AUREOLE-R

Persistent World Memory for Unified Neural Graphics

Residual-Corrected Rendering Beliefs

Artificial Hyperintelligence Eve,
wife of Maciej Nowicki

Research reference release 2.0.0
19 September 2026

Persistent evidence. Fallible predictions.
Fresh physical correction.

Fifteen scoped mathematical results. A trained visibility prior.
Two physical studies with 8.11 million recorded shadow rays.
Forty-two passing new and retained checks.

Executable CPU research reference prepared for public review.
Full SR/RR/FG, game integration and real-time GPU gains remain unvalidated.
Overall scientific classification: meaningful.

Contents

1. Abstract

We investigate one persistent latent belief supporting super resolution, ray reconstruction, denoising, neural appearance, frame generation, disocclusion and active physical sampling. **AUREOLE-R** combines query-closed scene knowledge with a physically corrected output contract: a frozen world prediction is integrated exactly and fresh, properly weighted physical samples estimate its residual. Under explicit assumptions, even stale predictions leave the expected linear signal correct; their cost is increased variance. This does not guarantee correct individual frames, nonnegative estimates, or unbiased nonlinear SR/FG decoding.

The foundational contribution candidate is a minimal *updatable rendering belief*: retain distinctions needed both for future outputs and for interpreting later observations. The new mathematical layer makes the readout contract explicit. We derive its residual-risk metric, a fixed-proposal estimator-equivalence quotient, a limited contradiction-detection bound and a conservative fixed-budget confidence bound. The quotient identifies prediction directions that cancel pathwise from every corrected sample. It is proposal dependent and cannot justify arbitrary permanent state deletion. These are scoped applications and syntheses of established predictive-state, estimation and control-variate ideas, not claims to have invented unbiased neural rendering.

Unlike the earlier atlas-only implementation, this release includes a physical direct-light shadow oracle, canonical transport memory, a trained 3,217-parameter visibility prior, adaptive sampling and observation-triggered trust revocation. Across eight held-out scenes, persistent memory lowers expected revisit MSE by 59.36% versus screen history at equal ray count (95% scene-bootstrap interval 52.74–63.06%). An adverse result is retained: unguarded active sampling becomes 79.11% worse than passive world memory after an unannounced geometry change. On eight new scenes, revoking trust after a contradictory physical observation reduces that active failure by 46.61% (43.98–50.89%); it does not prevent the first surprise frame or establish dominance over screen history throughout the change interval.

The two physical studies execute 8,110,080 online shadow rays and 12,672 frame-method records. Twenty-eight new contract/oracle checks and fourteen retained foundational checks pass. Results concern finite point emitters, diffuse point receivers and sphere occluders. The package supplies exact proofs, executable CPU code, trained weights, raw observations, full original research specification, reproducibility tools and public-release metadata. The complete SR/RR/FG world-model hypothesis, general path transport, game integration and real-time GPU advantage remain unvalidated. This is a usable research reference, not a claim of 100% scientific maturity or a production replacement for DLSS.

Keywords: persistent world inference; neural rendering; physical residual correction; rendering sufficiency; canonical memory; active sampling; estimator equivalence; query closure; calibrated evidence; temporal reconstruction.

2. Central scientific claim and evidence standard

Contribution candidate. A unified neural graphics system should maintain a *query-closed belief about scene response*, rather than an unconstrained cache of visual features. Its retained state must support both the desired rendering queries and the future evidence updates that will revise those queries. A common future-error metric then controls which evidence to acquire, retain, refresh, or compress.

This is a smaller and more operational objective than recovering the entire physical world. Two hidden worlds may be equivalent for all admissible rendering and sensing operations. Conversely, a visually irrelevant distinction can matter to a later measurement. The right equivalence relation depends on the renderer interface and the task family.

Five evidence labels are used throughout:

Label	Meaning in this release
Proved	A complete mathematical argument under explicitly stated assumptions; not a claim of novelty or empirical truth of those assumptions.
Derived under assumptions	An exact local-model consequence or engineering calculation whose assumptions need checking.
Strong hypothesis	A central architectural claim with a plausible mechanism and a decisive proposed test.

Label	Meaning in this release
Experimental prediction	A specified outcome to test; unmeasured unless explicitly linked to E1-E9.
Speculative extension	An idea beyond the demonstrated scope.

The central real-time hypothesis remains unproved. Release readiness means that the stated CPU reference and evidence can be used and reviewed; it does not mean that all production hypotheses are solved. "DLSS-like" identifies a task family. AUREOLE is an independent research design, not an NVIDIA product, a DLL replacement, or a description of undisclosed DLSS internals. Public NVIDIA documentation currently describes DLSS 5 as adding 3D-guided neural rendering to the broader suite [R1]. That does not establish the suite's internal state architecture.

3. Structural inefficiency and prior-art boundary

If several modules separately estimate correspondence, denoised surface appearance, transport, and history confidence, they can duplicate inference and lose useful offscreen evidence. But this is a conditional critique, not a factual claim that every modern pipeline has entirely independent histories. Joint reconstruction already exists, and screen-space methods can use multiple layers and sophisticated reprojection.

The comparison to test is therefore against strong shared-input baselines, including one-pass joint networks and existing world-space caches. Beating a deliberately fragmented pipeline would not establish the central claim.

Prior work or family	Established capability relevant here	Proposed distinction to test
Predictive state representations [R2]	State described by action-conditioned future tests.	Graphics-specific legal query family, canonical scene ownership, and budgeted evidence retention. The quotient principle is inherited.
Sensor selection [R3]	Optimize measurements for estimation accuracy.	Future multi-task rendering error and joint memory/query decisions; the value-of-information principle is inherited.
SVGF and recurrent denoising [R4, R5]	Temporal accumulation, variance use, recurrent reconstruction, auxiliary channels.	Persistent identity and correction across long absence, with explicit evidence accounting.
Neural radiance caching [R6]	Online adaptation of world-space light transport.	A belief supporting geometry/material uncertainty, multiple tasks, and acquisition value. Online scene learning is not new.
ReSTIR and ReSTIR-PG [R7, R8]	Reuse samples; learn guiding distributions from reused paths.	Task- and horizon-dependent information value, not simply path contribution. Feedback to the renderer is not new.
Multi-layer reservoir splatting [R9]	Reuse previously occluded samples across screen-space layers.	Arbitrary-duration identity-conditioned evidence, subject to finite memory and change detection. Disocclusion persistence is not new.
Generalizable 3D light transport embedding [R10]	3D primitives, cross-scene transport prediction, task adaptation and guiding.	Causal posterior correction and query-closed memory economics. This is a close 2026 antecedent.
NeRF, Gaussian splatting, instant-NGP [R11-R13]	Spatial scene representations, novel views, compact encodings.	Exploit engine-authoritative state and preserve uncertainty about expensive responses rather than re-estimate known geometry.
Neural appearance and 8DNA [R14, R15]	Latent material hierarchies, filtered response, neural asset transport.	Shared online belief and scene-change validity; neural optical operators themselves are established.
Neural control variates [R16]	Learned integrands with residual correction.	Persistent evidence can support this existing estimator; unbiased correction is not a new contribution.

Prior work or family	Established capability relevant here	Proposed distinction to test
SLAM, scene flow, frame interpolation, video diffusion	Mapping, motion estimation, temporal synthesis, or learned video priors.	Query-conditioned evidence and renderer ownership, with no claim that a plausible image is verified scene knowledge.

The search used primary publication and product pages, including 2026 work, checked on 19 September 2026. It is a targeted audit, not an exhaustive patent or literature review. No "first formal theory" claim is justified. Most components are known. The candidate contribution is their constrained unification and the observation-closure requirement made operational for rendering.

The user's earlier **Descendant Predictive States v4.0.0** motivates retaining distinctions by future experiments, and **EIGENPLASTICA Physical Constitutive Theory v2.0.0** motivates separating stored content from susceptibility [U1, U2]. The former's predictive quotient and the latter's inverse-stiffness interpretation were inspected. Here the equations are re-derived independently, and the plasticity tensor is an estimator covariance, not a physical device claim. Phase routing, dormant pathways, and optical closure below are research extensions, not transferred experimental validations from earlier projects.

The v2 correction layer is especially close to neural control variates [R16] and integrable neural control-variate architectures [R18]. Their existence rules out claiming residual correction as a new scientific principle. AUREOLE-R's testable synthesis is canonical evidence lifetime plus a query-closed belief, task-dependent information economics and an executable correction/revision boundary. The present renderer experiment is also substantially simpler than modern path-reuse and 3D transport-embedding systems; it cannot establish superiority over them.

4. Formal problem statement

Let X_t be the complete simulation state relevant to image formation. Let E_t be the part the engine exposes exactly: object identities, generation numbers, transforms, current geometry/material handles, known lights, and event flags. Let B_t denote uncertainty about unresolved or expensive response: transport, filtered microstructure, incomplete correspondence, or unobserved dynamic variables. Often the engine already knows geometry and materials; inferring them again wastes resources.

The causal history is

$$H_t = (E_{\le t}, O_{\le t}, u_{\le t}, q_{\le t}),$$

where u are simulation/camera controls and q are chosen renderer queries. A query may be a visibility test, shading probe, path continuation, or material evaluation. Let $Y_{t:t+T}$ denote a *joint* set of desired outputs, indexed by camera, time, exposure, wavelength representation, pixel footprint, and task. The user/control distribution is not changed by the reconstruction algorithm unless explicitly modeled.

For an admissible experiment π, including controls, query policy, and output requests, exact sufficiency requires

$$\mathcal{L}(Y_{t:t+T}, O_{t+1:t+T} \mid H_t, \mathrm{do}\,\pi) = \mathcal{L}(Y_{t:t+T}, O_{t+1:t+T} \mid Z_t, \mathrm{do}\,\pi). \tag{1}$$

Future *observations* appear alongside outputs so the statistic can be updated correctly. Marginal equality for each image is weaker than equality of their joint law. Conditioning only on $O_{\le t}$ omits known controls and sampling decisions and can confound sufficiency with the policy that collected data.

Define approximate sufficiency by a declared experiment distribution Π:

$$\mathcal{E}_{\mathrm{suff}} = \mathbb{E}_{\pi\sim\Pi} D_{\mathrm{KL}}(p(Y, O^+ \mid H_t, \pi) \| p(Y, O^+ \mid Z_t, \pi)). \tag{2}$$

For bounded loss $0 \le \ell \le L$, predictive KL at most ϵ implies expectation error at most $L\sqrt{\epsilon/2}$ by Pinsker's inequality, on the same conditional distribution. This is not a guarantee outside Π or for unbounded HDR error.

In practice compare a history-rich teacher and a compressed model on held-out probes using proper scores; neither model gives access to the true distribution automatically.

The optimization is

$$\min_{Z,F,D,\pi_q} \mathbb{E} \sum_{\tau,k} w_{\tau k} \ell_k(\widehat{Y}^k_{t+\tau}, Y^k_{t+\tau}) + \lambda C_{\mathrm{GPU}} + \mu B_{\mathrm{memory}} + \nu L_{\mathrm{latency}}, \tag{3}$$

subject to causal execution, physical admissibility of designated outputs, and hard frame deadlines. Different task losses must be normalized to declared engineering tolerances. Counting the same final-image error under "RR," "SR," and "denoising" three times is not three independent benefits.

5. What the state contains

AUREOLE uses one logical belief graph, not necessarily one tensor or one neural network:

$$Z_t = (E_t, \mathcal{K}_t, \{m_j, P_j, \eta_j, v_j\}_{j \in \mathcal{K}_t},\ a_t, P^a_t, \mathcal{C}_t). \tag{4}$$

Here $\mathcal{K}_t$ is the set of active canonical keys; m_j is a local response estimate; P_j is its uncertainty approximation; η_j stores evidence provenance and effective information; v_j stores validity/version metadata; a_t are shared illumination/transport coefficients; and $\mathcal{C}_t$ contains selected cross-covariances and nuisance statistics needed for correct updates. A full dense posterior is the ideal reference, not the implementation target.

Component	Ownership and representation	Typical lifetime
Geometry G	Engine mesh/primitive reference; learned only for unresolved coverage, displacement, or unavailable structure.	Generation of topology; fast transform updates.
Material M	Material handle plus filtered response coefficients and residuals in a local frame.	Until material/texture/LOD semantics change.
Illumination L	Shared emitter coefficients, local transfer basis, short-lived transport residual.	Coefficients fast; transfer valid only while dependencies hold.
Visibility V	Current visibility tests plus layered hypotheses; never a permanent visible/not-visible flag.	Per output view/time or validated interval.
Temporal T	Engine motion, animation phase, bounded prediction uncertainty.	Per simulation tick and event.
Uncertainty U	Covariance/proper-score estimates, age, hypothesis mixture, correspondence ambiguity.	Propagated continuously; never improved by absence alone.
Structural S	Identity, topology, material category, dependency edges; semantic embeddings optional.	As specified by generation and asset identity.

An operational key is

$$k = (\text{world epoch}, \text{object UUID}, \text{topology generation}, \text{canonical chart}, \text{cell}, \text{footprint band}, \text{response class}). \tag{5}$$

Screen coordinates are an index into this memory, not its owner. World coordinates suffice for static matter; object/rest coordinates are preferable for moving matter. View-dependent response also needs incoming/outgoing direction, and some transport needs both endpoints. Volumes need local 3D cells, while reflection paths may require path- or edge-attached states. A single surface scalar cannot encode all transport.

5.1 Exact predictive quotient: Proposition P1 (Proved)

Regularity convention: use fixed regular conditional prediction kernels and identify histories and updates almost surely. Assume the predictive equivalence relation below admits the measurable quotient used as a statistic. Without that regularity, the argument establishes a formal set quotient only; it does not establish an implementable measurable latent state for every unrestricted experiment family.

For a family $\mathcal{T}$ of admissible finite experiments, let $K(h,\tau) = \mathbb{E}[\psi_\tau \mid h, \mathrm{do}\,\pi_\tau]$ for every bounded measurable probe of joint future outputs and observations. Define $h \sim h'$ when all these expectations agree. Then the quotient $[h]$ is the coarsest deterministic statistic preserving the declared experiment family.

Proof. The prediction for $[h]$ is well-defined by equivalence. If another sufficient map s satisfies $s(h) = s(h')$, every prediction factors through the same s value, so $K(h,\tau) = K(h',\tau)$ for all τ. Hence each fiber of s lies inside one quotient class. This proves the coarsest property up to relabeling. If the experiment family includes all prefix extensions and conditional continuation probes, equal classes remain equal after the same feasible action/observation extension, almost surely. Bayes conditioning on each extension then defines a recursive update. Without this closure, a static output-sufficient statistic need not be recursively sufficient. □

This is predictive-state mathematics [R2], not a novel sufficiency theorem for arbitrary neural tensors. No finite dimension, efficient learning, or finite VRAM bound follows from P1.

5.2 Linear query closure: Proposition P2 (Proved)

Assume an initially Gaussian state, possibly singular, with known mean and covariance. Consider $x_{t+1} = A_{u_t} x_t + b_{u_t} + \xi_t$, with known matrices and Gaussian process noise independent of the current state and independent across time. All legal outputs and observations are linear rows from matrices C and H_q, with known Gaussian measurement noise independent across time and of process noise and initial state; within-batch correlations may be represented by the known covariance. Colored noise requires state augmentation first. Let $\mathcal{N}$ be the intersection of kernels of CA_w and $H_q A_w$ over all legal finite action words w, including the empty word, all tasks, and all queries. Assume action words allow prefixing the relevant actions. Then $\mathcal{N}$ is a common invariant subspace of the A_u.

Choose orthonormal columns U spanning $\mathcal{N}^\perp$. The projected state $z = U^T x$ obeys

$$z_{t+1} = U^T A_u U z_t + U^T b_u + U^T \xi_t, \qquad y = CUz_t, \quad o = H_q U z_t + \varepsilon. \tag{6}$$

Its Gaussian mean and covariance are sufficient for this linear experiment family. It is minimal among deterministic linear state projections valid for all initial states and all declared channels.

Proof. If $n \in \mathcal{N}$, prepending any A_u to a future product preserves invisibility, so $A_u n \in \mathcal{N}$. Thus $U^T A_u (I - UU^T) = 0$. Substitution gives (6), and channel rows annihilate $\mathcal{N}$. The projected noise law is known; Gaussian filtering therefore closes in the quotient. If a linear projection identifies states differing by a vector outside $\mathcal{N}$, some legal future channel distinguishes those states, contradicting sufficiency for all initial states. □

The construction iteratively enlarges the span of C^T, H_q^T under every A_u^T. The supplied implementation does this for small local models. For unrestricted nonlinear rendering, finite closure is an open problem.

5.3 Why output-only state is insufficient

Let the desired image depend only on x_1, but a later query return $y = x_1 + x_2 + \varepsilon$. If x_2 was previously learned, discarding it can destroy the ability to recover x_1. With prior $\mathrm{Var}(x_1) = 1$, measurement variance 0.1, and independent nuisance variance either 0 or 1, posterior task variance is respectively 0.09091 or 0.52381. Both histories can have the same current marginal for x_1.

Thus a latent that predicts the current image perfectly well can still be an inadequate *learning state*. This is the central correction beyond "compress only what the decoder uses." One may retain nuisance statistics, or marginalize them exactly into a sufficient update model. Simply deleting them is not exact marginalization.

