CAUSTIC SPECTROSCOPY: IDENTIFIABILITY AND ACQUISITION LIMITS OF TOPOLOGICAL EVENT TIMING

Anonymous authors
Paper under double-blind review

ABSTRACT

Caustics carry the shape of a transparent surface, but their intensities are unstable under exposure and pose. We study a sparser, analytically tractable observable: sweep one knob of the illumination and record only the control values at which the caustic changes topology, a *catastrophe spectrum*. Isolated nondegenerate events are Morse critical points of the fold function, so their thresholds move with a closed-form differential, and the rank of the resulting event Jacobian certifies local identifiability: outright where it is full, which needs screen positions, and up to the sweep's symmetry group for thresholds alone, whose Jacobian that symmetry keeps below full rank. Paraxially the thresholds are the critical values of the surface's curvature fields. On simulated glass reliefs the differential matches finite differences to 9×10^{-11}, and the singular values order a trained network's per-direction error across twelve nominal reliefs after one held-out calibration. The analysis also marks its own limits, and we measure each. Umbilic events are degenerate and leave the signature unchanged. Nine events per relief let a trained inverse recover half the shape variance up to rotation, measured against the best data-free predictor and carried mostly by their counts and types. Adaptive bisection localizes matched events eighteen times more precisely than a uniform sweep, but only on the idealized signature: injecting spurious transitions into it reproduces the end-to-end collapse, which is complete by a precision of 0.67. All evidence is simulated.

1 INTRODUCTION

Glass, water and polished metal are usually measured through the light patterns they project. The dominant approaches fit those patterns: analysis-by-synthesis with differentiable light transport (Kassubeck et al., 2021; 2023; Nimier-David et al., 2019; Zhang et al., 2020; Lyu et al., 2020) including per-scene inverse rendering of transparent objects (Deng et al., 2024; Huang et al., 2025), or learned regression from images (Li et al., 2020; Sajjan et al., 2020). For both, caustics (the bright curves where rays focus) are the hardest part of the image: their intensity diverges, depends on exposure, sensor response and source size, and moves under any change of the camera.

Catastrophe theory gives a different view of the same curves (Thom, 1975; Berry & Upstill, 1980; Arnol'd, 1992; Nye, 1999). A caustic is the singular value set of a ray map. If one knob of the illumination is varied slowly (for example, a lamp moved along its axis), the caustic deforms smoothly almost everywhere and changes its topology only at isolated knob values. This suggests measuring a surface spectroscopically: sweep the light and record only *when* the pattern changes topology (Figure 1). We call the resulting short list of control values, with their topological types, a *catastrophe spectrum*. As a continuum object it is defined by the ray map alone, so it does not change when the camera's response curve, exposure or pose changes.

A few numbers instead of millions of pixels raises a question that is not specific to optics: *what can be learned from the timing of qualitative changes of a controlled system?* It arises wherever transitions are the informative part of a record, whether they are all one observes (Szep et al., 2021) or what one is trying to anticipate (Bury et al., 2021). Our answer is a differential: the map from shape to event timings has a closed-form derivative, and that derivative says which deformations are

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.

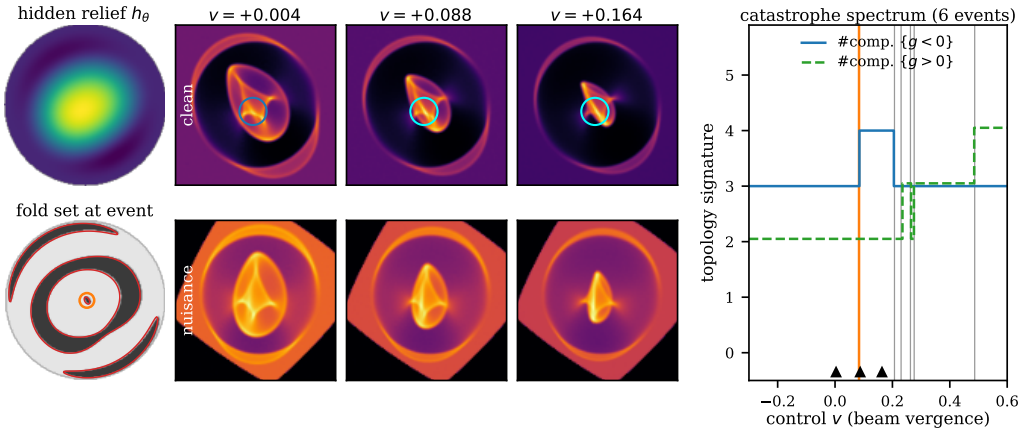

Figure 1: **A catastrophe spectrum.** A hidden glass relief (left) is lit by a beam whose vergence v is swept. The caustic network on the screen (top row) deforms smoothly almost everywhere; its topology changes only at isolated values of v (right: component counts of the fold set $\{g < 0\}$ and of its complement; gray lines are the six events, the orange one is a birth, whose fold set is shown bottom left and whose screen position is circled). The same frames seen through an unknown camera response and pose (bottom row: $\gamma = 0.5$, gain, noise, roll, tilt and lens distortion; relative pixel change above 125%) have the same thresholds under the gain, gamma and camera warp, because events are defined by the ray map and not by the pixels (Proposition 1). The added noise lies outside that proposition, which assumes a strictly increasing response; what a detector on pixels then keeps is measured in Section 5.5.

measurable, how accurately a trained estimator can recover them, and when an adaptive sweep can find the events at all. Each statement is tested against the prediction it makes, including where it fails.

- **Event timing has a closed-form differential, and its rank is an identifiability certificate.** Isolated nondegenerate events are Morse critical points of the fold function $g = \det D_{\boldsymbol{u}} F$ at level zero; they persist under perturbations of the shape and of the optical model, and their thresholds move with the envelope gradient $\nabla_{\boldsymbol{\theta}} \lambda^* = -\partial_{\boldsymbol{\theta}} g / \partial_\lambda g$ (Theorem 1). The rank of the resulting event Jacobian then certifies local identifiability: outright where it is full, which needs screen positions, and up to the sweep's symmetry group for thresholds alone, whose Jacobian that symmetry keeps below full rank (Theorem 2). We verify the differential to 9×10^{-11} and the predicted blind direction to our numerical floor, and report how often each rank condition actually holds under the prior (Section 5.1).
- **In a local regime the same Jacobian predicts the error profile of a learned inverse.** Around a nominal relief, the singular values order a network's error direction by direction after one effective noise scale, and calibrating that scale on *other* nominal reliefs still predicts the ordering, over the identified directions as well as over all of them (Corollary 1; Section 5.2). The exact statement is about a linear–Gaussian model; what we test is the hypothesis that a trained regressor follows it locally. Across a broad prior it stops holding (Section 5.3).
- **Acquisition succeeds exactly where the theory says it can, on the observable the theory assumes.** Bisecting on topology changes localizes events separated by more than the coarse spacing exponentially in the extra frames, while closer events are read as one aggregate jump and can cancel (Proposition 3); sweeping the separation reproduces both sides of that boundary. Read from rendered frames the gain nearly vanishes, and a controlled injection of spurious transitions into the idealized signature reproduces that collapse, so precision alone accounts for it (Section 5.4).

The theory also draws its own limits, and Section 6 reports each one measured: degenerate events, a prior over which the ordering stops holding, and a detector on pixels that is less accurate than

a pixel network. The spectrum is therefore not a more accurate estimator than pixels, which is not the claim. The claim is that topological transition times form a differentiable sensing modality whose Jacobian exposes, in closed form and in one place, which deformations are identifiable, how a learned inverse will distribute its error, and when an adaptive sweep is justified. The tools are classical (Morse theory, the implicit function theorem, the Cramér–Rao bound); the object is not, and every prediction it makes is tested here, including the ones that fail.

2 Related work

Caustics as signal. Caustic images recover 3D-printed glass by differentiable light transport (Kassubeck et al., 2021), accelerated with learned optimization (Kassubeck et al., 2023); caustic design shapes a surface to produce a target image (Schwartzburg et al., 2014); intentional caustics serve as wavefront sensors (Ribak, 2001); caustic-crossing times constrain binary microlensing (Cassan, 2008); and transparent-object reconstruction fits photometric or correspondence data (Ihrke et al., 2010; Morris & Kutulakos, 2011; Li et al., 2020; Deng et al., 2024; Huang et al., 2025), by differentiable refraction tracing (Lyu et al., 2020; Li et al., 2023) or neural fields (Zhao et al., 2024). All fit image content; we discard the intensities, keep only the control values at which the ray map's singularity changes topology, and ask what that list identifies.

Singularities and catastrophes in sensing. Caustics are Lagrangian singularities classified by catastrophe theory (Berry & Upstill, 1980; Nye, 1999; Arnol'd, 1992); Dangelmayr & Güttinger (1982) classified those that unknown surfaces imprint on sensing wavefields; aspect graphs catalogue viewpoint-indexed visual events (Koenderink & van Doorn, 1976; Petitjean et al., 1992); specular flow is singular at parabolic lines (Adato et al., 2010). That work classifies which singularities appear; we take their control values as the measurement and differentiate the map from shape to them, turning a classification into an identifiability certificate, an error prediction and an acquisition rule.

Topological transforms. Persistent homology and Euler characteristic transforms sweep a filtration of an observed shape and are injective descriptors (Turner et al., 2014; Ghrist et al., 2018; Curry et al., 2022), now differentiable layers (Röell & Rieck, 2024) and graph descriptors (von Rohrscheidt & Rieck, 2025) in topological deep learning (Papamarkou et al., 2024), with persistence used as features or losses (Hofer et al., 2017; Hu et al., 2019) and stability due to Cohen-Steiner et al. (2007). There the sweep filters an *observed* object; here it is a physical control acting through an *unobserved* map, whose symmetries the descriptor inherits and whose differential decides what is identifiable.

Bifurcations, identifiability and design. Szep et al. (2021) fit ODE parameters to bifurcation diagrams with implicit gradients and networks anticipate tipping points (Bury et al., 2021); identifiability as a rank condition is classical (Rothenberg, 1971), checked in practice by profile likelihood (Raue et al., 2009); amortized inference over simulators is simulation-based inference (Cranmer et al., 2020); choosing measurements is experimental design (Chaloner & Verdinelli, 1995; Rainforth et al., 2024; Foster et al., 2021; Huang et al., 2024; Iollo et al., 2025), also for illumination (Kellman et al., 2019); and invariance usually comes from augmentation or canonicalization (Kaba et al., 2023; Puny et al., 2022) rather than from the observable itself.

3 Catastrophe spectra

Setup. A transparent relief $z = h_{\boldsymbol{\theta}}(\boldsymbol{u})$, $\boldsymbol{\theta} \in \mathbb{R}^d$, on a flat glass plate is lit by a beam with control $\lambda \in \Lambda$ (the vergence v) and projects onto a screen; $F(\,\cdot\,;\boldsymbol{\theta},\lambda) : \mathbb{R}^2 \to \mathbb{R}^2$ maps a surface point $\boldsymbol{u}$ to its screen point (exact vector Snell refraction at both faces; Appendix B). Below, J is the event Jacobian with singular values σ_j, and NRMSE_q is the shape error up to rotation and reflection, normalised so that the prior mean scores 1; Table 2 lists every symbol. Screen irradiance is $I(\boldsymbol{x}) = \sum_{\boldsymbol{u}\in F^{-1}(\boldsymbol{x})} E(\boldsymbol{u})/|g(\boldsymbol{u})|$ with

$$g(\boldsymbol{u};\boldsymbol{\theta},\lambda) = \det D_{\boldsymbol{u}} F(\boldsymbol{u};\boldsymbol{\theta},\lambda), \tag{1}$$

so I diverges on the caustic $F(\Sigma_\lambda)$, $\Sigma_\lambda = \{g = 0\}$. The fold set $\{g < 0\}$ is where the ray sheets overlap in reversed orientation, and its *topology signature* $\beta(\lambda) = (\#\text{components of } \{g_\lambda < 0\}, \#\text{components of } \{g_\lambda > 0\})$ summarises its topology.

Definition 1 (Event, observed transition, catastrophe spectrum)**.** $(\boldsymbol{u}^*, \lambda^*)$ is a *geometric event* of $\boldsymbol{\theta}$ if $g = 0$ and $\nabla_{\boldsymbol{u}} g = 0$ there, i.e. a critical point of $g_\lambda = g(\cdot; \boldsymbol{\theta}, \lambda)$ at level zero. It is *nondegenerate* if $\det \nabla^2_{\boldsymbol{u}} g \neq 0$ (the critical point is a Morse point) and $\partial_\lambda g \neq 0$ (its critical value crosses zero at rate $\partial_\lambda g$), and *isolated* if no other geometric event shares its control value. An *observed transition* at λ^* is the jump $\tau = \beta(\lambda^* + 0) - \beta(\lambda^* - 0)$ of the signature, an aggregate over all geometric events at that control value. The *catastrophe spectrum* is $S(\boldsymbol{\theta}) = \{(\lambda_i^*, \tau_i, \boldsymbol{x}_i^*)\}_{i=1}^K$ over isolated nondegenerate events, with $\boldsymbol{x}_i^* = F(\boldsymbol{u}_i^*; \boldsymbol{\theta}, \lambda_i^*)$, and the *threshold spectrum* is $S_\lambda(\boldsymbol{\theta}) = \{(\lambda_i^*, \tau_i)\}$. Events that fail either nondegeneracy condition are excluded; those with $\partial_\lambda g = 0$ exist and are not rare (umbilics are one family, Proposition 2), and Appendix I shows that they leave β unchanged.

