMM, NeTMY). The cross spectra are deterministic (no added measurement noise, unlike the $\mathcal{F}_3$ benchmark), and all methods run to convergence within fixed per-scene budgets (Tikhonov 3,000, ADMM 500, NeTMY 20,000). GMSD, Hungarian F_1, and SWD are primary and MSE is secondary. Under a directional cross, MSE is additionally confounded: the observation-only scale correction must reconcile quadratic generation with linear inversion or the reverse (Appendices C.4 and E.5), with per-scene corrections spanning six orders of magnitude. We therefore report MSE without ranking in both directions (Tables 10 and 11).

Table 10: Directional cross: spectra generated with $\mathcal{F}_1$, inverted with $\mathcal{F}_2$. Untrimmed means over deterministic (noiseless) cross spectra. Best primary metric in bold; MSE is reported without ranking because the directional cross mixes it with the scale reconciliation described in the text.

Method	GMSD↓	Hungarian F_1 ↑	SWD↓	MSE↓
Tikhonov	0.1973	0.7940	0.09693	0.1845
ADMM	0.1916	0.05261	0.1359	0.3800
NeTMY	**0.1198**	**0.8571**	**0.05030**	1.8794

Table 11: Directional cross: spectra generated with $\mathcal{F}_2$, inverted with $\mathcal{F}_1$. Untrimmed means over deterministic (noiseless) cross spectra. Best primary metric in bold; MSE is reported without ranking because the directional cross mixes it with the scale reconciliation described in the text.

Method	GMSD↓	Hungarian F_1 ↑	SWD↓	MSE↓
Tikhonov	**0.3038**	0.6768	0.07412	0.01013
ADMM	0.3758	**0.7929**	**0.03453**	0.01020
NeTMY	0.3720	0.3512	0.04639	0.00984

Untrimmed means are the predeclared summary; medians and paired bootstrap comparisons follow. For $\mathcal{F}_1 \to \mathcal{F}_2$, NeTMY is best on both mean and median for GMSD (0.1198 / 0.1027) and SWD (0.0503 / 0.0446), with paired bootstrap intervals excluding zero against both comparators. Its Hungarian F_1 advantage is mean-only (0.8571 vs. 0.7940 for Tikhonov) and reverses on medians (0.8819 vs. 0.9091); consistently, the paired interval against Tikhonov crosses zero.

The $\mathcal{F}_2 \to \mathcal{F}_1$ direction reverses the ranking: Tikhonov leads mean GMSD (0.3038), and ADMM leads mean Hungarian F_1 (0.7929; median 0.899) and SWD (0.03453). The inversion operator changes which parameterization leads the primary metrics. The asymmetry between the two directions tracks independent evidence on the operators. $\mathcal{F}_1$ introduces nonphysical cross terms (Eq. 12, Appendix B.5). On the real data it is disfavored on the depth-amplitude and conditioning axes and yields a degenerate valley (condition number $301{,}139$ vs. 931 for $\mathcal{F}_2$; Section 5.5). The interpolation path that crosses an $h \approx 1.12$ barrier under $\mathcal{F}_2$ is monotone under $\mathcal{F}_1$ (Section 5.3), a landscape that favors free-density solvers. The components of NeTMY target pathologies diagnosed under the tensor operator (Sections 3.3 and 4.1), so its advantage concentrates in the direction inverted under $\mathcal{F}_2$. The comparison that controls for realism remains the $\mathcal{F}_3$-generated benchmark of Table 1, where NeTMY leads all primary metrics under $\mathcal{F}_2$ inversion.

F.7 Mechanism Aggregate and Per-Method Diagnostics

32-run ADMM aggregate sweep. To isolate the operator effect from solver-specific choices on the centered-collapse phenomenon, we run 32 ADMM configurations (4 samples $\times$ 8 loss-baseline settings) under both $\mathcal{F}_1$ and $\mathcal{F}_2$ inversion and report the center-mass ratio at termination. Mean center-mass ratio rises from 0.553 ± 0.344 under $\mathcal{F}_1$ to 0.776 ± 0.369 under $\mathcal{F}_2$ (Table 12), confirming that the more faithful operator drives a stronger center collapse for free-density solvers. The 32-run sweep is ADMM-only by construction; the cross-method evidence is the per-sample comparison of Figure 4.

