Intrinsic Pixel Complexity: A Mechanism-Grounded Ground Truth for Per-Pixel Rendering Difficulty

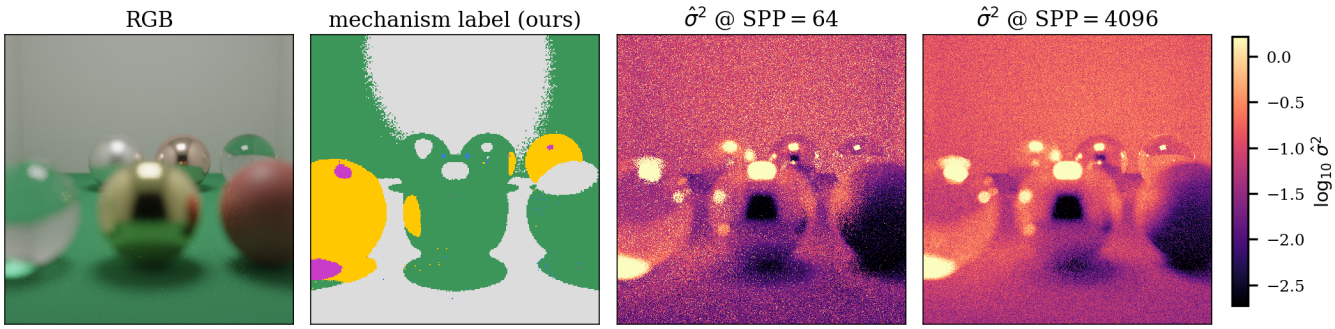

Figure 1: *Our discrete mechanism label (second panel) is determined at SPP= 64 and barely changes at SPP= 4096, with the residual disagreement concentrated at the boundary between two light-tailed buckets (§4). The classical variance signal (right two panels) drifts substantially across the same 64× budget change.* Variance-based per-pixel difficulty is not a stable ground truth under heavy-tailed transport; path mechanism is.

Abstract

Per-pixel rendering difficulty has no agreed-upon ground truth. The default across adaptive sampling, variance-guided denoising, and efficiency-aware path tracing is the sample variance $\hat{\sigma}^2(p)$ of a Monte Carlo estimator. We argue that this default is mathematically unsound under heavy-tailed light transport and propose to replace it with a discrete, mechanism-based label assigned to each contribution event in a path-traced integrator. The variance-based default has three defects: (i) the relative error of $\hat{\sigma}^2_N$ scales as $\sqrt{(\kappa-1)/N}$ where κ is the integrand's kurtosis, which is large or infinite on caustic, near-specular, and firefly-prone pixels; (ii) empirically, the rank correlation of $\hat{\sigma}^2$ between SPP=64 and SPP=4096 drops to $\rho = 0.26$ on caustic-dominated scenes; (iii) any feature predictor evaluated against a $\hat{\sigma}^2$-derived ground truth is upper-bounded by a $\rho = 0.23$–0.29 noise floor. Our mechanism representation partitions contribution events into seven exhaustive, mutually exclusive buckets by (end-vertex BSDF lobe, presence of delta-S in the path sequence). We show: (a) the partition closes with 0% of events falling outside the seven buckets across nine (scene × SPP) runs; (b) discrete labels are SPP-stable where $\hat{\sigma}^2$ is not, with cross-SPP mismatches concentrated at the boundary between two light-tailed buckets and structural-bucket pixel shares drifting by less than 1.2%; (c) the labels are robust not only across an order-of-magnitude SPP change but across estimator completion: restoring the MIS half of the estimator leaves the per-pixel argmax unchanged on 99.9–100% of pixels, with structural buckets immune by construction and only energy-argmax boundaries flickering; (d) a cross-scene correlation analysis yields two falsifiable structural findings — the caustic–glossy correlation reverses sign with object-space separation (shown on a controlled scene built to falsify it), and the depth–purity correlation reverses sign with material topology — and a controlled roughness sweep shows that end-vertex lobe width monotonically governs the firefly intensity of the glossy bucket, the regime in which continuous cost saturates. On a representative scene 51.8% of the (NEE-defined) glossy bucket is unconverged at 2048 SPP; we show this rate is bucket-definition-dependent and that the discrete label keeps discriminating where continuous cost has gone flat.

CCS Concepts

• ***Computing methodologies*** → ***Ray tracing;*** *Rendering; Visibility;*

1. Introduction

Per-pixel rendering difficulty has no agreed-upon ground truth (GT). For thirty-five years the operational signal in adaptive Monte Carlo rendering has been one scalar — the per-pixel sample variance $\hat{\sigma}^2(p)$ of an unbiased pixel estimator — which adaptive samplers refine on, denoisers steer their filter footprint by, efficiency-aware Russian roulette branches off, and inverse-rendering pipelines now differentiate through. The default has hardened into a methodological assumption: *the per-pixel sample variance, estimated from a pilot render, is a reliable measurement of how hard each pixel is to render*. We argue this is mathematically unsound under heavy-tailed light transport. The defect is not that $\hat{\sigma}^2$ is a little noisy; it is that no quantity built on a finite-sample variance estimator can serve as a difficulty reference at the budgets the community actually uses.

1.0.0.1. The variance-based ground truth and its inheritors. Adaptive samplers since Mitchell [Mit87] refine where a per-pixel contrast/variance threshold is exceeded, or in the modern formulation distribute samples to greedily minimise per-pixel mean-squared error [RKZ11]; real-time denoising steers a spatio-temporal filter footprint by a per-pixel variance buffer [SKW*17]; efficiency-aware Russian roulette iteratively estimates per-path second moments [RGH*22]; Bayesian direct illumination drives light-selection probabilities from online regression over per-light statistics [VKK18]; and the Eurographics State-of-the-Art Report [ZJL*15] surveys variance-driven allocation as the dominant family. We do not dispute these methods' utility; we dispute that the quantity they all rely on can serve as a ground truth.

1.0.0.2. Problem 1: heavy-tail instability. $\hat{\sigma}^2$ is a sample-variance estimator, and such estimators are unstable on heavy-tailed integrands. The instability shows in the simplest measurement: rank-correlate per-pixel $\hat{\sigma}^2$ at a low pilot SPP against the same quantity at a much higher SPP. On caustic-dominated scenes we measure Spearman $\rho(\hat{\sigma}^2_{64}, \hat{\sigma}^2_{4096})$ as low as 0.26 (§4). Prior work documents the same instability under the proxy of firefly outliers — density-based rejection [DWR10], sample re-weighting [ZHD18] — but treats it as a removable nuisance at the *mean* estimator rather than a foundational defect of variance as a ground-truth signal.

1.0.0.3. Problem 2: kurtosis regress. The instability is a mathematical necessity, not an implementation defect. The relative error of $\hat{\sigma}^2_N$ is governed by the integrand's fourth standardised moment: $\sqrt{\mathrm{Var}(\hat{\sigma}^2_N)}/\sigma^2 \approx \sqrt{(\kappa-1)/N}$ [CB02, vdV98]. On caustic, near-specular and firefly-prone integrands κ is large and may be infinite — the regime where sample variance does not converge to a finite-rate Gaussian limit [EKM97]. The natural workaround — estimate κ from a pilot, compute an error bar, budget accordingly — fails by infinite regress: estimating κ needs the eighth moment, which converges still more slowly than $\hat{\sigma}^2$. The rendering literature has long known this pathology under other names ("weak singularities" [KK06], unbounded contribution from delta paths [Vea97], infinite Hardy–Krause variation [Owe06]), but the implications for $\hat{\sigma}^2$ *as a ground-truth signal* have not been examined. To our knowledge, no rendering paper cites the $\sqrt{(\kappa-1)/N}$ bound or quantifies the resulting estimator-of-variance instability on caustic pixels.

1.0.0.4. Problem 3: noise floor on the GT itself. Any ground truth derived from a finite-sample MC estimator inherits an unmeasured noise floor. Concretely: we fit a per-pixel convergence exponent $\hat{\alpha}(p)$ from a four-level SPP sweep with ten seeds per level ($\sim$40,000 samples per pixel), then split the seeds in half and refit on each. The Spearman correlation of the two halves — 1.0 for a noiseless GT — is only 0.23–0.29 on diffuse scenes (§4). *Any* cheap-feature predictor evaluated against a $\hat{\sigma}^2$- or $\hat{\alpha}$-derived ground truth is upper-bounded by this floor, however good the predictor. Recent deep adaptive-sampling work acknowledges that variance and contrast cannot be reliably computed at low sample counts [KKR18], but treats this as a problem of better estimation rather than a defect of the signal itself.

1.0.0.5. Problem 4: structural property versus ground-truth signal. One might object that $\sigma^2(x,y)$ is, after all, a stable structural property of the scene — a recent line treats it as a differentiable functional of scene parameters [YPD*24], and denoising-aware adaptive sampling proposes ever more sophisticated variance estimators on the premise that σ^2 is a stable per-pixel property worth estimating accurately [FFJ23]. The objection needs a careful answer, because there is a sense in which it is correct and a sense in which it is not, and the two are easy to conflate.

What is stable. Where σ^2_∞ exists, the σ^2-field is a deterministic functional of the integrand, and the estimator obeys $\mathbb{E}[\hat{\sigma}^2_N] = \sigma^2_\infty/N$ — a global rescaling. Correlations of $\hat{\sigma}^2$ buffers across budgets are invariant to this rescaling and so measure only the field's relative *shape*, which is well-defined and stable where σ^2_∞ is finite. Cross-SPP experiments reporting near-unity values in this regime are not in error.

What is not stable. The absolute *level* of $\hat{\sigma}^2$ at a pixel, and its per-pixel rank in a regime that includes heavy-tailed pixels, are governed by Problems 1–3. Problem 2's $\sqrt{(\kappa-1)/N}$ bound concerns the variance *of* the estimator, not of the field; Problem 1's $\rho = 0.26$ on caustic-dominated scenes measures a rank instability that survives the $1/N$ rescaling, because heavy-tailed pixels do not contract uniformly with N. This matters because a difficulty ground truth must rank pixels consistently, not merely reproduce field shape on the well-behaved subset.

Hence the narrow form of our objection: ground truth for per-pixel difficulty must be well-defined and stable *across* the entire scene, including the heavy-tailed pixels where σ^2_∞ either fails to exist (Problem 2) or is recoverable only through estimators whose own variance is set by the fourth moment. The σ^2-field shape on the light-tailed subset is preserved as budget grows; it says nothing about the regime where $\hat{\sigma}^2$ is least informative — precisely where the hardest pixels live.

1.0.0.6. Our position. The community is not missing a better way to *estimate* per-pixel variance; it is missing the right object to take as ground truth. The variance *estimate* we obtain depends on the estimator and the sampler in use; per-pixel difficulty, to serve as a ground truth across regimes, should be defined as a property of

the integrand. We therefore propose to build the GT on a quantity that is (i) a deterministic function of the light-transport integrand, (ii) discrete — so that the noise floor on continuous estimators does not propagate into it — and (iii) cross-scene invariant. The candidate we develop is the *transport mechanism of each contribution event*: a small label set defined by the BSDF lobe at the contribution's end-vertex and whether the path sequence to that vertex passes through a delta (specular) interface. The seven labels that result from this two-axis partition (Section 3) are exhaustive and mutually exclusive by construction, are assigned per contribution event from quantities the path tracer already maintains, and are — empirically — stable across sample budgets where $\hat{\sigma}^2$ is not (Fig. 1).