Lemma 1 (Isolated nondegenerate events are the topology changes)**.** *If $g > 0$ near the rim of the relief, the fold sets $\{g_\lambda < 0\}$ are isotopic on every interval of Λ free of events, and at an isolated nondegenerate event exactly one of the component counts of $\{g_\lambda < 0\}$ and $\{g_\lambda > 0\}$ changes by one, so $\tau \in \{(\pm 1, 0), (0, \pm 1)\}$. Several events at one control value contribute the sum of their jumps, which can cancel.*

Lemma 1 (Ehresmann plus the Morse lemma, Appendix A) is what turns the spectrum into a measurement rather than a definition: the events are exactly the jumps of one integer-valued function of the control. Reading β off a frame is a separate detection problem that we do not solve; the detector of Section 5.5 counts components of bright quantile masks, a related but different quantity, and Section 5.4 measures what the difference costs. Three caveats travel with the lemma and reappear as measured effects: the type is not a function of the Morse index alone (a saddle either merges two components of $\{g < 0\}$ or closes a loop, depending on the global configuration); events sharing a control value are read as one jump and can cancel (Section 5.4); and degenerate events are invisible to β (Appendix I).

Proposition 1 (Invariance)**.** *Let the camera observe $\varphi\big(a\, I \circ \psi^{-1}\big)$ with ψ any diffeomorphism of the screen onto the image, φ strictly increasing and $a > 0$. Thresholds and types are unchanged (if ψ reverses orientation, the two entries of β swap roles), and positions map by ψ.*

The proof is one line: $\det D(\psi \circ F) = (\det D\psi \circ F)\, g$ shares the zero set, critical points and sublevel topology of g. This is a statement about the continuum object; what a detector on pixels inherits is measured in Section 5.5.

4 What a catastrophe spectrum identifies

We now differentiate the map from shape to spectrum, read identifiability off the rank of that differential, turn its singular values into a risk for a learned inverse, and say what the spectrum means paraxially and how to acquire it.

Theorem 1 (Stability and envelope gradient)**.** *Let $(\boldsymbol{u}^*, \lambda^*)$ be a nondegenerate event of $\boldsymbol{\theta}_0$ and $F \in C^k$, $k \geq 4$. On a neighbourhood of $\boldsymbol{\theta}_0$ there is a unique C^{k-2} branch of nondegenerate events, with the same Morse index and sign of $\partial_\lambda g$, and*

$$\nabla_{\boldsymbol{\theta}} \lambda^* = -\frac{\partial_{\boldsymbol{\theta}} g}{\partial_\lambda g}, \qquad D_{\boldsymbol{\theta}} \boldsymbol{u}^* = -\big(\nabla^2_{\boldsymbol{u}} g\big)^{-1} \big(\partial_{\boldsymbol{\theta}} \nabla_{\boldsymbol{u}} g + \partial_\lambda \nabla_{\boldsymbol{u}} g\, \nabla_{\boldsymbol{\theta}} \lambda^{*\top}\big), \tag{2}$$

all evaluated at the event. If every event of $\boldsymbol{\theta}_0$ in Λ is nondegenerate and interior, the number and types of events are locally constant, and so is their order when the thresholds $\lambda_i^*(\boldsymbol{\theta}_0)$ are pairwise distinct; a model perturbation $F + \epsilon G$ with $\|G\|_{C^3} \leq 1$ moves each threshold by $O(\epsilon)$.

The key step is that at an event the Jacobian of $(g, \nabla_{\boldsymbol{u}} g)$ in $(\boldsymbol{u}, \lambda)$ has determinant $\partial_\lambda g \det \nabla^2_{\boldsymbol{u}} g$, and $\nabla_{\boldsymbol{u}} g = 0$ removes the $\boldsymbol{u}$-dependence from the threshold gradient. The theorem is qualitative, but the radius can be estimated, and what binds is crowding rather than degeneracy: the spectrum keeps its event count only while no two thresholds meet, which on our reliefs means $\|\delta\boldsymbol{\theta}\| \lesssim 0.01$ (Appendix A, measured in Section 5.2). Equation (2) gives the event Jacobian $J(\boldsymbol{\theta}) = \partial s / \partial \boldsymbol{\theta}$ of any stacked event vector s in closed form, without differentiating through an event solver.

Theorem 2 (Identifiability)**.** *Let s collect event coordinates on a neighbourhood where the spectrum is stable, and $J = Ds$. (a)* (sufficient) *If $\operatorname{rank} J(\boldsymbol{\theta}_0) = d$, then s is injective near $\boldsymbol{\theta}_0$ and $\|\boldsymbol{\theta} - \boldsymbol{\theta}'\| \leq (1 + o(1))\, \|s(\boldsymbol{\theta}) - s(\boldsymbol{\theta}')\| / \sigma_{\min}(J)$; under Gaussian event-coordinate noise of covariance Σ every unbiased estimator has covariance $\succeq (J^\top \Sigma^{-1} J)^{-1}$, which is $\sigma_\epsilon^2 (J^\top J)^{-1}$ for $\Sigma = \sigma_\epsilon^2 \boldsymbol{I}$. (b)*

(modulo the acquisition symmetry) *Let a compact group G of dimension p act on Θ with $s(g \cdot \boldsymbol{\theta}) = s(\boldsymbol{\theta})$ and with trivial stabiliser at $\boldsymbol{\theta}_0$, so that the tangent space $\mathcal{T}$ to the orbit of $\boldsymbol{\theta}_0$ has dimension p. If J restricted to a complement of $\mathcal{T}$ has rank $d-p$, then s separates orbits near $\boldsymbol{\theta}_0$ and the bound of (a) holds for the quotient distance with $\sigma_{\min}$ taken on that complement. This, not (a), is the statement available for the threshold spectrum, because (d) caps* $\operatorname{rank} J_\lambda$ *at $d-1$. (c)* (persistent deficiency) *If* $\operatorname{rank} J = r < d$ *on an open set, the level sets of s are $(d-r)$-dimensional manifolds and only the quotient is identifiable. By (a), rank deficiency at $\boldsymbol{\theta}_0$ is* necessary *for s to be non-injective in every neighbourhood of $\boldsymbol{\theta}_0$, but it is not sufficient ($s(\boldsymbol{\theta}) = \boldsymbol{\theta}^3$ on $\mathbb{R}$). (d)* (symmetry) *If a group acts on Θ with $s(g \cdot \boldsymbol{\theta}) = s(\boldsymbol{\theta})$, every infinitesimal generator ξ satisfies $J\xi = 0$. For an axial sweep of a radially windowed relief the threshold spectrum is $O(2)$-invariant, so the rotation generator lies in* $\ker J_\lambda$ *and* $\operatorname{rank} J_\lambda \le d-1$*; recording the screen position of at least one event at nonzero radius removes that generator from the kernel and generically raises the rank, but does not by itself make J full rank.*

Corollary 1 (Predicted risk of a learned inverse)**.** *In the linearization $s = s(\boldsymbol{\theta}_0) + J(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \boldsymbol{\epsilon}$ with prior $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}, \rho^2 \boldsymbol{I})$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \boldsymbol{I})$, the Bayes risk of the posterior mean is*

$$R(\boldsymbol{\theta}_0) = \sum_{j=1}^{d'} \frac{\rho^2 \sigma_\epsilon^2}{\sigma_\epsilon^2 + \rho^2 \sigma_j(J)^2}, \tag{3}$$

with $\sigma_j = 0$ beyond the rank and $d' = d$; the dependence on the nominal is through $J = J(\boldsymbol{\theta}_0)$. For the quotient risk, $d' = d-1$ and the singular values are those of J restricted to the complement of the orbit direction.

Equation (3) is exact for the stated linear–Gaussian model. Applying it to a trained network is a separate, empirical hypothesis, which we state so that it can be tested: *within a local regime in which the event set is stable, the singular values of $J(\boldsymbol{\theta}_0)$ order the directional error of a squared-loss regressor after one free scale σ_ϵ, and directions in* $\ker J$ *keep their prior variance.* Because the networks see exact thresholds, σ_ϵ is not measurement noise but the regressor's approximation error in threshold units, so Section 5.2 calibrates it on *other* nominal reliefs before predicting. Section 5.3 shows where the unimodality the corollary assumes fails.

Proposition 2 (Paraxial spectroscopy)**.** *In the paraxial thin-element model with vergence control v, $g \propto (c(v) + \kappa_1)(c(v) + \kappa_2)$ where $\kappa_1 \ge \kappa_2$ are the eigenvalues of $\nabla^2 h$ and $c(v) = (1 + Lv)/(Lk)$ ($k = n - 1$, L the effective screen distance). Away from umbilics the nondegenerate events are exactly the Morse critical points of κ_1 and κ_2, with thresholds $v^* = -k\,\kappa^* - 1/L$ at the critical value κ^*. At an umbilic ($\kappa_1 = \kappa_2$) the matrix $c\boldsymbol{I} + \nabla^2 h$ vanishes, so $g = 0$,* $\nabla_{\boldsymbol{u}} g = 0$ and $\partial_v g = 0$*: umbilics solve the event equations but are degenerate, and the signature does not jump there.*

Thus a vergence sweep reads the Morse critical values of the surface's curvature fields on its convex (focusing) part, like spectral lines, and umbilics are not among them. The number of such values bounds what thresholds alone can identify: fewer than $d-1$ lines cannot determine a d-parameter relief up to rotation. Umbilics are not rare (Appendix I: 1.4 per relief fall inside our window), so it matters that they are excluded by Definition 1 rather than counted: the critical value of g touches zero quadratically instead of crossing it, a discrete detector reports a cancelling pair of transitions whose separation shrinks with its resolution, and both effects are measured in Appendix I. The exact optics adds non-paraxial corrections that vanish as all ray angles shrink (Section 5.1).

Proposition 3 (Catastrophe-seeking acquisition)**.** *Let a coarse pass of N_0 frames have spacing $\Delta_0 = (\lambda_{\max} - \lambda_{\min})/(N_0 - 1)$, and let B be the total budget. (a) Every event separated from its neighbours by more than Δ_0 is alone in a coarse interval, is bracketed by a signature change, and is localized by bisection to error $\le \Delta_0\, 2^{-\lfloor (B-N_0)/K_b \rfloor - 1}$, where $K_b \le K$ is the number of brackets the coarse pass opens. (b) Events closer than Δ_0 may share a coarse interval, and those that do are read as one aggregate jump (Lemma 1), which may cancel; they can then be missed by any strategy that reads only β. A uniform sweep of B frames has worst-case localization error at least $(\lambda_{\max} - \lambda_{\min})/(2(B-1))$.*

The predicted operating regime is therefore a separation condition rather than a budget: exponential for isolated events, and no better than cancellation allows for crowded ones. Section 5.4 sweeps the ratio δ/Δ_0 across this boundary. Where the threshold Jacobian has maximal quotient rank, Theorem 2(b) turns the threshold error into a shape error up to the symmetry, so that error inherits the exponential decay.

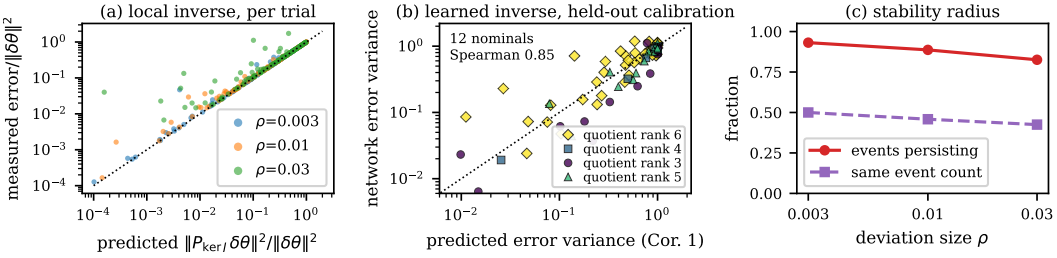

Figure 2: **Local regime.** (a) One-step local inverse from persisting thresholds: measured squared error against the null-space prediction of Theorem 2, per trial. (b) Networks trained on deviations around 12 nominal reliefs: error variance along each right-singular direction of the analytic threshold Jacobian against Corollary 1, with the single free scale σ_ϵ calibrated on the other 11 nominals (84 directions; points at 1 are directions the Jacobian declares blind). (c) Fraction of events that persist, and of spectra that keep their event count, as the deviation grows. Dotted lines: equality.

5 Experiments

Setup (details in Appendix B). Reliefs are windowed Zernike expansions of radial order 2–3 ($d = 7$) or 2–4 ($d = 12$) with $\boldsymbol{\theta} \sim \mathcal{N}(0, \boldsymbol{I})$ on a glass plate ($n = 1.5$), lit by a beam whose vergence sweeps $v \in [-0.3, 0.6]$, with exact vector Snell refraction at both faces. A median relief has 9 events in the window (mean 9.9, 10–90% range 3–17). The *oracle* spectrum is read from the fold function on a 161^2 grid (Lemma 1); the *image* spectrum from rendered 96^2 frames (46 per sweep) by a calibration-free detector, a heuristic superset of the oracle events (Section 5.5). Spectra go to DeepSets encoders (Zaheer et al., 2017), pixels to frame-stack CNNs. We report NRMSE_q, the shape error up to rotation and reflection normalised so that the prior mean scores 1, over 3 seeds unless stated. That is not the no-information level: the minimisation over $O(2)$ aligns a nonzero constant with each target, so the best data-free constant scores 0.702 for $d = 7$ and 0.808 for $d = 12$ (Table 1), and explained variance is read against it. Writing NRMSE_0 for that level, a model at NRMSE_q explains $1 - (\mathrm{NRMSE}_q/\mathrm{NRMSE}_0)^2$ of the variance.

