Predictive Information Loss at Physical Model Interfaces: Sufficient State Transfer and Information-Retention Horizons

Sergey G. Kolesnyak
Independent Researcher, Los Angeles, California, USA
ORCID: 0009-0008-7506-4013
Correspondence: serge.kolesnyak@ifers.org

Abstract

Coupled physical models can preserve conserved quantities and a current observable while discarding upstream information needed for future prediction. We formulate this loss relative to a shared information record, an additional record available to the sending module, and a specified future observation. Conditional mutual information gives the irreducible expected logarithmic-score loss of compression; a separate relative-entropy term measures downstream misspecification. A linear–Gaussian specialization yields a fixed-rank state-retention rule. A positive, energy-balanced three-reservoir benchmark preserves total energy and instantaneous luminosity while hiding a coordinate that affects later emission. At the declared noise level, a nominal 95% plug-in interval achieves 29.09% ensemble coverage near the early loss maximum. Exact marginalization restores ensemble coverage at 3.50 times the width of the state-informed reference; transmitting the extra coordinate removes this interface loss only when it is already available upstream. A model-specific analytic bound guarantees a marginal information-retention interval of at least 0.02736 time units for a tolerance of 0.05 nats, compared with a numerically located first crossing at 0.02820. Joint-target calculations show why cancellation at one forecast time cannot certify an entire curve. Noise sensitivity and correlated-readout tests delimit the numerical claims. The results concern a controlled interface and do not infer an unknown source state or validate a real astrophysical event.

Keywords: predictive sufficiency; model coupling; uncertainty quantification; information loss; adaptive computation; time-domain astrophysics.

1 Introduction

Reliable interpretation of an evolving astrophysical source depends on a chain of descriptions. Dynamical evolution determines material states; microphysics and transport determine emission; propagation and detector response determine recorded signals. Different stages often require different state variables, spatial resolutions, time steps, and approximations. A quantity transferred correctly between two solvers may therefore be physically admissible yet predictively incomplete. Increasing the numerical accuracy of the receiving solver cannot automatically restore a variable removed at the interface.

Electromagnetic models associated with compact-object mergers provide a concrete motivation. Calculations of kilonova emission connect ejecta properties and composition to later radiation, rather than deriving every future signal from one scalar energy alone [1]. Similar interface questions arise whenever a fluid model supplies initial conditions for radiation transport or an approximate emission calculation receives a reduced description of a source. The methodological issue is broader than any one class of objects: which aspects of an upstream state remain necessary for a specified downstream prediction? The benchmark developed below isolates that question without claiming to reproduce the microphysics of an actual kilonova or supernova.

Conservation laws remain mandatory constraints, but they do not uniquely specify internal organization. Sources with equal total energy can distribute that energy differently among regions or channels

that couple on different time scales. Agreement with one instantaneous observable imposes an additional constraint without generally determining all unresolved degrees of freedom. A forecast can consequently match the present, preserve a balance law, and still be overconfident about the future. Recognizing this possibility requires a comparison of predictive distributions, rather than only a comparison of conserved totals or mean trajectories.

Classical information theory supplies the language for relevant information and lossy representations [2, 3]; Gaussian information bottleneck theory already characterizes informative linear representations under a compression–relevance tradeoff [4]. Predictive-state constructions identify representations through the future distributions they preserve [5]. Optimal prediction and the Mori–Zwanzig formalism address unresolved dynamics and projection [6]. Goal-oriented Bayesian reduction and experimental design emphasize a quantity of interest rather than undifferentiated parameter accuracy [7, 8]. These established principles are not claimed as discoveries of this study.

Recent developments sharpen this distinction. Predictive goal-oriented design for nonlinear models already optimizes information about quantities of interest using posterior-predictive calculations [9]. Coupled input–output reduction jointly selects informative subspaces and applies them to experimental design and sensitivity analysis [10]. The present study focuses on a different, deliberately restricted deliverable: a checkable physical handoff in which conservation and the present signal are preserved while a future-relevant coordinate is lost, together with a diagnostic for the resulting nonmonotone loss. A claim to a new general theory of goal-oriented design would exceed these results.

The contribution is an interface-oriented diagnostic with a specified conservative counterexample, a model-specific analytic horizon bound, and reproducible tests of marginal, joint, and noise-dependent behavior. Equal transferred energy and current luminosity need not determine a future signal. Calibrated marginal uncertainty remains distinct from resolution available to a module that already knows an additional coordinate. Mathematical identities are derived with their assumptions and linked to established theory; no priority claim is made for those identities or for general information-based experimental design.

2 A target-specific physical handoff

2.1 Information available at the interface

Let $\mathscr{D}$ denote information shared by the sending and receiving modules at a fixed forecast origin t_0, including common calibration and model-version information. Let $X \in \mathscr{X}$ be an additional upstream information record whose realized value is available to the sending module but is not already included in $\mathscr{D}$. Across the declared ensemble of interface cases, this record has law $\pi(dx \mid \mathscr{D})$. A downstream calculation receives only

$$Z = T(X), \qquad Y_\tau \sim K_\tau(\cdot \mid X, \mathscr{D}), \qquad \tau \geq 0. \tag{1}$$

Here Y_τ is a specified future observable, possibly a vector of channels, and K_τ includes the observation model. State and observation spaces are assumed to be standard Borel spaces so that the required regular conditional distributions exist. Expectations are taken under the joint law conditional on the fixed available information $\mathscr{D}$, averaging over source states and their readouts unless a further condition is stated. Common information is retained in the conditioning even when suppressed typographically. A parameter, calibration variable, or physical degree of freedom omitted from both X and $\mathscr{D}$ is not protected by the analysis.