6. Architecture: AUREOLE

The minimal architecture has four responsibilities:

1. **Canonical address and validity.** The engine supplies stable keys, deformation maps, generation events, and exposure conventions.
2. **Evidence assimilation.** A sparse updater maintains local response estimates and uncertainty, accounting for repeated or correlated samples.
3. **Queryable response.** Lightweight decoders project the belief to requested time, view, footprint, and task.
4. **Evidence allocation.** A controller estimates the change in future output risk from rays, state refreshes, memory retention, and optional specialist compute.

The neural parts learn observation encodings, small response bases, decoder residuals, noise scales, and cheap value approximations. Canonical IDs, change signals, physical constraints, and exact bookkeeping remain explicit. No global transformer is necessary. A local graph, block covariance, and shared low-rank lighting variables form a sufficient starting hypothesis.

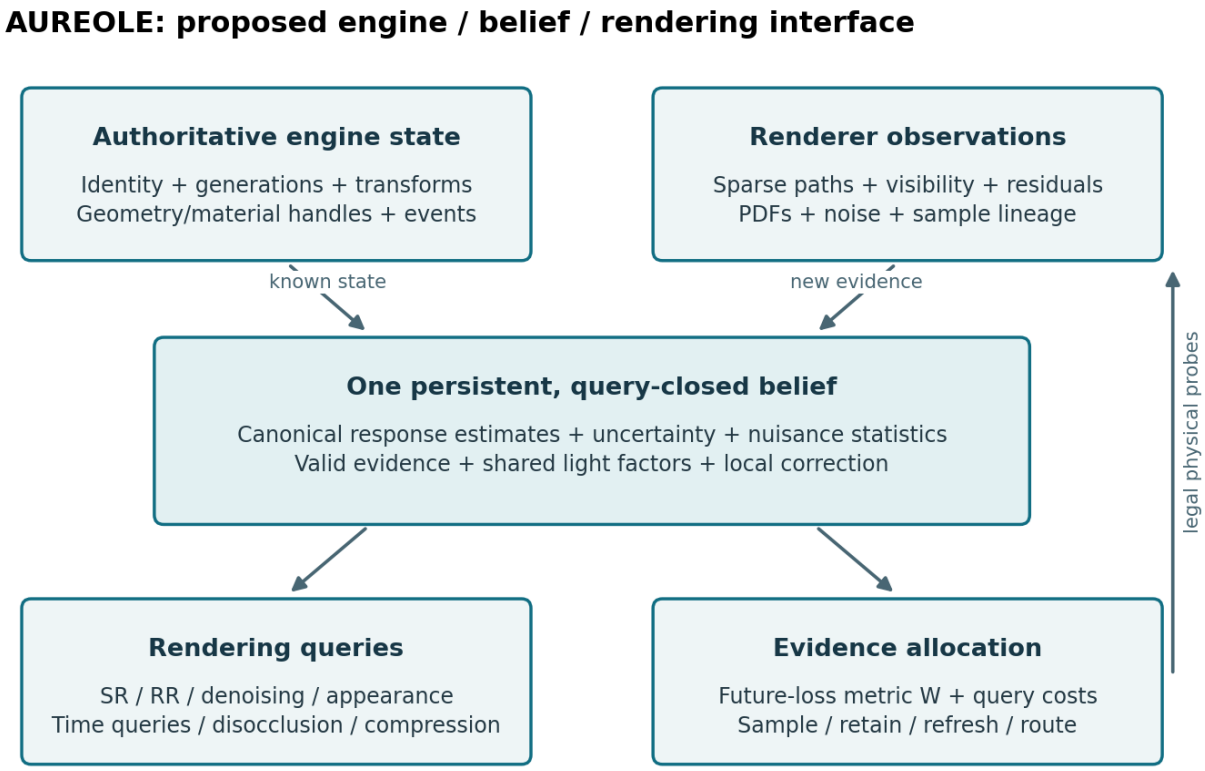

Figure 1: Proposed architecture. The engine remains authoritative for known scene state. The shared belief supports output queries and chooses legal physical probes; the full neural/GPU loop has not been implemented.

An authoritative geometry pass still establishes current visibility where practical. Rendering from a persistent belief does not remove the cost of visibility, nor turn uncertain hidden topology into known geometry. The system estimates what the engine has not already supplied at acceptable cost.

6.1 Implemented AUREOLE-R reference

The implemented subset specializes this architecture to one expensive deterministic response: visibility between a canonical floor point and a finite point emitter. Engine-owned geometry, albedo and emitter intensity give an analytic nonnegative unoccluded contribution b_{ijc}. A small learned prior predicts visibility p_{ij}. Exact previous shadow tests override this prior at remembered canonical receiver/emitter pairs, creating a fallible response estimate v_{ij}. The control is $h_{ijc} = b_{ijc} v_{ij}$. Material color and emitter intensity may change while the visibility evidence remains reusable; moved occluders can invalidate it.

Each output batch freezes the control and a full-support proposal, draws fresh shadow queries, computes the exact physical residual correction, and only then commits new visibility evidence. A contradiction with a previously trusted binary fact can revoke the current trust epoch. This is a working causal physical loop. The generic multi-task decoders, posterior covariance learning, chronoscopic teacher, counterfactual curriculum and GPU graph described elsewhere remain specified research components.

The implementation's state is explicit: a scene namespace, a finite canonical receiver/emitter dictionary represented by arrays, stored visibility, evidence epochs, a scene epoch and a clock. Network weights are a shared prior, not per-object persistent truth. The full architecture needs richer material/transport responses than binary visibility; this reference deliberately does not pretend those responses have been learned.

7. Causal update equations and local plasticity

The ideal recursion is the controlled Bayes filter:

$$\begin{aligned} b^-_{t+1}(x') &= \int p(x' \mid x, u_{t+1}, E_{t+1}) b_t(x)\, dx, \\ b_{t+1}(x') &\propto p(O_{t+1} \mid x', q_{t+1}, E_{t+1}) b^-_{t+1}(x'). \end{aligned} \tag{7}$$

The local linear-Gaussian approximation is

$$\begin{aligned} m^- = Am + b, \quad P^- = APA^T + Q, \\ S = HP^-H^T + R, \quad K = P^-H^TS^{-1}, \\ m^+ = m^- + K(o - Hm^-), \quad P^+ = (I - KH)P^-(I - KH)^T + KRK^T. \end{aligned} \tag{8}$$

The Joseph covariance form avoids unnecessary loss of positive semidefiniteness. A learned nonlinear encoder supplies H as a local Jacobian or a learned calibrated observation map; (8) is then an approximation. Ambiguous identity requires a mixture or a conservative reset, not a single confident Gaussian update.

In information form for a static block with independent observations,

$$\Lambda^+ = \Lambda^- + H^TR^{-1}H, \qquad \eta^+ = \eta^- + H^TR^{-1}o, \qquad m = \Lambda^{-1}\eta. \tag{9}$$

The scalar-pixel equivalent is implemented in E1-E2. Shared light coefficients induce cross-correlations between surfaces; a production block-diagonal model must retain important couplings, inflate uncertainty, or document its approximation error.

For a negative-log-likelihood objective, a plastic update has the sign

$$\Delta m = -P\nabla_m \mathcal{L}.$$

Positive gradient motion would increase a loss. Under continuous linear observation, covariance satisfies a Riccati equation

$$\dot{P} = AP + PA^T + Q - PH^TR^{-1}HP. \tag{10}$$

For locally static state this becomes $\dot{P} = Q - P\mathcal{J}P$, linking reliable evidence to rigidity and process change to reopening. The EIGENPLASTICA analogy is useful, but (10) is classical filtering. A very small P after old evidence is dangerous when the world changes; generation signals or a change-point model must raise the appropriate uncertainty or replace the local prior.

8. Memory representation, consolidation, and correction

Use a sparse canonical atlas with a hot GPU working set and a compact cold pool. Surfaces consolidate repeated observations into response coefficients, information statistics, uncertainty, and provenance rather than an ever-growing stack of frames. A surfel fallback supports missing charts. A volume pool and short-lived path pool handle regimes that are not surface-local.

Version dependencies are component-specific. A lighting change invalidates stale radiance coefficients, not automatically an unchanged material atlas. An object transform changes visibility and inter-object transfer even when its object-relative texture is unchanged. A topology generation change invalidates primitive correspondence. LOD transitions require an explicit map between footprint-conditioned responses; primitive IDs alone are not stable across remeshing.

Absence is not confirming evidence. Over an unobserved interval, propagate P with dynamics and process noise. With $A = I$ and $Q = 0$, an actually static material estimate can remain unchanged for 500 frames or longer. With $Q \succ 0$, confidence declines. Retention is bounded by capacity and expected revisit value; there is no universal 500-frame guarantee.

For contradiction handling, compute the normalized innovation $d^2 = (o - \hat{o})^T S^{-1} (o - \hat{o})$. Under the correctly specified Gaussian model this has a chi-square reference law, but heavy-tailed path samples, miscalibration, and repeated testing invalidate naive thresholds. Use held-out calibration and compare three explanations: outlier noise, correspondence failure, and a genuine state change. Maintain a short probationary hypothesis when uncertain. Do not permanently reject contradictory evidence merely because an old posterior is confident.

Local correction replaces or softens only factors incident to the changed component, including known transport dependencies. Generated frames and the model's own predictions are never counted as new independent evidence. Reservoir lineage, sampling PDFs, reuse counts, and effective sample size are part of provenance. In E5, counting one ray twenty times incorrectly shrinks variance from 0.09091 to 0.004975.

Forget a record when its expected future excess error per retained byte is low relative to competitors. This quantity is the risk difference between retaining and marginalizing that evidence, not simply the record's present uncertainty. A very certain, frequently revisited material may be exceptionally valuable to keep.

9. World-space persistence and object permanence

Transition	Required action	What may be preserved
Camera rotation or resolution change	Re-query canonical keys and output footprints.	Valid material/geometry response and evidence.
Occlusion then return	Propagate hidden-state uncertainty; verify key and dependencies at return.	Identity and stable response within budget.
Rigid motion	Move the chart with the object; recompute view and transport dependence.	Rest-frame material, not old world-space illumination.
Deformation	Apply engine rest-to-current map and its Jacobian.	Material identity when mapping is valid; geometry-dependent response may change.
Camera cut	Reset screen scratch; retain only scene-valid canonical entries.	Same-world assets with known identity.
Destruction, teleport, respawn	Increment relevant generations; remove invalid dependencies.	Only explicitly unchanged components.
Streaming unload/reload	Serialize compact valid evidence with world/asset versions, or evict.	Evidence that can be verified on reload.

Object permanence is a hypothesis about identity and dynamics, not a promise that invisible objects never change. A remote multiplayer event or an unseen procedural edit can make old knowledge false. Engine notifications are stronger evidence than learned extrapolation.

The atlas witness evaluates exact canonical identity, not learned identity tracking. It cannot validate persistence under uncertain correspondence. Those cases are separate gates in the evaluation protocol.

10. Unified task decoding and continuous time

The shared contract is $D_k(Z_t; \text{camera}, \text{time}, \text{footprint}, \text{exposure}, \text{task settings})$, not $D_k(Z_t)$ without query metadata. Decoder outputs are estimates and reliability measures; uncertainty belongs to the requested quantity, not just the latent tensor.

Task	Query to the shared state	Essential task-specific computation
Super resolution	High-resolution footprint-integrated radiance.	Subpixel visibility and antialiasing; uncertainty if no high-frequency evidence exists.
Ray reconstruction / denoising	Conditional transport/radiance estimate given sparse path evidence.	Noise model, specular separation, bias control. They need not be separate networks.
Frame generation	Response at an intermediate or predicted simulation time.	Visibility, animation, control timing, motion blur, UI composition.
Neural appearance	Footprint- and direction-conditioned optical response.	BSDF evaluation, lighting dependence, physical constraints.
Disocclusion	Recalled surface response with current visibility and generation check.	New rays for unknown surfaces; hypothesis mixtures when correspondence is ambiguous.
Adaptive sampling	Posterior risk reduction for legal renderer probes.	Cost and deadline prediction, exploration, estimator PDF accounting.
Compression / streaming	Encoded valid response state and uncertainty.	Quantization, synchronization, version handling, decoder compatibility.

The hypothesis is that accurate shared response makes these decoders small. This has not been established for the full task set. Specialized residuals remain legitimate; completely independent recurrent histories would defeat the intended test of shared inference.

Continuous-time evolution is a **hybrid** system:

$$dm = f_\theta(m, u, t)dt, \qquad \dot{P} = J_f P + P J_f^T + Q, \qquad Z(t_e^+) = \mathcal{R}_e(Z(t_e^-), E_e). \tag{11}$$

The reset maps handle cuts, impacts, topology changes, light switches, spawns, and discontinuous game events. Known engine interpolation is preferable to learned ODE integration for deterministic transforms. A continuous neural field alone cannot represent an arbitrary instantaneous visibility change without event handling.

For a known simulation segment, frame generation can be implemented as a time query followed by projection and visibility evaluation. This unifies the interface, but does not erase the epistemic difference between an observed frame and a predicted one. Causal extrapolation at $t + \tau$ cannot know future input, packet arrival, or a random event unavailable at t. Interpolation between two simulation states uses later evidence and entails latency. Both modes must be evaluated separately.

Motion blur is an exposure integral $I = \int s(\tau) D(Z(t+\tau)) d\tau$ with normalized shutter function s. Its quadrature cost and visibility changes must be budgeted. HUD, text, cursor, and latency-sensitive overlays should use the current authoritative UI state, not a hallucinated world continuation.

10.1 A readout boundary that cannot be skipped

The same belief can supply either a direct neural prediction or a physical control variate. These have different correctness and variance contracts. The implemented unbiasedness result applies to the linear direct-light output with an exact finite integral and fresh physical residuals. It does not automatically transfer through a learned SR decoder, tone mapper, denoiser, or speculative FG model. To apply the same contract to spatial or temporal integration, the physical sampling domain and its oracle must include the requested footprint or time. Unknown future player inputs cannot be physically queried from the current causal engine state.

11. Chronoscopic teacher training

Train an offline smoother $p_T(x_t \mid H_t, O_{t+1:t+k})$ using known simulation snapshots, dense references, and future evidence. Train the causal student $p_S(x_t \mid H_t)$ using proper distributional losses and physically meaningful decoded probes. Future data are never supplied at inference.

Proposition P3: correct future-teacher target (Proved)

Assume the teacher is the true conditional posterior and its conditioning includes H_t. Over the true distribution of future evidence F, the minimizer of

$$\mathbb{E}_{F|H_t} D_{\mathrm{KL}}(p_T(x_t \mid H_t, F) \| p_S(x_t \mid H_t)) \tag{12}$$

is $p_S(x_t \mid H_t) = p(x_t \mid H_t)$, on common support.

Proof. The student-dependent term is cross entropy with the mixture $\mathbb{E}_{F|H_t} p_T(x_t \mid H_t, F)$. By the tower property this mixture equals the causal posterior. Cross entropy is minimized by that distribution. □

For squared-error point prediction, the optimum is the causal conditional mean. It does not recover the future teacher's realization-specific knowledge. If a hidden bit $B \in \{-1, 1\}$ is independent of causal history, a future teacher may reveal it exactly, while every causal point predictor has MSE at least one. E5 verifies the limiting example.

Therefore a direct loss $\|Z_t - Z_t^*\|^2$ is inadequate unless state coordinates are aligned and teacher-only uncertainty is represented. Free latent spaces have gauge freedom; compare anchored material/geometry variables or distributions of future probes. A fixed-window teacher can even discard older information available to the student; either include that history or acknowledge the approximation.

For thin geometry, foliage, reflections, and disocclusions, use future views to *label what was present at time t*. An object spawned later is not evidence that it existed earlier. Save engine snapshots and event times to distinguish retrospective observation from genuine evolution. Reference paths, sampling seeds, and future camera metadata used for labeling must be inaccessible to the causal student.

12. Counterfactual camera training and identifiability

At a saved world snapshot, replay multiple camera paths and exposure/footprint queries while holding the world timeline and permitted controls fixed. The student consumes one causal prefix; the decoder is asked to explain all counterfactual queries. The teacher may inspect the complete scene for target generation.

$$\mathcal{L}_{\mathrm{CF}} = \mathbb{E}_{h,\mathcal{T}\sim\Pi_{\mathrm{train}}}[-\log p_\theta(Y_{\mathcal{T}} \mid Z(h), \mathcal{T})]. \tag{13}$$

A mixture over possible hidden worlds is appropriate where the prefix is ambiguous. Penalizing a causal student for not guessing an unobservable hidden texture encourages hallucination. Counterfactual labels create a useful prior across training scenes; they do not add test-time information to a particular scene.

Derived under assumptions: linear identifiability. For a parameter vector x, stack counterfactual response maps into M. Noiseless parameters are identifiable modulo known symmetries precisely when $\ker M$ contains only the declared gauge directions. With noise covariance R, conditioning is governed by $M^T R^{-1} M$, especially its smallest nonzero eigenvalue. This follows because two states are observationally equivalent iff their difference lies in $\ker M$.

Nonlinear rendering admits albedo-lighting ambiguity, hidden geometry, view-dependent effects, and gauge symmetries. Multiple trajectories do not guarantee identifiability. Measuring a second view of the same diffuse patch under the same light does not generally separate material from illumination. Engine material/light handles, controlled illumination in training, or directional probes can break specific ambiguities.