5.1 Checking the differential, the blind direction and the paraxial limit

Appendix F reports these with Figure 4; two of the three are consistency checks on the implementation rather than falsifiable predictions, and we label them so. The gradient matches central finite differences to 9×10^{-11}; $\|J_\lambda \xi\|/\|J_\lambda\|$ sits at 6×10^{-9}, the numerical floor of our pipeline, and rises to 0.031 under a 0.4 rad beam tilt that breaks the symmetry; and scaling the ray angles drives the exact spectrum to the curvature law of Proposition 2. Both rank conditions are reported: the threshold Jacobian cannot exceed $d - 1 = 6$ and attains it for 50% of these reliefs (the hypothesis of Theorem 2b), while with screen positions it reaches full rank 7 (the hypothesis of a) for 85%.

5.2 Local regime: inspecting deviations from a nominal relief

This is the regime Theorem 2 speaks to: a nominal relief $\boldsymbol{\theta}_0$ is known (a design, or a previous measurement) and a deviation $\delta\boldsymbol{\theta}$ must be estimated, as in quality control of printed glass (Kassubeck et al., 2021). We test two predictions before looking at the data.

Prediction 1 (model-based). The one-step inverse $J_m^+ \delta s$ from the m persisting events should err by the component of $\delta\boldsymbol{\theta}$ in $\ker J_m$. To first order that is an identity, so what it measures is the $O(\rho^2)$ remainder: over 360 trials the mean relative error is 0.561 against a predicted 0.558 at $\rho = 0.003$, and 0.646 against 0.618 at $\rho = 0.03$ (Figure 2a). The informative panel is Figure 2(c): 93% of events persist at $\rho = 0.003$ and 83% at $\rho = 0.03$, but the event count survives for only 50% and 43% of spectra.

Prediction 2 (learned). A network trained on deviations $\delta\boldsymbol{\theta} \sim \mathcal{N}(0, \rho^2 \boldsymbol{I})$ with ordinary squared loss approximates the posterior mean of Corollary 1 (no quotient here: the nominal fixes the orientation), so its error variance along the j-th right-singular direction of $J(\boldsymbol{\theta}_0)$ should be $\rho^2\sigma_\epsilon^2/(\sigma_\epsilon^2 + \rho^2\sigma_j^2)$,

Table 1: **Global regime**: NRMSE_q (mean $\pm$ s.d. over 3 seeds, 500 test reliefs) of DeepSets encoders on oracle spectra. The first row is the no-information level: the best constant fitted with the same $O(2)$-quotient loss, which is what explained variance should be read against (the prior mean scores 1 by construction). The middle block holds attribution controls that keep the event counts and types; the last row (two seeds) shows what the rendered images carry. The d=12 and reflection columns use the fast oracle (Appendix B), which gives 0.496 ± 0.003 for refraction with thresholds and types.

Input to the network	refraction, d=7	refraction, d=12	reflection, d=7
no input (best constant, same loss)	0.702	0.808	0.692
thresholds + types (the spectrum S_λ)	**0.498** $\pm$ 0.002	**0.719** $\pm$ 0.003	**0.500** $\pm$ 0.003
types only (order-free multiset)	0.555 $\pm$ 0.003	0.730 $\pm$ 0.003	0.551 $\pm$ 0.000
thresholds only	0.619 $\pm$ 0.001	0.774 $\pm$ 0.001	0.615 $\pm$ 0.002
thresholds + types + screen positions	0.501 $\pm$ 0.003	–	–
positions + types (no thresholds)	0.556 $\pm$ 0.006	–	–
number of events only (no timing, no type)	0.638 $\pm$ 0.002	–	–
thresholds shuffled across reliefs	0.557 $\pm$ 0.003	–	–
thresholds drawn at random	0.570 $\pm$ 0.002	–	–
reference: CNN on 23 rendered frames (pixels)	0.355 $\pm$ 0.010	–	–

and the prior variance along null directions. For each of 12 nominals spanning the quotient ranks the prior produces (3 to 6) we train a DeepSets encoder on 490 thresholds-only spectra ($\rho = 0.03$) and measure its error on 140 held-out deviations. Figure 2(b) shows the result with σ_ϵ calibrated on the *other* eleven nominals: over the 84 directions, predicted and measured variances agree within a factor of two for 83% of them (Pearson 0.88 on log variance, Spearman 0.85); per-nominal calibration, the best case, gives 0.97 and 96%. The 25 directions the Jacobian declares blind sit at the prior level (median 0.95, range 0.76 to 1.13). Those are correct by construction, so we also report the ordering without them: over the 59 identified directions it is unchanged (Pearson 0.85, Spearman 0.85, 76% within a factor of two).

Because the networks see exact thresholds, σ_ϵ is the estimator's approximation error in threshold units rather than measurement noise, which is why it is calibrated and not assumed. The regime is local but not strictly fixed-event: at $\rho = 0.03$ the event count survives for 43% of draws (Figure 2c), and that shows in the result, since the deficit in null directions (down to 0.76) is the network using exactly what the linearisation ignores, namely changes in the count. Since $\rho = 0.03$ is about eight times the crowding radius of Appendix A, we repeated six nominals inside it: at $\rho = 0.01$ and 0.003 *every* one of the 42 directions agrees within a factor of two, against 83% at $\rho = 0.03$ (Appendix F), so the prediction improves where its hypothesis holds. What is tested remains a local surrogate, not an exact description of the estimator.

5.3 Global regime: what a spectrum carries across the prior

Without a nominal, a network must invert the spectrum over the whole prior. Table 1 reports NRMSE_q on 500 held-out reliefs. Thresholds with their types reach $\mathrm{NRMSE}_q = 0.498$ for $d = 7$ with 2000 training reliefs, against 0.702 for the best data-free constant: a median of nine thresholds explains half of the shape variance up to rotation. The same holds for a mirror relief (0.500 against 0.496 for refraction read with the same oracle), whose forward map differs but whose event system does not. The attribution controls separate the three things a spectrum carries: how many events occur, of which types, and when. The number of events alone already gives 0.638 and the multiset of types (which also reveals the number) 0.555, that is 0.17 and 0.38 of the variance against the 0.50 the full spectrum reaches; thresholds without types give 0.619, and keeping counts and types while shuffling thresholds across reliefs or drawing them at random costs 0.06 to 0.07 in NRMSE_q, that is 0.13 to 0.16 of the variance; drawing them at random is worse than withholding them, since a wrong threshold is worse than none. Timing is informative, but three quarters of what the spectrum carries is in how many catastrophes occur and of which kind. With $d = 12$ parameters and a similar number of events the spectrum explains only a fifth of the variance (0.719 against a no-input 0.808), the regime Proposition 2 describes when there are fewer spectral lines than unknowns. Screen positions

identify orientation (plain NRMSE 0.69 instead of 1.24) but add nothing to the shape up to rotation at this training-set size, and the level is set by data rather than capacity (Table 4).

A negative result. Here the local risk of Corollary 1 barely predicts which reliefs the thresholds-only network gets wrong: the Spearman correlation between $R(\boldsymbol{\theta}_0)$ and the per-relief error is 0.20 over 100 reliefs (0.08 with σ_ϵ held out); with positions it rises to 0.50. We looked for the collisions that would explain it. Among the 2750 reliefs, 21 pairs agree in event count, type multiset and every threshold to within one oracle step (1.25×10^{-3} in v) while their shapes differ by a median of 0.63; but recomputing every event with Newton refinement separates them (3.8×10^{-5} to 2.7×10^{-3}, median 7.3×10^{-4}, none coinciding). What the prior contains is ambiguity at the acquisition resolution, not non-injectivity of the continuous map, and we claim only the former. The pairs are not arbitrary either: 41 of the 42 reliefs carry a single event, the regime Proposition 2 already counts as hopeless, since fewer than $d - 1 = 6$ spectral lines cannot fix a seven-parameter relief up to rotation. No estimator reading the spectrum at this resolution separates them, however much data it sees; a finer sweep would. Global identifiability of topological sweeps, known for filtrations of observed shapes (Turner et al., 2014; Curry et al., 2022), is open for physical sweeps.

5.4 Acquisition succeeds on the signature exactly where the theory says it can

Proposition 3 makes a two-sided prediction: events separated by more than the coarse spacing Δ_0 are bracketed and localized exponentially in the extra frames, while closer events are read as one aggregate jump and can cancel. We test both sides at two levels of observation: the *idealized signature* computed from the fold function, which needs the unknown ray map, and *end to end*, with the signature read by the detector of Section 5.5 from one rendered 96^2 frame per control value. Thresholds are matched to the precise oracle spectrum within 0.1 in v, the coarse pass uses $N_0 = 10$ frames ($\Delta_0 = 0.1$), and the budget runs from 12 to 48 frames on 40 reliefs.

Budget. On the idealised signature seeking spends its extra frames inside brackets and localises far better than a uniform sweep: at $B = 48$ the median error over matched events is 2.7×10^{-4} against 5.0×10^{-3}, a factor of 18, widening with the budget (Figure 3a), at equal recall (0.66), since the coarse pass decides what is bracketed and bisection only sharpens it. Read from rendered frames the advantage nearly disappears, 9.1×10^{-3} against 1.02×10^{-2}, both near the 46-frame spacing. The reason is precision, not recall: only 18 to 22% of the detector's transitions correspond to a fold-set event, against 97% for the signature.

Boundary. The prior is dominated by the crowded regime that (b) describes: of 392 events, 327 lie below $\Delta_0/2$ and 39 above Δ_0. Seeking recovers 0.61 of the crowded events and 0.88, 1.00, 1.00 in the three separated bins (26, 25 and 14 events), localising what it finds to 1–4×10^{-4} (Figure 3b,c). The uniform sweep rises with separation too, as (b) implies, but not identically: between $\Delta_0/2$ and Δ_0 it reaches 1.00 against seeking's 0.88, because there seeking has spent only its ten-frame coarse pass against the sweep's 46. Cancellation in β sets the floor, and the coarse pass decides how much of the rest seeking recovers.

Is precision the cause? Holding the idealized signature, the events, the separations, the budget and the policy fixed, we inject spurious transitions at a controlled density so that precision is the only thing that varies (Appendix H). The gain falls from 18.3 at precision 1 to 4.9 at 0.80 and 1.0 at 0.67, never exceeding 1.1 below that: it is gone far above the detector's 0.18 to 0.22, so precision on its own suffices to remove the gain, and the missing piece is a detector rather than a policy. The sweep also explains why the detector's recall looks flat while its errors do not improve: matched recall *rises* with the false-positive rate, from 0.68 to 0.99, because spurious brackets leave an estimate near every true event by chance.

5.5 What a detector on pixels keeps

Proposition 1 is exact for the continuum spectrum, and the detector is not the spectrum: it reports 18.5 events per relief against 9.9 oracle events, recovering 76% of them within ± 2 frames at a precision of 0.53 (the stricter matching of Section 5.4 gives 0.18 to 0.22). Models are trained on clean renders, or with random nuisances where stated, and tested under radiometric and camera shifts (Appendix D; Figure 5 and Table 3 in Appendix G list every condition). In distribution pixels win

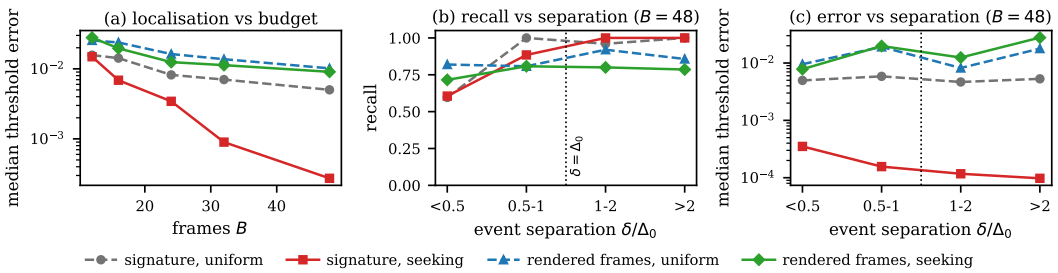

Figure 3: **Acquisition and its predicted boundary** (40 reliefs; coarse pass $N_0 = 10$, so $\Delta_0 = 0.1$). (a) Median threshold error against the frame budget, for the idealized signature and for signatures read from single rendered frames. (b, c) Recall and error against an event's separation from its nearest neighbour, in units of Δ_0; dotted line: the condition of Proposition 3(a).

clearly, 0.355 for a rank-normalised CNN against 0.553 for the image spectrum, because images keep what the spectrum discards. Rank normalisation makes both exactly invariant to gain and gamma; clipping at saturation is not strictly monotone and costs the spectrum 0.003, and noise makes the masks flicker (0.582 at $\sigma = 0.05$). Under tilt and zoom the spectrum's error rises by at most 0.08 while pixel errors rise by 0.16 to 0.42: invariance buys a flat profile, not a lower error, and augmentation buys more, since a CNN trained with random nuisances stays between 0.421 and 0.456 everywhere. On the 20 reliefs where we can afford it, a per-scene analysis-by-synthesis fit is more accurate than either (0.321 clean against 0.370 and 0.559) and collapses only when the camera is left unmodelled (Table 5). The spectrum is thus the more analysable observable and the least accurate one; its gap to the invariant oracle spectrum (0.553 against 0.498) bounds what a better detector could recover.

6 Discussion and Limitations

What closes. Around a nominal relief the chain is complete: events persist with a closed-form differential, its rank says which deformations are measurable and which are blind, its singular values order a trained inverse's error with the scale calibrated on other nominals, and the predicted acquisition boundary appears where Proposition 3 puts it. Table 6 lists every claim with its scope and failure case.

