# FOURIER-DUAL GAUSSIAN REPRESENTATIONS FOR 3D NON-CARTESIAN PARALLEL MRI RECONSTRUCTION

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Gaussian primitives are an attractive continuous representation for MRI reconstruction because their Fourier transform is available in closed form, enabling fitting directly to the measured spatial-frequency ($k$-space) samples. However, existing Gaussian-based MRI reconstruction methods do not fully exploit this: they either rasterize an image inside the forward model or use parallel imaging formulation whose cost grows with the number of coils. They also parameterize each primitive by its amplitude, which couples its integral to its spatial extent and degrades the conditioning of the fit in $k$-space.

We propose a Fourier-dual representation for 3D non-Cartesian parallel MRI, where optimization never rasterizes the image: (i) Integral-Parameterized Gaussians (IPG) decouple each primitive's integral from its scale: for a single primitive, we prove that this lowers the best condition number reachable by diagonal preconditioning, most strongly for center-weighted trajectories; (ii) coil sensitivities are expressed in a compact Fourier basis for direct measurement at arbitrary $k$-space locations from a single shared Gaussian representation; (iii) A continuous total variation prior is estimated by Monte Carlo sampling of analytical spatial gradients.

On 19 simulated 3D brain volumes, the method obtains the highest PSNR and SSIM against compressed sensing, a scan-specific implicit neural representation (INR) and parameter-matched Gaussian baselines. A 3T phantom acquisition confirms the lowest held-out $k$-space error.

## 1 INTRODUCTION

### 1.1 MR IMAGE RECONSTRUCTION

The MRI acquisition process can be modeled as follows:

$$y_{c,m} = \int_{\mathbb{R}^3} s_c(\boldsymbol{r})\rho(\boldsymbol{r})e^{-i2\pi \boldsymbol{k}_m^\top \boldsymbol{r}}\,\mathrm{d}\boldsymbol{r} + \varepsilon_{c,m}, \qquad m = 1,\dots,M, \tag{1}$$

where $\rho : \mathbb{R}^3 \to \mathbb{C}$ denotes the continuous complex-valued object to be reconstructed and $s_c : \mathbb{R}^3 \to \mathbb{C}$ the receive sensitivity of coil $c \in \{1,\dots,C\}$. $\boldsymbol{r}$ and $\boldsymbol{k}$ are the normalized spatial and spectral coordinates, respectively. During an MRI exam, we sample the coordinates $\boldsymbol{K} = (\boldsymbol{k}_m)_{1\dots M}$ and get the k-space measurements $\boldsymbol{y} = (y_{c,m})_{c,m} \in \mathbb{C}^{C\times M}$, corrupted by additive noise $\boldsymbol{\varepsilon} = (\varepsilon_{c,m})_{c,m}$.

Non-Cartesian sampling patterns allow for more degrees of freedom in undersampling of $k$-space, and combined with suitable reconstruction algorithms, maintain image clarity even at high acceleration factors. More precisely, (1) is usually discretized and solved on a finite grid using a nonuniform fast Fourier transform (NUFFT, Fessler & Sutton (2003); Comby et al. (2025)), optimizing the following loss in the image domain:

$$\widehat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}\in\mathbb{C}^N} \frac{1}{2}\sum_{c=1}^{C} \left\| \mathcal{F}_{\boldsymbol{K}} S_c \boldsymbol{\rho} - \boldsymbol{y_c} \right\|_2^2 + g_\lambda(\boldsymbol{\rho}), \tag{2}$$



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.

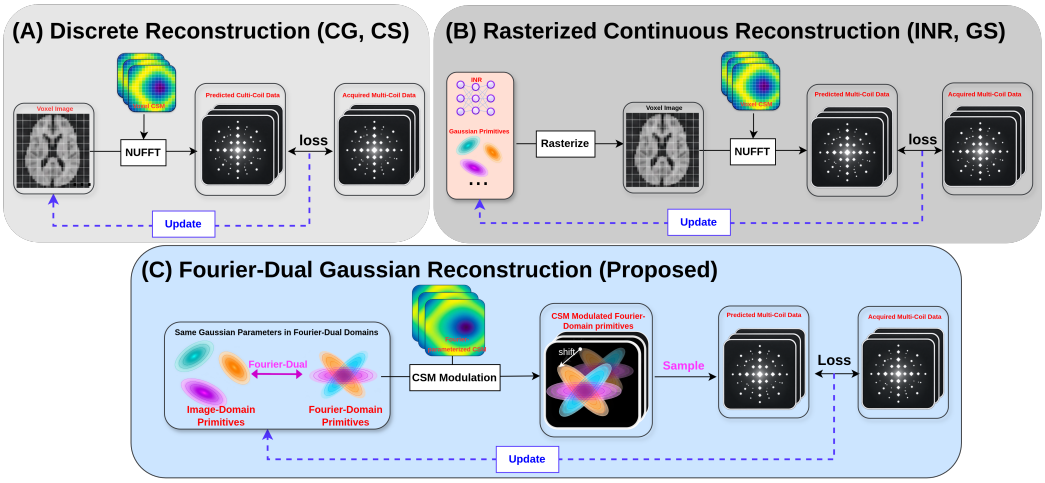



Figure 1: Overview of three reconstruction pathways. (A) Voxel-based methods optimize a discrete image and evaluate the sampled multi-coil Fourier operator. (B) Typical continuous-representation pipeline (INR, GS) first rasterizes onto a voxel grid before Fourier-domain data consistency. (C) The proposed Fourier-dual formulation evaluates analytically in the Fourier domain. Coils sensitivities act as spectral shifts, and the signal is evaluated only at the acquired coordinates. Gradients propagate directly to the continuous primitive parameters without an intermediate rasterization step.

with $g_\lambda$ a regularization term parameterized by $\lambda$. Eq. (2) ties the reconstruction to a fixed grid, setting both the resolution and the number of unknowns, and only solve an discrete approximation of Eq. (1).

### 1.2 Continuous Representation and Related Work

An alternative is to parameterize the image directly as a continuous function. Implicit neural representations (INRs) use a coordinate-based neural network for this purpose (Müller et al., 2022; Shen et al., 2021; Li et al., 2026),and can estimate coil sensitivities jointly (Feng et al., 2024) but their MRI forward model still requires a numerical Fourier transform. Gaussian representations provide an explicit alternative (Kerbl et al., 2023), and has already been adapted to CT (Zha et al., 2024) and MRI (Peng et al., 2025), It can be traced back to classical blob bases(Lewitt, 1990) and off-the-grid reconstruction(Ongie & Jacob, 2016). An image is a collection of Gaussian primitives with learnable centers, covariances, orientations, and complex coefficients. Its Fourier transform is closed form: the covariance sets a Gaussian spectral envelope, and the center act as a complex phasor. The data consistency term in Eq. (2) can then be evaluated directly at the measured $k$-space samples without discretization.

Analytical parallel imaging has also been explored, but with different trade-offs. MR-BLOB (Singh et al., 2024) models coil sensitivities with isotropic Gaussians at the cost of a restrictive sensitivity representation and increasing complexity as more components are used. Back-to-Basis (B2B) (Singh et al., 2026) instead assigns coil-specific complex coefficients to shared Gaussian geometry, but parameter count grows with the number of coils.

Low-order spectral coil models are classical in parallel imaging, e.g. Sobolev-weighted Fourier coefficients in NLINV (Uecker et al., 2008) and sinusoids in analytical phantoms (Guerquin-Kern et al., 2012); we use such a model to keep Fourier evaluation analytical without duplicating object-side parameters across coils.

### 1.3 Contributions and Paper Organization

We therefore formulate the multi-coil image as a Fourier-dual continuous representation, combining integral-parameterized Gaussian primitives for the object with a compact Fourier representation of the coil sensitivities. This preserves closed-form evaluation from the continuous object to

the acquired non-Cartesian multi-coil measurements, while continuous spatial regularization avoids reintroducing a voxel grid during optimization. Figure 1 contrasts the resulting reconstruction path with conventional discrete reconstruction and continuous representations that are rasterized before Fourier-domain data consistency. The resulting framework has three main components.

**Integral-parameterized Gaussians.** Gaussian primitives can be parameterized in several equivalent ways. In Section 2.1, we recall the conventional parameterization by peak amplitude (APG), and instead propose to parameterize by integral (IPG). We also prove both theoretically and in simulation that for a single primitive under the Gauss-Newton approximation, IPG lowers by an order magnitude the best condition number reachable by diagonal preconditioning for center-weighted trajectories (Section 2.2, Appendix E).

**Fourier sensitivity encoding.** To preserve analytical Fourier evaluation under parallel imaging, we represent each coil sensitivity using a compact low-frequency Fourier expansion, following the sinusoidal sensitivity model of Guerquin-Kern et al. (2012). Sensitivity modulation then becomes a set of spectral shifts of the same analytical object representation, allowing Gaussian evaluations to be shared across coils. The resulting multi-coil forward model is derived in Section 2.3.

**Continuous spatial regularization.** Spatial priors such as total variation (TV) are usually evaluated on voxel grids, which would reintroduce rasterization during optimization. Because the Gaussian representation has analytical spatial derivatives, in Section 2.4, we instead compute an unbiased Monte-Carlo estimator of the continuous TV. In our setting it is $2.8\times$ faster than rasterized TV, with a slightly higher PSNR.

Finally, Section 3 compares against variational, INR, and Gaussian-based reconstruction methods, and decompose the gain over each contribution with single-factor ablations. Further results are in Appendices G and H.

## 2 Methods

### 2.1 Fourier-Dual Integral-Parameterized Gaussians

We parameterize $\rho(\boldsymbol{r})$ as a sum of Gaussian primitives $\rho(\boldsymbol{r}) = \sum_{n=1}^{N_g} \psi_n(\boldsymbol{r})$, with

$$\psi_n(\boldsymbol{r}) = \psi(a_n, \boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n, \boldsymbol{r}) = a_n \exp\left[-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{\mu}_n)^\top \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{r}-\boldsymbol{\mu}_n)\right] \qquad \text{(APG-image)}$$

Hereafter we refer to this formulation as *Amplitude Parameterized Gaussian* (APG). Following standard Gaussian splatting practice (Kerbl et al., 2023), we further parameterize the covariance matrix by its principal standard deviations and its orientation using quaternions:

$$\boldsymbol{\Sigma}_n = \boldsymbol{R}(\boldsymbol{q}_n)\,\mathrm{diag}\big(\sigma_{x,n}^2, \sigma_{y,n}^2, \sigma_{z,n}^2\big)\,\boldsymbol{R}(\boldsymbol{q}_n)^\top, \tag{3}$$

where $\boldsymbol{q}_n = \frac{\widetilde{\boldsymbol{q}}_n}{\|\widetilde{\boldsymbol{q}}_n\|}$ is the unit quaternion obtained from the trainable quaternion $\widetilde{\boldsymbol{q}}_n$. The principal standard deviations $(\sigma_{d,n})_{d=x,y,z}$ are parameterized as $\sigma_{d,n} = \exp(\zeta_{d,n})$ with the constraint $\zeta_{d,n} \in [\log(\sigma_{\min}), \log(\sigma_{\max})]$.

Overall, APG optimizes the Gaussian mixture parameters $\Theta = (\theta_1, \ldots, \theta_{N_g})$ with $\theta_n = (a_n, \boldsymbol{\zeta}_n, \widetilde{\boldsymbol{q}}_n, \boldsymbol{\mu}_n)$, and its primitives admits the following Fourier transform

$$\widehat{\psi}_n(\boldsymbol{k}) = a_n \underbrace{(2\pi)^{\frac{3}{2}}|\boldsymbol{\Sigma}_n|^{\frac{1}{2}}}_{Z_n} \exp\left[-i2\pi\boldsymbol{k}^\top\boldsymbol{\mu}_n\right] \exp\left[-2\pi^2\boldsymbol{k}^\top\boldsymbol{\Sigma}_n\boldsymbol{k}\right] \qquad \text{(APG-kspace)}$$

In APG setup, the zero-frequency value $\hat{\psi}_n(0) = a_n Z_n$, depends jointly on its amplitude and covariance (Fig 2, E). Section 2.2 shows that this coupling degrades the conditioning of a $k$-space fit. Instead, prior Gaussian MRI methods fits a rasterized image (Peng et al., 2025).