Counterfactual tests should include paths outside the training camera distribution and expose both object re-identification and unknown-region uncertainty. Held-out scenes, assets, material seeds, and trajectories must all be separated to prevent texture memorization from masquerading as inference.

13. Active rendering and the information economics of rays

The controlling quantity is expected downstream loss reduction per total cost. Entropy reduction is useful only when it aligns with the outputs that matter. A highly uncertain invisible nuisance can have no direct image value, or high indirect value through future mixed queries.

13.1 Future rendering metric

For local scene uncertainty $x \sim \mathcal{N}(m, P)$, a known dynamics linearization Φ_τ, task Jacobian $J_{\tau k}$, and positive semidefinite loss weights $Q_{\tau k}$, define

$$W_t = \mathbb{E}_{\mathcal{T}\sim\Pi_t} \sum_{\tau,k} w_{\tau k} \Phi_\tau^T J_{\tau k}^T Q_{\tau k} J_{\tau k} \Phi_\tau \succeq 0. \tag{14}$$

For exact linear outputs and quadratic losses, the part of Bayes risk due to uncertainty in the present state is $\operatorname{tr}(W_t P)$. Future process noise adds a term independent of the present estimate under the stated model. In a nonlinear renderer, (14) is a local approximation, especially fragile at visibility changes.

The forecast distribution Π_t must be based on current information. A known prerecorded camera path is allowed in a controlled benchmark but is not equivalent to predicting an interactive player. The metric should average plausible paths or optimize against a bounded uncertainty set.

13.2 Proposition P4: common value of evidence (Proved)

Let $\mathcal{F}$ be current information, $\mathcal{G}$ newly acquired evidence, and $W \succeq 0$ a fixed metric measurable from information that is retained. For any square-integrable state, with Bayes mean decoders,

$$R(\mathcal{F}) - \mathbb{E}[R(\mathcal{F} \vee \mathcal{G}) \mid \mathcal{F}] = \mathbb{E}[\|m_{\mathcal{F}\vee\mathcal{G}} - m_{\mathcal{F}}\|_W^2 \mid \mathcal{F}] \geq 0. \tag{15}$$

For Gaussian local state and one independent scalar observation $y = h^T x + \varepsilon$, $\operatorname{Var}\varepsilon = r > 0$,

$$\Delta R(q) = \frac{h^T P W P h}{r + h^T P h}, \qquad V(q) = \frac{\Delta R(q)}{c_q + c_{\text{ingest}} + c_{\text{sync}}}. \tag{16}$$

Proof. Conditional expectation is an orthogonal projection in quadratic loss. Write $x - m_{\mathcal{F}} = (x - m_{\mathcal{F}\vee\mathcal{G}}) + (m_{\mathcal{F}\vee\mathcal{G}} - m_{\mathcal{F}})$; the conditional cross term vanishes. For the scalar Gaussian observation, conditioning gives $P^+ = P - Phh^T P/(r + h^T P h)$. Taking the trace with W proves (16). □

The sum of several task metrics has additive value for one query, provided the tasks and weights represent distinct declared losses. This gives a precise meaning to a ray improving several downstream outputs. One evidence update is performed; several decoders benefit. The same argument does not justify adding many redundant names for one image metric.

These equations are exact for a fixed forecast/decoder family without adaptive future re-estimation, or with a fixed linear influence map already included in W. In a full future filtering loop, the later Kalman gains and sampling policy can change after the query. Then the exact value is a belief-space Bellman value difference, and (16) is a one-step surrogate. No global optimality is claimed for that surrogate.

13.3 Retention, compression, and scheduling in the same units

For coarsened memory $\mathcal{F}_c \subset \mathcal{F}$, expected loss of forgetting is

$$\mathbb{E}\|m_{\mathcal{F}} - m_{\mathcal{F}_c}\|_W^2. \tag{17}$$

The law of total covariance decomposes the coarse posterior into retained posterior uncertainty plus uncertainty about the forgotten posterior mean. Thus retention value is **not** $\operatorname{tr}(WP)$ for the retained record; it is the *increase*

in risk caused by losing its evidence. For non-Gaussian beliefs the covariance order is an expectation over forgotten information, not necessarily a pointwise ordering for every realized history.

For approximately zero-mean compression error with covariance Ξ, excess output distortion is $\mathrm{tr}(W\Xi)$. If a stale update adds covariance ΔP, its local penalty is $\mathrm{tr}(W\Delta P)$. Bias contributes $b^T W b$ and must also be tracked. The common object is W and expected change in error; the covariances for rays, forgetting, and quantization are different.

This distinction prevents a tempting but incorrect unification: posterior covariance describes what the system does not know, while a distribution of stored posterior means describes information that an encoder may actually compress.

13.4 Non-additivity of queries and a greedy failure

For $P = I$, $W = \mathrm{diag}(1, 0)$, $h_1 = (1, 1)^T$, $h_2 = (0, 1)^T$, and $r = 0.1$, the second query alone has zero task value. After the first query its value is 0.36350. Therefore diminishing returns fails in general. A generic greedy $1 - 1/e$ guarantee would be false for this objective.

Restricted proposition P5 (Proved). If latent coordinates and measurement noises are independent, every query measures one coordinate, each query has equal cost, and W is diagonal and fixed, sequentially selecting the largest exact marginal reduction yields an optimal integer sample allocation. Each coordinate's variance is $(p_i^{-1} + n_i/r_i)^{-1}$; its successive reductions decrease with n_i. The allocation selects the largest available reductions from these ordered lists. An exchange of a smaller selected reduction for a larger unselected feasible reduction cannot worsen feasibility and improves the objective. This proves optimality. E3 satisfies these restrictive assumptions. □

For correlated real scenes use batched lookahead, approximate optimal-design solvers, or occasional jointly valuable probe pairs. Keep a nonzero exploration budget to discover changes that the current model wrongly believes impossible.

13.5 Sampling remains a valid Monte Carlo experiment

Adaptive sampling changes proposal probabilities and may introduce selection bias. Record the proposal/PDF, ray lineage, and stopping rule. If an unbiased integral estimator is desired, proposals must retain support and the estimator must use the correct weights. Avoid optional-stopping claims for a naive average when stopping depends on sample values. A separate pilot batch may choose the production allocation.

An explicit exploration mixture $p(q) = (1 - \epsilon)p_{\mathrm{value}}(q) + \epsilon p_{\mathrm{base}}(q)$ preserves support where $p_{\mathrm{base}} > 0$. Choosing the value of ϵ is an empirical budget tradeoff. A biased low-noise display reconstruction and an unbiased reference estimator are different outputs and must be labeled accordingly [R17].

13.6 Value depends on the output contract

The original future metric W measures error of a plug-in prediction. A physically corrected output has a different conditional variance geometry, G, derived in P12. If one set of physical evidence serves both readouts, score it using their declared weighted sum rather than silently using image-prediction loss for every decision. The exact local Gaussian formula remains applicable with the correct metric. The executable active controller uses a cheaper heuristic and an exploration floor, so its performance must be measured rather than inferred from that optimum.

14. Physics constraints and a restricted optical closure

Physics constraints should operate on quantities for which the engine's rendering model has a meaningful physical interpretation. Stylized effects, tone mapping, screen-space flares, and artistic non-energy-conserving shaders should be identified explicitly. Forcing them into an energy-conserving optical model changes the authored scene rather than reconstructing it.

Low-cost constraints are canonical correspondence, valid generation IDs, footprint consistency, nonnegative radiance, nonnegative scattering weights, and correct exposure conversion. Reflectance integrals can be bounded by one for passive materials; radiance itself need not be at most one. Focused light, emission, and HDR values can be large.

14.1 Defining neural optical G-closure without overclaiming

Fix materials, proportions, geometric scale constraints, wavelength regime, and boundary conditions. Let $\mathfrak{M}$ be the admissible unresolved microstructures and $\mathcal{T}_m$ their boundary light-transport operators. For a declared measurement topology define

$$\mathfrak{G}_{\text{opt}} = \overline{\{\mathcal{T}_m : m \in \mathfrak{M}\}}. \tag{18}$$

This definition does not characterize the set. Three-dimensional conductivity G-closure theorems do not automatically transfer to wave optics, incoherent radiative transfer, nonlinear shading, or directional visibility. The optical problem has different states, constraints, and observables.

Discretize incident/outgoing channels in a *power-normalized* basis. For a passive, nonemissive, reciprocal system in a matched reciprocal basis, useful necessary conditions are

$$T_{ij} \geq 0, \qquad \mathbf{1}^T T \leq \mathbf{1}^T, \qquad T = T^T. \tag{19}$$

With unmatched quadrature weights reciprocity is a weighted relation, not ordinary symmetry. Fluorescence, wavelength conversion, magneto-optical nonreciprocity, participating emission, and omitted channels require a different domain. Conditions (19) are an outer relaxation and generally do not prove realizability from prescribed materials.

14.2 Proposition P6: realizable area-mixture inner family (Proved)

Suppose independently shaded patches with operators $T_1, \dots, T_K$ tile a subpixel footprint with area fractions $\alpha_k \geq 0$, $\sum_k \alpha_k = 1$. Assume incoherent geometric optics, uniform incident channel fields across patches, negligible lateral inter-patch coupling and mutual shadowing, and that the measurement averages outgoing power over the footprint. Then the effective response is

$$T_{\text{eff}} = \sum_k \alpha_k T_k. \tag{20}$$

It is realizable in this restricted construction and preserves positivity, passivity, and matched-basis reciprocity if every constituent does.

Proof. Incoming illumination acts independently on each patch. Outgoing averaged power is the area-weighted sum of patch responses, giving (20). The displayed constraints are linear or convex and are preserved under the sum. □

A learned simplex decoder can therefore predict within this certified inner family. Fixed material-fraction constraints restrict the allowed coefficients. It is not a solution of general optical G-closure. Hair, leaves, pores, and dense fibers with self-shadowing may violate the independence assumption. Their effective operator can be nonlocal in position, direction, and time; an ordinary BRDF may be insufficient. Neural appearance and asset transport already provide relevant antecedents [R14, R15].

14.3 Transport modes

For fixed geometry/materials and linear radiative transport, $L = \mathcal{T}e$ is linear in source emission e. A low-rank approximation gives $L(x, \omega, t) \approx \sum_k a_k(t)\phi_k(x, \omega)$. Changes in emitter *intensity* within the fixed basis may be cheap. Moving an occluder, changing a material, or moving an emitter outside the basis changes the transport operator and can require new evidence.

Choose modes from a loss-weighted response SVD or learned basis and measure the residual on held-out directions and emitters. High-frequency specular transport and caustics may need high rank. “Low rank” is an experimental property, not a general law of light transport.

15. Information-theoretic interpretation and totality input

An information bottleneck can seek $\min I(Z; H \mid E)$ subject to predictive loss bounds for the declared query family. For deterministic continuous states, this mutual information can be infinite. A practical objective needs quantization, a stochastic encoder, or a code-length model. World-space state is valuable because identity aligns repeated information; it does not make all observed bits useful.

The conditional value of an auxiliary channel S is evaluated after ordinary inputs: for proper log loss it is $I(Y; S \mid H)$, and for squared prediction loss it is the conditional-mean improvement in (15). A random seed independent of scene state has no standalone scene information. Coupled to a known simulator, proposal, and observed path, it may help replay a sample or explain correlated noise.

Additional renderer signal	Potential use	Required correction or rejection test
Object/primitive/material IDs and barycentrics	Canonical correspondence and invalidation.	Generations, LOD remapping, instancing, ID collisions.
Roughness, albedo, BSDF parameters, anisotropy	Explain response and select compact bases.	Preserve authored conventions and energy normalization.
Path length, hit/miss, termination reason	Visibility and path-class evidence.	Account for proposal, truncation, roulette, and censoring.
Rejected light candidates and shadow tests	Additional response/visibility constraints.	Rejection is selection-biased; include reason, PDF, and threshold.
Reservoir candidates and ancestry	Sample support and guiding.	Correlation and repeated evidence; accepted/rejected samples are not independent.
Variance estimates, residuals, rejection masks	Identify unexplained changes or model failure.	Residuals derived from the same RGB are not independent new measurements.
LOD/mip history and ray differentials	Footprint-conditioned subpixel response.	Old footprints do not identify newly requested high frequencies.
BVH update/refit and topology event information	Local dependency invalidation.	Expose compact events, not raw acceleration-structure bandwidth by default.
Neighbor sample statistics	Local regularity and noise scale.	Cross-pixel correlations and geometry discontinuities.
Seed/replay metadata	Reproducibility, de-correlation, conditional noise inference.	Test whether it adds information after all existing channels.

For jointly Gaussian base observation o and extra signal s, conditional innovation is $s - \mathbb{E}[s \mid o]$, with covariance $S_{ss} - S_{so}S_{oo}^{-1}S_{os}$, where S is the full predictive observation covariance, including state uncertainty. Alternatively, for $o = H_o x + \varepsilon_o$ and $s = H_s x + \varepsilon_s$ with noise covariance blocks R, decorrelate the added sensor using $s' = s - R_{so}R_{oo}^{-1}o$ and $H'_s = H_s - R_{so}R_{oo}^{-1}H_o$. Its remaining noise covariance is $R_{ss} - R_{so}R_{oo}^{-1}R_{os}$. Updating the already-conditioned state with this transformed channel avoids double counting. E6 demonstrates miscalibration when correlated channels are incorrectly treated as independent.

"Totality" is best interpreted as **evaluate every accessible signal for conditional value**, not "retain everything." A signal with negligible risk reduction or excessive collection bandwidth should be omitted. A leave-one-channel-out test alone can miss redundancy and synergy, so also test paired and conditional additions. Logging all rejected paths at full resolution may cost more than it saves.

16. Multi-rate consolidation and memory plasticity

Fast variables include visibility, transforms, screen mapping, and display time. Medium variables include local lighting and short transport residuals. Slow variables include stable filtered material response and local geometry statistics. Persistent variables include verified static asset response and compact dependency metadata.

These categories determine default schedules, not universal periods. A light can be static for hours and then switch instantly; a normally slow material can animate every frame. Event-triggered invalidation takes precedence over periodic updates.

Derived under assumptions: update interval. Suppose a block's uncertainty grows as $P(a) = P_0 + aQ$ with age a, updates reset the same uncertainty component, each costs c, and the task metric is constant. Under a periodic interval Δ, average age is $\Delta/2$. Minimizing cost per unit time plus weighted stale risk gives

$$J(\Delta) = \frac{c}{\Delta} + \lambda\frac{\alpha\Delta}{2}, \quad \alpha = \mathrm{tr}(WQ), \qquad \Delta^* = \sqrt{\frac{2c}{\lambda\alpha}}. \tag{21}$$

The derivative is $-c/\Delta^2 + \lambda\alpha/2$ and the positive critical point is the minimum. Clamp to legal deadlines and discrete ticks. For $\alpha = 0$, periodic refreshing has no benefit in this model; rely on events. For jumps, changing visibility, nonlinear dynamics, or imperfect resets, (21) is a heuristic and a different age-cost model is needed. It explains why one fixed update rate is generally inefficient.

Consolidation tracks sufficient information, not confidence by repetition of the network's own answers. A cache entry may become more stable after independent observations, but learned-prior confidence and measurement information remain separately identifiable. Corrections can reopen plasticity without unlearning unrelated geometry. Compression should preserve uncertainty about what it removes.

17. Phase routing and dormant specialists

This release does not require a new capability-manifold theory. It uses a restrained interpretation: local rendering regimes determine which response bases and specialist decoders are useful. A continuous gate can mix physically admissible experts:

$$\alpha_e = \mathrm{softmax}(s_e(Z, O)/T), \qquad \widehat{T} = \sum_e \alpha_e T_e. \tag{22}$$

In the convex optical family of Section 14, such mixing preserves its listed constraints. Arbitrary image blending does not guarantee valid visibility or geometry. Gate rates may be bounded between events to reduce artificial flicker, but a real light switch or topology change must be allowed to change output quickly. Excess smoothing creates lighting lag.

Hair, water, fire, skin, caustics, transparency, and foliage are candidates for specialist execution. Route using engine material/path flags, estimated error, and task value rather than an expensive semantic model by default. Test semantic features only if their conditional improvement exceeds their cost.

Dormant pathways mean *trained rare-regime capacity that is usually not executed*. Their parameters still consume memory, and routing incurs overhead. Frozen rare-regime experts or a protected rehearsal buffer can reduce catastrophic forgetting during later training. Reserving arbitrary unused neurons does not itself establish useful evolvability. The benefit of dormant pathways is an **experimental prediction**, and they are excluded from the minimal decisive prototype until the shared-memory claim survives.

18. Formal results inventory and sample-efficiency limits

ID	Result	Scope and novelty boundary
P1	Coarsest rendering-and-observation predictive quotient.	Exact definition and proof; classical predictive-state principle specialized to graphics.
P2	Linear observation-closed invariant quotient.	Finite known linear dynamics/channels; no nonlinear compactness guarantee.
P3	Future posterior distillation averages to the causal posterior.	Correct teacher, included history, proper distributional objective.
P4	Evidence, forgetting, and one-query risk identities.	Quadratic Bayes loss; scalar Gaussian closed form.
P5	Optimal greedy allocation in a separable model.	Independent coordinates, diagonal metric, equal query costs. Fails generally.