Table 12: ADMM-only center-collapse aggregate over 32 runs (4 samples $\times$ 8 loss-baseline configurations). Higher center-mass ratio means more centered collapse.

Operator	Center-mass ratio (mean$\pm$std)	Sample count
$\mathcal{F}_1$ (scalar)	0.553 ± 0.344	32
$\mathcal{F}_2$ (tensor)	0.776 ± 0.369	32

NeTMY first-step image-space update. Lemma 2 predicts that the realized iter-0 NeTMY update $|\Delta\rho|$ should be smoothed and structurally coupled across pixels rather than executed verbatim as a singular center spike. We measure this directly on the same many-source sample (sample 024): the iter-0 image-space update $|\Delta\rho|$ is spatially distributed with a center-to-outer ratio of approximately

$1.6\times$ (vs $18.29\times$ for the raw density-space gradient). The center-to-outer ratio is therefore $\sim 11\times$ lower than for the raw gradient, as Lemma 2 predicts.

F.7.1 Operator-Spectral Diagnostic

Lemma 2 states that the parameterization filters the raw density gradient through $G_\theta = J_\theta J_\theta^\top$. We quantify this filtering in the singular coordinates of the measurement map. On four stratified scenes, at the same initial density state and with ω_L frozen at ground truth, we linearize the processed $\mathcal{F}_2$ map and compare three image-space directions: the raw direction $-g$ with $g = \nabla_\rho \mathcal{L}$, the parameterization-induced direction $-J_\theta J_\theta^\top g$, and the realized one-step AdamW change $\Delta\rho$. A rank-128 randomized SVD of the linearized map defines a resolved band of singular directions with $\sigma/\sigma_{\max} \geq 10^{-2}$. Because the SVD is truncated at rank 128, we call the complement of this band the unresolved residual subspace; it is not a certified null space.

We evaluate the data-fidelity term and the full objective separately. With the data-fidelity term alone, the raw gradient places 99.9914–99.9966% of its energy in the resolved rank-128 band. After equal-norm comparison, $-J_\theta J_\theta^\top g$ places 5.95–10.43% in the unresolved residual and $\Delta\rho$ places 5.82–10.24%. The attenuation ratios are $\left\|J_\theta J_\theta^\top g\right\| / \|g\| = 0.0276$–$0.0310$ and $\|\Delta\rho\| / \|g\| = 0.00523$–$0.0101$. Because the raw gradient of the data-fidelity term already lies almost entirely in the resolved band, the parameterization here redistributes a direction that is already resolved. Under the full objective of Eq. 6, which NeTMY optimizes, the filtering appears directly: the raw gradient places 16.4–38.3% of its energy in the unresolved residual, $-J_\theta J_\theta^\top g$ only 5.5–10.1%, and $\Delta\rho$ 7.0–9.6%, a 2–$5\times$ suppression of unresolved-residual energy. The diagnostic therefore shows objective-dependent structured filtering, consistent with Lemma 2.

F.8 Component Ablation: Per-Class and Hyperparameter Sweeps

The cumulative ablation of Table 3 aggregates over 512 samples. The largest per-class failure under any ablation is case *medium/medium*, where $-\ell_1$ raises density MSE by $\times \sim 2.7$ and degrades SSIM from 0.96 to 0.82. The smallest is case *few/close*, where most ablations leave density MSE essentially unchanged (because a single source is reconstructed correctly by every variant).

Hyperparameter sweeps. Table 13 reports three single-axis sweeps (learning rate, hidden width, PE octave count) on held-out samples. NeTMY is robust to learning rate and hidden width within roughly an order of magnitude of the defaults; PE octave count $K = 12$ is the smallest setting that recovers high-frequency sparse detail without divergent peaks.

Table 13: Hyperparameter sweeps over a 64-sample held-out set. Best per axis in bold.