1.0.0.7. Contributions. This paper makes four contributions.

1. **A methodological argument** that variance-based per-pixel ground truth is unsound under heavy-tailed transport, supported by the kurtosis-regress bound (Problem 2), the empirical heavy-tail instability (Problem 1), and the noise-floor measurement (Problem 3) above (§1).
2. **A mechanism-grounded GT** consisting of seven exhaustive, mutually exclusive labels assigned per contribution event, derived from a two-axis partition (end-vertex lobe, presence of delta-S) with no continuous threshold parameter on the main axis (§3).
3. **Two-pronged validation:** (a) exhaustiveness — the partition closes with 0% of contribution events falling outside the seven labels across nine (scene×SPP) runs; (b) cross-SPP stability — discrete labels agree at 89.5%–99.6% between SPP=64 and SPP=4096, with 67%–82% of disagreements concentrated at the boundary between the two light-tailed buckets (`direct` ↔ `diffuse-indirect`) and structural-bucket pixel shares drifting by less than 1.2%. We further show the labels are robust to a second, independent perturbation — completing the estimator with MIS leaves the argmax unchanged on 99.9–100% of pixels, with structural buckets immune by construction (§4).
4. **A cross-scene correlation analysis of the mechanism-derived difficulty vector** on five scenes, yielding two falsifiable structural findings: (a) the correlation between caustic and glossy contributions reverses sign monotonically with object-space separation between specular and glossy materials, demonstrated on a controlled scene triple including a scene constructed specifically to falsify the prediction (§5.1); and (b) the correlation between bounce depth and mechanism purity reverses sign with the relative abundance of specular-chain versus glass-transmission paths (§5.2). A controlled roughness sweep (§5.4) shows end-vertex lobe width monotonically controls the firefly intensity of the glossy bucket — the regime in which the time-to-quality analysis (§6) finds continuous cost saturating (51.8% of the NEE-defined glossy bucket unconverged at 2048 SPP on a representative scene, a rate we show to be bucket-definition-dependent) while the discrete labels still discriminate.

1.0.0.8. Roadmap. Section 2 surveys the variance-as-GT camp, the covariance / frequency a-priori prediction camp, and prior uses of path mechanism inside (rather than as the target of) Monte Carlo estimators. Section 3 develops the mechanism GT. Section 4 reports the exhaustiveness and cross-SPP stability results. Section 5 reports the cross-scene correlation structure of the difficulty vector and its two sign-reversal findings. Section 6 closes with the time-to-quality saturation argument and a per-dimension analysis of how the difficulty vector relates to convergence cost.

2. Related Work

The methodological context for our claim falls into three camps, each producing or relying on a per-pixel quantity that has been treated — implicitly or explicitly — as a difficulty signal. The first uses a sample-statistic-based estimate of variance; the second traces an analytically propagated continuous descriptor of the radiance signal; the third already uses path-transport mechanism, but exclusively inside the estimator or for compositing. We discuss each in turn.

2.1. Variance as the de facto difficulty signal

Three and a half decades of adaptive Monte Carlo rendering use $\hat{\sigma}^2(p)$ as the operational signal: image-space adaptive sampling refines where contrast or variance exceeds a threshold [Mit87] or minimises per-pixel MSE [RKZ11]; variance-guided denoising builds its filter footprint from a variance buffer [SKW*17]; efficiency-aware path tracing iterates over second-moment estimates [RGH*22]; Bayesian direct illumination regresses over per-light statistics [VKK18]; and the Eurographics STAR [ZJL*15] surveys this family as the dominant paradigm. Recent inverse-rendering work treats $\sigma^2(x,y)$ as a differentiable functional [YPD*24], and denoising-aware adaptive sampling refines variance estimators on the premise that σ^2 is a stable property worth estimating accurately [FFJ23].

2.1.0.1. Heavy-tailed instability is known but not foundationalised. The community has noticed that $\hat{\sigma}^2$ behaves badly under firefly-prone transport: density-based outlier rejection [DWR10] and firefly re-weighting [ZHD18] both target the fingerprint of an unstable estimator, and deep adaptive sampling states that variance and contrast cannot be reliably computed at low sample counts [KKR18]. In each case the instability is suppressed at the *mean* estimator, and its implications for $\hat{\sigma}^2$ *as a ground-truth signal* remain unexamined.

2.1.0.2. Our distinction. All these methods build on $\hat{\sigma}^2$ as an estimated continuous quantity and inherit its instability. We replace it with a deterministic discrete label, read from path structure rather than estimated from samples.

2.2. A-priori prediction via traced continuous descriptors

A second line — closer in spirit, forgoing sample variance to predict difficulty analytically — traces a continuous descriptor of the radiance signal through the transport operators. The frequency-analysis framework [DHS*05] treats radiance as a band-limited signal in space–angle frequency, occlusion as Fourier convolution by the blocker spectrum, free-space propagation as a shear. Later

)

work specialised it to depth-of-field sampling via local lens bandwidth [SSD*09] and generalised it to a 5×5 covariance matrix propagated through transport, BRDF, occlusion, lens, and motion [BSS*13] — the closest prior work to ours by goal, both targeting an a-priori per-pixel prediction without sample variance. A non-Fourier antecedent expresses the same philosophy through first-order sensitivities of lighting, shading, and shadow [RMB07].

2.2.0.1. Our distinction. The object of prediction here — a covariance matrix, spectrum, or gradient field — is itself continuous: it must be traced, propagated under Gaussian-family assumptions, and approximated as it accumulates. Our mechanism label is a deterministic function of path structure, read from a single contribution event, not traced or estimated.

2.3. Path mechanism as an estimator-internal construct

Path-transport mechanism is not new in light-transport research; it has been load-bearing for decades, but only inside the estimator or as a compositing convenience. Heckbert's `L(D|S)*E` regex [Hec90] described paths and motivated a hybrid algorithm tied to particular path classes (the glossy state `G` was a later production addition); Veach [Vea97] formalised the regex as a classifier of which algorithms can sample which configurations. Photon mapping separates a caustic from a global map for density estimation [Jen96] — an estimator structure, not a per-pixel label. Path guiding learns incident-radiance proposals from a particle stream, as a parametric mixture [VKŠ*14] or an SD-tree [MGN17]; the neural variant trains a normalising-flow sampler on KL or χ^2 divergence [MMR*19] — a loss internal to sampler training, driving a sampler toward the target density, the opposite direction from our use of a mechanism label as the predicted target. Specular-manifold samplers target the SDS mechanism to render refractive caustics [HDF15] or unify caustics and glints [ZGJ20]; there the mechanism is what the sampler targets, not how the pixel is labelled. Variance-aware MIS injects variance into the balance heuristic [GGSK19] — the closest prior work combining mechanism and variance, though at sampler-family granularity and for sampler weighting, not per-pixel difficulty.

2.3.0.1. Production light-path expressions. The closest analog to our seven-bucket taxonomy is the light-path-expression facility of production renderers [Pix24]: a regex over `L/D/G/S/E` selecting events into named AOVs. But the expressions are user-defined, not a closed exhaustive taxonomy; framed as compositing conveniences, not a difficulty signal; and overlap is permitted, so the exhaustive-and-mutually-exclusive property our taxonomy rests on (§3) is not part of their design.

2.3.0.2. Our distinction. The path-mechanism abstraction has appeared in every role — Heckbert's notation, Veach's classifier, path-guiding proposals, manifold samplers, production light-path expressions — except as the supervision target for a per-pixel difficulty ground truth. That is the role we propose.

3. Mechanism-Grounded Ground Truth

We replace the per-pixel sample variance $\hat{\sigma}^2(p)$ with a discrete label assigned per contribution event. The label is a deterministic function of the lobe sampled at the event's end-vertex and the lobe sequence leading to it; nothing is estimated, and nothing on the main axis is thresholded. We give the construction in three pieces — the unit of analysis (§3.1), the two-axis partition into seven buckets (§3.2), and the continuous sidefields that retain the per-pixel mechanism mixture (§3.3) — and then address in §3.4 three design choices (discreteness; continuous rather than boolean sidefields; and glossy as an independent bucket) that anticipate reviewer objection.

3.1. The contribution event as the unit of analysis

A path-traced estimator at pixel p accumulates radiance from many sources, but each contribution to the pixel value reaches the sensor through a specific sequence of vertices and BSDF events. We call each such accumulation event a *contribution event*. In a standard path tracer with next-event estimation, contribution events arise in three forms:

- a successful NEE shadow ray connecting the camera-traced subpath to an emitter,
- a BSDF-sampled bounce that lands directly on an emitter, or
- the trivial case in which the camera primary ray intersects an emitter with no intervening surface bounce.

For each contribution event we observe two quantities: (a) the BSDF lobe at the *end-vertex* V_n, the last surface vertex before the emitter, and (b) whether the light-to-end-vertex sequence contains any delta-specular interaction. Both are computed from lobe types already sampled during path construction; no additional rays are cast.

3.1.0.1. End-vertex lobe under NEE. The two NEE-path-tracer branches need different end-vertex rules, one requiring a convention. For a BSDF-sampled bounce landing on an emitter, V_n is the last surface vertex before the emitter, and its lobe is read from the BSDF's reported sampled lobe. For an NEE shadow-ray connection, V_n is the same vertex but the BSDF is only *evaluated*, not sampled, so we assign V_n a lobe from the BSDF's flags: **D** if it advertises a diffuse component, **G** otherwise. Multi-lobe BSDFs advertising both (e.g. `roughplastic`) are labelled **D**, conservative in the sense of §3.2: it routes the event to the light-tailed side rather than the long-tailed glossy bucket. Delta-specular surfaces do not participate in NEE (zero connection probability), so an NEE end-vertex is never **S**; consequently the `specular-direct` bucket (Table 1) receives only BSDF-sampling contributions.

The per-pixel ground truth is the energy-weighted argmax over the contribution-event labels arriving at that pixel: each event is classified individually, and the pixel takes the bucket with the largest energy share.

This differs from prior uses of path-mechanism information (§2.3) in two ways. First, the unit is a single contribution event, not a whole path: one path-tracer step may contribute through both NEE and BSDF branches, which we label separately. Second, the label is the supervision target of a ground truth — not consumed inside an estimator (photon-map split, path-guiding proposal, manifold-sampler target) nor user-defined for compositing (light-path expression).

3.2. The two-axis partition

Each contribution event maps into one of seven buckets through a partition along two axes (Table 1). The first axis is the end-vertex BSDF lobe, which takes one of three values: **D** (diffuse), **G** (glossy, non-delta), or **S** (delta specular). The second axis is binary — whether the light-to-end-vertex sequence contains any delta-S event — and a seventh row covers the trivial primary-ray-into-emitter case, disjoint from all surface-bounce cases.