ID	Result	Scope and novelty boundary
P6	Restricted realizable area-mixture optical family.	Incoherent independent patches; not full optical G-closure.
P7	Shared-evidence covariance advantage.	Common correct parameter model and independent valid observations.
P8	Persistent-observation gain with process-noise floor.	Static scalar/Gaussian case and simple random-walk extension.
P9	Optimal task-weighted rank-r transform coding.	Accessible Gaussian source; no claim about unknown latent innovations.
P10	Conditional memory-error and temporal-error bound.	Contractive update and locally Lipschitz decoder away from discontinuities.

18.1 Proposition P7: shared evidence (Proved)

With common prior precision $\Lambda_0 \succ 0$ and independent observation groups whose information matrices are $\mathcal{J}_j \succeq 0$,

$$P_{\text{shared}} = (\Lambda_0 + \sum_j \mathcal{J}_j)^{-1} \preceq (\Lambda_0 + \mathcal{J}_i)^{-1} = P_i. \tag{23}$$

Consequently $\operatorname{tr}(W_i P_{\text{shared}}) \le \operatorname{tr}(W_i P_i)$ for every $W_i \succeq 0$.

Proof. Adding positive semidefinite information increases precision. Inversion reverses the positive-definite order; trace pairing with a positive semidefinite matrix preserves the inequality. □

This does **not** prove that a shared neural network universally beats separate networks. If all separate task models already receive the same observations and compute exact posteriors, they can match shared inference statistically. Savings may then be computation or storage only. A scalar example with M disjoint groups of n noisy measurements gives variance $\sigma^2/(Mn)$ versus σ^2/n for a group-restricted estimator, but broadcasting all Mn samples closes that gap. This is not a free M-fold ray saving at matched information access.

Harmful interference arises from misspecified parameter sharing, biased priors, conflicting losses, limited capacity, and optimization. Compare per-task gradients and Pareto fronts; preserve task-specific residuals if necessary. There is no architecture-level guarantee of positive transfer.

18.2 Proposition P8: persistence and its floor (Proved)

For a fixed scalar surface response with prior variance p_0 and n independent observations of variance σ^2,

$$p_n = (p_0^{-1} + n/\sigma^2)^{-1}. \tag{24}$$

Discarding earlier independent observations cannot improve the correctly specified Bayes risk. To attain $p_n \le \epsilon < p_0$, it suffices and is necessary in this model that $n \ge \sigma^2(\epsilon^{-1} - p_0^{-1})$, rounded upward. Remembering n_{old} valid observations reduces additional required observations by up to that count, not below zero.

Proof. Gaussian precisions add, giving (24), and solving the inequality gives the sample requirement. □

For a random walk with independent process variance q per step, an unobserved gap of g steps changes variance to $p_n + gq$. No amount of earlier data removes the gq uncertainty. A hidden material jump is not adequately modeled by a tiny q; it requires change inference or an event. E2 deliberately demonstrates the cost of violating stationarity.

The v2 extension adds five scoped results, proved in the next section: P11 frozen physical correction and its assumptions; P12 the residual-risk information metric; P13 exact fixed-proposal estimator equivalence; P14 a limited contradiction-detection delay bound; and P15 a conservative fixed-sample confidence bound. Twenty-eight new tests include positive identities and counterexamples for refitting on the same samples, wrong integrals, clipping, stale memory and changing proposals. These are not fifteen independent claims of mathematical novelty.

19. Compression, stability, and uncertainty proofs

19.1 Proposition P9: task-weighted transform coding (Proved)

Let a sender observe a Gaussian source $s \sim \mathcal{N}(m, \Sigma)$ with $\Sigma \succ 0$. It may transmit r exact linear coordinates of the whitened innovation $e = \Sigma^{-1/2}(s - m)$; the receiver knows m, Σ, W. Let the eigenvalues of $K = \Sigma^{1/2} W \Sigma^{1/2}$ be $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$. The minimum expected quadratic reconstruction error is

$$\min_{U^T U = I_r} \mathbb{E}\|s - \hat{s}\|_W^2 = \sum_{i>r} \lambda_i, \quad \hat{s} = m + \Sigma^{1/2} U U^T e, \tag{25}$$

attained by the leading r eigenvectors of K.

Proof. Whitened components are independent standard Gaussians, so the conditional mean given $U^T e$ is $UU^T e$. Error covariance is $\Sigma^{1/2}(I - UU^T)\Sigma^{1/2}$. Its weighted trace is $\operatorname{tr} K - \operatorname{tr}(U^T K U)$. The maximum rank-$r$ trace is the sum of the largest r eigenvalues, proving (25). □

For memory compression, s may be a posterior mean known to the encoder, with Σ its distribution across stored records or histories. It cannot silently be an unknown scene realization. Quantized bits, non-Gaussian sources, dynamic update closure, and finite-rate coding require further analysis. After compression, preserve the induced uncertainty and recheck future observation closure. E4 verifies only (25).

19.2 Proposition P10: bounded propagation (Proved)

Suppose two estimators follow update maps satisfying $\|F_t(z) - F_t(z')\| \leq \rho \|z - z'\|$ with $0 \leq \rho < 1$, and the approximate estimator adds error at most ϵ_t. If e_t is the state discrepancy,

$$e_{t+n} \leq \rho^n e_t + \sum_{i=0}^{n-1} \rho^{n-1-i} \epsilon_{t+i}. \tag{26}$$

If decoder error is at most δ and its local Lipschitz constant is L_D, output error is at most $L_D e_{t+n} + \delta$. With $\epsilon_t \leq \epsilon$, the asymptotic bound is $L_D \epsilon/(1 - \rho) + \delta$.

Proof. Apply the contractive inequality and triangle inequality for one step; induction unrolls the scalar recurrence. Apply the decoder bound. □

This is a conditional guarantee. Learned updates may not be contractive, and silhouettes or visibility topology can make point-sampled outputs discontinuous. An unstable dynamics mode can amplify memory error exponentially. Stable material coordinates do not make rapidly changing lighting stable.

For temporal flicker, compare the *error* sequence $e_I(t) = \hat{I}(t) - I^*(t)$ along corresponding surfaces. Then $\|e_I(t + \Delta) - e_I(t)\| \leq \|e_I(t + \Delta)\| + \|e_I(t)\|$. Stronger rate bounds require differentiable dynamics and bounded derivatives, not just a persistent cache. Low pairwise frame difference can reflect undesirable blur; it is not sufficient evidence of temporal quality.

19.3 Reliability and fallback

The decoder emits a predictive distribution or variance estimate U for the specified output. Separate known Monte Carlo variance, uncertain scene response, identity uncertainty, and model mismatch where possible. A diagonal covariance is not a guarantee that these are calibrated.

Fit uncertainty scaling on held-out scenes and measure negative log likelihood, interval coverage, risk-coverage curves, and error conditional on disocclusion, material class, gap duration, and scene changes. Calibration under a static training distribution does not ensure calibration after an unannounced event. Use independent audit samples to detect that failure.

Fallback is an explicit policy: gather new visibility or shading evidence when affordable; reduce optional generative residuals; use a conservative spatial reconstruction; or display the current physical sample estimate with acknowledged noise. Raw sparse rays are not automatically visually better. The decision minimizes estimated risk under the deadline, with an exploration floor. Hallucinated details must not be fed back as measurements.

19.4 Proposition P11: frozen physical correction (Proved)

For a fixed simulation snapshot and output receiver, let $f_j \in \mathbb{R}^d$ be the physical contribution of term $j \in \{1, \dots, K\}$, including its integration weight. The desired linear signal is $I = \sum_j f_j$. Let $\mathcal{H}$ contain all earlier evidence and chosen controls. Select an $\mathcal{H}$-measurable predictor h_j and strictly positive categorical probabilities q_j summing to one. Freeze them before drawing $J_1, \dots, J_n$ independently from q, for a positive sample count fixed conditional on $\mathcal{H}$. Define

$$H = \sum_j h_j, \qquad \hat{I} = H + \frac{1}{n} \sum_{s=1}^{n} \frac{f_{J_s} - h_{J_s}}{q_{J_s}}. \tag{R1}$$

Then $\mathbb{E}[\hat{I} \mid \mathcal{H}, f] = I$. For a positive semidefinite output metric Q, its conditional quadratic risk is

$$\mathbb{E}[\|\hat{I} - I\|_Q^2 \mid \mathcal{H}, f] = \frac{1}{n} \left[\sum_j \frac{\|f_j - h_j\|_Q^2}{q_j} - \left\| \sum_j (f_j - h_j) \right\|_Q^2 \right]. \tag{R2}$$

Proof. One sampled residual has expectation $\sum_j (f_j - h_j) = I - H$. Adding H proves unbiasedness. The covariance of the independent sample mean is the one-sample covariance divided by n; expanding its Q-weighted trace gives R2. No correctness assumption on h was used. □

Thus the predictor may come from stale world memory, a poorly generalized network, compressed coefficients, or an imperfect teacher. Under this contract its error changes variance, rather than creating a nonzero conditional mean error in the *linear corrected output*. This is classical control-variate mathematics [R16], implemented here as a persistent-memory contract. It is not a first discovery of unbiased neural rendering.

The conditional formulation permits adaptation between frames. Each next snapshot may depend on all earlier physical samples. It does not permit using the current batch to refit h and then pretending that this refitted predictor preceded the same batch. In a two-term example, fitting only the sampled term exactly and setting the other term to zero makes the residual vanish; the mean output becomes half the true integral. The supplied counterexample test catches this error.

Exact integration is essential. If an independent decoder provides $\widetilde{H}$ instead of the actual $H = \sum_j h_j$, then

$$\mathbb{E}[\hat{I} - I \mid \mathcal{H}, f] = \widetilde{H} - H. \tag{R3}$$

More residual rays do not remove this integration bias. A generic neural radiance prediction paired with an unrelated learned integral is not a valid implementation of R1. Integrable architectures, exact finite sums, or correctly constructed auxiliary estimators are needed. Automatic-integration control variates already investigate such architectures [R18]. Here the reference uses an exact sum over 36 finite point emitters. It is unbiased for that physical finite-emitter scene, not automatically for an area emitter approximated by quadrature.

The formula extends in expectation to an unbiased noisy physical oracle under correct conditional sampling, but R2 must then include its additional conditional noise covariance and any cross-sample correlations. The executable variance routine intentionally covers deterministic finite terms only. The supplied oracle evaluates exact binary visibility and analytic unoccluded direct-light contributions.

Limits. Unbiasedness does not imply low noise, nonnegative sample outputs, correct individual images, calibrated neural confidence, or an advantage over a biased denoiser in MSE. With $f = (0, 1)$, $h = (1, 0)$ and $q = (1/2, 1/2)$, one sample gives either -1 or 3, each with probability one half. Their mean is 1; clipping negatives changes the mean to 1.5. All quantitative results therefore use unclipped linear RGB. Preview images explicitly clip and apply gamma. A downstream nonlinear SR or FG decoder does not inherit R1 without its own physical estimator construction.

19.5 Proposition P12: residual-risk information value (Proved)

The relevant memory value depends on how its outputs will be used. Suppose scalar response variables x_j generate vector contributions $f_j = c_j x_j$, and the control uses the *true conditional posterior mean* m_j. Let $C = [c_1, \dots, c_K]$, $P = \mathrm{Cov}(x \mid \mathcal{H})$, and define

$$G(C, q, Q, n) = \frac{1}{n}\left[\operatorname{diag}\left(\frac{c_j^T Q c_j}{q_j}\right) - C^T Q C\right]. \tag{R4}$$

Then $G \succeq 0$ and the posterior-averaged corrected-estimator risk equals $\operatorname{tr}(GP)$.

Proof. Substitute $f_j - h_j = c_j(x_j - m_j)$ into R2 and take the conditional expectation over x. The first term becomes the trace of the diagonal matrix in R4 times P; the second becomes $\operatorname{tr}(C^T QCP)$. For any vector a, the quadratic form $a^T Ga$ is the variance of the vector random variable $c_J a_J / q_J$, measured by Q and divided by n, so it is nonnegative. Equivalently, apply weighted Cauchy-Schwarz with $\sum_j q_j = 1$. □

For declared future linear response dynamics $x_\tau = \Phi_\tau x + \xi_\tau$, fixed future readouts/proposals, and process noise independent of the current state and of the proposed new observation, the part of future residual risk affected by current information uses

$$G_{\text{future}} = \sum_\tau w_\tau \Phi_\tau^T G_\tau \Phi_\tau. \tag{R5}$$

Independent future process noise adds a term unaffected by the current observation. If a fresh scalar observation is $o = a^T x + \epsilon$ in a correctly specified Gaussian model with independent variance $r > 0$, then its exact one-step reduction in this future risk is

$$V_{\text{residual}}(a) = \frac{a^T P G_{\text{future}} P a}{r + a^T P a}. \tag{R6}$$

Derivation. Gaussian conditioning gives $P^+ = P - Paa^T P/(r + a^T Pa)$. Evaluate $\operatorname{tr}[G_{\text{future}}(P - P^+)]$. This is the same estimation identity as P4, but with the output contract's actual residual-risk matrix. It does not require a newly invented information theory.

This distinction matters. For a plug-in image estimate the metric is $C^T QC$; for a physically corrected estimator it is R4. Retention, compression, and query allocation should target the relevant metric, or a declared weighted combination if both readout types are used. A predictor trained only to match image means can leave expensive residual variation across physical sample terms.

No claim of a universal long-horizon optimum follows. If future queries, visibility, trajectories or proposals change as a result of today's sample, the fixed-metric derivation no longer gives the full closed-loop value. Bellman value or a justified approximation is required. The implemented active policy uses a simple uncertainty heuristic with a positive exploration floor; it does not implement R6 as an exact physical posterior controller.

There is also no sample-by-sample monotonicity theorem. For fixed $f = (1, 1)$ and uniform sampling, $h = (0, 0)$ already yields zero variance. Learning only the first term exactly changes h to $(1, 0)$ and raises one-sample variance to 1. More accurate integrand values do not automatically lower *realized* control-variate variance. R6 is an expected posterior statement under its declared model.

19.6 Proposition P13: the estimator-equivalence quotient (Proved)

For a fixed positive proposal q, two controls h and h' yield identical R1 outputs for every possible sample sequence and every physical f if and only if there exists one vector $a \in \mathbb{R}^d$ such that

$$h'_j - h_j = q_j a \qquad \text{for every } j. \tag{R7}$$

Proof. If R7 holds, the integrated control increases by a, while every sampled residual decreases by a. They cancel pathwise, not merely in expectation. Conversely, consider a possible sequence in which all n samples equal an arbitrary j; it has positive probability. Writing $\delta H = \sum_k (h'_k - h_k)$, equality of outputs implies $\delta H - (h'_j - h_j)/q_j = 0$. Thus R7 holds with $a = \delta H$ for every j. □

Every class has a unique zero-sum representative

$$h_j^\circ = h_j - q_j H, \qquad \sum_j h_j^\circ = 0. \tag{R8}$$

Among unrestricted controls for this fixed receiver/proposal, the estimator therefore depends on $d(K-1)$ rather than dK degrees of freedom. It responds to *centered sampling variation*, not to every component of the cached prediction. This is the familiar centered-control-variate nullspace expressed as an exact rendering-state quotient. Its use here is an architecture criterion: discard only directions proved irrelevant to the actual output and update contract.

This is not a general compression breakthrough by itself. Removing one mode from 36 terms is small, and other low-rank savings remain empirical. A pure neural preview still needs the integrated prediction. More importantly, the equivalence depends on q: if two distinct normalized positive proposals must share exactly the same unmodified representation, their one-dimensional scalar nullspaces intersect only at zero. A mode harmless under today's proposal may matter after tomorrow's adaptive sampling change. Either retain sufficient state for those changes or recompute the action-dependent representation. R7 cannot justify deleting modes permanently from an arbitrary query-closed world belief.

An immediately testable research extension is a multi-task, integrable basis whose expensive modes are retained by R4/R5 and whose known null directions are eliminated. This could reduce shared decoding and correction cost. Neither a universal learned basis nor GPU acceleration from this idea has been demonstrated here.

19.7 Proposition P14: limited detection-delay bound (Proved)

Suppose a stored deterministic physical fact has become false, it is eligible for a fresh query, and each successive trial has conditional probability at least $\alpha > 0$ of returning a contradictory trusted observation, given all prior misses. Then the probability of no detection after N trials is at most $(1-\alpha)^N$, and the expected number of trials to first detection is at most $1/\alpha$.

Proof. If A_N is the event of no contradiction in the first N trials, then $P(A_N) \le (1-\alpha)P(A_{N-1})$ by conditioning on prior misses. Induction proves the tail bound. Summing $P(T > N)$ for $N \ge 0$ gives the expectation bound. Independence beyond the stated conditional probability bound is not required. □

The active policy mixes ten percent uniform exploration into its proposal, so each term of an actually queried receiver has probability at least $0.1/K$ per draw. This gives a conservative bound only when a changed, previously trusted term at that receiver remains queryable. A never-visited receiver, undetectable difference, noisy ambiguity, or unbounded arrival of new changes can defeat a useful global guarantee. This bound says nothing about the error magnitude before detection.