Axis	Setting				
Learning rate η_0	10^{-4}	5×10^{-4}	$\mathbf{10^{-3}}$	5×10^{-3}	10^{-2}
density MSE	0.0049	0.0040	**0.0037**	0.0038	0.0067
Hidden width	128	256	**320**	384	512
density MSE	0.0046	0.0040	**0.0037**	0.0038	0.0040
PE octaves K	4	8	**12**	16	20
density MSE	0.0083	0.0048	**0.0037**	0.0040	0.0046

F.9 Supervised Baselines and Distribution Shift

Setup. We train four supervised baselines (GAN-SL+physics, U-Net-SL, U-Net-SL+physics, HybridNeTMY) on dense scenes (300 paired samples drawn from a dense-source generator with 30–60 sources per scene) and evaluate on the sparse test scenes used by NeTMY. The dense-to-sparse shift models the label-scarce setting, in which simulated dense scenes are the available training data: training scenes have nearly continuous mass distributions while test scenes have 1–8 off-center delta-like sources.

Failure observation. Across the four supervised baselines, the dense-to-sparse evaluation produces three failure patterns. **GAN-SL+physics** concentrates its prediction on a few dominant responses and misses weaker peaks. **U-Net-SL+physics** produces low-amplitude diffuse predictions with residual background activation. **U-Net-SL** produces exaggerated peak amplitudes, spurious activations, and broad background artifacts. The **HybridNeTMY** model, despite its coordinate-feature input, behaves

like a supervised CNN under the dense-trained regime: its density MSE on the matched-operator $\mathcal{F}_2/\mathcal{F}_2$ benchmark is 1.011 (Table 8), the worst entry in that table. NeTMY is unaffected because it is optimized per measurement without paired density labels.

What the failure shows. The supervised distribution-shift failure shows that coordinate features alone do not produce NeTMY's behavior in this setting: HybridNeTMY shares the coordinate-feature input but, when trained supervised on dense scenes, behaves like the other supervised baselines. The advantage of NeTMY is in the per-measurement optimization-against-physics protocol of Section 4, not in the feature space alone.

F.10 Real-Data Cross-Check: Full Four-Step Protocol

The cross-check of Section 5.5 evaluates four quantities.

Step 1: SRIM-anchored amplitude calibration. For each NV $i \in \{2, 3, 4, 6, 7, 8, 9, 10\}$, we compute the implied NV depth $d_i^\star$ under each operator by inverting $\Gamma_{\mathrm{Ru}}^{\mathrm{pred}}(d_i^\star) = \Gamma_{\mathrm{Ru}}^{i,\mathrm{meas}}$. Under $\mathcal{F}_2$, 1 NV (NV4) lands in $[9, 20]$ nm at $d_4^\star = 14.49$ nm; under $\mathcal{F}_1$, no NVs land in the window because $\Gamma_{\mathrm{Ru}}^{\mathcal{F}_1}(d) > \Gamma_{\mathrm{Ru}}^{\mathrm{meas}}$ for all $d > 9$ nm at the SRIM-prior amplitude. The point amplitude ratio at $d = 14.5$ nm is $\Gamma^{\mathcal{F}_1}/\Gamma^{\mathcal{F}_2} = 11{,}353.5/54.0 \approx 210\times$.

Step 2: Power-law analysis. The independent quantity $\Gamma_{\mathrm{int}} \propto d^{-\alpha}$ with $\alpha = 1.06 \pm 0.22$ provided by Kumar et al. (2024) lets us eliminate d and convert each operator's depth scaling into an internal-rate scaling: $\mathcal{F}_2$ predicts $b_{\mathrm{pred}} = 4/\alpha = 3.77$, $\mathcal{F}_1$ predicts $b_{\mathrm{pred}} = 2/\alpha = 1.89$. The measured slope is $b_{\mathrm{exp}} = 1.43 \pm 0.21$. Slope alone is closer to the $\mathcal{F}_1$ prediction; we therefore explicitly report this axis as ambiguous, since slope and intercept are independent.

Step 3: Spectral calibration. Magnetic-field-dependent T_1 measurements at $B \in \{2, 313, 383.5\}$ Gauss (single NV) calibrate the spin-bath correlation time $\tau_c = 0.41$ ns and bath linewidth $\gamma_{\mathrm{bath}} = 2.45$ GHz. Both operators share the Lorentzian spectral response, so this calibration is operator-neutral; it fixes τ_c and γ_{bath} for both operators.