Table 1: *The seven contribution-event buckets, defined by end-vertex BSDF lobe and presence of a delta-specular event in the light-to-end-vertex sequence.*

end-vertex lobe	δ-S in seq.	bounces	bucket
(no surface bounce)	—	$=0$	`emitter-direct`
D or G	no	$=1$	`direct`
D	no	>1	`diffuse-indirect`
G	no	>1	`glossy`
D or G	yes	any	`caustic`
S	—	—	`specular-direct`
other	—	—	`other`

The partition is exhaustive and mutually exclusive *by construction*: every contribution event has a unique (end-vertex lobe, δ-S presence) pair, and the trivial primary-ray case is disjoint from all surface-bounce cases. Validation in §4 confirms empirically that the `other` catch-all receives 0% of contribution-event energy across all nine (scene×SPP) runs of our test matrix.

The seven labels are not arbitrary categories: they enumerate a structural partition along two well-defined axes, the structure preceding the labels. The partition admits no continuous threshold on the main axis — every cell boundary is a discrete predicate on a lobe type or sequence property.

3.3. Purity and energy-fraction sidefields

The dominant-class field captures only the energy-argmax bucket. To retain the per-pixel mechanism mixture — and support the cross-scene analysis of §5 — we add two continuous sidefields:

- *purity* $\pi(p) \in [0,1]$: the energy fraction of the dominant bucket. $\pi = 1$ marks a mechanically pure pixel; lower values indicate mixing.
- *glossy energy fraction* $\gamma(p) \in [0,1]$: the energy share of events whose light-to-end-vertex sequence contains any glossy bounce, regardless of dominant class.

A derived mixed-flag $m(p) = [\pi(p) < \theta]$ ($\theta = 0.7$) marks low-purity pixels for downstream consumers wanting a binary signal. The threshold θ is post-hoc and does not enter dominant class: any continuous threshold in our framework is confined to sidefields, never the main partition.

3.4. Three design choices

3.4.0.1. Why discrete. A discrete label is robust to small pipeline perturbations. As §4 shows, dominant class agrees at 89.5%–99.6% between SPP= 64 pilot and SPP= 4096 reference renders, disagreement concentrated at the `direct`↔`diffuse-indirect` boundary and structural-bucket shares stable to within 1.2%. A continuous variance estimate carries no such robustness — any perturbation crossing the heavy-tail threshold reshapes the σ^2-field non-trivially (Problem 1, §1). Discreteness is what prevents the GT from inheriting the variance estimator's instability.

3.4.0.2. Why γ is continuous, not a boolean flag. A simpler design for the soft-caustic sidefield is a single bit: "does any contribution to this pixel pass through a glossy interaction?" We tested it. On the three scenes of §4.1, the boolean `has_glossy_in_path` saturates on every scene (0.0, 0.0, 1.0 at SPP=4096 on `glass_caustic`, `cubes_dof`, `snooker`), with no intermediate value separating "occasionally hits a glossy surface" from "glossy energy dominates." The continuous fraction $\gamma(p) = L_{\text{has-glossy}}(p)/L_{\text{total}}(p)$ recovers this distinction without a threshold, and the path tracer already accumulates the glossy-touching stream for free: the boolean saved no work and lost the distribution. Booleanising at any threshold would also re-introduce a continuous parameter into a sidefield — exactly what the main axis avoids (§3.2). Continuous storage, threshold-at-use-site.

3.4.0.3. Why glossy stays its own bucket. One might fold glossy specular focusing — informal "soft caustics" — into the `caustic` bucket. We resist this for two reasons. First, glossy-vs-specular is a continuous roughness spectrum, not a discrete transition; a roughness-threshold rule would re-introduce a continuous parameter on the main axis, violating the principle above. Second, the empirical caustic–glossy relationship has no fixed sign: it reverses monotonically with object-space separation between specular and glossy objects (§5). Any rule pre-committing glossy to caustic, or to diffuse-indirect, would mislabel at least one regime. We therefore keep glossy separate and preserve $\gamma(p)$ as the continuous handle through which the caustic/glossy interaction is recoverable downstream.

4. Validation: Exhaustiveness and Cross-SPP Stability

This section reports two empirical results on the same data: (i) the seven-bucket partition closes — the `other` catch-all takes zero energy on every scene, at every sample budget tested; (ii) the discrete mechanism labels are stable across an order-of-magnitude sample-budget change in a regime where the variance-derived ground truth is not. The contrast is the empirical core of our claim that mechanism, not variance, is the right object on which to ground a per-pixel difficulty signal.

4.1. Test matrix

Three scenes span the difficulty regime of interest:

- `cubes_dof` — three coloured cubes on a ground plane viewed through a thin-lens camera; Lambertian materials only, no specular or glossy surfaces. A reference scene where the partition is expected to be dominated by `direct` and `diffuse-indirect`.
- `glass_caustic` — a Cornell-box variant with a glass sphere casting an L-S-D caustic on the floor; the prototypical heavy-tailed transport scene.

)

- `snooker` — six balls of mixed materials (matte plastic, polished gold/silver/copper, glass) on a felt table; mixed glossy and specular surfaces, with inter-ball reflections producing a broad range of mechanism mixtures.

We render each scene at three sample budgets, SPP $\in \{64, 512, 4096\}$, for nine (scene×SPP) runs total. The SPP=4096 render serves as the reference; the partition is computed at each SPP independently. Our implementation is a Mitsuba 3 custom `SamplingIntegrator` that, for every contribution event, records the lobe sampled at V_n and a single bit indicating whether any earlier vertex along the light-to-end-vertex sequence used a delta-specular lobe. These two quantities determine the bucket (Table 1) without any additional ray casting.

4.2. The partition closes

The `other` catch-all takes 0% of contribution-event energy on all nine runs. The two-axis partition is therefore empirically exhaustive on the scenes tested. Per-scene bucket distributions agree with physical expectation (Fig. 2): `cubes_dof` is 90.9% `direct`, with the remainder in `diffuse-indirect`; `glass_caustic` places 7.4% of pixels (10.0% of energy) into `caustic`, localised on the floor below the glass sphere; `snooker` sorts its glass and chrome balls into `caustic` and its gold ball into `glossy`, with the felt as `direct`.

An empirically exhaustive partition (no events fall outside the seven labels) combined with physical agreement (the bucket each region of each scene belongs to is the one a human would assign by inspection) is direct evidence that the (lobe, δ-S) two-axis structure is the right partition, not an arbitrary 7-way grouping that happens to fit.

4.3. Cross-SPP stability

We measure cross-SPP stability in two ways: per-pixel dominant-class agreement across SPP levels, and per-scene bucket-fraction drift (Table 2).

Table 2: *Cross-SPP stability of `dominant_class`. "agree" is the percentage of pixels whose dominant class matches between the two SPP levels. "max drift" is the largest change in any bucket's pixel share between SPP= 64 and SPP= 4096.*

scene	agree (64 vs 4096)	agree (512 vs 4096)	max drift
`cubes_dof`	99.58%	99.82%	0.12%
`glass_caustic`	95.64%	98.70%	1.19%
`snooker`	89.55%	95.77%	0.58%

The headline observation — the strongest single piece of empirical evidence here — is that the label's cross-SPP agreement survives chance-correction on every scene, and does so most informatively where disagreement is largest. Raw agreement could in principle be inflated by a dominant majority bucket; it is not. Cohen's κ, which removes majority-class agreement, stays 0.83–0.98 across all five scenes (Table 4); on `snooker` (lowest raw at 89.6%), $\kappa = 0.828$ is still "almost perfect" on the Landis–Koch scale. Crucially, on the two heavy-tailed scenes the *weighted* κ — which forgives the near-neighbour `direct`↔`diffuse-indirect` swap — *rises* above Cohen's κ (0.828 → 0.885 on `snooker`; 0.890 → 0.900 on `glass_caustic`), confirming those disagreements are fungible boundary flicker, not structural collapse (struct% of 17% and 33%).

Table 3: *Same-scale comparison: cross-SPP (64 vs 4096) agreement of the discretised variance signal versus the mechanism label. $\hat{\sigma}^2$ is binned into quantile bins (3 and 7 bins; bin edges are data-driven quantiles of the per-scene $\hat{\sigma}^2$ distribution, no fixed threshold), then scored by the same agreement metric used for the mechanism label. The mechanism label is 1.5–2× more stable than 3-bin $\hat{\sigma}^2$ and 3–4× more stable than 7-bin.*

scene	3-bin $\hat{\sigma}^2$	7-bin $\hat{\sigma}^2$	mechanism
`cubes_dof`	0.484	0.264	0.996
`glass_caustic`	0.424	0.212	0.956
`snooker`	0.560	0.307	0.896
`tabletop_mixed`	0.596	0.286	—
`tabletop_separated`	0.652	0.335	—

Table 4: *Chance-corrected cross-SPP (64 vs 4096) agreement of the mechanism label. Cohen's κ removes majority-class agreement; weighted κ additionally down-weights the near-neighbour `direct`↔`diffuse-indirect` swap relative to swaps into structural buckets. "struct%" is the fraction of disagreeing pixels involved in a structural-bucket swap.*

scene	raw	Cohen κ	weighted κ	struct%
`cubes_dof`	0.996	0.975	0.975	0.0%
`glass_caustic`	0.956	0.890	0.900	32.9%
`snooker`	0.896	0.828	0.885	17.0%
`tabletop_mixed`	0.987	0.964	0.953	70.3%
`tabletop_separated`	0.994	0.974	0.972	47.2%

The more specific "disagreement concentrates at the near-neighbour boundary" claim holds in the heavy-tailed regime but is scene-dependent. On the original heavy-tailed scenes, 67%–82% of mismatches fall on the `direct`↔`diffuse-indirect` boundary (Fig. 3), and the structural buckets drift by $\leq 1.2\%$ in pixel share. The two tabletop scenes tilt the other way (struct% of 47% and 70%), but this must be read with the absolute rate: they agree at 98.7% and 99.4% raw, so a 70% structural share of a 1.3% disagreement is an absolute structural-swap rate below 1%, and their weighted κ stays ≥ 0.95. The honest summary: in the heavy-tailed regime disagreement is dominated by the near-neighbour boundary; in the light-tailed regime it is rare but not purely near-neighbour — and in both, the absolute structural-swap rate stays below the 1.2% drift bound.

Against the same metric, the variance signal does not come close. Discretising $\hat{\sigma}^2$ onto quantile bins and scoring it with the identical metric (Table 3) — the apples-to-apples comparison the raw "89.5% vs ρ=0.26" juxtaposition does not provide — it agrees with itself across the 64× change at only 0.42–0.65 (3-bin) or 0.21–0.34 (7-bin), against the mechanism label's 0.90–0.996: mechanism is 1.5–2× more stable than coarse-binned variance, 3–4× more than fine-binned.