What does not, and why. Umbilics solve the event equations, occur 1.4 times per relief and leave the signature unchanged (Appendix I). Across a broad prior the local risk stops predicting per-sample error (Spearman 0.20), as the unimodality of Corollary 1 warns, and the reliefs whose spectra collide at our resolution are those Proposition 2 counts as underdetermined. On pixels the detector keeps the radiometric invariance and part of the geometric one, and its false positives, alone, break the acquisition gain (Appendix H).

Scope. All data are simulated (exact vector refraction; no diffraction, dispersion, scattering or interreflection), the reliefs are smooth and low-dimensional ($d \leq 12$), and the control is one-dimensional; a mirror in place of the plate leaves the event system intact (Appendix E). A finite source is a superposition of tilted beams, and a 0.1 rad tilt already moves thresholds by twenty-six times the seeking precision, so source size is the first thing a physical sweep must answer. Events cluster here (53% of neighbouring thresholds within one frame spacing, mostly near the window's rim), the regime Proposition 3(b) describes.

7 Conclusion

A caustic changes topology at a short list of control values, and the map from shape to that list has a differential. What it buys is not accuracy, which pixels win, but the ability to say in advance what a measurement resolves and what it is blind to. We show this for one optical system, and offer the rest as a proposal: differentiate the event system of a controlled experiment, read identifiability modulo its symmetries from the rank, and measure at the transitions.

AI USE STATEMENT

In this paper, AI helps to aid and polish writing. We used a large-language-model assistant throughout the work and disclose its uses here. *Writing:* drafting and polishing prose from our notes and results, and reviewing drafts for clarity, consistency and claims that outrun their evidence. *Code:* as a coding agent, writing and refactoring the simulator, the oracle, the detector, the models and the analysis scripts to our specifications, and generating the plots in this paper from saved experiment outputs. *Discussion:* as an interlocutor while we developed the framework, chose the experiments and interpreted the results. *Presentation:* preparing slides for our internal meetings. We also used it to search and summarise related literature and to format references.

We have reviewed all AI-assisted work and take full responsibility for it. Every statement that admits a numerical check was checked: the envelope gradient against finite differences, the predicted blind direction, the paraxial limit, the degeneracy of umbilics, and the no-information level of the reported metric. Every number in the paper is produced by a script that reads saved experiment outputs, and every reference was verified against its publisher or DOI record.

ETHICS STATEMENT

This work uses only simulated optical data; it involves no human subjects, personal data or deployed systems. Its application domain (non-contact metrology of transparent and specular surfaces) has no foreseeable harmful use beyond that of existing optical inspection methods.

REPRODUCIBILITY STATEMENT

The optical model, relief family, sweep, oracle, image detector, network architectures, training protocol, seeds and evaluation metric are specified in Appendix B; nuisance conditions are listed in Appendix D; complete proofs are in Appendix A, and the degenerate-event analysis in Appendix I. All datasets are generated deterministically from the stated seeds by the simulator described there, and the simulator, the oracle, the detector, the models and the analysis scripts that produce every number and figure are included in the supplementary material. Every result states its sample size (reliefs, trials or events) and, for trained models, the number of seeds with the spread across them; single-seed results are marked as such. Appendix J maps every claim to the statement that supports it and the experiment that tests it, and lists each assumption with the check we ran against our own data.

REFERENCES

Marc Kassubeck, Florian Bürgel, Susana Castillo, Sebastian Stiller, and Marcus Magnor. Shape from caustics: Reconstruction of 3D-printed glass from simulated caustic images. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2876–2885, 2021.

Marc Kassubeck, Moritz Kappel, Susana Castillo, and Marcus Magnor. N-SfC: Robust and fast shape estimation from caustic images. In *Vision, Modeling, and Visualization (VMV)*, pp. 33–41. The Eurographics Association, 2023.

Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. Mitsuba 2: A retargetable forward and inverse renderer. *ACM Transactions on Graphics*, 38(6):203:1–203:17, 2019.

Cheng Zhang, Bailey Miller, Kai Yan, Ioannis Gkioulekas, and Shuang Zhao. Path-space differentiable rendering. *ACM Transactions on Graphics*, 39(4):143:1–143:19, 2020.

Jiahui Lyu, Bojian Wu, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Differentiable refraction-tracing for mesh reconstruction of transparent objects. *ACM Transactions on Graphics*, 39(6):195:1–195:13, 2020.

Weijian Deng, Dylan Campbell, Chunyi Sun, Shubham Kanitkar, Matthew E. Shaffer, and Stephen Gould. Differentiable neural surface refinement for modeling transparent objects. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20268–20277, 2024.

Letian Huang, Dongwei Ye, Jialin Dan, Chengzhi Tao, Huiwen Liu, Kun Zhou, Bo Ren, Yuanqi Li, Yanwen Guo, and Jie Guo. TransparentGS: Fast inverse rendering of transparent objects with gaussians. *ACM Transactions on Graphics*, 44(4):1–17, 2025.

Zhengqin Li, Yu-Ying Yeh, and Manmohan Chandraker. Through the looking glass: Neural 3D reconstruction of transparent shapes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1259–1268, 2020.

Shreeyak Sajjan, Matthew Moore, Mike Pan, Ganesh Nagaraja, Johnny Lee, Andy Zeng, and Shuran Song. ClearGrasp: 3D shape estimation of transparent objects for manipulation. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3634–3642, 2020.

René Thom. *Structural Stability and Morphogenesis: An Outline of a General Theory of Models*. W. A. Benjamin, 1975.

M. V. Berry and C. Upstill. Catastrophe optics: morphologies of caustics and their diffraction patterns. In *Progress in Optics*, volume 18, pp. 257–346. Elsevier, 1980.

V. I. Arnol'd. *Catastrophe Theory*. Springer, third edition, 1992.

J. F. Nye. *Natural Focusing and Fine Structure of Light: Caustics and Wave Dislocations*. Institute of Physics Publishing, 1999.

Gregory Szep, Neil Dalchau, and Attila Csikász-Nagy. Parameter inference with bifurcation diagrams. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Thomas M. Bury, R. I. Sujith, Induja Pavithran, Marten Scheffer, Timothy M. Lenton, Madhur Anand, and Chris T. Bauch. Deep learning for early warning signals of tipping points. *Proceedings of the National Academy of Sciences*, 118(39):e2106140118, 2021.

Yuliy Schwartzburg, Romain Testuz, Andrea Tagliasacchi, and Mark Pauly. High-contrast computational caustic design. *ACM Transactions on Graphics*, 33(4):74:1–74:11, 2014.

Erez N. Ribak. Harnessing caustics for wave-front sensing. *Optics Letters*, 26(23):1834–1836, 2001.

Arnaud Cassan. An alternative parameterisation for binary-lens caustic-crossing events. *Astronomy & Astrophysics*, 491(2):587–595, 2008.

Ivo Ihrke, Kiriakos N. Kutulakos, Hendrik P. A. Lensch, Marcus Magnor, and Wolfgang Heidrich. Transparent and specular object reconstruction. *Computer Graphics Forum*, 29(8):2400–2426, 2010.

Nigel J. W. Morris and Kiriakos N. Kutulakos. Dynamic refraction stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1518–1531, 2011.

Zongcheng Li, Xiaoxiao Long, Yusen Wang, Tuo Cao, Wenping Wang, Fei Luo, and Chunxia Xiao. NeTO: Neural reconstruction of transparent objects with self-occlusion aware refraction-tracing. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 18501–18511, 2023.

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

G. Dangelmayr and W. Güttinger. Topological approach to remote sensing. *Geophysical Journal of the Royal Astronomical Society*, 71(1):79–126, 1982.

J. J. Koenderink and A. J. van Doorn. The singularities of the visual mapping. *Biological Cybernetics*, 24(1):51–59, 1976.

Sylvain Petitjean, Jean Ponce, and David J. Kriegman. Computing exact aspect graphs of curved objects: Algebraic surfaces. *International Journal of Computer Vision*, 9(3):231–255, 1992.

Yair Adato, Yuriy Vasilyev, Todd Zickler, and Ohad Ben-Shahar. Shape from specular flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):2054–2070, 2010.

Katharine Turner, Sayan Mukherjee, and Doug M. Boyer. Persistent homology transform for modeling shapes and surfaces. *Information and Inference: A Journal of the IMA*, 3(4):310–344, 2014.

Robert Ghrist, Rachel Levanger, and Huy Mai. Persistent homology and Euler integral transforms. *Journal of Applied and Computational Topology*, 2(1–2):55–60, 2018.

Justin Curry, Sayan Mukherjee, and Katharine Turner. How many directions determine a shape and other sufficiency results for two topological transforms. *Transactions of the American Mathematical Society, Series B*, 9:1006–1043, 2022.

Ernst Röell and Bastian Rieck. Differentiable Euler characteristic transforms for shape classification. In *International Conference on Learning Representations*, 2024.

Julius von Rohrscheidt and Bastian Rieck. Diss-l-ECT: Dissecting graph data with local Euler characteristic transforms. In *International Conference on Machine Learning*, pp. 61790–61809, 2025.

Theodore Papamarkou, Tolga Birdal, Michael M. Bronstein, Gunnar E. Carlsson, Justin Curry, Yue Gao, Mustafa Hajij, Roland Kwitt, Pietro Liò, Paolo Di Lorenzo, Vasileios Maroulas, Nina Miolane, Farzana Nasrin, Karthikeyan Natesan Ramamurthy, Bastian Rieck, Simone Scardapane, Michael T. Schaub, Petar Veličković, Bei Wang, Yusu Wang, Guowei Wei, and Ghada Zamzmi. Position: Topological deep learning is the new frontier for relational learning. In *International Conference on Machine Learning*, pp. 39529–39555, 2024.

Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. Topology-preserving deep image segmentation. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.

Thomas J. Rothenberg. Identification in parametric models. *Econometrica*, 39(3):577–591, 1971.

Andreas Raue, Clemens Kreutz, Thomas Maiwald, Julie Bachmann, Marcel Schilling, Ursula Klingmüller, and Jens Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

Tom Rainforth, Adam Foster, Desi R. Ivanova, and Freddie Bickford Smith. Modern Bayesian experimental design. *Statistical Science*, 39(1):100–114, 2024.

Adam Foster, Desi R. Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential Bayesian experimental design. In *International Conference on Machine Learning*, pp. 3384–3395, 2021.

Daolang Huang, Yujia Guo, Luigi Acerbi, and Samuel Kaski. Amortized Bayesian experimental design for decision-making. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Jacopo Iollo, Christophe Heinkelé, Pierre Alliez, and Florence Forbes. Bayesian experimental design via contrastive diffusions. In *International Conference on Learning Representations*, 2025.

Michael R. Kellman, Emrah Bostan, Nicole A. Repina, and Laura Waller. Physics-based learned design: Optimized coded-illumination for quantitative phase imaging. *IEEE Transactions on Computational Imaging*, 5(3):344–353, 2019.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *International Conference on Machine Learning*, pp. 15546–15566, 2023.

Omri Puny, Matan Atzmon, Heli Ben-Hamu, Ishan Misra, Aditya Grover, Edward J. Smith, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations*, 2022.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R. Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

John Milnor. *Morse Theory*. Princeton University Press, 1963.

E. L. Lehmann and George Casella. *Theory of Point Estimation*. Springer, second edition, 1998.

John M. Lee. *Introduction to Smooth Manifolds*. Springer, second edition, 2012.

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*, pp. 3744–3753, 2019.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

A PROOFS

Throughout, $F(\,\cdot\,;\boldsymbol{\theta},\lambda)$ is C^k ($k \geq 4$) jointly in $(\boldsymbol{u},\boldsymbol{\theta},\lambda)$ on $\Omega \times \Theta \times \Lambda$, $g = \det D_{\boldsymbol{u}} F$ is C^{k-1}, and the relief is flat near $\partial\Omega$ so that $g > 0$ on a neighbourhood N of $\partial\Omega$ for all $(\boldsymbol{\theta},\lambda)$ considered (in our optics the normalised ray map is the identity on the flat surround, so $g = 1$ there). Write $\Omega' = \Omega \setminus N$ and $\Phi(\boldsymbol{u},\lambda;\boldsymbol{\theta}) = (g, \partial_1 g, \partial_2 g) \in \mathbb{R}^3$.

