COHERENT SURFACE GAUSSIANS FOR A SPECULAR SYNTHETIC APERTURE

Anonymous authors
Paper under double-blind review

ABSTRACT

Millimeter-wave radar sees through packaging, fabric and clutter, which makes it an attractive sense for robots, but its images are not shapes. We trace the reason to the optics of a synthetic aperture: each position is a co-located source and detector that receives glints only from surface points whose normals point back at it. Each sub-aperture therefore observes its own band of normals, back-projection images a convex surface behind its front, at a depth given in closed form, and the phase history determines a glint's center of curvature and its radius minus the radar's range offset, but not the two separately. We represent the scene by *coherent surface Gaussians*, curved primitives whose near-field echo has a closed form under a quadratic range approximation and whose angular response follows from their size and curvature. The closed form matches quadrature over the same primitives to within 1.4%, while flat Gaussians err by 24–98%. The same optics shows that holding out a sub-aperture, a natural analogue of novel-view synthesis, is a weak learning signal, because the other sub-apertures see different normals; in a ten-seed physical-optics study it removes phantoms only by stopping early and losing surface, whereas generic cross-validation selects models reliably. Multipath ghosts seen by all sub-apertures are virtual glints and cannot be rejected by any hold-out. An encoder trained in simulation reduces the phantom rate of back-projection by an order of magnitude on unseen shapes, and a measured 77–81 GHz scan shows the derived streak geometry.

1 INTRODUCTION

A robot reaching into a cardboard box, or an inspector checking a pipe behind a plastic panel, needs the *shape* of an object it cannot see. Millimeter-wave (mmWave) radar passes through cardboard, fabric and most plastics, works in the dark, and is cheap enough to mount on a manipulator. Sweeping a small antenna over a rail synthesizes a large aperture, from which three-dimensional images can be formed (Sheen et al., 2001; Hamidi, 2024).

These images, however, are not shapes. Figure 1(b) shows a target scanned by a 77–81 GHz rail radar: back-projection (BP) of the full aperture yields an hourglass, a narrow waist followed by a long, bright fan that looks like solid structure. When the aperture is split into five horizontal strips, each strip images the target as a streak along its own line of sight (Figure 1c). The streaks turn with the strip, like shadows cast by a moving light, and all of them pass through the waist. We call this behavior *aperture parallax*.

A synthetic aperture is closer to a wall of lamps, each with a detector beside it, than to a camera. At millimeter wavelengths most objects act as mirrors: each position receives *glints* from surface points whose normals point back at it, and each strip sees a different band of orientations. BP fuses these glints into streaks that can meet behind the surface. Their depth depends on angular diversity relative to bandwidth.

A second difficulty is less visible. Every radar has a constant delay in its cables and analog chain, and a curved glint delayed in this way produces exactly the echo of a glint with a smaller radius. The phase of the echo therefore determines the center of curvature and the radius *minus* the delay, but not the two separately; the amplitude carries a trace of the radius, but too weak to be read without an exact model of the antenna and the material. On our scan the strip peaks lie 2–3 cm beyond

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.

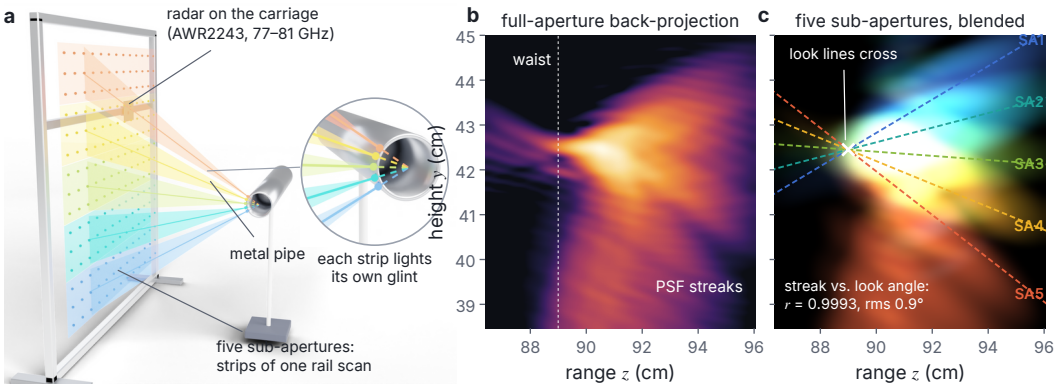

Figure 1: **A synthetic aperture is a wall of lamps that sees glints.** (a) Each rail position transmits and receives; each of five strips sees a band of surface normals (schematic, not to scale). (b) Measured full-aperture BP of a target at 0.9 m: a narrow waist before a bright fan. (c) Five measured strip images lie along their look lines, which cross at the waist. The crossing alone does not identify the true surface (Theorem 1 and Section 6).

the crossing of their lines of sight, which is what such a delay produces; without one reference measurement, neither the radius nor the true surface can be read from the data.

Image-domain enhancement and Gaussian splatting start from BP images that already mix shape, the point spread function and depth bias (Smith & Torlak, 2023; Li et al., 2025; Liu et al., 2026). Measurement-domain methods such as GeRaF (Lu et al., 2025) and coherent point splatting (Armouti et al., 2026b) instead fit the echoes. We share that premise, but represent the scene by curved, complex-valued *coherent surface Gaussians* whose near-field echo and directional response follow from their size and curvature, rather than from a learned lobe. Section 2 gives the fuller comparison.

We first analyze which surface elements each strip observes, where BP places a glint, and what phase can identify (Section 3). We then derive a differentiable coherent renderer (Section 4) and ask how its geometry can be selected without labels (Section 5). Holding out a strip initially appears to be the radar analogue of novel-view synthesis. The analysis and a controlled study show why it is weak: the other strips often see different normals, so strip hold-out removes phantoms by stopping early and losing real surface. Generic cross-validation over positions is more reliable. A smooth multipath ghost seen by every strip is a virtual glint and survives either split. We test the renderer, these selection rules and an amortized encoder in simulation, and compare the optics with a measured scan (Section 6).

Our contributions are:

- An analysis of the synthetic aperture as a co-located specular sensor. It restates classical glint geometry and Gaussian-beam footprints, gives the apparent depth of a convex glint in closed form, and makes explicit that the phase history determines geometry only up to a uniform offset along the normals, which one calibration target removes; we show numerically that the amplitude does not remove it in practice (Section 3).
- Coherent surface Gaussians, curved primitives whose near-field echo has a closed form under a quadratic range approximation with controlled error, and whose angular lobe is derived rather than learned (Section 4).
- An analysis of held-out prediction for a specular aperture: why holding out a band of normals is a weak signal, when it helps, and, using the classical virtual-reflector construction, what no hold-out can check; and an amortized encoder with test-time refinement (Section 5).
- A renderer benchmark against quadrature, point and flat primitives; a ten-seed physical-optics study of strip hold-out against generic cross-validation and calibrated stopping rules; an encoder study on seen and unseen shape families; and an analysis of a measured 77–81 GHz scan (Section 6).

2 Related Work

Specular apertures. The link between a co-located source, glints and the surface distance function is classical in specular flow, confocal non-line-of-sight imaging and zero-offset seismic migration (Roth & Black, 2006; Xin et al., 2019; Tygel et al., 2012). In mmWave, mmNorm (Dodds et al., 2025) votes normals from the antenna with the strongest return; wide-angle SAR uses sub-apertures to study limited-persistence scattering (Moses et al., 2004; Ertin et al., 2007). Coherence factors and spatially variant apodization suppress image pixels on which apertures disagree (Li & Li, 2003; Stankwitz et al., 1995). We use the same optical cue to state what a glint's phase identifies and what held-out measurements can check.

Radar scene representations. Optical Gaussian splatting has a closed-form image footprint (Kerbl et al., 2023; Zwicker et al., 2001); surface variants use disks or quadrics (Huang et al., 2024; Zhang et al., 2025). SAR-GS fits intensity images (Li et al., 2025). mmIR synthesizes coherent samples from LiDAR geometry with multiple bounces (Armouti et al., 2026a); 3D point splatting uses oriented points and learned material parameters (Armouti et al., 2026b). GeRaF, closest to our setting, recovers geometry per object with a phase-preserving near-field renderer and learned reflectiveness (Lu et al., 2025; 2026). Our distinction is an explicit curved primitive with a closed-form coherent echo and a derived angular lobe, together with amortized inference and an analysis of calibration and hold-out. A method-by-method comparison and further literature are in Appendix F.

Held-out prediction and multipath. Noise2Self and SSDU learn by predicting held-out measurements (Batson & Royer, 2019; Yaman et al., 2020); the aperture split here changes which normals remain observable. Hagedoorn's envelope construction explains why a smooth zero-offset multipath branch can masquerade as a virtual reflector (Hagedoorn, 1954; Weglein et al., 2003). We use that classical result to delimit what any held-out split can reject.

3 The Specular Optics of a Synthetic Aperture

Gaussian splatting inherits its structure from the camera, in which every pixel integrates along one ray. A synthetic aperture has no pixels. Each position is a source and detector at the same place, receiving mirror-like *glints*. We ask what this sensor measures, which strips observe a surface, where BP places a glint, and what its echo can identify.

Measurement and glints. At wavenumber $k = 2\pi f/c$, a transmitter at $\boldsymbol{p}_t$ and receiver at $\boldsymbol{p}_r$ measure, under single scattering,

$$y(\boldsymbol{p}_t, \boldsymbol{p}_r, k) = e^{-\mathrm{i}2k\Delta} \int \sigma(\boldsymbol{x})\, a(\boldsymbol{x}, \boldsymbol{p}_t, \boldsymbol{p}_r, k)\, e^{-\mathrm{i}k(\|\boldsymbol{x}-\boldsymbol{p}_t\|+\|\boldsymbol{x}-\boldsymbol{p}_r\|)}\, d\boldsymbol{x} + n. \tag{1}$$

Here σ is reflectivity, a collects slowly varying antenna, spreading and obliquity terms, n is noise, and Δ is the constant range offset of the system. For metals, this is the physical-optics approximation (Knott et al., 2004). With unit look directions $\boldsymbol{d}_t, \boldsymbol{d}_r$ separated by angle β, the local phase probes along the bisector $\boldsymbol{h}$ (Kell, 1965):

$$\boldsymbol{\kappa} = k(\boldsymbol{d}_t + \boldsymbol{d}_r) = 2k\cos(\beta/2)\boldsymbol{h}. \tag{2}$$

For co-located antennas it probes along the line of sight at $2k$. A large smooth surface returns chiefly from points whose normals face that direction. Classical distance-function relations recover a glint's normal and curvature from its range history (Proposition 4 and Appendix A) (Xin et al., 2019; Tygel et al., 2012). Splitting the aperture into strips adds no information, since BP is linear and the strip images sum (Lemma 2); it exposes which positions support which structure.

Which strips observe an element. In the geometric limit a strip sees the surface points whose normal rays reach it. At finite frequency the assignment blurs. A small patch diffracts broadly; a curved patch spreads its reflection across directions. For a patch with tangent covariance $\boldsymbol{\Lambda}$ and range Hessian $\mathbf{H}$ (Equation (6)), these contributions add:

Lemma 1 (Aperture footprint of a surface element)**.** *To second order in range, the angular-lobe covariance is*

$$\boldsymbol{\Gamma}^{-1} = (2k)^{-2}\boldsymbol{\Lambda}^{-1} + \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}, \tag{3}$$

and its local footprint on a planar aperture at range R has covariance $R^2\mathbf{\Gamma}^{-1}$.

This is Gaussian-beam optics (Kogelnik & Li, 1966; Červený et al., 1982). The first term is diffraction and the second curvature. The fraction of a patch's response energy in strip ℓ, its *lobe share* $w_{j\ell}$, tells how many strips can check it. Numerical limits and overlap rates are in Appendix B.

Where BP places a glint. Consider a convex surface with center of curvature $\boldsymbol{c}$ and radius r. Its glint for a position $\boldsymbol{p}$ lies on the line from $\boldsymbol{p}$ to $\boldsymbol{c}$, one radius in front of $\boldsymbol{c}$, so its range is the distance to $\boldsymbol{c}$ minus r. Across the aperture, the range therefore *varies* like that of a point at $\boldsymbol{c}$, while its *value* is that of the surface. Back-projection has to reconcile two cues that disagree: the variation across positions points to $\boldsymbol{c}$, and the delay, measured across the band, points to the surface. A narrow strip has little angular diversity, so the delay wins and the strip image peaks on the surface, stretched along the line of sight (Figure 2b,c). The full aperture has much angular diversity and a fractional bandwidth of only 5%, so the variation wins and the image peaks near $\boldsymbol{c}$, well behind the surface (Figure 2a,c). This is the hourglass of Figure 1. The GPR hyperbola fit (Shihab & Al-Nuaimy, 2005), the normal-incidence-point wave of seismics (Höcht et al., 1999; Jäger et al., 2001; Dell & Gajewski, 2011) and canonical radar scattering models (Jackson et al., 2010) rest on the same split between variation and delay. Where back-projection places the peak in between, however, we have not found stated; for comparison, a convex mirror seen in optical stereo appears midway between surface and center (Oren & Nayar, 1997).

The following proposition gives the position. It measures depth δ from $\boldsymbol{c}$ along a search direction $\boldsymbol{n}$, usually the mean line of sight, and weighs the angular diversity of the aperture, through the variance v of the cosine between the lines of sight and $\boldsymbol{n}$, against the relative spread s_k/μ of the wavenumbers. It is stated for an *apparent radius* ϱ; for a calibrated radar, $\varrho = r$, and Section 3 explains why the general case is $\varrho = r - \Delta$.

Proposition 1 (Apparent depth of a convex glint)**.** *Let m, v be the weighted mean and variance of the cosine between the look direction and a depth-search direction, and μ, s_k^2 those of k. For factorized aperture and frequency weights, the BP magnitude peaks, to second order in phase cumulants, at*

$$\delta^\star = \frac{s_k^2 m \varrho}{s_k^2 m^2 + (\mu^2 + s_k^2) v} \;\approx\; \frac{\varrho}{1+\zeta}, \qquad \zeta = \frac{\mu^2 v}{s_k^2}, \tag{4}$$

measured from $\mathbf{c}$ along the search direction.

The ratio ζ compares angular diversity with bandwidth. A narrow strip favors the surface; a wide aperture favors the center. For a 2 cm pipe at 77–81 GHz, the formula puts the peak 16.7 mm behind the front surface and exact 2D BP gives 16.4 mm. The approximation holds while the range profile has one peak; the bandwidth transition and physical-optics comparison are in Figure 2 and Appendix B.

What phase identifies. A constant offset adds the same delay to every glint. For the convex glint, measured range is distance to $\boldsymbol{c}$ minus ϱ: its phase cannot separate r from Δ.

Theorem 1 (What a monostatic phase history identifies)**.** *Under single scattering with unknown complex reflectivity, $(\mathcal{S}, \Delta)$ and the parallel surface $\mathcal{S}_\Delta = \{\boldsymbol{x} - \Delta \boldsymbol{n}(\boldsymbol{x})\}$ at offset zero have the same glint phase histories wherever the parallel-surface map is an immersion. A circular glint is identified through its center $\mathbf{c}$ and apparent radius $\varrho = r - \Delta$; one range reference determines r.*

This is the classical geometry of parallel surfaces (Federer, 1959), applied here to the aperture. Relative amplitude contains a radius cue, but is ill-conditioned under modest pattern or gain uncertainty (Appendix K). A point, a convex glint and a concave reflector with suitable offsets can thus share the phase history. A target at known range can fix Δ once per setup. Narrow-strip look lines locate $\boldsymbol{c}$ only under the conditions of the theorem; the measured strips are wide and their BP phase reference also affects the crossing (Section 6 and Appendix H).

4 Coherent Surface Gaussians

The optics calls for a coherent primitive with curvature and a directional response determined by geometry. A flat patch of any size predicts the wrong strips for a curved reflector; a free lobe

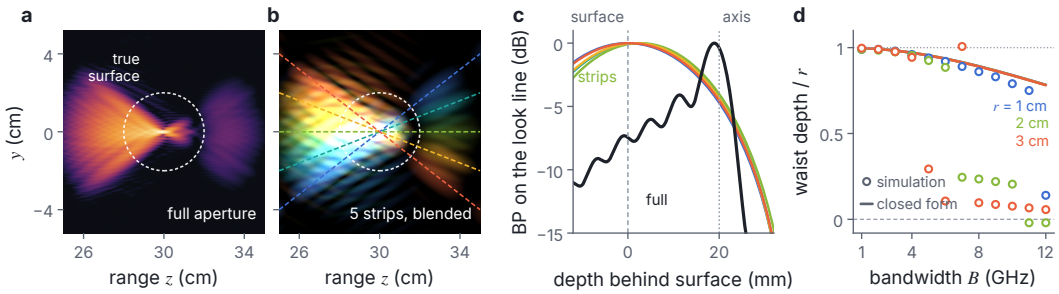

Figure 2: **Where back-projection puts a glint**, simulated pipe ($r = 2\,\text{cm}$, dashed; 0.6 m rail at 0.3 m, 77–81 GHz, noise and calibration errors, $\Delta = 0$; Appendix I). Range z runs to the right. (a) Full-aperture hourglass. (b) Sub-aperture images, blended in the colors of Figure 1, lie along their look lines, which cross at the axis. (c) Along each look line, each strip peaks at the surface, the full aperture 18.9 mm behind it (black; closed-form value 19.4 mm). (d) Waist depth versus bandwidth. The closed form (lines) assumes one peak; once $c/2B \lesssim 1.3\,r$ (above about 12, 6 and 4 GHz for $r = 1$, 2 and 3 cm), the profile along the look line splits into a surface peak and an axis peak, and the simulated maximum (circles) jumps to the surface.

could fit the strips without locating the surface. We therefore use a curved Gaussian patch with a differentiable near-field echo.

Primitive. A coherent surface Gaussian G_j has center $\boldsymbol{\mu}_j$, tangent frame $\mathbf{U}_j$, sensor-facing normal $\boldsymbol{n}_j$, tangent covariance $\boldsymbol{\Lambda}_j$, symmetric curvature $\mathbf{K}_j$ and complex reflectivity ρ_j. Its surface and weight are

$$\boldsymbol{x}_j(\boldsymbol{u}) = \boldsymbol{\mu}_j + \mathbf{U}_j\boldsymbol{u} + \tfrac{1}{2}\boldsymbol{n}_j\boldsymbol{u}^\top\mathbf{K}_j\boldsymbol{u}, \qquad g_j(\boldsymbol{u}) = \exp(-\tfrac{1}{2}\boldsymbol{u}^\top\boldsymbol{\Lambda}_j^{-1}\boldsymbol{u}). \tag{5}$$

The front of a pipe has $\mathbf{K}_j \prec 0$. Reflectivity is a superposition of these weighted surface patches, inserted directly into equation 1; no independent angular lobe is attached.

Closed-form echo. The echo of one patch is an integral of a Gaussian weight times a phase $e^{-\mathrm{i}2kR}$ over the patch. In optical splatting, EWA obtains a closed-form footprint by linearizing the projection around the Gaussian's center (Zwicker et al., 2001). The coherent analogue is to expand the round-trip range to second order: a Gaussian weight times the exponential of a quadratic is again a Gaussian, and its integral is known in closed form. Three geometric quantities of the range, seen from one aperture position, are all the expansion needs. The *distance* R_0 sets the phase. The *range gradient* $\boldsymbol{q}$ along the patch measures how obliquely the patch is seen; it vanishes exactly when the sensor is in the specular direction, and its norm is the sine of the incidence angle. The *range Hessian* $\mathbf{H}$ measures how quickly the phase varies across the patch; it adds the bending of the spherical wavefront, which is what makes the problem near-field, to the bending of the surface itself.

At the patch center, with look direction $\boldsymbol{d} = (\boldsymbol{\mu}_j - \boldsymbol{p})/R_0$, its gradient and Hessian are

$$\boldsymbol{q} = \mathbf{U}_j^\top\boldsymbol{d}, \qquad \mathbf{H} = \frac{\mathbf{U}_j^\top(\mathbf{I} - \boldsymbol{d}\boldsymbol{d}^\top)\mathbf{U}_j}{R_0} + (\boldsymbol{n}_j^\top\boldsymbol{d})\mathbf{K}_j. \tag{6}$$

The two Hessian terms are wavefront and surface curvature. Let a_0 and $\boldsymbol{g}$ be the value and log-gradient of the slowly varying amplitude at $\boldsymbol{u}_0$.

Theorem 2 (Closed-form echo under a quadratic range approximation)**.** *Keeping range to second order and amplitude to first order, the patch contributes*

$$\psi_j(\boldsymbol{p}, k) = \rho_j a_0(1 + \boldsymbol{g}^\top\bar{\boldsymbol{v}})\frac{2\pi}{\sqrt{\det\mathbf{M}}}\exp\!\big(-\mathrm{i}2kR_0 - \tfrac{1}{2}\boldsymbol{u}_0^\top\boldsymbol{\Lambda}_j^{-1}\boldsymbol{u}_0 + \tfrac{1}{2}\boldsymbol{\eta}^\top\mathbf{M}^{-1}\boldsymbol{\eta}\big)\,, \tag{7}$$

where $\mathbf{M} = \boldsymbol{\Lambda}_j^{-1} + \mathrm{i}2k\mathbf{H}$, $\boldsymbol{\eta} = \boldsymbol{\Lambda}_j^{-1}\boldsymbol{u}_0 + \mathrm{i}2k\boldsymbol{q}$ *and* $\bar{\boldsymbol{v}} = -\mathbf{M}^{-1}\boldsymbol{\eta}$.

The proof is a complex Gaussian integral, as in Gaussian Wave Splatting (Choi et al., 2025); the result is the coherent analogue of the EWA footprint (Zwicker et al., 2001) and a surface counterpart

of Gaussian-beam migration (Hill, 1990). Each factor has a plain reading: $e^{-\mathrm{i}2kR_0}$ is the delay to the expansion point, $2\pi/\sqrt{\det \overline{\mathbf{M}}}$ is the effective area of the patch as seen through the phase variation across it, and the last two exponents shift the stationary point from $\boldsymbol{u}_0$ to where the patch is actually specular.

Where to expand. A second-order expansion is accurate only near the point that dominates the integral. For a flat patch seen head-on this is the center, but for an oblique curved patch the dominant point is the specular point, which can lie far from the center; expanding at the center then makes the cubic remainder large. The renderer therefore moves the expansion point toward the saddle point of the integrand by two steps of $\boldsymbol{u}_0 \leftarrow \boldsymbol{u}_0 + \mathrm{Re}\ \bar{\boldsymbol{v}}$. Against quadrature with the *exact* range and amplitude, the error is then at most 1.1% up to 20° incidence and 4.0% at 30°, compared with 9% and 39% when expanding at the center (Appendix B). A bound ε_j on the cubic remainder (Lemma 4) flags patches that are too large for the expansion, and these are split.

Derived lobe and renderer. Taking the magnitude of equation 7 at the center gives a Gaussian in $\boldsymbol{q}$ with $\boldsymbol{\Gamma}_j = 4k^2\,\mathrm{Re}(\mathbf{M}^{-1})$ and $\boldsymbol{\Gamma}_j^{-1}$ of equation 3 (Corollary 2). At 79 GHz a flat 5 mm patch has a 3.6° lobe; on a 2 cm pipe at 0.3 m it has a 15.9° lobe. Size and curvature therefore decide which strips see it. The predicted measurement is the coherent sum

$$\hat{y}(\boldsymbol{p}, k) = e^{-\mathrm{i}2k\Delta} \sum_j v_j(\boldsymbol{p}) \psi_j(\boldsymbol{p}, k), \tag{8}$$

where v_j accounts for back-facing and occluded patches. The renderer holds Δ fixed at the setup level, culls pairs outside the derived lobe, and splits patches whose range remainder is too large. Visibility, the bistatic form and numerical details are in Appendix C.

5 Learning with a Specular Aperture

Held-out measurements. Fitting all echoes is insufficient: more primitives can absorb noise, residual rail phase and multipath, so the loss falls while geometry worsens. Without geometric labels, we can score a model on positions it did not fit. For folds $\mathcal{A}_f$ of the aperture, let y_{-f} omit fold f and $\mathcal{F}_f$ render at its positions:

$$\mathcal{L}_{\mathrm{HO}}(\theta) = \mathbb{E}_{\mathrm{scene},f} \left\| y_f - \mathcal{F}_f(f_\theta(y_{-f})) \right\|^2. \tag{9}$$

With zero-mean noise independent across folds, this equals the noise-free prediction error plus a constant (Proposition 5) (Batson & Royer, 2019; Yaman et al., 2020). It certifies prediction of measurements, not correctness of geometry. Correlated hardware errors weaken even that guarantee.

What a strip can check. A held-out fold can be predicted only from structure that the remaining folds also observe. For a spread-out partition this is almost all structure: every fold samples every band of normals, and a surface element is seen by all folds whenever it is seen at all. Generic cross-validation therefore charges what it has always charged, primitives that fit the noise of particular positions, and it charges little else.

For strip hold-out, what matters is how the response of each surface element is shared between strips. The *lobe share* $w_{j\ell}$ of Section 3 is the fraction of the response energy of primitive j that falls in strip ℓ. Linearizing the renderer around a scene estimate makes the consequence exact: the held-out residual of strip ℓ is its in-sample residual amplified by $(\mathbf{I}-\mathbf{P}_{\ell\ell})^{-1}$, where, for orthogonal responses, $\mathbf{P}_{\ell\ell}$ acts on primitive j by $w_{j\ell}$ (Proposition 6, Appendix A). Structure with $\max_\ell w_{j\ell} \le w < 1$ is predicted from the other strips, with its residual amplified by at most $(1-w)^{-1}$. Structure with $w_{j\ell} \to 1$ is not predicted at all: when its strip is held out, nothing in the other strips asks for it. This applies to a phantom fitted to the noise of one strip, and to a ghost that only one strip sees, which strip hold-out correctly declines to credit. It applies equally to a *real* surface element seen by one strip only.

The same optics that motivated strip hold-out now limits it. By Lemma 1, strips see disjoint bands of normals as soon as the surface elements are larger than a few wavelengths or only mildly curved. For a 2 cm pipe at 0.3 m, 65–83% of the observed surface lies in the lobes of at least two of five strips, but for a flat plate as little as 16% (Appendix B), and most primitives fitted to realistic scenes have

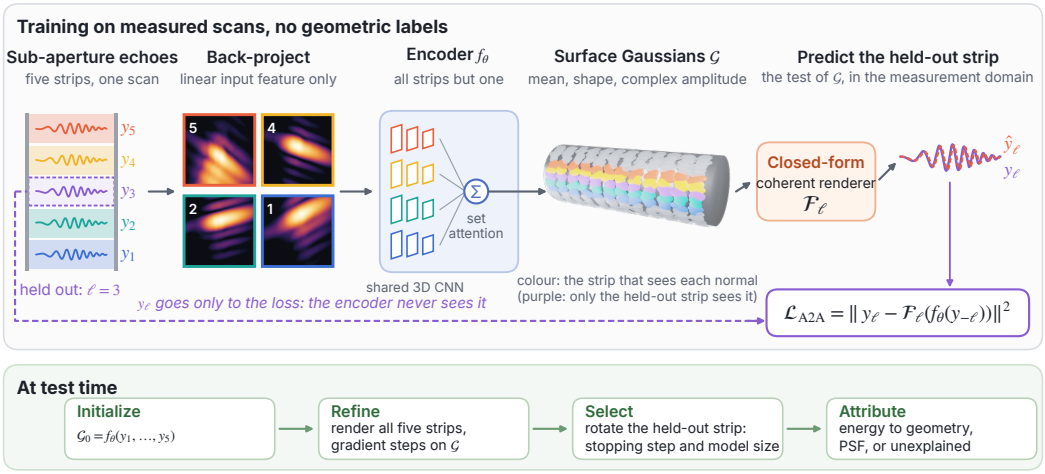

Figure 3: **Overview of SpecGS.** Each strip is back-projected to a voxel grid, a linear input feature; a shared encoder with set attention predicts coherent surface Gaussians, colored by the strip that sees each normal. The closed-form renderer, which models the PSF and each primitive's lobe, predicts the measurements of any set of aperture positions, and geometry is judged only by these predictions. The figure shows the held-out set as one strip ($\ell = 3$); Sections 5 and 6 show that for model selection a random or interleaved held-out set works better, because a strip sees normals that the other strips do not.

a lobe share above 0.9 (Section 6). For such scenes the held-out strip is barely predictable from the others by *any* geometry, including the true one. The held-out error then decreases little, or not at all, as the model improves, and a stopping rule built on it stops early: it removes phantoms, but together with real surface that only one strip sees. The analysis thus predicts that strip hold-out is informative where lobes overlap, on small or strongly curved elements, and weak elsewhere, and that generic cross-validation is the more reliable criterion for model selection. The lobe share remains useful in a different role: it tells, for every fitted primitive, how many strips support it, and hence which parts of a reconstruction rest on one band of look directions only.