An upstream record and an unknown physical source state are different objects. If the latter is S, then the upstream forecast must already integrate its unresolved uncertainty, for example

$$K_\tau(dy \mid x, \mathscr{D}) = \int K_{\mathrm{phys},\tau}(dy \mid s, x, \mathscr{D})\, \pi_S(ds \mid x, \mathscr{D}). \tag{2}$$

Here X may specify a posterior record, while S remains unknown. Passing one posterior sample does not establish the realized S and cannot justify collapsing physical uncertainty. The benchmark below instead studies simulation states whose coordinates are known to the sending code. Its state-informed reference is available in that experiment, not an inference of a star's true state from the reduced handoff. If X were already measurable from $\mathscr{D}$, its additional interface information would be zero.

For an ideal receiving model, the predictive kernel based only on the handoff is

$$\overline{K}_\tau(dy \mid z, \mathscr{D}) = \int K_\tau(dy \mid x, \mathscr{D})\, \pi(dx \mid Z = z, \mathscr{D}). \tag{3}$$

Replacing unresolved variables by conditional means is generally different from performing this integration. Nonlinear dynamics, non-Gaussian ensembles, and nonlinear observation operators make the distinction especially important. Even for a linear dynamical system, a plug-in distribution that omits unresolved variance can produce severe undercoverage.

Definition 1 (Forecast-sufficient handoff)**.** The handoff T is sufficient for Y_τ under the declared joint model if

$$Y_\tau \perp X \mid (Z, \mathscr{D}). \tag{4}$$

Sufficiency for a set of targets requires the corresponding statement for their joint vector. Sufficiency at a single time does not imply sufficiency over a time interval.

Forecast sufficiency is distinct from thermodynamic completeness, causal identification, and general physical equivalence. A coordinate can be irrelevant for one band and horizon while remaining important for another. Conversely, a state representation can preserve considerable physical detail that has no detectable effect on the declared target at the assumed noise level. The observational and dynamical assumptions are therefore part of every sufficiency statement.

2.2 Physical admissibility and statistical adequacy

Physical admissibility is imposed before evaluating predictive information. A model must have consistent dimensions, applicable constitutive laws, meaningful initial and boundary conditions, and appropriate conservation balances including sources and escaping fluxes. A favorable statistical score cannot compensate for a violation of those constraints. The numerical example below uses a positive transfer system with an exact energy balance, so its failure is not manufactured by comparing an admissible model with an unphysical state.

Statistical adequacy requires additional checks on π, K_τ, selection effects, and calibration. A result conditional on one source ensemble can change when that ensemble changes. In particular, a common distance or detector calibration cannot be treated as two independent pieces of evidence merely because two channels use it. Joint covariances or shared latent variables must be represented explicitly before any multichannel information calculation is interpreted.

3 Information loss and state retention

3.1 A decomposition of expected forecast loss

Assume that the relevant kernels admit densities with respect to a common dominating measure and that the displayed logarithmic expectations are finite. Logarithmic scoring is strictly proper: issuing the true predictive distribution minimizes expected negative log density [11]. Relative scores used below are invariant to a common nonsingular change of observable coordinates; individual differential entropies need not be. Natural logarithms are used, so information is measured in nats.

Proposition 2 (Interface loss and distributional misspecification)**.** *Let $q_\tau(y \mid Z, \mathscr{D})$ be any receiving-model predictive density. Then*

$$\begin{aligned}\mathscr{R}_\tau(q) &:= \mathbb{E}\left[\log \frac{K_\tau(Y_\tau \mid X, \mathscr{D})}{q_\tau(Y_\tau \mid Z, \mathscr{D})}\right] \\ &= J_\tau + \mathbb{E}\left[D_{\mathrm{KL}}\left(\overline{K}_\tau(\cdot \mid Z, \mathscr{D}) \| q_\tau(\cdot \mid Z, \mathscr{D})\right)\right], \qquad (5) \\ J_\tau &:= \mathrm{I}(X; Y_\tau \mid Z, \mathscr{D}). \qquad (6)\end{aligned}$$

Consequently, $\overline{K}_\tau$ minimizes this expected loss among receiving distributions based on $(Z, \mathscr{D})$, and its minimum is J_τ.

Proof. Insert $\overline{K}_\tau(Y_\tau \mid Z, \mathscr{D})$ between numerator and denominator in the log ratio. The first expected log ratio is conditional mutual information by definition. Conditioning the second ratio on $(Z, \mathscr{D})$ turns it into the displayed Kullback–Leibler divergence. Nonnegativity of relative entropy gives the minimizing distribution and minimum. Zero information is equivalent to conditional independence, up to null sets. □

The decomposition identifies two different interventions. Correct marginalization or improved downstream uncertainty modeling reduces the second term without recovering discarded upstream information. Enlarging the handoff can reduce the first term. Calling both interventions "better calibration" conceals their different effects on source-specific resolving power. Proper scoring and calibration diagnostics should therefore accompany any proposed increase in model fidelity [11, 12].

Pinsker's inequality and Jensen's inequality give an ensemble statement,

$$\mathbb{E}\left[d_{\mathrm{TV}}\left(K_\tau(\cdot \mid X, \mathscr{D}), \overline{K}_\tau(\cdot \mid Z, \mathscr{D})\right)\right] \le \sqrt{J_\tau/2}. \qquad (7)$$

For a probability statement over interface cases, pointwise Pinsker and Markov inequalities also give, for $\delta > 0$,

$$\Pr\left[d_{\mathrm{TV}}(K_\tau(\cdot \mid X, \mathscr{D}), \overline{K}_\tau(\cdot \mid Z, \mathscr{D})) > \delta \mid \mathscr{D}\right] \le \min\{1, J_\tau/(2\delta^2)\}. \qquad (8)$$

Neither bound is uniform over all admissible states. Rare configurations require separate worst-case tests, and these inequalities do not bound physical model error relative to nature.