`snooker`, with the lowest per-pixel agreement (89.55%), has

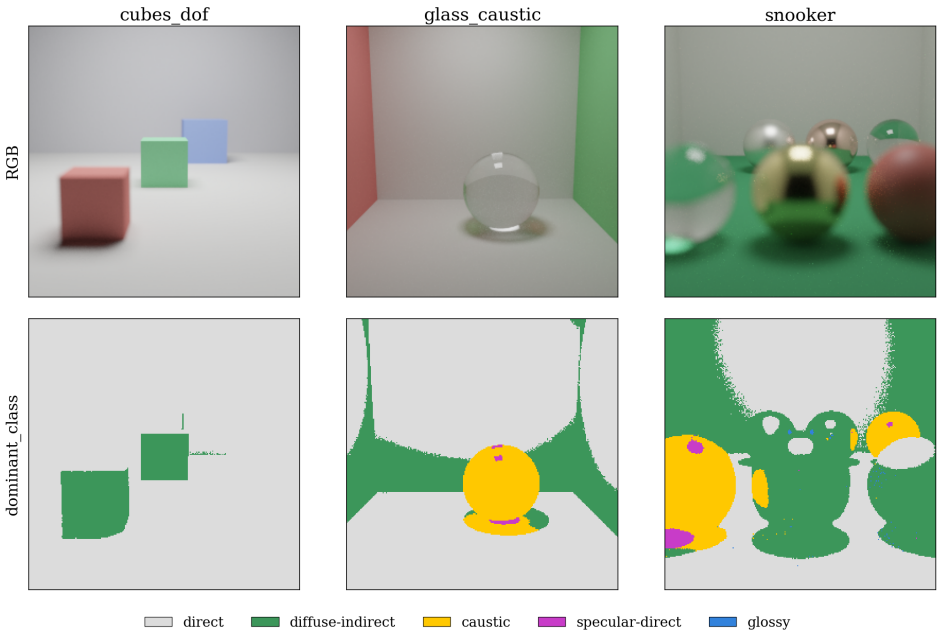

Figure 2: *Per-scene RGB renders (top) and* `dominant_class` *maps (bottom) for the three test scenes.* `cubes_dof` *is 90.9%* `direct` *(Lambertian only).* `glass_caustic` *places 7.4% of pixels into the* `caustic` *bucket on the floor below the glass sphere.* `snooker` *sorts its glass and chrome balls into* `caustic`*, its gold ball into* `glossy`*, and the felt into* `direct`*.*

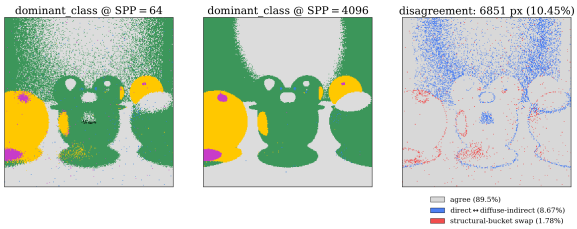

Figure 3: *Cross-SPP stability of* `dominant_class` *on* `snooker`*. Left/middle:* `dominant_class` *at SPP*$=64$ *and SPP*$=4096$*. Right: per-pixel agreement map (grey = agree, blue =* `direct`$\leftrightarrow$`diffuse-indirect` *swap, red = swap to or from a structural bucket* `caustic`/`specular-direct`/`glossy`*).* 89.55% *of pixels agree across the* $64\times$ *budget change; the* 10.45% *disagreement is dominated by the light-tailed boundary (blue,* 8.67%*), with only* 1.78% *of pixels participating in any swap to or from a structural bucket. Structural-bucket pixel shares remain stable to within* 1.2% *(Table 2).*

the smallest bucket-fraction drift (0.58%): boundary pixels flicker along the direct/diffuse-indirect edge of the felt, but the bucket sizes are unmoved. The boundary absorbing most disagreement is the one whose sides are closest by construction — `direct` and `diffuse-indirect` share both axes (no δ-S, non-delta end-vertex) and differ only in bounce count. Residual structural swaps occur — the largest single transition is 0.8% (`glass_caustic`, `specular-direct`→`direct`) — but stay balanced enough that no structural bucket's share moves by more than 1.2%. Discrete labels are SPP-stable where it matters: the structural content of each scene is fixed at SPP$=64$, and only the most fungible boundary needs more budget.

The continuous sidefields drift slightly with SPP, as expected for continuous quantities. The *purity* median drifts by 0.94% to 2.4% across SPP levels; the *glossy energy fraction* is more stable, drifting by less than 0.5% in the `snooker` scene where it is non-zero. The continuous information layered on top of the main partition moves only slightly with budget, while the discrete labels are fully stable in size and overwhelmingly stable in location.

4.4. Robustness to estimator completeness

Cross-SPP stability establishes that the labels do not move when the *sample budget* changes. A second, independent perturbation is available: changing the *sampling strategy* itself. Our pilot integrator resolves the `glossy` and `caustic` buckets through next-event estimation, with the BSDF-sampling-into-emitter contribution suppressed by an emitter-acceptance rule (§3.1) that avoids double-counting in the absence of multiple importance sampling. This means the energy accounting on those two buckets is, by construction, only half of a complete MIS estimator. If the `dominant_class` argmax were sensitive to this, the label would inherit an estimator dependence of exactly the kind we fault $\hat{\sigma}^2$ for. We test it directly.

We implement a restricted power-heuristic MIS on the emitter-hit branch of the `glossy` and `caustic` buckets only — the only two buckets whose end-vertex is non-delta and therefore admits a genuine NEE-versus-BSDF competition; the `emitter-direct`

)

and `specular-direct` buckets have no NEE counterpart and are left at unit weight — and recompute `dominant_class` at SPP= 4096 against the original NEE-only labelling. Two checks confirm the implementation is clean: the `diffuse-indirect` bucket, which the restricted MIS does not touch, drifts by 0.0000 on all four scenes, and the single `specular-direct`→`caustic` transition that survives is a known artefact of an upper-bound test rather than a classification ambiguity.

Table 5: *Robustness of `dominant_class` to estimator completeness. "MIS agree" is the per-pixel argmax agreement between the NEE-only pilot and the MIS-completed estimator at SPP= 4096. "cross-SPP (MIS)" is the 64 vs 4096 agreement recomputed under MIS, to be compared against the NEE-only range of Table 2.*

scene	MIS agree (4096)	cross-SPP (MIS)
`tabletop_separated`	100.00%	99.42%
`tabletop_mixed`	100.00%	98.64%
`glass_caustic`	100.00%	95.64%
`snooker`	99.86%	89.58%

The argmax is almost completely insensitive to restoring the missing estimator half (Table 5): three of four scenes are unchanged to the pixel, and `snooker` moves by 0.14%. The cross-SPP stability recomputed under MIS lands inside the NEE-only range of Table 2 (89.55 → 89.58% on `snooker`; 95.64 → 95.64% on `glass_caustic`): the structural content of each scene is fully determined regardless of whether the estimator is NEE-only or MIS-complete. The structural buckets carry this robustness most strongly — `caustic` agrees at 99.7% or better across the perturbation, because its membership is decided by the presence of a δ-S event in the path sequence, a discrete predicate that no amount of energy re-weighting can move. The buckets that do flicker are, once again, the energy-argmax boundaries (chiefly `glossy` ↔ `diffuse-indirect`), which respond to energy perturbation whatever its source. This is the same structural-versus-boundary split we observe under SPP perturbation, now reproduced on a second, independent perturbation axis: *the discrete label's stability is not an artefact of the particular estimator used to compute it.*

4.4.0.1. The cost of the boundary's sensitivity. The one place the perturbation does move the label is informative. On `snooker`, of the 83 pixels the NEE-only pilot calls `glossy`-dominant, 44 relabel as `diffuse-indirect` under MIS, leaving 37. These 44 are the narrowest-lobe pixels in the bucket: on a near-specular lobe the BSDF PDF dominates the NEE PDF, so the power heuristic correctly drives the NEE weight toward zero, and the NEE-only estimate — which retained that half at full weight — had overstated the glossy energy by a factor that pushed the argmax across the `glossy`/`diffuse-indirect` boundary. The swap is therefore not noise but a correction: the NEE-only pilot systematically over-assigned the narrowest-lobe pixels to `glossy`. We return to the consequence of this for the time-to-quality figures in §6.

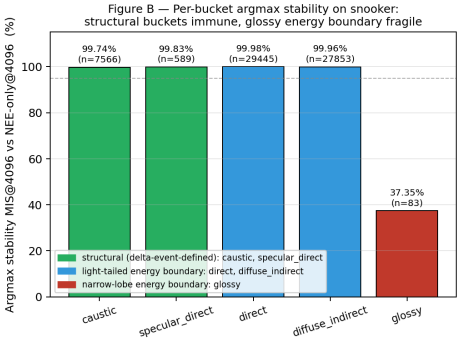

Figure 4: *Per-bucket argmax stability on `snooker` under estimator completion (MIS@4096 vs NEE-only@4096). The structural buckets, whose membership is fixed by a δ-event predicate (`caustic`, `specular-direct`), and the light-tailed energy boundaries (`direct`, `diffuse-indirect`) are all stable above 99.7%; only the narrow-lobe `glossy` energy boundary moves, for the reason analysed below. The dashed line marks 95%.*

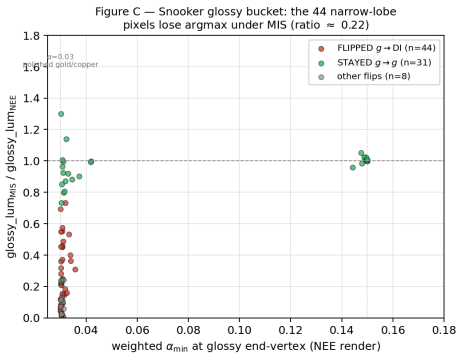

Figure 5: *Mechanism of the `glossy` re-labelling on `snooker`. Each point is a pixel from the NEE-only `glossy` bucket, plotted by its end-vertex lobe width (weighted $\alpha_{\min}$) against the ratio of its glossy energy under MIS to under NEE-only. The 44 FLIPPED pixels (red) cluster at the narrowest lobe ($\alpha \approx 0.03$, polished gold/copper), where MIS down-weights the NEE half and the glossy energy contracts to ≈ 0.22 of its NEE-only value — enough to lose the argmax. The 31 STAYED pixels (green) span wider lobes and keep a ratio near unity. The swap is a correction concentrated on exactly the pixels the kurtosis argument flags as worst-behaved.*

4.5. The variance baseline

The same data, judged by the variance-derived ground truth that current practice would use, is self-inconsistent in a way no amount of post-processing repairs — but, crucially, only in the regime the thesis is about. We measure two split-half self-consistency quantities. First, a per-pixel convergence exponent $\hat{\alpha}(p)$ fit from a four-level SPP sweep ($\hat{\sigma}^2$ at SPP $\in \{64, 256, 1024, 4096\}$) with ten independent seeds per level (∼40,000 samples per pixel): we randomly partition the ten seeds into two halves, refit $\hat{\alpha}(p)$ on each half, and Spearman-correlate the two fields. Second, and more directly rele-

vant to our thesis, the same split-half procedure applied to $\hat{\sigma}^2$ itself at each fixed SPP level. For a noiseless ground truth either correlation would be 1.0.