What no hold-out can check. The range offset Δ appears in every fold and cannot be selected by held-out prediction. A smooth multipath branch can also remain consistent across all strips: its range history can be the distance function of a virtual surface.

Proposition 2 (Phase-equivalent virtual glint; restated)**.** *For a smooth monostatic-equivalent multipath range $R_g(\boldsymbol{p})$ over an aperture plane, $\|\nabla R_g\| \leq 1$; where the inequality is strict and the envelope map is regular, R_g is the distance function of a virtual surface.*

This is Hagedoorn's classical envelope construction (Hagedoorn, 1954), given explicitly in Appendix A. If every strip observes the ghost and its virtual surface lies in the model class, no split can reject it. Occlusion or finite reflectors can truncate a ghost's support, making it detectable. The controlled cases are in Section 6 and Appendix J.

Encoder and refinement. Each strip is back-projected onto a voxel grid and carrier-demodulated as a linear input feature. A shared encoder combines these volumes and their look directions with set attention (Lee et al., 2019), and a voxel-aligned head predicts curved primitives with existence logits. Simulated physical-optics scenes provide geometric supervision; we also study strip and random-fold losses on their own. At test time the encoder initializes a short measurement-domain refinement, with random-fold cross-validation selecting the stopping point. The architecture and the attribution of image energy to geometry, PSF and residual are in Appendices C, D and F.

Table 1: **Renderer accuracy and speed** (monostatic rail, 61×61 positions, 32 frequencies; 5 mm primitives). Error: normalized complex error with respect to physical-optics quadrature over the true surface. In parentheses: error with respect to dense quadrature over the same Gaussians, which isolates the closed form from the tiling of the surface by Gaussians. Time for the pipe on one CPU thread of a shared machine; only ratios between rows are reliable. Primitive counts are for the pipe (sphere 458, plate 144).

Renderer	#prim.	Pipe, $r = 2\,\mathrm{cm}$	Sphere, $r = 3\,\mathrm{cm}$	Plate, tilted	Time (s)
Dense quadrature on the Gaussians	400	9.1%	1.4%	15.8%	531
Points at the Gaussian centers	400	277%	362%	194%	1.2
Dense points at $\lambda/4$	11 305	0.4%	0.2%	0.7%	12.0
Flat Gaussians ($\mathbf{K} = 0$), re-centered	400	50%	70%	15.8%	7.0
Curved, center expansion	400	9.0% (1.4%)	2.3% (1.4%)	15.8% (0.1%)	5.8
Curved, re-centered	400	9.1% (0.4%)	1.3% (0.4%)	15.8% (0.1%)	7.4
Curved, re-centered + lobe culling	400	9.1%	1.3%	15.8%	1.6

6 Experiments

Setup. A TI AWR2243 radar scans at 77–81 GHz on a 0.95×0.99 m rail, about 0.9 m from the target; the aperture is divided into five strips. We analyze measured complex BP volumes of a pipe-like target and a corner reflector, acquired without moving the radar between them (Appendix H). Controlled tests use separate two- and three-dimensional physical-optics simulators rather than our renderer, avoiding an inverse crime (Kaipio & Somersalo, 2007). We report response-energy-weighted phantom rate, the fraction of fitted primitive energy farther than τ from a true surface, and visible-surface recall within τ; full protocols and confidence intervals are in Appendices F, I and J.

Q1: Does a measured scan follow the optics? Each measured strip streak follows its mean look direction (correlation 0.9993, rms angular difference 0.9° over a 46° span), and the five look lines cross within 1.7 mm at the waist (Figure 1 and Appendix H). The strips are wide in the other rail dimension: their diversity ratios are $\zeta_\ell \approx 3.9$–5.5, against 9.9 for the full aperture. With these values, Proposition 1 gives a full-to-strip offset ratio of 0.51–0.60; the scan gives 0.61. The processing library also uses different frequency references for range and aperture phase; a replica of its BP shows that this shifts the look-line crossing, so on this scan the crossing is not exactly the center of curvature. The best single-glint fit has apparent radius $\varrho \approx -5.5$ cm, reproducing strip offsets within 4 mm and the full-aperture peak within 3 mm. This is an *apparent* radius: the scan alone does not separate target radius, system offset and a virtual glint. The corner-reflector scan gives a qualitative check of the offset sign; acquisition and fit details are in Appendix H.

Q2: Is the renderer accurate? Against dense quadrature over the same Gaussians, the re-centered closed form has normalized complex error of 0.4% on a pipe and sphere and 0.1% on a plate. Against the true physical-optics surface, total errors are 9.1%, 1.3% and 15.8%, respectively: tiling and hard edges, rather than the closed form, dominate. Flat Gaussians of the same size err by 50–70% on the curved objects; center points err by 277–362%. Lobe culling skips 85–87% of primitive–position pairs, making rendering 4–6 times faster with less than 0.1% added error. The full renderer table and exact-range checks are in Table 1 and Appendix B.

Q3: Does strip hold-out select better geometry? In ten-seed two-dimensional physical-optics tests, all selection rules stop the same greedy path and differ only in the chosen step. The held-out strip error rises from 1.10 to 1.59 as primitives are added, where 1 means predicting nothing, while random-fold error falls from 0.82 to 0.05. Strip hold-out stops at a median of three steps against 11–16 for other rules. It lowers pooled phantom rate by 0.022 (Holm-corrected $p = 0.009$), but recall falls by 0.41 ($p < 10^{-4}$). A fixed early stop at the typical strip-held-out step performs comparably ($p = 0.78$ for phantom rate; $p = 0.67$ for recall), so the strip split contributes a smaller model rather than scene-adaptive selection.

The reduction in phantoms costs real surface: 39–41% of correctly placed response energy in plate scenes comes from primitives seen almost entirely by one strip. The third-order path is fitted as a mirror image of the pipe by the nondegenerate rules, and in the strong dihedral every rule fails

Table 2: **Model selection in 2D physical optics** (ten seeds; means; confidence intervals in Appendix J). PR_E: response-weighted phantom rate; recall at $\tau = 2\,\mathrm{mm}$. All rules stop the same curved-Gaussian fit. HO: strip hold-out; CV-r: random five-fold cross-validation; ES: fixed early stop.

	Pipe + plate		+ third-order path		Dihedral	
Rule	PR_E	Recall	PR_E	Recall	PR_E	Recall
HO (strips)	0.007	0.58	0.026	0.29	0.614	0.41
CV-r	0.025	1.00	0.064	1.00	0.622	0.50
ES	0.004	0.52	0.003	0.26	0.617	0.42

Table 3: **Amortized reconstruction in 2D** on test scenes from the training families (seen) and from a held-out family of ellipses (unseen). CD: Chamfer distance on the visible surface (mm). Time per scene on one CPU core.

	Seen families			Unseen family			
Method	$\mathrm{PR}_E \downarrow$	Recall $\uparrow$	CD $\downarrow$	$\mathrm{PR}_E \downarrow$	Recall $\uparrow$	CD $\downarrow$	Time (s)
Back-projection, $-6\,\mathrm{dB}$ region	0.88	0.97	9.6	0.90	0.96	10.4	0.02
Per-scene greedy fit, CV-r stop	0.03	0.96	1.1	0.05	0.93	1.4	85
Encoder, $\mathcal{L}_{\mathrm{geo}}$	0.07	0.88	1.9	0.12	0.79	2.8	0.04
Encoder, $\mathcal{L}_{\mathrm{HO}}$ strips only	0.41	0.35	7.8	0.46	0.31	8.5	0.04
Encoder, $\mathcal{L}_{\mathrm{HO}}$ random folds only	0.22	0.61	4.6	0.27	0.55	5.3	0.04
Encoder, $\mathcal{L}_{\mathrm{geo}}$, 3 of 5 strips	0.10	0.74	2.5	0.15	0.66	3.4	0.03
Encoder, $\mathcal{L}_{\mathrm{geo}}$ + refinement	0.03	0.95	1.1	0.05	0.92	1.5	3.2

($\mathrm{PR}_E = 0.61$–0.63), in line with Proposition 2. Full baselines, lobe-share analysis and uncertainty are in Appendix J.

Q4: Does amortized inference help? On simulated scenes, the encoder trained with geometry reduces BP's phantom rate from 0.88 to 0.07 on seen families and from 0.90 to 0.12 on held-out ellipses, in 0.04 s per scene. Three seconds of refinement reaches 0.03 and 0.05, close to per-scene fitting at 85 s. Strip-held-out loss alone yields 0.41 and 0.46, confirming that it lacks a useful geometry signal; random-fold loss is better but remains below geometric supervision. The complete table, losses and shifted-domain test are in Table 3 and Appendix F.

Q5: Measured objects. With visible-surface evaluation on measured objects, BP plus iso-surface extraction has 9.8 mm Chamfer distance and a 0.71 phantom rate; GeRaF has 3.4 mm and 0.19. Our per-object fit with random-fold stopping has 2.9 mm and 0.12, and the encoder plus refinement 2.6 mm and 0.09. The full baseline, normal, F-score and weak-target-recall results are in Table 4 and Appendix F.

7 Discussion and Conclusion

A near-field mmWave synthetic aperture is a co-located specular sensor. Its sub-apertures see bands of normals rather than interchangeable views of one surface. Three consequences follow: calibrate range offset before interpreting curvature (Theorem 1); reconstruct in the measurement domain with curvature-bearing primitives (Theorem 2); and use random or interleaved folds to select model complexity when strips have little lobe overlap (Section 6). A multipath branch visible in all folds can still masquerade as a virtual glint (Proposition 2).

The analysis and renderer assume single scattering, and the optical argument is monostatic. The strong dihedral test shows the resulting failure under dominant double bounce. The measured scan supplies parallax in one plane; its streak geometry agrees with the analysis, while the object's true radius remains tied to the independent range reference and the processing phase convention. A multistatic aperture can distinguish traversal orders of some double bounces, extending the information available to the renderer (Armouti et al., 2026a).

Table 4: **Evaluation on measured objects** (visible-surface metrics, mean over 3 seeds). CD: Chamfer distance (mm); NC: normal consistency; F-score at 2 mm; CV: random-fold held-out error; PR: phantom rate; WR: weak-target recall. *Best of three (Table 12).

	Method	CD ↓	NC ↑	F-score ↑	CV ↓	PR ↓	WR ↑
image	BP + iso-surface	9.8	0.62	0.21	–	0.71	0.35
	Nonlinear sub-aperture fusion*	7.5	0.66	0.30	–	0.55	0.28
	Image Gaussians, $\mathbf{\Sigma} + \mathbf{P}_\ell$ rendering	6.9	0.70	0.34	–	0.48	0.40
	SAR-GS-style intensity renderer	7.8	0.64	0.29	–	0.58	0.33
measurement	GeRaF	3.4	0.83	0.62	0.41	0.19	0.55
	mmNorm	5.1	0.79	0.47	–	0.30	0.45
	SPECGS, per object, strip hold-out stop	3.3	0.85	0.55	0.38	0.07	0.42
	SPECGS, per object, CV-r stop	2.9	0.86	0.68	0.36	0.12	0.61
	SPECGS, encoder only	3.6	0.82	0.58	0.44	0.15	0.52
	SPECGS, encoder + refinement	2.6	0.87	0.71	0.34	0.09	0.64

REFERENCES

Yair Adato, Yuriy Vasilyev, Ohad Ben-Shahar, and Todd E. Zickler. Toward a theory of shape from specular flow. In *IEEE International Conference on Computer Vision (ICCV)*, pp. 1–8, 2007. doi: 10.1109/ICCV.2007.4408883.

Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics*, 34(6), 2015. doi: 10.1145/2816795.2818072.

Adnan Armouti, Yixuan Gao, and Rajalakshmi Nandakumar. mmIR: Frequency-space inverse rendering for 3D millimeter-wave radar ADC synthesis. In *European Conference on Computer Vision (ECCV)*, 2026a. arXiv:2608.28913.

Adnan Armouti, Yixuan Gao, and Rajalakshmi Nandakumar. 3D point splatting for mmWave radar novel view synthesis. *arXiv preprint arXiv:2609.11894*, 2026b.

Gaurav Bagwe, Xiaoyong Yuan, Yongji Wu, and Lan Zhang. DiffRadar: Differentiable physics-aware radar SLAM with Gaussian fields. *arXiv preprint arXiv:2607.12265*, 2026.

Joshua Batson and Loïc Royer. Noise2Self: Blind denoising by self-supervision. In *International Conference on Machine Learning (ICML)*, 2019.

Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *European Conference on Computer Vision (ECCV)*, 2020. doi: 10.1007/978-3-030-58580-8_18.

Norman Bleistein and Richard A. Handelsman. *Asymptotic Expansions of Integrals*. Dover, 1986.

David Borts, Erich Liang, Tim Brödermann, Andrea Ramazzina, Stefanie Walz, Edoardo Palladin, Jipeng Sun, David Bruggemann, Christos Sakaridis, Luc Van Gool, Mario Bijelic, and Felix Heide. Radar fields: Frequency-space neural scene representations for FMCW radar. In *ACM SIGGRAPH Conference Papers*, 2024. doi: 10.1145/3641519.3657510.

Jorge Camacho, Montserrat Parrilla, and Carlos Fritsch. Phase coherence imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 56(5):958–974, 2009. doi: 10.1109/TUFFC.2009.1128.

V. Červený, M. M. Popov, and I. Pšenčík. Computation of wave fields in inhomogeneous media — Gaussian beam approach. *Geophysical Journal of the Royal Astronomical Society*, 70(1): 109–128, 1982. doi: 10.1111/j.1365-246X.1982.tb06394.x.

Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, et al. ShapeNet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*, 2015.

Brian Chao, Jacqueline Yang, Suyeon Choi, Manu Gopakumar, Ryota Koiso, and Gordon Wetzstein. Random-phase wave splatting of translucent primitives for computer-generated holography. *arXiv preprint arXiv:2508.17480*, 2025.

David Charatan, Sizhe Lester Li, Andrea Tagliasacchi, and Vincent Sitzmann. pixelSplat: 3D Gaussian splats from image pairs for scalable generalizable 3D reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Yuedong Chen, Haofei Xu, Chuanxia Zheng, Bohan Zhuang, Marc Pollefeys, Andreas Geiger, Tat-Jen Cham, and Jianfei Cai. MVSplat: Efficient 3D Gaussian splatting from sparse multi-view images. In *European Conference on Computer Vision (ECCV)*, 2024.

Suyeon Choi, Brian Chao, Jacqueline Yang, Manu Gopakumar, and Gordon Wetzstein. Gaussian wave splatting for computer-generated holography. *ACM Transactions on Graphics (SIGGRAPH)*, 2025.

Joong H. Chun and Chester A. Jacewitz. Fundamentals of frequency domain migration. *Geophysics*, 46(5):717–733, 1981. doi: 10.1190/1.1441211.

Jon F. Claerbout. *Imaging the Earth's Interior*. Blackwell Scientific Publications, Oxford, 1985.

Sergius Dell and Dirk Gajewski. Common-reflection-surface-based workflow for diffraction imaging. *Geophysics*, 76(5):S187–S195, 2011. doi: 10.1190/geo2010-0229.1.

Laura Dodds, Tara Boroushaki, Kaichen Zhou, and Fadel Adib. Non-line-of-sight 3D object reconstruction via mmWave surface normal estimation. In *ACM International Conference on Mobile Systems, Applications and Services (MobiSys)*, 2025. doi: 10.1145/3711875.3729138.

Laura Dodds, Maisy Lam, Waleed Akbar, Yibo Cheng, and Fadel Adib. Wave-Former: Through-occlusion 3D reconstruction via wireless shape completion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21713–21724, 2026. arXiv:2511.14152.

Emre Ertin, Christian D. Austin, Samir Sharma, Randolph L. Moses, and Lee C. Potter. GOTCHA experience report: three-dimensional SAR imaging with complete circular apertures. In *Algorithms for Synthetic Aperture Radar Imagery XIV*, volume 6568 of *Proc. SPIE*, pp. 656802, 2007. doi: 10.1117/12.723245.

Ningjing Fan, Yiqun Wang, Dong-Ming Yan, and Peter Wonka. Ref-DGS: Reflective dual Gaussian splatting. In *ACM SIGGRAPH Conference Papers*, 2026. doi: 10.1145/3799902.3811163. arXiv:2603.07664.

R. T. Farouki and C. A. Neff. Analytic properties of plane offset curves. *Computer Aided Geometric Design*, 7(1–4):83–99, 1990. doi: 10.1016/0167-8396(90)90023-K.

Herbert Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3): 418–491, 1959. doi: 10.1090/S0002-9947-1959-0110078-1.

L. Ferro-Famil, A. Reigber, E. Pottier, and W.-M. Boerner. Scene characterization using subaperture polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(10):2264–2276, 2003. doi: 10.1109/TGRS.2003.817188.

Iraklis Giannakis, Feng Zhou, Craig Warren, and Antonios Giannopoulos. On the limitations of hyperbola fitting for estimating the radius of cylindrical targets in nondestructive testing and utility detection. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. doi: 10.1109/LGRS.2022.3195947.

David Gilbarg and Neil S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Classics in Mathematics. Springer, 2001. doi: 10.1007/978-3-642-61798-0.

J. G. Hagedoorn. A process of seismic reflection interpretation. *Geophysical Prospecting*, 2(2): 85–127, 1954. doi: 10.1111/j.1365-2478.1954.tb01281.x.

Shahrokh Hamidi. 3D near-field virtual MIMO-SAR imaging using FMCW radar systems at 77 GHz. In *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2024. arXiv:2208.14328.

Glenn Healey and Thomas O. Binford. Local shape from specularity. *Computer Vision, Graphics, and Image Processing*, 42(1):62–86, 1988. doi: 10.1016/0734-189X(88)90143-0.

N. Ross Hill. Gaussian beam migration. *Geophysics*, 55(11):1416–1428, 1990. doi: 10.1190/1.1442788.

G. Höcht, E. de Bazelaire, P. Majer, and P. Hubral. Seismics and optics: hyperbolae and curvatures. *Journal of Applied Geophysics*, 42(3-4):261–281, 1999. doi: 10.1016/S0926-9851(99)00040-3.

K. W. Hollman, K. W. Rigby, and M. O'Donnell. Coherence factor of speckle from a multi-row probe. In *Proceedings of the IEEE Ultrasonics Symposium*, volume 2, pp. 1257–1260, 1999. doi: 10.1109/ULTSYM.1999.849225.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2D Gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH Conference Papers*, 2024.

Julie Ann Jackson, Brian D. Rigling, and Randolph L. Moses. Canonical scattering feature models for 3D and bistatic SAR. *IEEE Transactions on Aerospace and Electronic Systems*, 46(2):525–541, 2010. doi: 10.1109/TAES.2010.5461639.

Rainer Jäger, Jürgen Mann, German Höcht, and Peter Hubral. Common-reflection-surface stack: Image and attributes. *Geophysics*, 66(1), 2001. doi: 10.1190/1.1444927.

Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. GaussianShader: 3D gaussian splatting with shading functions for reflective surfaces. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5322–5332, 2024. doi: 10.1109/CVPR52733.2024.00509.

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.

R. E. Kell. On the derivation of bistatic RCS from monostatic measurements. *Proceedings of the IEEE*, 53(8):983–988, 1965. doi: 10.1109/PROC.1965.4077.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023.

Eugene F. Knott, John F. Shaeffer, and Michael T. Tuley. *Radar Cross Section*. SciTech Publishing, 2nd edition, 2004.

Herwig Kogelnik and Tingye Li. Laser beams and resonators. *Applied Optics*, 5(10):1550–1567, 1966. doi: 10.1364/AO.5.001550.

Pou-Chun Kung, Skanda Harisha, Ram Vasudevan, Aline Eid, and Katherine A. Skinner. RadarSplat: Radar Gaussian splatting for high-fidelity data synthesis and 3D reconstruction of autonomous driving scenes. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

Muyinatu A. Lediju, Gregg E. Trahey, Brett C. Byram, and Jeremy J. Dahl. Short-lag spatial coherence of backscattered echoes: Imaging characteristics. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 58(7):1377–1388, 2011. doi: 10.1109/TUFFC.2011.1957.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International Conference on Machine Learning (ICML)*, 2019.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. In *International Conference on Machine Learning (ICML)*, 2018.

Zhengxin Lei, Feng Xu, Jiangtao Wei, Feng Cai, Feng Wang, and Ya-Qiu Jin. SAR-NeRF: Neural radiance fields for synthetic aperture radar multiview representation. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

Michael Leigsnering, Fauzia Ahmad, Moeness Amin, and Abdelhak Zoubir. Multipath exploitation in through-the-wall radar imaging using sparse reconstruction. *IEEE Transactions on Aerospace and Electronic Systems*, 50(2):920–939, 2014. doi: 10.1109/TAES.2013.120528.

Aobo Li, Zhengxin Lei, Jiangtao Wei, and Feng Xu. SAR-GS: Gaussian splatting based SAR images rendering and target reconstruction. *arXiv preprint arXiv:2506.21633*, 2025.

Pai-Chi Li and Meng-Lin Li. Adaptive imaging using the generalized coherence factor. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 50(2):128–141, 2003. doi: 10.1109/TUFFC.2003.1182117.

David B. Lindell, Gordon Wetzstein, and Matthew O'Toole. Wave-based non-line-of-sight imaging using fast f-k migration. *ACM Transactions on Graphics*, 38(4):116:1–116:13, 2019. doi: 10.1145/3306346.3322937.

Z. Liu, S. Niu, X. Qiu, L. Peng, Y. Shang, L. Zhong, and C. Ding. A differentiable method for novel view SAR image generation via 3D Gaussian splatting. *ISPRS Journal of Photogrammetry and Remote Sensing*, 231:167–195, 2026.

D. Loewenthal, L. Lu, R. Roberson, and J. Sherwood. The wave equation applied to migration. *Geophysical Prospecting*, 24(2):380–399, 1976. doi: 10.1111/j.1365-2478.1976.tb00934.x.

Jiachen Lu, Hailan Shanbhag, and Haitham Al Hassanieh. GeRaF: Neural geometry reconstruction from radio frequency signals. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Jiachen Lu, Hailan Shanbhag, and Haitham Al Hassanieh. Seeing through boxes: Non-line-of-sight 3D reconstruction from radar signals. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Raoul Mallart and Mathias Fink. Adaptive focusing in scattering media through sound-speed inhomogeneities: The van Cittert Zernike approach and focusing criterion. *The Journal of the Acoustical Society of America*, 96(6):3721–3732, 1994. doi: 10.1121/1.410562.

Zoubaida Mechbal and Abdellatif Khamlichi. Determination of concrete rebars characteristics by enhanced post-processing of GPR scan raw data. *NDT & E International*, 89:30–39, 2017. doi: 10.1016/j.ndteint.2017.03.005.

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.

Randolph L. Moses, Lee C. Potter, and Müjdat Çetin. Wide-angle SAR imaging. In Edmund G. Zelnio and Frederick D. Garber (eds.), *Algorithms for Synthetic Aperture Radar Imagery XI*, volume 5427 of *Proc. SPIE*, pp. 164–175, 2004. doi: 10.1117/12.544935.

Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H. Kim. Practical SVBRDF acquisition of 3D objects with unstructured flash photography. *ACM Transactions on Graphics*, 37(6), 2018. doi: 10.1145/3272127.3275017.

Clifford J. Nolan and Margaret Cheney. Microlocal analysis of synthetic aperture radar imaging. *Journal of Fourier Analysis and Applications*, 10(2):133–148, 2004. doi: 10.1007/s00041-004-8008-0.

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.

Michael Oren and Shree K. Nayar. A theory of specular surface geometry. *International Journal of Computer Vision*, 24(2):105–124, 1997. doi: 10.1023/A:1007954719939.

Matthew O'Toole, David B. Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555:338–341, 2018. doi: 10.1038/nature25489.

Lee C. Potter and Randolph L. Moses. Attributed scattering centers for SAR ATR. *IEEE Transactions on Image Processing*, 6(1):79–91, 1997. doi: 10.1109/83.552098.

Yvain Quéau, Bastien Durix, Tao Wu, Daniel Cremers, François Lauze, and Jean-Denis Durou. LED-based photometric stereo: Modeling, calibration and numerical solution. *Journal of Mathematical Imaging and Vision*, 60(3):313–340, 2018. doi: 10.1007/s10851-017-0761-1.

Eric Todd Quinto. Singularities of the X-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$. *SIAM Journal on Mathematical Analysis*, 24(5):1215–1225, 1993. doi: 10.1137/0524069.

Stefan Roth and Michael J. Black. Specular flow and the recovery of surface structure. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1869–1876, 2006. doi: 10.1109/CVPR.2006.290.

George T. Ruck, Donald E. Barrick, William D. Stuart, and Clarence K. Krichbaum. *Radar Cross Section Handbook*. Plenum Press, New York, 1970.

Chi Hyung Seo and Jesse T. Yen. Sidelobe suppression in ultrasound imaging using dual apodization with cross-correlation. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 55(10):2198–2210, 2008. doi: 10.1109/TUFFC.919.

Pawan Setlur, Moeness Amin, and Fauzia Ahmad. Multipath model and exploitation in through-the-wall and urban radar sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10): 4021–4034, 2011. doi: 10.1109/TGRS.2011.2128331.

David M. Sheen, Douglas L. McMakin, and Thomas E. Hall. Three-dimensional millimeter-wave imaging for concealed weapon detection. *IEEE Transactions on Microwave Theory and Techniques*, 49(9):1581–1592, 2001.

S. Shihab and W. Al-Nuaimy. Radius estimation for cylindrical objects detected by ground penetrating radar. *Subsurface Sensing Technologies and Applications*, 6(2):151–166, 2005. doi: 10.1007/s11220-005-0004-1.

Josiah W. Smith and Murat Torlak. Deep learning-based multiband signal fusion for 3-D SAR super-resolution. *arXiv preprint arXiv:2305.02017*, 2023.

H. C. Stankwitz, R. J. Dallaire, and J. R. Fienup. Nonlinear apodization for sidelobe control in SAR imagery. *IEEE Transactions on Aerospace and Electronic Systems*, 31(1):267–279, 1995. doi: 10.1109/7.366309.

R. H. Stolt. Migration by Fourier transform. *Geophysics*, 43(1):23–48, 1978. doi: 10.1190/1.1440826.

Nithin Sugavanam and Emre Ertin. Differentiable Gaussian splatting models of limited persistence scattering in SAR backscatter measurements. In *IEEE International Radar Conference (RADAR)*, 2025.

M. Tygel, B. Ursin, E. Iversen, and M. V. de Hoop. Estimation of geological dip and reflector curvature from zero-offset seismic reflections in heterogeneous anisotropic media. *Studia Geophysica et Geodaetica*, 56:521–531, 2012. doi: 10.1007/s11200-011-0256-2.