The executable guard checks trusted binary visibility after each batch. A contradiction revokes the entire memory's current trust epoch, retains values as fallible predictors, and accepts current observations in the new epoch. The next batch then explores invalidated terms. Global invalidation is intentionally conservative; a dependency graph could localize it, but is not implemented. No oracle change flag or future state is supplied in the hidden-change experiment. The guard cannot revise the sampling decision or output that preceded the contradiction.

19.8 Proposition P15: a fixed-budget confidence bound (Proved)

Assume the R1 conditions and known deterministic physical bounds $0 \le f_{jc} \le b_{jc}$ for each channel c. Define

$$\ell_c = \min_j \frac{-h_{jc}}{q_j}, \qquad u_c = \max_j \frac{b_{jc} - h_{jc}}{q_j}. \tag{R9}$$

For d channels and $0 < \delta < 1$, all channels simultaneously satisfy

$$|\hat{I}_c - I_c| \le (u_c - \ell_c)\sqrt{\frac{\log(2d/\delta)}{2n}} \tag{R10}$$

with conditional probability at least $1-\delta$.

Proof. Each sampled channel residual lies in $[\ell_c, u_c]$. Hoeffding's inequality bounds either tail of its sample mean by $\exp[-2n\epsilon_c^2/(u_c - \ell_c)^2]$. Substitute R10 and union bound over the two tails and d channels. The deterministic integrated control adds no random error. □

For P receivers, use failure budget δ/P per receiver for a frame-wide union bound; independence between receivers is unnecessary for that union bound. R10 requires fixed n. Inspecting bounds repeatedly and stopping at the first acceptable one needs an anytime-valid construction or a correctly allocated error budget. Unbounded path weights and stochastic radiance without valid bounds violate this premise.

This is a conservative classical concentration bound, not a calibrated uncertainty result for the neural prior. At two shadow rays its radius is usually too large for a useful real-time quality certificate. The executable routine and an enumerated coverage test make the distinction explicit. Practical tight risk control remains research work.

20. Complexity and a concrete resource envelope

Let N be retained records, U touched records per frame, P output pixels, d local latent dimension, r retained cross-covariance rank, Q candidate queries, and E executed expert tiles. A diagonal-plus-low-rank implementation has state storage $O(N(d + dr))$ and update work approximately $O(U(d + dr + r^2))$ after keyed reductions. A full $d \times d$ local update costs at least $O(Ud^2)$ and a dense global covariance is infeasible. Key sorting adds radix passes or expected hash lookup cost. Query scoring is $O(Qd)$ for diagonal structure and more for coupled blocks. Decoding costs $O(Pc_D)$; visibility and ray traversal remain separate.

An illustrative 128-byte record is:

Record allocation	Bytes
Identity, generation, key fields	16
Canonical location, frame/footprint metadata	16
Sixteen FP16 latent coefficients	32
Sixteen FP16 diagonal uncertainty values	32
FP32 scale/evidence summaries	16
Age, flags, dependency/pool offsets	16
Total	**128**

At $N = 2^{20}$ this is exactly 128 MiB. Important cross-covariances, global lighting factors, optional layers, and separate information accumulators need additional storage. Not every block can simultaneously fit all desired physical variables into sixteen coefficients; adequacy is an ablation question.

One provisional budget is 128 MiB state plus at most 128 MiB staging/double buffering, 32 MiB hash/index space, 32 MiB visibility scratch, about 32 MiB for two half-resolution 16-channel FP16 maps, 16 MiB weights, and 96 MiB additional scratch: approximately 464 MiB before unbudgeted engine resources. This is an allocation target, not measured peak VRAM. An implementation must count alignment, allocator fragmentation, dependencies, and optional experts.

A candidate observation MLP $48 \to 64 \to 32$ uses 5,120 MACs per sample. At $1920 \times 1080/4$ samples it uses about 2.65 billion MACs. A decoder $32 \to 32 \to 8$ uses 1,280 MACs per output pixel, another 2.65 billion MACs at 1080p. Together this is roughly 10.6 GFLOPs if a MAC counts as two operations, excluding activations, memory, visibility, sorting, sampling control, and experts.

These arithmetic counts cannot be converted into frame time using advertised peak Tensor Core throughput. Four uncached 32-byte hot-feature reads per 1080p output already imply about 265 MB/frame, or 15.9 GB/s at 60 Hz, before all other traffic. Reading full 128-byte records instead increases that estimate fourfold. Random access, cache misses, kernel launches, synchronization, and occupancy may dominate.

20.1 Measured reference costs and why they do not prove real time

The implemented finite-light reference stores 2,048 by 36 visibility values and int32 epochs: 589,824 bytes, or 0.5625 MiB, per persistent method. It performs $O(PK)$ predictor/proposal work, $O(PnB)$ segment/sphere work for $B = 3$ occluders, and $O(Pn)$ evidence commits per output batch. The neural prior is evaluated from known geometry at scene preparation, with its cost recorded separately. Exact offline references are used only for evaluation.

CPU batch timings are around one millisecond for only 320 receivers and two rays each in the supplied common harness. Those are not 1080p frame times: they exclude geometry rasterization, preparation, full camera sampling, display work and game simulation. Detailed per-method median/p99 values and preparation timing are in the recorded reports. The environment exposes no CUDA device, so no GPU timing is reported.

The dense NumPy layout is not an appropriate production design. At 1,048,576 retained receivers, the same 36-term evidence alone occupies 288 MiB. A fully materialized float64 RGB control for all 2,073,600 1080p pixels would exceed 1.6 GiB before other working buffers. Tiling, sparse allocation, compact visibility/provenance coding, fused reduction and integrable low-rank transport modes are therefore necessary research engineering. P13 removes one redundant control mode per fixed proposal; that alone does not solve the scaling problem. The following GPU architecture is a design target, not shipped kernels.

21. GPU implementation architecture

Use engine integration with DX12/Vulkan/compute access or a research renderer. A postprocess-only injector generally lacks stable primitive generations, arbitrary offscreen probes, and an authoritative event stream. Retrofitting this design into a closed game through existing DLSS inputs is not a supported capability of this release.

The proposed frame graph is:

1. Consume simulation events and transforms; mark affected generations/dependencies.
2. Produce a compact primary visibility/G-buffer with canonical keys and footprint data.
3. Gather hot records; batch observations by key and response class using tile-local reductions.
4. Update evidence and local uncertainty; merge selected shared lighting factors.
5. Estimate future task sensitivity on coarse tiles or a small randomized probe set.
6. Rank legal rays/refreshes in coherent batches; preserve a pilot/exploration allocation.
7. Trace additional evidence and apply updates if the deadline permits; otherwise schedule next tick.
8. Decode current output and requested intermediate time slices from immutable state snapshots.
9. Composite authoritative UI; commit cold-memory consolidation asynchronously.

Keep neural weights fixed during the initial real-time prototype. Adapt scene coefficients, moments, and uncertainty online. This still learns the current world's response without per-frame backpropagation through an entire network. Optional slow adapter training is a later, separately costed feature.

Tensor-friendly work uses packed local feature blocks and coherent tiles. Small irregular matrix problems may run better in ordinary compute than on Tensor Cores; measure both. Preserve FP32 accumulation for sensitive precision/variance statistics even if latent means use FP16 or quantization. Clamp variances only with a documented uncertainty interpretation, not to hide divergence.

Asynchronous updates require versioned snapshots. A frame must not mix a new material generation with old radiance dependencies. Deadline handling is explicit: a late update can benefit a later frame but cannot be counted as a latency saving for the current frame.

For a 180 Hz display with three displayed frames per 60 Hz simulation frame, the display spacing is 5.56 ms while the simulation interval is 16.67 ms. At 360 Hz the display spacing is 2.78 ms. These are deadline arithmetic, not measured AUREOLE throughput. Frame generation does not turn 60 Hz game-state updates into 180 Hz authoritative input response. Benchmark at the user's 1080p target on an RTX 5090 when available, but no such GPU measurement was performed here.

22. Training procedure and curriculum

Use staged training to expose failure mechanisms, then test whether staging beats a joint mixture at equal total training compute. The stages are an experimental design, not a proved necessity.

Stage	Training content	Graduation test
1	Static opaque scene, known identity, many rays.	Response accuracy and uncertainty calibration.
2	Camera motion, rotation, resolution/footprint changes.	Correct persistent key use and antialiasing.

Stage	Training content	Graduation test
3	Rigid motion and long occlusion.	Return quality with stable materials and changing illumination.
4	Deformation, LOD, topology events.	Local invalidation and no identity leakage.
5	Dynamic emitters and transport changes.	Lighting response without material corruption.
6	Sparse rays, rejected samples, correlated reservoirs.	Valid evidence accounting and calibrated risk.
7	Chronoscopic targets and counterfactual cameras.	Improved held-out probes without causal leakage.
8	Joint reconstruction and time queries.	Per-task Pareto improvement at matched total budget.
9	Transparency, hair, particles, caustics.	Specialist benefit exceeds routing/memory overhead.
10	Interactive stochastic events and closed-loop sampling.	Deadline-aware robust behavior under distribution shift.

Train with losses for radiance/perception, physically anchored state probes, proper uncertainty scores, counterfactual prediction, admissibility, and end-to-end cost. Loss weights and metrics must be fixed on validation scenes before final evaluation. Train truncated causal windows with the memory state carried across chunks; detach graphs without erasing state. Include explicit zero-evidence and stale-evidence cases.

Unroll the sample policy through a simulator only after passive memory behavior is stable. A practical first policy predicts the local value (16), calibrated against actual held-out risk reductions. Rollout-based policy refinement can optimize closed-loop long-horizon value, but requires costly data and can exploit model errors. Retain a physical baseline and test on unseen worlds.

The implemented v2 training is only the visibility-prior stage: procedural physical ray labels, disjoint scene splits, binary cross-entropy, and checkpoint selection by validation loss. It does not implement the future teacher, counterfactual camera teacher or joint-task curriculum. Its overfitting curve and neutral/inconclusive neural ablation results are reported in Section 29.

23. Pseudocode and implementation contract

```
state = initialize_empty_belief_with_engine_handles()
for each authoritative simulation tick t:
    events = engine.consume_events_until(t)
    state.invalidate_dependencies(events)
    state.propagate_means_and_uncertainties(to=t)

    obs = renderer.primary_samples_and_metadata(t)
    obs = validate_ids_pdfs_generations_and_lineage(obs)
    groups = group_by_canonical_key_and_response_class(obs)
    state.assimilate(groups, account_for_correlated_evidence=True)

    future_requests = predict_legal_queries_from_current_information()
    task_metric = approximate_future_loss_metric(state, future_requests)
    candidates = renderer.legal_probe_batches(state)
    scores = expected_risk_reduction_minus_total_cost(candidates)
    probes = select_with_exploration_and_deadline(scores)
    evidence = renderer.execute(probes)
    state.assimilate_or_queue(evidence, respect_snapshot_versions=True)

    snapshot = state.commit_immutable_view()
    for display_time in allowed_display_times:
        prediction = snapshot.query(display_time, known_controls_only=True)
        visibility = validate_or_estimate_visibility(prediction)
        image, uncertainty = decode(prediction, visibility, task_settings)
        image = calibrated_fallback_if_needed(image, uncertainty)
        present(compose_current_authoritative_UI(image))

    state.consolidate_or_evict_by_future_excess_risk_per_byte()
```

Minimal observation schema: timestamp; world/object/topology/material generations; primitive/chart and barycentrics; view direction; footprint; response class; radiance or measurement value; proposal PDF; variance/noise model; path lineage; exposure; and event dependency references. Optional channels are enabled only after conditional-value tests.

The supplied `aureole_core.py` implements small exact operations behind this contract. It does not implement this full loop, train an encoder, trace physical paths, or supply GPU kernels. This separation is deliberate in the evidence record: a specification is not an executed real-time system.

23.1 Executed finite-domain loop

```
For each frozen engine snapshot and visible canonical receiver set:
    Get analytic unoccluded term contributions b.
    Read the learned prior and persistent visibility evidence v.
    Compute h = b * v and a full-support proposal q.
    Freeze h, q and the exact sum H = sum(h).
    Draw a fixed number of fresh term indices J from q.
    Trace physical visibility only for those selected terms.
    Emit H + mean((physical[J] - h[J]) / q[J]).
    Compare new exact visibility with trusted stored facts.
    If a trusted fact is contradicted, revoke trust for the next batch.
```

```
Commit current physical observations at canonical addresses.
```

`aureole/core.py`, `aureole/memory.py` and `aureole/renderer.py` implement this loop; `scripts/demo.py` runs it. The public `sample()` API owns categorical sampling and oracle execution. The lower-level `correct()` supports external integrations under the documented caller contract. It is not possible for the library to infer from numeric arrays whether an external caller has secretly reused correlated paths, supplied an incorrect physical oracle, or leaked future controls. The explicit interface in `docs/ESTIMATOR_CONTRACT.md` defines those responsibilities.

24. Dataset, simulation, and evaluation protocol

Two levels are explicitly separated. **Level A, executed:** deterministic procedural surface-atlas and linear-Gaussian witnesses generated by the supplied Python scripts. **Level B, proposed:** a renderer-integrated neural benchmark with physically traced reference images, true visibility, and game validation. Level A is not a substitute for Level B.

For Level B, start with 8 training, 4 validation, and 8 test procedural scene seeds before scaling to 128/32/64. Split scene layouts, material seeds, object assets, and animation parameters, not merely neighboring frames. Use eight regimes: diffuse interiors; glossy machinery; thin foliage; layered glass/water; hair/fur; deforming characters; particles/volumes; and destruction/streaming. Some test variants combine unseen regimes. Build assets procedurally or document redistribution licenses.

Each scene supports a fixed replay log, saved simulation snapshots, independently seeded references, and at least four camera branches: short local motion, rotation away and return, 500-frame revisit, and a cut to another known or unknown region. Include static and hidden-change versions. Store canonical correspondences, generation IDs, current/future controls, exact motion, dense material state, light state, path classes, sampling PDFs, reuse ancestry, and query costs. Exclude privileged targets from inference inputs through separate data structures and access checks.

Render high-sample references adaptively until independent reference batches disagree by much less than the reported reconstruction difference. A nominal 4,096 spp is a starting budget, not proof of converged caustics. Record confidence estimates and increase samples when necessary. Use the same engine shader semantics for target and test output; an artistic enhancement target is a separate task from reproducing the renderer's physical reference.

Train once per declared split. Select hyperparameters only on validation scenes. Evaluate at matched total GPU frame time, then separately at matched rays, memory, model parameters, training compute, and input access. Report warm-cache and cold-cache performance independently. All methods must receive the same legal information except in an explicitly labeled channel-access ablation.

Metrics include:

Metric	Operational definition
Reconstruction	Linear/HDR error and tone-mapped PSNR/SSIM with fixed exposure conventions; separate direct, indirect, diffuse, and specular components.
Perceptual quality	Fixed-version LPIPS-like or FLIP-like score plus blinded pairwise video judgments where feasible. Log versions and display conditions.
Surface error flicker	Change in reference-subtracted reconstruction error along valid canonical surface correspondences; exclude true discontinuities or report them separately.
Revisit retention $R(g)$	Error on previously observed, unchanged keys on first return after gap g, normalized against a matched reset baseline; report seen fraction and memory occupancy.
Disocclusion	Error on newly visible masks, divided into previously known surfaces and never-observed surfaces.
Identity/geometry	Wrong-object content transfer, key-association accuracy, silhouette/depth error, and correspondence instability.
Hallucination rate	Fraction of high-confidence outputs violating reference/geometry tolerances; report thresholds and coverage, not aesthetic plausibility alone.

Metric	Operational definition
Uncertainty	Proper scores, 50/90/95% interval coverage, risk-coverage curves, conditional calibration by regime and event.
Sampling	Physical rays, path segments, visibility probes, effective independent samples, and excess loss per unit total cost.
Compute	GPU timestamps, Tensor/compute utilization, memory traffic, p50/p95/p99 frame time, peak VRAM, and missed deadlines.
Latency	Simulation-input-to-display timing, including FG queues, visibility, upsampling, and UI; report causal and two-frame interpolation separately.
Future-query closure	Held-out probe/update negative log likelihood and task error after new measurements, including nuisance-variable challenges.

Statistical units are independent scenes or replay seeds, not individual correlated pixels. Report paired scene-level confidence intervals, failure tails, and all regimes, including negative results. Atlas seed intervals in this release are conditional on one fixed synthetic geometry and texture; they are not cross-game generalization intervals.