Step 4: Loss-landscape identifiability. On the Gaussian density ansatz $\rho(\mathbf{r}; A, \sigma_g) = A \exp(-\|\mathbf{r}\|^2 / 2\sigma_g^2)$, where σ_g denotes the Gaussian width, we evaluate the log-MSE loss $L(A, \sigma_g) = |N_{\mathrm{NV}}|^{-1} \sum_i (\log \Gamma_{\mathrm{Ru}}^{\mathrm{pred}}(A, \sigma_g, d_i) - \log \Gamma_{\mathrm{Ru}}^{i,\mathrm{meas}})^2$ on a 90×90 grid in $(\log_{10} A, \sigma_g)$. The Hessian condition number at the minimum is $\kappa_{\mathcal{F}_2} = 931$ and $\kappa_{\mathcal{F}_1} = 301{,}139$ (ratio $\sim 323\times$). The $\mathcal{F}_2$ landscape is a parabolic bowl; the $\mathcal{F}_1$ landscape is a degenerate valley along $A^2\sigma_g^2 = \mathrm{const}$, which is the algebraic consequence of $\Gamma \propto A^2$ in $\mathcal{F}_1$ and is non-identifiable on this dataset.

Scope. Two of three axes (depth-amplitude, conditioning) disfavor $\mathcal{F}_1$; the third (slope) is ambiguous. The data comprise 8 NVs with one depth coordinate each, and the density enters through a Gaussian ansatz, so the check compares the two operators on these measurements and does not involve a spatial density map.

F.11 Runtime

NeTMY's per-instance wall-clock on the 512-sample matched-operator benchmark on a single NVIDIA A6000 GPU is long-tailed: median 273 s, trimmed mean (drop top 5 samples) 426 s, untrimmed mean 780 s. The top-5 samples are dense *many/close* configurations where the 64×64 stage requires more iterations to converge on the full source set. Free-density baselines (Tikhonov 1.55 s, ADMM 1.10 s) are roughly 200–700$\times$ faster on the same hardware; DeepDecoder (259 s) is within an order of magnitude of NeTMY. In the regime that motivated this work (per-instance reconstruction without paired density labels), the relevant comparison is to the supervised amortized baselines, which are fast at inference but require a labeled training set that is generally not available for NV-relaxometry samples; in regimes where labels are abundant, amortized methods are the appropriate choice.

Total-cost accounting on matched hardware. The timings above were measured on an A6000. Table 14 reports a separate timing set in which data generation, amortized training and inference, and Tikhonov, ADMM, and NeTMY are timed on one NVIDIA L40 over a stratified 32-scene set. Including generation of all 300 $\mathcal{F}_3$ training pairs and training, the one-time cost of the U-Net is 368.989 s, after which inference costs 0.001487 s per query. The compute-only crossover occurs after 102.57, 65.48, and 2.61 queries versus Tikhonov, ADMM, and NeTMY, respectively. These

crossovers count compute on matched hardware only; accuracy is governed by training data, and supervised amortized models degrade under the dense-to-sparse shift of Appendix F.9. Matched paired data are scarce in this setting: the ground-truth spin-source density is never directly observed, NV depth requires an independent SRIM implantation prior (Section 5.5), the real dataset provides 8 NV sites in total (Kumar et al., 2024), and relaxometry spectra are acquired site by site with extensive averaging (Mzyk et al., 2022). NeTMY therefore targets offline scientific reconstruction, in which each spectrum follows a lengthy physical acquisition. With abundant matched pairs and many queries, amortized inference is cheaper, and Table 14 gives the query count at which that happens.

Table 14: Measured median wall time per component on one NVIDIA L40 over a stratified 32-scene set. U-Net training is paid once; all other entries are per scene.

Component	Measured median wall time
$\mathcal{F}_2$ generation	0.000542 s
$\mathcal{F}_3$ generation	0.013664 s
U-Net inference	0.001487 s
Tikhonov solve	3.599 s
ADMM solve	5.636 s
NeTMY solve	141.342 s
U-Net training, 300 scenes $\times$ 200 epochs	364.890 s once

F.12 Failure Modes

Three failure modes remain.