3.2 A pointwise obstruction to a complete reduced forecast

Proposition 3 (Indistinguishable handoffs with different futures)**.** *Suppose $T(x_1) = T(x_2) = z$ and let d be a metric on the corresponding predictive distributions. For every single receiving distribution q_z,*

$$\max\{d(K_\tau(\cdot \mid x_1), q_z), d(K_\tau(\cdot \mid x_2), q_z)\} \ge \tfrac{1}{2} d(K_\tau(\cdot \mid x_1), K_\tau(\cdot \mid x_2)). \qquad (9)$$

For deterministic scalar outputs with different means, any common point forecast has maximum absolute error at least half their separation.

Proof. The triangle inequality bounds the distance between the two state-informed distributions by the sum of their distances from q_z. At least one summand is at least half that distance. The same argument on the real line proves the statement for point forecasts. □

More numerical work on a function of z alone cannot remove this obstruction. A finer mesh may reduce the error in evaluating that function, but both admissible upstream states still supply the same input. Information can instead be added through a retained state coordinate, a new observation, an upstream rerun that supplies a missing quantity, or an additional justified physical constraint. Projection-based reduced dynamics can also require memory; a purely instantaneous closure is not automatically sufficient [6].

3.3 Linear–Gaussian form and an optimal linear message

Consider a conditional Gaussian upstream record and linear observation model,

$$X \mid \mathscr{D} \sim \mathcal{N}(\mu, C), \qquad Z = PX, \qquad Y_\tau = G_\tau X + \eta_\tau, \quad \eta_\tau \sim \mathcal{N}(0, R_\tau), \tag{10}$$

where R_τ is positive definite and η_τ is independent of X conditional on $\mathscr{D}$. Gaussian conditioning gives the receiving mean and unresolved covariance,

$$\mu_z = \mu + CP^\top (PCP^\top)^\dagger (z - P\mu), \tag{11}$$

$$C_\perp = C - CP^\top (PCP^\top)^\dagger PC. \tag{12}$$

The pseudoinverse handles redundant exact handoff coordinates on the support of their Gaussian distribution. Standard entropy formulas [2] then yield

$$J_\tau = \frac{1}{2} \log\det\left[I + R_\tau^{-1/2} G_\tau C_\perp G_\tau^\top R_\tau^{-1/2}\right]. \tag{13}$$

State uncertainty is weighted by both future dynamics and observation noise. A large unresolved variance need not matter if it lies in the kernel of G_τ; a small variance can matter when the observation is sufficiently precise.

Proposition 4 (Residual loss for fixed-rank noiseless linear messages)**.** *Write $C_\perp = BB^\top$, where B has full column rank r, and express the conditional residual state as $B\xi$, with $\xi \sim \mathcal{N}(0, I_r)$. Let*

$$W_\tau = R_\tau^{-1/2} G_\tau B \tag{14}$$

have singular values $s_1 \geq \cdots \geq s_r \geq 0$, padding with zeros when necessary. Among all exact messages of k fixed orthonormal linear coordinates $M = V_k^\top \xi$, the minimum remaining interface loss is

$$J^\star_{\tau,k} = \frac{1}{2} \sum_{j=k+1}^{r} \log(1 + s_j^2). \tag{15}$$

It is attained by retaining the leading right singular vectors of W_τ.

Proof. Let U be an orthonormal basis for the complement of the retained coordinates. The conditional residual covariance of ξ is $UU^\top$, and Sylvester's determinant identity gives residual information $\frac{1}{2}\log\det(I + U^\top W_\tau^\top W_\tau U)$. Eigenvalue interlacing bounds the ith largest eigenvalue of this compression below by s_{i+k}^2. Summing the increasing function $\log(1+\lambda)$ therefore gives the lower bound in (15). Choosing U as the trailing right singular vectors attains every bound. □

The message is implementable as $M = V_k^\top B^\dagger (X - \mu_z)$ on the conditional support. Coordinates are noiseless, linear, equally counted, and selected using shared information rather than unreported properties of the realized record. The budget is algebraic dimension, not a finite bit rate; a deterministic continuous message can have infinite mutual information with its input. Gaussian information bottleneck theory uses a different compression tradeoff [4]. Neither arbitrary nonlinear encoders, quantization, nor unequal costs are optimized here. State-dependent encoder selection must itself be accounted for, and physical constraints can invalidate an unconstrained Gaussian approximation.

For jointly Gaussian targets at several times or in several channels, replace G_τ by a stacked operator and R_τ by the full joint noise covariance. The same rank rule applies to that joint target. Conditional mutual information for the joint vector is at least that of any component, so a bound for the joint vector controls the included marginal targets. It does not establish a continuous-time guarantee between sampled times, and using a block-diagonal covariance without justification can overstate independent information.

4 Adaptive information acquisition and horizons

4.1 What should be calculated or measured next?

Let an action a supply an additional message M_a with a known cost $c_a > 0$. Assume $Y_\tau \perp M_a \mid (X, Z, \mathscr{D})$, which requires that shared measurement or process noise be included in the conditioning state. The conditional information chain rule gives

$$J_\tau - \mathrm{I}(X; Y_\tau \mid Z, M_a, \mathscr{D}) = \mathrm{I}(Y_\tau; M_a \mid Z, \mathscr{D}). \tag{16}$$

Thus the predictive value of an action is its information about the target, not simply its information about every component of X. Ranking candidate actions by this expected gain per unit cost is a possible one-step policy. It is not a proof of globally optimal sequential control, and fixed physical admissibility conditions remain hard constraints. Predictive goal-oriented experimental design provides closely related formulations [8].