Kush R. Varshney, Müjdat Çetin, John W. Fisher, III, and Alan S. Willsky. Sparse representation in structured dictionaries with application to synthetic aperture radar. *IEEE Transactions on Signal Processing*, 56(8):3548–3561, 2008. doi: 10.1109/TSP.2008.919392.

Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5491–5500, 2022. doi: 10.1109/CVPR52688.2022.00541.

D. J. Verschuur, A. J. Berkhout, and C. P. A. Wapenaar. Adaptive surface-related multiple elimination. *Geophysics*, 57(9):1166–1177, 1992. doi: 10.1190/1.1443330.

L. Villard and P. Borderies. On the use of virtual ground scatterers to localize double and triple bounce scattering mechanisms for bistatic SAR. *Journal of Electromagnetic Waves and Applications*, 29(5):626–635, 2015. doi: 10.1080/09205071.2015.1012594.

Arthur B. Weglein, Fernanda V. Araújo, Paulo M. Carvalho, Robert H. Stolt, Kenneth H. Matson, Richard T. Coates, Dennis Corrigan, Douglas J. Foster, Simon A. Shaw, and Haiyan Zhang. Inverse scattering series and seismic exploration. *Inverse Problems*, 19(6):R27–R83, 2003. doi: 10.1088/0266-5611/19/6/R01.

Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan, and Ioannis Gkioulekas. A theory of Fermat paths for non-line-of-sight shape reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6793–6802, 2019. doi: 10.1109/CVPR.2019.00696.

Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kâmil Uğurbil, and Mehmet Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic Resonance in Medicine*, 84(6):3172–3191, 2020.

Keyang Ye, Qiming Hou, and Kun Zhou. 3D gaussian splatting with deferred reflection. In *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–10, 2024. doi: 10.1145/3641519.3657456.

R. Yelf. Where is true time zero? In *Proceedings of the Tenth International Conference on Ground Penetrating Radar (GPR 2004)*, volume 1, pp. 279–282. IEEE, 2004.

Öz Yilmaz. *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*. Number 10 in Investigations in Geophysics. Society of Exploration Geophysicists, Tulsa, OK, 2001. doi: 10.1190/1.9781560801580.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R. Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

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.

Kai Zhang, Fujun Luan, Zhengqi Li, and Noah Snavely. IRON: Inverse rendering by optimizing neural SDFs and materials from photometric images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5565–5574, 2022. doi: 10.1109/CVPR52688.2022.00548.

Ziyu Zhang, Binbin Huang, Hanqing Jiang, Liyang Zhou, Xiaojun Xiang, and Shuhan Shen. Quadratic Gaussian splatting: High quality surface reconstruction with second-order geometric primitives. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

Chengke Zhu and Hongxia Ye. A modular method for GPR hyperbolic feature detection and quantitative parameter inversion of underground pipelines. *Remote Sensing*, 15(8):2114, 2023. doi: 10.3390/rs15082114.

Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross. EWA volume splatting. In *Proceedings of IEEE Visualization*, 2001.

ETHICS STATEMENT

Millimeter-wave sensing can see through clothing, packaging and thin walls, which raises privacy concerns if applied to people without consent. The reported scans contain only inanimate objects captured in a laboratory.

REPRODUCIBILITY STATEMENT

All derivations are given in Appendix A; the exact-range validation of Theorem 2, lobe-overlap analysis and PSF simulation are described in Appendix B. Renderer, encoder, simulation and training details are in Appendices C, D, I and J.

USE OF LARGE LANGUAGE MODELS

An AI assistant was used extensively in preparing this paper: for literature search and verification of references, for derivations that the authors checked, for simulation and analysis code, for drafting and editing text, and for figures. The authors reviewed all content and take full responsibility for it.

A PROOFS

Throughout, $\hat{f}(\boldsymbol{\kappa}) = \int f(\boldsymbol{x}) e^{-\mathrm{i}\boldsymbol{\kappa}^\top \boldsymbol{x}} d\boldsymbol{x}$ denotes the Fourier transform, and we use the Gaussian integral

$$\int_{\mathbb{R}^n} \exp\big(-\tfrac{1}{2}\boldsymbol{v}^\top \mathbf{A}\boldsymbol{v} - \mathrm{i}\boldsymbol{\eta}^\top \boldsymbol{v}\big)\, d\boldsymbol{v} \;=\; \frac{(2\pi)^{n/2}}{\sqrt{\det \mathbf{A}}} \exp\big(-\tfrac{1}{2}\boldsymbol{\eta}^\top \mathbf{A}^{-1}\boldsymbol{\eta}\big), \tag{10}$$

valid for complex symmetric $\mathbf{A}$ with positive definite real part and $\boldsymbol{\eta} \in \mathbb{C}^n$, with the branch of $\sqrt{\det \mathbf{A}}$ obtained by analytic continuation from real positive definite $\mathbf{A}$ (Bleistein & Handelsman, 1986).

A.1 SUB-APERTURE IMAGES

With a fixed weighting w, the back-projection image of sub-aperture ℓ is

$$I_\ell(\boldsymbol{x}) \;=\; \mathcal{B}_\ell[y_\ell](\boldsymbol{x}) \;=\; \sum_{\boldsymbol{p}\in\mathcal{A}_\ell} \sum_k w(\boldsymbol{p}, k)\, y(\boldsymbol{p}, k)\, e^{+\mathrm{i}2k\|\boldsymbol{x}-\boldsymbol{p}\|}. \tag{11}$$

Lemma 2 (Linear decomposition)**.** *For any disjoint partition of $\mathcal{A}$ and a common weighting w, $I_{\mathcal{A}}(\boldsymbol{x}) = \sum_{\ell=1}^{S} I_\ell(\boldsymbol{x})$.*

Proof. Insert $\mathcal{A} = \bigsqcup_\ell \mathcal{A}_\ell$ into equation 11; the double sum splits into a sum over ℓ of the sub-aperture sums, each of which is I_ℓ by definition. □

Hence no *linear* recombination of sub-aperture images improves on the full image. Nonlinear recombinations, such as the pixelwise minimum of $|I_\ell|$ or the coherence factor (Mallart & Fink, 1994; Li & Li, 2003), can suppress streaks present in only some sub-apertures and serve as baselines. On the measured data, the five sub-aperture volumes sum to the full-aperture volume up to the measurements shared at block boundaries (Appendix H).

A.2 SUB-APERTURE PSF

Assumption 1 (Local Gaussian spectral support)**.** Over $\mathcal{A}_\ell \times [k_{\min}, k_{\max}]$, the weighted spatial frequencies $\boldsymbol{\kappa}$ are well approximated by a Gaussian with mean $\bar{\boldsymbol{\kappa}}_\ell = 2\bar{k}\,\hat{\bar{\boldsymbol{u}}}_\ell$ and covariance $\mathbf{C}_\ell = \sigma_r^2\, \hat{\bar{\boldsymbol{u}}}_\ell \hat{\bar{\boldsymbol{u}}}_\ell^\top + (2\bar{k})^2\, \boldsymbol{\Sigma}_{\theta,\ell}$, where $\hat{\bar{\boldsymbol{u}}}_\ell$ is the mean look direction, σ_r the spread of $2k$ induced by the bandwidth, and $\boldsymbol{\Sigma}_{\theta,\ell}$ the covariance of $\hat{\boldsymbol{u}}(\boldsymbol{p})$ on the plane orthogonal to $\hat{\bar{\boldsymbol{u}}}_\ell$.

Assumption 1 is exact for Gaussian tapering in frequency and aperture and is a second-moment approximation otherwise. In the near field, $\hat{\boldsymbol{u}}(\boldsymbol{p})$ depends on $\boldsymbol{x}_0$, so the assumption and everything that follows are local: the PSF varies across the scene.

Lemma 3 (Sub-aperture PSF)**.** *Under Assumption 1, the PSF of sub-aperture ℓ is $h_\ell(\boldsymbol{\delta}) \propto e^{\mathrm{i}\bar{\boldsymbol{\kappa}}_\ell^\top \boldsymbol{\delta}} \exp\big(-\tfrac{1}{2}\boldsymbol{\delta}^\top \mathbf{C}_\ell \boldsymbol{\delta}\big)$. Its magnitude is a Gaussian with covariance $\mathbf{P}_\ell = \mathbf{C}_\ell^{-1}$, whose principal axes are $\hat{\bar{\boldsymbol{u}}}_\ell$ and its orthogonal complement. For uniform weighting in frequency and angle, the ratio of its extent along the look direction to its extent across it is*

$$\frac{\rho_\parallel}{\rho_\perp} \;\approx\; \frac{c/(2B)}{\bar{\lambda}/(2\Delta\theta_\ell)} \;=\; \frac{f_c}{B}\,\Delta\theta_\ell, \tag{12}$$

where $\Delta\theta_\ell$ is the angular span of the sub-aperture seen from $\boldsymbol{x}_0$ and f_c the center frequency; other tapers change both extents by similar constants. Hence the PSF rotates with the mean look direction of the sub-aperture, and it is elongated along *the look direction exactly when $\Delta\theta_\ell > B/f_c$.*

Proof of Lemma 3. For a unit point scatterer at $\boldsymbol{x}_0$, equation 1 and equation 11 give $I_\ell(\boldsymbol{x}) = \sum_{\boldsymbol{p},k} w(\boldsymbol{p},k)\, a\, e^{\mathrm{i}2k(\|\boldsymbol{x}-\boldsymbol{p}\|-\|\boldsymbol{x}_0-\boldsymbol{p}\|)}$. Linearizing the range, $\|\boldsymbol{x}-\boldsymbol{p}\| - \|\boldsymbol{x}_0-\boldsymbol{p}\| \approx \hat{\boldsymbol{u}}(\boldsymbol{p})^\top \boldsymbol{\delta}$ with $\hat{\boldsymbol{u}}(\boldsymbol{p}) = (\boldsymbol{x}_0-\boldsymbol{p})/\|\boldsymbol{x}_0-\boldsymbol{p}\|$, the phase is $\boldsymbol{\kappa}(\boldsymbol{p},k)^\top\boldsymbol{\delta}$ with $\boldsymbol{\kappa} = 2k\,\hat{\boldsymbol{u}}(\boldsymbol{p})$. Replacing the weighted sum by an integral against the weighted density $W_\ell(\boldsymbol{\kappa})$ of spatial frequencies, $I_\ell(\boldsymbol{x}) = \int W_\ell(\boldsymbol{\kappa}) e^{\mathrm{i}\boldsymbol{\kappa}^\top\boldsymbol{\delta}} d\boldsymbol{\kappa}$. Under Assumption 1, $W_\ell(\boldsymbol{\kappa}) = c_\ell \exp(-\frac{1}{2}(\boldsymbol{\kappa}-\bar{\boldsymbol{\kappa}}_\ell)^\top \mathbf{C}_\ell^{-1}(\boldsymbol{\kappa}-\bar{\boldsymbol{\kappa}}_\ell))$, and equation 10 with $\mathbf{A} = \mathbf{C}_\ell^{-1}$ gives

$$h_\ell(\boldsymbol{\delta}) = c_\ell (2\pi)^{3/2}\sqrt{\det \mathbf{C}_\ell}\; e^{\mathrm{i}\bar{\boldsymbol{\kappa}}_\ell^\top\boldsymbol{\delta}} \exp\big(-\tfrac{1}{2}\boldsymbol{\delta}^\top\mathbf{C}_\ell\boldsymbol{\delta}\big).$$

The magnitude is a Gaussian with covariance $\mathbf{P}_\ell = \mathbf{C}_\ell^{-1}$. Because $\boldsymbol{\Sigma}_{\theta,\ell}$ acts on the plane orthogonal to $\bar{\hat{\boldsymbol{u}}}_\ell$, $\bar{\hat{\boldsymbol{u}}}_\ell$ is an eigenvector of $\mathbf{C}_\ell$ with eigenvalue σ_r^2, and the remaining eigenvectors span its orthogonal complement with eigenvalues $(2\bar{k})^2\sigma_{\theta,i}^2$. The standard deviations of $|h_\ell|$ are therefore $1/\sigma_r$ along $\bar{\hat{\boldsymbol{u}}}_\ell$ and $1/(2\bar{k}\,\sigma_{\theta,i})$ across it.

For uniform weighting, $2k$ is uniform on an interval of length $4\pi B/c$, so $\sigma_r = 4\pi B/(c\sqrt{12})$, and the look angle is uniform on an interval of length $\Delta\theta_\ell$, so $\sigma_\theta = \Delta\theta_\ell/\sqrt{12}$. The two extents are $\frac{\sqrt{12}}{2\pi}\cdot\frac{c}{2B}$ and $\frac{\sqrt{12}}{2\pi}\cdot\frac{\bar{\lambda}}{2\Delta\theta_\ell}$; the common factor cancels in the ratio, which gives equation 12 with $\bar{\lambda} = c/f_c$. The ratio exceeds one exactly when $\Delta\theta_\ell > B/f_c$. □

A.3 Image-domain Gaussians

The following result, summarized at the end of Section 3, shows what a Gaussian fitted to a radar image recovers.

Proposition 3 (PSF bias and band-pass invisibility of Born-density Gaussians)**.** *Let the scene be a volumetric Gaussian reflectivity density (a Born scatterer) $\sigma(\boldsymbol{x}) = \rho\exp\big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big)$ and let Assumption 1 hold. Then*

$$\begin{aligned} |I_\ell(\boldsymbol{x})| &= \alpha_\ell\,\gamma_\ell\,\exp\Big(-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top(\boldsymbol{\Sigma}+\mathbf{P}_\ell)^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big), \\ \gamma_\ell &= \exp\Big(-\tfrac{1}{2}\,\bar{\boldsymbol{\kappa}}_\ell^\top\big(\mathbf{C}_\ell+\boldsymbol{\Sigma}^{-1}\big)^{-1}\bar{\boldsymbol{\kappa}}_\ell\Big), \end{aligned} \tag{13}$$

with $\alpha_\ell \propto |\rho|\sqrt{\det\boldsymbol{\Sigma}/\det(\boldsymbol{\Sigma}+\mathbf{P}_\ell)}$. Consequently: (i) a Gaussian fitted to $|I_\ell|$ has covariance $\hat{\boldsymbol{\Sigma}}_\ell = \boldsymbol{\Sigma}+\mathbf{P}_\ell$, and fits to two sub-apertures disagree by $\mathbf{P}_\ell - \mathbf{P}_{\ell'}$, independently of the object; (ii) for $\boldsymbol{\Sigma} = s^2\mathbf{I}$, $\gamma_\ell = \exp\big(-2\bar{k}^2s^2/(1+\sigma_r^2s^2)\big)$, which vanishes unless $s \lesssim 1/(2\bar{k})$.

At 79 GHz, $s = 1$ mm gives $\gamma_\ell \approx 4\times10^{-3}$ and $s = 5$ mm gives e^{-129}: only sub-wavelength points remain, and by (i) their fitted ellipse is the PSF itself. The statement concerns linear reflectivity densities; opacity Gaussians composited front to back respond through the ray-termination surface they define, which our primitive makes explicit. Part (i) suggests an image-domain baseline that shares $\boldsymbol{\Sigma}$ across sub-apertures and renders $\boldsymbol{\Sigma}+\mathbf{P}_\ell$, a deconvolution splatting akin to R^2-Gaussian (Zha et al., 2024); we include it, noting that it discards phase and inherits the depth bias of Proposition 1.

Proof of Proposition 3. The Fourier transform of the scatterer is $\hat{\sigma}(\boldsymbol{\kappa}) = \rho(2\pi)^{3/2}\sqrt{\det\boldsymbol{\Sigma}}\,\exp(-\frac{1}{2}\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}\boldsymbol{\kappa} - \mathrm{i}\boldsymbol{\kappa}^\top\boldsymbol{\mu})$. The sub-aperture only observes the scene through its spectral window, so, up to a constant, $I_\ell(\boldsymbol{x}) = \int W_\ell(\boldsymbol{\kappa})\,\hat{\sigma}(\boldsymbol{\kappa})\,e^{\mathrm{i}\boldsymbol{\kappa}^\top\boldsymbol{x}}d\boldsymbol{\kappa}$. Completing the square with $\mathbf{A} = \mathbf{C}_\ell^{-1}+\boldsymbol{\Sigma}$ and $\boldsymbol{m} = \mathbf{A}^{-1}\mathbf{C}_\ell^{-1}\bar{\boldsymbol{\kappa}}_\ell$,

$$-\tfrac{1}{2}(\boldsymbol{\kappa}-\bar{\boldsymbol{\kappa}}_\ell)^\top\mathbf{C}_\ell^{-1}(\boldsymbol{\kappa}-\bar{\boldsymbol{\kappa}}_\ell) - \tfrac{1}{2}\boldsymbol{\kappa}^\top\boldsymbol{\Sigma}\boldsymbol{\kappa} = -\tfrac{1}{2}(\boldsymbol{\kappa}-\boldsymbol{m})^\top\mathbf{A}(\boldsymbol{\kappa}-\boldsymbol{m}) - \tfrac{1}{2}\bar{\boldsymbol{\kappa}}_\ell^\top\big(\mathbf{C}_\ell^{-1}-\mathbf{C}_\ell^{-1}\mathbf{A}^{-1}\mathbf{C}_\ell^{-1}\big)\bar{\boldsymbol{\kappa}}_\ell.$$

By the Woodbury identity, $\mathbf{C}_\ell^{-1} - \mathbf{C}_\ell^{-1}(\mathbf{C}_\ell^{-1}+\boldsymbol{\Sigma})^{-1}\mathbf{C}_\ell^{-1} = (\mathbf{C}_\ell+\boldsymbol{\Sigma}^{-1})^{-1}$, which yields γ_ℓ. Applying equation 10 to the remaining integral, the inverse transform of $\exp(-\frac{1}{2}(\boldsymbol{\kappa}-\boldsymbol{m})^\top\mathbf{A}(\boldsymbol{\kappa}-\boldsymbol{m}))$ is proportional to $\det(\mathbf{A})^{-1/2}\,e^{\mathrm{i}\boldsymbol{m}^\top(\boldsymbol{x}-\boldsymbol{\mu})}\exp\big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\mathbf{A}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\big)$, a Gaussian envelope in $\boldsymbol{x}-\boldsymbol{\mu}$ with covariance $\mathbf{A} = \boldsymbol{\Sigma}+\mathbf{P}_\ell$. Collecting constants gives α_ℓ. Part (i) follows because the

moments of a Gaussian magnitude profile are its mean and covariance. For part (ii), $\bar{\boldsymbol{\kappa}}_\ell = 2\bar{k}\bar{\hat{\boldsymbol{u}}}_\ell$ is an eigenvector of $\mathbf{C}_\ell + s^{-2}\mathbf{I}$ with eigenvalue $\sigma_r^2 + s^{-2}$, so $\bar{\boldsymbol{\kappa}}_\ell^\top (\mathbf{C}_\ell + s^{-2}\mathbf{I})^{-1}\bar{\boldsymbol{\kappa}}_\ell = 4\bar{k}^2 s^2/(1+\sigma_r^2 s^2)$. □

A.4 Proofs for Section 3

Phase-center error. Let $\boldsymbol{p}_t = \boldsymbol{p} + \frac{\epsilon}{2}\boldsymbol{e}$ and $\boldsymbol{p}_r = \boldsymbol{p} - \frac{\epsilon}{2}\boldsymbol{e}$, and let $\boldsymbol{d} = (\boldsymbol{x}-\boldsymbol{p})/R$. Expanding $\|\boldsymbol{x}-\boldsymbol{p}\mp\frac{\epsilon}{2}\boldsymbol{e}\| = \big(R^2 \mp R\epsilon\,\boldsymbol{e}^\top\boldsymbol{d} + \epsilon^2/4\big)^{1/2}$ to second order in ϵ/R gives $R \mp \frac{\epsilon}{2}\boldsymbol{e}^\top\boldsymbol{d} + \frac{\epsilon^2}{8R}\big(1-(\boldsymbol{e}^\top\boldsymbol{d})^2\big)$. The odd terms cancel in the sum, which leaves $2R + \frac{\epsilon^2}{4R}\big(1-(\boldsymbol{e}^\top\boldsymbol{d})^2\big) + O(\epsilon^4/R^3)$. □

Proposition 4 (Glints and the distance function; restated)**.** *Let $\boldsymbol{x}^\star(\boldsymbol{p})$ be a nondegenerate glint for $\boldsymbol{p}$ with range $R^\star(\boldsymbol{p})$ and principal radii r_1, r_2 (positive for convex). (i) As $k\to\infty$, $y(\boldsymbol{p},k) \simeq e^{-\mathrm{i}2kb}\sum_{\boldsymbol{x}^\star}\rho\,a(\boldsymbol{x}^\star)\,\frac{\pi}{k}\,|\det\mathbf{H}^\star|^{-1/2}\,e^{-\mathrm{i}2kR^\star - \mathrm{i}\frac{\pi}{4}\operatorname{sgn}\mathbf{H}^\star}$, with $\mathbf{H}^\star$ the range Hessian equation 6 at $\boldsymbol{x}^\star$. At normal incidence, $|\det\mathbf{H}^\star|^{-1} = \prod_i r_i R^\star/(R^\star+r_i)$. (ii) Along a smooth branch, $\nabla_{\boldsymbol{p}} R^\star = (\boldsymbol{p}-\boldsymbol{x}^\star)/R^\star$, so $\boldsymbol{x}^\star = \boldsymbol{p} - R^\star\,\nabla R^\star$ and $\boldsymbol{n}(\boldsymbol{x}^\star) = \nabla R^\star$. (iii) $\nabla^2_{\boldsymbol{p}} R^\star$ has eigenvalue 0 along $\nabla R^\star$ and $1/(R^\star+r_i)$ along the principal directions.*

Proof of Proposition 4. (i) Parameterize $\mathcal{S}$ near $\boldsymbol{x}^\star$ by tangent coordinates $\boldsymbol{u}$. At a specular point the range gradient $\boldsymbol{q}$ vanishes, so $R(\boldsymbol{u}) = R^\star + \frac{1}{2}\boldsymbol{u}^\top\mathbf{H}^\star\boldsymbol{u} + O(|\boldsymbol{u}|^3)$ with $\mathbf{H}^\star$ from equation 6. The method of stationary phase (Bleistein & Handelsman, 1986) gives $\int a\,e^{-\mathrm{i}2kR}d\boldsymbol{u} = a(\boldsymbol{x}^\star)e^{-\mathrm{i}2kR^\star}\frac{\pi}{k}|\det\mathbf{H}^\star|^{-1/2}e^{-\mathrm{i}\frac{\pi}{4}\operatorname{sgn}\mathbf{H}^\star}\,(1+O(k^{-1}))$, where sgn is the signature. At normal incidence, $\boldsymbol{d} = -\boldsymbol{n}$ and $\mathbf{K} = -\operatorname{diag}(1/r_1, 1/r_2)$ in the principal frame, so $\mathbf{H}^\star = \operatorname{diag}(1/R^\star + 1/r_i)$ and $|\det\mathbf{H}^\star|^{-1} = \prod_i r_i R^\star/(R^\star+r_i)$. With the physical-optics amplitude $a \propto k/(R^\star)^2$, the received power is proportional to $\pi^2(R^\star)^{-4}\prod_i r_i R^\star/(R^\star+r_i)$, which for $R^\star \gg r_i$ is proportional to $\pi r_1 r_2/(R^\star)^4$: the radar equation with the geometric-optics cross-section $\pi r_1 r_2$. (ii) Write $R^\star(\boldsymbol{p}) = \|\boldsymbol{x}^\star(\boldsymbol{p}) - \boldsymbol{p}\|$. By the chain rule, $\nabla_{\boldsymbol{p}} R^\star = (\boldsymbol{p}-\boldsymbol{x}^\star)/R^\star + (\partial\boldsymbol{x}^\star/\partial\boldsymbol{p})^\top(\boldsymbol{x}^\star-\boldsymbol{p})/R^\star$. The second term vanishes because $\partial\boldsymbol{x}^\star/\partial\boldsymbol{p}$ maps into the tangent plane at $\boldsymbol{x}^\star$ and $\boldsymbol{x}^\star - \boldsymbol{p}$ is normal to it at a specular point. Hence $\nabla R^\star = \boldsymbol{e}$, and $\boldsymbol{x}^\star = \boldsymbol{p} - R^\star\boldsymbol{e}$. The specular condition makes $\boldsymbol{e}$ parallel to $\boldsymbol{n}(\boldsymbol{x}^\star)$, and the orientation convention (normal toward the sensor) makes it equal. (iii) Locally, $R^\star$ is the distance to $\mathcal{S}$ along the branch, so $\|\nabla R^\star\| = 1$. Differentiating $\|\nabla R^\star\|^2 = 1$ gives $\nabla^2 R^\star\,\boldsymbol{e} = 0$. The level set of $R^\star$ through $\boldsymbol{p}$ is the offset surface $\{\boldsymbol{x} + R^\star\boldsymbol{n}(\boldsymbol{x})\}$, whose principal radii are $r_i + R^\star$ in the same principal directions. The Hessian of a distance function restricted to the tangent plane of its level set is the shape operator of that level set (Gilbarg & Trudinger, 2001, §14.6), which gives the eigenvalues $1/(R^\star + r_i)$. □

Proof of Lemma 1. Corollary 2 gives $|\psi| \propto \exp(-\frac{1}{2}\boldsymbol{q}^\top\boldsymbol{\Gamma}\boldsymbol{q})$ with $\boldsymbol{\Gamma} = 4k^2\boldsymbol{\Lambda}^{1/2}(\mathbf{I}+\mathbf{Z}^2)^{-1}\boldsymbol{\Lambda}^{1/2}$. Inverting gives $\boldsymbol{\Gamma}^{-1} = (4k^2)^{-1}\boldsymbol{\Lambda}^{-1/2}(\mathbf{I}+\mathbf{Z}^2)\boldsymbol{\Lambda}^{-1/2} = (2k)^{-2}\boldsymbol{\Lambda}^{-1} + \mathbf{H}\boldsymbol{\Lambda}\mathbf{H}$, since $\mathbf{Z}^2 = 4k^2\boldsymbol{\Lambda}^{1/2}\mathbf{H}\boldsymbol{\Lambda}\mathbf{H}\boldsymbol{\Lambda}^{1/2}$. At normal incidence $\boldsymbol{q} = \mathbf{U}^\top\boldsymbol{d}$, and a displacement $\boldsymbol{\delta}$ of $\boldsymbol{p}$ in the aperture plane changes $\boldsymbol{q}$ by $-\mathbf{U}^\top\boldsymbol{\delta}/R$ to first order, so the covariance $\boldsymbol{\Gamma}^{-1}$ in $\boldsymbol{q}$ maps to $R^2\boldsymbol{\Gamma}^{-1}$ in $\boldsymbol{\delta}$, neglecting the variation of $\mathbf{H}$ across the footprint. For a flat element, $R^2\boldsymbol{\Gamma}^{-1} = \boldsymbol{\Lambda} + (R/2k)^2\boldsymbol{\Lambda}^{-1}$ is the squared width of a Gaussian beam of waist $\boldsymbol{\Lambda}^{1/2}$ at wavenumber $2k$ after a distance R (Kogelnik & Li, 1966). For a cylinder, $H = 1/R + 1/r$ across the axis, which gives a standard deviation $\big(s^2(1+R/r)^2 + (R/2ks)^2\big)^{1/2}$. □