A.1 Events are exactly the topology changes (Lemma 1)

Proof. Let $I = [\alpha, \beta] \subset \Lambda$ contain no event. Then $\nabla_{\boldsymbol{u}} g \neq 0$ on $Z = \{(\boldsymbol{u}, \lambda) \in \Omega' \times I : g = 0\}$, so Z is a compact embedded surface and the projection $\pi(\boldsymbol{u}, \lambda) = \lambda$ restricted to Z is a submersion: the tangent space $\ker(\nabla_{\boldsymbol{u}} g, \partial_\lambda g)$ contains a vector with $d\lambda = 1$ because $\nabla_{\boldsymbol{u}} g \neq 0$. Since 0 is a regular value of g_λ for every $\lambda \in I$, the gradient flow of the sublevel function carries $\Sigma_\lambda = \{g_\lambda = 0\}$ to $\Sigma_{\lambda'}$ diffeomorphically (Milnor, 1963, Thm. 3.1), and integrating the vector field $X = (V, 1)$ with $V = -\partial_\lambda g \, \nabla_{\boldsymbol{u}} g / \|\nabla_{\boldsymbol{u}} g\|^2$ near Z (extended by a cut-off) yields an ambient isotopy carrying $\{g_\alpha \leq 0\}$ onto $\{g_\lambda \leq 0\}$. Hence all topological invariants of the fold set $\{g_\lambda < 0\}$ are constant on I. At a nondegenerate event $(\boldsymbol{u}^*, \lambda^*)$ the Morse lemma with parameters gives, for λ near λ^*, local coordinates $\boldsymbol{y}$ in which $g = c(\lambda) + Q(\boldsymbol{y})$, where $c(\lambda)$ is the value of g_λ at its critical point near $\boldsymbol{u}^*$, $c(\lambda^*) = 0$, $c'(\lambda^*) = \partial_\lambda g \neq 0$, and Q is a nondegenerate quadratic form of index $\operatorname{ind} \nabla^2_{\boldsymbol{u}} g$; as λ crosses λ^* the level $-c(\lambda)$ of the sublevel set $\{Q \leq -c(\lambda)\}$ crosses 0 transversally, so a handle of that index is attached or removed (Milnor, 1963), changing the Euler characteristic of $\{g_\lambda \leq 0\}$ by $\pm(-1)^{\mathrm{ind}}$. For index 0 (index 2) a component of $\{g < 0\}$ (of $\{g > 0\}$) is born or dies. For index 1 the handle either joins two components of $\{g \leq 0\}$, which leaves the complement connected as before, or closes a loop; since the complement contains the flat surround, Alexander duality in the plane turns the new loop into one new component of $\{g > 0\}$. In every case exactly one of the two counts changes, by one. Away from events 0 is a regular value, so $\{g_\lambda < 0\}$ is the interior of $\{g_\lambda \leq 0\}$ and has the same topology. If several events share the control value λ^*, the handles are attached at distinct points of the same level set, so the jump of β across λ^* is the sum of the individual jumps and may cancel; the statement is therefore about isolated events, and the aggregate is what a sweep observes. □

A.2 Invariance (Proposition 1)

Proof. The camera observes $I_{\mathrm{obs}}(\boldsymbol{y}) = \varphi\big(a\, I(\psi^{-1}(\boldsymbol{y}))\big)$ with ψ a diffeomorphism from the screen to the image plane, φ strictly increasing and $a > 0$. The composite observation map is $\tilde{F} = \psi \circ F$, whose critical function is $\tilde{g} = (\det D\psi \circ F)\, g = c\, g$ with $c \neq 0$ of constant sign (positive if ψ preserves orientation). Hence $\{\tilde{g} = 0\} = \{g = 0\}$; on this set $\nabla \tilde{g} = c \nabla g$, so the critical points coincide; there $\nabla^2 \tilde{g} = c \nabla^2 g$ (the terms $\nabla c \, \nabla g^\top$ vanish) and $\partial_\lambda \tilde{g} = c\, \partial_\lambda g$. Therefore events and thresholds are identical, and so is the sublevel topology: $\{\tilde{g} < 0\} = \{g < 0\}$ if $c > 0$, while for an orientation-reversing ψ ($c < 0$) the sets $\{\tilde{g} < 0\}$ and $\{g > 0\}$ coincide, so the two entries of β and of every type swap. The radiometric map φ and the gain a do not enter F at all. Positions transform as $\tilde{\boldsymbol{x}}^* = \psi(\boldsymbol{x}^*)$. □

A.3 Stability and the envelope formula (Theorem 1)

Proof. At an event the Jacobian of Φ with respect to $(\boldsymbol{u}, \lambda)$ is

$$D_{(\boldsymbol{u},\lambda)} \Phi = \begin{pmatrix} \nabla_{\boldsymbol{u}} g^\top & \partial_\lambda g \\ \nabla^2_{\boldsymbol{u}} g & \partial_\lambda \nabla_{\boldsymbol{u}} g \end{pmatrix} = \begin{pmatrix} 0 \;\; 0 & \partial_\lambda g \\ \nabla^2_{\boldsymbol{u}} g & \partial_\lambda \nabla_{\boldsymbol{u}} g \end{pmatrix}, \qquad \det D_{(\boldsymbol{u},\lambda)} \Phi = \partial_\lambda g \, \det \nabla^2_{\boldsymbol{u}} g,$$

which is nonzero by nondegeneracy. The implicit function theorem yields neighbourhoods $U \ni \boldsymbol{\theta}_0$, $W \ni (\boldsymbol{u}^*, \lambda^*)$ and a unique C^{k-2} map $\boldsymbol{\theta} \mapsto (\boldsymbol{u}^*(\boldsymbol{\theta}), \lambda^*(\boldsymbol{\theta})) \in W$ with $\Phi = 0$. By continuity $\det \nabla^2_{\boldsymbol{u}} g$ and $\partial_\lambda g$ keep their signs on U, so the event stays nondegenerate with the same Morse index and sign of $\partial_\lambda g$. Differentiating $g(\boldsymbol{u}^*(\boldsymbol{\theta}); \boldsymbol{\theta}, \lambda^*(\boldsymbol{\theta})) \equiv 0$ gives $\nabla_{\boldsymbol{u}} g^\top D_{\boldsymbol{\theta}} \boldsymbol{u}^* + \nabla_{\boldsymbol{\theta}} g + \partial_\lambda g \, \nabla_{\boldsymbol{\theta}} \lambda^* = 0$, and the first term vanishes because $\nabla_{\boldsymbol{u}} g = 0$ at events; this is the envelope formula. Differentiating $\nabla_{\boldsymbol{u}} g \equiv 0$ gives the formula for $D_{\boldsymbol{\theta}} \boldsymbol{u}^*$, and the chain rule for $\boldsymbol{x}^* = F(\boldsymbol{u}^*; \boldsymbol{\theta}, \lambda^*)$ gives $D_{\boldsymbol{\theta}} \boldsymbol{x}^*$.

For the whole sweep: the event set $E(\boldsymbol{\theta}_0) = \Phi^{-1}(0) \cap (\Omega' \times \Lambda)$ is compact and consists of isolated points (each is a regular zero of Φ), hence is finite, $\{e_1, \ldots, e_K\}$, and by assumption none lies on $\partial\Lambda$. Choose disjoint neighbourhoods $W_i \ni e_i$ as above. On the compact set $C = (\Omega' \times \Lambda) \setminus \cup_i W_i$ we have $\|\Phi(\cdot; \boldsymbol{\theta}_0)\| \geq \eta > 0$, and by uniform continuity $\|\Phi(\cdot; \boldsymbol{\theta})\| > 0$ on C for $\boldsymbol{\theta}$ close to $\boldsymbol{\theta}_0$; together with uniqueness in each W_i this shows that the number of events is locally constant. Types are locally constant too: for λ outside the small intervals $\pi(W_i)$, 0 is a regular value of $g(\cdot; \boldsymbol{\theta}, \lambda)$ for all $\boldsymbol{\theta}$ on the segment from $\boldsymbol{\theta}_0$, and the isotopy argument of Lemma 1 with $\boldsymbol{\theta}$ as the parameter makes the fold sets of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ isotopic there, so β agrees between consecutive events and so do its jumps. The order of thresholds is preserved whenever the $\lambda_i^*(\boldsymbol{\theta}_0)$ are distinct, and each λ_i^* is locally Lipschitz

because $\nabla_{\boldsymbol{\theta}}\lambda_i^*$ is continuous. For a model perturbation $F_\epsilon = F + \epsilon G$ apply the same argument with ϵ as an extra parameter (Φ is C^1 in ϵ when $G \in C^3$): $\partial_\epsilon \lambda^* = -\partial_\epsilon g / \partial_\lambda g = O(\|G\|_{C^1})$. □

Reliefs with an umbilic in the window. The whole-sweep part assumes every event of $\boldsymbol{\theta}_0$ in Λ is nondegenerate, and Appendix I reports that 72% of reliefs have an umbilic threshold inside the window, so the hypothesis as stated covers a minority of the prior. The conclusion survives because the two event sets are separated. Umbilics are the zeros of the two-component map $\boldsymbol{\theta} \mapsto$ (deviatoric part of $c\boldsymbol{I} + \nabla^2 h$) in the two surface coordinates; for generic reliefs this map is transverse to zero, so umbilic points are isolated and persist under small perturbations with their thresholds moving continuously. They contribute no jump of β (Appendix I), so excluding a neighbourhood of each from Λ leaves the signature and its jumps unchanged, and the argument above applies on the complement, which contains every nondegenerate event. What is *not* covered is a relief at which a nondegenerate event collides with an umbilic; those form a codimension-one set that our prior meets with probability zero, and we do not analyse it.

How large is the neighbourhood? Theorem 1 is qualitative, but two radii can be written down, and in our relief family the binding one is not the one about degeneracy. For a *single* event, quantifying the implicit function theorem gives existence and uniqueness of the branch for $\|\delta\boldsymbol{\theta}\| \lesssim \sigma^2/(LC)$, where $\sigma = \sigma_{\min}(D_{(\boldsymbol{u},\lambda)}\Phi)$ at the event and L, C bound $\|D^2\Phi\|$ and $\|\partial_{\boldsymbol{\theta}}\Phi\|$ nearby: an event is fragile exactly when it is close to degenerate, since both $\partial_\lambda g \to 0$ and $\det \nabla^2_{\boldsymbol{u}} g \to 0$ shrink σ. The *spectrum* needs more. Its event count and order survive only while no two thresholds meet, that is while $\|\delta\boldsymbol{\theta}\| < \delta_{\min}/(2\max_i \|\nabla_{\boldsymbol{\theta}}\lambda_i^*\|)$, where $\delta_{\min}$ is the smallest gap between neighbouring thresholds and the gradients come from Equation (2). Both quantities are available on the theory set: the median relief has $\delta_{\min} = 0.015$ in v, three quarters of a frame spacing, and $\max_i \|\nabla_{\boldsymbol{\theta}}\lambda_i^*\| = 0.73$, so the crowding radius is $\|\delta\boldsymbol{\theta}\| \approx 0.010$, with an interquartile range of 0.003 to 0.028. That is the scale at which Figure 2(c) shows event counts breaking down ($\rho = 0.003$ per coefficient is $\|\delta\boldsymbol{\theta}\| = 0.008$ in $d = 7$) even though 93% of individual events persist. The bound covers threshold collisions only, so it is not the whole story: events also leave through the ends of the sweep or the rim of the relief, which is why the measured fraction is not simply the fraction of reliefs whose radius exceeds $\|\delta\boldsymbol{\theta}\|$. Crowding rather than degeneracy sets the practical radius, and it is the same crowding that limits recall in Section 5.4.

A.4 Identifiability (Theorem 2)

Proof. (a) Let $\sigma = \sigma_{\min}(J(\boldsymbol{\theta}_0)) > 0$ and let ω be a modulus of continuity of Ds at $\boldsymbol{\theta}_0$. For $\boldsymbol{\theta}, \boldsymbol{\theta}'$ in the ball $B_r(\boldsymbol{\theta}_0)$, $s(\boldsymbol{\theta}) - s(\boldsymbol{\theta}') = \bar{J}(\boldsymbol{\theta} - \boldsymbol{\theta}')$ with $\bar{J} = \int_0^1 Ds(\boldsymbol{\theta}' + t(\boldsymbol{\theta} - \boldsymbol{\theta}'))\, dt$ and $\|\bar{J} - J(\boldsymbol{\theta}_0)\| \le \omega(r)$, hence $\sigma_{\min}(\bar{J}) \ge \sigma - \omega(r)$ and $\|\boldsymbol{\theta} - \boldsymbol{\theta}'\| \le \|s(\boldsymbol{\theta}) - s(\boldsymbol{\theta}')\|/(\sigma - \omega(r))$, which proves local injectivity for $\omega(r) < \sigma$. For the local least-squares estimator $\hat{\boldsymbol{\theta}} \in \arg\min_{B_r} \|s(\boldsymbol{\theta}) - \hat{s}\|$ with $\hat{s} = s(\boldsymbol{\theta}) + \boldsymbol{\epsilon}$: $\|s(\hat{\boldsymbol{\theta}}) - s(\boldsymbol{\theta})\| \le 2\|\boldsymbol{\epsilon}\|$, so $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\| \le 2\|\boldsymbol{\epsilon}\|/(\sigma - \omega(r))$; a first-order expansion sharpens the constant to $\|J^+\boldsymbol{\epsilon}\| \le \|\boldsymbol{\epsilon}\|/\sigma$. Under $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \Sigma)$ the Fisher information is $J^\top \Sigma^{-1} J$, which gives the Cramér–Rao bound (Lehmann & Casella, 1998). (b) The stabiliser at $\boldsymbol{\theta}_0$ is trivial, so the orbit is a p-dimensional submanifold near $\boldsymbol{\theta}_0$; choose a slice $S \ni \boldsymbol{\theta}_0$ transverse to it, of dimension $d - p$; by hypothesis $D(s|_S)$ has full rank $d - p$, so (a) applies on S and $s|_S$ is injective with the stated constant. Invariance makes s constant on orbits, and near $\boldsymbol{\theta}_0$ every orbit meets S in one point, so s separates orbits and the bound transfers to the quotient distance. (c) If $\operatorname{rank} Ds \equiv r < d$ on an open set V, the constant-rank theorem (Lee, 2012) provides local coordinates in which $s(\boldsymbol{\theta}) = (\theta_1, \dots, \theta_r, 0, \dots, 0)$; the fibre $s^{-1}(s(\boldsymbol{\theta}))$ is locally a $(d - r)$-dimensional submanifold and every function of s is constant on it. (d) Differentiating $s(\exp(tA)\cdot\boldsymbol{\theta}) = s(\boldsymbol{\theta})$ at $t = 0$ gives $J(\boldsymbol{\theta})\,\xi_A(\boldsymbol{\theta}) = 0$. For the axial acquisition, rotating the relief by $R \in SO(2)$ gives $F_{R\boldsymbol{\theta}}(\boldsymbol{u}) = RF_{\boldsymbol{\theta}}(R^{-1}\boldsymbol{u})$ because the beam, the plate and the normalised screen coordinates are rotationally symmetric; thus $g_{R\boldsymbol{\theta}}(\boldsymbol{u}) = g_{\boldsymbol{\theta}}(R^{-1}\boldsymbol{u})$, events map to $(R\boldsymbol{u}^*, \lambda^*)$ with the same type, thresholds are invariant and positions rotate. The same computation with a reflection P (where $\det P \, \det P^{-1} = 1$) gives $O(2)$ invariance. □