Cross-shaped artifacts on dense scenes. As shown in Figure 8, on samples in the *many/medium* class with $\sim$30–40 sources spaced at the resolution limit, NeTMY occasionally produces axis-aligned cross-shaped artifacts in the reconstructed density. The artifact morphology is consistent with source merging, which permits low-cost axis-aligned explanations (Section 3.3): when many sources lie within w_{psf} of each other, the optimizer converges to a low-cost density that places mass along low-cost axis-aligned tracks rather than at the true source positions. Anti-artifact regularizers (isotropic TV, Laplacian, Fourier-axis penalties) reduce visible cross artifacts but slightly worsen aggregate density MSE; we keep the standard anisotropic TV in the main runs.

High-frequency leakage on small clusters. The medium/medium row of Figure 9 shows that NeTMY occasionally leaks small-cluster mass into nearby pixels. This is the high-frequency tail of the annealed PE schedule: at full $\beta = K$ the network can fit non-source-supported pixels, and the ℓ_1/gate combination suppresses this leakage. On samples with $\sim$8–15 closely spaced sources the leakage is visible; on simpler sparse scenes it is suppressed.

Centered minimum trapping under random initialization. For $\sim 5\%$ of the matched-operator samples, NeTMY's stage-1 minimum is in the centered-collapse basin documented in Section 5.3. Stage-2 refinement at higher resolution and lower learning rate escapes this basin via the β-schedule reset, but the fraction of samples in which this happens is sensitive to the random initialization of the MLP and to stage-1 epoch count. We did not observe any systematic correlation with scene class.

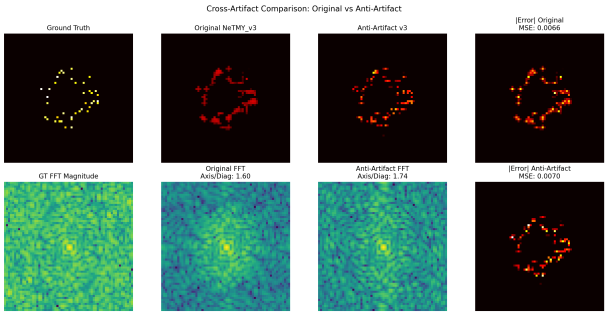


Figure 8: The reconstruction captures the support of the active region but has cross-like, axis-aligned artifacts around high-intensity locations; the anti-artifact variant reduces them but increases density MSE.

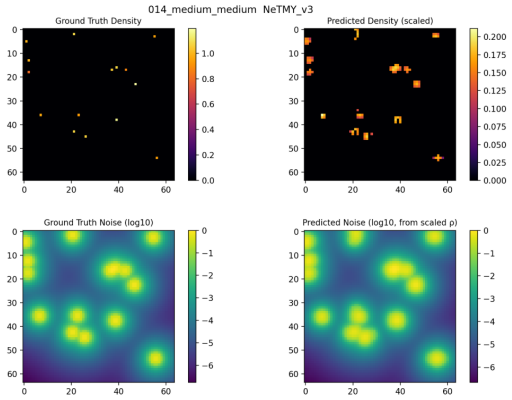


Figure 9: High-frequency leakage on a medium/medium sample. The predicted density spreads small-cluster mass into neighboring pixels beyond the true source locations, a leakage artifact driven by the high-frequency tail of the annealed PE schedule. The ℓ_1/gate combination partially suppresses this effect, but residual leakage remains visible given the $\sim$8–15 closely spaced sources in this scene.