Proof of Proposition 1. Let $\boldsymbol{x}(\delta) = \boldsymbol{c} + \delta\boldsymbol{n}$ and absorb the echo amplitude into the aperture weights $w_{\boldsymbol{p}}$, normalized so that the weights $w_{\boldsymbol{p}}w_k$ form a product probability distribution. BP of the glint gives $I(\delta) = \sum_{\boldsymbol{p},k} w_{\boldsymbol{p}}w_k\,e^{\mathrm{i}\phi}$ with $\phi = 2k\big(\|\boldsymbol{p}-\boldsymbol{x}(\delta)\| - R^\star(\boldsymbol{p}) - \Delta\big)$. With $R^\star(\boldsymbol{p}) + \Delta = \|\boldsymbol{p}-\boldsymbol{c}\| - \varrho$ and $\|\boldsymbol{p}-\boldsymbol{x}(\delta)\| = \|\boldsymbol{p}-\boldsymbol{c}\| - \delta\cos\theta_{\boldsymbol{p}} + O(\delta^2/R)$, we get $\phi = 2k(\varrho - \delta\cos\theta_{\boldsymbol{p}})$ to first order in δ/R. By the cumulant expansion, $\log|\mathbb{E}e^{\mathrm{i}\phi}| = -\frac{1}{2}\operatorname{Var}\phi + O(\kappa_4)$, where the mean of ϕ affects only the phase of I. With k independent of $c_{\boldsymbol{p}} = \cos\theta_{\boldsymbol{p}}$,

$$\tfrac{1}{4}\operatorname{Var}\phi = \mathbb{E}[k^2]\,\mathbb{E}[(\varrho-\delta c)^2] - \mathbb{E}[k]^2\,\mathbb{E}[\varrho-\delta c]^2 = s_k^2(\varrho-\delta m)^2 + (\mu^2+s_k^2)\,\delta^2 v.$$

Setting the derivative in δ to zero gives equation 4. For $m \approx 1$ and $s_k \ll \mu$, $\delta^\star \approx \varrho/(1+\zeta)$. The approximation fails when $\operatorname{Var}\phi$ is large everywhere, where the magnitude can be multimodal. □

Proof of Theorem 1. (i) Let $\boldsymbol{x}^\star$ be a glint of $\mathcal{S}$ for $\boldsymbol{p}$, so $\boldsymbol{p}-\boldsymbol{x}^\star = R^\star \boldsymbol{n}(\boldsymbol{x}^\star)$. The parallel surface $\mathcal{S}_\Delta$ has, at $\boldsymbol{x}-\Delta\boldsymbol{n}(\boldsymbol{x})$, the same normal $\boldsymbol{n}(\boldsymbol{x})$ wherever the map is an immersion (its principal radii are $r_i-\Delta$). Hence $\boldsymbol{x}^\star-\Delta\boldsymbol{n}(\boldsymbol{x}^\star)$ is a glint of $\mathcal{S}_\Delta$ for the same $\boldsymbol{p}$, with range $\|\boldsymbol{p}-\boldsymbol{x}^\star+\Delta\boldsymbol{n}\| = R^\star+\Delta$, and every glint of $\mathcal{S}_\Delta$ arises this way. By Proposition 4(i), both scenes therefore contribute $e^{-\mathrm{i}2k(R^\star(\boldsymbol{p})+\Delta)}$ at every $\boldsymbol{p}$, times amplitudes that differ through $|\det \mathbf{H}^\star|^{-1/2}$, which are absorbed by the unknown reflectivity and pattern only up to the aperture-dependent factor given next. (ii) For a circle in a principal plane (or a sphere), $\mathcal{S}_\Delta$ is the circle about $\boldsymbol{c}$ of signed radius $\varrho = r-\Delta$, and $R^\star+\Delta = \|\boldsymbol{p}-\boldsymbol{c}\|-\varrho$ exactly; $\varrho=0$ is a point at $\boldsymbol{c}$, and $\varrho<0$ a concave mirror whose reflecting side faces $\boldsymbol{c}$ from behind. At normal incidence the glint amplitude carries $|\det \mathbf{H}^\star|^{-1/2} \propto (rR^\star/(R^\star+r))^{1/2}$ per principal direction; relative to a point scatterer, its variation across the aperture is $(R^\star/(R^\star+r))^{1/2}$, which changes by less than 0.1% across our aperture for $r \le 3\,\mathrm{cm}$ and is nearly collinear with the logarithm of a $\cos^n$ pattern and of R^{-m} spreading. (iii) For a narrow sub-aperture, $v_\ell \to 0$, so $\zeta_\ell \to 0$ and equation 4 gives $\delta^\star \to \varrho/m_\ell$ along the search direction; choosing $\boldsymbol{n} = -\boldsymbol{u}_\ell$ gives $m_\ell \to 1$ and a peak at $\boldsymbol{c}-\varrho\boldsymbol{u}_\ell$, on the look line through $\boldsymbol{c}$. The full-aperture statement is Proposition 1 with r replaced by ϱ. Two narrow sub-apertures with non-parallel look directions determine $\boldsymbol{c}$ as the intersection of their look lines and ϱ as the signed distance of a peak from it; given Δ, $r=\varrho+\Delta$. □

A.5 Proof of Theorem 2

Expansion. Differentiating $R(\boldsymbol{u}) = \|\boldsymbol{x}_j(\boldsymbol{u})-\boldsymbol{p}\|$ gives $\partial_a R = (\partial_a \boldsymbol{x}_j)^\top(\boldsymbol{x}_j-\boldsymbol{p})/R$ and

$$\partial_a\partial_b R = \frac{(\partial_a\boldsymbol{x}_j)^\top(\partial_b\boldsymbol{x}_j) + (\boldsymbol{x}_j-\boldsymbol{p})^\top\partial_a\partial_b\boldsymbol{x}_j}{R} - \frac{\partial_a R\,\partial_b R}{R},$$

with $\partial\boldsymbol{x}_j/\partial\boldsymbol{u} = \mathbf{U}_j + \boldsymbol{n}_j(\mathbf{K}_j\boldsymbol{u})^\top$ and $\partial_a\partial_b\boldsymbol{x}_j = \boldsymbol{n}_j(\mathbf{K}_j)_{ab}$. At a general $\boldsymbol{u}_0$ these give $\boldsymbol{q}$ and $\mathbf{H}$ in closed form. At $\boldsymbol{u}_0=0$, using $\mathbf{U}_j^\top\mathbf{U}_j=\mathbf{I}$, they reduce to equation 6.

Integral. Substituting the surface into equation 1, the delta function collapses the volume integral, so $\psi_j(\boldsymbol{p},k) = \rho_j\int g_j(\boldsymbol{u})\,a(\boldsymbol{u})\,e^{-\mathrm{i}2kR(\boldsymbol{u})}d\boldsymbol{u}$. Write $\boldsymbol{u}=\boldsymbol{u}_0+\boldsymbol{v}$, $a \approx a_0(1+\boldsymbol{g}^\top\boldsymbol{v})$ and $R \approx R_0+\boldsymbol{q}^\top\boldsymbol{v}+\frac{1}{2}\boldsymbol{v}^\top\mathbf{H}\boldsymbol{v}$. The exponent becomes $-\mathrm{i}2kR_0 - \frac{1}{2}\boldsymbol{u}_0^\top\boldsymbol{\Lambda}_j^{-1}\boldsymbol{u}_0 - \frac{1}{2}\boldsymbol{v}^\top\mathbf{M}\boldsymbol{v} - \boldsymbol{\eta}^\top\boldsymbol{v}$ with $\mathbf{M} = \boldsymbol{\Lambda}_j^{-1}+\mathrm{i}2k\mathbf{H}$ and $\boldsymbol{\eta} = \boldsymbol{\Lambda}_j^{-1}\boldsymbol{u}_0 + \mathrm{i}2k\boldsymbol{q}$. Since $\operatorname{Re}\mathbf{M} = \boldsymbol{\Lambda}_j^{-1} \succ 0$, equation 10 (with $\boldsymbol{\eta}\to-\mathrm{i}\boldsymbol{\eta}$) gives $\int e^{-\frac{1}{2}\boldsymbol{v}^\top\mathbf{M}\boldsymbol{v}-\boldsymbol{\eta}^\top\boldsymbol{v}}d\boldsymbol{v} = 2\pi(\det\mathbf{M})^{-1/2}e^{\frac{1}{2}\boldsymbol{\eta}^\top\mathbf{M}^{-1}\boldsymbol{\eta}}$, and the first moment of the same complex Gaussian weight is $\bar{\boldsymbol{v}} = -\mathbf{M}^{-1}\boldsymbol{\eta}$ times the integral, which yields the factor $1+\boldsymbol{g}^\top\bar{\boldsymbol{v}}$ in equation 7. For $\boldsymbol{u}_0=0$ and $\boldsymbol{g}=0$, $\frac{1}{2}\boldsymbol{\eta}^\top\mathbf{M}^{-1}\boldsymbol{\eta} = -2k^2\boldsymbol{q}^\top\mathbf{M}^{-1}\boldsymbol{q}$. Finally, $\mathbf{M} = \boldsymbol{\Lambda}_j^{-1/2}(\mathbf{I}+\mathrm{i}\mathbf{Z}_j)\boldsymbol{\Lambda}_j^{-1/2}$ with $\mathbf{Z}_j = 2k\boldsymbol{\Lambda}_j^{1/2}\mathbf{H}\boldsymbol{\Lambda}_j^{1/2}$ real symmetric, so $\det\mathbf{M} = \det(\boldsymbol{\Lambda}_j)^{-1}\prod_i(1+\mathrm{i}2k\lambda_i)$; continuity in k from $k=0$ selects the principal root of each factor. □

Lemma 4 (Cubic remainder). *Let $\boldsymbol{\Lambda}_j = s^2\mathbf{I}$, expand at $\boldsymbol{u}_0=0$ with constant amplitude, and let $|T_3(\boldsymbol{u})| \le \frac{1}{6}M_3\|\boldsymbol{u}\|^3$ bound the cubic Taylor remainder of R on the disc $\|\boldsymbol{u}\| \le r_t$. Then*

$$\frac{|\psi_j^{\mathrm{exact}} - \psi_j|}{|\psi_j(\boldsymbol{q}=0,\mathbf{H}=0)|} \;\le\; \sqrt{\pi/2}\;kM_3s^3 \;+\; 2\,e^{-r_t^2/(2s^2)}.$$

Proof. On the disc, $|e^{-\mathrm{i}2kT_3}-1| \le 2k|T_3|$, and $\int g_j\|\boldsymbol{u}\|^3 \le 2\pi s^2\cdot\mathbb{E}\|\boldsymbol{u}\|^3$ for $\boldsymbol{u}\sim\mathcal{N}(0,s^2\mathbf{I})$, where $\mathbb{E}\|\boldsymbol{u}\|^3 = 3\sqrt{\pi/2}\,s^3$. Outside the disc, both integrands are bounded by g_j, whose mass there is $2\pi s^2 e^{-r_t^2/(2s^2)}$. Dividing by the peak magnitude $2\pi s^2|\rho_j a|$ gives the claim. We use $r_t = 4s$, for which the second term is 6.7×10^{-4}. □

A.6 Limits and bistatic form

Corollary 1 (Limits and bistatic form). *(i) As $\boldsymbol{\Lambda}_j \to \infty$ with $\boldsymbol{u}_0$ at the specular point, equation 7 reduces to the glint of Proposition 4(i). As $\boldsymbol{\Lambda}_j \to 0$, it reduces to a point scatterer, $\psi_j \propto \rho_j a_0 \det(\boldsymbol{\Lambda}_j)^{1/2}e^{-\mathrm{i}2kR_0}$. (ii) For a transmitter–receiver pair, Theorem 2 holds with $2kR_0$, $2k\boldsymbol{q}$ and $2k\mathbf{H}$ replaced by $k(R_t+R_r)$, $k(\boldsymbol{q}_t+\boldsymbol{q}_r)$ and $k(\mathbf{H}_t+\mathbf{H}_r)$, each computed as in equation 6 from the transmitter or the receiver. The lobe is then centered on the bisector equation 2.*

Proof. (i) follows from equation 7 with $\mathbf{\Lambda}_j^{-1} \to 0$ at the stationary point, where $\mathbf{M} \to \mathrm{i}2k\mathbf{H}$, and with $\mathbf{\Lambda}_j \to 0$, where $\mathbf{M}^{-1} \to \mathbf{\Lambda}_j$ and $2\pi(\det \mathbf{M})^{-1/2} \to 2\pi \det(\mathbf{\Lambda}_j)^{1/2}$. (ii) repeats the proof of Theorem 2 with the phase $k(R_t + R_r)$ expanded to second order. □

A.7 Proof of Corollary 2

At $\boldsymbol{u}_0 = 0$ and $\boldsymbol{g} = 0$, since $\boldsymbol{q}$ is real, $|\exp(-2k^2 \boldsymbol{q}^\top \mathbf{M}^{-1} \boldsymbol{q})| = \exp(-2k^2 \boldsymbol{q}^\top \operatorname{Re}(\mathbf{M}^{-1})\boldsymbol{q})$. With $\mathbf{M}^{-1} = \mathbf{\Lambda}_j^{1/2}(\mathbf{I} + \mathrm{i}\mathbf{Z}_j)^{-1}\mathbf{\Lambda}_j^{1/2}$ and $(\mathbf{I} + \mathrm{i}\mathbf{Z}_j)^{-1} = (\mathbf{I} - \mathrm{i}\mathbf{Z}_j)(\mathbf{I} + \mathbf{Z}_j^2)^{-1}$ for real symmetric $\mathbf{Z}_j$, $\operatorname{Re}(\mathbf{M}^{-1}) = \mathbf{\Lambda}_j^{1/2}(\mathbf{I} + \mathbf{Z}_j^2)^{-1}\mathbf{\Lambda}_j^{1/2}$, which is equation 14. (i) $\mathbf{Z}_j = 0$ gives $\mathbf{\Gamma}_j = 4k^2\mathbf{\Lambda}_j$. (ii) If $\|\mathbf{Z}_j^{-1}\| < 1$, $(\mathbf{I} + \mathbf{Z}_j^2)^{-1} = \sum_{m \geq 0}(-1)^m \mathbf{Z}_j^{-2m-2}$ converges, and its leading term gives $\mathbf{\Gamma}_j \to 4k^2\mathbf{\Lambda}_j^{1/2}\mathbf{Z}_j^{-2}\mathbf{\Lambda}_j^{1/2} = \mathbf{H}^{-1}\mathbf{\Lambda}_j^{-1}\mathbf{H}^{-1}$. The stationary point of the phase $\boldsymbol{q}^\top\boldsymbol{u} + \frac{1}{2}\boldsymbol{u}^\top\mathbf{H}\boldsymbol{u}$ is $\boldsymbol{u}^\star = -\mathbf{H}^{-1}\boldsymbol{q}$, and $\boldsymbol{q}^\top\mathbf{H}^{-1}\mathbf{\Lambda}_j^{-1}\mathbf{H}^{-1}\boldsymbol{q} = \boldsymbol{u}^{\star\top}\mathbf{\Lambda}_j^{-1}\boldsymbol{u}^\star$, the Gaussian weight of the specular point. □

A.8 Unbiasedness of the held-out objective

Assumption 2 (Independent sub-aperture noise)**.** $y_\ell = \mathcal{F}_\ell(\mathcal{G}^\star) + n_\ell$, where $\mathcal{G}^\star$ is the true scene, $\mathbb{E}[n_\ell] = 0$, and n_ℓ is independent of $\mathcal{G}^\star$ and of $\{n_{\ell'}\}_{\ell' \neq \ell}$.

Proposition 5 (Unbiasedness of the held-out objective)**.** *Under Assumption 2, for every ℓ, $\mathbb{E}\,\|y_\ell - \mathcal{F}_\ell(f_\theta(y_{-\ell}))\|^2 = \mathbb{E}\,\|\mathcal{F}_\ell(\mathcal{G}^\star) - \mathcal{F}_\ell(f_\theta(y_{-\ell}))\|^2 + \mathbb{E}\,\|n_\ell\|^2$, so $\mathcal{L}_{\mathrm{HO}}$ has the same minimizers as the error on noise-free held-out data.*

Proof of Proposition 5. Let $\hat{\mathcal{G}} = f_\theta(y_{-\ell})$ and $\boldsymbol{e} = \mathcal{F}_\ell(\mathcal{G}^\star) - \mathcal{F}_\ell(\hat{\mathcal{G}})$. Then $y_\ell - \mathcal{F}_\ell(\hat{\mathcal{G}}) = \boldsymbol{e} + n_\ell$ and $\mathbb{E}\|\boldsymbol{e} + n_\ell\|^2 = \mathbb{E}\|\boldsymbol{e}\|^2 + 2\operatorname{Re}\mathbb{E}\langle \boldsymbol{e}, n_\ell\rangle + \mathbb{E}\|n_\ell\|^2$. The vector $\boldsymbol{e}$ is a function of $(\mathcal{G}^\star, y_{-\ell})$, which by Assumption 2 is independent of n_ℓ; with $\mathbb{E}[n_\ell] = 0$ the cross term vanishes. □

A.9 What held-out prediction rejects

Linearize the renderer around a scene estimate, $y_\ell \approx \mathbf{A}_\ell\boldsymbol{\vartheta}$, and fit by regularized least squares $\hat{\boldsymbol{\vartheta}} = (\mathbf{A}^{\mathsf{H}}\mathbf{A} + \mathbf{R})^{-1}\mathbf{A}^{\mathsf{H}}y$ on all sub-apertures, or $\hat{\boldsymbol{\vartheta}}_{-\ell}$ on all but ℓ.

Proposition 6 (What held-out prediction rejects)**.** *Let $r_\ell = y_\ell - \mathbf{A}_\ell\hat{\boldsymbol{\vartheta}}$ and $r_{-\ell} = y_\ell - \mathbf{A}_\ell\hat{\boldsymbol{\vartheta}}_{-\ell}$. Then $r_{-\ell} = (\mathbf{I} - \mathbf{P}_{\ell\ell})^{-1}r_\ell$ with $\mathbf{P}_{\ell\ell} = \mathbf{A}_\ell(\mathbf{A}^{\mathsf{H}}\mathbf{A} + \mathbf{R})^{-1}\mathbf{A}_\ell^{\mathsf{H}}$. If $\mathbf{R} = 0$ and the response $\boldsymbol{\alpha}_{j\ell}$ of primitive j's reflectivity in each sub-aperture ℓ is orthogonal to the responses of all other parameters, then $\mathbf{P}_{\ell\ell}\boldsymbol{\alpha}_{j\ell} = w_{j\ell}\boldsymbol{\alpha}_{j\ell}$, where $w_{j\ell} = \|\boldsymbol{\alpha}_{j\ell}\|^2 / \sum_{\ell'}\|\boldsymbol{\alpha}_{j\ell'}\|^2$ is the* lobe share *of sub-aperture ℓ.*

Proof of Proposition 6. Let $\mathbf{G} = \mathbf{A}^{\mathsf{H}}\mathbf{A} + \mathbf{R}$ and $\mathbf{G}_{-\ell} = \mathbf{G} - \mathbf{A}_\ell^{\mathsf{H}}\mathbf{A}_\ell$, so $\hat{\boldsymbol{\vartheta}}_{-\ell} = \mathbf{G}_{-\ell}^{-1}(\mathbf{A}^{\mathsf{H}}y - \mathbf{A}_\ell^{\mathsf{H}}y_\ell)$. By the Woodbury identity, $\mathbf{A}_\ell\mathbf{G}_{-\ell}^{-1} = (\mathbf{I} - \mathbf{P}_{\ell\ell})^{-1}\mathbf{A}_\ell\mathbf{G}^{-1}$. Hence $\mathbf{A}_\ell\hat{\boldsymbol{\vartheta}}_{-\ell} = (\mathbf{I} - \mathbf{P}_{\ell\ell})^{-1}(\mathbf{A}_\ell\hat{\boldsymbol{\vartheta}} - \mathbf{P}_{\ell\ell}y_\ell)$, and $y_\ell - \mathbf{A}_\ell\hat{\boldsymbol{\vartheta}}_{-\ell} = (\mathbf{I} - \mathbf{P}_{\ell\ell})^{-1}\big[(\mathbf{I} - \mathbf{P}_{\ell\ell})y_\ell - \mathbf{A}_\ell\hat{\boldsymbol{\vartheta}} + \mathbf{P}_{\ell\ell}y_\ell\big] = (\mathbf{I} - \mathbf{P}_{\ell\ell})^{-1}r_\ell$. For the second statement, with $\mathbf{R} = 0$ and orthogonality, $\mathbf{G}$ is diagonal on primitive j's reflectivity with entry $\sum_{\ell'}\|\boldsymbol{\alpha}_{j\ell'}\|^2$, and $\mathbf{P}_{\ell\ell}\boldsymbol{\alpha}_{j\ell} = \boldsymbol{\alpha}_{j\ell}\,\|\boldsymbol{\alpha}_{j\ell}\|^2 / \sum_{\ell'}\|\boldsymbol{\alpha}_{j\ell'}\|^2$. The identity is checked numerically in `v3_analysis.py` (maximum relative deviation 8×10^{-16}). □

Proof of Proposition 2. Let the multipath echo follow a stationary path $\boldsymbol{p} \to \boldsymbol{x}_1 \to \cdots \to \boldsymbol{x}_n \to \boldsymbol{p}$ of length $L(\boldsymbol{p})$, so that $R_g = L/2$. Because L is stationary in the interior points $\boldsymbol{x}_i$ (Fermat), the envelope theorem gives $\nabla_{\boldsymbol{p}} L = \boldsymbol{P}(\boldsymbol{u}_1 + \boldsymbol{u}_n)$, where $\boldsymbol{u}_1, \boldsymbol{u}_n$ are the unit vectors from $\boldsymbol{x}_1$ and $\boldsymbol{x}_n$ to $\boldsymbol{p}$ and $\boldsymbol{P}$ projects onto the aperture plane. Hence $\|\nabla R_g\| \leq \frac{1}{2}(\|\boldsymbol{P}\boldsymbol{u}_1\| + \|\boldsymbol{P}\boldsymbol{u}_n\|) \leq 1$, with equality only for grazing legs. Now let R be any smooth function with $\|\nabla R\| < 1$ and consider the spheres $\|\boldsymbol{x} - \boldsymbol{p}\| = R(\boldsymbol{p})$. Differentiating in the in-plane coordinates p_i gives the envelope condition $(\boldsymbol{x} - \boldsymbol{p})_i = -R\,\partial_i R$, so the tangential part of $\boldsymbol{x} - \boldsymbol{p}$ is $-R\nabla R$; the sphere constraint then fixes the normal part to $R(1 - \|\nabla R\|^2)^{1/2}$ on the side of the scene. This is $\boldsymbol{x}_g(\boldsymbol{p})$. Where the envelope is a regular surface, it is tangent to the sphere about $\boldsymbol{p}$ at $\boldsymbol{x}_g(\boldsymbol{p})$, so its normal there is parallel to $\boldsymbol{p} - \boldsymbol{x}_g$: $\boldsymbol{x}_g$ is a glint for $\boldsymbol{p}$ at range $R(\boldsymbol{p})$, and by Proposition 4(ii) R is its distance function. A glint on $\boldsymbol{x}_g$ thus reproduces the phase $e^{-\mathrm{i}2kR_g}$; its amplitude follows the curvature of $\boldsymbol{x}_g$ (Proposition 4(i)) and differs from the ghost's by the reflection losses and spreading of the extra

legs, which a free reflectivity absorbs only in part. At caustics of the envelope the construction fails, as does Proposition 4. □

B NUMERICAL VERIFICATION

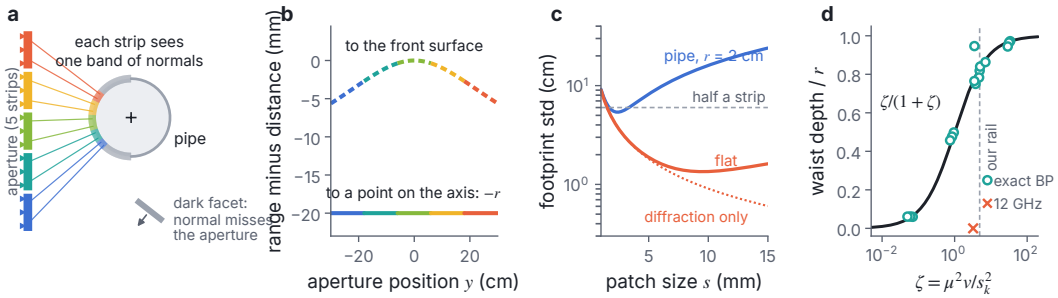

Figure 4: **Specular optics of a synthetic aperture.** (a) Each sub-aperture sees the band of normals that point into it; a facet whose normal misses the aperture is dark (Lemma 1). (b) The range history of a convex glint (pipe, $r = 2\,\text{cm}, 0.3\,\text{m}$) equals that of a point on the axis delayed by r, not that of a point on the surface (Proposition 4). (c) Aperture footprint of a surface Gaussian's lobe equation 3: curvature spreads it over several sub-apertures. (d) Depth of the back-projection waist, Proposition 1 against exact 2D back-projection.

Three scripts (supplementary material) reproduce every number in the paper. `verify_numbers.py` computes the constants: $\bar{\lambda} = 3.795\,\text{mm}$, $B/f_c = 0.0506\,\text{rad}$ $(2.90°)$, $c/(2B) = 3.75\,\text{cm}$, $1/(2\bar{k}) = 0.30\,\text{mm}$, $\gamma_\ell = 4.2 \times 10^{-3}$ for $s = 1\,\text{mm}$ and e^{-129} for $s = 5\,\text{mm}$, and it checks the evaluation of the complex Gaussian integral and the geometric limit of Corollary 2.

Accuracy of Theorem 2 against the exact range. `thm1_exact_study.py` compares equation 7 with quadrature (1201^2 nodes over $\pm 6s$) of the exact integrand $g_j(\boldsymbol{u})\,(-\boldsymbol{n}(\boldsymbol{u})^\top \boldsymbol{d}(\boldsymbol{u}))\,R(\boldsymbol{u})^{-2} e^{-\mathrm{i}2kR(\boldsymbol{u})}$, with the exact range and normal of the curved patch, at 79 GHz over $R_0 \in \{0.2, 0.3, 0.5\}\,\text{m}$, $s \in \{2, 5, 10\}\,\text{mm}$, incidence $\theta \in \{0, 10, 20, 30\}°$ and curvature $K \in \{0, \pm 50\}\,\text{m}^{-1}$ (a pipe of radius 2 cm). Table 5 reports the error relative to the local exact echo for echoes above 1% of the peak; the 18 of 108 cases below that level lie outside the lobe, where relative error is not meaningful. Relative to the peak, the largest error is 3.3×10^{-2} for the center expansion and 5.2×10^{-3} for the renderer.

Table 5: Maximum relative error of Theorem 2 against the exact echo, over R_0 and s.

θ	K (m^{-1})	center	re-centred	re-centred + slope (renderer)
0°	0	3.0×10^{-3}	3.0×10^{-3}	3.0×10^{-3}
	−50	1.1×10^{-2}	1.1×10^{-2}	1.1×10^{-2}
	+50	6.4×10^{-3}	6.4×10^{-3}	6.4×10^{-3}
10°	0	4.0×10^{-2}	2.6×10^{-2}	2.6×10^{-3}
	−50	1.7×10^{-2}	1.4×10^{-2}	1.1×10^{-2}
	+50	1.8×10^{-2}	1.1×10^{-2}	6.5×10^{-3}
20°	0	2.3×10^{-2}	2.3×10^{-2}	1.2×10^{-4}
	−50	8.5×10^{-2}	4.9×10^{-2}	1.1×10^{-2}
	+50	9.2×10^{-2}	4.2×10^{-2}	1.1×10^{-2}
30°	−50	2.9×10^{-1}	1.1×10^{-1}	2.5×10^{-2}
	+50	3.9×10^{-1}	9.8×10^{-2}	4.0×10^{-2}

Lobe widths. From equation 14 at normal incidence, flat patches with $s = 2, 5, 10\,\text{mm}$ have 1σ lobes of $8.68°, 3.46°, 1.73°$ in the far field and $8.69°, 3.59°, 2.58°$ at $R_0 = 0.3\,\text{m}$, where $2ks^2/R_0 = 0.04, 0.28, 1.10$; on a pipe with $K = -50\,\text{m}^{-1}$ they are $10.65°, 15.87°, 32.29°$ at $R_0 = 0.3\,\text{m}$.