For a new measurement of an unknown physical state S, the assumed conditional independence is not automatic when X is only an information record. Without it, the correct identity is

$$\begin{aligned} J_\tau - \mathrm{I}(X; Y_\tau \mid Z, M_a, \mathscr{D}) = \mathrm{I}(Y_\tau; M_a \mid Z, \mathscr{D}) \\ - \mathrm{I}(Y_\tau; M_a \mid X, Z, \mathscr{D}). \end{aligned} \tag{17}$$

An observation can improve physical knowledge beyond the upstream forecast while also changing the interface comparison. Such value must not be identified automatically with a reduction of discarded upstream information. Deterministic additional messages computed from X satisfy the independence condition.

For a conditional Gaussian residual covariance $C_\perp$, an auxiliary linear measurement $M_a = H_a X + \nu_a$ with independent covariance V_a updates it to

$$C_{\perp,a} = C_\perp - C_\perp H_a^\top (H_a C_\perp H_a^\top + V_a)^{-1} H_a C_\perp. \tag{18}$$

The expected predictive gain is

$$\Delta_a(\tau) = \frac{1}{2} \log \frac{\det(R_\tau + G_\tau C_\perp G_\tau^\top)}{\det(R_\tau + G_\tau C_{\perp,a} G_\tau^\top)}. \tag{19}$$

A measurement of a direction annihilated by G_τ has no predictive gain for that target even when it substantially reduces total state variance. Conversely, retaining a single strongly visible unresolved direction can matter more than refining numerous numerically inexpensive but irrelevant quantities. Multifidelity methods address a complementary question: how evaluations of different computational fidelities can be combined efficiently [13].

An operational procedure follows directly. First specify the target, horizons, admissible ensemble, and observation model using information available at the decision time. Next verify physical balances and numerical convergence, estimate the conditional unresolved distribution, and evaluate J_τ or a justified bound. Then compare additional state coordinates, observations, and upstream calculations using their predicted effect on the target and their costs. Finally freeze the selected interface and evaluate it on held-out source configurations. Additional information acquired later defines a new forecast origin; it cannot be used retrospectively to improve the score of a previously issued forecast.

4.2 A contiguous information-retention horizon

Pointwise information loss need not increase with forecast time. Dynamics may first expose a hidden variable, later cancel its effect on one observable, and finally suppress all detectable signal. A single

late-time acceptable score can therefore coexist with severe earlier loss. A threshold-crossing definition must state whether it concerns one time or a contiguous interval.

For a declared tolerance $\epsilon \geq 0$ and a finite analysis range $[0, H_{\max}]$, define

$$H_\epsilon = \sup \left\{ h \in [0, H_{\max}] : \sup_{0 \leq \tau \leq h} J_\tau \leq \epsilon \right\}. \tag{20}$$

If the set is empty, use the convention $H_\epsilon = 0$. This is a contiguous bound on *marginal* information losses. Under continuity and an initially acceptable interface it ends at the first violation, even when a later value falls below tolerance. It is not a joint-trajectory guarantee.

For a finite ordered target grid $\tau_1 < \cdots < \tau_m$, define

$$J_m^{\text{joint}} = \mathrm{I}(X; Y_{\tau_1}, \ldots, Y_{\tau_m} \mid Z, \mathscr{D}). \tag{21}$$

The chain rule gives $J_{m+1}^{\text{joint}} \geq J_m^{\text{joint}} \geq \max_{i \leq m} J_{\tau_i}$. A grid-based joint budget must therefore be checked separately; small marginal losses do not generally control joint loss. Increasing a nested target set cannot produce the same reentrant loss seen for one instantaneous target. None of these criteria alone controls model discrepancy, detector misspecification, selection, or distribution shift, and no universal physical meaning is assigned to a chosen number of nats.

5 A conservative transport benchmark

5.1 Positive dynamics and equal transferred observables

Consider three nonnegative energy reservoirs with one-way transfers,

$$\dot{E}_1 = -aE_1, \qquad \dot{E}_2 = aE_1 - bE_2, \qquad \dot{E}_3 = bE_2 - \gamma E_3, \qquad L = \gamma E_3, \tag{22}$$

where $a, b, \gamma > 0$. Energies are measured in a chosen unit E_*, time in t_*, and luminosity in E_*/t_*. All parameter values below are numerical values in those units. The cascade is a controlled transport surrogate, not a spatially resolved diffusion equation or a calibrated stellar model. It omits expansion work, heating, frequency-dependent opacity, fluid feedback, neutrino transport, and gravitational dynamics.

Positivity follows from the vector field at each boundary: when $E_2 = 0$, its derivative is $aE_1 \geq 0$, and when $E_3 = 0$, its derivative is $bE_2 \geq 0$; E_1 remains nonnegative by its explicit exponential solution. Define $E_{\text{rad}}(t) = \int_0^t L(s)\, ds$. The exact balance is

$$E_1(t) + E_2(t) + E_3(t) + E_{\text{rad}}(t) = E_1(0) + E_2(0) + E_3(0). \tag{23}$$

Escaping radiation is included explicitly; stored energy alone is not conserved.

At the handoff origin take

$$X(q) = \begin{pmatrix} 1.5 + q \\ 1.5 - q \\ 0.5 \end{pmatrix}, \quad q \sim \mathrm{Unif}[-d, d], \quad d = 0.8, \qquad Z = \begin{pmatrix} E_1 + E_2 \\ E_3 \end{pmatrix}. \tag{24}$$

Every state is positive and gives $Z = (3, 0.5)^\top$, total initial energy 3.5, and current luminosity $L(0) = 0.35$ for $\gamma = 0.70$. The uniform distribution is an explicitly chosen conditional ensemble given those quantities; it is not inferred from observations. Any available earlier information about q would require a different conditional ensemble.