Instead, we propose to "flip the duality" and normalize the image space amplitude by the integral of the unit primitive, which depends solely on the covariance. We called this the *Integral parameterized*

*Gaussian* (IPG), with parameters $\theta_n = (w_n, \boldsymbol{\zeta}_n, \tilde{\boldsymbol{q}}_n, \boldsymbol{\mu})$, where $w_n$ is then the integral (DC value) of $\phi_n$, independent of $\boldsymbol{\Sigma}_n$: $\widehat{\phi}_n(0) = w_n$. Figure 2 illustrates the two parameterizations in 1D.

$$\phi_n(\boldsymbol{r}) = \frac{w_n}{Z_n} \exp\left[-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{\mu}_n)^\top \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{r}-\boldsymbol{\mu}_n)\right], \qquad \text{(IPG-image)}$$

$$\widehat{\phi}_n(\boldsymbol{k}) = w_n \exp\left[-i2\pi \boldsymbol{k}^\top \boldsymbol{\mu}_n\right] \exp\left[-2\pi^2 \boldsymbol{k}^\top \boldsymbol{\Sigma}_n \boldsymbol{k}\right]. \qquad \text{(IPG-kspace)}$$

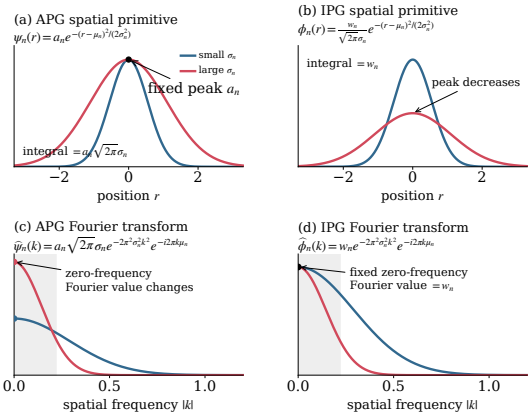



Figure 2: One-dimensional comparison of APG and IPG under two spatial scales. For APG, changing scale at fixed peak coefficient changes both the spatial integral and zero frequency Fourier value. For IPG, changing scale at fixed integral preserves the zero frequency value while modifying the spatial peak and spectral bandwidth.

### 2.2 Conditioning of APG and IPG formulation

APG and IPG describe the same images; they differ only in the trainable amplitude, the peak $a$ or the integral $w = aZ$. For a single isotropic primitive and the Gauss–Newton model of an unweighted $\ell_2$ $k$-space loss (assumptions A1–A3, Appendix E), this choice only changes the coupling between the primitive's log-magnitude and its log-scale. Table 1 reports this coupling and the smallest condition number $\kappa^\star$ reachable by any diagonal preconditioner, such as Adam's. IPG is never worse than APG, and an order of magnitude better for center-weighted trajectories such as 3D radial. This is a per-primitive result: interactions between primitives affect both parameterizations alike and are not captured (Appendix E.4).

Table 1: Energy–scale correlation $|c|$ and optimal diagonally-scaled condition number $\kappa^\star = (1+|c|)/(1-|c|)$ of the energy–scale block (Proposition 2, Appendix E) for IPG and APG. Model for a sampling density $\|\boldsymbol{k}\|^{-p}$ as a function of the product $\sigma k_{\max}$, When $\sigma k_{\max} \geq 1/2$, the primitive envelop lies well inside the acquired frequencies. Lower is better; $\kappa^\star = 1$ means complete decoupling.

| | $\lvert c\rvert$, $\sigma k_{max} \geq 1/2$ | | $\kappa^\star$, $\sigma k_{\max} \geq 1/2$ | | $\kappa^\star$, $\sigma k_{\max} = 1/4$ | |
|---|---|---|---|---|---|---|
| Sampling density | IPG | APG | IPG | APG | IPG | APG |
| Uniform ($p = 0$) | 0.78 | 0.78 | 7.9 | 7.9 | **12.6** | 39.0 |
| Variable density ($p = 1$) | **0.71** | 0.89 | **5.8** | 17.9 | **8.1** | 53.8 |
| 3D radial / cones ($p = 2$, used here) | **0.58** | 0.96 | **3.7** | 52.0 | **4.5** | 102 |

### 2.3 Analytical Fourier Coil-Sensitivity Encoding

Spatially smooth coil sensitivities motivate compact spectral models and local convolutional relations in multi-coil $k$-space, as exploited by GRAPPA and SPIRiT (Griswold et al., 2002; Lustig & Pauly, 2010). Following the sinusoidal coil sensitivity model of Guerquin-Kern et al. (2012), we represent each coil sensitivity in a shared set of $L$ spatial frequencies, where $L$ denotes the total number of basis functions per coil,

$$s_c(\boldsymbol{r}) = \sum_{\ell=1}^{L} \beta_{c,\ell}\, e^{\,i2\pi \boldsymbol{k}_\ell^\top \boldsymbol{r}}, \qquad (4)$$

where $\boldsymbol{k}_\ell$ is the shared frequency support across coils and $\beta_{c,\ell} \in \mathbb{C}$ is coil specific. Substituting Eq. (4) into Eq. (1) and applying the Fourier modulation identity $\mathcal{F}\left\{e^{i2\pi \boldsymbol{k}'^\top \boldsymbol{r}}\rho(\boldsymbol{r})\right\}(\boldsymbol{k}) = \widehat{\rho}(\boldsymbol{k}-\boldsymbol{k}')$ gives

$$y_{c,m}^{\text{pred}}(\Theta) = \sum_{\ell=1}^{L} \beta_{c,\ell}\widehat{\rho}_\Theta(\boldsymbol{k}_m - \boldsymbol{k}_\ell)\,. \tag{5}$$

For the Fourier sensitivity model, the multi-coil operator requires evaluating Eq. (IPG-kspace) at the shifted coordinates $\widetilde{\boldsymbol{k}}_{\ell m} = \boldsymbol{k}_m - \boldsymbol{k}_\ell$.These coordinates depend only on the acquisition trajectory and the fixed sensitivity-frequency dictionary and are therefore reused throughout optimization. The object spectrum is evaluated once per shifted coordinate and shared by all coils, costing $\mathcal{O}(LM(N_g + C))$ instead of $\mathcal{O}(LMN_gC)$ for a per-coil evaluation. (Appendix H).

We estimate coil-sensitivity maps $S_c^{\text{ref}}$ using ESPIRiT (Uecker et al., 2014) and fit their coefficients on a uniform low-frequency Fourier lattice:

$$\forall c = 1, ..., C \quad \boldsymbol{\beta}_c = \arg\min_{\boldsymbol{\beta}\in\mathbb{C}^L} \|\boldsymbol{\beta} - \boldsymbol{P}_L\boldsymbol{F}S_c^{\text{ref}}\|_2^2 + \lambda_\beta\|\boldsymbol{\beta}\|_2^2 = \frac{\boldsymbol{P}_L\boldsymbol{F}S_c^{\text{ref}}}{1+\lambda_\beta} \tag{6}$$

Where $F$ if the Cartesian FFT matrix, and $\boldsymbol{P}_L$ select $L$ spatial frequencies via masking. In practice we used a spherical mask with $L = 125$ and $\lambda_\beta = 10^{-3}$.

### 2.4 Continuous Reconstruction Objective

We optimize the Gaussian parameters with an objective combining data in $k$-space and a continuous image-domain regularizer, evaluated from analytical spatial derivatives of the IPG representation: $\mathcal{L}(\Theta) = \mathcal{L}_{\text{DC}}(\Theta) + \lambda_{\text{TV}}\mathcal{R}_{\text{TV}}^{\text{MC}}(\Theta)$.

**Data consistency.** Let $r_{c,m}(\Theta) = y_{c,m}^{\text{pred}}(\Theta) - y_{c,m}$ denote the complex residual for coil $c$ and sample $m$. We apply the Huber penalty $\psi_\delta$ to the residual magnitude, with an adapted threshold for the residual scale at each optimization step, and normalize by the measurement magnitude:

$$\mathcal{L}_{\text{DC}}(\Theta) = \frac{\sum_{c,m}\psi_\delta\left(|r_{c,m}(\Theta)|\right)}{\delta\sum_{c,m}|y_{c,m}|}, \qquad \delta = \alpha_{\text{H}}\ \text{sg}[\max(\text{median}_{c,m}\,|r_{c,m}|, \epsilon_{\text{H}})]\,, \tag{7}$$

where $\alpha_{\text{H}} = 1$ is the Huber threshold factor, $\epsilon_{\text{H}} = 10^{-12}$, and $\text{sg}[\cdot]$ denotes stop-gradient, so that the threshold follows the residual scale but is constant during differentiation. Because $\delta$ tracks the median residual, a fixed fraction of residuals stays in the linear Huber regime throughout optimization.

**Regularization using Monte Carlo Total Variation.** The continuous TV of a Gaussian mixture has no closed form, but its spatial gradient does (Eq. (24), Appendix D), so we estimate the TV by Monte Carlo sampling without finite differences or rasterization, with $\|\nabla_{\boldsymbol{r}}\rho\|_2^2 = \sum_d |\partial_d\rho|^2$ for complex $\rho$:

$$\mathcal{R}_{\text{TV}}^{\text{MC}}(\Theta) = \frac{1}{N_{\text{TV}}}\sum_{j=1}^{N_{\text{TV}}}\sqrt{\|\nabla_{\boldsymbol{r}}\rho_\Theta(\boldsymbol{r}_j)\|_2^2 + \epsilon_{\text{TV}}^2} \to \mathcal{R}_{\text{TV}}(\Theta) = \frac{1}{|\Omega|}\int_\Omega\sqrt{\|\nabla_{\boldsymbol{r}}\rho_\Theta(\boldsymbol{r})\|_2^2 + \epsilon_{\text{TV}}^2}\,\mathrm{d}\boldsymbol{r} \tag{8}$$

with $\boldsymbol{r}_j \sim \text{Unif}(\Omega)$ drawn independently at each step over the reconstruction domain $\Omega$, so $\mathbb{E}[\mathcal{R}_{\text{TV}}^{\text{MC}}(\Theta)] = \mathcal{R}_{\text{TV}}(\Theta)$. Since $\nabla_{\boldsymbol{r}}\rho_\Theta \neq 0$ almost everywhere and the norm is Lipschitz, the stochastic gradient is also unbiased. We use $N_{\text{TV}} = 8192$, the best runtime–accuracy compromise (Fig. 5b, Appendix H).

### 2.5 Optimization and Efficient Evaluation

**Staged optimization** From an early-stopped, grid-based conjugate-gradient (CG) image, we sample the primitive centers $\boldsymbol{\mu}$ and estimate the initial complex coefficients $w$ by ridge least squares. Then, we optimize the parameters in stages with Adam, solely through the continuous k-space model (IPG-kspace) (see Appendix C for the schedule and Table 4 for hyperparameter values).

**Efficient Fourier evaluation.** We evaluate the Gaussian mixture using fused CUDA kernels that stream over primitive and $k$-space tiles without materializing the $L \times M \times N_g$ interaction tensor. The fused backward pass computes vector–Jacobian products with respect to the complex coefficient, center, and covariance of each primitive; derivatives with respect to log-scale and orientations are then propagated through the differentiable covariance parameterization.

# 3 Experiments and Results

## 3.1 Experimental Setting

**Simulation data** We use 20 BrainWeb anatomies at $128^3$ resolution(Kwan et al., 1999), reserving one subject exclusively for development and hyperparameter selection, and evaluating all reported cohort results on the remaining 19 subjects. Measurements are generated analytically at the non-Cartesian coordinates using a 15-channel Biot–Savart coil sensitivities (Guerquin-Kern et al., 2012; Ngo et al., 2015). We evaluate 256, 1024, and 4096 in-out radial spokes of 128 samples each (nominal acceleration factors (AFs) of 64, 16, and 4 relative to the $128^3$ Cartesian grid). Complex Gaussian noise is added to the multi-coil samples at 50 and 100 dB SNR, relative to their mean power (Appendix F), so that the per-sample noise level, is the same for all spoke counts. Because radial sampling concentrates on the center of $k$-space, 50 dB on the radial samples corresponds to about 13 dB on the fully sampled Cartesian grid (Appendix F.3), and matches empirical noise level measurements on our 3T system. The 100 dB setting is effectively noiseless and isolates undersampling errors: it is not a physically realistic noise level.

**Experimental 3T validation** We scanned the NIST/ISMRM Phantom (Russek et al., 2021) on a 3T MR system (Magnetom Cima.X, Siemens Healthcare) at 2-mm isotropic resolution (matrix size $160^3$) with a Siemens 20-channel head/neck Coil. The protocol consisted of two radial gradient-recalled echo (GRE) scans (TR/TE/FA: 50 ms/13 ms/10°): A reference scan with 22940 radial spokes (19min7sec); then, a 4096-shot scan (3min24sec). From the 4096-spokes can, we retrospectively selected 1024 uniformly distributed spokes for reconstruction (AF=22 compared to the reference), and kept the other 3072 for held-out evaluation. Coil sensitivity maps were estimated from the 1024 spokes with ESPiRIT (Uecker et al., 2014). IPG gives the lowest held-out $k$-space error and, unlike SENSE–CG and CS–TV, no visible signal outside the phantom (Fig. 4). Because this error is dominated by low frequencies and passes through each method's own coil model, we use it as a consistency check rather than an image-quality metric.

**Baselines.** We compare our results against density-compensated adjoint NUFFT (iNUFFT), SENSE–CG, TV-regularized compressed sensing (CS–TV), a multiresolution-hash implicit neural representation (HashINR, Müller et al. (2022); Feng et al. (2025)), a "Raster Gaussian" that extends 3DGSMR (Peng et al., 2025) to non-Cartesian using NUFFT, with APG parameterization fixed number of primitives and IPG s optimization schedule, and Back to Basis (B2B, Singh et al. (2026)) in its published form, which does not use spatial regularization(B2B represents coil images rather than a single object and an object-domain TV is therefore not part of its formulation). Trainable representation baselines are matched to $\approx 96,000$ parameters (Appendix F.5 and Table 4 for hyper parameters values). We additionally report B2B with 8000 primitives (320,000 parameters) to separate parameter-budget matching from primitive-count matching.