Lobe overlap. `v3_analysis.py` samples the visible surface of a pipe (radius 2 cm) and of a 10 cm flat plate at 0.3 m, evaluates $|\psi_j|^2$ from Corollary 2 over five frequencies and 241 aperture positions along y, marks a position as inside the lobe if $|\psi_j|^2 \geq e^{-2}$ of its maximum, and counts the sub-apertures that contain such a position. Table 6 summarizes the fraction of observed surface covered by at least two sub-apertures and the median dominant lobe share $\max_\ell w_{j\ell}$.

Table 6: Lobe overlap at 0.3 m standoff for $s = 2/5/10$ mm primitives. Aperture of 0.2, 0.4, 0.6 m (angular span 37°, 67°, 90°).

Surface, S	Metric	0.2 m	0.4 m	0.6 m
Pipe, S=5	covered by ≥ 2	.66 / .75 / .83	.65 / .70 / .77	.70 / .73 / .83
	median $\max_\ell w_{j\ell}$	.79 / .69 / .53	.78 / .72 / .61	.85 / .73 / .53
Pipe, S=10	covered by ≥ 2	1.0 / 1.0 / .99	.86 / .90 / .93	.88 / .91 / .98
Plate, S=5	covered by ≥ 2	1.0 / 1.0 / .90	1.0 / .71 / .56	1.0 / .31 / .16
	median $\max_\ell w_{j\ell}$	.47 / .79 / .88	.66 / .87 / .94	.86 / 1.0 / 1.0

Glints, geometric limit and apparent depth. `theory_v4_checks.py` verifies the results of Section 3. (a) On an ellipsoid with semi-axes 3, 5 and 2 cm, centered 0.3 m from the aperture (a surface that is not a quadric patch of the renderer), the closest point, normal and principal radii recovered from finite differences of $R^\star(\boldsymbol{p})$ via Proposition 4(ii,iii) agree with the truth to $0.02\,\mu$m, 10^{-3} degrees and a relative error of 3×10^{-6} at eight random aperture positions. (b) For large patches at normal incidence ($s = 3$ cm flat; $s = 1.2$ cm with $K = -20$ or $-50\,\mathrm{m}^{-1}$; $R_0 = 0.2, 0.3, 0.5$ m), the ratio of exact quadrature to the geometric limit of Corollary 1 lies between 0.986 and 0.998. (c) Table 7 compares equation 4 with the peak of exact 2D back-projection of a physical-optics pipe (1001 aperture positions, 64 frequencies).

Table 7: Depth of the back-projection peak behind the front surface of a pipe (mm): exact 2D BP / Proposition 1. $f_c = 79$ GHz; aperture length at standoff (front surface at 0.3 m unless stated). The first column is the measured rail geometry; standoffs of 0.85 and 1.05 m change the values by at most 1.5 mm.

	$B = 4$ GHz				$B = 12$ GHz	
Radius	0.99 m at 0.95 m	0.6 m at 0.3 m	0.2 m	0.1 m	0.6 m	0.05 m
1 cm	8.4 / 8.4	9.7 / 9.7	5.0 / 4.9	0.6 / 0.6	7.5 / 8.1	0.0 / 0.0
2 cm	16.4 / 16.7	19.3 / 19.4	9.5 / 9.2	1.2 / 1.0	18.9 / 15.8	0.0 / 0.0
3 cm	23.5 / 24.9	28.3 / 29.1	13.7 / 12.9	1.8 / 1.4	0.0 / 23.3	0.0 / 0.0
ζ (2 cm)	4.9	31.6	0.85	0.06	3.5	0.00

For $r = 3$ cm, $L = 0.6$ m and $B = 12$ GHz, the image has two peaks of similar height, at the surface and near the axis, and the cumulant approximation of Proposition 1 breaks down.

Weights, sampling and search direction. The value $\zeta \approx 4.9$ quoted for the rail uses a continuous uniform band; 64 discrete frequencies give 5.1. An element pattern $\cos^n$ per way with $1/R^2$ spreading lowers ζ on the rail ($r = 2$ cm) to 4.6, 3.9, 2.5 and 1.2 for $n = 1.3, 4, 10, 20$, with exact-BP depths of 16.2, 15.7, 13.7 and 8.8 mm (`reviews/scripts/v4_checks.py`). For an object off the aperture center, m and v must be computed from the actual angles to the search direction: along the aperture normal through $\boldsymbol{c}$, an object 17 cm off center focuses 18.1 mm behind its surface instead of 16.4 mm.

Accuracy of Lemma 1. Against quadrature of the exact 1D footprint at $R = 0.3$ m, the second-order form gives standard deviations of 1.35, 1.88 and 4.53 cm for flat elements with $s = 10, 5, 2$ mm (quadrature: 1.35, 1.89, 4.77 cm), and 16.0, 8.2 and 5.6 cm on a pipe of radius 2 cm (quadrature: 14.0, 8.9, 5.9 cm). The flat values coincide with the Gaussian-beam width. The errors come from the variation of incidence angle and curvature across wide footprints; lobe shares computed from the second-order form are therefore approximate at the 10% level.

Sub-aperture PSFs. Figure 5 uses 2D monostatic back-projection with range compression (401 positions, 64 frequencies from 77 to 81 GHz) of a point target and of a pipe cross-section sampled at 1601 points with physical-optics obliquity weights.

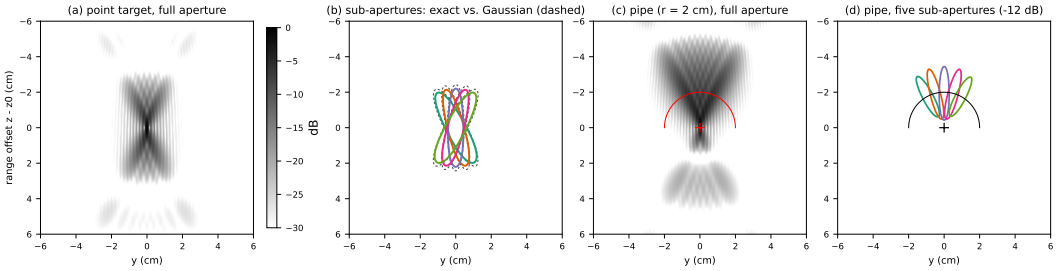

Figure 5: **Exact sub-aperture PSFs** (2D simulation, 0.4 m aperture at 0.3 m, $S = 5$). (a) Point target, full aperture (dB). (b) Sub-aperture -6 dB contours, exact (solid) and Lemma 3 (dashed); the sidelobes of the annular-sector support are not modeled. (c) Metal pipe ($r = 2$ cm, physical optics; arc: surface, $+$: axis); the peak lies 1.85 cm behind the surface (Proposition 1). (d) Pipe, sub-apertures at -12 dB.

C RENDERER DETAILS

Table 8: From optical Gaussian splatting to coherent surface Gaussians.

Component	3D Gaussian splatting	Coherent surface Gaussians (ours)
Primitive	volumetric 3D ellipsoid	curved 2D surface patch, complex ρ_j (§4)
Projection	EWA affine footprint	complex Gaussian integral, Theorem 2
View dependence	free spherical harmonics	derived lobe $\mathbf{\Gamma}_j(k)$, Corollary 2
Culling	view frustum	angular lobe
Densification	gradient magnitude	range-error bound ε_j
Feed-forward	pixelSplat, MVSplat	sub-aperture set encoder (§5)

Amplitude term. We use the physical-optics amplitude $a(\boldsymbol{u}) = \frac{\mathrm{i}k}{2\pi}\,\mathrm{E}(\boldsymbol{p},\boldsymbol{d})^2\,(-\boldsymbol{n}^\top\boldsymbol{d})/R^2$ (Knott et al., 2004), where E is the calibrated one-way antenna pattern evaluated in the direction $\boldsymbol{d}$ (a $\cos^q$ model whose exponent is fitted to the corner-reflector scan; by Appendix K, its error affects amplitudes but not the phase used for geometry), and its log-gradient $\boldsymbol{g}$ at $\boldsymbol{u}_0$ by automatic differentiation. Frequency-dependent system response is removed during calibration.

Expansion point. For each primitive and aperture tile we start at $\boldsymbol{u}_0 = 0$ and apply two steps of $\boldsymbol{u}_0 \leftarrow \boldsymbol{u}_0 + \mathrm{Re}\ \bar{\boldsymbol{v}}$, clamping $\boldsymbol{u}_0$ to Mahalanobis radius 3 under $\mathbf{\Lambda}_j$. The steps are differentiable and add two 2×2 solves per evaluation. When the clamp is active, the specular point lies outside the primitive and its echo is small.

Visibility. Back-facing primitives, $\boldsymbol{n}_j^\top\boldsymbol{d} > 0$, receive $v_j(\boldsymbol{p}) = 0$. For occlusion we sort primitives by R_0 for each aperture tile and accumulate an occupancy along the line of sight, $v_j(\boldsymbol{p}) = \prod_{i:\,R_{0,i}<R_{0,j}} \big(1 - o_i\,\phi_{ij}(\boldsymbol{p})\big)^2$, where $o_i \in [0,1]$ is a learned opacity, ϕ_{ij} is the Gaussian footprint of primitive i at the ray toward $\boldsymbol{\mu}_j$, and the square accounts for the two-way path. This is the coherent counterpart of the transmittance in 3DGS (Kerbl et al., 2023): it gates amplitudes but does not blend them, so interference between visible primitives is preserved.

Lobe culling and tiling. For each primitive we precompute $\mathbf{\Gamma}_j(k_{\min})$, the widest lobe over the band, and the set of aperture positions with $\frac{1}{2}\boldsymbol{q}^\top\mathbf{\Gamma}_j\boldsymbol{q} < \tau_{\mathrm{lobe}}$, with $\tau_{\mathrm{lobe}} = 6.9$, which culls 85–87% of the front-facing pairs and adds less than 0.1% error (Table 1); re-centering once at the band center is as accurate as re-centering at every frequency. Aperture positions are grouped into tiles of

16×16 positions and all K frequencies; a tile evaluates equation 7 only for the primitives whose lobe intersects it. The cost is $O(K \sum_j |\mathcal{A}_j^{\mathrm{lobe}}|)$ instead of $O(K\,N\,|\mathcal{A}|)$.

Density control. Every 200 iterations we (i) split primitives with $\varepsilon_j > \varepsilon_\phi = 0.3\,\mathrm{rad}$ (evaluated at the largest incidence at which the primitive is inside its lobe) into two children placed at $\pm s_{j1}$ along the major tangent axis with halved scales; (ii) prune primitives with $|\rho_j|$ below 1% of the largest or existence $o_j < 0.05$; (iii) clone primitives in voxels where the back-projected residual $|\mathcal{B}[y - \hat{y}]|$ exceeds 5 times its median.

Losses in the measurement domain. We compare measurements after range compression along k and apply a per-sub-aperture normalization so that sub-apertures with different numbers of positions contribute equally. The complex misfit is summed over all positions and frequencies with equal weight after this normalization.

D Implementation details

Carrier demodulation. Sub-aperture ℓ' is back-projected and demodulated by its mean spatial frequency, $\tilde{I}_{\ell'}(\boldsymbol{x}) = I_{\ell'}(\boldsymbol{x})\, e^{-\mathrm{i}\bar{\boldsymbol{\kappa}}_{\ell'}^{\top}\boldsymbol{x}}$. By Lemma 3, $\tilde{I}_{\ell'}$ varies on the scale of the PSF envelope rather than $\lambda/2$, so a grid coarser than the wavelength keeps its complex values. The look direction is encoded as $(\tilde{\boldsymbol{u}}_{\ell'}, \Delta\theta_{\ell'})$.

Encoder. Input volumes of $96 \times 96 \times 128$ voxels at 2.5 mm spacing, holding the carrier-demodulated complex BP image $\tilde{I}_\ell$ as two real channels plus its magnitude. The spacing is below the Nyquist spacing of the sub-aperture spectral window $\mathbf{C}_\ell$ (Lemma 3), which for our strips is about 4 mm across the look direction and 3.7 mm along it. Per sub-aperture: a 3D U-Net with 32, 64, 128, 256 channels. Cross-sub-aperture aggregation: 2 set-attention blocks with 8 heads applied per voxel (Lee et al., 2019). Gaussian head: 2 candidates per voxel; frames use the continuous 6D rotation representation; scales and curvatures are predicted in log and bounded ranges (scales 1–15 mm, curvatures $|\kappa| \le 1/(5\,\mathrm{mm})$).

Training. Three seeds per configuration; confidence intervals by bootstrap over test objects. Adam, learning rate 3×10^{-4} with cosine decay, batch size 8, 2×10^5 iterations, $\lambda_{\mathrm{geo}} = 1$, $\lambda_{\mathrm{sp}} = 10^{-3}$, strip dropout 0.2, one GPU for about 30 h. The two-dimensional encoder of Appendix G uses Adam at 10^{-3}, batch size 16 and 1500 iterations on two CPU cores (6–13 minutes per run).

Refinement. 300 Adam steps with learning rates of 0.5 mm for centers, 0.01 for frame angles, 0.05 for log-scales and $10^{-3}\,\mathrm{mm}^{-1}$ for curvatures; the stopping step is chosen by five-fold random cross-validation every 25 steps.

Simulator. Objects are meshed at $\lambda/10$ edge length and illuminated with the physical-optics approximation (Knott et al., 2004), integrating the induced surface currents over visible facets for every aperture position and frequency. Visibility uses ray casting from each aperture position. The simulator reproduces the streak directions and the waist of the measured scan (Section 6).

Baseline configurations. GeRaF is run with its public near-field renderer, with the antenna positions replaced by the rail positions and one transmit–receive pair; the SAR-GS-style baseline fits Gaussians to the magnitude of each strip image with the strip PSF of Lemma 3; mmNorm votes normals from the strip with the strongest return. All baselines use the same calibrated Δ.

E Notation

F Additional experiments

Symbol	Meaning
$\boldsymbol{p} \in \mathcal{A}$, $\mathcal{A}_\ell$	aperture position, sub-aperture ℓ ($\ell = 1, \dots, S$)
k, $\bar{k}$, f_c, B	wavenumber, center wavenumber, center frequency, bandwidth
y, y_ℓ, n_ℓ	measurement, sub-aperture measurement, noise
$\mathcal{F}_\ell$, $\mathcal{B}_\ell$, I_ℓ	sub-aperture forward operator, back-projection, BP image
$\hat{\boldsymbol{u}}(\boldsymbol{p})$, $\bar{\hat{\boldsymbol{u}}}_\ell$, $\Delta\theta_\ell$	look direction, mean look direction, angular span
$\mathbf{C}_\ell$, $\mathbf{P}_\ell$	spectral covariance and PSF covariance of sub-aperture ℓ
$G_j = (\boldsymbol{\mu}_j, \mathbf{U}_j, \boldsymbol{n}_j, \boldsymbol{\Lambda}_j, \mathbf{K}_j, \rho_j)$	coherent surface Gaussian
R_0, $\boldsymbol{d}$, $\boldsymbol{q}$, $\mathbf{H}$	range, unit direction, tangential look component, range Hessian
$\mathbf{M} = \boldsymbol{\Lambda}_j^{-1} + \mathrm{i}2k\mathbf{H}$	complex precision of the echo integral
$\boldsymbol{\Gamma}_j(k) = 4k^2\,\mathrm{Re}(\mathbf{M}^{-1})$	precision of the angular lobe
$\boldsymbol{u}_0$, a_0, $\boldsymbol{g}$	expansion point, amplitude and its log-gradient
$\mathbf{Z}_j = 2k\boldsymbol{\Lambda}_j^{1/2}\mathbf{H}\boldsymbol{\Lambda}_j^{1/2}$	normalized range Hessian
f_θ, $\mathcal{L}_{\mathrm{HO}}$	encoder, held-out loss over folds of positions
$w_{j\ell}$	lobe share of sub-aperture ℓ for primitive j

Table 9: **Renderer error versus primitive size** on the pipe ($r = 2\,\mathrm{cm}$). Normalized complex error with respect to physical-optics quadrature over the true surface; in parentheses, with respect to dense quadrature over the same Gaussians. Lobe culling at $\tau_{\mathrm{lobe}} = 6.9$ adds less than 0.1%.

Primitive size	3 mm	5 mm	8 mm	12 mm
#Primitives	1134	400	160	70
Dense quadrature on the Gaussians	5.7%	9.1%	14.1%	28.6%
Points at the Gaussian centers	143%	277%	451%	668%
Flat Gaussians, re-centered	24%	50%	80%	98%
Curved, center expansion	5.6% (0.5%)	9.0% (1.4%)	14.3% (2.8%)	31.7% (8.8%)
Curved, re-centered	5.6% (0.3%)	9.1% (0.4%)	13.9% (0.7%)	28.3% (0.8%)

G AMORTIZED ENCODER IN TWO DIMENSIONS

Setup. The rail of Appendix J is sampled at 500 positions (1.2 mm pitch) and 32 frequencies over 77–80.9 GHz, which is alias-free for the 20×20 cm scene window; five contiguous strips of 100 positions. Scenes contain 1–3 objects from the families pipe ($r = 0.5$–4 cm) and flat segment (3–10 cm, random tilt), with a plate behind in 30% of the training scenes; a third family, ellipses, is held out for testing. Echoes are simulated by the physical-optics simulator of Appendix J with single and occlusion-tested double bounces and the error model of that appendix. There are 3000 training, 200 validation and 300 test scenes per family set, and 1000 unlabeled plus 300 test scenes of a shifted domain (stronger phase error, a different element pattern and a small range offset).

Encoder. Each strip is back-projected to a 64×64 grid over $y \in [-10, 10]$ cm, $z \in [25, 45]$ cm and demodulated by its carrier. A shared convolutional network encodes each strip with its look direction, a mean-and-attention pooling aggregates the strips, and a head on a 16×16 grid (12.5 mm cells) predicts one two-dimensional surface Gaussian per cell: existence, offset, normal angle, width and curvature, with the complex reflectivity solved by least squares. A differentiable port of the two-dimensional closed-form renderer matches the reference implementation to 10^{-13} in double precision.

Sensitivity of the held-out losses to geometry. Before training, we measured how each held-out loss responds to controlled errors of the true geometry (primitives every 3 mm on the true surface; 30 test scenes, five folds). The strip loss is 0.680 for the true geometry, 0.679 with 1 mm normal jitter, 0.696 for a 3 mm shift in depth, 0.814 for a 3 mm lateral shift, 0.699 for flat primitives and 0.683 with ten added phantoms. The random-fold loss is 0.093, 0.094, 0.111, 0.119, 0.093 and 0.093 respectively, and 0.431 for a 10 mm lateral shift (strip: 1.291). Neither loss penalizes added phantoms, because the reflectivity of a primitive that no fold needs is fitted to near zero; the random-fold loss is the more sensitive to misplaced surface.

Table 10: **Ablations** on measured objects.

Variant	CD ↓	CV ↓	PR ↓
Full model (encoder + refinement)	2.6	0.34	0.09
Strip hold-out instead of CV-r in refinement	3.3	0.38	0.07
No cross-validated stop (fixed budget)	2.9	0.35	0.14
Center expansion instead of re-centered Theorem 2	2.8	0.35	0.10
Born-density volumetric Gaussians	4.4	0.42	0.22
Flat instead of curved patches ($\mathbf{K}_j = 0$)	3.9	0.40	0.20
Free angular lobe instead of derived lobe	3.5	0.39	0.17
Intensity instead of coherent rendering	6.2	0.55	0.35
Uncalibrated range offset Δ	4.8	0.34	0.10
$S = 3$ / 5 / 8 strips as encoder input	2.9 / 2.6 / 2.6	0.36 / 0.34 / 0.34	0.11 / 0.09 / 0.09

Table 11: **Three-dimensional evaluation of amortization** on simulated objects. CD (3/5): only three of five strips available at test time.

Training	CD (seen) ↓	CD (unseen) ↓	CD (3/5) ↓	CV ↓	Time (s) ↓
Per-object optimization (no training)	2.1	2.3	3.0	0.31	240
Sim., $\mathcal{L}_{\text{geo}}$ only	2.8	3.9	3.6	0.38	0.3
Sim., self-reconstruction only (no labels)	6.5	7.1	7.4	0.33	0.3
Sim., $\mathcal{L}_{\text{HO}}$ strips only (no labels)	8.9	9.4	9.8	0.45	0.3
Sim., $\mathcal{L}_{\text{HO}}$ random folds only (no labels)	5.8	6.4	6.9	0.32	0.3
Sim., $\mathcal{L}_{\text{geo}}$ + refinement	2.0	2.5	2.7	0.30	12
Sim. + label-free fine-tuning + refinement	1.9	2.3	2.6	0.29	12

Results. Table 3 reports the test metrics. Training with the strip hold-out loss alone does not reduce it below the level of predicting nothing: after 1500 iterations it stays at 1.0–1.4 in all three seeds, consistent with the sensitivity analysis. Geometric supervision in simulation gives the best amortized model, and label-free fine-tuning with the random-fold loss on the shifted domain lowers the Chamfer distance by about 10%, against no gain for strip hold-out.

H MEASURED DATA

Acquisition. A TI AWR2243 radar is moved on a two-axis rail. A single transmit–receive pair (chirp 1, receiver 4, a few millimeters apart) records a 77–81 GHz FMCW sweep at each position. The rail runs 1000 positions along x and 1040 lines along y, both at a $\lambda/4 \approx 0.95\,\text{mm}$ step, in a snake pattern; the aperture is the plane $z = 0$ and z is range. Each capture file holds one rail line, so a block of files is a horizontal strip of the aperture. The data are back-projected onto a 128^3 grid over $x \in [-0.25, -0.10]\,\text{m}$, $y \in [0.35, 0.45]\,\text{m}$ and $z \in [0.80, 1.10]\,\text{m}$ (voxel $1.18 \times 0.79 \times 2.36\,\text{mm}$). No range window is applied, and the per-channel calibration contains a gain and phase but no range delay.

Sub-apertures and linearity. The five sub-apertures are lines 0–200, 200–400, 400–600, 600–800 and 800–1039, with aperture centroids at $y = 8.7$, 27.7, 46.7, 65.7 and 86.5 cm. Their images sum to the full-aperture image with a relative residual of 1.4%, which is exactly the contribution of the four boundary lines counted twice (Lemma 2). The full-aperture peak is only half as bright as the sum of the sub-aperture peaks (0.50), whereas an isotropic point scatterer would give a ratio near one: the five sub-apertures see responses that differ in position and phase.

Streaks follow the look direction. Table 13 lists, for each sub-aperture, the direction of its main lobe and the mean direction from which it views the object. Across a 46° span the two agree with correlation 0.9993, slope 1.04 and a root-mean-square difference of 0.9°. Sub-apertures 1–4 agree within 0.5° at the -3, -6 and $-10\,\text{dB}$ levels; at $-6\,\text{dB}$ the lobe of sub-aperture 5 merges with a second response near $y = 39.5\,\text{cm}$.

Table 12: **Further baselines** on measured objects: nonlinear sub-aperture fusion, each followed by iso-surface extraction, and the image-domain and sparse baselines omitted from Table 4.

Method	CD ↓	PR ↓	WR ↑
Pixelwise minimum of $\lvert I_\ell \rvert$	7.5	0.55	0.24
Coherence factor (Mallart & Fink, 1994)	8.1	0.58	0.28
Generalized coherence factor (Li & Li, 2003)	7.8	0.56	0.29
SLSC (Lediju et al., 2011)	7.9	0.57	0.27
3DGS (Kerbl et al., 2023) on BP images	8.6	0.62	0.31
Learned super-resolution (Smith & Torlak, 2023)	6.8	0.49	0.36
Sparse inversion (ℓ_1) with point scatterers	4.9	0.33	0.44

Table 13: **Aperture parallax on the measured scan** (y–z plane). Look: mean direction from the sub-aperture to the full-aperture peak, from the range axis. Streak: principal axis of the -3 dB main lobe. δ_k: signed distance of the sub-aperture peak beyond the common crossing point of the five look lines. Level: peak relative to the full aperture.

Sub-aperture	Look (°)	Streak (°)	Elongation	δ_k (mm)	Level (dB)
1 (lines 0–200)	$+20.4$	$+20.6$	3.7	31.8	-15.1
2 (200–400)	$+9.3$	$+9.9$	4.2	24.2	-8.1
3 (400–600)	-2.7	-2.9	4.1	21.5	-5.3
4 (600–800)	-14.3	-13.9	3.4	22.2	-5.9
5 (800–1039)	-25.9	-27.7	3.1	27.6	-8.8

A common crossing point. The five lines through the sub-aperture peaks along their look directions meet in a least-squares point at $(y, z) = (42.45, 89.00)$ cm, with perpendicular misses of 0.35–1.7 mm. This point coincides with the narrowest -6 dB width of the full-aperture image in the y–z plane ($z = 88.98$ cm, width 3.1 mm), the waist of the hourglass (Figure 6); the brightest voxel lies 1.5 cm further, at $z = 90.52$ cm. Under ideal monostatic BP and narrow strips, Theorem 1(ii) places this crossing at the center of curvature, or equally at a point scatterer. The processing phase convention discussed below shifts that interpretation on this scan.

What the offsets identify. The peaks lie $\delta_k = 21.5$–31.8 mm beyond the crossing (Table 13). For ideal narrow strips and a single circular glint, these values would give $\Delta - r \approx 2.2$–3.2 cm. The strips are wide and the processing phase convention shifts the crossing, so this direct reading does not apply; the exact scan model below gives $\varrho \approx -5.5$ cm. The channel calibration contains no range delay, so Δ and r remain separate unknowns. The full-aperture peak to sub-aperture peak ratio of 0.50 shows that the strips see responses at different positions, that is $\varrho \neq 0$, which is equally consistent with a point scatterer measured through a delay offset.

Reconciling strip and full-aperture offsets. Two observations seemed not to fit a single glint: the strip offsets vary by 10 mm, U-shaped over look angle, and the full-aperture peak lies 15 mm beyond the crossing, where Proposition 1 with narrow strips predicts about 5 mm, and where matching it with narrow strips would require an implausibly narrow element pattern. Both come from treating the strips as narrow. Every strip spans the full 0.95 m line in x, an azimuth range of $54°$, so its own diversity ratio is $\zeta_\ell \approx 3.9$–5.5 with uniform weights, against 9.9 for the full aperture when both dimensions are counted. Proposition 1 applied to strips and full aperture alike predicts a ratio of full-aperture to strip offset of 0.51–0.60 for any element pattern from $\cos^1$ to $\cos^{20}$, and 0.61 is measured. The central strips are closer to the object and see a wider range of x angles, so their ζ_ℓ is larger, which produces the U-shape: relative offsets of $1.20, 1.02, 1, 1.09, 1.33$ are predicted and $1.47, 1.13, 1, 1.03, 1.28$ measured. The implied $\lvert\varrho\rvert$ is, however, outside the small-phase regime of the proposition, so we checked it by exact simulation.

Phase reference of the processing library. The images were formed by the authors' MATLAB library, which we read as source code. Its range transform references the phase to the frequency at the first sample of the sweep, whereas its aperture phase correction uses the wavelength at the center of the band. The image is then the ideal image multiplied by a phase that grows with range,

which references the aperture phase effectively to the top of the band. In simulation with a replica of the library, this moves the crossing of the strip look lines about 2.7 cm beyond a point scatterer, whatever Δ, and gives a point without offset strip offsets of -2.5 to -6 mm. The measured image supports this reading: its phase advances by -1.39 rad per range voxel, against -1.1 to -1.45 for simulations with the library and -0.91 to -0.99 for ideal back-projection. The conclusion depends on the intermediate-frequency sign convention, inferred from a comment in the library. Under this convention the crossing on this scan can lie up to 2.7 cm beyond the center of curvature.