Table 6: *Split-half self-consistency of* $\hat{\sigma}^2$ *(Spearman* ρ *between two disjoint five-seed halves, per fixed SPP level). A noiseless reference quantity would score* 1.0*. Self-consistency is* tail-dependent*: it collapses on the caustic-dominated scene (where the hardest pixels live) and is stable on the light-tailed scene.*

scene	SPP=64	256	1024	4096
glass_caustic	0.180	0.143	0.205	0.260
snooker	0.482	0.518	0.548	0.578
tabletop_mixed	0.511	0.546	0.600	0.668
tabletop_separated	0.688	0.699	0.707	0.713
cubes_dof	0.310	0.312	0.309	0.310

The two measurements agree, and localise the instability rather than asserting it globally. The $\hat{\alpha}$-derived ranking on the diffuse scenes (cubes_dof, glass_caustic) has split-half $\rho = 0.23$–0.29 — the Problem 3 noise floor. The $\hat{\sigma}^2$ measurement (Table 6) is sharper: its self-consistency is *a function of the scene's tail*. On caustic-dominated glass_caustic it sits at 0.18–0.26 — at or below the $\hat{\alpha}$ floor, and non-monotone in SPP (more samples do not buy stability, itself a heavy-tail signature). On light-tailed tabletop_separated it sits at 0.69–0.71, and on Lambertian cubes_dof a flat $\sim$0.31 across the whole range. The contrast between cubes_dof (flat) and glass_caustic (low and jittering) is exactly the §1 prediction: $\hat{\sigma}^2$-field shape is preserved on the light-tailed subset and least informative where the hardest pixels are.

This is the precise, narrower-and-stronger form of "variance is unstable." Where the integrand is light-tailed, $\hat{\sigma}^2$ is serviceable and nobody should stop using it. Where it is heavy-tailed — the regime holding the hardest pixels, the one a difficulty GT cannot afford to be wrong about — $\hat{\sigma}^2$ fails even to agree with itself across a seed split, and any predictor evaluated against a $\hat{\sigma}^2$- or $\hat{\alpha}$-derived GT there is bounded above by this floor.

The contrast with the mechanism GT is decisive on the same scenes. With a single SPP= 64 pilot, our discrete labels agree with the SPP= 4096 reference at 89.5–99.6% (Table 2); on glass_caustic the label agrees with itself across the 64$\times$ change at 95.6% while $\hat{\sigma}^2$ agrees across a seed split at $\rho \leq 0.26$. Discrete labels are stable exactly where $\hat{\sigma}^2$ is not.

The tail-dependence of the previous paragraph is a scene-level statement; resolving it per bucket localises it to the pixel level and closes the loop with the mechanism partition (Table 7). The single sharpest measurement in this paper is the caustic bucket of glass_caustic: there, $\hat{\sigma}^2$ has a split-half Spearman of 0.13 at SPP= 4096 and 0.02 at SPP= 64 — on the pixels the mechanism label marks as caustic in a caustic-dominated scene, the variance estimator does not agree with itself across a seed split at all. This is $\hat{\sigma}^2$ failing as a ground-truth signal in its most complete form, and it happens precisely on the pixels the partition isolates into a bucket of their own, computed without ever looking at $\hat{\sigma}^2$.

The effect is not that caustic is uniformly the least-self-consistent bucket — in the tabletop scenes specular-direct

Table 7: *Split-half self-consistency of* $\hat{\sigma}^2$ *at SPP=* 4096*, resolved* per mechanism bucket. *"—" marks buckets with no energy in that scene. The* caustic *bucket's self-consistency degrades monotonically with how dominant caustic transport is in the scene: it floors at* 0.13 *on the caustic-dominated* glass_caustic *(and at* 0.02 *there at SPP=* 64*, where the two seed halves are essentially uncorrelated), rises to* 0.39 *where caustic is prominent (*snooker*), and to* 0.50–0.54 *where caustic is a minority mechanism (tabletop scenes).*

scene	dir.	diff.	caus.	spec.	gloss.
snooker	0.49	0.65	0.39	0.57	0.82
glass_caustic	0.18	0.42	0.13	0.44	—
tabletop_mixed	0.67	0.58	0.54	0.49	—
tabletop_separated	0.71	0.56	0.50	0.47	—
cubes_dof	0.20	0.72	—	—	—

is comparable or lower (Table 7). It is that the caustic bucket's self-consistency *tracks how heavy-tailed the scene is*: it collapses where caustic transport dominates and recovers where caustic is incidental — the scene-level tail-dependence of Table 6 at bucket granularity. A variance-based GT has no representation to flag, in advance, which pixels carry this risk; the mechanism label does, by construction. The caustic bucket is, in effect, a pre-computed map of where the variance baseline cannot be trusted.

One entry appears, at first reading, to contradict our cost analysis (§6): the glossy bucket of snooker has the *highest* split-half self-consistency (0.82) yet §6 reports glossy as *slowest-converging*. There is no contradiction at the level of failure modes: self-consistency measures whether the $\hat{\sigma}^2$ *estimator* agrees with itself, time-to-quality whether the *pixel* converges. A glossy pixel can have a stably-estimated, consistently-high variance — $\hat{\sigma}^2$ reliably reports "hard" (high self-consistency), consistent with slow convergence — whereas the caustic failure is that $\hat{\sigma}^2$ cannot even report a consistent value. The partition separates these two failure modes into two buckets where a single continuous number would conflate them. We caution, though, that the bucket-level 0.82 is itself an average over two opposite sub-populations: as §5.4 shows, the narrowest-lobe glossy pixels have *negative* ρ, and the high mean is carried by a wider-lobe, multi-path sub-population. The glossy bucket is not internally homogeneous in $\hat{\sigma}^2$ stability — itself a reason to prefer the discrete label over any bucket-averaged continuous statistic.

5. Cross-Scene Correlation Structure

The partition equips each pixel with not only a discrete label but a small continuous vector — the caustic, glossy, and bounce-depth energy fractions and the inverse purity — defined deterministically on the same path-tracer state. We examine how these components *correlate across pixels* in each scene and how that structure *varies across scenes*. Two questions drive the analysis. First, local to §3.4: does the empirical caustic–glossy relationship justify keeping them as independent buckets, or should they merge? Second, structural: does the correlation structure expose scene-level geometric variables a variance-based GT cannot see?

)

We compute per-pixel Spearman correlations between the four continuous components — `caustic_frac`, `glossy_frac`, `depth` (energy-weighted average bounce count to the end-vertex), and `inv_purity` $(1-\pi)$ — on five scenes: the three of §4.1 plus two new tabletop scenes `tabletop_mixed` and `tabletop_separated` introduced below. Table 8 reports the six pairwise correlations across all five scenes. Two pairs are stably independent across every non-degenerate scene; two are scene-dependent with stable sign; and two are scene-dependent with their *sign* reversing across scenes. The two sign-reversing pairs are the substantive content of this section: in each, the sign is controlled by an identifiable scene-level geometric variable that the mechanism representation makes visible. We promote these to named findings in §5.1 and §5.2.

Table 8: *Per-pixel Spearman ρ among the four continuous components of the difficulty vector, across five scenes. Components are abbreviated as $C =$ `caustic_frac`, $G =$ `glossy_frac`, $D =$ `depth`, $P =$ `inv_purity`. Scene columns are `cubes_dof` (cb), `glass_caustic` (gc), `snooker` (sn), `tabletop_mixed` (tm), `tabletop_separated` (ts). Entries marked "deg." have one component identically zero in that scene (no caustic energy in cb; no glossy energy in gc and cb). Symbol key: ★ stable independent across all non-degenerate scenes; † scene-dependent, sign-stable; ‡ sign-reversal across scenes (Findings 1 and 2, §5.1–§5.2); ○ open observation (§5.3).*

pair	cb	gc	sn	tm	ts	note
$C \leftrightarrow P$	*deg.*	indep.	indep.	indep.	indep.	★
$P \leftrightarrow G$	*deg.*	*deg.*	indep.	indep.	indep.	★
$C \leftrightarrow G$	*deg.*	*deg.*	−0.12	+0.44	−0.27	‡ F1
$D \leftrightarrow C$	*deg.*	+0.73	+0.45	+0.43	−0.21	○
$D \leftrightarrow G$	*deg.*	*deg.*	+0.20	+0.42	+0.23	†
$D \leftrightarrow P$	+0.24	+0.52	+0.66	−0.37	+0.50	‡ F2

5.0.0.1. Stable and scene-dependent structure. Two pairs read as independent across every non-degenerate scene: `caustic_frac` versus `inv_purity`, and `inv_purity` versus `glossy_frac`. Two further pairs are scene-dependent but sign-stable. The remaining two pairs invert sign across scenes; both are examined below. Each sign-reversal exposes a scene-level geometric variable that controls the relationship between two mechanism components — a variable a variance-based per-pixel signal has no representation for and therefore cannot detect.

5.1. Finding 1: caustic/glossy sign reversal with object-space separation

The decision to keep glossy independent of caustic (§3.4) was motivated by a falsifiable structural prediction. If the two buckets reflected a fixed relational structure, the per-pixel correlation between `caustic_frac` and `glossy_frac` should have a consistent sign across scenes. The mechanism representation predicts otherwise. A contribution event lands in the `caustic` bucket if its light-to-end-vertex sequence contains a δ-S event; the same path may also pass through a glossy interaction, in which case γ, the glossy energy fraction (§3.3), is non-zero on the same pixel. The two fractions therefore *co-occur* when the scene's object-space topology makes such combined light paths common, and *compete* when geometry forces a path to pick a specular or a glossy interaction but not both. The sign of the correlation should depend on the object-space separation between specular and glossy materials, and should cross zero somewhere along the separation axis.

We test this on three scenes whose only relevant difference is the object-space layout of specular and glossy surfaces. `tabletop_mixed` has glass and metal balls placed in close proximity on a tabletop, so a single light path frequently encounters both within one bounce sequence (entanglement regime). `snooker` has six balls of mixed materials arranged on a felt surface with irregular packing. `tabletop_separated` shares its geometry with `tabletop_mixed` but adds an opaque partition forcing spatial separation between the glass and metal groups (competition regime); this scene was constructed specifically to falsify the prediction. If the sign reversal does not appear here, the mechanism account of the caustic–glossy relationship is wrong.

Table 9: *Per-pixel Spearman correlation between the caustic and glossy energy fractions, on three scenes ordered by object-space separation between specular and glossy materials. The full correlation matrix appears in Table 8.*

scene	configuration	$\rho_{\text{caustic–glossy}}$
`tabletop_mixed`	entangled	+0.441
`snooker`	irregular	−0.121
`tabletop_separated`	separated	−0.270

The correlation drops monotonically along the separation axis and crosses zero between `tabletop_mixed` and `snooker` (Table 9, Fig. 6). The prediction holds, including on the controlled scene constructed to falsify it. The sign of the caustic–glossy correlation is therefore not a property of the GT; it is a continuous function of scene geometry that swings through zero. Any rule that fixed glossy's relationship to caustic in advance — either by merging them into a single bucket, or by giving glossy a dominant ordering against caustic — would mislabel at least one of the three regimes. The finding closes the design question left open in §3.4: glossy stays its own bucket because the alternative is provably wrong on at least one scene.