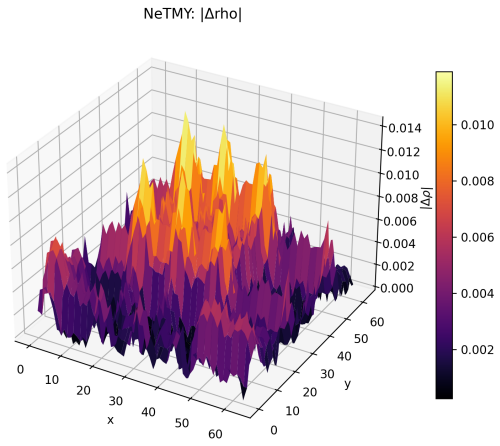


Figure 10: Realized first-step image-space update $|\Delta\rho|$ produced by NeTMY, where $\Delta\rho = \rho_1 - \rho_0$. The surface is spatially distributed, with a center-to-outer ratio of approximately $1.6\times$ against $18.29\times$ for the iter-0 free-density gradient under the same forward operator (Section 5.3). This is the empirical signature of the Lemma 2 filtering kernel $G_\theta = J_\theta J_\theta^\top$ redistributing the raw density-space gradient through the parameterization.

F.13 Qualitative Examples

This subsection presents additional per-sample reconstruction grids that complement the cross-fidelity benchmark of Section 5.2 and the matched-operator benchmark of Appendix F.4. Figures 11–14 show four cross-fidelity samples ($\mathcal{F}_3$-generated, inverted under both $\mathcal{F}_1$ and $\mathcal{F}_2$) and illustrate the operator-fidelity reshaping discussed in Section 5.2. Figures 15 and 16 show four matched-operator samples each under $\mathcal{F}_1/\mathcal{F}_1$ and $\mathcal{F}_2/\mathcal{F}_2$, respectively, drawn from the matched-operator benchmark of Appendix F.4. Within each figure, an observation panel is followed by (left to right) the ground-truth density, Tikhonov, ADMM, HybridNeTMY, and NeTMY reconstructions; the centered iter-0 collapse of free-density solvers under $\mathcal{F}_2$ is visible as a concentrated central blob in the Tikhonov / ADMM panels, while NeTMY recovers off-center sparse structure consistent with the ground truth.

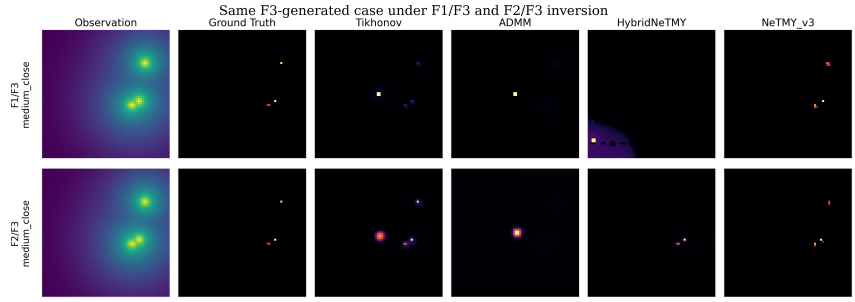


Figure 11: Cross-fidelity sample 008 ($\mathcal{F}_3$-generated). Reconstructions under $\mathcal{F}_1$ and $\mathcal{F}_2$ inversion. Free-density solvers (Tikhonov, ADMM) center-collapse under the more faithful $\mathcal{F}_2$, while NeTMY recovers the off-center support. See Section 5.2 for aggregate metrics.

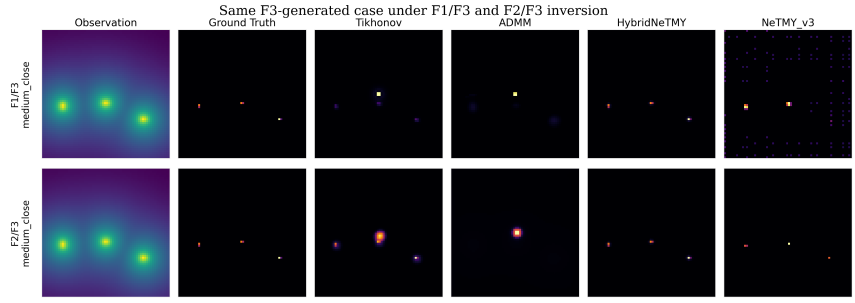


Figure 12: Cross-fidelity sample 009 ($\mathcal{F}_3$-generated). Same layout as Figure 11: $\mathcal{F}_1$ inversion (top row) vs. $\mathcal{F}_2$ inversion (bottom row). The operator-dependent difference between $\mathcal{F}_1$ and $\mathcal{F}_2$ inversion discussed in Section 5.2 is visible at the per-sample level.