**Model selection and metrics.** All tunable hyperparameters were selected using the development subject (s54) and then remained fixed for the 19 evaluation subjects. For each subject and acquisition setup, all methods received identical trajectories and noise realizations. Continuous reconstructions and the reference use matched voxel averaging (256 cubed to 128 cubed). Grid reconstructions retain their native grid. PSNR and SSIM use the complete 3D volume after one global complex-LS alignment. PSNR uses complex residuals; SSIM uses magnitudes. Cohort values are reported as mean $\pm$ sample standard deviation over the 19 evaluation subjects. Exact metric definitions and evaluation protocols are given in Appendix F.

Table 2: Reconstruction accuracy at 50 and 100 dB SNR for acceleration factors (AF) 64 and 16 with radial trajectories on simulation data. Values are mean ± standard deviation across the 19 evaluation subjects; higher values are better for both metrics. Best per condition in **bold**, second best underlined. All non-IPG methods differ significantly from IPG (Holm-adjusted two-sided paired Wilcoxon, $p < 0.01$). For each method, we report the number of parameters optimized to represent the complex-valued image; a fully parameterized image has $2 \times 128^3 = 4,194 \times 10^3$ parameters.

| | | SNR=50 dB | | | | SNR= 100dB | | | |
|---|---|---|---|---|---|---|---|---|---|
| | params | AF-64 | | AF-16 | | AF-64 | | AF-16 | |
| Method | # | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| iNUFFT | – | 17.21 ±0.40 | .147 ±.007 | 18.09 ±0.29 | .178 ±.009 | 17.22 ±0.40 | .148 ±.007 | 18.14 ±0.29 | .179 ±.009 |
| SENSE-CG | 4194k | 26.25 ±0.79 | .334 ±.016 | 30.08 ±0.56 | .400 ±.007 | 26.42 ±0.81 | .394 ±.025 | 31.24 ±0.62 | .612 ±.019 |
| CS-TV | 4194k | 28.38 ±0.74 | .598 ±.025 | 32.69 ±0.66 | .902 ±.011 | 28.66 ±0.76 | .691 ±.027 | 32.98 ±0.68 | .949 ±.009 |
| HashINR | 92.16k | 28.51 ±0.90 | .619 ±.057 | 31.12 ±0.63 | .930 ±.010 | 28.62 ±0.93 | .636 ±.063 | 31.29 ±0.66 | .939 ±.009 |
| Raster Gauss. | 96k | 33.32 ±0.57 | .977 ±.002 | 36.42 ±0.49 | .990 ±.001 | 34.10 ±0.60 | .982 ±.001 | 37.19 ±0.52 | .992 ±.001 |
| B2B ($N$=2400) | 96k | 28.85 ±0.37 | .950 ±.002 | 34.46 ±0.50 | .985 ±.001 | 29.51 ±0.45 | .956 ±.002 | 34.59 ±0.49 | .986 ±.001 |
| B2B ($N$=8000) | 320k | 21.52 ±0.15 | .895 ±.003 | 34.55 ±0.32 | .985 ±.001 | 24.84 ±0.19 | .925 ±.002 | 35.78 ±0.40 | .989 ±.001 |
| IPG (ours) | 96k | **34.03** ±0.53 | **.980** ±.001 | **37.43** ±0.47 | **.992** ±.001 | **35.10** ±0.53 | **.986** ±.001 | **38.35** ±0.46 | **.994** ±.000 |

### 3.2 Reconstruction Performance

**Cohort-level quantitative performance.** Table 2 summarizes reconstruction accuracy over the 19 evaluation subjects at 50 and 100-dB SNR with AF-16 and AF-64 using radial trajectories. The improvement over the parameter-matched raster Gaussian holds in every tested configuration: from $33.32 \pm 0.57$ to $34.03 \pm 0.53$ dB at AF-64 and 50-dB SNR and from $37.19 \pm 0.52$ to $38.35 \pm 0.46$ dB at AF-16 and 100dB SNR. The margin over the raster Gaussian, CS–TV and HashINR is smaller at 50 than at 100 dB.

**Qualitative reconstruction.** Figure 3 compares one held-out evaluation subject (s04) at 100-dB SNR and AF-16 and AF-64. In the example shown, IPG demonstrates lower localized errors than the rasterized Gaussian and parameter-matched B2B reconstructions around narrow tissue boundaries and cortical structures, with the differences more apparent under AF-64 undersampling. Additional qualitative comparisons are provided in Appendix G.

### 3.3 Component Analysis

Tables 2 and 3 share the AF-16, 100-dB condition. The raster Gaussian (APG, ESPIRiT maps, raster TV) reaches 37.19 dB; the analytical pipeline with the same APG amplitude, Fourier coil maps and MC-TV reaches 37.08 dB; switching to IPG raises it to 38.35 dB. These two APG configurations differ in three components at once, so the comparison bounds their combined effect: removing rasterization does not by itself improve accuracy here. Its benefits are a shared analytical multi-coil model and grid-free regularization, while the accuracy gain comes mainly from the integral parameterization.

**Fourier sensitivity representation.** Keeping the IPG object fixed, replacing the proposed $L = 125$ Fourier CSM with a $K = 64$ isotropic-Gaussian mixture (Singh et al., 2024) decreases PSNR from 38.35 to 31.08 dB and SSIM from 0.994 to 0.970 (Table 3). A smaller parameter budget does not explain the difference: Fourier $L = 125$ uses 250 real parameters per coil and attains a relative map error of 0.0358, versus 384 parameters and an error of 0.0619 for Gaussian ($K = 64$), as detailed in Appendix H and Fig. 5a.

**Continuous regularization.** We compare no spatial regularization, rasterized finite-difference TV, and the proposed Monte Carlo TV at AF 64 and 50-dB SNR in Table 3. MC-TV improves PSNR by +2.25dB over no regularization for 11% extra time, whereas raster TV requires 216% extra time

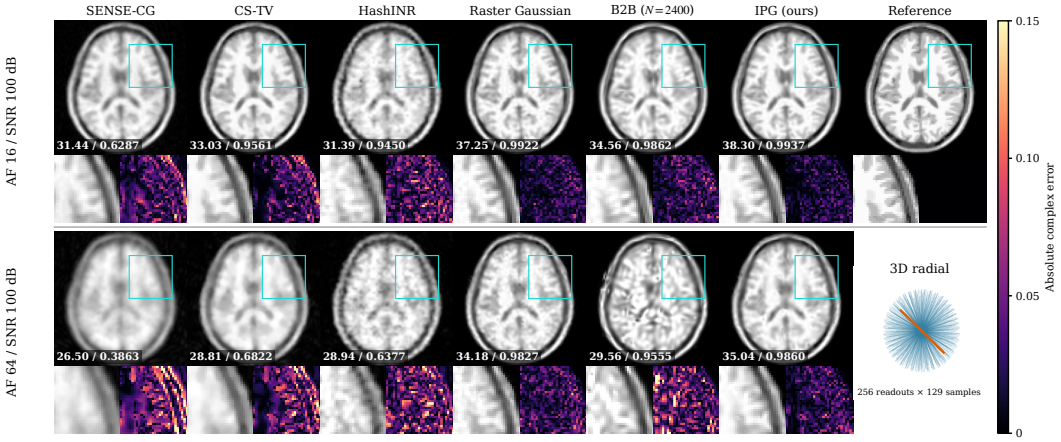



Figure 3: Qualitative reconstruction comparison for simulation subject s04 at $128^3$ resolution and 100-dB SNR. From top to bottom, groups show radial sampling at AF-16 and AF-64, with 15 coils. Columns show SENSE–CG, CS–TV, HashINR, the rasterized Gaussian baseline, parameter-matched B2B ($N = 2400$), the proposed IPG reconstruction, and the reference. All reconstruction panels share the display window, and color scales. Numeric annotations report full-volume PSNR/SSIM.

Table 3: Single-factor component ablations over the 19 evaluation subjects. Object and coil parameterizations are evaluated at AF=16 and 100-dB SNR; spatial regularization at AF=64 and 50-dB SNR. Entries are mean ± sample standard deviation; Time is the total reconstruction time on one RTX 6000 Ada, including the sensitivity calibration and CG initialization.

| Setting | Configuration | PSNR (dB) ↑ | SSIM ↑ | Time (s) ↓ |
|---|---|---|---|---|
| AF-16 100-dB SNR | APG + Fourier CSM | 37.08 ±.52 | .992 ±.001 | 1421 ±04 |
| | IPG + Gaussian CSM | 31.08 ±.71 | .970 ±.002 | 3833 ±21 |
| | **IPG + Fourier CSM** | **38.35** ±.46 | **.994** ±.000 | **1385** ±05 |
| AF-64 50-dB SNR | None | 31.78 ±.50 | .960 ±.003 | **386** ±02 |
| | Raster TV | 33.63 ±.53 | .978 ±.001 | 1219 ±06 |
| | **MC-TV** | **34.03** ±.53 | **.980** ±.001 | 429 ±19 |

for 1.85dB improvement. We further discuss the sensitivity to the number of Monte Carlo samples is shown in Fig. 5b.

## 4 Discussion

**Infinite support.** Because Gaussian primitives have infinite support, the analytical forward model formally allows signal outside the reconstruction FOV; in the tested reconstructions, the resulting out-of-FOV leakage was shown to be negligible ( 3.4e-7 to 8.9e-6 of the total energy in simulation, 8.1e-5 in phantom)

**Scope of grid-free reconstruction.** ESPIRiT sensitivity calibration and the CG reconstruction initialization and the evaluation operate on discrete grids. Once the Gaussian representation is initialized, neither the multi-coil forward model, data consistency, nor Monte Carlo TV requires rasterization; *grid free* refers to this optimization loop.

**Beyond MRI.** The conditioning analysis in Section 2.2 depends on the acquisition through primitive-weighted spectral moments. This provides a framework for studying IPG in other Fourier-based inverse problems, e.g., parallel-beam tomography via the Fourier-slice theorem and radio interferometric imaging under a planar Fourier measurement model.

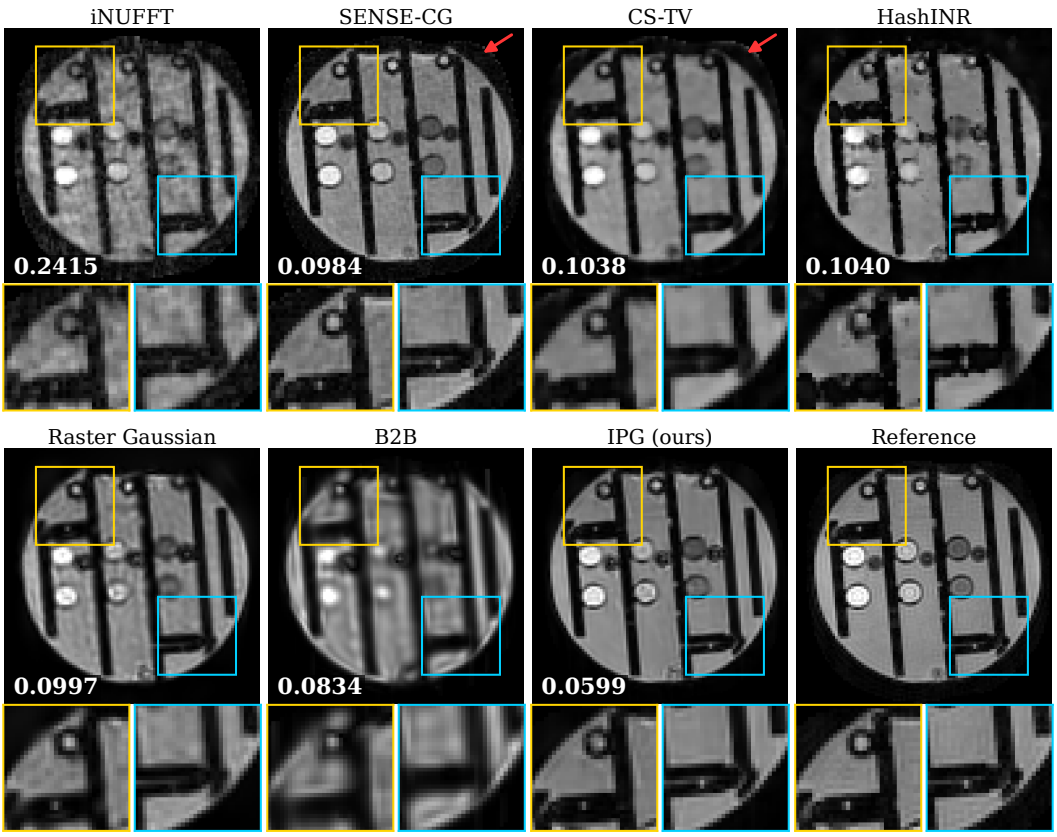
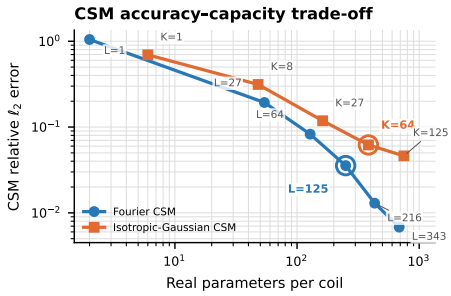
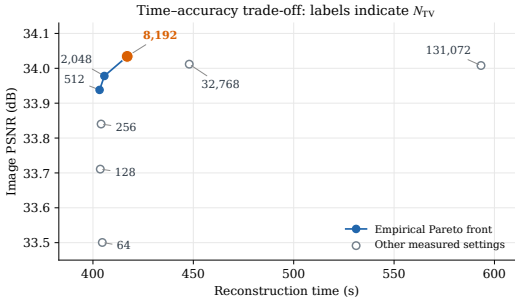



Figure 4: 3T phantom reconstructions from 1024 radial spokes ($160^3$, 2-mm voxels). Each panel shows the central sagittal magnitude image and two matched ROIs below it. Red arrows indicate residual signal outside the phantom in SENSE–CG and CS–TV. Numbers report relative held-out $k$-space $\ell_1$ error, $\sum |\alpha y^{\text{pred}} - y| / \sum |y|$, using one global complex least-squares scale and 3072 unused spokes; lower is better. The reference image is an iNUFFT reconstruction of an independent 22,940-spoke acquisition.