5.2. Finding 2: depth/purity sign reversal with material topology

A natural objection to the difficulty vector is that bounce-`depth` alone might already carry most of what the other components encode — "deeper paths are messier paths." The correlation between `depth` and `inv_purity` provides the direct test: if depth predicts mechanism mixing monotonically, the sign of $\rho(\texttt{depth}, \texttt{inv_purity})$ should be positive across scenes. The five-scene measurement (Table 8, last row) shows this is not the case: ρ takes values +0.24, +0.52, +0.66, −0.37, +0.50 on `cubes_dof`, `glass_caustic`, `snooker`, `tabletop_mixed`, and `tabletop_separated` respectively, including one negative outlier of substantial magnitude.

The mechanism explanation is geometric and not parametric. Two families of deep paths arise in our scenes. The first is specular-chain-dominated: light bounces multiple times among chrome and

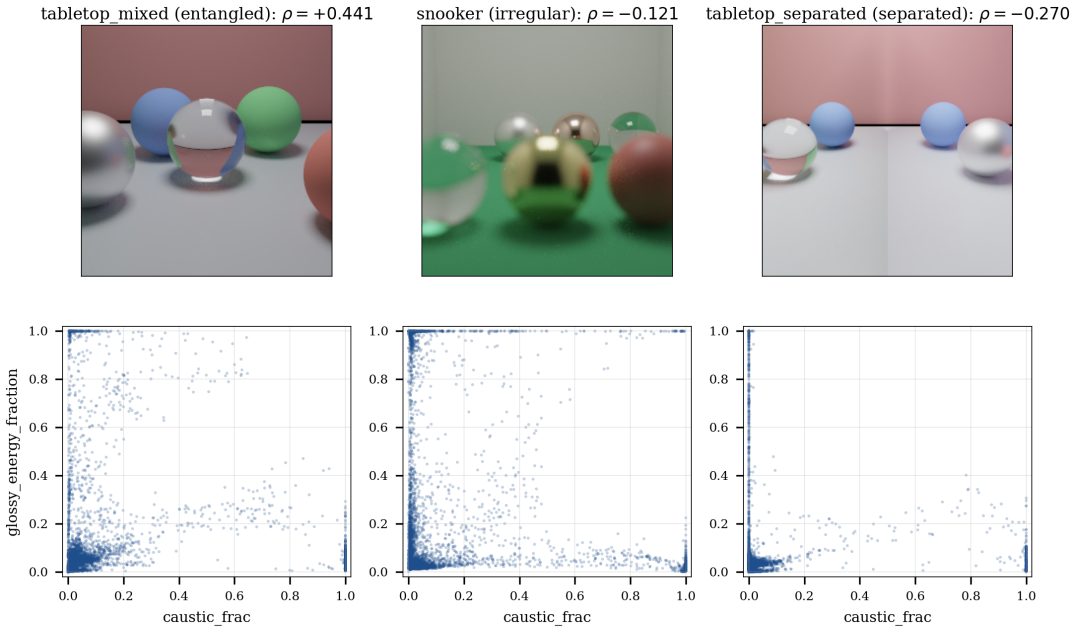

Figure 6: *Sign reversal of the caustic/glossy correlation with object-space separation. Top: RGB renders of the three scenes ordered by object separation. Bottom: per-pixel scatter of* `caustic_frac` *vs* `glossy_energy_fraction`*. ρ drops monotonically from* $+0.441$ *(entangled) through* -0.121 *(snooker, irregular) to* -0.270 *(separated).* `tabletop_separated` *was constructed specifically to falsify the prediction — if the sign reversal had not appeared there, the mechanism account of the caustic/glossy relationship would be wrong.*

glass surfaces before reaching a non-delta end-vertex (`snooker` between metal balls, `glass_caustic` through the glass sphere, `tabletop_separated` bouncing along the partition). Such paths accumulate length without accumulating lobe diversity — every intermediate bounce is the same δ-S event — so deep pixels remain mechanically *pure*, and depth correlates with π, hence *positively* with `inv_purity` in the weak sense that purer pixels also tend to be shallower in the diffuse parts of the scene. The second family is glass-transmission-dominated: light passes through a glass surface (one δ-S), then bounces among diffuse and glossy interior surfaces before exiting. These paths mix lobe types as they go, and depth correlates with mechanism *mixing* — but in `tabletop_mixed`, the lattice of nearby glass and metal balls makes most deep paths of this second family, while the shallow paths are the pure specular reflections on the near sides of the balls: shallow = mixed, deep = pure, yielding a negative ρ.

The two families coexist in every scene but in different proportions; the sign of $\rho(\texttt{depth}, \texttt{inv_purity})$ is determined by which family dominates. The implication for the difficulty representation is direct: any reduction of the difficulty vector that absorbs `depth` into `inv_purity`, or treats them as redundant, will mislabel at least one scene by sign. `depth` and `inv_purity` are independent dimensions *by data*, not merely by construction.

5.3. Observation and open item

One additional pair in Table 8 exhibits a sign flip: `depth` ↔ `caustic_frac` runs $+0.73, +0.45, +0.43, -0.21$ across the four non-degenerate scenes. The first three values are unsurprising — caustic paths include a specular bounce and so tend to be at least one vertex longer than direct paths. The negative outlier on `tabletop_separated` is harder to interpret confidently: the partition geometry constrains both `depth` (paths must detour around the wall) and `caustic_frac` (which surfaces the partition occludes) simultaneously, and a single-scene anomaly does not distinguish a genuine mechanism from a partition-induced artefact. We mark this as an open observation rather than a named finding: confirming or rejecting it would require a separation-axis sweep analogous to the one underlying Finding 1, which we leave to future work.

5.4. Finding 3: lobe width controls the glossy heavy tail

The `glossy` bucket is the one on which continuous cost saturates (§6), and the robustness analysis (§4.4) showed that its narrowest-lobe members are also the ones whose membership is least stable under estimator completion. Both point at the end-vertex roughness α as the controlling variable. We test this directly with a controlled single-sphere experiment that removes the confounds present in a multi-object scene like `snooker`, where lobe width and path multiplicity vary together.

5.4.0.1. Controlled roughness sweep. We render a single isotropic `roughconductor` sphere in an otherwise diffuse box — no glass or smooth specular surfaces, so no δ-S contaminates the bucket — and sweep the lobe width α over eight values while holding geometry, illumination, and camera fixed. The pre-registered prediction was that two quantities would co-vary with α: the per-

)

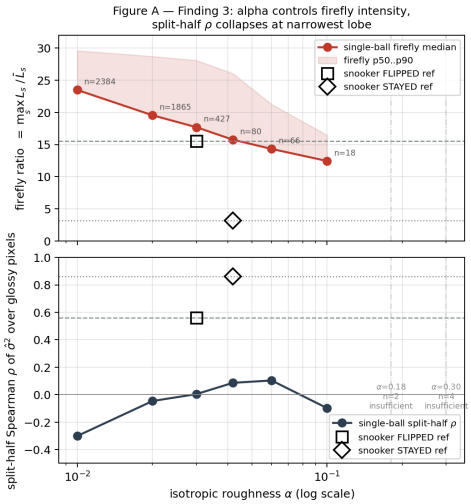

Figure 7: *Finding 3 on a controlled single-sphere roughness sweep.* Top: *the median firefly ratio falls monotonically as the lobe widens, from* 23.5 *at* α=0.01 *to* 12.4 *at* α=0.10 *(shaded p50–p90 band); the* `snooker` *FLIPPED reference (narrow-lobe,* □*) lands on the curve while the STAYED reference (wide-lobe, multi-path,* ◊*) sits far below it, a regime the single sphere does not reproduce.* Bottom: *the split-half* ρ *of* $\hat{\sigma}^2$ *is* not *monotone in* α *— it collapses to* -0.30 *at the narrowest lobe (seed halves anti-correlated) and otherwise sits in a noise band — while the STAYED reference's* $\rho \approx 0.86$ *floats well above the entire single-sphere curve. Lobe width controls the firefly tail; it does not, on its own, account for the stable-glossy regime. The two widest* α *levels yield too few* `glossy`*-dominant pixels to measure.*

pixel firefly ratio (the seed-wise ratio of the maximum to the mean glossy contribution) and the split-half self-consistency ρ of $\hat{\sigma}^2$. The firefly prediction is confirmed cleanly: the median firefly ratio falls monotonically from 23.5 at α=0.01 to 12.4 at α=0.10 across six populated α levels (the two widest lobes produce too few `glossy`-dominant pixels to measure, the bucket's energy having dispersed; we report the firefly trend over $\alpha \in [0.01, 0.10]$). Narrow lobes carry a heavier per-pixel contribution tail — exactly the signature the kurtosis bound of §1 predicts, with the narrowest lobe being the highest-kurtosis integrand.

5.4.0.2. $\hat{\sigma}^2$ self-consistency fails on the same pixels, but not monotonically. The split-half ρ of $\hat{\sigma}^2$ does not rise smoothly with α. It is most informative at the narrow end: at α=0.01 it is -0.30 — the two seed halves rank the pixels in *anti*-correlation, the unambiguous fingerprint of an estimator dominated by rare large samples — and elsewhere it sits in a noise band (-0.10 to $+0.10$) without a clean trend. We do not claim a monotone $\rho(\alpha)$ relationship; the honest reading is narrower and still on-thesis: on a controlled glossy bucket, $\hat{\sigma}^2$ fails to agree with itself across a seed split at every α we can measure, and on the narrowest lobe it agrees *in reverse*. A signal that cannot reproduce its own per-pixel ranking on a clean single-material sphere is not a ground truth on that material, irrespective of budget. This is the glossy-bucket counterpart of the caustic-bucket failure reported in §4.5.

5.4.0.3. What lobe width does not explain. A second controlled experiment isolates the limits of the lobe-width account. The `snooker glossy` bucket contains a sub-population (the 37 pixels that survive MIS, §4.4) with both a low firefly ratio (≈ 3) and a high split-half ρ (≈ 0.86) — a "stable" glossy regime that the single-sphere sweep never reproduces at any α. We tested whether per-pixel path multiplicity was the missing variable by holding α=0.03 fixed and sweeping the number of energy-conserving light sources from one to sixteen. It is not: the firefly ratio stayed flat (as predicted, since α was fixed), but ρ did not climb out of the noise band, because adding light *sources* did not increase per-pixel supplying-path *multiplicity* — each pixel still saw only one or two caustic hits. The stable-glossy regime of `snooker` is thus governed by some scene-topology variable — multi-object inter-reflection is our leading candidate — that neither controlled scene reproduces. We record it as an open item rather than over-claim a mechanism we have not isolated. What survives controlled test is the narrower, on-thesis statement: lobe width monotonically controls the firefly intensity of the glossy bucket, and $\hat{\sigma}^2$ is self-inconsistent across that bucket, most severely on the narrowest lobes.