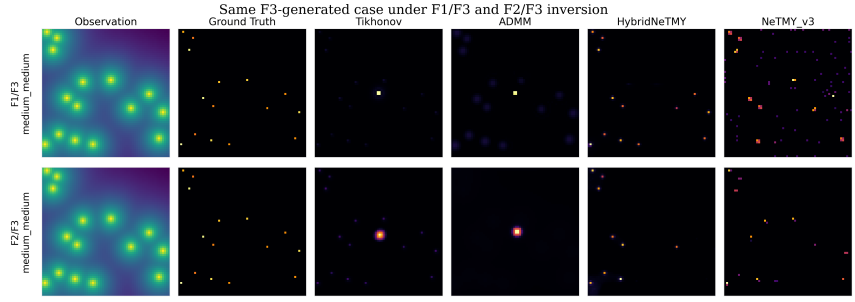


Figure 13: Cross-fidelity sample 010 ($\mathcal{F}_3$-generated). A *many/close* scene where source merging and the center collapse driven by center bias and peak coupling jointly stress the free-density solvers; NeTMY preserves the most peaks within the matching radius of Appendix F.3.

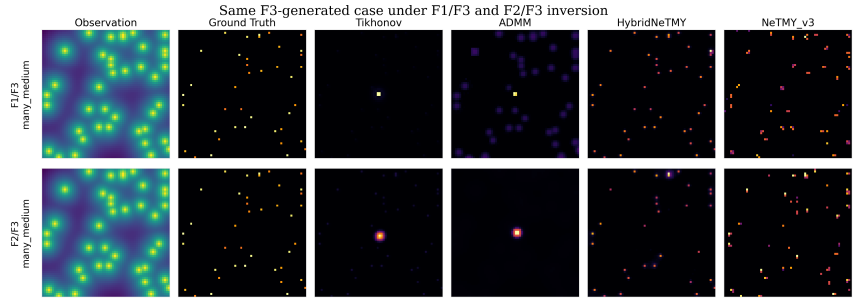


Figure 14: Cross-fidelity sample 018 ($\mathcal{F}_3$-generated). A *medium/medium* scene; Tikhonov and ADMM concentrate mass in a central peak under both inversions, while NeTMY distributes mass over the scattered off-center support, consistent with the table-level rankings of Section 5.2.