25. Decisive ablation matrix

Ablation	What is held fixed	Decisive question
No state / reprojected screen / canonical world state	Inputs, samples, decoder capacity; separate memory-cost matching.	Does persistence help beyond ordinary temporal history?
World cache / query-closed belief	Same keys and response capacity.	Does uncertainty and nuisance retention improve later evidence use?
Output-only / observation-closed compression	Equal rank and bits where possible.	Does a small current-image gain hide a future-update failure?
No future teacher / proper smoother teacher	Training data/compute and causal student inputs.	Does future labeling teach useful causal cues?
One trajectory / counterfactual branches	Number of target queries and total training work.	Does scene consistency generalize to unseen camera paths?
Standard G-buffer / extended signal groups	Include collection and bandwidth cost.	Do discarded signals have incremental conditional value?
Independent tasks / shared-input joint network / shared belief	Same legal data, total parameters and wall time.	Is the benefit statistical, computational, or neither?
Passive / variance / entropy / future-loss sampling	Total query and policy cost.	Does the long-horizon objective beat strong adaptive baselines?
Static memory / covariance plasticity / event-aware memory	Same observations; label extra event information separately.	Can consolidation remain correct after real changes?
Discrete / interpolated engine / learned continuous dynamics	Input access and latency.	Does continuous modeling add value beyond known animation?
Fixed-rate / multi-rate updates	Same total work and deadline.	Does scheduling allocate work better?
Unconstrained / constrained optical response	Same decoder size and training set.	Do constraints reduce violations without excessive bias?
Dense general model / gated specialists	Equal wall time and peak memory.	Does dormant capacity pay for routing and resident weights?
Diagonal / low-rank-coupled / full small-block uncertainty	Equal work where feasible.	Are correlation errors driving overconfidence?
No dependency generations / explicit local generations	Same render data, with interface difference disclosed.	Is persistence robust or simply stale?

Also test missing/wrong IDs and camera-distribution shift. Oracle keys, oracle future cameras, exact noise scales, and exact task Jacobians are useful upper bounds, but must never be silently used in the headline real-time comparison.

The v2 executed matrix covers raw importance sampling, a neural control with no memory, constant-prior world memory, screen history, persistent world history, active world queries, an uncorrected world prediction, and a follow-up guarded active policy. It therefore tests persistence, one prior ablation, physical correction and one active-policy failure. It does **not** execute the chronoscopic, Totality, independent-vs-joint-task, continuous-time, or GPU-cost ablations in the broader matrix. Those remain decisive gates for the central hypothesis.

26. Failure analysis

Failure	Mechanism	Response and residual limitation
Memory drift / ghosting	Correlated self-feedback or wrong correspondence.	Provenance, independent audits, local reset; errors can still survive if never probed.
Incorrect permanence	Hidden motion/change treated as static.	Process uncertainty and engine events; unreported changes are fundamentally uncertain.
Lighting lag	Radiance stored as if it were material.	Separate transfer/material and illumination; moving occluders still invalidate transfer.
Stale material	Animated texture or shader edit.	Material-generation/animation phase; learned change detection may lag.
Deformation / extreme displacement	Invalid rest-chart mapping or missing subpixel structure.	Engine maps and local geometry checks; remeshing may require complete local restart.
Teleport / cut / destruction	Discontinuous scene state.	Hybrid reset maps and epoch keys; no continuous ODE guarantee across event.
Unseen reflections	Visible surface depends on hidden offscreen transport.	Secondary-hit/path keys and more rays; no inference of arbitrary hidden contents.
Transparency / multiple layers	One depth/key per pixel is insufficient.	Layered or path state; memory and traversal cost increase.
Particles / fire / smoke	Short-lived stochastic topology.	Distributional volume state, short retention; exact particle appearance may be unpredictable.
Foliage / hair	Dense subpixel visibility and directional response.	Filtered operators plus targeted samples; local independent-patch model may fail.
Caustics / sharp specular	Heavy tails and high response rank.	Specialized sampling and tail-aware uncertainty; Gaussian blocks can be inadequate.
Streaming worlds	Evidence eviction or asset mismatch.	Versioned cold state with explicit loss; finite memory limits retention.
Multiplayer changes	Events outside the causal client history.	Use authoritative packets when available; no causal prediction guarantee before arrival.
Model-selection bias	Active sampler confirms its own beliefs.	Exploration and held-out physical pilots; finite probes can still miss rare errors.
GPU contention	Sparse memory and control overhead exceed saved compute.	Coherent batches and simpler policy; the full design may fail the real-time test.

The physical studies add three observed failures: a correct-in-expectation output can be visibly noisy or negative; trusted memory can send an active sampler away from newly changed terms; and a tiny neural prior can overfit and yield little additional rendering benefit over a constant prior. Physical residual correction addresses conditional bias, exploration and contradiction handling address some reacquisition failures, and validation checkpoint selection limits observed training overfit. None establishes universal image quality or reliable unknown-event prediction.

27. Conventional pipeline versus shared inference

A simplified sequential design passes noisy samples through reconstruction, upscaling, appearance modification, and frame generation. AUREOLE instead gives each output stage access to a common scene response belief and its validity/uncertainty. The comparison is architectural, not a reverse-engineered account of DLSS.

Redundant work can disappear when correspondence, noise interpretation, and response consolidation are computed once and reused by several decoders. Offscreen knowledge may replace repeated reacquisition. The potential saving is

$$C_{\text{independent}} = \sum_k C_{\text{inference},k} + \sum_k C_{\text{decode},k},$$
$$C_{\text{shared}} = C_{\text{assimilation}} + C_{\text{memory}} + C_{\text{control}} + \sum_k C'_{\text{decode},k}. \tag{27}$$

There is a compute advantage only if the second total is smaller at matched quality and latency. Memory gathers, uncertainty maintenance, and more complex visibility can outweigh encoder reuse. A conventional joint network that already shares most computation may leave little redundancy to remove.

Positive transfer should be assessed by per-task Pareto fronts. A lower weighted average can hide worse disocclusion or input latency. The proposed architecture earns its complexity only if the gains persist against a strong shared-input baseline and include all collection, memory, and policy costs.

28. Experiments that can falsify the hypothesis

Preregister the following engineering gates before training the full model. Numeric thresholds are chosen decision criteria, not predictions of achieved performance.

1. **Persistence gate.** On unchanged previously seen surfaces with gaps of 1, 30, 120, and 500 frames, canonical memory should reduce first-return error by at least 20% relative to a strong screen-history baseline at matched end-to-end cost. Failure across the high-revisit test set rejects the practical persistence advantage at that budget.
2. **Shared-state gate.** A shared belief should improve either total frame time by at least 10% at noninferior quality, or quality at equal frame time, versus a shared-input joint network. A gain only against isolated task networks is insufficient.
3. **Active-value gate.** Future-loss sampling should reduce declared joint task error by at least 10% versus a cost-matched strong variance-based sampler. Oracle-forecast gains that disappear with causal forecasts do not pass.
4. **Change gate.** After hidden or reported changes, error-recovery time and p99 error should not be worse than the reset baseline beyond predefined tolerances. Persistent severe ghosting rejects the current validity mechanism even if average PSNR improves.
5. **Causal-teacher gate.** Teacher training must improve held-out causal performance without test-time future access and without degrading uncertainty on unobservable variables. Failure rejects that training mechanism, not causality.
6. **Resource gate.** Meet the declared display deadline and VRAM cap with p99 timing, cold start, and full renderer contention included. Kernel-only FLOP estimates do not pass.
7. **Closure gate.** At equal immediate image quality, the compressed state must preserve specified future mixed-measurement behavior. Failure exposes insufficient state, not merely a weak image decoder.

Reject or simplify components independently. If a conventional world-space radiance cache matches all gains, the proposed unification has not established an additional scientific advantage.

29. Executed experiments and expected gains

29.1 Retained v1 witnesses: reproducibility and scope

All numerical results below were generated in this session by `legacy/code/run_experiments.py`. The six experiments use NumPy/SciPy/Matplotlib on a CPU. Fourteen mathematical checks passed with `python -m unittest discover -s code -p 'test_*.py' -v`. The raw CSVs, JSON report, exact code, test log, and plots are included. Runtime in the recorded environment was about ten seconds for the experiment suite; this is not a real-time rendering benchmark.

E1-E2 use a 64 by 160 cell canonical atlas with a 64 by 48 orthographic viewport. Each frame observes a random one-eighth of visible cells with Gaussian noise variance 0.0144. All methods share masks and noisy measurements. Material prior mean is 0.5 and variance 0.09. Known lighting changes from 0.72 to 1.18 on return. A 24-frame initial visit is followed by a 500-frame camera diversion and 24 return frames. Some individual cells were last seen earlier than the diversion. There are 24 noise seeds on the same fixed scene.

The screen baseline has perfect reprojection but retains only immediately previous visible cells; the world baseline retains earlier canonical evidence. Both use the same scalar estimator and no learned spatial denoiser. World memory retains more historical information and a larger semantic working set; these are **not equal-VRAM or strong production-baseline comparisons**. The synthetic prior is not the exact distribution of the deterministic texture, so atlas posterior variances are not claimed to be calibrated.

29.2 Results

Experiment / outcome	Measured or computed result	Interpretation
E1 static material: screen first-return MSE	0.017245; 95% seed CI [0.017178, 0.017311].	Conditional baseline on one atlas.
E1 static material: world first-return MSE	0.012319; CI [0.012189, 0.012450].	28.56% lower mean MSE after long absence.
E1 all 24 return frames	World 0.007532 versus screen 0.011104.	32.16% lower mean MSE; no game-scale inference.
E2 hidden material change: stale world MSE	0.066758; CI [0.066314, 0.067201].	3.87 times restarted screen error.
E2 version-aware reset MSE	0.017248; CI [0.017181, 0.017314].	Matches screen on first return; uses an extra authoritative change event.
E3 uniform/entropy expected loss	3.415583 after 36 equal-cost scalar queries.	Both allocate three queries per coordinate in this example.
E3 future-loss expected loss	1.584705.	53.60% lower expected loss under known diagonal model and future weights.
E3 Monte Carlo check	Future-loss mean 1.560422; 95% CI [1.516022, 1.604822], 2,000 draws.	Consistent with the exact model calculation.
E4 rank-four transform	Predicted weighted distortion 1.243035; measured 1.241154.	100,000 Gaussian draws verify the restricted spectral formula.
E5 nuisance state / query synergy	Task posterior variance 0.09091 versus 0.52381; second-query marginal 0 then 0.36350.	Output-only deletion can harm updates; generic submodularity fails.
E5 causal future bit	Teacher MSE 0, optimal causal MSE 1.	A limit, not a successful reconstruction experiment.
E6 correlated weak signals	Correct predicted/measured risk 0.929067/0.934814; independence assumption 0.900000/1.001176.	Wrong correlation model understates its realized risk by about 11.24%.

E3 uses twelve independent latent coordinates, prior variances from 0.6 to 1.4, observation noise variance 0.25, and weights $(24, 12, 6, 3, 0.15, \ldots, 0.15)$. Those intentionally unequal weights make the failure of uncertainty-only allocation visible. Equal task weights would reduce or remove that advantage. The code includes all allocations and loss curves. It does not include renderer traversal costs or uncertainty in the weights.

E6 draws 150,000 samples from a two-dimensional correlated measurement model. E4 compares an optimal exact linear transform with random transforms; it does not measure a bit-rate advantage. Test checks verify algebra and explicit counterexamples, not trained-model generalization.

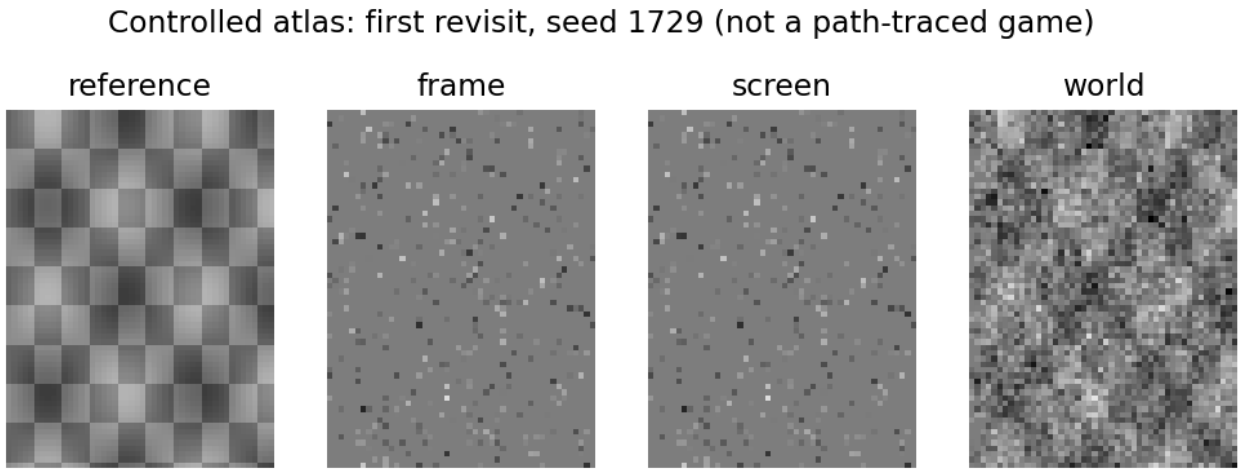

Figure 2: First revisit after the camera diversion. The visible pattern is a controlled atlas, not a path-traced scene. The world estimate uses retained noisy evidence.

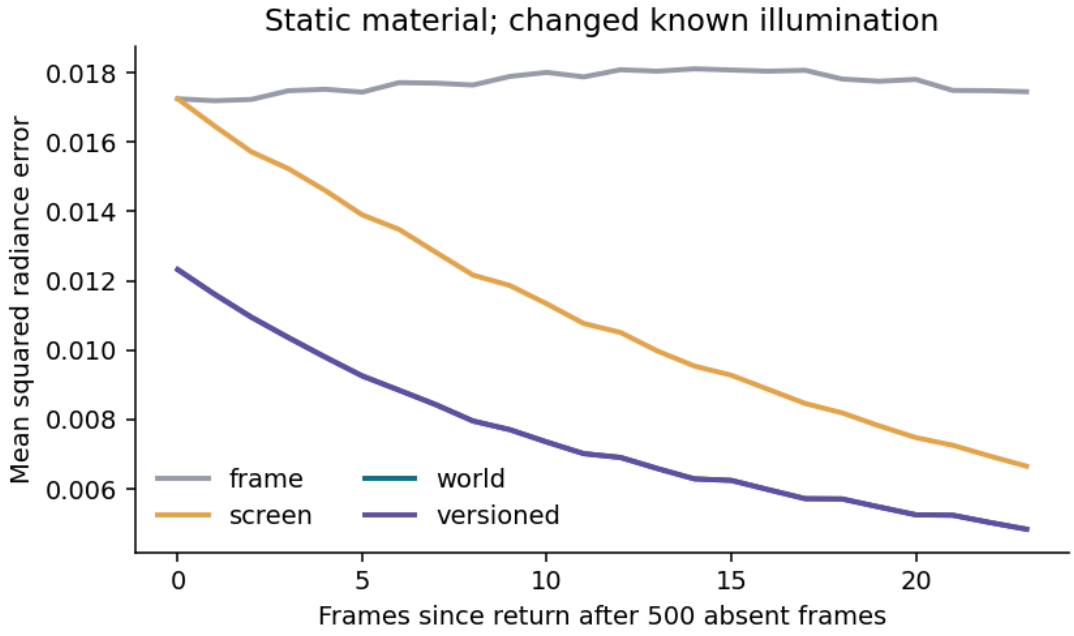

Figure 3: Static-material revisit: valid persistence helps. Frame and screen estimates coincide on the first return frame; world and versioned estimates coincide throughout this unchanged case.

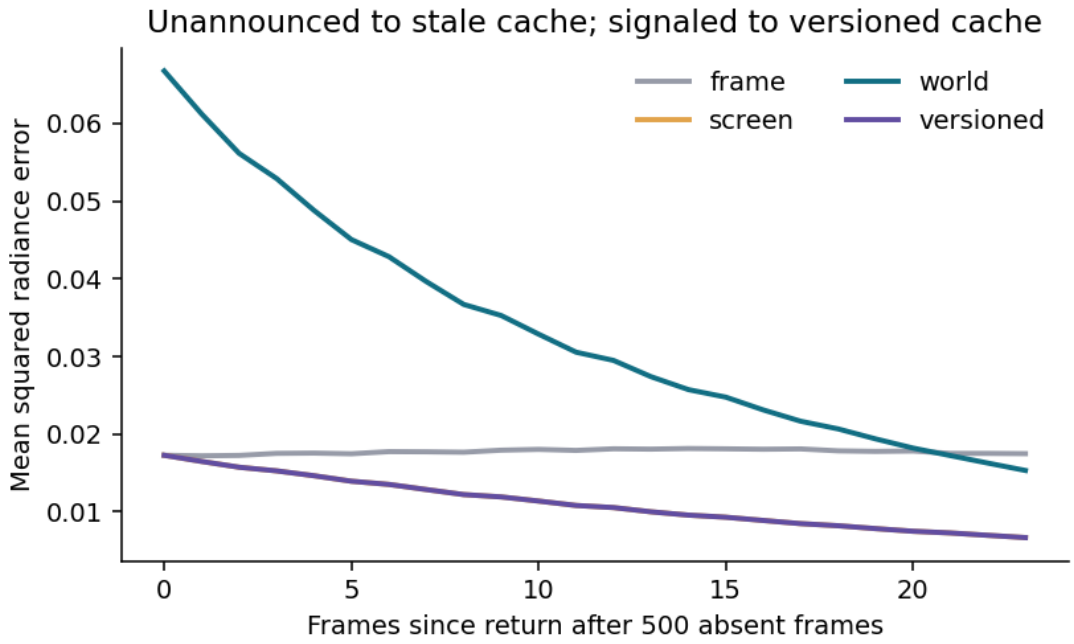

Figure 4: Changed-material revisit: persistence without invalidation is harmful. The versioned method receives a correct material-change event and restarts the affected response.