(a) CSM approximation error vs parameters count. Rings mark the reconstruction operating points, Fourier $L = 125$ and Gaussian $K = 64$.




(b) MC-TV time–accuracy trade-off (AF-64, 50-dB SNR). Labels give $N_{\text{TV}}$; $N_{\text{TV}} = 8192$ is the selected setting.

Figure 5: Component analysis of (a) the Fourier coil-sensitivity model and (b) the number of Monte Carlo TV samples.

**Summary.** We introduced a Fourier-dual Gaussian formulation for 3D non-Cartesian parallel MRI reconstruction. IPGs separate each primitive's integral from its covariance, Fourier coil-sensitivity modes turn spatial modulation into shifts of the same closed-form spectrum, and Monte Carlo TV regularizes without rasterization. Across the simulated condition and experimental scan, the formulation achieves lower reconstruction error than the evaluated baseline.

## 5 AI USE STATEMENT

AI tools were used to assist with implementation and to improve the clarity and presentation of the manuscript. All AI-assisted outputs were reviewed and supervised by the authors. The authors take full responsibility for the methods, results, interpretations, and content of the manuscript.

## REFERENCES


Pierre-Antoine Comby, Guillaume Daval-Frérot, Caini Pan, A. Tanabene, Léna Oudjman, Matteo Cencini, Philippe Ciuciu, and ChaithyaG R. Mri-nufft: Doing non-cartesian mri has never been easier. *Journal of Open Source Software*, 10(108):7743, 2025. doi: 10.21105/joss.07743.

Jie Feng, Ruimin Feng, Qing Wu, Xin Shen, Lixuan Chen, Xin Li, Li Feng, Jingjia Chen, Zhiyong Zhang, Chunlei Liu, et al. Spatiotemporal implicit neural representation for unsupervised dynamic MRI reconstruction. *IEEE Transactions on Medical Imaging*, 44(5):2143–2156, 2025. doi: 10.1109/TMI.2025.3526452.

Ruimin Feng, Qing Wu, Jie Feng, Huajun She, Chunlei Liu, Yuyao Zhang, and Hongjiang Wei. IMJENSE: Scan-specific implicit representation for joint coil sensitivity and image estimation in parallel MRI. *IEEE Transactions on Medical Imaging*, 43(4):1539–1553, 2024. doi: 10.1109/TMI.2023.3342156.

Jeffrey A. Fessler and Bradley P. Sutton. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing*, 51(2):560–574, 2003. doi: 10.1109/TSP.2002.807005.

Mark A Griswold, Peter M Jakob, Robin M Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (grappa). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002. doi: 10.1002/mrm.10171.

Matthieu Guerquin-Kern, Laurent Lejeune, Klaas P. Pruessmann, and Michael Unser. Realistic analytical phantoms for parallel magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 31(3):626–636, 2012. doi: 10.1109/TMI.2011.2174158.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and G. Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023. doi: 10.1145/3592433.

RK-S Kwan, Alan C Evans, and G Bruce Pike. Mri simulation-based evaluation of image-processing and classification methods. *IEEE transactions on medical imaging*, 18(11):1085–1097, 1999. doi: 10.1109/42.816072.

Robert M. Lewitt. Multidimensional digital image representations using generalized Kaiser–Bessel window functions. *Journal of the Optical Society of America A*, 7(10):1834–1846, 1990. doi: 10.1364/JOSAA.7.001834.

Qiaoxin Li, Caini Pan, Pierre-Antoine Comby, Chaithya Giliyar, and Philippe Ciuciu. Dd-inr: Dynamics-driven implicit neural representation for accelerated whole-brain functional mri reconstruction. *arXiv.org*, 2026. doi: 10.48550/arXiv.2606.10756.

Michael Lustig and John M Pauly. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Magnetic Resonance in Medicine*, 64(2):457–471, 2010. doi: 10.1002/mrm.22428.

Thomas Müller, Alex Evans, Christoph Schied, and A. Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):1–15, 2022. doi: 10.1145/3528223.3530127.

Tri M Ngo, George SK Fung, Shuo Han, Min Chen, Jerry L Prince, Benjamin MW Tsui, Elliot R McVeigh, and Daniel A Herzka. Realistic analytical polyhedral mri phantoms. *Magnetic resonance in medicine*, 76(2):663–678, 2015. doi: 10.1002/mrm.25888.

Frank Ong, Martin Uecker, and Michael Lustig. Accelerating non-Cartesian MRI reconstruction convergence using $k$-space preconditioning. *IEEE Transactions on Medical Imaging*, 39(5):1646–1654, 2020. doi: 10.1109/TMI.2019.2954121.

Greg Ongie and Mathews Jacob. Off-the-grid recovery of piecewise constant images from few Fourier samples. *SIAM Journal on Imaging Sciences*, 9(3):1004–1041, 2016. doi: 10.1137/15M1042280.

Tengya Peng, Ruyi Zha, Zhen Li, Xiaofeng Liu, and Qing Zou. Three-dimensional MRI reconstruction with 3D gaussian representations: Tackling the undersampling problem. *IEEE Transactions on Medical Imaging*, 45(5):1905–1917, 2025. doi: 10.1109/TMI.2025.3642134.

Stephen E. Russek, National Institute of Standards, and Technology. Nist/nibib medical imaging phantom lending library, 2 2021. URL https://data.nist.gov/od/id/mds2-2366.

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*, 2021. doi: 10.1109/TNNLS.2022.3177134.

Imraj Singh, Andrew Dupuis, Simran Kukran, Chaitra Badve, and Mark Griswold. Back to basis: Spatially continuous mrf with adaptive gaussians. In *2026 ISMRM-ISMRT Annual Meeting and Exhibition*, Cape Town, South Africa, May 2026. ISMRM. URL http://echo.ismrm.org/abstracts/view/df1ab44d-4965-468a-815c-a5d461de1dbb. Program Number: 430-03-009.

Imraj R. D. Singh, Željko Kereta, Alexander Denker, Riccardo Barbano, Bangti Jin, Kris Thielemans, and Simon Arridge. MR-blob: Coordinate-transformed blobs for parallel mri reconstruction. In *ISMRM Annual Meeting*, number 1911. ISMRM, 2024. doi: 10.58530/2024/1911.

A. Sluis. Condition numbers and equilibration of matrices. *Numerische Mathematik*, 14(1):14–23, 1969. doi: 10.1007/BF02165096.

Martin Uecker, Thorsten Hohage, Kai Tobias Block, and Jens Frahm. Image reconstruction by regularized nonlinear inversion—joint estimation of coil sensitivities and image content. *Magnetic Resonance in Medicine*, 60(3):674–682, 2008. doi: 10.1002/mrm.21691.

Martin Uecker, Peng Lai, Mark J. Murphy, Patrick Virtue, Michael Elad, John M. Pauly, Shreyas S. Vasanawala, and Michael Lustig. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3):990–1001, 2014. doi: 10.1002/mrm.24751.

Ruyi Zha, Tao Jun Lin, Yuanhao Cai, Jiwen Cao, Yanhao Zhang, and Hongdong Li. $R^2$-Gaussian: Rectifying radiative Gaussian splatting for tomographic reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

# APPENDIX

## A Reproducibility Statement

The forward model is specified in closed form in Section 2: Eq. (IPG-kspace) for the IPG spectrum and Eq. (5) for the multi-coil operator, with all analytical derivatives in Appendix D. The objective, including the adaptive Huber threshold and the Monte Carlo TV estimator with its scaling convention, is given in Section 2.4; initialization and the three-stage optimization schedule are in Appendix C, and every numerical hyperparameter,select on the development subjet are in Table 4.

Appendix F specifies the development/evaluation split, the noise model, the exact metric definitions and the optimization budget of every method. The anatomies are the public BrainWeb phantoms (Kwan et al., 1999) and the analytical simulation follows Guerquin-Kern et al. (2012); Ngo et al. (2015); trajectories and noise realizations are generated from fixed seeds and shared across methods. A minimal, self-contained implementation with data for reproducing the baselines and the proposed method is included in the supplementary material. The full source code will be released publicly upon acceptance.

## B Notation and Fourier Convention

We use spatial coordinates $\boldsymbol{r} \in \mathbb{R}^3$ and spatial-frequency coordinates $\boldsymbol{k} \in \mathbb{R}^3$, with frequencies measured in cycles per spatial unit. Our Fourier transform convention is

$$\widehat{\rho}(\boldsymbol{k}) = \int_{\mathbb{R}^3} \rho(\boldsymbol{r}) e^{-i2\pi \boldsymbol{k}^\top \boldsymbol{r}} \, \mathrm{d}\boldsymbol{r}.$$

| Symbol | Meaning |
|---|---|
| $\boldsymbol{r},\ \boldsymbol{k}$ | image and spatial-frequency coordinates |
| $\rho,\ \widehat{\rho}$ | continuous complex object and its Fourier transform |
| $s_c,\ \beta_{c,\ell}$ | sensitivity of coil $c$ and its Fourier coefficient |
| $\boldsymbol{k}_m,\ M$ | acquired non-Cartesian sample and number of samples |
| $\boldsymbol{k}_\ell,\ L$ | sensitivity frequency and number of Fourier modes |
| $N_g,\ C$ | number of Gaussian primitives and coils |
| $\boldsymbol{\mu}_n,\ \boldsymbol{\Sigma}_n$ | primitive center and covariance |
| $\widetilde{\boldsymbol{q}}_n,\ \boldsymbol{q}_n$ | trainable quaternion and its unit-normalized form |
| $\zeta_{d,n},\ \sigma_{d,n}$ | log-scale parameter and constrained principal standard deviation |
| $Z_n = (2\pi)^{3/2}\lvert\boldsymbol{\Sigma}_n\rvert^{1/2}$ | Gaussian integral factor |
| $w_n$ | complex spatial integral of IPG primitive $n$ |
| $\Theta$ | trainable object parameters |
| $\boldsymbol{B}$ | matrix of Fourier sensitivity coefficients |
| $\Omega$ | continuous reconstruction domain |
| $N_{\mathrm{TV}}$ | number of Monte Carlo TV samples |

## C Initialization and Optimization

The reconstruction is initialized in three steps.

**Image prior.** We first compute an early-stopped CG reconstruction from the measured $k$-space. This volume is used only to initialize the continuous representation and is not part of the subsequent forward model.

**Geometry initialization.** Primitive centers are sampled with replacement according to the magnitude-normalized initialization volume and perturbed by sub-voxel jitter. All principal standard deviations are initialized to the common value $\sigma_0 = 1.3$ nominal voxels, and all orientations are initialized to the identity rotation.

**Coefficient initialization.** For fixed initial geometry $\Gamma^{(0)} = \{\boldsymbol{\mu}_n^{(0)}, \boldsymbol{\zeta}_n^{(0)}, \boldsymbol{q}_n^{(0)}\}_{n=1}^{N_g}$, the multi-coil forward model is linear in $\boldsymbol{w} = (w_1, \ldots, w_{N_g})^\top$. Let $\boldsymbol{A}_{\Gamma^{(0)}}$ denote this linear operator evaluated on a subset of at most $2 \times 10^4$ acquired samples. We initialize

$$\boldsymbol{w}^{(0)} = \arg\min_{\boldsymbol{w}} \|\boldsymbol{A}_{\Gamma^{(0)}}\boldsymbol{w} - \boldsymbol{y}\|_2^2 + \lambda_w \bar{\eta} \|\boldsymbol{w}\|_2^2,$$

where $\boldsymbol{y}$ stacks the selected multi-coil measurements and $\bar{\eta} = \frac{1}{N_g} \operatorname{tr}\left(\boldsymbol{A}_{\Gamma^{(0)}}^H \boldsymbol{A}_{\Gamma^{(0)}}\right)$ scales the ridge relative to the Gram matrix.