A.5 Linear–Gaussian Bayes risk (Corollary 1)

Proof. Under $s = s(\boldsymbol{\theta}_0) + J(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \boldsymbol{I})$, and the prior $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{I})$, the posterior covariance is $P = (J^\top J/\sigma_\epsilon^2 + \boldsymbol{I})^{-1}$, whose eigenvalues are $\sigma_\epsilon^2/(\sigma_\epsilon^2 + \sigma_j^2)$ with σ_j the singular values of J (zero beyond the rank). The posterior mean minimises the expected squared error and its risk is $\operatorname{tr} P$. For the quotient error $\min_{g \in O(2)} \|g\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2$, the minimisation over the orbit removes, to leading order in the posterior spread, the error component along the orbit direction ξ; the remaining risk is the trace of P restricted to $\xi^\perp$, i.e. the sum over the $d-1$ singular values of J restricted to $\xi^\perp$ (for thresholds only, $J\xi = 0$ and these are the nonzero singular values of J, padded with zeros). □

A.6 Paraxial spectroscopy (Proposition 2)

Proof. Paraxially, a ray with slope $\boldsymbol{a} = \boldsymbol{t} + v\boldsymbol{u}$ refracts at the relief into slope $\boldsymbol{a}/n + (1 - 1/n)\nabla h$, crosses the plate and leaves the flat face with slope $\boldsymbol{a} + k\nabla h$ ($k = n - 1$), so the screen point is $\boldsymbol{x} = \boldsymbol{u} + L(\boldsymbol{a} + k\nabla h)$ with $L = D + T/n$. In normalised coordinates $\tilde{F}(\boldsymbol{u}) = (\boldsymbol{x} - L\boldsymbol{t})/m = \boldsymbol{u} + \frac{Lk}{m}\nabla h(\boldsymbol{u})$ with $m = 1 + Lv$, hence $g = \det(\boldsymbol{I} + \frac{Lk}{m}\nabla^2 h) = (\frac{Lk}{m})^2 (c + \kappa_1)(c + \kappa_2)$ with $c(v) = m/(Lk)$ and $\kappa_1 \geq \kappa_2$ the eigenvalues of $\nabla^2 h$. At a non-umbilic point with $-\kappa_2 = c > -\kappa_1$ the eigenvalues are smooth and $\nabla g = C(c + \kappa_1)\nabla\kappa_2$ with $C = (Lk/m)^2 > 0$ (the other term vanishes because $c + \kappa_2 = 0$), so the point is an event iff it is a critical point of κ_2 with value $-c(v^*)$, i.e. $v^* = -k\,\kappa_2(\boldsymbol{u}^*) - 1/L$; there $\nabla^2 g = C(c + \kappa_1)\nabla^2\kappa_2$. The case $-\kappa_1 = c$ is symmetric. At an umbilic ($\kappa_1 = \kappa_2 = -c$) the matrix $c\boldsymbol{I} + \nabla^2 h$ vanishes, hence so do g and ∇g (the adjugate of the zero 2×2 matrix is zero) and also $\partial_v g = c'(v) \operatorname{tr}\operatorname{adj}(c\boldsymbol{I} + \nabla^2 h) = 0$: the point solves the event equations but is degenerate, and Appendix I shows that the signature does not jump there. Thus the thresholds of the spectrum are the affine image $v^* = -k\kappa^* - 1/L$ of the critical values of the curvature fields on the convex part (where $-\kappa > 0$ can reach $c > 0$), and umbilics are excluded. □

A.7 Catastrophe-seeking acquisition (Proposition 3)

Proof. (a) If an event is farther than Δ_0 from every other event, its coarse interval contains no other event, so by Lemma 1 the signature differs at the interval's endpoints and the event is bracketed. Bisection keeps the invariant that the event lies in the current bracket, halving it per frame, so after m_i extra frames the midpoint is within $\Delta_0 2^{-m_i - 1}$ of λ_i^*. Round-robin allocation of the $B - N_0$ extra frames over the K_b brackets gives, for every event that is bracketed, $|\hat{\lambda}_i - \lambda_i^*| \leq \Delta_0\, 2^{-\lfloor (B - N_0)/K_b \rfloor - 1}$; events that are not bracketed are covered by part (b). A uniform sweep of B frames has worst-case localisation error at least $(\lambda_{\max} - \lambda_{\min})/(2(B - 1))$, since the event may lie anywhere inside its interval; a particular event may of course fall close to a sampled control value. (b) If two or more events lie in the same coarse interval, the sweep sees only the sum of their jumps (Lemma 1); when that sum vanishes the interval is not bracketed and the events are missed, and when it does not, one bracket is used for several events, so the bound of (a) applies to their aggregate rather than to each. Combining (a) with Theorem 2(b), the local shape error *up to the acquisition symmetry* is at most the threshold error times $(1 + o(1))\sqrt{K}/\sigma_{\min}^{\perp}(J_\lambda)$, where $\sigma_{\min}^{\perp}$ is the smallest singular value of J_λ restricted to a complement of the orbit direction ξ; so it inherits the exponential decay. Part (a) of Theorem 2 is unavailable here, since $J_\lambda \xi = 0$. □

B Optical model, data and training details

Reliefs. $h_{\boldsymbol{\theta}}(\boldsymbol{u}) = A\, w(r) \sum_{k=1}^{d} s_k \theta_k Z_k(\boldsymbol{u})$ on the unit disk, with Zernike polynomials Z_k of radial orders $2 \leq n \leq n_{\max}$ (piston and tilt only translate the caustic and are omitted), $s_k = 1/(n_k(n_k - 1))$ so that each mode contributes curvature of order one, the window $w(r) = \max(1 - r^4, 0)^4$ (which makes h vanish with three derivatives at $r = 1$, so that $g > 0$ on the flat surround), and $\boldsymbol{\theta} \sim \mathcal{N}(0, \boldsymbol{I}_d)$. We use $n_{\max} = 3$ ($d = 7$) and $n_{\max} = 4$ ($d = 12$), $A = 0.25$ for refraction and $A = 0.0625$ for reflection.

Optics. Refraction: a glass plate ($n = 1.5$, thickness $T = 0.2$, flat bottom), screen at distance $D = 2$ below the bottom face; exact vector Snell refraction at both faces, Fresnel transmittance for the irradiance. Reflection: a mirror relief and a screen at height $Z_s = 2$. The beam has vergence v (a point source at height $1/v$ for $v > 0$, a converging beam for $v < 0$, collimated at $v = 0$) and

optional tilt α; the control of the sweep is $v \in [-0.3, 0.6]$. Screen coordinates are divided by the flat-plate magnification $1 + Lv$ ($L = D + T/n$), a known diffeomorphism that keeps the pattern in view and changes neither topology nor thresholds.

Oracle spectra. The fold function g is evaluated on a 161×161 grid over the disk, either exactly (forward-mode differentiation through the ray map) or by second-order central differences of the ray map (relative deviation 4×10^{-3}; component counts agree on 98–100% of sweep steps). The topology signature is computed with 8-connectivity for $\{g < 0\}$ and 4-connectivity for its complement. The *fast* oracle reports each change of the signature between consecutive steps of a 91-step sweep at the interval midpoint (± 0.005 in v); the *precise* oracle bisects every change to 10^{-4} of the range (2×10^{-3} for the 2750-relief training and test sets, whose thresholds therefore land on a 1.25×10^{-3} grid in v; a jump of two units within this tolerance is recorded as two events at the same threshold), attaches it to the best-conditioned discrete critical point of g with value closest to zero, and refines $(\boldsymbol{u}^*, \lambda^*)$ by Newton's method on $(g, \nabla_{\boldsymbol{u}} g) = 0$ with exact derivatives; the event Jacobian is then evaluated with Equation (2).

Images. Frames are rendered by forward splatting of 512^2 jittered rays (bilinear), with a Gaussian point-spread of 0.7 pixel, at 96×96 pixels over the normalised screen window $[-1.25, 1.25]^2$, for 46 values of v.

Image detector. Each frame is rank-normalised, smoothed (Gaussian, 1 pixel) and thresholded at the 90% and 95% quantiles of the valid image region; the components of each mask (8-connectivity, at least 4 pixels) and of its complement (4-connectivity) give a four-number signature per frame, median-filtered over three frames; its changes are the image events $(v, \text{level}, \text{dimension}, \text{sign}, \text{magnitude})$.

Models. Set encoder (DeepSets, Zaheer et al., 2017): a per-event MLP (two layers of width 192, GELU), masked mean and max pooling concatenated with $\log(1 + K)$, and an MLP head (widths 384 and 192); 0.26M parameters. Oracle event features are the threshold rescaled to $[-1, 1]$, a one-hot type (which component count changes, and in which direction) and, where stated, the screen position; image events carry the threshold, the sign and size of the count change, and a one-hot of (quantile level, mask or complement). We use DeepSets rather than attention blocks (Lee et al., 2019), which were twenty times slower on our hardware; the capacity control of Section 5.3 (widths 96 to 384) shows that the global results are not capacity-limited. Pixel baseline: a convolutional network on every other frame (23 frames), average-pooled to 48×48 after the nuisance is applied, frames as channels (a stride-1 block and four stride-2 stages of widths 48 to 192, GroupNorm, global pooling; 0.91M parameters), with rank-normalised or $\log(1 + I)$ input. Set models are trained with AdamW (Loshchilov & Hutter, 2019) (learning rate 10^{-3}, weight decay 10^{-4}, batch 128, cosine schedule, gradient clipping at 1) for at most 300 epochs with early stopping (patience 50) on a validation split of 250 reliefs; pixel models use a one-cycle schedule (peak 2×10^{-3}, batch 32) for 30 epochs (20 for log input, 25 with augmentation) and keep the best validation checkpoint. The 2750 reliefs of a set split into 2000 for training, 250 for validation and 500 for test, disjoint throughout. Set models use three seeds; the rank-normalised pixel model two seeds and the other pixel variants one seed, because pixel runs dominate our compute budget (a laptop GPU). Every row of Table 1 is trained with the $O(2)$-quotient squared error $\min_{g \in O(2)} \|g\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2$ (72 rotations $\times$ reflection), so that all rows are trained for the metric the table reports. Inputs that carry screen positions do identify orientation, and training those with the plain squared error instead gives a plain NRMSE of 0.69 against 1.24 without positions (the pair quoted in Section 5.3) at the cost of the quotient error, which rises from 0.501 to 0.544: the network spends capacity on the orientation the metric quotients out. We report the quotient normalised RMSE $\mathrm{NRMSE}_q = (\mathbb{E} \min_g \|g\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 / \mathbb{E}\|\boldsymbol{\theta}\|^2)^{1/2}$ (1 for the constant prior-mean predictor).

Analysis-by-synthesis baseline. In the style of Shape from Caustics (Kassubeck et al., 2021), $\boldsymbol{\theta}$ is fitted by Adam (learning rate 0.05, 80 steps, 2 starts: zero and a random draw) so that differentiably splatted frames (4 of the 46 vergences, 256^2 rays, 1.5-pixel blur) match the observed ones in standardised log intensity, a loss invariant to gain and nearly invariant to gamma (the logarithm turns a power law into a scale factor that the standardisation removes), but not to camera geometry. It knows the optics and the camera and is run on the first 20 test reliefs.

Compute. Everything is implemented in PyTorch (Paszke et al., 2019) and ran on one laptop (Intel i7-10610U, NVIDIA Quadro P520 with 2 GB). The precise oracle takes about 1.5 s per relief on one

CPU thread; a set model trains in 1 to 9 minutes, a pixel model in 15 to 20 minutes per seed, and analysis-by-synthesis takes about 40 s per relief.

C NOTATION

Table 2 collects every symbol used in the main text, with the statement or section that introduces it.

Table 2: Notation.