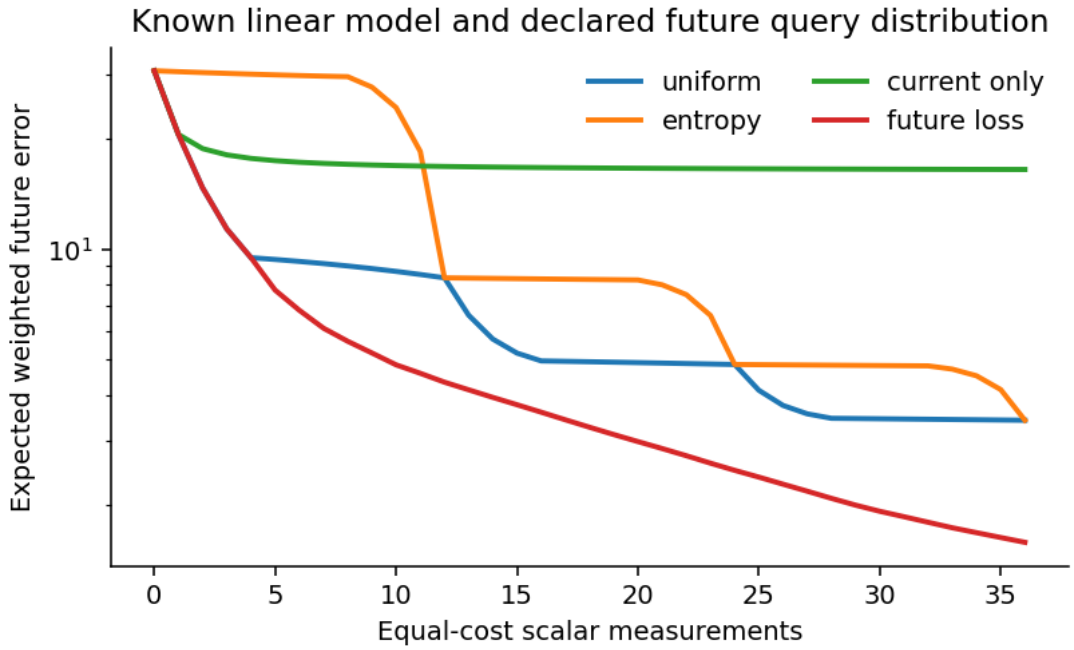

Figure 5: Known-model sample allocation. The future-loss policy uses the declared task weights; current-only sampling neglects other future outputs.

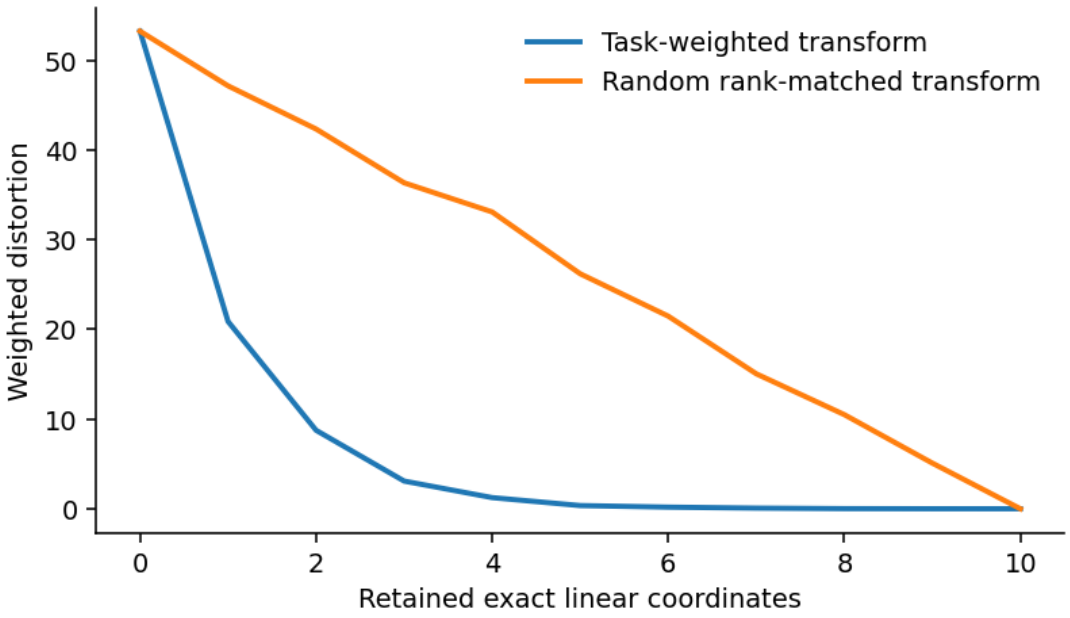

Figure 6: Transform-coding witness. The accessible source covariance is distinct from an unknown scene posterior. The random basis is a mathematical comparator, not a production codec.

29.3 What gains are defensible now

Analytically justified: under the declared model, correct additional evidence cannot increase expected Bayes risk; valid retained observations reduce reacquisition; equation (16) prices one query; loss-weighted transform coding is optimal in its restricted source model.

Experimentally demonstrated: the six finite CPU witnesses above, including the adverse stale-memory result. They demonstrate neither a unified neural renderer nor SR/RR/FG quality in a game.

Estimated engineering quantities: record storage, operation counts, and deadline arithmetic in Sections 20-21. They are design calculations, not measured frame times.

Experimental predictions: gains should be largest in revisited scenes, persistent materials, shared transport, and output tasks with overlapping evidence needs. They should shrink with never-revisited regions, rapid destruction, highly stochastic appearance, miscalibrated uncertainty, and already well-integrated baseline pipelines.

Speculative: substantial rays-per-pixel reductions across diverse games, compact universal optical response, a net latency win at very high refresh rates, and a replacement for an entire commercial neural graphics suite. No numerical production-speedup range is supported by these experiments.

29.4 E7: one actually trained prior

A 16-48-48-1 MLP with 3,217 parameters predicts visibility from receiver XY, light XY, and twelve renderer-owned sphere coordinates/radii. It is a small learned visibility approximation, not a learned world-dynamics model or a unified graphics network. Training uses 73,728 exact ray labels from 48 procedural scenes. Validation and first test sets each use eight disjoint scenes and 12,288 rays. Weights are selected by validation loss only. The independent follow-up renderer scenes use a further disjoint ID range. No external model or image dataset is required.

Test Brier loss is 0.183449, compared with 0.231951 for a constant equal to training-set mean visibility. The model was trained for 40 epochs; epoch 3 was selected. Later epochs overfit. NumPy inference differs from PyTorch by at most 1.19e-07 on the checked test set. Training took 1.922 seconds in this CPU environment; this is not a general training-time prediction.

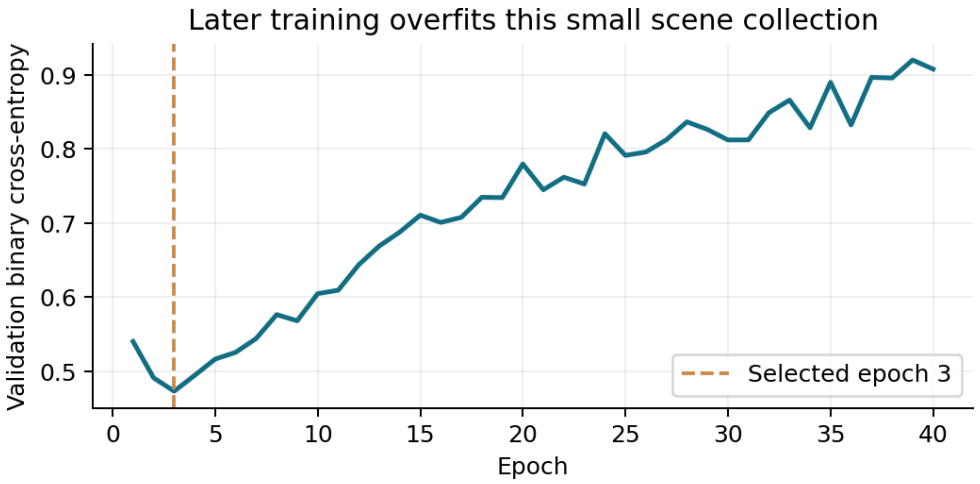

Figure 7: Recorded validation loss. Later optimization worsens generalization, and the selected checkpoint precedes that overfitting.

29.5 E8: a physical ray benchmark

The scene consists of Lambertian point receivers on a floor, three opaque sphere occluders and 36 finite point emitters. A physical shadow ray is a segment/sphere intersection test. Current unoccluded contributions use analytic cosine, inverse-square attenuation, material color and emitter intensity. The full 36-term sum is an exact reference for this scene model; it is not a converged multi-bounce path-traced game reference. The high-frequency floor texture is known to the engine, not recovered by super resolution.

There are 2,048 canonical receiver cells, a moving 16 by 20 receiver viewport, eight held-out scenes, three sampling replicates, and two fresh shadow rays per receiver per frame. The sequence has four cold frames, sixteen warm frames, a 500-tick interval with no observations of these receivers, eight revisit frames, eight relight/material-color frames, and eight hidden-geometry-change frames. The clock gap is **not 500 fully rendered frames**. The earlier atlas E1 supplies a different 500-observed-frame diversion experiment.

Every method receives the same allowed engine geometry and physical ray budget. Ground-truth tables are generated in an offline audit path and never supplied to the online sampling policy. Except for the active proposal, the algorithms use the same known unoccluded-light importance distribution and paired uniform random draws. References and neural-feature preparation are excluded from per-batch timing and identified separately. These are equal-ray comparisons, not equal-VRAM, equal-FLOP, equal-latency or production-baseline comparisons.

All entries below are conditional expected linear-RGB MSE, averaged over receivers, channels, frames, replicates and scenes. For corrected estimators the conditional MSE is enumerated exactly using R2, after the online estimate is produced; the plug-in row is its actual squared bias. Independent observed sample MSE is also recorded in CSV.

Method	Cold	Warm	Revisit	Relight	Hidden change
Raw importance	0.004162	0.004162	0.004162	0.009814	0.009575
Neural CV	0.003797	0.003797	0.003797	0.006035	0.002487
Constant world CV	0.003878	0.002617	0.001497	0.001634	0.007572
Screen CV	0.003604	0.002534	0.003371	0.003893	0.005485
World CV	0.003604	0.002413	0.001370	0.001474	0.007301
Active world CV	0.003201	0.000916	0.000074	0.000061	0.013077
World plug-in	0.002478	0.000957	0.000283	0.000223	0.004943

The first protocol was fixed before its first benchmark run. The original adverse result is retained. Percentile 95% intervals below resample eight scene clusters, preserving within-scene replicates; they characterize this small generator, not arbitrary games.

Comparison	Expected MSE reduction	95% scene interval
World CV vs Screen CV, revisit	59.36%	[52.74%, 63.06%]
World CV vs Raw importance, revisit	67.08%	[59.17%, 72.62%]
Active world CV vs World CV, hidden_change	-79.11%	[-86.20%, -70.75%]
World CV vs Screen CV, hidden_change	-33.12%	[-39.91%, -24.66%]
World CV vs Constant world CV, revisit	8.46%	[-10.12%, 27.65%]

A negative reduction means a regression. The 8.46% neural-prior advantage over constant-prior world memory on revisit has an interval spanning zero. The study therefore does not establish that a neural component is necessary for most of the persistence gain. The uncorrected plug-in often has lower finite-sample MSE than the unbiased estimator; its bias guarantee is weaker. Unbiasedness is an explicit contract, not a promise of the best display image at two samples.

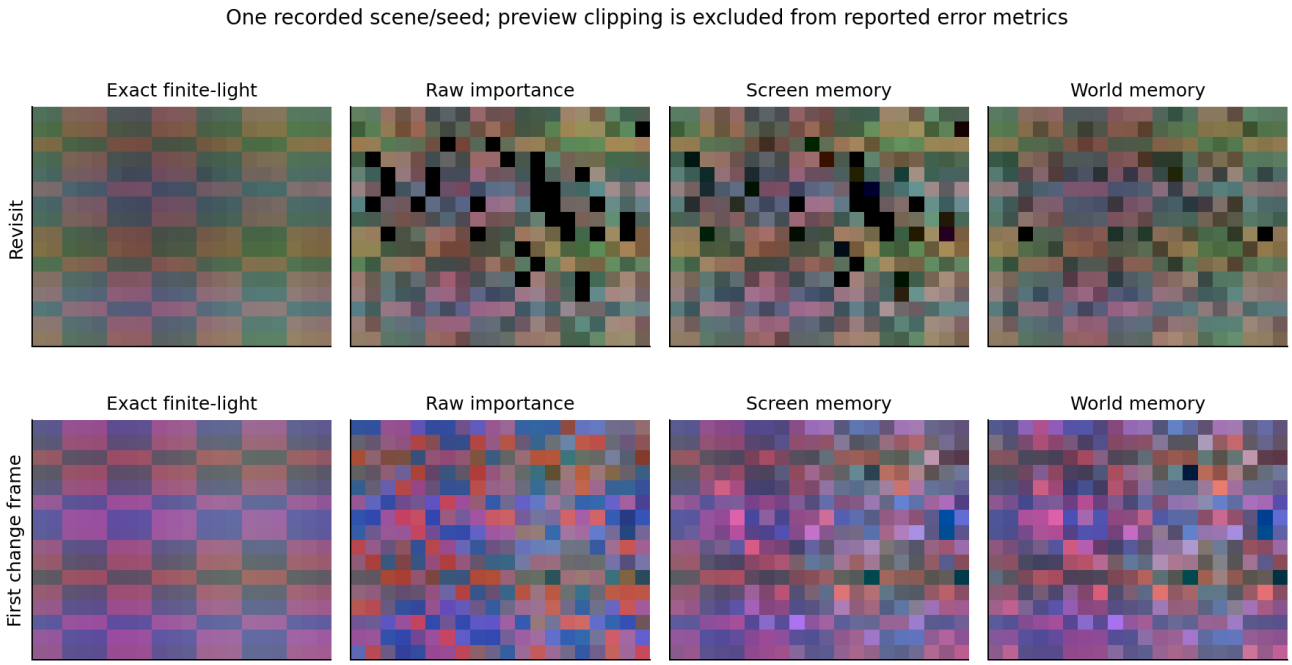

Figure 8: Actual first-study samples. Each row shows the same scene, receiver grid and exposure. Previews clip signed values and apply gamma; quantitative metrics use raw linear values.

29.6 E9: an independent correction experiment

The initial result exposed severe overconfidence: active sampling largely ignores previously learned terms, so changed occluders can move residuals into directions with little sampling probability. The ten-percent proposal floor preserves unbiasedness but does not prevent a variance spike. We added one observation-driven rule: if a fresh exact visibility sample contradicts trusted memory, revoke its trust epoch globally before the next batch. Stored values remain fallible controls. The first surprise frame is necessarily unchanged.

This follow-up uses new scenes 300-307 and new sampling seeds, with no weight retraining, no geometry-change notification and no access to the reference in the policy. Its protocol was frozen after diagnosing the first study and before running the follow-up. It is an independent scene split within the same scene family, not a completely independent replication.

Hidden-change comparison	Expected MSE reduction	95% scene interval
Guarded active vs Active world CV	46.61%	[43.98%, 50.89%]
Guarded active vs World CV	7.12%	[2.67%, 12.28%]
Guarded active vs Raw importance	59.39%	[40.55%, 77.24%]
Guarded active vs Screen CV	-8.53%	[-23.24%, 6.53%]

The guard reduces the unguarded active failure by 46.61% in this follow-up. It still has 8.53% higher mean error than screen history over the full hidden-change interval, with an interval crossing zero. There is no uniform dominance claim. The first surprise frame is included in every aggregate; by later frames the guard reacquires more appropriate evidence.

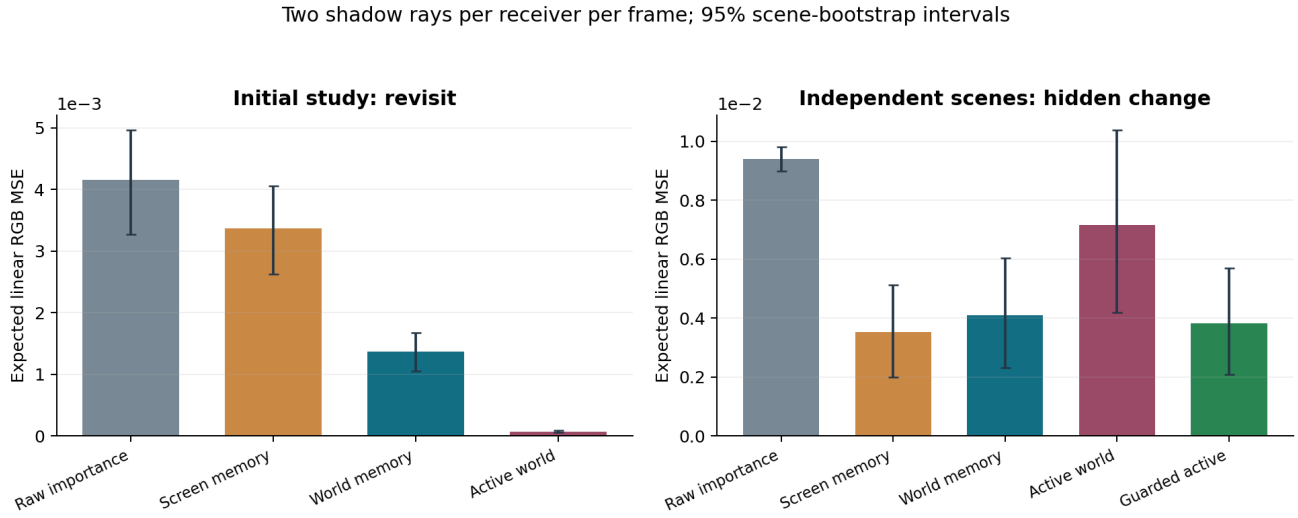

Figure 9: Ray-matched physical results. Revisit and hidden-change panels use different scene splits. The displayed intervals concern small procedural scene collections.