After initialization, Adam is applied in three stages:

$$\text{S1: } \{w_n\}, \tag{9}$$
$$\text{S2: } \{w_n, \boldsymbol{\mu}_n\}, \tag{10}$$
$$\text{S3: } \{w_n, \boldsymbol{\mu}_n, \boldsymbol{\zeta}_n, \widetilde{\boldsymbol{q}}_n\}. \tag{11}$$

Each stage uses a separate optimizer and cosine learning-rate schedule. After each update, centers and log-scales are projected to their prescribed bounds, and quaternions are renormalized to unit norm. The APG ablation and the raster Gaussian use the same stages with $a_n$ in place of $w_n$.

## D Analytical Derivatives in k-space

Suppressing the index $n$, we use Wirtinger derivatives for the complex coefficient $w$ and ordinary derivatives for the real geometric parameters. Vector derivatives are written as column vector. Let:

$$\hat{g}(\boldsymbol{k}) = \exp\left[-i2\pi \boldsymbol{k}^\top \boldsymbol{\mu}\right] \exp\left[-2\pi^2 \boldsymbol{k}^\top \boldsymbol{\Sigma} \boldsymbol{k}\right] \tag{12}$$

With the coordinate: $\boldsymbol{u} = 2\pi \operatorname{diag}(\sigma)\boldsymbol{R}^\top \boldsymbol{k}$, $s = \|\boldsymbol{u}\|_2^2$

$$|\hat{g}(\boldsymbol{k})|^2 = \exp[-\boldsymbol{u}^\top \boldsymbol{u}] = \exp[-u^2] \tag{13}$$

Below, $\boldsymbol{E}_{dd} = \boldsymbol{e}_d \boldsymbol{e}_d^\top$ and $\boldsymbol{r}_d(\boldsymbol{q})$ is the $d$-th column of $\boldsymbol{R}(\boldsymbol{q})$. We have the following derivatives:

$$\frac{\partial \hat{\phi}}{\partial w} = \hat{g}(\boldsymbol{k}) \tag{14}$$
$$\frac{\partial \hat{\phi}}{\partial \boldsymbol{\mu}} = -i2\pi w \hat{g}(\boldsymbol{k}) \cdot \boldsymbol{k} \tag{15}$$
$$d\hat{\phi} = -2\pi^2 w \hat{g}(\boldsymbol{k}) \boldsymbol{k}^\top (d\boldsymbol{\Sigma}) \boldsymbol{k} \tag{16}$$

Then to propagate in $\boldsymbol{\Sigma}$

$$\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_d} = 2\boldsymbol{R}(\boldsymbol{q}) \sigma_d \boldsymbol{E}_{dd} \boldsymbol{R}(\boldsymbol{q})^\top \tag{17}$$
$$\frac{\partial \sigma_d}{\partial \zeta_d} = \sigma_d = e^{\zeta_d} \tag{18}$$
$$\frac{\partial \boldsymbol{\Sigma}}{\partial q_i} = \frac{\partial \boldsymbol{R}}{\partial q_i} \operatorname{diag}(\boldsymbol{\sigma})^2 \boldsymbol{R}^\top + \boldsymbol{R} \operatorname{diag}(\boldsymbol{\sigma})^2 \frac{\partial \boldsymbol{R}}{\partial q_i}^\top \tag{19}$$
$$\frac{\partial \boldsymbol{\Sigma}}{\partial \widetilde{\boldsymbol{q}}} = \frac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{q}} \frac{\boldsymbol{I} - \boldsymbol{q}\boldsymbol{q}^\top}{\|\widetilde{\boldsymbol{q}}\|} \tag{20}$$

Chain rule:

$$\frac{\partial \hat{\phi}}{\partial \sigma_d} = -2\pi^2 w \hat{g}(\boldsymbol{k}) \boldsymbol{k}^\top \frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_d} \boldsymbol{k} = -4\pi^2 \sigma_d w \hat{g}(\boldsymbol{k}) (\boldsymbol{k}^\top \boldsymbol{r}_d(\boldsymbol{q}))^2 \tag{21}$$
$$\frac{\partial \hat{\phi}}{\partial \zeta_d} = -4\pi^2 \sigma_d^2 w \hat{g}(\boldsymbol{k}) (\boldsymbol{k}^\top \boldsymbol{r}_d(\boldsymbol{q}))^2 \tag{22}$$
$$\frac{\partial \hat{\phi}}{\partial \boldsymbol{\zeta}} = -4\pi^2 w \hat{g}(\boldsymbol{k}) (\operatorname{diag}(\boldsymbol{\sigma}) \boldsymbol{R}^\top \boldsymbol{k})^{\circ 2} = -w \hat{g}(\boldsymbol{k}) \boldsymbol{u}^{\circ 2} \tag{23}$$

Here $\circ 2$ denotes elementwise squaring.

Finally, we also have the following gradient in image space, used for computing the Monte-Carlo TV:

$$\nabla_{\boldsymbol{r}} \rho_\Theta(\boldsymbol{r}) = -\sum_{n=1}^{N_g} \phi_n(\boldsymbol{r}) \boldsymbol{\Sigma}_n^{-1}(\boldsymbol{r} - \boldsymbol{\mu}_n). \tag{24}$$

### D.1 Derivatives of APG in k-space

For APG ($\psi$) we also have:

$$w(a, \boldsymbol{\zeta}) = aZ = a(2\pi)^{\frac{3}{2}} |\Sigma|^{\frac{1}{2}} = a(2\pi)^{\frac{3}{2}} \exp(\mathbf{1}^\top \boldsymbol{\zeta})$$

Which gives the following reparametrization

$$\hat{\psi}(a, \boldsymbol{\zeta}, \boldsymbol{\mu}, \boldsymbol{q}) = \hat{\phi}(w(a, \boldsymbol{\zeta}), \boldsymbol{\zeta}, \boldsymbol{\mu}, \boldsymbol{q}) \tag{25}$$

which admits the following derivatives:

$$\frac{\partial w}{\partial a} = Z = (2\pi)^{\frac{3}{2}} \exp(\mathbf{1}^\top \boldsymbol{\zeta}) \tag{26}$$

$$\frac{\partial w}{\partial \boldsymbol{\zeta}} = a(2\pi)^{\frac{3}{2}} \exp(\mathbf{1}^\top \boldsymbol{\zeta}) \mathbf{1} = w\mathbf{1} \tag{27}$$

Following (25) we have

$$\frac{\partial \hat{\psi}}{\partial \boldsymbol{\mu}} = \frac{\partial \hat{\phi}}{\partial \boldsymbol{\mu}}, \quad \frac{\partial \hat{\psi}}{\partial \boldsymbol{q}} = \frac{\partial \hat{\phi}}{\partial \boldsymbol{q}}, \tag{28}$$

$$\frac{\partial \hat{\psi}}{\partial a} = \frac{\partial w}{\partial a} \frac{\partial \hat{\phi}}{\partial w} = Z\hat{g} \tag{29}$$

$$\frac{\partial \hat{\psi}}{\partial \boldsymbol{\zeta}} = \frac{\partial w}{\partial \boldsymbol{\zeta}} \frac{\partial \hat{\phi}}{\partial w} + \frac{\partial \hat{\phi}}{\partial \boldsymbol{\zeta}} = w\hat{g}(\boldsymbol{k})(\mathbf{1} - \boldsymbol{u}^{\circ 2}) \tag{30}$$

## E Conditioning of amplitude- and integral-parameterized Gaussians

APG and IPG describe the same images; they differ only in the trainable amplitude, the peak value, $a$ or the integral $w = aZ$. We quantify how this choice couples a primitive's energy to its scale in the data-consistency loss, and how the coupling depends on the sampling trajectory.

### E.1 Setting

We consider a single coil and the unweighted loss $\ell(\theta) = \frac{1}{2} \sum_m |y_m - \hat{\rho}_\theta(\boldsymbol{k}_m)|^2$, the negative log-likelihood under i.i.d. Gaussian noise With the Fourier coil model of Eq. (4), the analysis carries over when each sensitivity is approximately constant over a primitive's support. We then make the following assumptions:

- (A1) *Isotropic primitive*, $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$ with $\sigma = e^{\zeta}$, so that $Z = (2\pi)^{3/2} e^{3\zeta}$; this is the state of all primitives when the scales are released (stage S3, $\sigma_0 = 1.3$ voxels).
- (A2) *Single primitive*: couplings between primitives are neglected.
- (A3) *Gauss–Newton*:$\boldsymbol{H} = \mathrm{Re}(\boldsymbol{J}^H \boldsymbol{J})$ approximates the Hessian of $\ell$ (small residuals).

The adaptive Huber loss of Eq. (7) is *not* equivalent to $\ell$,as $\delta$ tracks the median residual, a fixed fraction of residuals (half if $\alpha_{\mathrm{H}} = 1$) stays in the linear regime at any residual scale. Its Gauss–Newton matrix is $\mathrm{Re}(\boldsymbol{J}^H \boldsymbol{W} \boldsymbol{J})$ with $W_m = \min(1, \delta/|r_m|)$, and all results below hold with the weights $\nu_m \propto W_m e^{-s_m}$. Table 1, computed with $W_m = 1$, carries over when the weights are uncorrelated with $\|\boldsymbol{k}_m\|$, as for noise-dominated residuals.

We further discuss Assumptions (A1) and (A2) in Appendix E.4.

**Jacobian.** Under (A1) both models have spectrum $\hat{\phi} = w\,\hat{g}$ with $\hat{g}(\boldsymbol{k}) = e^{-i2\pi\boldsymbol{k}^\top\boldsymbol{\mu}}e^{-s/2}$ and $s = (2\pi\sigma\|\boldsymbol{k}\|)^2$. Changing the scale at fixed trainable amplitude changes the energy by $\delta := \partial \log w/\partial\zeta$, equal to 0 for IPG and 3 for APG. Writing the amplitude in polar form (log-magnitude $\omega$, phase $\varphi$) and dropping constant factors, which do not affect the diagonally scaled condition numbers used below, the Jacobian columns are

$$\partial_\omega\hat{\phi} = w\hat{g}, \qquad \partial_\varphi\hat{\phi} = iw\hat{g}, \qquad \partial_\zeta\hat{\phi} = w(\delta - s)\hat{g}, \qquad \partial_{\boldsymbol{\mu}}\hat{\phi} = -i2\pi\boldsymbol{k}\,w\hat{g}, \tag{31}$$

consistent with Eqs. (23) and (30). All curvature entries are averages under the probability measure $\nu_m \propto e^{-s_m}$, i.e. the acquired samples weighted by the primitive's own spectral envelope.

### E.2 Energy–scale coupling

**Proposition 1** (Block structure)**.** *Under (A1)–(A3), with parameters ordered as* $(\omega, \varphi, \zeta, \boldsymbol{\mu})$*,*

$$\boldsymbol{H} \propto \begin{pmatrix} 1 & 0 & \mathbb{E}_\nu[\delta - s] & 0 \\ 0 & 1 & 0 & -2\pi\,\mathbb{E}_\nu[\boldsymbol{k}]^\top \\ \mathbb{E}_\nu[\delta - s] & 0 & \mathbb{E}_\nu[(\delta - s)^2] & \boldsymbol{0}^\top \\ 0 & -2\pi\,\mathbb{E}_\nu[\boldsymbol{k}] & \boldsymbol{0} & 4\pi^2\,\mathbb{E}_\nu[\boldsymbol{k}\boldsymbol{k}^\top] \end{pmatrix}. \tag{32}$$

*For centro-symmetric sampling ($\mathbb{E}_\nu[\boldsymbol{k}] = 0$: Cartesian, radial, cones), the only coupling is between energy $\omega$ and scale $\zeta$.*

*Proof.* Each entry is $\mathrm{Re}\sum_m \overline{J_{mp}}J_{mq}$ with the columns (31). The $(\zeta, \boldsymbol{\mu})$, $(\varphi, \zeta)$ and $(\omega, \boldsymbol{\mu})$ products are purely imaginary, e.g. $\overline{w(\delta - s)\hat{g}}\,(-i2\pi\boldsymbol{k}w\hat{g}) = -i2\pi|w|^2(\delta - s)e^{-s}\boldsymbol{k}$. □

**Lemma 1.** *For a* $2 \times 2$ *SPD matrix* $\boldsymbol{A}$*,*

$$\kappa^\star(\boldsymbol{A}) := \min_{\boldsymbol{D}\ diag>0} \kappa(\boldsymbol{D}\boldsymbol{A}\boldsymbol{D}) = (1 + |c|)/(1 - |c|) \quad c = A_{12}/\sqrt{A_{11}A_{22}} \tag{33}$$