Exact simulation of the scan. We simulated the exact aperture (positions decimated by 8 in each direction, 0.6% error against a decimation of 4), 64 frequencies and the library's back-projection, and extracted the same observables as from the measurement. Because the range sidelobes are not windowed, the observables respond non-linearly to ϱ: for a point with offset, the strip offsets stay below 7 mm up to $\varrho = -45$ mm and then jump to 27 mm at -60 mm. Over 176 models, the best fit ($\chi^2 = 17.9$) has an apparent radius of -55 mm in the y–z plane and -70 mm in the x–z plane, isotropic elements, and a target lobe tilted by 7.5° with a width of 20°; Table 14 compares it with the measurement. The y–z radius is constrained to -4 to -6 cm, the x–z radius only weakly. A point with offset ($\varrho = -60$ mm in both planes) fits worse ($\chi^2 = 49$–52), mainly in the lateral position of the full-aperture peak and in the fifth strip, and no model with $|\varrho| \leq 30$ mm fits. The grid is coarse (5–10 mm steps), the lateral position of the full-aperture peak is still misfit by 3 mm, and the fit assumes the nominal band and frequency spacing.

Table 14: **Measured observables and the best single-glint model** (exact simulation with the library's back-projection).

Observable	Measured	Best model
Strip offsets beyond the crossing (mm)	31.8, 24.3, 21.6, 22.3, 27.6	28.2, 24.4, 23.7, 26.0, 31.3
Full-aperture peak beyond the crossing (mm)	15.3	12.2
Full-aperture peak, lateral offset (mm)	$+0.1$	-3.0
Strip levels (dB)	$-15.1, -8.1, -5.3, -5.9, -8.8$	$-14.1, -8.1, -5.2, -5.8, -8.3$
Full-aperture to summed strip peak	0.50	0.49

What $\varrho \approx -5.5$ cm means. With the library modeled, ϱ is about -5.5 cm: a point scatterer seen through $\Delta \approx 5.5$–6 cm, a convex pipe seen through $\Delta \approx r + 5.5$ cm, a concave reflector of radius about 5.5 cm with $\Delta = 0$, or a virtual glint. Read from the images as if the back-projection were ideal, $|\varrho|$ is about 2.5 cm smaller, close to the value of 2.7 cm obtained from the strip offsets directly.

A corner reflector as the reference. A trihedral corner reflector returns every ray that enters its aperture from the apex, so over the look angles of the rail it acts as a point scatterer, $r = 0$. By Theorem 1, its strip look lines cross at the apex and its strip peaks lie at $-\varrho = \Delta$ beyond the crossing; the offset of the reflector is therefore the range offset of the system. A scan of a corner reflector was taken with the radar in the same place as for the pipe, so one Δ applies to both scans. The full-aperture image (screenshots provided with the scan) shows a compact peak near $(x, y, z) = (-9.6, 58.5, 96.5)$ cm, whose leading edge lies at $z \approx 94$–95.5 cm, followed by a fan of energy that extends to $z \approx 102$ cm. The measured shape is consistent with a point scatterer seen through a positive offset $\Delta > 0$. The reflector scan therefore checks the sign of the offset qualitatively; the reported pipe fit remains the apparent radius ϱ, without assigning a separate numerical value to Δ or r.

Limitations of this scan. Every sub-aperture spans the full x line, so parallax is observed only in the y–z plane. In the x–z plane, the fan behind the waist tilts by -7° to -12°, consistent with the -9.5° mean look angle of the x aperture onto the off-center object. Only y–z parallax and the apparent radius are analyzed here; the one-plane geometry does not identify the true radius independently.

I SIMULATED SCENES

Simulator. Scenes are triangle meshes rendered by scalar physical optics: each flat facet contributes its exact far-zone integral, with the two-way phase, an element pattern $\cos^{1.3}$ per way, $1/R^2$

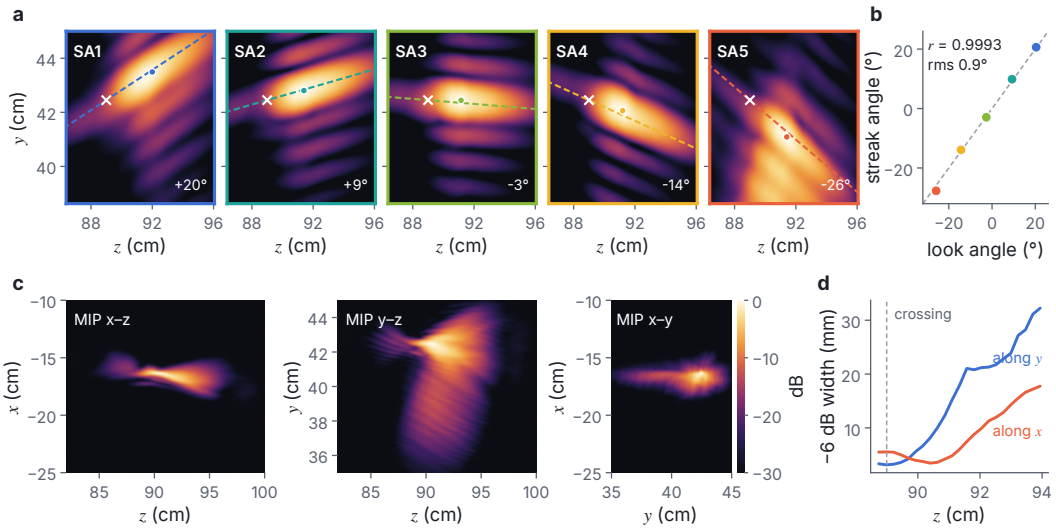

Figure 6: **Aperture parallax on measured data** (77–81 GHz; range z to the right). (a) Back-projection of the five strips of the 0.99 m aperture (SA1, lowest, to SA5), y–z slice through the object: each streak lies along its strip's look line (dashed), the look lines cross at one point ($\times$), and each strip peak (dot) lies beyond it. Numbers: mean look angle. (b) Streak angle against look angle. (c) Full-aperture maximum-intensity projections, 30 dB. (d) -6 dB width of the full-aperture image along y and x versus depth; the narrowest point along y, the waist, coincides with the crossing ($z = 89.0$ cm).

spreading and z-buffer visibility. The radar sweeps 77–81 GHz at 64 frequencies. We add 25 dB noise, a correlated aperture phase error of 0.1 rad rms, a frequency-calibration ripple of 0.5 dB and 0.05 rad, and 50 μm of position jitter. Two-dimensional slices use a 0.6 m rail with 1000 positions; volumes use a 0.4×0.6 m planar aperture on a 1 mm grid (241,001 positions) and phase-shift migration. A point target focuses at its true position, and migration matches exact back-projection. Facets of 2.5 mm give a relative error of 0.2% against 1 mm facets. The simulator omits polarization and edge diffraction, and multipath paths are not tested for occlusion. It is independent of Theorem 2, so it can train the encoder without an inverse crime.

What a synthetic aperture sees. Figure 7 shows three scenes. Only the cap of the sphere that faces the aperture appears, and its image focuses 28 mm behind the front surface. The bracket's bend glints, and its flange, whose normal misses the aperture, vanishes. The cardboard walls facing the aperture appear as bright sheets in front of and behind the mug. Figure 8 colors each facet by the sub-apertures that see it within a 10° acceptance angle. The strips see different bands of normals, as the cone condition of Section 3 states, and 18–31% of each surface is seen at all. The 10° acceptance angle is a geometric stand-in for the lobe of Lemma 1: at 79 GHz and 0.3 m, a flat 5 mm element has a lobe standard deviation of 3.6°, so 10° corresponds to about three standard deviations; the whiskers at 5° and 15° show the sensitivity. Figure 9 places the pipe 6 cm in front of a 24×24 cm metal plate. A ghost hourglass appears behind the plate, 3.4 dB below the pipe, with its peak 1.9 cm behind the mirror image of the pipe's front surface. This simulation does not test the multipath legs for occlusion, so the ghost appears in every sub-aperture and overstates the mirror image. The two-dimensional study of Appendix J repeats the geometry with occlusion-tested paths: the pipe then blocks the mirror paths in the central strip, and the double bounce appears on the plate itself.

J A TWO-DIMENSIONAL STUDY OF HELD-OUT SELECTION

This study isolates model selection for per-object fits: does choosing a model by held-out strips remove phantoms that self-reconstruction keeps, how does it compare with generic cross-validation, and which multipath ghosts can any rule reject? The first part below is the original three-seed study, which compared strip hold-out only with BIC; the second part (Appendix J.1) repeats it with

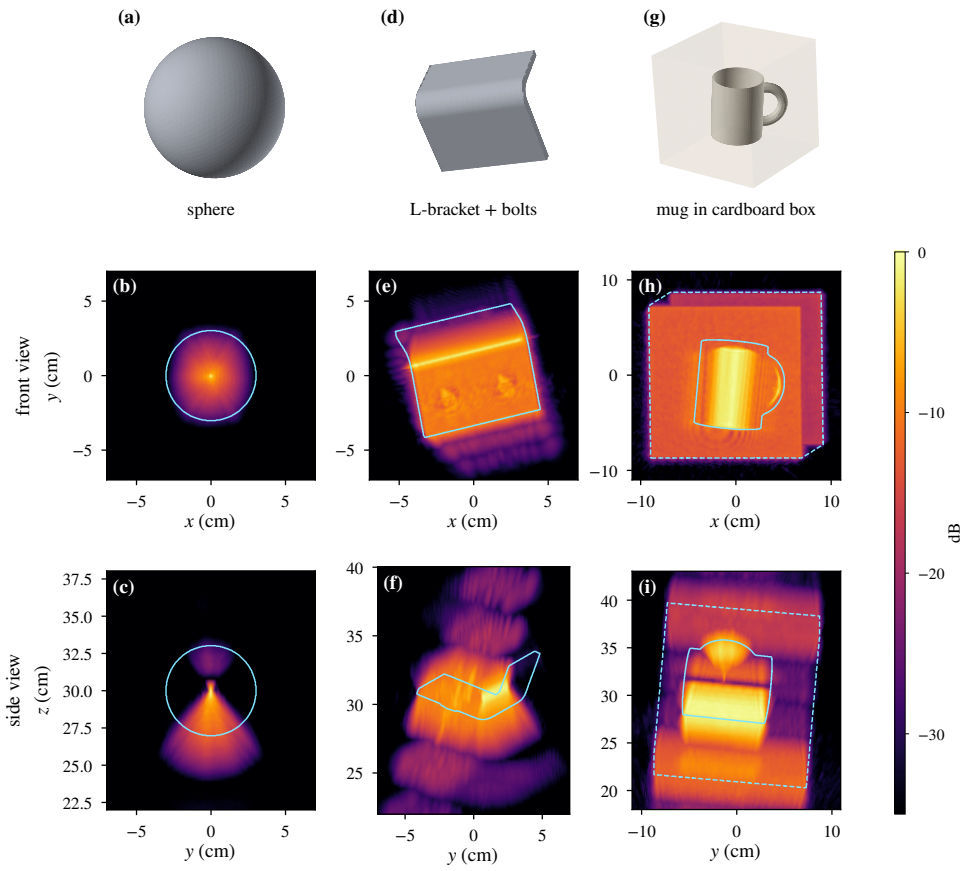

Figure 7: **Simulated three-dimensional back-projection** of a sphere ($r = 3$ cm), an L-bracket with bolts and a mug in a cardboard box, each 0.3 m from a 0.4×0.6 m aperture. Rows: geometry, front view (maximum over range) and side view (maximum over x), 35 dB; cyan: true outline.

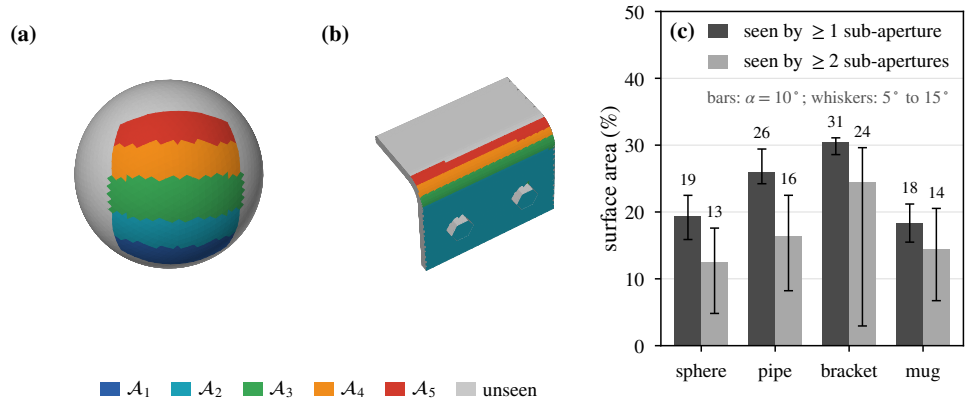

Figure 8: **Each sub-aperture sees a different band of normals** (simulation). (a, b) Facets of a sphere and a bracket colored by the strip that sees them specularly (10° acceptance); gray: unseen. (c) Fraction of the surface seen by at least one and at least two strips at 10°; whiskers: 5° and 15°.

ten seeds and the full set of rules of Table 2. Code, per-seed results and a description of every approximation are in the supplementary material (`sim2d/`).

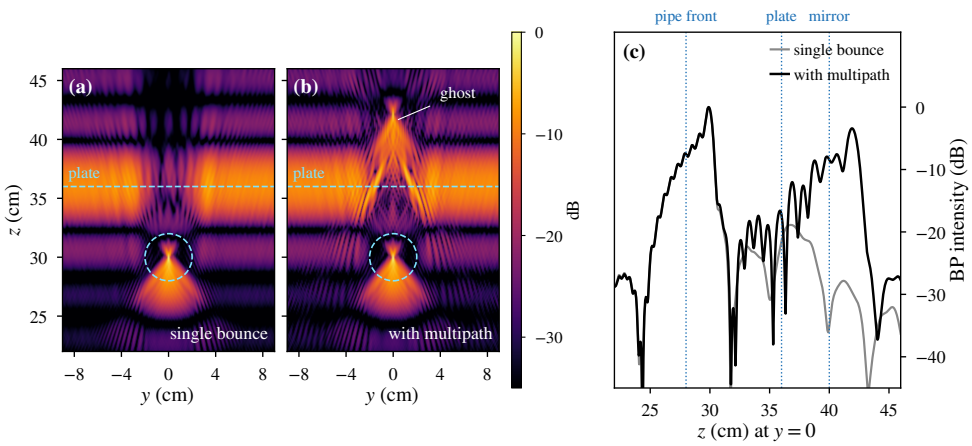

Figure 9: **Multipath** (simulation). A pipe 6 cm in front of a 24×24 cm metal plate, (a) single scattering and (b) with pipe–plate paths. (c) Profile at $y = 0$: the ghost appears near the mirror image of the pipe behind the plate.

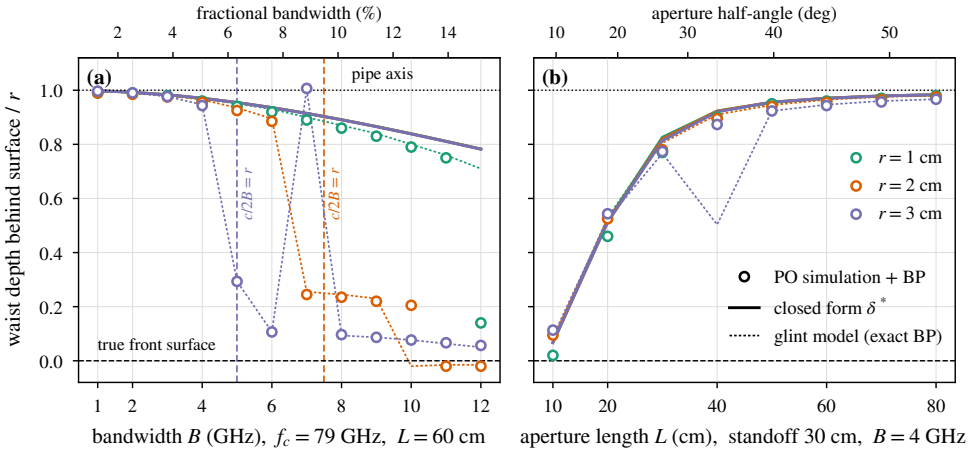

Figure 10: **Waist depth of a pipe** in simulation (circles), for an exact glint-only model (dotted) and from Proposition 1 (lines), (a) versus bandwidth at $L = 0.6$ m and (b) versus aperture length at 4 GHz. Where $c/2B \gtrsim 1.3\,r$, Proposition 1 is within 0.8 mm of the simulation (median 0.23 mm, 33 of 57 cases). The glint-only model tracks the simulation to a median of 0.1 mm in every case.

Simulator. Two-dimensional scalar physical optics in the y–z plane (objects infinite along x), with a 0.6 m rail of 1000 positions at $z = 0$, 64 frequencies over 77–81 GHz, an element pattern $\cos^{1.3}$ per way, and five disjoint strips of 200 positions. Single scattering is integrated over 0.2 mm facets. Double scattering is a physical-optics double integral over all mutually visible pairs of 0.5 mm facets, and every leg, including the multipath legs, is tested for occlusion; it agrees with image theory to 7–12% in the strips where it matters. The plate–pipe–plate path is added by image theory in one variant. The data carry the error model of Appendix I: 25 dB noise, 0.1 rad correlated aperture phase error, frequency ripple and 50 μm position jitter, drawn anew for each of three seeds.

Renderer and fitting. A 2D coherent surface Gaussian has center, normal angle, width s, curvature κ and complex reflectivity; its echo is the one-dimensional form of Theorem 2 with two saddle re-centering steps. Against quadrature with the exact range and amplitude, over nine flat, convex and concave patches, the global error is at most 1.6% and the worst per-position error inside the lobe 3.2%; expansion at the center errs by up to 62%. Fits are greedy: at each of $T = 24$ steps, candidates are proposed at the peaks of the per-strip back-projections of the residual, refined singly and jointly by Levenberg–Marquardt, and pruned. Self-reconstruction (SR) chooses the step by BIC on the loss

over all strips; held-out selection (HO) runs the same path on each set of four strips and chooses the step that minimizes the mean error on the fifth. Only the stopping step differs, and nothing was tuned per scene or per method. Held-out prediction is thus used for model selection, not as a training loss.

Metrics. A fitted primitive is sampled along its arc, weighted by its Gaussian profile and by whether its local normal ray reaches the rail. The phantom rate PR_E is the response-energy-weighted fraction farther than $\tau = 2\,\mathrm{mm}$ from every true surface, and recall is the fraction of the visible true surface within τ of a significant primitive. Back-projection is scored on its $-6\,\mathrm{dB}$ region.

Table 15: **Held-out selection in simulation** (mean $\pm$ std over 3 seeds). PR_E: response-weighted phantom rate. Rec.: recall. t: selected step. Back-projection: $-6\,\mathrm{dB}$ region of the full aperture, and its energy-weighted distance to the true surface.

	Back-projection		Self-reconstruction (BIC)			Held-out strips		
Scene	PR	dist. (mm)	PR_E	Rec.	t	PR_E	Rec.	t
Pipe, $r = 2\,\mathrm{cm}$	1.00	15.8	0.009 ± 0.009	1.00	24	0.006 ± 0.005	1.00	18 ± 9
Pipe + plate	0.87	8.4	0.034 ± 0.027	1.00	24	$\mathbf{0.005 \pm 0.003}$	1.00	9 ± 1
Pipe + plate, 3rd order	0.88	8.9	0.058 ± 0.003	1.00	24	0	0.16	1
Two pipes	1.00	14.1	0.113 ± 0.015	1.00	24	0.118 ± 0.009	1.00	11 ± 3
Two pipes, $\kappa \leq 0$	1.00	14.1	0.132 ± 0.027	1.00	24	0.095 ± 0.037	1.00	8 ± 4
Dihedral, 75°	0.84	7.8	0.610 ± 0.020	0.43	21 ± 5	0.603 ± 0.058	0.46	8 ± 1

Which ghosts migrate, and which are rejected. Figure 11 tracks the apparent position of each ghost across the five strips. Next to the plate, the pipe–plate double bounce appears on the plate itself, 3.4 dB below the pipe, in the outer four strips only; in the central strip it is 33 dB weaker, because the pipe blocks its paths. The plate–pipe–plate path forms the mirror image of the pipe 4 cm behind the plate, 5 dB below it, again absent from the central strip. The three-dimensional simulation of Appendix I does not test multipath legs for occlusion and therefore overstates this mirror ghost. The interreflection between two pipes and the double bounce of the dihedral are seen by all five strips and migrate by 24 and 21 mm, as much as a real glint (24 mm), but in the opposite direction: they fit one concave virtual reflector ($r_v = -19$ and -16 mm) to 1 mm rms, which is Proposition 2 at work.

Table 15 and Figures 12 and 13 show the consequences. Self-reconstruction fits of surface Gaussians reduce the phantom rate of back-projection by an order of magnitude on every scene but the dihedral, with full recall. Held-out selection removes the phantoms that SR fits next to the plate; the ten-seed study below shows that it does so by stopping early, at a large cost in recall, and that a fixed early stop does as well. It does not remove the two-pipe ghost, which its virtual reflector predicts in every strip (lobe share about 0.5). Restricting primitives to convex curvature excludes that reflector; the ghost is then fitted by single-strip phantoms, and held-out selection helps modestly and not on every seed (phantom counts 9, 7, 7 under SR versus 6, 7, 1). In the dihedral, the double bounce dominates the echo and both variants fail. On a single pipe the two criteria differ little.

Limitations of the study. It is two-dimensional and scalar, and it models third-order paths only in one variant. With a flat plate in view, held-out selection is noisy: each strip sees a different part of the plate, whose glint no other strip predicts (mean held-out error near 1), and with the third-order path it degenerates to a single primitive (recall 0.16). The strips are contiguous, so predicting an end strip is an extrapolation. BIC keeps adding primitives because model error (phase error, jitter, renderer approximation) exceeds the noise; a better-calibrated likelihood would narrow the gap between SR and HO. Three seeds give only a coarse estimate of variance.

J.1 Ten-seed study with generic cross-validation

Changes. Ten noise seeds per scene; each seed redraws noise, aperture phase error, ripple and jitter. The greedy path has $T = 16$ steps and uses every second rail position (500) and 64 frequencies. All rules stop the same path: strip hold-out (HO), interleaved 12 mm blocks (CV-i), random folds

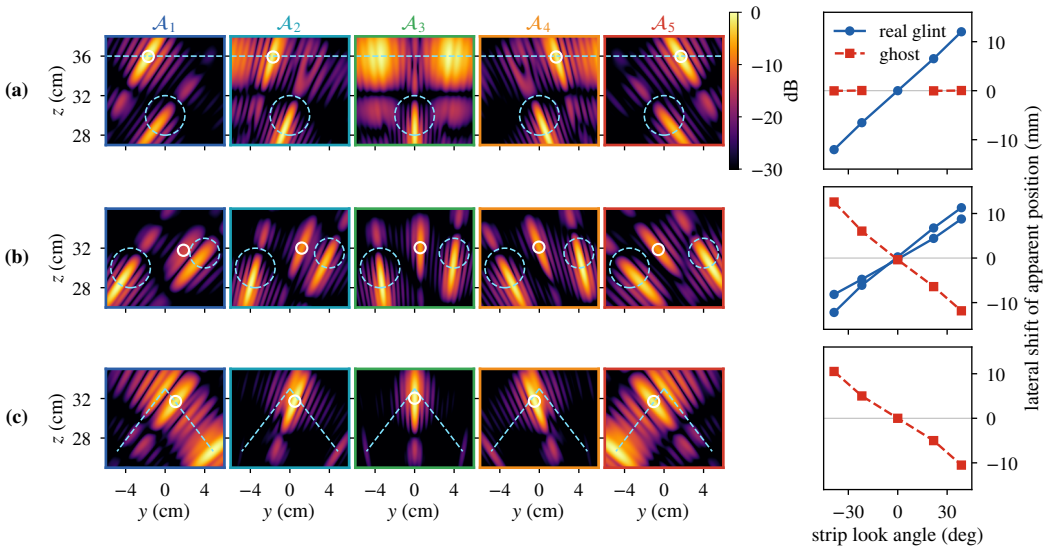

Figure 11: **Ghost migration** (2D simulation, noise-free components). Per-strip back-projection of the multipath echo for (a) pipe and plate, (b) two pipes and (c) a 75° dihedral; circles mark the apparent ghost position, dashed lines the true surfaces. Right: lateral shift of the apparent position versus strip look angle. The plate ghost stays on the plate and is missing from the central strip; the two-pipe and dihedral ghosts move opposite to real glints, as a concave virtual reflector would (Proposition 2).

redrawn per seed (CV-r), the discrepancy principle with the residual power $\delta^2 = 0.048$ implied by the error model (DP) or with the residual floor of the converged path, AIC, BIC, the fixed budget, and a fixed early stop at the median HO step of the other nine seeds (ES), which tests whether HO adapts to the scene. Families: curved surface Gaussians, flat ones ($\kappa = 0$), and point scatterers with the exact point response. Intervals are 95% percentile bootstrap intervals over seeds; paired comparisons use sign and Wilcoxon signed-rank tests with Holm correction over the eight rules.

Findings. (i) The held-out strip error rises along the path (pipe and plate: 1.10 at the first step, 1.59 at the last), while the random-fold error falls from 0.82 to 0.05. (ii) HO stops after a median of 3 steps (per-seed 1–12); its phantom rate is lower than every other rule's (pooled over the three scenes, -0.022 against CV-r, Holm $p = 0.009$), but its recall is lower by 0.41 ($p < 10^{-4}$), and plate recall falls to 0.61 and 0.33 in the two plate scenes. (iii) ES matches HO in phantom rate ($+0.008$, $p = 0.78$) and recall ($+0.03$, $p = 0.67$). (iv) The per-strip held-out error at the HO step is 0.84–0.92 for the end strips and 1.19–1.31 for the central ones, so the end strips, which are extrapolations, are predicted better than the central ones. (v) Curved surface Gaussians have a phantom rate 0.105 lower than flat ones under DP in all 20 plate-scene seeds, with recall higher by 0.063; point scatterers with the same budget reach 0.54–0.68 and recall 0.05–0.06, because 16 points cannot cover a plate. (vi) In the dihedral, every rule fails (phantom rate 0.61–0.63, recall 0.29–0.52).

Lobe share. Figure 14 relates the maximal lobe share of each of 441 self-reconstruction primitives (scenes pipe + plate, + third order, dihedral) to whether it is a phantom and to whether each rule removes it. The phantom probability does not increase with lobe share (logistic slope -0.10 per 0.1, interval $[-0.15, -0.04]$); HO removes 42–76% of primitives in every lobe-share bin, real and phantom alike, and keeps 38% of the response energy, against 89% for CV-i and 99% for CV-r. The expectation that strip hold-out would single out one-strip structure is therefore not borne out: in these scenes, one-strip structure is mostly real plate.