Symbol	Meaning
$\boldsymbol{\theta} \in \mathbb{R}^d$	relief parameters (windowed Zernike coefficients), prior $\mathcal{N}(0, \boldsymbol{I}_d)$
$h_{\boldsymbol{\theta}}(\boldsymbol{u})$	relief height at surface point $\boldsymbol{u} \in \mathbb{R}^2$ (unit disk, flat outside)
λ, v	control of the sweep; in our optics the beam vergence $v \in [-0.3, 0.6]$
$F(\boldsymbol{u}; \boldsymbol{\theta}, \lambda)$	ray map from surface point to normalised screen point
$g = \det D_{\boldsymbol{u}} F$	fold function; the caustic is the image of $\Sigma_\lambda = \{g = 0\}$
$(\boldsymbol{u}^*, \lambda^*)$	event: critical point of g_λ at level 0 (Definition 1)
τ	event type: the jump of β across the event (which component count changes, and in which direction)
$\boldsymbol{x}^* = F(\boldsymbol{u}^*)$	screen position of an event
$S(\boldsymbol{\theta})$, $S_\lambda(\boldsymbol{\theta})$	catastrophe spectrum and threshold spectrum
K	number of events in the swept window
$\beta(\lambda)$	topology signature: component counts of $\{g_\lambda < 0\}$ and $\{g_\lambda > 0\}$
$s(\boldsymbol{\theta})$	stacked event coordinates (thresholds, optionally positions)
$J = \partial s / \partial \boldsymbol{\theta}$	event Jacobian; σ_j its singular values; J_λ its threshold rows alone
$\xi(\boldsymbol{\theta})$	infinitesimal generator of the in-plane rotation of the relief
$R(\boldsymbol{\theta})$	linear-Gaussian Bayes risk (Equation (3)); σ_ϵ effective threshold noise
$\boldsymbol{\theta}_0$, $\delta\boldsymbol{\theta}$, ρ	nominal relief, deviation from it, and deviation scale per coefficient (Section 5.2)
J_m, δs	Jacobian of the m events that persist, and the change of their coordinates (Section 5.2)
$\kappa_1 \geq \kappa_2$	eigenvalues of $\nabla^2 h$ (principal curvatures, small-slope)
n, $k = n - 1$	refractive index and refractive power
$L = D + T/n$	effective screen distance (T plate thickness, D plate-to-screen distance)
$m = 1 + Lv$	flat-plate magnification used to normalise screen coordinates
$c(v) = m/(Lk)$	curvature level swept by the vergence (Proposition 2)
B, N_0, Δ_0, δ, K_b	frame budget, coarse frames, coarse spacing, threshold gap, number of brackets (Proposition 3)
ψ, φ, a	camera diffeomorphism, radiometric response, source power (Proposition 1)
NRMSE_q	$O(2)$-quotient normalised RMSE; 1 for the prior-mean predictor

D NUISANCE CONDITIONS

Radiometric nuisances act on the (warped) irradiance I as $I' = \min\{\max\{a\, I^\gamma + b + \eta, 0\}, s\}$ with gain a, gamma γ, offset b, saturation s, and i.i.d. Gaussian pixel noise η whose standard deviation is σ times the frame mean of $(aI^\gamma + b)^{1/2}$ (a shot-noise scale). Geometric nuisances are applied first and resample the frames under a smooth map of the image plane composed of a roll by ϕ, a tilt (perspective foreshortening along one axis by $\tan\beta$), a scale, radial distortion $x \mapsto x(1 + k_1 r^2)$, and a smooth random elastic field (bicubic interpolation of a 5×5 Gaussian grid with standard deviation e). The single-factor ladder is: gain 0.25 with offset 0.05; $\gamma = 0.5$; $\gamma = 2$; saturation at $3\times$ the flat level; noise $\sigma = 0.02$ and 0.05; roll 90°; tilt 20° with $k_1 = 0.1$; elastic $e = 0.03$; scale 0.8. The combined severity $\ell \in \{1, \ldots, 5\}$ uses $s_\ell \in \{0.25, 0.5, 0.75, 1, 1.5\}$ and sets $\gamma = 2^{s_\ell}$, $a = 4^{-s_\ell}$, $\sigma = 0.02 s_\ell$, roll $60^\circ s_\ell$, tilt $15^\circ s_\ell$, $k_1 = 0.1 s_\ell$, $e = 0.015 s_\ell$. Augmented training draws, per sample, a and γ log-uniformly in $[1/4, 4]$ and $[1/2, 2]$, $\sigma \in [0, 0.03]$, roll in $[-180^\circ, 180^\circ]$, tilt in $[-20^\circ, 20^\circ]$, $k_1 \in [-0.15, 0.15]$, $e \in [0, 0.02]$ and scale $e^{\pm 0.15}$.

E REFLECTION

Replacing the glass plate by a mirror relief ($A = 0.0625$, screen at height $Z_s = 2$, same vergence sweep) changes the forward map but not the event system: with the same (fast) oracle, the median relief has 8 events for reflection and 8 for refraction (2750 reliefs each), and Table 1 shows that the learned inverse from the threshold spectrum reaches the same accuracy as for refraction. The paraxial analysis of Proposition 2 carries over with $k = 2$ and reversed sign: a mirror focuses where the relief is concave.

F CHECKS OF THE THEORY

Two of these are consistency checks on the implementation rather than falsifiable predictions, and we label them as such. Figure 4(a): over 125 events of 20 reliefs, envelope-formula gradients (Equation (2)) match central finite differences of Newton-refined thresholds to a median relative error of 9×10^{-11}, so event Jacobians need no differentiation through an event solver. (b): $\|J_\lambda \xi\|/\|J_\lambda\|$ has median 6×10^{-9}, the numerical floor of our pipeline, over the 293 of 300 theory reliefs with a refined event, while the position rows give 0.40 (Theorem 2d). Tilting the beam by α breaks the symmetry only through non-paraxial effects, and the ratio rises to 0.011, 0.020 and 0.031 for $\alpha = 0.1, 0.2, 0.4$: orientation becomes identifiable, with a small singular value. The two rank conditions differ and we report both: the threshold Jacobian cannot exceed $d - 1 = 6$, and attains it for 50% of these reliefs, which is the hypothesis of Theorem 2(b); with screen positions it reaches full rank 7, the hypothesis of (a), for 85%. (c) is a real test: scaling every ray angle by s should drive the exact spectrum to the curvature law of Proposition 2, and the deviation falls by a factor of 8 between $s = 1$ and $s = 0.25$ before meeting the discretisation floor (from 0.8% of the window, 0.0075 in v, at our operating scale). The expansion predicts order s^2; with two points clearly above the floor we can confirm the decay but not the exponent.

Corollary 1 inside the stability radius. Section 5.2 trains at $\rho = 0.03$, which is about eight times the crowding radius estimated in Appendix A, so the stable event set that Corollary 1 assumes holds for only 43% of draws there. Repeating six of the twelve nominals at $\rho = 0.01$ and $\rho = 0.003$ puts the deviation inside that radius (the event count survives for 68% and 56% of draws). Agreement improves: *all* 42 directions fall within a factor of two at both sizes, against 83% at $\rho = 0.03$, with Pearson correlations of 0.97 and 0.84 on log variance. The rank correlation is weaker (0.68 and 0.44 against 0.85), which is expected rather than contradictory: as ρ shrinks the identified directions are shrunk less and their measured variances bunch towards the prior level, so the ordering is harder to resolve even though every value is predicted well. The prediction therefore holds better, not worse, where its hypothesis is satisfied.

G ADDITIONAL RESULTS

Observation shift. Table 3 lists every condition. The rank-normalised pixel model computes its ranks over the valid image region (the part of the frame that sees the rendered screen window), the same side information the detector receives; with ranks over the whole frame, the zero padding introduced by a warp shifts every rank, and the errors under tilt with lens distortion and under zoom rise to 0.711 and 0.791 (column "pix. rank (frame)"). The combined ladder mixes radiometric and geometric shifts. For the log-input model its error at severity 4 (0.476) is lower than under its radiometric part (0.635) or its geometric part (0.592) alone, an interaction of the zero padding with the network's normalisation layers rather than robustness, so the main text draws conclusions from single nuisances.

Capacity and data. Table 4 varies the encoder width and the number of training reliefs for the global regime of Table 1, always on the same 500 test reliefs; the additional reliefs are drawn from the same prior with independent random streams.

Same-subset comparison. The optimisation baseline is expensive, so it is run on the first 20 test reliefs. Table 5 evaluates every method on exactly those 20 reliefs. Paired bootstrap intervals over the 20 reliefs (10,000 resamples) give, against the fixed-camera baseline: image spectrum $+0.238$ $[+0.115, +0.395]$ on clean frames and -0.346 $[-0.406, -0.291]$ under the $0.8\times$ zoom; augmented

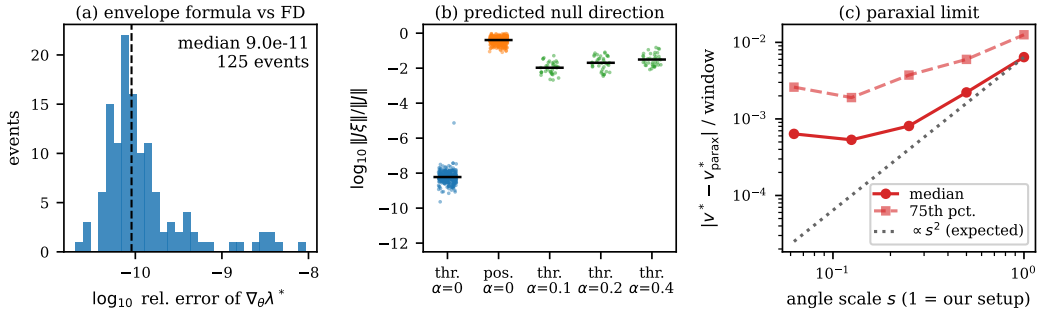

Figure 4: **The theory holds to numerical precision.** (a) Relative error of envelope-formula threshold gradients against central finite differences (125 events, 20 reliefs). (b) Weight of the rotation generator in the event Jacobian: threshold rows of axial sweeps annihilate it (float32 floor), position rows do not, and tilting the beam by α lifts it slightly. (c) Deviation of exact-optics thresholds from the curvature law of Proposition 2 when every ray angle is scaled by s; it falls toward the discretisation floor, with the expected s^2 slope shown for reference.

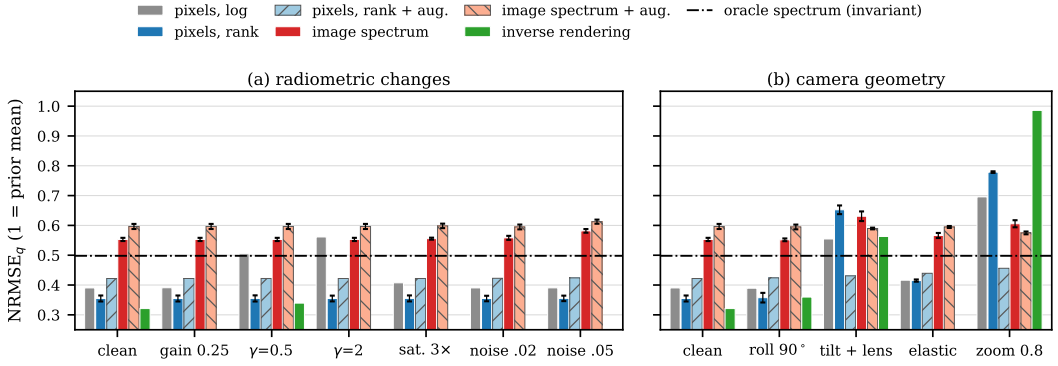

Figure 5: **Observation shift** (refraction, $d = 7$, 500 test reliefs, analysis-by-synthesis on 20; error bars: s.d. over seeds). Models are trained on clean renders unless hatched (augmented) and tested under (a) radiometric changes and (b) camera geometry. The oracle spectrum (dash-dotted) is exactly invariant by Proposition 1; the detector is not the spectrum, and inherits that invariance only in part.

pixels -0.535 $[-0.633, -0.442]$ under the zoom and -0.151 $[-0.243, -0.051]$ under tilt with lens distortion. The advantage of the calibrated optimisation on clean frames and its collapse under an unmodelled camera are both significant; so is the spectrum's advantage in that regime.

Self-calibrating optimisation. The fixed-camera baseline collapses under an unmodelled camera by construction, so we also fit three camera unknowns (log-zoom, tilt, one radial coefficient) jointly with the shape through a differentiable warp, at the same budget (Appendix B). It recovers them accurately: under a 20° tilt with $k_1 = 0.1$ it returns 0.348 rad and 0.102, and under the $0.8\times$ zoom a log-zoom of -0.207 against the true -0.223. An unmodelled camera is therefore a modelling choice rather than a limit of analysis-by-synthesis, and the fixed-camera row of Table 5 should be read as the cost of not modelling it. We report the variant's shape errors in Table 5 but draw no further conclusion from them, because they contain an anomaly we cannot explain: on clean frames it reaches 0.293, close to the fixed-camera 0.321, yet under both geometric conditions it reaches 0.039 and 0.049, six times better than its own clean result. A plausible cause is that resampling smooths the observation and the render alike and so conditions the optimisation, which would be a property of the fitting procedure rather than of the camera model; we did not pursue it, as the variant is a baseline and no claim in this paper rests on it.