*attained by Jacobi scaling* $D = \mathrm{diag}(A)^{-1/2}$*.*

*Proof.* Jacobi scaling gives eigenvalues $1 \pm |c|$; $\mathrm{tr}^2/\det = \kappa + 2 + \kappa^{-1}$ increases with $\kappa$, $\det(\boldsymbol{D}\boldsymbol{A}\boldsymbol{D}) = d_1^2d_2^2\det\boldsymbol{A}$, and at fixed $d_1^2d_2^2$ the trace is minimal iff $d_1^2A_{11} = d_2^2A_{22}$. □

**Proposition 2** (Energy–scale conditioning)**.** *Under (A1)–(A3), let* $t = (\delta - \mathbb{E}_\nu[s])/\sqrt{\mathrm{Var}_\nu[s]}$*. The* $\{\omega, \zeta\}$ *block has correlation* $c = t/\sqrt{1 + t^2}$ *and*

$$\kappa^\star = \left(|t| + \sqrt{1 + t^2}\right)^2, \qquad t_{\mathrm{IPG}} = \frac{-\mathbb{E}_\nu[s]}{\sqrt{\mathrm{Var}_\nu[s]}}, \quad t_{\mathrm{APG}} = \frac{3 - \mathbb{E}_\nu[s]}{\sqrt{\mathrm{Var}_\nu[s]}}. \tag{34}$$

*Hence IPG is better conditioned than APG if and only if* $\mathbb{E}_\nu[s] \le 3/2$*.*

*Proof.* By Proposition 1, $c = \mathbb{E}_\nu[\delta - s]/\sqrt{\mathrm{Var}_\nu[s] + (\delta - \mathbb{E}_\nu[s])^2}$; Lemma 1 and $(\sqrt{1 + t^2} + |t|)(\sqrt{1 + t^2} - |t|) = 1$ give (34). $\kappa^\star$ increases with $|t|$, so the comparison reduces to $\mathbb{E}_\nu[s] \le 3 - \mathbb{E}_\nu[s]$. □

**Relation to the optimizer and schedule.** $\kappa^\star$ is the best condition number reachable by *any* diagonal preconditioner. Adam's preconditioner is diagonal, so $\kappa^\star$ is a floor for it rather than a prediction of its behavior; near a noise-dominated optimum Adam behaves closer to a $\mathrm{diag}(\boldsymbol{H})^{1/2}$ scaling. The stages of Appendix C must be distinguished. In S1–S2 the scales are frozen, and APG and IPG differ only by the per-primitive rescaling $w = aZ_n$, with a common $Z_0 = (2\pi)^{3/2}\sigma_0^3 \approx 35$ (voxel units) at initialization; this leaves $\kappa^\star$ unchanged, so any difference in these stages comes from the effective step size ($Z_0$ times the APG coefficient learning rate, in $w$ units). Only in S3, when the scales are trained, is the change of variables $(a, \zeta) \mapsto (w, \zeta)$ a shear, whose coupling no per-parameter step size can remove.

**Density compensation.** Weighting the data term to equalize the sampling density (effective $p = 0$) makes APG and IPG equivalent in Table 1 ($\kappa^\star = 7.9$), but also reduces the long-range $1/\|\boldsymbol{r}\|$ coupling between primitives discussed in Appendix E.4, which motivates $k$-space preconditioning for non-Cartesian MRI (Ong et al., 2020). It also changes the effective problem being solved, and can make the reconstruction highly sensitive to high frequency noise. We therefore use no density compensation.

### E.3 Dependence on the sampling density

$\kappa^\star$ depends on the trajectory only through the first two moments of $u$ under $\nu$.

**Proposition 3** (Closed form for conditioning number)**.** *If the samples have radial density $\varrho(\boldsymbol{k}) \propto \|\boldsymbol{k}\|^{-p}$ on $\|\boldsymbol{k}\| \le k_{\max}$:*

*Then:*

1. *The law of $s$ under the measure $\nu$ is a Gamma law of shape $q = (3-p)/2$ truncated to $[0, T]$ with $T = 4\pi^2\sigma^2 k_{\max}^2 = \pi^2\sigma_{\text{vox}}^2$, for $k_{\max}$ at the Nyquist limit and $\sigma_{\text{vox}}$ the scale in voxel unit.*
2. *$\mathbb{E}_\nu[s] < q \le 3/2$ for every $p \ge 0$, so IPG is never worse than APG.*
3. *When the primitive envelope is well contained in the sampled frequencies ($T \to \infty$),*

$$\kappa^\star_{\text{IPG}} \to \left(\sqrt{q} + \sqrt{1+q}\right)^2, \qquad \kappa^\star_{\text{APG}} \to \left(\tfrac{3-q}{\sqrt{q}} + \sqrt{1 + \tfrac{(3-q)^2}{q}}\right)^2. \tag{35}$$

*Proof outline.* In 3D the density $\|\boldsymbol{k}\|^{-p}$ and the volume element $\|\boldsymbol{k}\|^2 \,\mathrm{d}\|\boldsymbol{k}\|$ give $u = 2\pi\sigma\|\boldsymbol{k}\|$ the density $\propto u^{2-p}$ on $[0, \sqrt{T}]$; with $s = u^2$ this becomes $\propto s^{(1-p)/2}$, and weighting by $e^{-s}$ gives a Gamma$(\xi, 1)$ law truncated to $[0, T]$, integrable for $p < 3$. Truncation lowers the mean, so $\mathbb{E}_\nu[s] < \xi \le 3/2$ for $p \ge 0$, and Proposition 2 gives item 2. Item 3 follows from $\mathbb{E}[s] = \mathrm{Var}[s] = \xi$ for the untruncated law. □

The two coincide for uniform sampling ($p = 0$): the advantage of IPG comes from the concentration of samples at low frequencies. Table 1 reports this continuous model with a ball-shaped sampled band. For 3D radial sampling every spoke has the same $\|\boldsymbol{k}\|$ profile, so the discrete trajectory gives the same values, independently of the number of spokes. For the trajectories used in our experiments, APG's magnitude–scale coupling is an order of magnitude worse than IPG's.

### E.4 Beyond (A1) and (A2)

**Anisotropic primitives.** At the isotropic state the orientation columns vanish, and with three log-scales the energy–scale block is $4 \times 4$. If $\nu$ is invariant under permutations of the axes, an optimal diagonal scaling can be taken permutation invariant, and in the basis $\{\omega, (\zeta_x + \zeta_y + \zeta_z)/\sqrt{3}, \text{two shape directions}\}$ the block splits into the $2 \times 2$ block of Proposition 2 and a shape block that is uncoupled from $\omega$ and identical for APG and IPG. The values of Table 1 therefore carry over (radial, $\sigma_0$: 3.7 and 52.0). Away from isotropy the orientation couples to the scales and no closed form is available; Jacobi scaling remains within a factor $n$ of optimal (Sluis, 1969, Thm. 4.1).

**Several primitives.** Split the full Gauss–Newton matrix as $\boldsymbol{H} = \boldsymbol{B} + \boldsymbol{O}$, with $\boldsymbol{B}$ the per-primitive diagonal blocks and $\boldsymbol{O}$ the blocks between different primitives. Let $\boldsymbol{D} = \mathrm{diag}(\boldsymbol{H})^{-1/2}$, $c_{\max} = \max_n |c_n|$ and $\eta = \|\boldsymbol{DOD}\|_2$. For sampling symmetric under each reflection $k_d \mapsto -k_d$,

$$\max_n \frac{1+|c_n|}{1-|c_n|} \;\le\; \kappa^\star(\boldsymbol{H}) \;\le\; \frac{1 + c_{\max} + \eta}{1 - c_{\max} - \eta} \quad (c_{\max} + \eta < 1). \tag{36}$$

The lower bound follows from Cauchy interlacing, since each $\{\omega_n, \zeta_n\}$ block is a principal submatrix of $\boldsymbol{D}'\boldsymbol{H}\boldsymbol{D}'$ for any diagonal $\boldsymbol{D}'$. The upper bound follows from Weyl's inequality, since $\boldsymbol{DBD}$ has eigenvalues $1 \pm |c_n|$ and 1. Thus the single-primitive value is a floor for the full problem under any diagonal preconditioner: at initialization, APG cannot be brought below 52 on our trajectories.

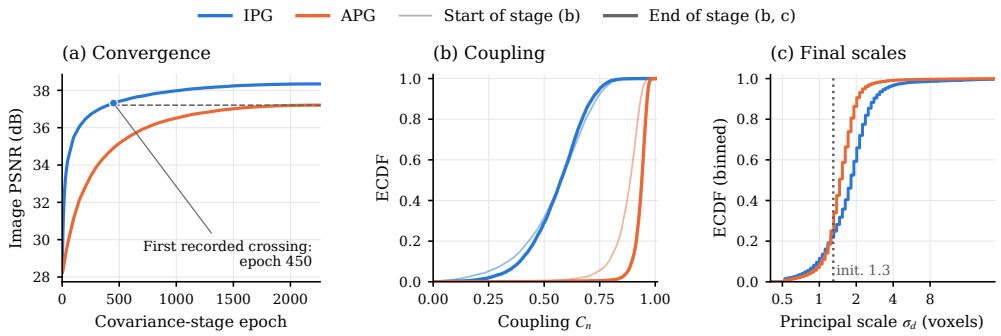



Figure 6: Coupling and scale adaptation during covariance-stage optimization on subject s06 at $128^3$ resolution, 1024 radial spokes, 15 coils, 100-dB SNR, and $N_g = 8000$. (a) Image PSNR versus covariance-stage epoch; the marker indicates the first recorded IPG snapshot reaching the final APG PSNR. (b) Empirical cumulative distribution functions (ECDFs) of the amplitude–scale gradient alignment (Appendix E.5) each arm's training coordinates: the fraction of primitives with $C_n \leq x$, at the start (thin) and end (thick) of the stage. (c) Final scale ECDFs estimated from the archived histograms: the fraction of the 24000 principal axes with $\sigma_d \leq x$. The scale axis is logarithmic; the dotted line marks the common 1.3-voxel initialization. All panels use the same training replays.

It is accurate when the normalized coupling between primitives satisfies $\eta \ll 1 - c_{\max}$, a tolerance of 0.42 for IPG against 0.04 for APG. For unweighted radial sampling the point-spread function decays only as $1/\|\boldsymbol{r}\|$, so $\eta$ can be large; this long-range coupling comes from the acquisition and affects both parameterizations. Applying the same interlacing argument to the $\{\omega_i, \omega_j\}$ block of two nearby primitives gives $\kappa^\star(\boldsymbol{H}) \geq (1 + |c_{ij}|)/(1 - |c_{ij}|)$, with $c_{ij}$ the correlation of their spectra under the sampling weights. At fixed scale the $\omega$-columns of APG and IPG coincide, so this floor is the same for both and, for close primitives under radial sampling, can exceed the single-primitive values.

### E.5 In practice Training dynamics and reconstruction quality

We next track coupling and scale adaptation during covariance-stage optimization on subject s06. The APG and IPG replays start from shared isotropic geometry, with all 8000 primitives initialized at a principal scale of 1.3 voxels, and follows the best learning rates selected above. Coupling is evaluated along each arm's own optimization trajectory, while scale histograms collect all 24000 principal axes.

For each primitive with coefficient $c_n$ ($a_n$ for APG, $w_n$ for IPG), let $u_{nb}$ and $v_{nb}$ be the derivatives of the minibatch Huber data loss $\mathcal{L}_b$ along a log-dilation of $c_n$ and along an isotropic log-dilation of its three scales. We report the uncentered gradient alignment

$$C_n = |\sum_b u_{nb} v_{nb}| / \sqrt{\sum_b u_{nb}^2 \sum_b v_{nb}^2} \tag{37}$$

over $B = 32$ minibatches of 10% of the samples. $C_n$ generalizes the Gauss–Newton correlation $c$ of Proposition 2: it includes all coils, the Huber weights and the other primitives, which is why its initial distribution is broad although all primitives start isotropic.

Figure 6(a) shows the image-PSNR trajectories. The coupling distributions at the start and end of the stage are shown in Figure 6(b), with lower coupling for IPG. IPG also develops a wider final scale distribution, particularly toward larger scales (Figure 6(c)). These observations connect the amplitude–scale coupling difference to the behavior observed during reconstruction.

Figure 7 complements the cohort-level results in Table 3 with reconstructions of held-out subject s06 at two noise levels. With the same number of primitives, covariance parameterization, Fourier sensitivities, prior, and optimization schedule, IPG yields lower errors in the displayed ROIs and higher full-volume PSNR/SSIM under both conditions.