5.4.0.4. Implications for the GT. The same continuous handles (γ, `depth`, `inv_purity`) that decorate the mechanism partition at the pixel level expose, when correlated across pixels and scenes, the geometric variables that govern interactions between the mechanism buckets. A variance-based ground truth, which has no representation of light-path topology, could not have produced these predictions, and would not detect that the difficulty vector's effective dimensionality is scene-dependent. Two *specific* variables — object-space separation between specular and glossy materials (Finding 1) and the relative abundance of specular-chain versus glass-transmission paths (Finding 2) — are what the mechanism representation makes visible. Both reverse the sign of a pair-wise correlation; neither can be detected by inspecting per-pixel $\hat{\sigma}^2$.

6. Closure: Continuous Cost Saturates Where Mechanism Does Not

We close with a second route to the same thesis. The variance instability of §1 concerned a sample *statistic*; the failure was that an estimator of σ^2 cannot be stable on heavy-tailed integrands. The same heavy-tailed regime should also make any continuous *cost* measurement saturate — pixels that fail to converge within a budget cap show up as right-censored, not as informative cost values — while the mechanism representation should remain functional on the same pixels. We test this directly, in two parts: a per-bucket median cost (§6.1) that identifies the bucket in which continuous measurement saturates, and a per-dimension correlation between cost and the difficulty vector (§6.2) that quantifies the consequences.

For each pixel we measure a per-pixel time-to-quality cost $c(p)$, defined as the smallest SPP at which $|I_N(p) - I_{\text{ref}}(p)| / \max(\text{lum}(I_{\text{ref}}), \epsilon) < \tau$, with I_{ref} the SPP=4096 reference image and τ a fixed relative-error threshold. Pixels that fail to reach the threshold by $\text{SPP}_{\max} = 2048$ are right-censored — their true cost is unknown but lies above the cap.

6.1. Per-bucket cost and the glossy long tail

Table 10 reports the median $c(p)$ per `dominant_class` bucket on each of four scenes.

Table 10: *Median time-to-quality cost (SPP) per `dominant_class` bucket. "—" marks buckets with no pixels. The single right-censored entry is shown as* 2048† *with the censoring rate. The `glossy` bucket is the NEE-defined population of the pilot integrator; its membership is estimator-dependent (§4.4).*

bucket	snooker	t._mixed	t._sep.	g._caustic
`direct`	16	16	16	16
`diffuse-indirect`	32	32	32	16
`caustic`	32	32	32	32
`specular-direct`	128	256	128	256
`glossy`	2048†	—	—	—

† 51.8% of the NEE-defined `glossy` bucket right-censored at $\mathrm{SPP}_{\max} = 2048$.

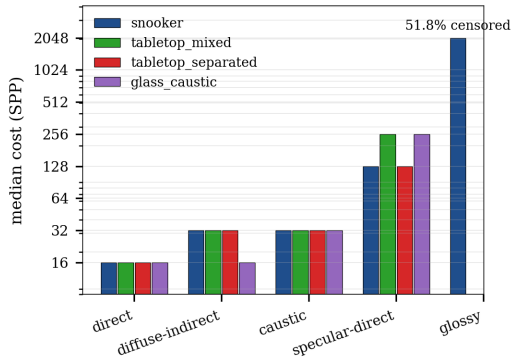

Figure 8: *Time-to-quality median cost per `dominant_class` bucket. The (NEE-defined) `glossy` bucket on `snooker` requires SPP≥ 2048 for median pixels to converge, with 51.8% right-censored at that budget. The continuous cost signal saturates on the glossy bucket while the discrete mechanism label remains operational on the same pixels; §4.4 shows the bucket population is itself estimator-dependent.*

The critical entry is in the bottom-left: on `snooker` the `glossy` bucket has a median cost of 2048 SPP, with 51.8% of its pixels right-censored at that budget (Fig. 8). The continuous cost measurement saturates — half of these pixels never converge in our test, so the reported median is in effect a lower bound on the true median. The mechanism classifier, on the exact same pixels, remains operational: discrete labels are still discriminating in a regime where time-to-quality has ceased to. The qualitative claim of this table — that a continuous cost signal saturates while the discrete label does not — is robust; the precise magnitude of the censoring rate is, however, sensitive to how the `glossy` bucket is defined, which we examine next.

6.1.0.1. The 51.8% figure is bucket-definition-dependent. The cost above is measured over the population of pixels our pilot integrator labels `glossy`-dominant. That label is an energy argmax (§3.1), and an energy argmax is only as trustworthy as the energy accounting beneath it. Our pilot integrator resolves the `glossy` and `caustic` buckets through next-event estimation only; the BSDF-sampling-into-emitter half of the estimator, which carries substantial energy on glossy end-vertices, is suppressed by the emitter-acceptance rule that avoids double-counting in the absence of MIS. The robustness analysis of §4.4 quantifies the consequence: when the missing half is restored through a power-heuristic MIS computed on the `glossy` and `caustic` buckets, the `snooker` `glossy` population contracts from 83 to 37 pixels — the 44 pixels that leave are the narrowest-lobe pixels, on which MIS correctly down-weights the high-variance NEE half, and they re-label as `diffuse-indirect`. The 51.8% censoring rate is therefore measured over an NEE-only-defined population that is itself a superset of the MIS-defined `glossy` bucket. We report it as the convergence cost of the *NEE-only-defined* `glossy` bucket and flag the definition dependence explicitly; the convergence cost of the MIS-defined bucket is left to a clean recomputation. The argument of this section does not rest on the particular value: under either definition, the `glossy` bucket is the one on which $c(p)$ saturates and the discrete label keeps discriminating.

A second consequence of the table is that the common assumption that `caustic` is the worst case is, at least on these scenes, not borne out: `caustic` converges at 32 SPP across all four scenes, while the `glossy` bucket carries a right-censored long tail at many times that budget. The mechanism partition exposes this because the two regimes live in different buckets by construction; a continuous difficulty signal computed from $\hat{\sigma}^2$ or from $c(p)$ blends them into one tail. We are deliberately more cautious than to call the glossy reflections the single "hardest" pixels in absolute terms: as the robustness analysis shows, part of the apparent hardness of the NEE-only `glossy` population is the high variance of the NEE estimator on narrow lobes rather than an intrinsic property of the transport — itself an instance of the estimator-dependence this paper argues against in the variance signal.

6.2. Cost versus the difficulty vector, dimension by dimension

The per-bucket view answers "which mechanism is most expensive." A complementary question is whether the continuous components of the difficulty vector correlate with cost in a stable way — whether `depth`, `caustic_frac`, `inv_purity`, and `glossy_frac` would each serve, on their own, as a cheap proxy for time-to-quality. The answer, shown in Table 11, is that they would not, and the reasons are informative about both the cost measurement and the structure of the vector itself.

Table 11: *Per-pixel Spearman correlation between time-to-quality $c(p)$ and the four continuous components of the difficulty vector, on the four non-degenerate scenes. "n/a" marks `glossy_frac` on `glass_caustic`, which has no glossy energy.*

GT dim	snooker	t._mixed	t._sep.	g._caustic
`depth`	+0.272	+0.439	**+0.040**	+0.221
`caustic_frac`	+0.070	+0.214	+0.006	+0.168
`inv_purity`	−0.160	+0.069	−0.101	+0.063
`glossy_frac`	+0.229	+0.218	+0.203	n/a

Three observations follow. We label them to mirror the structure of §5.

)

6.2.0.1. Observation A: `glossy_frac` is the only sign-stable dimension, and its magnitude is structurally capped. $\rho(\texttt{glossy_frac}, c)$ is positive on all three non-degenerate scenes, in the narrow range $+0.20$ to $+0.23$. No other component is sign-stable. But $\rho \approx 0.21$ explains only $\sim 4\%$ of the variance in c, and this number is not the intrinsic strength of the relationship: the bottom-left entry of Table 10 is the cause. With 51.8% of the NEE-defined glossy bucket right-censored, Spearman ρ on c assigns them all the same maximal rank, compressing the top of the distribution into a plateau that systematically pulls correlations toward zero. The weak ρ does not mean the difficulty vector fails to track cost; it means *cost itself has gone flat on the very pixels the vector flags as glossy-dominant*. The discrete-vs-continuous contrast of §4.5 recurs here at the level of a second continuous signal: where mechanism labels would let the predictor isolate the glossy long-tail by bucket, cost has nothing left to say about it.

6.2.0.2. Observation B: `inv_purity` versus cost is sign-unstable. $\rho(\texttt{inv_purity}, c)$ is -0.160, $+0.069$, -0.101, $+0.063$ across the four scenes — uniformly small, signs disagreeing. The instability is not noise (it is consistent within each scene) but needs care: the link between mechanism mixing and cost is mediated by which bucket the mixed pixels fall into and which buckets are expensive there (Table 10). Where caustic pixels are pure and cheap, `inv_purity` on the non-caustic remainder carries the variation, its sign reflecting whichever non-caustic mechanism dominates locally. The component does the work of §3.3 — flagging mixture — but is not, alone, a stable cost proxy.

6.2.0.3. Observation C: `depth` versus cost is unstable, and the failure matches Finding 2. $\rho(\texttt{depth}, c)$ is $+0.27$, $+0.44$, $+0.22$ on `snooker`, `tabletop_mixed`, `glass_caustic`, but **collapses to** $+0.04$ on `tabletop_separated` — the same scene where $\rho(\texttt{depth}, \texttt{inv_purity})$ inverts sign (Finding 2), for the same reason: the partition geometry forces a family of long paths that does not coincide with the family driving cost in the other three scenes. `depth` is the component whose relationship with cost is most geometry-sensitive; any predictor using it as a cost proxy must do so per-scene.

6.2.0.4. Reading the table. A naive reading of Table 11 is that the difficulty vector tracks cost weakly. The closer reading is that *the continuous-cost signal itself* is locally degenerate on the hardest pixels (A), sign-unstable on the mixture component (B), and geometry-varying on the depth component (C). The mechanism partition sits ahead of all three: it separates the saturating pixels into their own bucket before averaging, and need not correlate with cost to be informative — the bucket identity is the signal.

6.2.0.5. Closing the loop. Two consequences. First, the variance-instability argument of §1 is not specific to sample variance: any continuous difficulty measurement reaches a saturation regime on the hardest pixels, and the empirical hardest bucket — the glossy long tail — is where it hurts most. Second, the mechanism classifier does not hit this regime change: the `glossy` boundary is computed from path structure, not estimated cost, so it does not saturate when cost does. A discrete representation built on the integrand's deterministic structure is robust where the continuous estimators are not, at both ends of the argument this paper makes.

7. Conclusion

Per-pixel rendering difficulty has had thirty-five years of operational consensus and no agreed-upon ground truth. The default — the per-pixel sample variance of the path-traced estimator — is mathematically unsound under heavy-tailed light transport, empirically unstable on the hardest pixels of typical scenes, and inherits a noise floor of $\rho \approx 0.25$ that upper-bounds any predictor evaluated against it. The community's response has been to suppress the symptom at the mean estimator and refine the variance estimator itself; we instead take the instability as evidence that the wrong object has been chosen as ground truth.