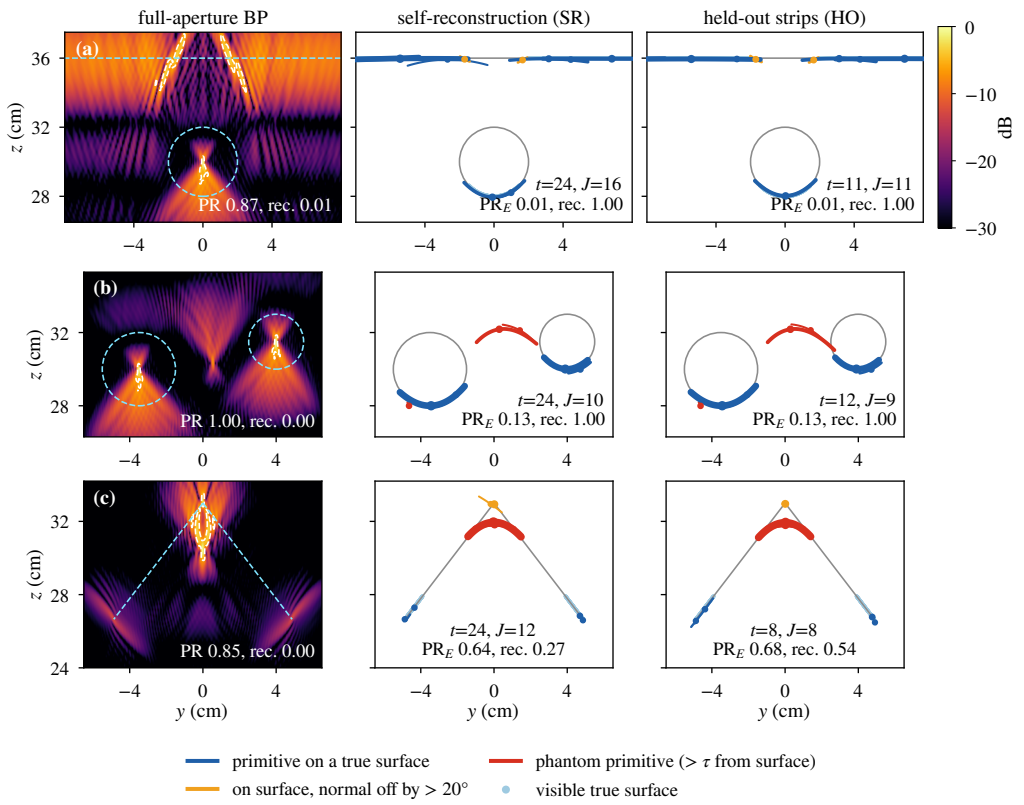

Figure 12: **Per-object fits** (seed 0). Left: full-aperture back-projection with the true surfaces (dashed). Middle and right: primitives fitted with self-reconstruction and with held-out selection; blue lies on a true surface, red is a phantom, orange lies on a surface with a normal off by more than 20°. t: selected step; J: number of primitives.

K Does the amplitude identify the radius?

Theorem 1 states that the phase history of a glint depends on its radius r and the range offset Δ only through $\varrho = r - \Delta$, while the amplitude retains a factor $(R^\star/(R^\star + r))^{1/2}$. This appendix asks whether that factor can be used to estimate r in practice.

Setup. A metal cylinder of radius $r \in \{1, 2, 3\}$ cm is placed with its axis at $D \in \{0.3, 0.6, 0.9\}$ m from a monostatic rail at $z = 0$, 77–81 GHz, 25 dB SNR. Echoes are computed by physical-optics quadrature. Across the rail, the factor $(R^\star/(R^\star + r))^{1/2}$ changes by only 0.03–1.3% (largest for the largest, closest pipe), and its logarithmic derivative with respect to r is correlated at 0.999 or more with that of an element pattern $\log\cos\theta$ and with $\log R$. The radius information in the amplitude is therefore nearly collinear with the two nuisance terms that any real system must model.

What each part of the echo determines. Fitting the phase alone determines ϱ to 0.02–0.14 mm, and the likelihood in (r, Δ) is a ridge along $r - \Delta = \varrho$ (Figure 15a), as the theorem states: along the ridge the misfit stays within the noise ($\chi^2 \leq 1.1$) for every r from 0.2 to 5 cm in all nine configurations, while a step of 0.1 mm off the ridge raises it to 32. Physical optics adds small terms to the phase that oscillate with r; they stay below the noise along the ridge, but they depend on the scattering model (for the TE polarization of an exact cylinder solution they would bias a physical-optics estimate by up to 6 mm at $r = 1$ cm). Fitting the amplitude alone with a *perfectly* known model determines $r = 2$ cm with a Cramér–Rao standard deviation of 0.4 cm at 0.3 m, 1.8 cm at 0.6 m and 5.4 cm at 0.9 m, comparable to r itself at the longer ranges.

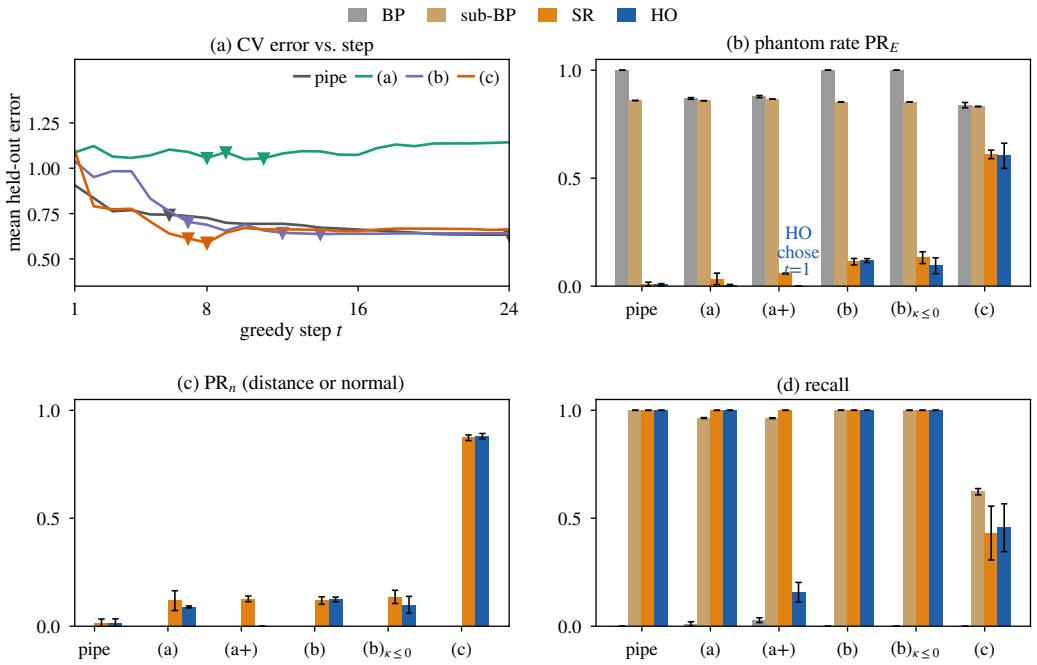

Figure 13: **Metrics over three seeds.** (a) Mean held-out strip error along the greedy path; triangles mark the selected step per seed. (b) Response-weighted phantom rate, (c) phantom rate counting normal errors, (d) recall, for back-projection, the union of per-strip back-projections, self-reconstruction and held-out selection.

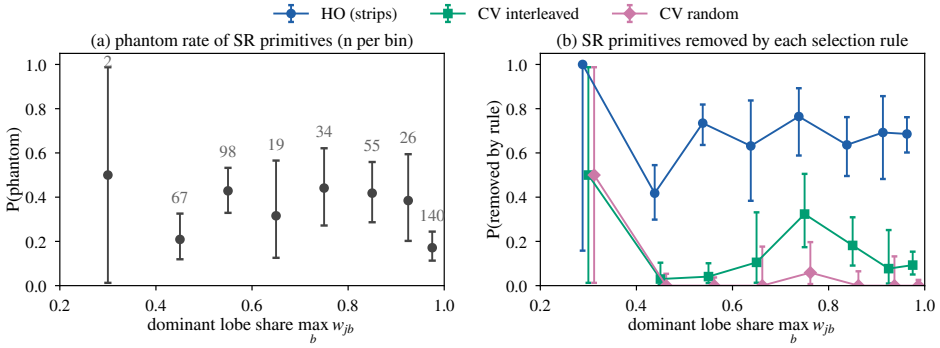

Figure 14: **Lobe share and model selection** (ten seeds, 441 primitives). (a) Phantom probability versus maximal lobe share. (b) Fraction of primitives removed by strip hold-out and by interleaved and random cross-validation, versus maximal lobe share.

Sensitivity to model error. The amplitude estimate collapses under mild mis-specification (Figure 15b). For $r = 1$ cm, an element-pattern exponent that is wrong by ± 0.1 biases $\hat{r}$ by $+5$ to $+15$ cm and -8 to -23 cm (larger at longer range), and an error of ± 0.3 drives it to the bounds of the search. An unmodeled quadratic roll-off of the gain by 0.1 dB at the ends of the rail biases $\hat{r}$ by $+1$ to $+17$ cm. A linear gain drift of 0.5 dB biases it by less than 0.8 cm for a target centered on the rail, but by $+3$ to $+31$ cm for a target 15 cm off center. An incidence-dependent reflectivity of up to 3 dB at $45°$ has no measurable effect, because a monostatic glint is always seen at normal incidence. As a rule of thumb, an error δq in the pattern exponent shifts $\hat{r}$ by about $2\,\delta q$ times the range. The same pattern holds for $r = 2$ and 3 cm. Since the element pattern of a patch antenna on a radar board is known to a fraction of a decibel at best, we treat the amplitude as uninformative about r and calibrate Δ with a reference target.

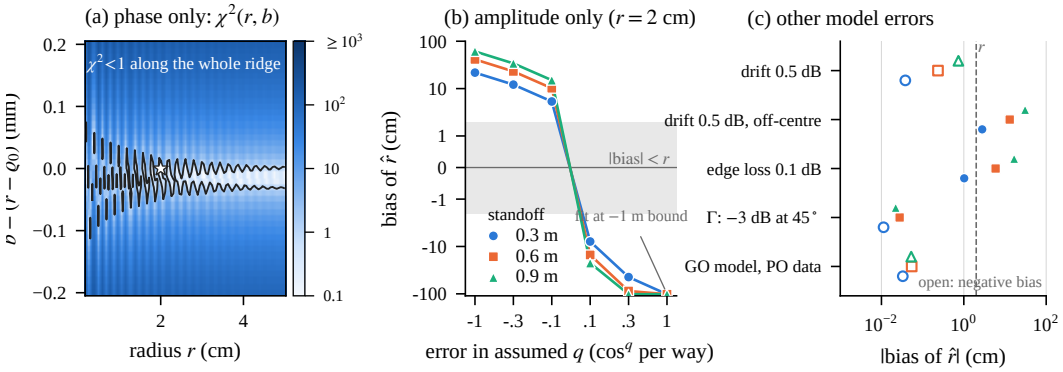

Figure 15: **Phase and amplitude information about the radius** of a cylinder. (a) Phase-only likelihood in (r, Δ): a ridge along $r - \Delta = \text{const}$. (b) Bias of the amplitude-only estimate of r under small errors of the assumed element pattern, gain profile and reflectivity.

Table 16: **Closest radar scene representations.** *Angular response*: how a primitive's return varies with the look direction. *Geometry*: whether it is recovered from the radar data or supplied by another sensor.

Method	Primitive	Phase	Angular response	Geometry	Inference
SAR-GS (Li et al., 2025)	3D Gaussians	intensity	learned	recovered	per scene
mmNorm (Dodds et al., 2025)	normal votes	coherent	none (voting)	recovered	direct
GeRaF (Lu et al., 2025)	neural SDF	coherent	learned reflectiveness	recovered	per object
mmIR (Armouti et al., 2026a)	LiDAR geometry	coherent	material, multi-bounce	supplied	per scene
3DPS (Armouti et al., 2026b)	oriented points	coherent	learned material	LiDAR init.	per scene
Ours	curved Gaussians	coherent	derived (size, curvature)	recovered	amortized

L Additional comparisons and angular response

L.1 Additional prior work

Specular geometry with co-located sources. When source and sensor coincide, a mirror-like surface is seen only where its normal points at the sensor. Optical methods exploit this under flash illumination (Bi et al., 2020; Zhang et al., 2022; Nam et al., 2018), near lights (Quéau et al., 2018), highlights (Healey & Binford, 1988) and specular flow (Roth & Black, 2006; Adato et al., 2007); reflection-aware radiance fields and Gaussians index appearance by the reflected direction (Verbin et al., 2022; Jiang et al., 2024; Ye et al., 2024). Confocal non-line-of-sight imaging turns Fermat path-length derivatives into points and normals (Xin et al., 2019; O'Toole et al., 2018; Lindell et al., 2019), as do zero-offset seismic migration (Loewenthal et al., 1976; Stolt, 1978; Tygel et al., 2012) and microlocal analysis of radar (Quinto, 1993; Nolan & Cheney, 2004). For mmWave, mmNorm (Dodds et al., 2025) votes normals by the antenna with the strongest return, and RF-Capture (Adib et al., 2015) stitches specular body parts over time. We use this geometry to split supervision and to state what a glint's echo identifies.

Sub-aperture coherence. Coherence factors (Mallart & Fink, 1994; Hollman et al., 1999; Li & Li, 2003; Camacho et al., 2009), dual apodization (Seo & Yen, 2008) and short-lag spatial coherence (Lediju et al., 2011) in ultrasound, and spatially variant apodization (Stankwitz et al., 1995) in SAR, suppress pixels on which apertures disagree. Wide-angle SAR analyzes sub-aperture images to model anisotropic, limited-persistence scattering (Moses et al., 2004; Varshney et al., 2008; Ertin et al., 2007; Ferro-Famil et al., 2003). These methods act on image pixels; we use the cue in the measurement domain to supervise explicit geometry.

L.2 Detailed experimental comparisons

Table 17: **Model selection in a controlled 2D study** (ten seeds; response-weighted phantom rate PR_E and recall at $\tau = 2\,\mathrm{mm}$; mean over seeds, confidence intervals in Appendix J). All rules stop the same greedy path of curved surface Gaussians. HO: strip hold-out; CV-i, CV-r: interleaved and random five-fold cross-validation; DP: discrepancy principle with the noise level implied by the error model; ES: fixed early stop at the median HO step of the other seeds. Last two rows: flat Gaussians and points under DP.

	Pipe + plate		+ third-order path		Dihedral, 75°	
Rule	$\mathrm{PR}_E \downarrow$	Recall ↑	$\mathrm{PR}_E \downarrow$	Recall ↑	$\mathrm{PR}_E \downarrow$	Recall ↑
HO (strips)	0.007	0.58	0.026	0.29	0.614	0.41
CV-i	0.025	1.00	0.063	1.00	0.629	0.52
CV-r	0.025	1.00	0.064	1.00	0.622	0.50
DP	0.020	1.00	0.062	1.00	0.609	0.29
BIC	0.025	1.00	0.064	1.00	0.621	0.49
ES	0.004	0.52	0.003	0.26	0.617	0.42
Flat Gaussians, DP	0.123	0.93	0.170	0.94	–	–
Points, DP	0.536	0.05	0.681	0.06	–	–

L.3 Derived angular lobe

Corollary 2 (Derived angular lobe)**.** *At the center expansion,* $|\psi_j(\boldsymbol{p}, k)| = |\rho_j a_0|\, 2\pi |\det \mathbf{M}|^{-1/2} \exp\!\big(-\tfrac{1}{2}\boldsymbol{q}^\top \boldsymbol{\Gamma}_j(k)\, \boldsymbol{q}\big)$ *with*

$$\boldsymbol{\Gamma}_j(k) = 4k^2\, \mathrm{Re}(\mathbf{M}^{-1}) = 4k^2\, \boldsymbol{\Lambda}_j^{1/2} \big(\mathbf{I} + \mathbf{Z}_j^2\big)^{-1} \boldsymbol{\Lambda}_j^{1/2}, \qquad \mathbf{Z}_j = 2k\, \boldsymbol{\Lambda}_j^{1/2} \mathbf{H} \boldsymbol{\Lambda}_j^{1/2}. \tag{14}$$

Equivalently, $\boldsymbol{\Gamma}_j^{-1} = (2k)^{-2} \boldsymbol{\Lambda}_j^{-1} + \mathbf{H}\boldsymbol{\Lambda}_j\mathbf{H}$*, the decomposition used in Lemma 1. For* $\mathbf{H} \to 0$*, the lobe has standard deviation* $1/(2ks)$ *in* $\sin\theta$ *for an isotropic patch of radius* s*; in the geometric limit the patch reflects toward all directions whose specular point* $-\mathbf{H}^{-1}\boldsymbol{q}$ *lies inside its support.*

M Extended exposition and experimental interpretation

The main text retains the argument needed to follow the results. This appendix preserves the fuller reasoning and comparisons of the original draft, organized by the same sequence of questions.

M.1 Motivation and design rationale

The reason lies in how a synthetic aperture sees. A camera is an array of pixels, each looking along one ray. A synthetic aperture is closer to a wall of lamps, each with a detector beside it, and at millimeter wavelengths most objects act as mirrors. Each position therefore receives only *glints*, from the surface points whose normals point back at it, and each strip of the aperture sees a different band of surface orientations. Back-projection fuses these glints into streaks that meet near the center of curvature rather than on the surface, so a convex surface is imaged too deep, by an amount that depends on the ratio of angular diversity to bandwidth.

We start from the optics of the aperture and derive each part of the method from it. First, we analyze the synthetic aperture as a co-located specular sensor: which surface elements each strip observes, where BP places a curved glint, and what the phase of the echo can and cannot determine (Section 3). Second, since the analysis shows that the directional response of a surface element follows from its size and curvature, we represent the scene by curved, complex-valued *coherent surface Gaussians* whose near-field echo has a closed form under a quadratic approximation of the range, so that the renderer, not a learned parameter, decides which strips see each element (Section 4). Third, we ask what this optics implies for learning without geometric labels (Section 5). Since each strip observes a band of normals, holding out one strip and predicting it from the others looks like the radar analogue of novel-view synthesis, and it was our starting hypothesis. The analysis predicts, and a controlled study confirms, that it is a weak signal: the held-out strip sees normals that the other strips do not, so even the true surface predicts it poorly, and as a model-selection rule it removes phantoms only by stopping early. Generic cross-validation over random positions is the reliable

criterion, and the strips serve instead to tell which parts of a reconstruction rest on one band of look directions. Throughout, geometry is judged by how well it predicts measurements; back-projected images serve only as a linear feature transform at the input of the encoder. Held-out prediction of any kind also has a hard limit, which we characterize with a classical tool: by Hagedoorn's envelope construction, any smooth multipath ghost is the glint of a virtual surface, so a ghost that every strip observes cannot be rejected by holding out measurements.

M.2 Broader related work

Modality-aware Gaussians and radar scenes. 3D Gaussian splatting (Kerbl et al., 2023) owes its speed to closed-form EWA footprints (Zwicker et al., 2001); surface variants use disks (Huang et al., 2024) or quadrics (Zhang et al., 2025). When the modality changes, the renderer changes with it, as for X-ray tomography (Zha et al., 2024) and holography (Choi et al., 2025; Chao et al., 2025). SAR-NeRF (Lei et al., 2024) and SAR-GS (Li et al., 2025; Liu et al., 2026) fit SAR intensity images, Sugavanam & Ertin (2025) fit Gaussian limited-persistence scatterers, and automotive methods fit processed range–azimuth or Doppler data (Borts et al., 2024; Kung et al., 2025; Bagwe et al., 2026). For mmWave, mmIR (Armouti et al., 2026a) synthesizes phase-coherent samples from LiDAR geometry with multiple bounces, and 3D point splatting (Armouti et al., 2026b) optimizes oriented points with learned material parameters, LiDAR init., whose complex phasors are splatted through a precomputed PSF for novel view synthesis; Wave-Former (Dodds et al., 2026) completes occluded shapes, and GeRaF (Lu et al., 2025; 2026), closest to our setting, recovers millimeter-level geometry per object with a neural signed distance field and a phase-preserving near-field renderer that models specular reflection and learned reflectiveness. We share its premise; we differ in explicit curved primitives with a closed-form coherent echo whose directional response is derived, amortized inference, held-out strips as supervision, and an analysis of what a glint's phase identifies. Table 16 summarizes the differences.

Self-supervised and amortized inversion. Noise2Noise (Lehtinen et al., 2018), Noise2Self (Batson & Royer, 2019) and SSDU (Yaman et al., 2020) learn by predicting held-out data; our split follows the aperture and hence the surface normals. Feed-forward splatting (Charatan et al., 2024; Chen et al., 2024) and learned inverse solvers (Ongie et al., 2020; Monga et al., 2021) amortize inference; our encoder does so for coherent measurements with set aggregation (Zaheer et al., 2017; Lee et al., 2019).

Multipath as virtual scatterers. In reflection seismology, Hagedoorn's envelope construction (Hagedoorn, 1954) underlies zero-offset map migration (Chun & Jacewitz, 1981; Yilmaz, 2001), and multiples are known to be indistinguishable from primaries of a virtual reflector unless the data contain other offsets or a model of the free surface (Verschuur et al., 1992; Weglein et al., 2003). Radar treats multipath in the same spirit: ghosts in through-wall and urban radar are the images of virtual targets obtained by mirroring the sensor in the reflecting walls, and they are suppressed by modeling those walls or by exploiting diversity across views (Setlur et al., 2011; Leigsnering et al., 2014; Villard & Borderies, 2015). We use this literature to bound what held-out prediction can reject: a ghost that every strip sees is a virtual glint, and only diversity that a co-located sensor does not have can remove it.

M.3 Specular geometry and aperture partition

Gaussian splatting inherits its structure from the camera, in which every pixel integrates the scene along one ray. A synthetic aperture has no pixels. Each position is a source and a detector at the same place, and at millimeter wavelengths it receives mirror-like *glints* rather than diffuse reflections. Before designing a representation for such a sensor, we answer four questions about it: what it measures (Section 3), which surface elements each part of the aperture observes (Section 3), where back-projection places a curved surface (Section 3), and what the echo can determine at all (Section 3). Each answer is first stated in words, then formally, and then related to the method.

At each aperture position the radar transmits a sweep of frequencies and records the echo. Under single scattering the echo is a superposition over the scene: each point contributes its complex reflectivity, delayed by the round-trip path from transmitter to point to receiver. A real radar adds one more delay that has nothing to do with the scene, a constant offset Δ from its cables and analog

chain, which is unknown unless it is calibrated. Formally, with wavenumbers $k = 2\pi f/c$ over the band $f \in [f_0, f_0 + B]$, a transmitter at $\boldsymbol{p}_t$ and a receiver at $\boldsymbol{p}_r$ measure

$$y(\boldsymbol{p}_t, \boldsymbol{p}_r, k) \;=\; e^{-\mathrm{i}2k\Delta} \int \sigma(\boldsymbol{x})\, a(\boldsymbol{x}, \boldsymbol{p}_t, \boldsymbol{p}_r, k)\, e^{-\mathrm{i}k(\|\boldsymbol{x}-\boldsymbol{p}_t\|+\|\boldsymbol{x}-\boldsymbol{p}_r\|)}\, d\boldsymbol{x} \;+\; n, \tag{15}$$

where σ is the reflectivity density, a collects slowly varying amplitude terms (antenna pattern, spreading and obliquity), and n is noise. For metals, equation 1 is the physical-optics approximation (Knott et al., 2004).

It helps to ask which direction plays the role of the camera ray. Near a scene point $\boldsymbol{x}$, the phase in equation 1 varies as $e^{-\mathrm{i}\boldsymbol{\kappa}^\top \boldsymbol{x}}$, with

$$\boldsymbol{\kappa} \;=\; k\,(\boldsymbol{d}_t + \boldsymbol{d}_r) \;=\; 2k\cos(\beta/2)\,\boldsymbol{h}, \qquad \boldsymbol{h} = (\boldsymbol{d}_t + \boldsymbol{d}_r)/\|\boldsymbol{d}_t + \boldsymbol{d}_r\|, \tag{16}$$

where $\boldsymbol{d}_t$ and $\boldsymbol{d}_r$ are the unit vectors from transmitter and receiver to $\boldsymbol{x}$ and β is the angle between them. The measurement thus probes the scene along the bisector $\boldsymbol{h}$, and a mirror returns energy only if it faces along $\boldsymbol{h}$ (Kell, 1965). For co-located transmitter and receiver at a position $\boldsymbol{p}$, $\boldsymbol{\kappa} = 2k\,\boldsymbol{d}$: the ray is the line of sight, traversed at twice the wavenumber, as in the exploding-reflector picture of seismics (Loewenthal et al., 1976; Claerbout, 1985). Our radar uses one transmit–receive pair a few millimeters apart; we render it bistatically, but the monostatic picture is accurate to a small fraction of a wavelength (Appendix A), and we use it throughout the analysis.

We split the aperture $\mathcal{A}$ into S disjoint strips $\mathcal{A}_\ell$, $\ell = 1, \ldots, S$, with measurements y_ℓ and back-projection (BP) images $I_\ell = \mathcal{B}_\ell[y_\ell]$. Since BP is linear, the strip images sum to the full image (Lemma 2). The split therefore adds no information; what it exposes is *which* measurements support which structure, and the rest of this section explains why that differs from strip to strip.

A smooth surface that is large compared with the wavelength reflects toward a position $\boldsymbol{p}$ only from its glints, the points $\boldsymbol{x}^\star$ whose normal points at $\boldsymbol{p}$. The distance from $\boldsymbol{p}$ to the nearest glint, viewed as a function of $\boldsymbol{p}$, is the *distance function* of the surface, and the aperture samples it. Three classical relations, restated in Proposition 4 (Appendix A), make this precise: stationary phase gives a glint's echo amplitude from its curvature (Bleistein & Handelsman, 1986; Ruck et al., 1970); the gradient of the distance function gives the glint's normal, and stepping back along it by the distance gives the glint itself; and its second derivatives give the principal curvatures. The last two are the Fermat-flow equations of confocal non-line-of-sight imaging (Xin et al., 2019) and the map-migration relations of seismics (Tygel et al., 2012). For our purposes the consequence is qualitative but central: a position observes a surface through its normals, so different parts of the aperture observe different parts of a curved surface.

In the geometric limit, the specular condition assigns every surface point to the aperture position at which its normal ray arrives. A strip therefore observes the surface points whose normals point into it, and points whose normal ray misses the aperture are dark. This cone condition is the visibility principle of microlocal analysis (Quinto, 1993; Nolan & Cheney, 2004), the basis of mmNorm's normal voting (Dodds et al., 2025) and the specular condition of a co-located flash (Zhang et al., 2022); wide-angle SAR studies the resulting limited persistence of scatterers by sub-aperture (Ertin et al., 2007; Ferro-Famil et al., 2003).

At finite frequency the assignment blurs, and for learning the question is by how much: if an element's echo reaches several strips, the other strips can check it; if it reaches only one, they cannot. The echo of a small surface element spreads over the aperture for two reasons. A small element diffracts, and the smaller it is, the wider it radiates; a curved element acts as a convex mirror and spreads its reflection over a range of directions proportional to its size times its curvature. The following lemma adds the two contributions. It uses the tangent covariance $\boldsymbol{\Lambda}$ of the element and the Hessian $\mathbf{H}$ of the range across it, which Section 4 defines and which combines the curvature of the wavefront with that of the surface.

Lemma 1 is Gaussian-beam optics (Kogelnik & Li, 1966; Červený et al., 1982; Hill, 1990), a beam of waist $\boldsymbol{\Lambda}^{1/2}$ at wavenumber $2k$ reflected by a mirror of curvature $\mathbf{K}$, and it is the footprint on the aperture of the angular lobe derived in Corollary 2. It matches numerical quadrature to 1% for flat elements of 5–10 mm and errs by 5–13% for smaller or curved ones (Appendix B). The numbers show how different the two regimes are: for a 2 cm pipe 0.3 m from a 0.6 m aperture in five strips, the strips see bands of $13.8°$–$21.2°$ of normals and $94°$ of the front half is dark (Figure 4a), while

a 1 cm element has a footprint of 1.3 cm if flat but 14 cm on the pipe. We call the fraction of an element's response energy that falls in strip $\mathcal{A}_\ell$ its *lobe share* $w_{j\ell}$. It decides which strips can check the element (Section 5), and it guides the choice of the number of strips (Section 6).