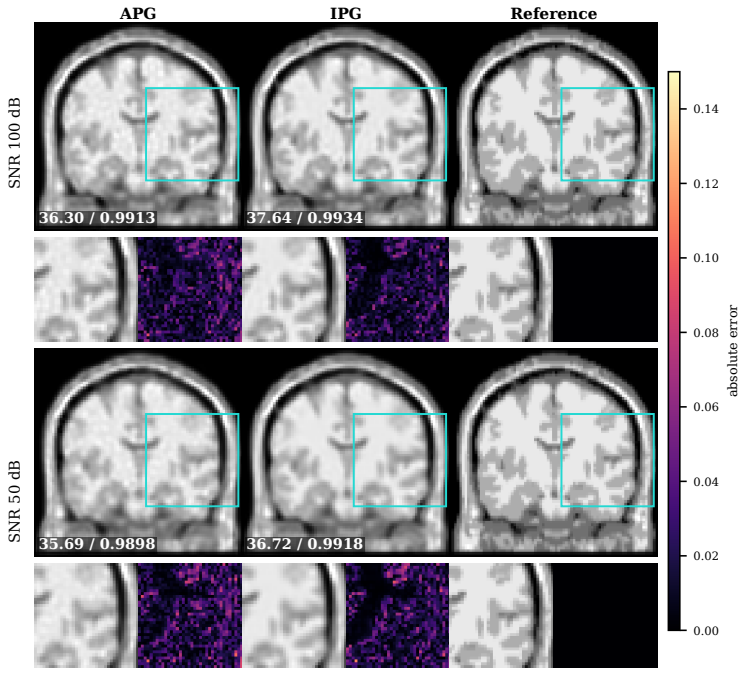



Figure 7: APG and IPG reconstructions for held-out subject s06 at 1024 radial spokes and 100-dB SNR (top) or 50-dB SNR (bottom). Columns show APG, IPG, and the reference. Central $x$–$z$ slices, ROI details, and absolute complex-error maps use matched locations and display scales. Numeric annotations report full-volume PSNR/SSIM.

### E.6 Sensitivity to learning rates

We compare both parameterizations on the same nine-setting learning-rate grid. The base rates for complex coefficients, centers, log-scales, and quaternions are $(3 \times 10^{-3}, 10^{-3}, 3 \times 10^{-3}, 3 \times 10^{-3})$. In addition to the base setting, we test coefficient-only multipliers $\{1/3, 3, 10\}$, log-scale-only multipliers $\{1/3, 3\}$, and joint multipliers $\{1/3, 3, 10\}$ for all parameter groups. Final PSNR is U-shaped for both arms.

Each diagnostic run uses the same cached prior, analytical forward model, TV weight $(0.0879)$, and stage-wise cosine schedule: 280 coefficient-only updates, 460 coefficient-and-center updates, and 2260 updates of all parameters. IPG achieves higher image PSNR in all nine paired settings and a higher best-observed PSNR when each parameterization is selected separately within this grid.

## F Evaluation Protocol and Baseline Configuration

### F.1 Development and evaluation split

Of the 20 BrainWeb anatomies, subject s54 is designated as the development subject. All tunable hyperparameters, regularization strengths, and optimization schedules are selected on this subject and then fixed before evaluation on the remaining 19 anatomies. The development subject is excluded from all cohort statistics and statistical comparisons.

### F.2 Simulation and acquisition protocol

All methods are evaluated using identical trajectories and noise realizations for each subject and acquisition condition. Each $128^3$ anatomy is sampled with a 15-coil array using 3D radial trajectories with 128 samples per spoke, at nominal acceleration factors of 64, 16, and 4.

Data generation follows the analytical phantom framework of Guerquin-Kern et al. (2012); Ngo et al. (2015). Each BrainWeb anatomy is converted into a continuous-domain phantom by extracting polygonal intensity-level contours from axial slices and linearly interpolating between the resulting contour-defined layers, following Ngo et al. (2015). This construction defines the object at arbitrary spatial coordinates and provides an analytical Fourier representation of the piecewise geometric model.

Coil sensitivities are generated from the Biot–Savart fields of 15 circular loops (Guerquin-Kern et al., 2012). Each loop has radius $0.55$ and is oriented toward the FOV center. Loop centers follow a Fibonacci distribution on a sphere of radius $2.0$, with independent uniform perturbations in $[-0.04, 0.04]$ along each coordinate. All lengths are expressed in units of the FOV half-width. A polynomial expansion is fitted to the resulting Biot–Savart fields for analytical data generation.

The multi-coil measurements are then evaluated from the analytical object and sensitivity representations without rasterizing their product. The simulation sensitivities are used only for data generation and are not provided to any reconstruction method.

### F.3 Noise model

Let $y_{c,m}^{\text{clean}}$ denote the clean acquired $k$-space samples and define their mean power by

$$P_y = \frac{1}{CM} \sum_{c=1}^{C} \sum_{m=1}^{M} \left| y_{c,m}^{\text{clean}} \right|^2 .$$

We add circular complex Gaussian noise

$$\varepsilon_{c,m} = \sigma \left( \eta_{c,m}^{\text{R}} + i \eta_{c,m}^{\text{I}} \right), \qquad \eta_{c,m}^{\text{R}}, \eta_{c,m}^{\text{I}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1),$$

with

$$\sigma^2 = \frac{P_y}{2\, 10^{\text{SNR}_{\text{dB}}/10}}.$$

Since $\mathbb{E}|\varepsilon_{c,m}|^2 = 2\sigma^2$, this corresponds to

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \frac{P_y}{\mathbb{E}|\varepsilon_{c,m}|^2}.$$

A single noise variance is used across all coils and trajectory locations. The reported SNR is therefore a global average over the acquired multi-coil $k$-space samples rather than a pointwise $k$-space SNR.

**Cartesian-equivalent SNR.** Let $P_{\text{rad}} = P_y$ be the mean clean power on the radial trajectory and $P_{\text{Cart}}$ the mean clean power, for the same object and coils, over a fully sampled $128^3$ Cartesian grid of the same FOV. For the same per-sample noise variance $\mathbb{E}|\varepsilon_{c,m}|^2 = P_{\text{rad}}\, 10^{-\text{SNR}_{\text{dB}}/10}$,

$$\text{SNR}_{\text{dB}}^{\text{Cart}} = 10 \log_{10} \frac{P_{\text{Cart}}}{\mathbb{E}|\varepsilon_{c,m}|^2} = \text{SNR}_{\text{dB}} - 10 \log_{10} \frac{P_{\text{rad}}}{P_{\text{Cart}}}, \qquad P_{\text{Cart}} = \frac{1}{C} \sum_{c=1}^{C} \sum_{j=1}^{N} \left| s_c(\boldsymbol{r}_j) \rho(\boldsymbol{r}_j) \right|^2,$$

where the expression of $P_{\text{Cart}}$ follows from Parseval's identity for the unnormalized DFT, with the voxel volume absorbed in $\rho$. Every spoke crosses the low-frequency lobe that carries most of the object's power, so the fraction of radial samples in this lobe scales as $N^{-1/3}$, versus $N^{-1}$ on the Cartesian grid. For our phantoms $P_{\text{rad}}/P_{\text{Cart}} \approx 37$ dB, independently of the number of spokes since every spoke has the same $\|\boldsymbol{k}\|$ profile, which gives $\text{SNR}^{\text{Cart}} \approx 13$ dB at $\text{SNR} = 50$ dB.

### F.4 Image-domain evaluation

Before evaluation, each complex reconstruction $\boldsymbol{x}$ is aligned to the reference $\boldsymbol{x}_{\text{ref}}$ using one global complex scalar,

$$\alpha_x^\star = \arg\min_{\alpha \in \mathbb{C}} \|\alpha \boldsymbol{x} - \boldsymbol{x}_{\text{ref}}\|_2^2 = \frac{\boldsymbol{x}^H \boldsymbol{x}_{\text{ref}}}{\boldsymbol{x}^H \boldsymbol{x}}, \qquad \widetilde{\boldsymbol{x}} = \alpha_x^\star \boldsymbol{x}.$$

This removes only the global complex scale ambiguity.

PSNR uses the complex mean-squared error and the peak reference magnitude,

$$\mathrm{PSNR} = 10 \log_{10} \frac{\left(\max_j |x_{\mathrm{ref},j}|\right)^2}{\frac{1}{N}\sum_{j=1}^{N} |\widetilde{x}_j - x_{\mathrm{ref},j}|^2}.$$

SSIM is computed in 3D between $|\widetilde{\boldsymbol{x}}|$ and $|\boldsymbol{x}_{\mathrm{ref}}|$, using the reference magnitude range as the data range.

### F.5 Parameter budgets

The IPG and raster-Gaussian models use $8000$ primitives. Each primitive has three center parameters, three log-scales, four quaternion parameters, and one complex coefficient, giving $8000 \times (3 + 3 + 4 + 2) = 96{,}000$ real trainable parameters.

HashINR uses a six-level hash-grid encoding with two features per level, a maximum table size of $2^{13}$ entries per level, base resolution 16, and per-level scale 1.35, followed by a two-hidden-layer, width-32 ReLU MLP with two real outputs; this configuration contains $92{,}160$ trainable real parameters. In B2B, each primitive contains ten shared geometric parameters and one complex coefficient for each of the $C = 15$ coils, giving $10 + 2C = 40$ real parameters per primitive. The parameter-matched B2B configuration therefore uses $2400 \times 40 = 96{,}000$ parameters, whereas the primitive-count-matched configuration uses $8000 \times 40 = 320{,}000$.

### F.6 Reconstruction hyperparameters

Table 4 lists the configurations used for the reported comparisons. IPG and raster Gaussian first optimize coefficients, then add centers, and finally add scales and rotations. Their initial scale is 1.3 voxels, with scale bounds $[0.512, 25.6]$ voxels. The learned representations use an adaptive Huber data loss whose threshold is $\alpha_H = 1$ times the median residual magnitude as described in Eq. 7. The APG and Gaussian-CSM ablations retain the IPG reconstruction settings, changing only the object coefficient convention or the sensitivity model ($K = 64$), respectively. Methods requiring explicit sensitivity maps use the same separately calibrated ESPIRiT estimates.

Table 4: Reconstruction hyperparameters. Settings are fixed after development on s54. Learning rates are initial values; iterative neural and Gaussian fits use Adam with cosine decay during each reconstruction stage. TV weights refer to each method's implemented loss normalization.

| Method | Configuration |
|---|---|
| iNUFFT | Density-compensated adjoint NUFFT; no iterative optimization. |
| SENSE–CG | For radial AF $(64, 16, 4)$, iteration counts are $(200, 100, 100)$ at 50-dB SNR and $(800, 600, 200)$ at 100-dB SNR. |
| CS–TV | 400 FISTA iterations; 20 Chambolle iterations per proximal step; SENSE–CG initialization. Relative TV weight: $(0.005, 0.001, 0.0002)$ for AF $(64, 16, 4)$. |
| HashINR | Six hash-grid levels, two features per level, table size at most $2^{13}$, base resolution 16, scale factor 1.35; two width-32 ReLU hidden layers and two real outputs. 600 prior-fitting updates, then 1200 full-data updates; learning rate $5 \times 10^{-3}$; raster-TV weight $5 \times 10^{-3}, (10^{-3}), 2 \times 10^{-4}$ for AF $(64, 16, 4)$. |
| Raster Gaussian | 8000 primitives; coefficient, center, log-scale, and quaternion learning rates $(3, 1, 3, 3) \times 10^{-3}$; stage lengths $(280, 460, 2260)$; 10% of acquired locations per update. Relative TV weight $(0.12, 0.0879, 0.0594)$ for AF $(64, 16, 4)$. |
| B2B | 2400 primitives (8000 in the additional radial comparison); complex coefficient, center, and covariance-factor learning rates $(3, 1, 6) \times 10^{-3}$; stage lengths $(280, 460, 2260)$ for coefficients, then isotropic geometry, then full covariance; 10% k-space batches. |
| IPG (proposed) | 8000 primitives; Fourier CSM with $L = 125$ modes per coil; coefficient, center, log-scale, and quaternion learning rates $(3, 1, 3, 3) \times 10^{-3}$; stage lengths $(280, 460, 2260)$; 10% k-space batches; Relative MC-TV weight $(0.1184, 0.0879, 0.0534)$ for AF $(64, 16, 4)$ and $N_{\mathrm{TV}} = 8192$ samples per update. |