We propose to ground the difficulty signal in the discrete transport mechanism of each contribution event, defined by a two-axis partition — end-vertex BSDF lobe by presence of δ-S in the light-to-end-vertex sequence — yielding seven exhaustive, mutually exclusive labels with no continuous threshold on the main axis. The partition closes empirically (`other` = 0% on nine (scene×SPP) runs), is stable across a 64× change in sample budget on the same data where the variance-derived ground truth agrees with itself at $\rho \approx 0.25$, and exposes a five-scene correlation structure of the difficulty vector in which two pair-wise correlations reverse sign monotonically with identifiable geometric variables (object-space separation between specular and glossy materials; relative abundance of specular-chain versus glass-transmission paths). One of the two predictions holds on a controlled scene constructed specifically to falsify it. The labels are stable not only across the 64× budget change but across estimator completion: restoring the MIS half of the estimator leaves the per-pixel argmax unchanged on 99.9–100% of pixels, the structural buckets immune by construction. In the regime where time-to-quality itself saturates — 51.8% of the NEE-defined glossy bucket unconverged at 2048 SPP on a representative scene, a rate we show to be bucket-definition-dependent — the discrete labels remain discriminating where the continuous cost signal has gone flat. Variance-based rendering difficulty is not a stable ground truth under heavy-tailed transport; path mechanism is.

7.0.0.1. Future work. Several extensions follow naturally. First, one regime resists our controlled account: the stable-glossy subpopulation of `snooker` (low firefly, high $\hat{\sigma}^2$ self-consistency) is reproduced by neither the single-sphere roughness sweep nor the light-count sweep of §5.4, which together rule out lobe width and light-source count as its cause; isolating the responsible scene-topology variable — multi-object inter-reflection is our leading candidate — requires a controlled two-object experiment we leave to future work. Second, the cost-saturation argument of §6 was made against L_2 image error; a perceptual-MSE variant (FLIP, FovVideoVDP) might saturate differently, and an analogous analysis under perceptual ground truth is a clean follow-up. Third, combining the stability of mechanism labels with the local responsiveness of sample-derived signals in a Bayesian shrinkage formulation is a natural direction, with mechanism providing the prior and the sample-derived quantity serving as the likelihood. Finally, the

bounce-depth structure consumed implicitly by our partition admits an explicit path-length axis — for instance, a per-pixel Sobol-axes effective dimension — that would complement the lobe-based labels along a quantitative dimension orthogonal to mechanism type.

References

[BSS*13] BELCOUR L., SOLER C., SUBR K., HOLZSCHUCH N., DURAND F.: 5d covariance tracing for efficient defocus and motion blur. *ACM Transactions on Graphics 32*, 3 (July 2013), 31:1–31:18. doi:10.1145/2487228.2487239. 4

[CB02] CASELLA G., BERGER R. L.: *Statistical Inference*, 2nd ed. Duxbury / Thomson Learning, Pacific Grove, CA, 2002. 2

[DHS*05] DURAND F., HOLZSCHUCH N., SOLER C., CHAN E., SILLION F. X.: A frequency analysis of light transport. *ACM Transactions on Graphics 24*, 3 (July 2005), 1115–1126. Proceedings of SIGGRAPH 2005. doi:10.1145/1186822.1073320. 3

[DWR10] DECORO C., WEYRICH T., RUSINKIEWICZ S.: Density-based outlier rejection in Monte Carlo rendering. *Computer Graphics Forum 29*, 7 (2010), 2119–2125. Proceedings of Pacific Graphics 2010. doi:10.1111/j.1467-8659.2010.01799.x. 2, 3

[EKM97] EMBRECHTS P., KLÜPPELBERG C., MIKOSCH T.: *Modelling Extremal Events for Insurance and Finance*, vol. 33 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 1997. 2

[FFJ23] FIRMINO A., FRISVAD J. R., JENSEN H. W.: Denoising-aware adaptive sampling for Monte Carlo ray tracing. In *ACM SIGGRAPH 2023 Conference Proceedings (SIGGRAPH '23)* (New York, NY, USA, 2023), ACM, pp. 32:1–32:11. doi:10.1145/3588432.3591537. 2, 3

[GGSK19] GRITTMANN P., GEORGIEV I., SLUSALLEK P., KŘIVÁNEK J.: Variance-aware multiple importance sampling. *ACM Transactions on Graphics 38*, 6 (Nov. 2019), 152:1–152:9. Proceedings of SIGGRAPH Asia 2019. doi:10.1145/3355089.3356515. 4

[HDF15] HANIKA J., DROSKE M., FASCIONE L.: Manifold next event estimation. *Computer Graphics Forum 34*, 4 (2015), 87–97. Proceedings of EGSR 2015. doi:10.1111/cgf.12681. 4

[Hec90] HECKBERT P. S.: Adaptive radiosity textures for bidirectional ray tracing. *ACM SIGGRAPH Computer Graphics 24*, 4 (Sept. 1990), 145–154. doi:10.1145/97880.97895. 4

[Jen96] JENSEN H. W.: Global illumination using photon maps. In *Rendering Techniques '96 (Proceedings of the 7th Eurographics Workshop on Rendering)* (Porto, Portugal, 1996), Pueyo X., Schröder P., (Eds.), Springer-Verlag, pp. 21–30. doi:10.1007/978-3-7091-7484-5_3. 4

[KK06] KOLLIG T., KELLER A.: Illumination in the presence of weak singularities. In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, Niederreiter H., Talay D., (Eds.). Springer, Berlin, Heidelberg, 2006, pp. 245–257. doi:10.1007/3-540-31186-6_15. 2

[KKR18] KUZNETSOV A., KALANTARI N. K., RAMAMOORTHI R.: Deep adaptive sampling for low sample count rendering. *Computer Graphics Forum 37*, 4 (2018), 35–44. Proceedings of EGSR 2018. doi:10.1111/cgf.13473. 2, 3

[MGN17] MÜLLER T., GROSS M., NOVÁK J.: Practical path guiding for efficient light-transport simulation. *Computer Graphics Forum 36*, 4 (2017), 91–100. Proceedings of EGSR 2017. doi:10.1111/cgf.13227. 4

[Mit87] MITCHELL D. P.: Generating antialiased images at low sampling densities. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '87)* (1987), vol. 21 of *Computer Graphics*, ACM, pp. 65–72. doi:10.1145/37401.37410. 2, 3

[MMR*19] MÜLLER T., MCWILLIAMS B., ROUSSELLE F., GROSS M., NOVÁK J.: Neural importance sampling. *ACM Transactions on Graphics 38*, 5 (Oct. 2019), 145:1–145:19. doi:10.1145/3341156. 4

[Owe06] OWEN A. B.: Quasi-Monte Carlo for integrands with point singularities at unknown locations. In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, Niederreiter H., Talay D., (Eds.). Springer, Berlin, Heidelberg, 2006, pp. 403–417. doi:10.1007/3-540-31186-6_24. 2

[Pix24] PIXAR ANIMATION STUDIOS: Light path expressions — RenderMan 26 documentation. https://rmanwiki-26.pixar.com/space/REN26/19661883/Light+Path+Expressions, 2024. Accessed: 2026-05-27. 4

[RGH*22] RATH A., GRITTMANN P., HERHOLZ S., WEIER P., SLUSALLEK P.: EARS: Efficiency-aware Russian roulette and splitting. *ACM Transactions on Graphics 41*, 4 (2022). Proceedings of SIGGRAPH 2022. doi:10.1145/3528223.3530168. 2, 3

[RKZ11] ROUSSELLE F., KNAUS C., ZWICKER M.: Adaptive sampling and reconstruction using greedy error minimization. *ACM Transactions on Graphics 30*, 6 (2011), 159:1–159:12. Proceedings of SIGGRAPH Asia 2011. doi:10.1145/2070781.2024193. 2, 3

[RMB07] RAMAMOORTHI R., MAHAJAN D., BELHUMEUR P.: A first-order analysis of lighting, shading, and shadows. *ACM Transactions on Graphics 26*, 1 (Jan. 2007), 2:1–2:21. doi:10.1145/1186644.1186646. 4

[SKW*17] SCHIED C., KAPLANYAN A., WYMAN C., PATNEY A., CHAITANYA C. R. A., BURGESS J., LIU S., DACHSBACHER C., LEFOHN A., SALVI M.: Spatiotemporal variance-guided filtering: Real-time reconstruction for path-traced global illumination. In *Proceedings of High Performance Graphics (HPG '17)* (2017), ACM, pp. 1–12. doi:10.1145/3105762.3105770. 2, 3

[SSD*09] SOLER C., SUBR K., DURAND F., HOLZSCHUCH N., SILLION F.: Fourier depth of field. *ACM Transactions on Graphics 28*, 2 (May 2009), 18:1–18:12. doi:10.1145/1516522.1516529. 4

[vdV98] VAN DER VAART A. W.: *Asymptotic Statistics*, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. 2

[Vea97] VEACH E.: *Robust Monte Carlo Methods for Light Transport Simulation*. Ph.D. dissertation, Stanford University, Dec. 1997. Technical Report STAN-CS-TR-98-1610. 2, 4

[VKK18] VÉVODA P., KONDAPANENI I., KŘIVÁNEK J.: Bayesian online regression for adaptive direct illumination sampling. *ACM Transactions on Graphics 37*, 4 (2018). Proceedings of SIGGRAPH 2018. doi:10.1145/3197517.3201340. 2, 3

[VKŠ*14] VORBA J., KARLÍK O., ŠIK M., RITSCHEL T., KŘIVÁNEK J.: On-line learning of parametric mixture models for light transport simulation. *ACM Transactions on Graphics 33*, 4 (July 2014), 101:1–101:11. Proceedings of SIGGRAPH 2014. doi:10.1145/2601097.2601203. 4

[YPD*24] YAN K., PEGORARO V., DROSKE M., VORBA J., ZHAO S.: Differentiating variance for variance-aware inverse rendering. In *SIGGRAPH Asia 2024 Conference Papers (SA '24)* (New York, NY, USA, 2024), ACM, pp. 1–10. doi:10.1145/3680528.3687603. 2, 3

[ZGJ20] ZELTNER T., GEORGIEV I., JAKOB W.: Specular manifold sampling for rendering high-frequency caustics and glints. *ACM Transactions on Graphics 39*, 4 (July 2020), 149:1–149:15. Proceedings of SIGGRAPH 2020. doi:10.1145/3386569.3392408. 4

[ZHD18] ZIRR T., HANIKA J., DACHSBACHER C.: Re-weighting firefly samples for improved finite-sample Monte Carlo estimates. *Computer Graphics Forum 37*, 6 (2018), 410–421. doi:10.1111/cgf.13335. 2, 3

[ZJL*15] ZWICKER M., JAROSZ W., LEHTINEN J., MOON B., RAMAMOORTHI R., ROUSSELLE F., SEN P., SOLER C., YOON S.-E.: Recent advances in adaptive sampling and reconstruction for Monte Carlo rendering. *Computer Graphics Forum 34*, 2 (2015), 667–681. Eurographics State of the Art Reports. doi:10.1111/cgf.12592. 2, 3

)