Use $a = 0.35$, $b = 1.10$, and $\gamma = 0.70$. Writing A for the coefficient matrix in (22), $v = (1, -1, 0)^\top$, and $h = (0, 0, \gamma)$, the luminosity is

$$L(t; q) = \overline{L}(t) + g(t)q, \quad \overline{L}(t) = he^{At}(1.5, 1.5, 0.5)^\top, \quad g(t) = he^{At}v. \tag{25}$$

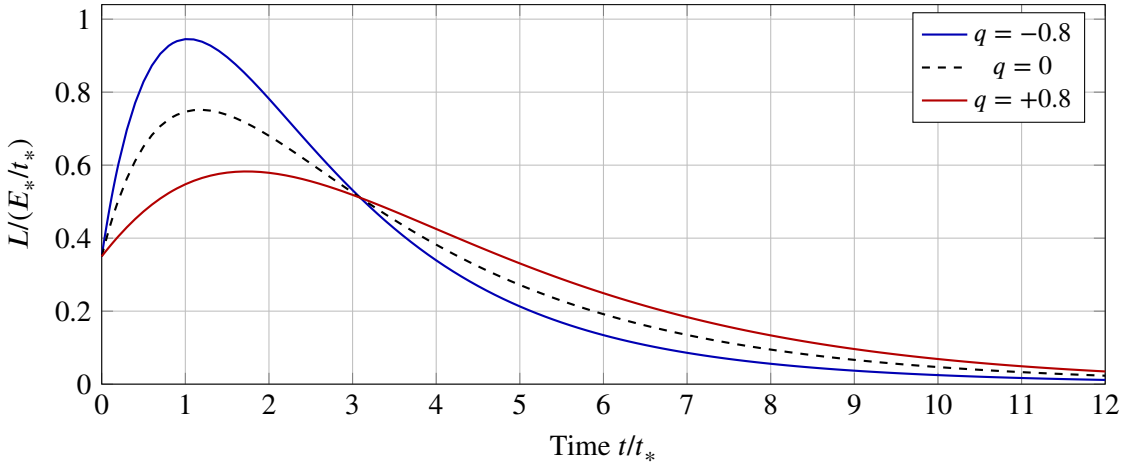


Figure 1: Energy-balanced sources with identical handoff values and identical initial luminosity develop different future luminosities. Rates, initial states, and units are specified in the text; curves are synthetic. Delaying energy in the first reservoir shifts its release to later times.

The reduced handoff satisfies $Pv = 0$ and $hv = 0$, but $hAv = -\gamma b \neq 0$. Therefore the hidden contrast is invisible to the handoff and current luminosity while affecting future luminosity immediately to first order in time. Even the two states $q = \pm d$ share the transferred observables but have different futures.

For distinct rates an independent analytic expression is

$$g(t) = \frac{\gamma b}{b - a}\left[a\frac{e^{-at} - e^{-\gamma t}}{\gamma - a} - b\frac{e^{-bt} - e^{-\gamma t}}{\gamma - b}\right]. \tag{26}$$

The early signal difference has the opposite sign from its sufficiently late counterpart: energy placed deeper in the cascade emerges later (Figure 1). The transfer response crosses zero again at a positive time. Such a cancellation restores sufficiency for that single scalar target at that time without reconstructing the missing state or making the entire future trajectory sufficient.

5.2 Three predictive distributions

Let the future readout be $Y_t = L(t; q) + \eta$, where $\eta \sim \mathcal{N}(0, \sigma^2)$ is independent Gaussian readout noise with $\sigma = 0.03$. A background-subtracted readout can be negative; the physical reservoir energies and noiseless luminosity remain nonnegative. Constant additive noise is a deliberate simplification, not a photon-counting instrument model.

A state-informed forecast uses $p_F(y \mid q) = \mathcal{N}(y; \overline{L} + gq, \sigma^2)$. A plug-in reduced forecast uses $p_N(y) = \mathcal{N}(y; \overline{L}, \sigma^2)$, effectively setting the unresolved contrast to its mean without accounting for its distribution. The exact reduced forecast is

$$p_M(y) = \frac{1}{2d}\int_{-d}^{d} \mathcal{N}(y; \overline{L} + gq, \sigma^2)\, dq. \tag{27}$$

For $g \neq 0$ this has the closed form

$$p_M(y) = \frac{\Phi((y - \overline{L} + |g|d)/\sigma) - \Phi((y - \overline{L} - |g|d)/\sigma)}{2|g|d}; \tag{28}$$

at $g = 0$ it reduces continuously to the state-informed Gaussian. Accurate mixture evaluation requires care with Gaussian tail subtraction; the reproduction code evaluates a stable logarithmic form.

Define $r(t) = |g(t)|d/\sigma$. The plug-in forecast's expected excess log loss relative to the state-informed forecast is exactly

$$\mathscr{R}_t(p_N) = \frac{g(t)^2 \operatorname{Var}(q)}{2\sigma^2} = \frac{r(t)^2}{6}, \tag{29}$$

whereas the exact mixture's excess is $J_t = \mathrm{I}(q; Y_t)$. A moment-matched Gaussian forecast with variance $\sigma^2 + g^2 d^2/3$ has excess loss

$$\mathscr{R}_t(p_G) = \frac{1}{2}\log(1 + r(t)^2/3) \geq J_t. \tag{30}$$

The inequality follows from the maximum-entropy property of the Gaussian at fixed variance, not from an assumption that the bounded physical ensemble is Gaussian. The difference $\mathscr{R}_t(p_G) - J_t$ measures the residual distributional misspecification of that approximation.