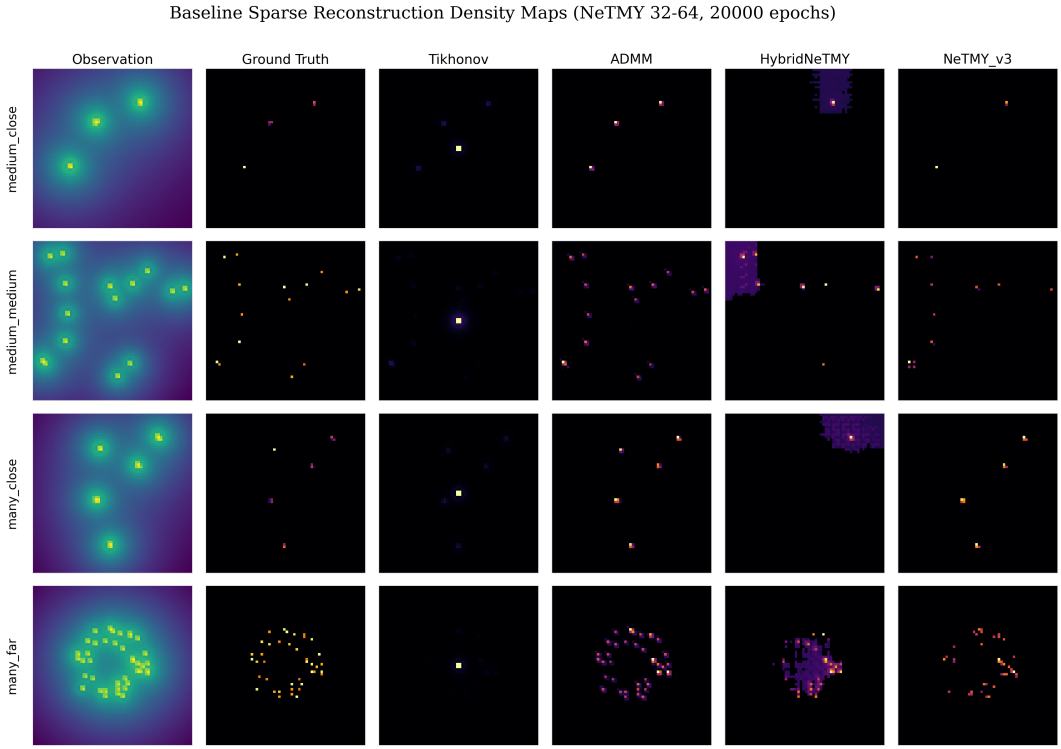


Figure 15: Matched-operator $\mathcal{F}_1/\mathcal{F}_1$ examples (samples 012, 016, 022, 029). Under the simplified scalar/coherent operator center bias does not induce collapse, and ADMM is the lowest-MSE method (Table 2); NeTMY produces cleaner peaks and the second-lowest MSE (0.009 vs 0.006 for ADMM).

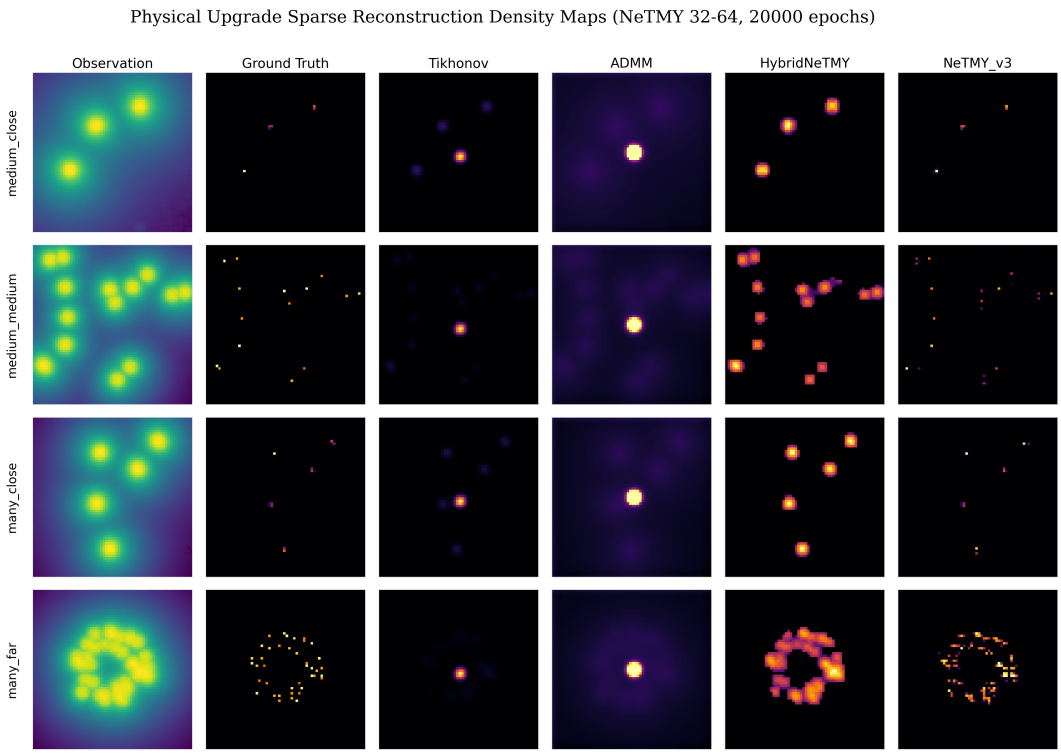


Figure 16: Matched-operator $\mathcal{F}_2/\mathcal{F}_2$ examples (samples 012, 016, 022, 029). Under the tensor power-summed operator the same ADMM run degrades to a centered collapse, while NeTMY produces the cleanest off-center reconstructions and tops every $\mathcal{F}_2/\mathcal{F}_2$ metric in Table 2.