Table 3: **Observation shift**: NRMSE_q of models trained on clean renders ("aug.": trained with random nuisances) on the 500 test reliefs; analysis-by-synthesis ("inv. render.") on 20 of them. Means over the seeds given in the second row (largest s.d. over seeds 0.026). The oracle spectrum (0.498) is exactly invariant.

	pixels (CNN)				image spectrum		inverse
condition	log	rank (frame)	rank	rank+aug.	plain	+aug.	rendering
seeds	1	2	2	1	3	3	1
clean	0.390	0.355	0.355	0.422	0.553	0.596	0.321
gain 0.25, offset 0.05	0.391	0.355	0.355	0.422	0.553	0.596	–
$\gamma = 0.5$	0.505	0.355	0.355	0.422	0.553	0.596	0.339
$\gamma = 2$	0.561	0.355	0.355	0.422	0.553	0.596	–
saturation at $3\times$ flat	0.407	0.355	0.355	0.422	0.556	0.598	–
noise $\sigma = 0.02$	0.390	0.355	0.355	0.423	0.558	0.595	–
noise $\sigma = 0.05$	0.390	0.355	0.355	0.424	0.582	0.612	–
roll 90°	0.389	0.358	0.358	0.424	0.552	0.595	0.360
tilt 20°, $k_1 = 0.1$	0.554	0.711	0.652	0.431	0.631	0.590	0.563
elastic $e = 0.03$	0.416	0.416	0.415	0.439	0.566	0.595	–
zoom 0.8	0.696	0.791	0.779	0.456	0.605	0.575	0.986
combined, severity 1	0.449	0.591	0.550	0.421	0.575	0.568	–
combined, severity 2	0.475	0.723	0.630	0.422	0.606	0.565	–
combined, severity 3	0.470	0.758	0.670	0.424	0.627	0.576	–
combined, severity 4	0.476	0.777	0.690	0.429	0.656	0.584	–
combined, severity 5	0.589	0.752	0.694	0.445	0.701	0.611	–

Table 4: **Capacity and data controls** for the global regime (refraction, $d = 7$, oracle spectra, NRMSE_q on the same 500 test reliefs; one seed unless stated). Width rows use 2000 training reliefs; the last column adds screen positions to thresholds and types.

setting	thresholds + types	types only	+ positions
width 96	0.499	0.554	0.490
width 192 (default; 3 seeds)	0.498 ± 0.002	0.555 ± 0.003	0.501 ± 0.003
width 384	0.498	0.560	0.505
500 training reliefs	0.532	0.583	0.553
1000 training reliefs	0.509	0.573	0.520
2000 training reliefs (3 seeds)	0.498 ± 0.002	0.555 ± 0.003	0.501 ± 0.003
3500 training reliefs	0.489	0.550	0.460
5000 training reliefs	0.473	0.548	0.440

H FALSE POSITIVES AND THE ADAPTIVE GAIN

Section 5.4 attributes the loss of the adaptive acquisition gain to the detector's false positives. To isolate that cause we keep the idealized signature and inject spurious transitions: for a relief with K events we draw $\lceil rK \rceil$ control values uniformly in the sweep, at least 0.002 away from any true event, and add one unit to the first component count at each of them. Every spurious location therefore creates one extra transition, the genuine events are untouched, and the precision of the signature the policy reads is $1/(1 + r)$. Budget, coarse pass, matching tolerance and the 40 reliefs are those of Figure 3.

Figure 6 gives the result. Seeking degrades steeply with precision while the uniform sweep is almost unaffected, as it must be: a uniform sweep spends the same frames wherever the transitions are, whereas bisection spends its budget on whichever brackets it is given. The gain is 18.3, 4.9, 1.0, 1.1, 0.8, 0.8 and 0.8 at precisions 1, 0.80, 0.67, 0.50, 0.33, 0.20 and 0.11. Two consequences. First, a detector does not need to be perfect for the policy to pay: precision above roughly 0.8 already recovers a factor of five. Second, matched recall is a misleading quality measure in this regime, since it climbs from 0.68 to 0.99 as precision falls, purely because denser spurious brackets leave an estimate near every true event.

Table 5: **Same 20 test reliefs for every method** (NRMSE_q). The oracle spectrum is unaffected by image nuisances by Proposition 1, so its clean value applies to every column. The self-calibrating variant models zoom, tilt and radial distortion only, so it is run on the conditions those parameters can absorb.

method	clean	$\gamma = 0.5$	roll 90°	tilt + lens	zoom 0.8
analysis-by-synthesis (fixed camera)	0.321	0.339	0.360	0.563	0.986
analysis-by-synthesis (self-calibrating)	0.293	–	–	0.049	0.039
pixels, rank-normalised	0.370	0.370	0.296	0.729	0.802
pixels, rank + augmentation	0.406	0.407	0.395	0.412	0.451
image spectrum	0.559	0.559	0.572	0.677	0.639
oracle spectrum (invariant)	0.514	0.514	0.514	0.514	0.514

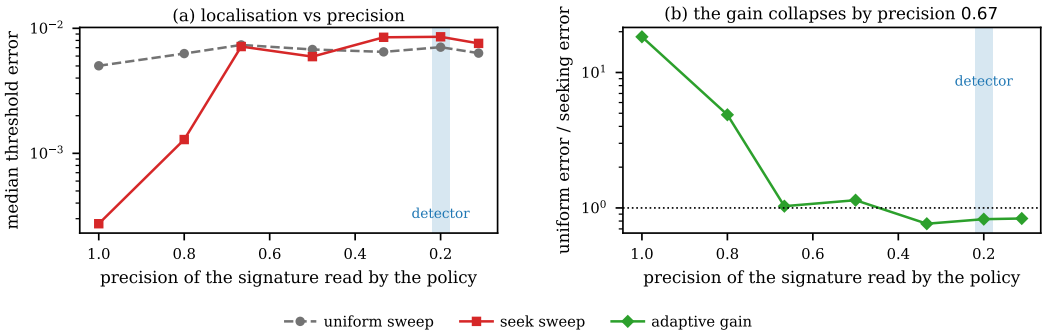

Figure 6: **Precision alone accounts for the end-to-end collapse.** Spurious transitions are injected into the idealized signature at a controlled density; everything else is held fixed. (a) Median threshold error at $B = 48$ for the uniform sweep and for seeking. (b) Their ratio, the adaptive gain. The shaded band is the precision the image detector actually achieves (0.18 to 0.22, Section 5.4); the gain has already vanished well above it.

I DEGENERATE EVENTS: UMBILICS

Paraxially $g \propto \det(c(v)\boldsymbol{I} + \nabla^2 h)$. At an umbilic point $\boldsymbol{u}_0$ of the relief the two principal curvatures coincide, $\nabla^2 h(\boldsymbol{u}_0) = \kappa^* \boldsymbol{I}$, so at the control value with $c(v^*) = -\kappa^*$ the matrix $c\boldsymbol{I} + \nabla^2 h$ vanishes at $\boldsymbol{u}_0$. Then $g = 0$ and $\nabla_{\boldsymbol{u}} g = 0$ (the adjugate of the zero matrix is zero), and also

$$\partial_v g = c'(v)\ \mathrm{tr}\,\mathrm{adj}\big(c\boldsymbol{I} + \nabla^2 h\big) = 0,$$

so an umbilic solves the event equations but is *degenerate* in the sense of Definition 1: Lemma 1 and Theorem 1 do not apply to it. Writing $c\boldsymbol{I} + \nabla^2 h = \epsilon \boldsymbol{I} + A(\delta \boldsymbol{u})$ near $\boldsymbol{u}_0$, with $\epsilon = c(v) - c(v^*)$ and A the linearisation of the Hessian, gives $g \propto (\epsilon + s)^2 - |D\delta\boldsymbol{u}|^2$ with $s = \langle \sigma, \delta\boldsymbol{u} \rangle$ linear and D the (generically invertible) linear map formed by the deviatoric part. The critical value of g near $\boldsymbol{u}_0$ is $\epsilon^2/(1-\alpha)$ with $\alpha = \langle \sigma, (D^\top D)^{-1}\sigma \rangle$, provided $\alpha \neq 1$: it touches zero quadratically at $\epsilon = 0$ instead of crossing it, so the sublevel sets on either side are diffeomorphic and the signature does not jump. The excluded case $\alpha = 1$ is where the quadratic term degenerates as well; it is a codimension-one condition on the relief, our prior meets it with probability zero, and the census below found no instance of it. Umbilic metamorphoses of the caustic are therefore invisible to a signature-based measurement, which is why Definition 1 keeps only nondegenerate events.

A detector on a finite grid sees something else. The vanishing component is lost once it is smaller than a pixel and returns afterwards, so a sweep reports a *cancelling pair* of spurious transitions around v^* whose separation is set by the resolution. On the exactly solvable monkey saddle $h = (x^3 - 3xy^2)/6$, for which $g = c^2 - r^2$ and the vanishing component is a disk of radius $|c|$, a sweep reports exactly two transitions, separated by 1.41 pixels at every resolution we tried (grids of 80 to 320 points, so a width of 0.036 down to 0.0089 in c): the pair closes in proportion to the pixel size and is therefore an artifact of the discretisation, not a transition.

Umbilics are not rare under our prior. Over 120 reliefs there are 5.9 umbilic points per relief (median 6), of which 1.4 have their threshold inside the swept window (median 1), and 72% of reliefs have at least one. Sweeping the paraxial fold function finely around those thresholds, a cancelling pair of transitions appears for 16 of the 41 we swept (39%, at a 121^2 grid, median width 0.003 in v); the remainder fall within 0.02 of a genuine event, which masks the pair. They are one source of the near-coincident pairs of recorded events that make the crowded regime of Section 5.4 denser than the relief geometry alone would.

J CLAIM LEDGER AND ASSUMPTIONS

Table 6 lists every headline claim, the scope in which it is asserted, the statement that supports it, the experiment that tests it, and the regime in which it is known to fail; Table 7 lists the assumptions those scopes rest on, what each means in words, and how each is checked against our own data.

Table 6: **Claim ledger.** The last column is measured in this paper rather than left open.

claim	scope	statement	evidence	where it fails
events are the topology changes	isolated, nondegenerate, $g > 0$ near the rim	Lemma 1	Figure 1; 97% of signature changes match a refined event (Section 5.4)	simultaneous events add and can cancel; degenerate events are invisible (Appendix I)
thresholds move with a closed-form differential	nondegenerate event	Theorem 1	Figure 4(a)	near-degenerate or crowded events shrink the neighbourhood (Figure 2c)
full quotient rank certifies identifiability up to the sweep symmetry	threshold Jacobian of maximal quotient rank $d-1$ at $\boldsymbol{\theta}_0$, reached by 50% of reliefs	Theorem 2(b)	rank census, Section 5.1; blind directions measured in Figure 2(b)	a rank drop is necessary for local non-injectivity but not sufficient; no experiment here tests injectivity itself
full rank certifies local identifiability outright	full event-Jacobian rank at $\boldsymbol{\theta}_0$, which needs screen positions and is reached by 85% of reliefs	Theorem 2(a)	rank census with positions, Section 5.1	thresholds alone can never satisfy it, since $J_\lambda \xi = 0$; untested as an injectivity claim
blind directions from acquisition symmetry	axial sweep, radial window, thresholds only	Theorem 2(d)	Figure 4(b)	tilting the beam lifts the degeneracy weakly (same panel)
singular values order the learned error	local fixed-event regime, unimodal posterior, squared loss	Corollary 1	Figure 2(b)	across the prior the ordering fails and spectra collide at the acquisition resolution (Section 5.3)
paraxial curvature law	small ray angles, away from umbilics	Proposition 2	Figure 4(c)	exact optics leaves a residual that vanishes with the angle scale
adaptive localisation is exponential	events separated by more than the coarse spacing, read from a signature of events separated by more than the coarse spacing	Proposition 3(a)	Figure 3	closer events share a bracket and can cancel; and below a signature precision of 0.67 the gain is gone, which is empirical rather than part of the proposition (Figure 6)
invariance to radiometry and camera	continuum spectrum	Proposition 1	exact for the oracle spectrum by construction; not separately measurable	the image spectrum's invariance to gain and γ comes from rank normalisation, which the pixel baseline shares (Table 3); clipping and noise fall outside the proposition

Table 7: **Assumptions, in plain words, and how each is checked against our own data.** The last column is the reason the scope statements in Table 6 read as they do.

Assumption	Plain meaning	Used in	Checked here
$\det \nabla^2_{\boldsymbol{u}} g \neq 0$ and $\partial_\lambda g \neq 0$ at an event	the fold function has a Morse critical point there, and its critical value crosses zero at nonzero rate	Lemma 1, Theorem 1	generic for one-parameter families, but umbilics violate it systematically: 1.4 per relief fall in our window and leave β unchanged (Appendix I)
no two events share a control value	each observed jump of the signature comes from one geometric event	Lemma 1, Proposition 3(a)	false for most events here: 53% of neighbouring thresholds lie within one frame spacing, the crowding radius is $\|\delta\boldsymbol{\theta}\| \approx 0.01$ (Appendix A), and Section 5.4 measures what that costs
$g > 0$ near the rim of the relief	the fold set stays inside the illuminated patch, so its components can be counted	Lemma 1	enforced by the radial window: $g = 1$ on the flat surround (Appendix B)
$\operatorname{rank} J(\boldsymbol{\theta}_0) = d$	the event coordinates move independently under every shape deformation	Theorem 2(a)	sufficient but not typical: over 300 theory reliefs the threshold Jacobian reaches the maximal quotient rank 6 for 50% of them (85% reach full rank 7 with screen positions), and Figure 2(b) reports the blind directions of the rest
unimodal posterior, linearised forward map, squared loss	one shape explains the spectrum, and the trained network approximates the posterior mean	Corollary 1	holds locally (Figure 2b), fails across the prior: 21 relief pairs share a spectrum to one oracle step yet differ by 0.63 in shape (Section 5.3)
σ_ϵ is a single effective noise scale	the network's approximation error is the same in every direction, in threshold units	Corollary 1	not assumed but calibrated out of sample: leave-one-nominal-out gives 83% of 84 directions within a factor of two (Section 5.2)
small ray angles	the paraxial thin-element model approximates exact refraction	Proposition 2	scaling every ray angle by s shrinks the deviation by a factor of 8 from $s = 1$ to $s = 0.25$ (Figure 4c)
the sweep reads β itself	the policy sees the fold-set signature, not a proxy for it	Proposition 3, Section 5.4	false: the detector reads quantile masks instead, at precision 0.18 to 0.22, and Appendix H shows that this alone removes the adaptive gain
ψ a diffeomorphism, ϕ strictly increasing	the camera may move and re-expose, but may not clip or fold the image	Proposition 1	clipping at saturation is not strictly monotone and costs the image spectrum 0.003 (Section 5.5)
exact vector refraction, a source of zero angular extent, no wave optics	geometric optics with a point-like source describes the measurement	everywhere	partly checked only: a 0.1 rad source tilt moves thresholds by a median of 0.007, twenty-six times the seeking precision (Section 6); diffraction, dispersion, scattering and interreflection are not modelled