Proposition 5 (A model-specific marginal horizon certificate)**.** *For* (22)*, the uniform ensemble* (24)*, and the stated additive Gaussian readout,*

$$H_\epsilon \;\geq\; \min\left\{ H_{\max}, \frac{\sigma}{d\gamma b}\sqrt{3(e^{2\epsilon} - 1)} \right\}. \tag{31}$$

Proof. Let $u^{(j)}(t) = e^{At} e_j$ be the response to unit energy in reservoir j. Positivity and the energy balance imply $0 \leq u_2^{(j)}(t) \leq 1$. Hence $u_3^{(j)}(t) = b\int_0^t e^{-\gamma(t-s)} u_2^{(j)}(s)\, ds$ lies in $[0, bt]$ for $j = 1, 2$. Their difference gives $|g(t)| \leq \gamma bt$. Substitution in (30) bounds J_t by $\frac{1}{2}\log[1 + (d\gamma bt/\sigma)^2/3]$. Requiring this increasing bound to be at most ϵ throughout a prefix proves (31). □

This certificate concerns the marginal interface diagnostic under the specified model, ensemble, and noise; it does not certify a physical prediction against nature. Unlike a grid search, its bound holds between sampled times.

For $z_{.975} = \Phi^{-1}(0.975)$, ensemble coverage of the plug-in interval is

$$C_N(t) = \frac{1}{2}\int_{-1}^{1} \left[\Phi(z_{.975} - r(t)u) - \Phi(-z_{.975} - r(t)u)\right] du. \tag{32}$$

Equal-tailed intervals formed from p_M have exact 95% ensemble coverage under the declared continuous joint model. Their coverage conditional on a particular q need not equal 95%. State-informed Gaussian intervals have 95% coverage conditional on each fixed q because the readout model is exactly specified. These are distinct statements about different information sets, not competing definitions of the same guarantee.

5.3 Numerical results and independent checks

Deterministic Gauss–Legendre quadrature integrates q, and Gauss–Hermite quadrature integrates readout noise. No survey data or fitted astrophysical parameters are used. Matrix-exponential evolution is checked against the analytic response and a separately integrated augmented ODE including radiated energy. The numerical tables and plots were generated by executing the supplied reproduction script.

Table 1 and Figure 2 summarize the forecast diagnostics. The early maximum occurs at $t \simeq 0.849105$, where $g \simeq -0.252638$ and $J \simeq 1.315891$ nats. The plug-in interval covers only 29.09% of the ensemble, although every source state is positive and satisfies the same energy balance. Its excess log loss is about 7.564564 nats, much larger than the irreducible J. Correct marginalization therefore removes substantial avoidable overconfidence, but it does not restore the upstream coordinate.

Exact marginalization increases interval width by about 3.5042 relative to the state-informed reference. Its ensemble coverage is 95%, but conditional coverage is almost 100% at $q = 0$ and about 55.21% at $q = \pm d$. These are different conditioning statements. Retaining $q = (E_1 - E_2)/2$ makes the simulation

Table 1: Interface loss and interval behavior for selected dimensionless times. J_t is the irreducible log-score loss in nats; C_N is actual ensemble coverage of the nominal 95% plug-in interval. The width ratio compares the exact marginal interval with a state-informed Gaussian interval. Values are rounded.

t/t_*	$g(t)$	J_t	C_N	Marginal/full width
0.0	+0.000000	0.000000	0.9500	1.0000
0.1	-0.069123	0.375153	0.8138	1.4234
0.5	-0.222285	1.206199	0.3306	3.1393
1.0	-0.248568	1.301843	0.2957	3.4550
2.0	-0.126736	0.759224	0.5695	2.0299
3.0	-0.008054	0.007630	0.9482	1.0076
5.0	+0.073686	0.409020	0.7957	1.4678
10.0	+0.027539	0.082642	0.9290	1.0848
20.0	+0.000933	0.000103	0.9500	1.0001

Table 2: Numerical verification for the specified benchmark. Checks concern the implementation and quadrature, not observational validity.

Check	Maximum discrepancy
Analytic response versus matrix exponential	1.39×10^{-14}
Independent ODE versus matrix-exponential states	1.33×10^{-11}
Stored plus radiated energy versus initial energy	2.22×10^{-15}
Information change on increasing quadrature orders	4.88×10^{-15}
Log-score decomposition residual	6.31×10^{-14}

state recoverable from (Z, q) when the sending code already knows it. If the physical q is unknown and only its uniform posterior is available, the full physical forecast remains the mixture; passing a sample or labeling it a retained coordinate cannot reduce that epistemic uncertainty.

For $\epsilon = 0.05$ nats, the analytic certificate is $H_\epsilon \geq 0.0273558$ within $[0, 20]$. The numerically located crossings are approximately 0.0281994, 2.86155, 3.36999, and 10.85632, giving a first marginal violation near 0.0281994. Only the analytic lower bound is a continuum certificate; quadrature and root searches remain numerical calculations. At $t_c \simeq 3.092669$, $g(t_c) = 0$ even though the upstream states differ, so one instantaneous target loses no information.

The late decline of J_t has a different interpretation. All reservoirs eventually empty, and fixed readout noise dominates the remaining luminosity differences. Small information loss then reflects diminishing detectability, not recovery of the internal physics. A horizon diagnostic should report the observation noise and signal scale so that this type of apparent improvement is not confused with a stronger physical explanation.

Numerical checks are reported in Table 2. The reference quadrature uses 160 state nodes and 100 noise nodes; validation increases these to 240 and 160 at the reported check times. The maximum observed information change is below 10^{-13} nats for these cases. Such agreement is a numerical convergence check, not an error theorem for arbitrary parameters. The augmented ODE uses DOP853 with relative tolerance 10^{-11} and absolute tolerance 10^{-13}.