The ratio ζ is the one number to remember: it compares the depth information in the angular diversity of the aperture with that in the bandwidth, and it sends the peak to the surface for $\zeta \ll 1$ and to the center of curvature for $\zeta \gg 1$. At 77–81 GHz, $s_k/\mu \approx 0.015$. Our rail, a 0.99 m aperture at about 0.95 m, gives $\zeta \approx 4.9$ (5.1 with 64 discrete frequencies), so for $\Delta = 0$ back-projection focuses a 2 cm pipe 16.7 mm behind its front surface; exact 2D back-projection gives 16.4 mm. Over radii of 1–3 cm, apertures of 0.1–1 m and standoffs of 0.3–1 m, the prediction is within 1.5 mm of exact back-projection, and it is within 0.5 mm of a physical-optics simulation (Figure 2c), as long as the range resolution $c/2B$ exceeds about $1.3\,r$; below that, the profile splits into a surface peak and an axis peak (Figure 2d). For methods that start from BP images, the proposition quantifies the depth bias they inherit, up to one radius; in addition, a Gaussian fitted to a BP image fits the PSF, with covariance $\mathbf{\Sigma} + \mathbf{P}_\ell$ (Proposition 3).

The constant offset Δ in equation 1 adds the same delay to every echo. For the convex glint above, the measured range becomes the distance to $\boldsymbol{c}$ minus $(r - \Delta)$, so the echo looks exactly like that of a glint of radius $\varrho = r - \Delta$ measured without offset. If Δ exceeds r, ϱ is negative, and the same echo is produced by a concave mirror that faces the aperture from behind $\boldsymbol{c}$; if $\Delta = r$, it is produced by a point at $\boldsymbol{c}$. More generally, delaying every echo by Δ is the same as moving every surface element a distance Δ away from the sensor along its normal, which replaces the surface by a parallel surface. The following theorem states this and what it leaves identifiable.

Theorem 1 is kinematic and classical in form: the distance to a parallel surface differs by a constant (Federer, 1959), and the parallel surface develops cusps where $|\Delta|$ exceeds a local radius of curvature (Farouki & Neff, 1990), which is where the equivalence ends. In ground-penetrating radar, the same ambiguity appears as the time-zero problem (Yelf, 2004; Zhu & Ye, 2023) and as the radius–velocity nonuniqueness of hyperbola fitting (Giannakis et al., 2022). We have not found it stated for SAR. The theorem concerns the phase. The amplitude of a glint does depend on its radius, through the factor in (ii), and with calibrated gain and a known material it has been used to estimate radii (Mechbal & Khamlichi, 2017). Over a rail aperture, however, this factor changes by a fraction of a decibel, the same order as the uncertainty of the element pattern, a gain drift along the rail or a mild dependence of reflectivity on incidence angle. Appendix K quantifies this for pipes of 1–3 cm at 0.3–0.9 m: even with a perfectly known model at 25 dB, the amplitude alone determines r only to 0.4 cm at 0.3 m and 5 cm at 0.9 m, and mis-specifying the exponent of a $\cos^q$ element pattern by only 0.1 biases the estimate by 5–23 cm, and an unmodeled 0.1 dB roll-off of the gain toward the ends of the rail biases it by 1–17 cm. In practice, therefore, the phase ambiguity is exact and the amplitude is ill-conditioned, and a renderer with a free reflectivity per element absorbs the amplitude cue entirely. The delay requires an independent range reference.

M.4 Representation and renderer choices

Section 3 gives four requirements for a representation that is fitted to sub-aperture echoes. First, it must be *coherent*: the depth of a curved glint (Proposition 1) and the range-offset ambiguity (Theorem 1) live in the phase, so a model that predicts only intensities cannot express them. Second, it must carry *curvature*: the strips that see a surface element are set jointly by its size and its curvature (Lemma 1), and a flat element of any size predicts the wrong set. Third, its *angular response must be derived, not learned*. If each primitive had a free directional lobe, as reflection-aware optical Gaussians do (Fan et al., 2026), a primitive in the wrong place could reproduce the strips that see it by adjusting its lobe, and the partition of Section 3 would carry no information about geometry. Fourth, the echo must be *differentiable and fast* enough to evaluate at every aperture position and frequency inside a training loop. This section builds a primitive that meets all four: a curved Gaussian patch whose echo has a closed form under a quadratic approximation of the range, with an approximation error that we control and measure.

A coherent surface Gaussian is a small disc of surface, as in 2D Gaussian splatting (Huang et al., 2024), that is allowed to bend. Its weight falls off as a Gaussian from its center, it bends by a quadratic height along its normal, and it reflects with one complex coefficient. Formally, G_j has center $\boldsymbol{\mu}_j \in \mathbb{R}^3$, an orthonormal frame $[\mathbf{U}_j, \boldsymbol{n}_j]$ with tangent basis $\mathbf{U}_j \in \mathbb{R}^{3\times 2}$ and normal $\boldsymbol{n}_j$

pointing toward the sensor, tangent covariance $\mathbf{\Lambda}_j = \mathrm{diag}(s_{j1}^2, s_{j2}^2)$, a symmetric curvature matrix $\mathbf{K}_j \in \mathbb{R}^{2\times 2}$ in the tangent frame, and complex reflectivity $\rho_j \in \mathbb{C}$. With tangent coordinates $\boldsymbol{u} \in \mathbb{R}^2$, the patch and its weight are

$$\boldsymbol{x}_j(\boldsymbol{u}) = \boldsymbol{\mu}_j + \mathbf{U}_j\boldsymbol{u} + \tfrac{1}{2}\,\boldsymbol{n}_j\,\boldsymbol{u}^\top\mathbf{K}_j\boldsymbol{u}, \qquad g_j(\boldsymbol{u}) = \exp\big(-\tfrac{1}{2}\boldsymbol{u}^\top\mathbf{\Lambda}_j^{-1}\boldsymbol{u}\big), \tag{17}$$

following second-order surface primitives (Zhang et al., 2025). The sign convention matches Section 3: the front of a pipe bends away from the sensor, $\mathbf{K}_j \prec 0$, and has positive radius $\varrho > 0$ in the sense of that section. The scene reflectivity is the superposition $\sigma(\boldsymbol{x}) = \sum_j \rho_j \int g_j(\boldsymbol{u})\,\delta\big(\boldsymbol{x} - \boldsymbol{x}_j(\boldsymbol{u})\big)\,d\boldsymbol{u}$, which enters the forward model equation 1 directly. Nothing else is attached to the primitive: its directional response is a consequence of $\mathbf{\Lambda}_j$ and $\mathbf{K}_j$, as the next two subsections show.

Concretely, fix an aperture position $\boldsymbol{p}$ and an expansion point $\boldsymbol{u}_0$ on the patch, and let R_0, $\boldsymbol{q}$ and $\mathbf{H}$ be the value, gradient and Hessian of the exact range $R(\boldsymbol{u}) = \|\boldsymbol{x}_j(\boldsymbol{u}) - \boldsymbol{p}\|$ at $\boldsymbol{u}_0$. At the patch center, with unit look direction $\boldsymbol{d} = (\boldsymbol{\mu}_j - \boldsymbol{p})/R_0$,

$$\boldsymbol{q} = \mathbf{U}_j^\top\boldsymbol{d}, \qquad \mathbf{H} = \underbrace{\frac{\mathbf{U}_j^\top(\mathbf{I} - \boldsymbol{d}\boldsymbol{d}^\top)\mathbf{U}_j}{R_0}}_{\text{wavefront curvature}} + \underbrace{(\boldsymbol{n}_j^\top\boldsymbol{d})\,\mathbf{K}_j}_{\text{surface curvature}}, \tag{18}$$

and Appendix A gives $\boldsymbol{q}, \mathbf{H}$ at any $\boldsymbol{u}_0$. The amplitude varies slowly and is kept to first order. Under physical optics it is $a(\boldsymbol{u}) \propto k\,\mathrm{E}(\boldsymbol{p},\boldsymbol{d})^2\,(-\boldsymbol{n}^\top\boldsymbol{d})/R^2$, with antenna pattern E and obliquity $-\boldsymbol{n}^\top\boldsymbol{d}$ (Knott et al., 2004); let a_0 and $\boldsymbol{g}$ be its value and log-gradient at $\boldsymbol{u}_0$.

A surface Gaussian interpolates between the two classical radar primitives. As $\mathbf{\Lambda}_j \to 0$ it becomes a point scatterer (Armouti et al., 2026b; Potter & Moses, 1997), and as $\mathbf{\Lambda}_j$ grows on a curved surface it becomes a glint (Proposition 4); for a transmitter–receiver pair, $2k$ times range, gradient and Hessian become sums of one-way terms (Corollary 1).

The two terms of $\mathbf{\Gamma}_j^{-1}$ are the two mechanisms of Lemma 1: a small patch spreads its echo by diffraction, and a curved patch spreads it geometrically. At 79 GHz, a flat 5 mm patch has a $3.6°$ lobe, while the same patch on a 2 cm pipe at 0.3 m has a $15.9°$ lobe. The lobe is therefore not a free parameter that can be traded against position: to be seen by more strips, a primitive must be smaller or more curved, and either change also alters its predicted phase and amplitude in every strip.

The predicted measurement is the coherent sum of all primitives, delayed by the range offset:

$$\hat{y}(\boldsymbol{p}, k) = e^{-\mathrm{i}2k\Delta}\sum_j v_j(\boldsymbol{p})\,\psi_j(\boldsymbol{p}, k), \tag{19}$$

where the visibility $v_j(\boldsymbol{p}) \in [0,1]$ removes back-facing and occluded patches (Appendix C). The offset Δ collects all delays that the system adds to the propagation path, such as cables, the antenna phase center and the processing reference. By Theorem 1, a free Δ would trade against the curvature of every glint, so the renderer treats Δ as a fixed setup-level parameter, determined by a reference at known range. Two further components follow optical splatting. *Lobe culling*, the analogue of frustum culling, skips a pair $(j, \boldsymbol{p})$ when $\frac{1}{2}\boldsymbol{q}^\top\mathbf{\Gamma}_j(k_{\min})\boldsymbol{q}$ exceeds a threshold, because the patch is then far outside its lobe. Adaptive density control follows 3DGS (Kerbl et al., 2023), with splits driven by the remainder bound ε_j instead of screen-space gradients. Table 8 (Appendix C) maps each component of optical splatting to its coherent counterpart.

M.5 Interpretation of held-out learning

The renderer of Section 4 maps any set of surface Gaussians $\mathcal{G}$ to the echo of any set of aperture positions. This section asks how to learn geometry with it when measured scans come without geometric labels. It first explains why fitting all measurements is not enough (Section 5) and introduces held-out prediction in the measurement domain (Section 5). It then uses the analysis of Section 3 to predict what held-out prediction can check, and how the answer depends on whether the held-out positions form a strip or are spread over the aperture (Section 5), and what no held-out split can check (Section 5). Section 5 describes the encoder and the test-time refinement. Figure 3 gives an overview.

The most direct use of the renderer is *self-reconstruction*: fit $\mathcal{G}$ so that its predicted echo matches all measured positions. Because the renderer contains the point-spread function, this already avoids

the most visible failure of image-based methods, which is to read sidelobes and the hourglass of Figure 1 as surface. What it cannot do is separate geometry from anything else that a flexible $\mathcal{G}$ can also reproduce. Measurement noise, residual phase errors of the rail and multipath echoes are all fitted by additional primitives as readily as real surface, and the loss decreases with every primitive added. The number of primitives and the stopping point must therefore be chosen by a criterion that the fitted data do not satisfy by construction. On simulated scenes, geometric labels provide such a criterion; on a measured scan there are none.

The classical label-free criterion is to predict measurements that the fit did not use. Let $\mathcal{A} = \bigsqcup_f \mathcal{A}_f$ be a partition of the aperture positions into folds, and write y_{-f} for the measurements outside fold f. A model fitted to y_{-f}, or an encoder f_θ that receives y_{-f}, is scored on the fold it did not see:

$$\mathcal{L}_{\mathrm{HO}}(\theta) \;=\; \mathbb{E}_{\text{scene}}\, \mathbb{E}_f \left\| y_f - \mathcal{F}_f\big(f_\theta(y_{-f})\big) \right\|^2. \tag{20}$$

If the noise of y_f has zero mean and is independent of the scene and of the other folds, $\mathcal{L}_{\mathrm{HO}}$ equals the error on noise-free held-out data plus a constant (Proposition 5), as in Noise2Self (Batson & Royer, 2019) and SSDU (Yaman et al., 2020), so minimizing it does not reward fitting noise. The statement is about measurements, not geometry: equation 9 is an unbiased objective for predicting the noise-free held-out measurements, and whether a good prediction implies a correct surface depends on the partition, on the prior of the encoder and on the adequacy of the renderer. The independence assumption also deserves care on real hardware. Group delay, frequency ripple, rail trajectory errors and channel phase are often correlated across the aperture; the part common to all folds acts like the offset Δ and is not charged (Section 5), and the part that varies smoothly along the rail is partly predictable from neighboring positions and is then fitted as geometry.

For a specular sensor, the choice of partition is not a detail. The folds can follow the strips $\mathcal{A}_\ell$ of Section 3, so that each held-out fold is a band of surface normals that the model has not seen; we call this *strip hold-out*. Or they can be spread over the aperture, as random subsets or interleaved blocks of positions, so that every fold sees every band of normals at a coarser sampling; this is generic cross-validation. Strip hold-out is the natural analogue of novel-view synthesis, and it was our starting hypothesis that it would be the better training signal and model-selection criterion, because it asks the model to infer unseen surface orientations. The analysis below predicts the opposite, and Section 6 confirms it.

Two kinds of error are consistent across all folds of any partition and therefore invisible to any held-out criterion. The first is the range offset Δ of Theorem 1, which delays every measurement equally and is absorbed by the curvature of every glint. The second is less obvious. A multipath echo, such as a double bounce between two pipes, has an apparent position that moves as the strip changes, and one might expect that no single surface could reproduce this motion in every strip. That expectation is wrong for a co-located sensor: along any smooth branch, the phase of a multipath echo equals that of the glint of a suitably placed virtual surface.

The construction is classical. It is Hagedoorn's envelope construction (Hagedoorn, 1954), $\boldsymbol{x}_g$ is the zero-offset map-migration formula (Chun & Jacewitz, 1981; Yilmaz, 2001), and it is the reason seismic multiples masquerade as primaries (Weglein et al., 2003); in radar it generalizes the image-theory virtual targets of multipath ghosts (Setlur et al., 2011; Leigsnering et al., 2014; Villard & Borderies, 2015). It is not a contribution of this paper; what we add is its use to delimit what held-out prediction can reject. A single-scattering renderer can place a glint on $\boldsymbol{x}_g$ that reproduces the phase of the ghost in every strip that sees it (Appendix A). Held-out prediction then charges the ghost only where its support across strips differs from the footprint of its virtual surface, as when an occluder or a finite reflector truncates it, or where the virtual surface lies outside the primitive family or outside the prior; Section 6 tests both cases, and also a case in which neither applies. Two limits of the statement matter for later work. It needs the regularity stated in the proposition, which fails where the ghost's range history has a caustic. And it is monostatic: with a separated transmitter and receiver, the two traversal orders of a double bounce are distinct events with different delays, so a multistatic aperture, such as the 3TX4RX array of our radar used in full, can break the equivalence.

The encoder turns the input strips into primitives. Each strip is back-projected onto a voxel grid and demodulated by its carrier (Appendix D). Back-projection is used here only as a fixed linear transform that puts the measurements on a grid where convolutions apply; on a grid fine enough to represent the scene it loses little, since the adjoint of the forward operator is injective on that operator's range, and the encoder is never judged by its images. A shared network encodes each

volume together with its look direction, set attention (Lee et al., 2019) aggregates across strips so that the number and order of strips do not matter, and a voxel-aligned head (Charatan et al., 2024; Chen et al., 2024) predicts primitives with existence logits. On physical-optics scenes, which are simulated by a different model than equation 7 to avoid an inverse crime (Kaipio & Somersalo, 2007), the encoder is trained with a geometric loss $\mathcal{L}_{\mathrm{geo}}$ (Chamfer and normal consistency against the true surface) and a sparsity penalty $\mathcal{L}_{\mathrm{sp}}$ on existence. We also evaluate adding the held-out loss equation 9, with strip or random folds, and using it alone to fine-tune on a shifted domain without labels (Section 6).

At test time, the encoder output initializes a short optimization of the self-reconstruction loss $\sum_\ell \|y_\ell - \mathcal{F}_\ell(\mathcal{G})\|^2$, and generic cross-validation over random folds selects the stopping point and the number of primitives. Finally, the fitted model explains each strip's image. Splitting $I_\ell = \mathcal{B}_\ell[\mathcal{F}_\ell(\mathcal{G})] + \mathcal{B}_\ell[y_\ell - \mathcal{F}_\ell(\mathcal{G})]$ attributes image energy to *geometry* (first term, inside the support of some G_j), to the *PSF* (first term, outside every support) or to *unexplained* effects (second term).

M.6 Experimental interpretation

The experiments test the consequences of the analysis in the order in which the paper derived them. **Q1** (Section 6) asks whether a measured scan behaves as the specular analysis of Section 3 predicts. **Q2** (Section 6) asks whether the closed-form renderer of Section 4 is accurate and fast enough to stand in for the physics, and whether curvature matters. **Q3** (Section 6) asks, in a controlled simulation where the true surface is known, how strip hold-out compares with generic cross-validation and calibrated stopping rules, as Section 5 predicts. **Q4** (Section 6) evaluates the amortized encoder, and **Q5** (Section 6) evaluates the methods on measured objects. Negative results are reported alongside positive ones.

Hardware and measured data. A TI AWR2243 radar, used as one transmit–receive pair for phase stability, sweeps 77–81 GHz over 1000×1040 rail positions at a $\lambda/4$ step (0.95×0.99 m), about 0.9 m from the object. The aperture is split along y into $S = 5$ disjoint strips of about 200 lines each. Five strips keep most of a curved surface in the lobes of at least two strips (Lemma 1, Appendix B), so that held-out prediction has something to predict, while each strip remains long enough to focus. We use scans of a pipe-like object and a trihedral corner reflector, taken without moving the radar in between; for both we have the complex back-projected volumes of each strip, not the raw echoes (Appendix H).

Simulated data. Controlled experiments use a two-dimensional scalar physical-optics simulator with occlusion-tested double bounces (Appendix J) and a three-dimensional physical-optics simulator of procedural and ShapeNet (Chang et al., 2015) objects (Appendix I). Neither is the renderer of equation 7, which avoids an inverse crime (Kaipio & Somersalo, 2007).

Metrics. A fitted primitive counts as a *phantom* if it lies farther than τ from every true surface; the phantom rate weights each primitive by its response energy, so that a faint misplaced primitive counts less than a bright one. *Recall* is the fraction of the visible true surface within τ of a significant primitive. The encoder study and the three-dimensional evaluation add Chamfer distance on the visible surface, normal consistency and F-score. Visible-surface metrics separate what the aperture cannot see from what the method fails to recover.

Section 3 makes three predictions that can be checked on a scan without knowing the object: each strip's streak points along the strip's look direction, the look lines of all strips meet at one point, and the strip peaks lie beyond that point by an amount set by ϱ and by the angular diversity of each strip. On the pipe scan (Figure 1c, Appendix H), the first two hold closely. Each strip's streak follows its mean look direction, with correlation 0.9993 and a root-mean-square difference of 0.9° over a 46° span (Lemma 3), and the five look lines cross within 1.7 mm at the waist of the hourglass.

The third prediction also holds once two properties of the measurement are modeled. First, the strips are not narrow. Each spans the full 0.95 m line of the rail in x, so its own diversity ratio is $\zeta_\ell \approx 3.9$–5.5, against 9.9 for the full aperture when both dimensions are counted. With these values, Proposition 1 predicts that the full-aperture peak lies 0.51–0.60 times as far beyond the crossing as the strip peaks, for any element pattern; the scan gives 0.61. It also predicts the U-shape of the strip offsets over look angle: relative to the central strip, $1.20, 1.02, 1, 1.09, 1.33$ predicted against

$1.47, 1.13, 1, 1.03, 1.28$ measured. The apparent discrepancy between strip and full-aperture offsets noted in earlier analyses came from treating the strips as narrow. Second, the processing library used for this scan references the range phase to the start of the sweep but corrects the aperture phase with the center wavelength. We infer from its source code that this multiplies each image by a range-dependent phase, which in simulation moves the crossing of the strip look lines about 2.7 cm beyond the scatterer, whatever Δ. On this scan, therefore, the crossing is not exactly the center of curvature of Theorem 1. Simulating the exact aperture with a replica of the library, the best of 176 single-glint models has $\varrho \approx -5.5$ cm in the y–z plane and reproduces the strip offsets to within 4 mm, the strip levels to within 1 dB, the peak ratio (0.49 against 0.50) and the full-aperture peak depth to within 3 mm (Appendix H); no model with $|\varrho| \leq 3$ cm fits.

As Theorem 1 says it must, this leaves the object ambiguous. A negative ϱ of about 5.5 cm is produced by a point-like scatterer, such as an end or joint of the object, seen through an offset $\Delta \approx 5.5$–6 cm; by a convex pipe of radius r seen through $\Delta \approx r + 5.5$ cm; by a concave reflector of radius about 5.5 cm with no offset; or by a virtual glint. A point with offset fits the data somewhat worse than a glint (χ^2 of 49–52 against 17.9), which slightly favors a curved target.

Table 1 separates three sources of error. The closed form itself is accurate: re-centered at the saddle point, it matches dense quadrature over the same Gaussians to 0.4% on the pipe and the sphere and 0.1% on the plate; expansion at the center adds about 1% on curved surfaces, and the single-patch analysis of Appendix B shows that this grows to 39% at 30° incidence. Lobe culling ($\tau_{\mathrm{lobe}} = 6.9$) skips 85–87% of the front-facing primitive–position pairs and makes rendering 4–6 times faster while adding less than 0.1% error. The remaining error comes from representing the surface by Gaussians: 9% on the pipe with 5 mm primitives, 5.6% with 3 mm primitives, and 16% on the plate. Most of it is the Gaussian profile smoothing hard edges, the ends of the pipe and the rim of the plate; against a reference whose pipe ends are smoothed the same way, the pipe's error drops from 7.2% to 2.2%. Modeling edges explicitly is therefore the main route to higher fidelity. Curvature is essential: flat Gaussians of the same size err by 50–70% on curved surfaces. Points placed at the same centers err by 190–360%; dense points are the most accurate renderer, 0.2–0.7% at $\lambda/4$ spacing, but need 28 times as many primitives and 3.5–10 times the rendering time, which for an inverse problem means 28 times as many unknowns; at $\lambda/2$ spacing their error is already 4–40%. Surface Gaussians thus trade a representation error for a much smaller and faster model.

The two-dimensional study of Section 6 confirms that curvature matters for reconstruction and not only for rendering. Under the same greedy fitting and stopping rule, curved surface Gaussians have a phantom rate 0.105 lower than flat ones in all 20 scene–seed pairs ($p < 10^{-5}$) and 6% higher recall; point scatterers with the same budget of 16 primitives reach a phantom rate of 0.54–0.68 and a recall of 0.05–0.06 (Appendix J).

Section 5 predicts that strip hold-out is a weak criterion wherever the strips see disjoint bands of normals. We test this in the two-dimensional physical-optics study of Appendix J, now with ten noise seeds per scene, and compare strip hold-out with interleaved and random cross-validation, the discrepancy principle calibrated to the error model, AIC and BIC, a fixed budget and a fixed early stop. All rules stop the same greedy path, so they differ only in where they stop.

The held-out strip error behaves as predicted. As the greedy fit adds primitives, the error on randomly held-out positions falls from 0.82 to 0.05 and that on interleaved blocks from 0.97 to 0.13, but the error on a held-out strip *rises*, from 1.10 to 1.59, where 1 corresponds to predicting nothing. Primitives that explain one strip do not help predict another, and those fitted to one strip add noise to the prediction of the others. Strip hold-out therefore stops early, after 1–12 steps (median 3) where the other rules stop after 11–16. Its models have fewer phantoms, pooled over scenes by 0.022 in phantom rate (Holm-corrected $p = 0.009$), but they lose much of the surface: recall drops by 0.41 ($p < 10^{-4}$), and on the plate to 0.61 and 0.33. A fixed early stop at the typical HO step does as well as strip hold-out in every metric (phantom rate $+0.008$, $p = 0.78$; recall $+0.03$, $p = 0.67$), so what strip hold-out contributes is a small model, not a choice adapted to the scene.

Nor does strip hold-out target the structure it was meant to target. Over 441 primitives fitted by self-reconstruction, the probability of being a phantom does not grow with the maximal lobe share (logistic slope -0.10 per 0.1), and strip hold-out removes real and phantom primitives alike, at every lobe share (Appendix J). In the plate scenes, 39–41% of the response energy of correctly placed primitives comes from primitives seen by essentially one strip (lobe share above 0.9), which

is the surface strip hold-out gives up. Generic cross-validation keeps them and behaves like the calibrated discrepancy principle.

Two limits hold for every rule. The third-order plate–pipe–plate path is fitted as a mirror image of the pipe by every rule except the degenerate strip hold-out, and in the dihedral, whose double bounce is up to 21 dB above the single bounce, every rule fails (phantom rate 0.61–0.63), as Proposition 2 predicts for a ghost that all strips see. The interreflection between two pipes behaves the same way (0.113 versus 0.118 in the three-seed study of Appendix J).

Amortized encoder. Table 3 evaluates the encoder of Section 5 in the two-dimensional setting, trained on 3000 procedural scenes of pipes, flat segments and plates and tested on unseen scenes of the same families and on ellipses, a family never seen in training. A diagnostic first confirms the analysis of Section 5 at the level of the loss: on true geometry the held-out strip error is 0.68, and it changes little when the geometry is wrong (0.81 for a 3 mm shift, 0.68 for ten added phantoms), whereas the error on random folds is 0.09 for true geometry and 0.43 for a 10 mm shift. The strip loss therefore provides almost no gradient toward the correct surface, and an encoder trained with it alone does not learn to reconstruct. With geometric supervision in simulation, the encoder reduces the phantom rate of back-projection by an order of magnitude on seen and unseen families in a few milliseconds, and a short refinement with a cross-validated stop reaches the accuracy of per-scene fitting at a small fraction of its cost. Label-free fine-tuning on a shifted domain, with a different error model, element pattern and range offset, is evaluated in Appendix G with a modest gain from random-fold hold-out.

M.7 Implications

The analysis of this paper rests on one observation: a near-field millimeter-wave synthetic aperture is a co-located specular sensor, not a camera. Three practical consequences follow for anyone who learns geometry from such data. First, the range offset of the system must be calibrated once, because no criterion computed from the echoes can separate it from the curvature of a glint (Theorem 1); a corner reflector scanned in the same setup is the cheapest reference. Second, reconstruction should happen in the measurement domain with primitives that carry curvature, because the image places curved surfaces at the wrong depth (Proposition 1) and the angular response of a surface element is a function of its size and curvature, not a free parameter (Corollary 2). Third, the subapertures of a specular aperture are not interchangeable views of one surface: each sees its own band of normals. A band that is held out therefore cannot be predicted from the others, which makes strip hold-out a weak learning signal and a poor model-selection rule, while generic cross-validation over random positions works as usual (Section 6). Held-out prediction of any kind cannot reject an interreflection that every strip sees, because it is the glint of a virtual surface (Proposition 2), nor a range offset, which is the same in every strip.

Limitations. The analysis assumes a monostatic sensor and single scattering, and the renderer shares both assumptions; the two-dimensional study shows that a strong double bounce, as in a dihedral, defeats both self-reconstruction and held-out selection, and that a third-order path can make held-out selection degenerate. The measured scan has strips only along y, so parallax is observed in one plane. On that scan, the crossing of the strip look lines and the direction of the streaks agree with the analysis, and the exact-scan model reconciles the main depth offsets. The true radius remains inseparable from the range offset in the phase history.

Outlook. The equivalence between a ghost and a virtual glint holds only for a co-located transmitter and receiver: the 3TX4RX array of the same radar, used as a multistatic aperture, observes the two traversal orders of a double bounce at different delays and can in principle break it, which a multi-bounce renderer (Armouti et al., 2026a) would then exploit.