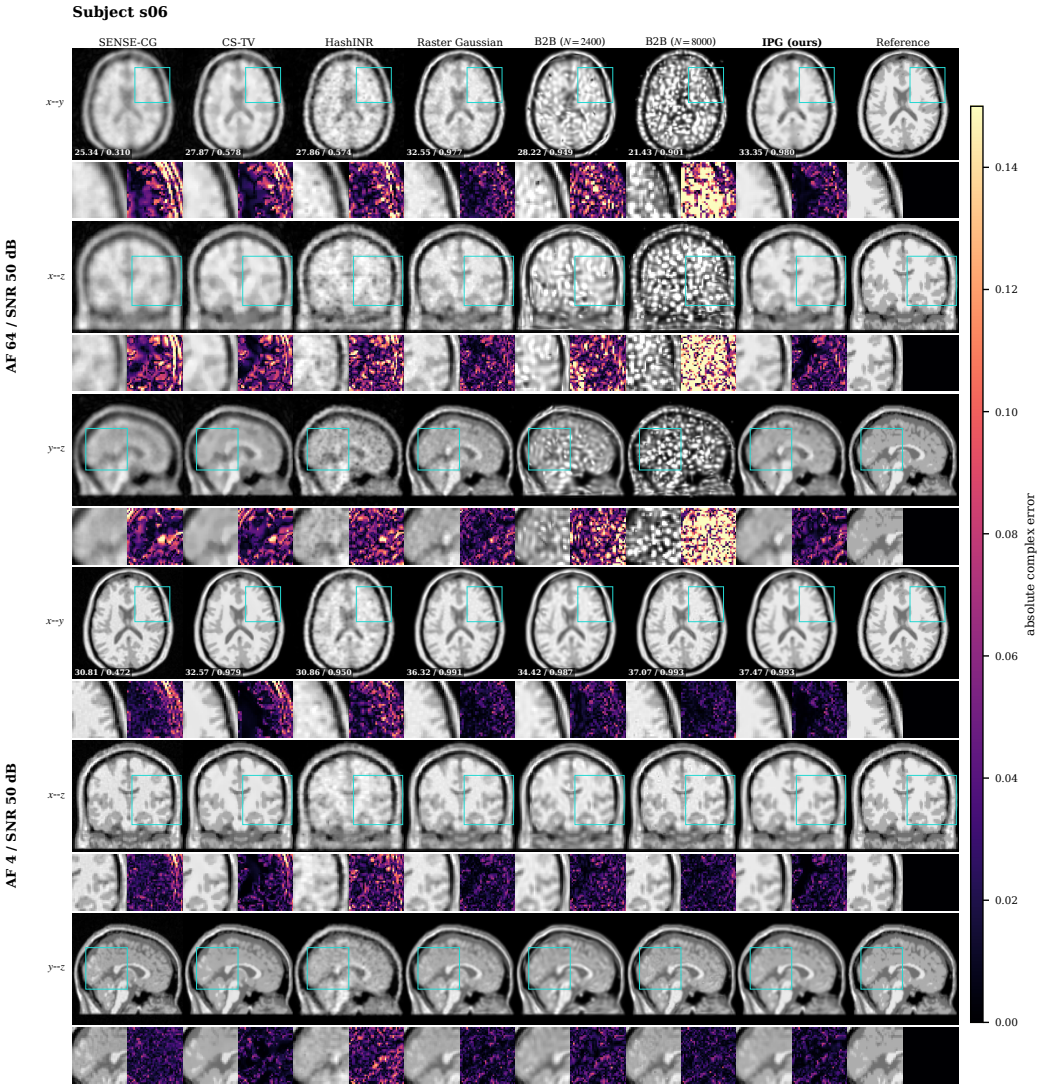



Figure 8: Additional qualitative comparison for held-out subject s06 at 50-dB SNR under AF 64 and AF 4 acquisitions. Columns show SENSE–CG, CS–TV, HashINR, the rasterized Gaussian baseline, parameter-matched B2B ($N = 2400$), primitive-count-matched B2B ($N = 8000$), IPG, and the reference. For each acquisition, matched central $x$–$y$, $x$–$z$, and $y$–$z$ views are shown with corresponding ROI details and absolute complex errors. Numeric annotations report full-volume PSNR/SSIM. The higher-capacity B2B configuration is less accurate than parameter-matched B2B at AF 64 in this example, but improves substantially at AF 4.

## G Additional Qualitative Reconstructions

Figures 8 and 9 provide additional qualitative comparisons on held-out evaluation subjects across orthogonal anatomical planes. We include both AF 64 and AF 4 acquisitions to complement the cohort-level results with a visual comparison of method behavior under strongly data-limited and data-rich regimes. Slice locations, ROIs, reconstruction display windows, and error scales are fixed across methods within each figure.

The two B2B configurations additionally illustrate a capacity–sampling interaction. B2B with $N = 8000$ uses the same number of primitives as IPG but substantially more trainable parameters because each primitive carries coil-specific coefficients. In the severely undersampled AF 64 example, this higher-capacity B2B model performs worse than its parameter-matched $N = 2400$ counterpart,

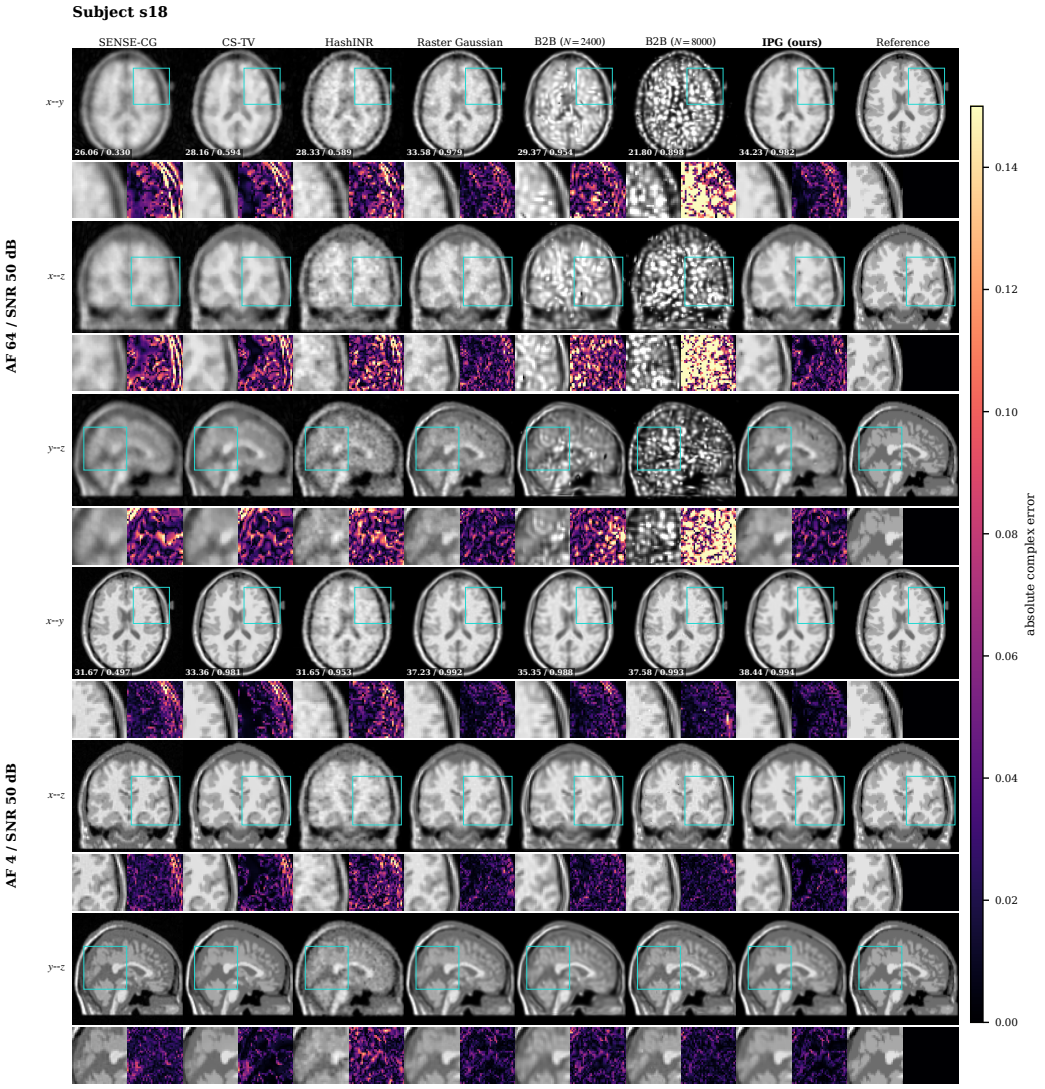



Figure 9: Additional qualitative comparison for held-out subject s18 at 50-dB SNR under AF 64 and AF 4 acquisitions. Same configuration as 8.

whereas at AF 4 its reconstruction improves substantially. This behavior is consistent with the larger set of coil-dependent degrees of freedom being more difficult to constrain when measurements are limited. We therefore report both B2B configurations throughout the comparison rather than treating primitive count and parameter count as interchangeable notions of model capacity.

## H Coil-Sensitivity Parameterization Study

We study the sensitivity representation independently of object reconstruction by fitting the Fourier and isotropic-Gaussian models to the same 15-coil calibration maps and evaluating both against the simulation-only sensitivities over the object support.

**Approximation accuracy and capacity.** The Fourier sensitivity model contains $L$ complex coefficients, or $2L$ real parameters, per coil. An isotropic-Gaussian mixture with $K$ components contains $6K$ real parameters per coil: three center coordinates, one scale, and one complex coefficient per component (Singh et al., 2024). Figure 5a reports relative complex $\ell_2$ error as these capacities vary.

At the reconstruction operating points, Fourier $L = 125$ uses 250 real parameters per coil and obtains relative CSM error $0.0358$, whereas the isotropic-Gaussian $K = 64$ model uses 384 parameters per coil and obtains error $0.0619$. Increasing the Gaussian model to $K = 125$ reduces its error to $0.0461$ but requires 750 real parameters per coil, remaining less accurate than Fourier $L = 125$ in this experiment.

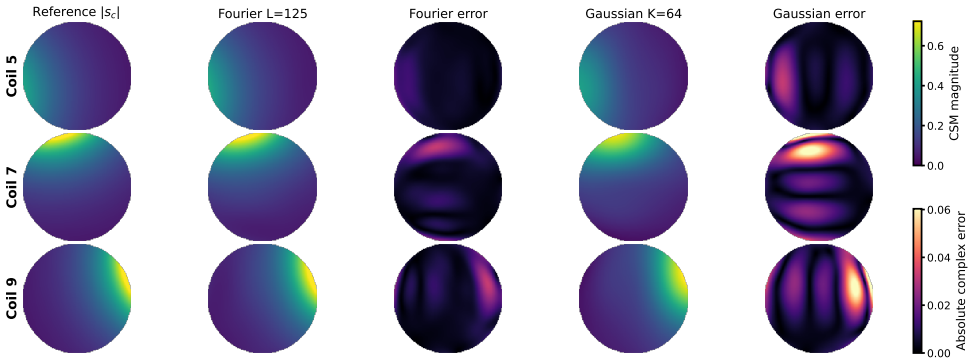



Figure 10: Central $x$–$y$ CSM magnitude and absolute complex error for three representative coils in the $128^3$, 15-coil simulation. Errors are computed against the simulation-only reference; magnitude and error maps use separate shared scales.

**Forward-model cost.** A Gaussian sensitivity representation also preserves analytical closure because the product of an object Gaussian and a sensitivity Gaussian is Gaussian. However, each coil-component–object-primitive pair produces a distinct Gaussian. Evaluating $M$ samples therefore has complexity $\mathcal{O}(LM(N_g + C))$ versus $\mathcal{O}(KMN_gC)$ in Singh et al. (2024) because the Fourier formulation evaluates the $N_g$-dependent shifted spectra once and subsequently combines them across coils, whereas the Gaussian-product form does not admit the same coil-wise factorization.

For the operating points used in Table 3, $N_g = 8000$, $L = 125$, $C = 15$, and $K = 64$. The primitive-dependent term therefore contains $LN_g = 1.00$ million primitive–mode interactions for the Fourier formulation, compared with $CKN_g = 7.68$ million primitive–coil-component interactions for the Gaussian model per sampled frequency. The corresponding complete reconstruction runs require 1382 s and 3837 s, respectively.

**Downstream reconstruction.** Figure 11 shows the corresponding reconstruction comparison on held-out subject s06. Both arms use the same IPG object representation and differ only in the fitted sensitivity model.

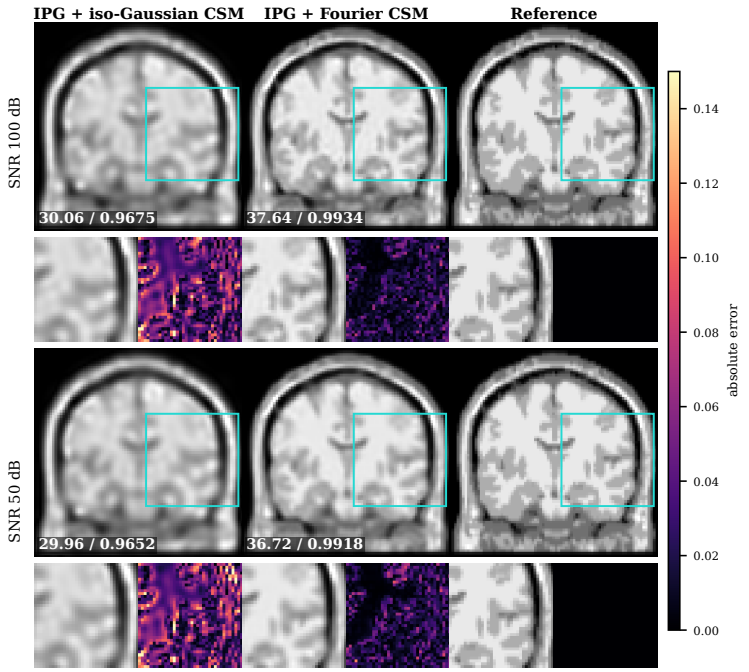



Figure 11: Fourier and isotropic-Gaussian sensitivity models for held-out subject s06 at radial AF 16 under 50- and 100-dB SNR. Both reconstructions use the same IPG object representation; only the coil-sensitivity parameterization differs. Numeric annotations report full-volume PSNR/SSIM.