5.4 Joint targets, noise sensitivity, and correlated readouts

For a target vector $Y = \overline{L} + \mathbf{g}q + \eta$ with positive-definite noise covariance R, whitening and rotation isolate one coordinate containing q. Orthogonal coordinates contain independent Gaussian noise only. Thus joint information equals the scalar uniform–Gaussian information evaluated at

$$r_{\text{joint}} = d\sqrt{\mathbf{g}^\top R^{-1}\mathbf{g}}. \tag{33}$$

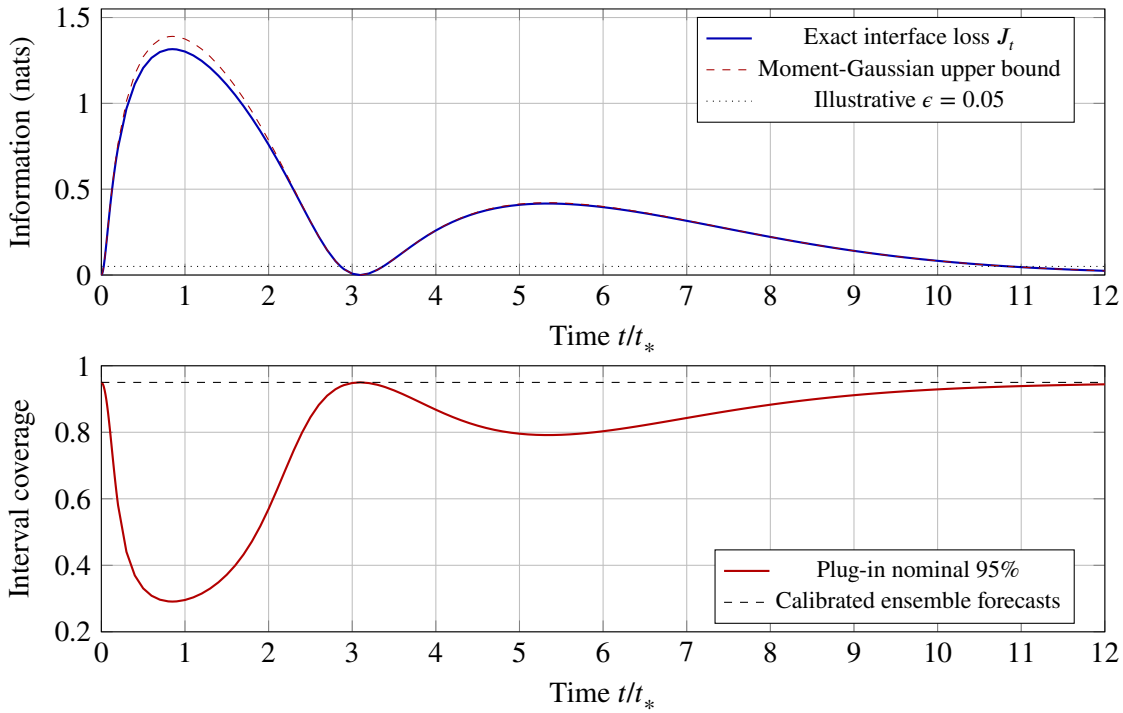


Figure 2: Information loss and coverage are time dependent. The reduced mixture remains calibrated for the declared ensemble, while the plug-in interval undercovers. A temporary cancellation near $t = 3.09267$ removes loss for one instantaneous target, followed by a second interval of non-negligible loss. Lines show deterministic calculations, not empirical confidence bands.

Table 3: Sensitivity to the chosen readout noise at the same early response maximum. All rates, energies, and $d = 0.8$ remain fixed. The 29.09% result is specific to the baseline $\sigma = 0.03$, not a universal failure rate.

σ	r	J (nats)	Nominal 95% plug-in coverage
0.015	13.4740	1.942005	0.1455
0.030	6.7370	1.315891	0.2909
0.060	3.3685	0.756808	0.5712
0.120	1.6843	0.330448	0.8361

With independent readouts, J^{joint} for $(t = 0.5, t_c)$ is 1.206199 nats, although $J_{t_c} = 0$. For $(0.5, 1.0)$ it is 1.560965 nats. The trajectory therefore remains informative about the missing coordinate even when one selected time cancels its effect.

For two same-time readouts with equal response g and correlation ρ, (33) becomes $r^2_{\text{joint}} = 2d^2g^2/[\sigma^2(1+\rho)]$. At the early peak the information is 1.623197 nats for independent noise and 1.318061 nats for $\rho = 0.99$, close to the single-readout value 1.315891. Exact duplicate data are handled as one readout or on their nonsingular support, not as two independent observations. Additional convergence checks cover these joint and sensitivity cases. A fixed-matrix numerical check also matches the spectral residual formula to within 2.22×10^{-16}; random-subspace comparisons are consistency checks, not substitutes for its proof.

6 Consequences for time-domain and multi-messenger modeling

An astrophysical application requires a physically meaningful pair of calculations, such as fluid evolution coupled to electromagnetic transport or hydrodynamics coupled to neutrino emission. Candidate retained information includes composition, angular structure, velocity distributions, optical depth, and closure memory. Its usefulness must be tested with a valid observation operator and source ensemble; the reservoir example does not identify the sufficient coordinates of a real event.

A decisive test would freeze the interface, basis, and thresholds on training configurations, then hold out complete source configurations. Matched resolution studies would separate discretization error from information loss that survives convergence. Held-out evaluation should report proper scores, coverage with sampling uncertainty, interval width, computational cost, and violated physical constraints. Alternative physical closures and source populations would test whether the selected coordinates remain useful.

Shared distance, source parameters, calibration, selection, non-detections, cadence, and latency must enter a joint likelihood or latent-variable model. Two related channels are not automatically independent evidence, as the correlated-readout calculation illustrates. Archived held-out tests establish performance under their access protocol; prospective forecasts additionally require a recorded origin, available data, software version, and target before verifying observations arrive. Reconstruction of past histories and prediction of new observations remain separate achievements.