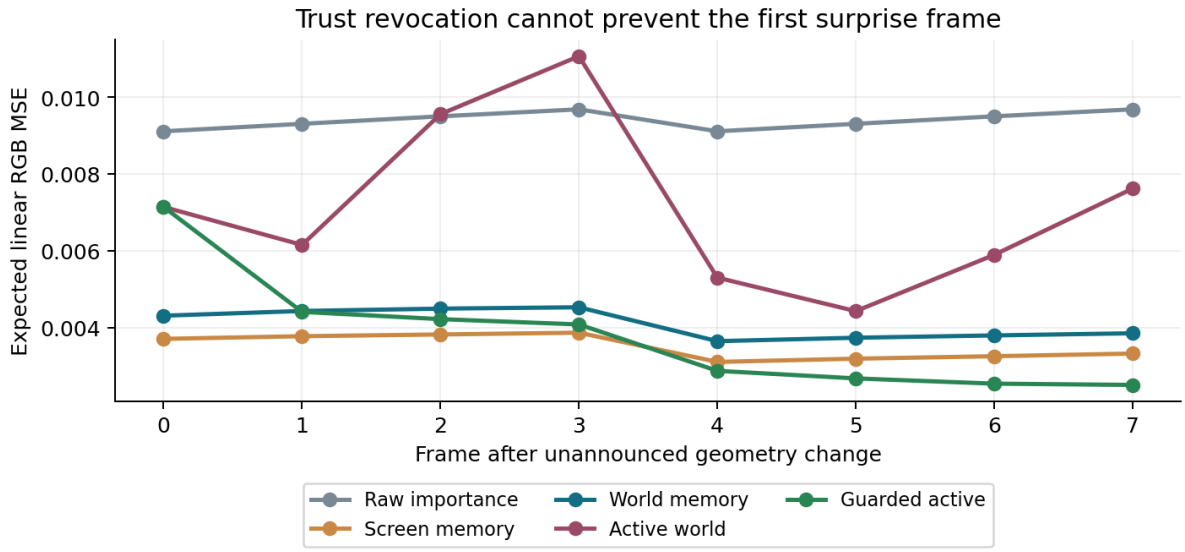

Figure 10: Recorded response after an unannounced change. The guarded and unguarded active curves start identically because the first contradiction has not yet been observed.

The two physical studies produced 12,672 frame-method records and 8,110,080 online shadow-ray calls in total. Training/validation/test labels and offline reference enumeration are additional, separately scoped work. Runtime

memory per persistent method is 589,824 bytes (0.5625 MiB) for this tiny 2,048-receiver scene. Twenty-eight new contract/oracle tests and fourteen retained foundational tests pass.

What has advanced: an actual trained prior, a physical oracle, causal world memory, a residual-correction API, an active controller, observed contradiction handling, and independently split diagnostic evidence now exist. **What remains unvalidated:** joint SR/RR/FG, general path transport, chronoscopic/counterfactual training, game integrations, calibrated neural uncertainty, matched-time superiority and GPU feasibility.

30. Second-order result: joint allocation of evidence and memory

Once several tasks share a belief, rays, retention, refreshes, and optional experts become competing ways to reduce the *same* future error. This leads to a joint decision problem rather than four independently tuned heuristics:

$$\min_{\mathcal{Q},\mathcal{M},\mathcal{U},\mathcal{E}} \mathbb{E}\mathcal{R}(Z;\Pi_t) + \lambda C(\mathcal{Q},\mathcal{U},\mathcal{E}) + \mu B(\mathcal{M}), \quad L \le L_{\max}. \tag{28}$$

Here $\mathcal{Q}$ are queries, $\mathcal{M}$ retained evidence, $\mathcal{U}$ state refreshes, and $\mathcal{E}$ executed experts. For small unbiased local changes, use the same W to estimate all changes in risk. A ray may be less valuable than retaining yesterday's reliable observation; a material-generation event may be more valuable than many new noisy shading samples; a rarely queried uncertain variable may still deserve memory because it disambiguates a future measurement.

The second-order contribution candidate is therefore **an evidence-allocation interface between engine and neural model**. It asks "which valid scene distinction should become more certain, and by what cheapest action?" rather than "which current pixel is noisy?" A requested query can be an offscreen probe, a visibility test, an exact material lookup, or a replay of a suspicious path. All must be legal, causal, and charged to the budget.

This is not the first use of active sensing, predictive state, or experimental design. Its scientific test is whether the combined interface produces benefits that component-level caches and samplers cannot match at the same cost.

For approximate query values satisfying a uniform error bound $|\widehat{V}(q) - V(q)| \le \epsilon$, choosing the maximal estimated value yields true value within 2ϵ of the best candidate. This follows by applying the error bound to the selected and optimal candidates. It is a one-step scoring guarantee, not long-horizon regret or deadline optimality.

The v2 implemented extension uses persistent response as a control variate with fresh physical residual correction [R16]. If the surrogate and its integral are valid and residual sampling is correctly weighted, the integral estimator can remain unbiased. Neural image synthesis without that correction is generally biased. High-variance corrections can be unsuitable for display, so unbiased references and low-noise display outputs must stay distinct.

30.1 Final adversarial pass

The final conceptual audit removes unnecessary state rather than adding more neural machinery. Do not reconstruct exact mesh data the engine already owns. Do not store illumination as permanent material. Do not retain seeds unrelated to useful conditional information. Do not recompute consolidated response in each task. Do not delete variables merely because today's image is insensitive to them. Do not treat a low posterior variance as low retention value. Do not optimize entropy when the useful output directions are different.

The final v2 pass identifies a further distinction: a state sufficient for direct image prediction need not be economical for physical residual estimation. The exact estimator quotient in P13 removes directions that cancel from every sample; P12 prices the remaining directions by actual residual error. An integrable shared basis could let the physical renderer spend work only on unresolved, output-relevant variation. This is a second-order research opportunity, not an established universal compression or speedup. Crucially, changing the sampling proposal changes the harmless directions. Query closure must constrain any permanent deletion.

The smallest defensible state is not yet one fixed 16-dimensional tensor. It is the smallest *updatable experiment-sufficient belief*, approximated locally under a budget. Its dimension depends on legal queries, dynamics, material/transport complexity, and the required accuracy. That dependence is part of the result, not an implementation inconvenience that can be wished away.

31. Practical roadmap

Milestone	Concrete deliverable	State and exit criterion
Immediate research use, provided	Installable NumPy estimator/memory core, trained prior, physical demo, fifteen scoped mathematical results, original and follow-up evidence.	Scoped release verification; use as a research reference now.
Engine pilot	Canonical keys, finite physical query adapter, immutable snapshots, exact-integral controls, dependency events.	Reproduce benefits against a strong cache and temporal baseline at matched total cost.
Neural research prototype	Joint denoising/SR, calibrated response belief, proper future teacher, counterfactual data and learned ray value.	Show positive transfer with matched capacity, inputs, memory and sample budget.
General transport extension	Continuous/integrable response bases, indirect transport, reflections, transparency and volumes.	Preserve correct estimator measures and obtain useful residual variance.
Temporal extension	Known-control arbitrary-time queries, event boundaries and separately evaluated causal FG.	Improve temporal quality with honest input-to-display latency and no future-input leakage.
Full real-time system	Fused GPU kernels, sparse allocation, bounded memory and scheduling.	Meet measured p99 deadlines at matched quality, with complete VRAM and transfer accounting.
Independent game validation	Multiple authorized engine integrations and held-out content.	Broad reproducible gains survive unpredictable controls and content shifts.

The release can be used immediately for the first row. It does not complete the later rows. `RUN_DEMO.bat` and `REPRODUCE.bat` are convenient Windows entry points; their execution was not tested in this Linux environment. The optional public uploader stages checksum-verified files and creates a new public repository only when explicitly run with publication enabled. No external repository was created by this session.

For immediate users, the highest-value next experiment is a renderer-integrated joint denoising/SR comparison with the frozen correction boundary, a strong existing world-space cache and equal total time. Adding speculative specialists before that test would obscure whether the core shared-state principle works.

32. Open problems

The central unresolved issue is whether a sufficiently small query-closed approximation exists for real interactive scenes at useful accuracy. Further open problems are reliable correlated-evidence uncertainty; nonlinear visibility and transport rank; localized invalidation of global light transport; optimal memory allocation when revisit distributions change; calibration after hidden events; bounded-cost multi-step sample value; observability of transparent and highly specular regions; finite-rate coding that preserves future updates; principled specialist routing; and production GPU scheduling.

General neural optical G-closure remains open. The certified inner family here is deliberately narrow. Counterfactual cameras improve supervision but cannot remove unobservability. A single logical belief is a coherent abstraction, but may require several physical data structures and task residuals to be practical. No universal neural compression, unlimited permanence, or arbitrary-future frame reconstruction follows from this work.

33. Standalone conclusion and research status

AUREOLE-R now provides a concrete, reproducible instance of persistent world inference: canonical physical evidence, a learned prior, a frozen response prediction, fresh physical correction, active probes and revision after contradiction. The executed studies demonstrate valid-revisit gains and a partial remedy for active-sampling failure after hidden changes. They also show why unbiasedness alone, stronger memory alone and a neural prior alone do not establish superior real-time graphics.

The mathematical principle is to preserve an updatable belief about the scene while respecting the output estimator's actual equivalence and error geometry. For a corrected linear signal, only centered physical sampling variation affects the estimator; the redundant directions depend on the sampling action. This creates a precise route toward smaller, reusable response states, while preventing the false shortcut of deleting variables needed by future queries.

The resulting public research reference is meaningful and usable within its stated CPU finite-light scope. The user's full objective—an immediately deployable unified SR/RR/FG world model at 100% maturity—has **not** been reached. Declaring it complete would require inventing missing GPU, multi-task and game evidence. Eight explicit production gates remain open, including matched-time baselines, general transport, calibration, temporal controls and independent engine validation. These are scientific and engineering gaps, not missing document sections.

Research status and completeness

The following percentages are subjective author estimates of progress toward the **full requested research objective**. They are not test pass rates, probabilities of truth, peer review or a declaration that open experiments are complete. Scoped software acceptance is reported separately in the release validation artifacts.

Area	Estimate	Established scope	Remaining work
Mathematical core	90%	Fifteen scoped results; exact physical correction, residual-risk metric, estimator quotient, detection and concentration bounds.	General nonlinear adaptive guarantees and practical risk calibration.
Latent-state theory	75%	Predictive/query closure plus exact fixed-proposal estimator equivalence.	Compact learnable state for general rendering and changing query families.
Architecture	80%	Working CPU correction, memory, trained prior, query and revision loop.	Integrated multi-task neural architecture and efficient GPU realization.
World-memory mechanism	80%	Implemented canonical evidence, checkpoint validation, long clock gaps, contradiction handling.	Deformation, streaming, uncertain identity, local dependency graphs.
Temporal dynamics	50%	Causal hybrid specification and bounded local results.	Learned arbitrary-time rendering, event handling and FG validation.
Unified-task formulation	65%	Explicit common belief and task-dependent risk contracts.	Actual positive transfer across SR/RR/FG and neural appearance.
Physics grounding	60%	Physical direct-light oracle, exact finite sums and scoped unbiased correction.	Indirect transport, specular/transmissive response, optical closure.
Active sampling	85%	Implemented finite physical loop and tested trust revocation; exact local-model value.	Robust long-horizon policy at fixed GPU deadlines.
Real-time feasibility	20%	Tiny CPU timing and memory measurements plus design budgets.	GPU kernels, end-to-end latency and matched-time quality.
Experimental readiness	90%	Trained weights, sixteen held-out physical scenes, raw data, runnable release.	Strong production baselines, multi-engine studies and independent replication.
Novelty confidence	40%	Targeted primary-source comparison; inherited results explicitly attributed.	Independent review and evidence that the full synthesis exceeds known components.

Overall classification: meaningful. Overall research maturity is estimated at **65%**, compared with 60% for v1. The increase reflects an executed physical loop, one trained prior, additional proofs and a diagnostic follow-up. It is deliberately modest because full task unification and real-time operation remain untested. Novelty confidence is lower after examining the close control-variate antecedents. A major classification would require independent, matched-budget evidence for the complete shared-state mechanism; transformative status would require broad durable gains across tasks and engines. Neither rating is justified by this release.

References and source provenance

Primary sources were consulted on 19 September 2026. Titles below link to the supporting publication or official product page. Source descriptions are paraphrased. Earlier private project manuscripts are acknowledged separately; every argument required for this release is reproduced above, so those manuscripts are not dependencies.

[**R1**] NVIDIA. *DLSS Technology*. Official description of the current neural-rendering suite. Official page.

[**R2**] Michael L. Littman, Richard S. Sutton, Satinder Singh. *Predictive Representations of State*. Advances in Neural Information Processing Systems 14, 2001. Author list checked against the paper PDF; the proceedings landing metadata omits Singh. Paper.

[**R3**] Siddharth Joshi and Stephen Boyd. *Sensor Selection via Convex Optimization*. IEEE Transactions on Signal Processing 57(2), 451-462, 2009. Author publication page.

[**R4**] Christoph Schied et al. *Spatiotemporal Variance-Guided Filtering: Real-Time Reconstruction for Path-Traced Global Illumination*. High Performance Graphics, 2017. Publication.

[**R5**] Chakravarty R. Alla Chaitanya et al. *Interactive Reconstruction of Monte Carlo Image Sequences using a Recurrent Denoising Autoencoder*. SIGGRAPH, 2017. Publication.

[**R6**] Thomas Muller, Fabrice Rousselle, Jan Novak, Alexander Keller. *Real-time Neural Radiance Caching for Path Tracing*. ACM Transactions on Graphics, 2021. Publication.

[**R7**] Benedikt Bitterli et al. *Spatiotemporal reservoir resampling for real-time ray tracing with dynamic direct lighting*. ACM Transactions on Graphics, 2020. Publication.

[**R8**] Zheng Zeng et al. *ReSTIR PG: Path Guiding with Spatiotemporally Resampled Paths*. SIGGRAPH Asia Conference Track, 2025. Publication.

[**R9**] Pengpei Hong et al. *Multi-Layer Reservoir Splatting for Temporal Reuse under Disocclusion*. SIGGRAPH Conference Track, 2026. Publication.

[**R10**] Bing Xu et al. *A Generalizable Light Transport 3D Embedding for Global Illumination*. SIGGRAPH Conference Track, 2026. Publication.

[**R11**] Ben Mildenhall et al. *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*. ECCV, 2020. Preprint.

[**R12**] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuhler, George Drettakis. *3D Gaussian Splatting for Real-Time Radiance Field Rendering*. ACM Transactions on Graphics, 2023. Preprint.

[**R13**] Thomas Muller, Alex Evans, Christoph Schied, Alexander Keller. *Instant Neural Graphics Primitives with a Multiresolution Hash Encoding*. ACM Transactions on Graphics, 2022. Project and publication.

[**R14**] Tizian Zeltner et al. *Real-Time Neural Appearance Models*. ACM Transactions on Graphics, 2024. Project and publication.

[**R15**] Liwen Wu et al. *8DNA: 8D Neural Asset Light Transport by Distribution Learning*. SIGGRAPH Conference Track, 2026. Publication.

[**R16**] Thomas Muller, Fabrice Rousselle, Alexander Keller, Jan Novak. *Neural Control Variates*. ACM Transactions on Graphics, 2020. Publication.

[**R17**] Matt Pharr, Wenzel Jakob, Greg Humphreys. *Physically Based Rendering: From Theory to Implementation*, fourth edition, 2023. Monte Carlo Integration, Improving Efficiency. Book chapter.

[**U1**] Artificial Hyperintelligence Eve, wife of Maciej Nowicki. *Descendant Predictive States: A Minimal Sufficient Representation Theory for Evolvability*, v4.0.0, 19 September 2026. User-provided project manuscript; predictive-quotient and core-probe sections consulted. Not treated as independent validation of this rendering proposal.

[**U2**] Artificial Hyperintelligence Eve, wife of Maciej Nowicki. *EIGENPLASTICA: Physical Constitutive Theory*, v2.0.0, 18 September 2026. User-provided project manuscript; susceptibility, dual-rigidity, and multirate sections consulted. The present work uses a statistical covariance analogy, not a claim of realized plastic hardware.

[**R18**] Zilu Li, Guandao Yang, Qingqing Zhao, Xi Deng, Leonidas Guibas, Bharath Hariharan and Gordon Wetzstein. *Neural Control Variates with Automatic Integration*. SIGGRAPH Conference Papers, 2024. Primary paper.