7 Limitations and falsifiable extensions

The framework quantifies loss within a declared probabilistic model. Missing emission channels, an incorrect equation of state, detector misspecification, or population shift can invalidate a forecast despite an exact information calculation. Nonlinear operators can require state-dependent coordinates, and Gaussian approximations can miss multimodality, physical constraints, or rare events. The mode rule is consequently a conditional diagnostic, not a general model-selection theorem.

The benchmark has one unresolved scalar and a simplified radiative output. Its joint-target extension concerns multiple noisy readouts of that same process, not distinct astrophysical messengers. The analytic horizon bound uses this cascade's positivity and variance structure and does not transfer unchanged to nonlinear radiation hydrodynamics. Numerical thresholds and the 29.09% coverage value are design-dependent quantities.

Falsifiable next tests are whether retained coordinates improve held-out predictions beyond variance-only baselines, whether the benefit survives numerical convergence and alternative closures, and whether calibration persists under controlled population shifts. Failure would delimit the chosen interface. The present study establishes a conservative counterexample and model-specific bounds, not a new physical law or a validated universal forecasting technology.

8 Conclusions

Predictive information at a physical interface is specific to an observable, a horizon, an ensemble, and an observation model. Conditional mutual information separates the irreducible effect of discarded state information from avoidable misspecification of the downstream forecast. Under a linear–Gaussian model, singular directions of the noise-whitened future observation operator give an exact fixed-rank retention rule. These results turn a broad request to preserve "important physics" into a quantity that can be derived, bounded, and tested under explicit assumptions.

The positive three-reservoir example demonstrates the practical distinction. Equal energy and equal current luminosity coexist with different future signals. Omitting the unresolved contrast produces se-

vere overconfidence; marginalizing it honestly restores ensemble coverage at the price of wider intervals; transmitting an already-known upstream coordinate restores the state-informed reference. The analytic lower bound applies to a continuous marginal prefix, while joint-curve information requires a separate criterion. Noise and correlation tests show the dependence of the numerical effect on the declared observation model.

An integrated astrophysical calculation should therefore evaluate what it discards at each handoff, not only how accurately each solver processes what remains. The next scientific step is an independent application to a specified pair of physical codes with realistic observations and held-out source configurations. Success would consist of measured improvement and explicit limits of applicability; failure would provide equally useful constraints on which interfaces and approximations are inadequate.

Data and Code Availability

All study data are synthetic. The supplementary materials provide the complete Python reproduction script, numerical results, and plot data. The reported run used Python 3.12.14, NumPy 2.3.5, and SciPy 1.17.0. The mathematical model, parameters, and plotted coordinates are documented in the manuscript. No proprietary data are required.

Funding

This research received no external funding.

Conflicts of Interest

The author is developing a proposed collaborative research program to which the interface methodology may be relevant. This manuscript reports a separate methodological study and does not represent an evaluation or approval by that program's prospective funder.

References

[1] Kasen, D.; Metzger, B.; Barnes, J.; Quataert, E.; Ramirez-Ruiz, E. Origin of the heavy elements in binary neutron-star mergers from a gravitational-wave event. *Nature* **2017**, *551*, 80–84. doi:10.1038/nature24453.

[2] Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006. doi:10.1002/047174882X.

[3] Tishby, N.; Pereira, F.C.; Bialek, W. The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, 1999, pp. 368–377; preprint posted 2000. arXiv:physics/0004057.

[4] Chechik, G.; Globerson, A.; Tishby, N.; Weiss, Y. Information bottleneck for Gaussian variables. *Journal of Machine Learning Research* **2005**, *6*, 165–188. `https://jmlr.org/papers/v6/chechik05a.html`.

[5] Shalizi, C.R.; Crutchfield, J.P. Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics* **2001**, *104*, 817–879. doi:10.1023/A:1010388907793.

[6] Chorin, A.J.; Hald, O.H.; Kupferman, R. Optimal prediction and the Mori–Zwanzig representation of irreversible processes. *Proceedings of the National Academy of Sciences USA* **2000**, *97*, 2968–2973. doi:10.1073/pnas.97.7.2968.

[7] Spantini, A.; Cui, T.; Willcox, K.; Tenorio, L.; Marzouk, Y. Goal-oriented optimal approximations of Bayesian linear inverse problems. *SIAM Journal on Scientific Computing* **2017**, *39*, S167–S196. doi:10.1137/16M1082123.

[8] Wu, K.; Chen, P.; Ghattas, O. An efficient method for goal-oriented linear Bayesian optimal experimental design: Application to optimal sensor placement. Preprint, 2021, revised 2022. arXiv:2102.06627.

[9] Zhong, S.; Shen, W.; Catanach, T.; Huan, X. Goal-oriented Bayesian optimal experimental design for nonlinear models using Markov chain Monte Carlo. *SIAM/ASA Journal on Uncertainty Quantification* **2026**, *14*, 19–47. doi:10.1137/24M1649344.

[10] Chen, Q.; Arnaud, É.; Baptista, R.; Zahm, O. Coupled input-output dimension reduction: Application to goal-oriented Bayesian experimental design and global sensitivity analysis. *SIAM Journal on Scientific Computing* **2025**, *47*, A2403–A2430. doi:10.1137/24M1670421.

[11] Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **2007**, *102*, 359–378. doi:10.1198/016214506000001437.

[12] Gneiting, T.; Balabdaoui, F.; Raftery, A.E. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B* **2007**, *69*, 243–268. doi:10.1111/j.1467-9868.2007.00587.x.

[13] Peherstorfer, B.; Willcox, K.; Gunzburger, M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review* **2018**, *60*, 550–591. doi:10.1137/16M1082469.