The Finite-Observer Ledger: Boundary-Relative Accounting for Numerical Evidence

Jeffery Huckstead
Cerebral Graphix ORCID 0009-0007-0234-2177

Working draft v0.3 (production), 2026-07-13

Abstract

Words like *loss*, *absorption*, *escape*, *storage*, and *return* are ambiguous until an observer, a boundary, and a conserved quantity are declared. We propose a discipline that removes the ambiguity for numerical evidence: every claimed transfer is defined relative to a declared boundary and separated into named channels, a quantified closure test, a deterministic receipt, an explicit certification state, and explicit nonclaims. The individual ingredients are established—verification and validation, provenance, claim-to-evidence linkage, and limitation reporting all predate this—and our contribution is their integration and operationalization, inside a single executable notebook, into a per-claim structure specialized to boundary-aware physical accounting. We did not find, in the literature examined, an executable research object that co-locates all six fields as first-class per-claim structure, nor one that uses them to force a reduced-model-versus-full-experiment discrepancy into declared physical accounting channels. The discipline's value is demonstrated by what it caught: three distinct numerical error classes, and a flagship case in which two configurations sharing a reduced coordinate ($h/\sigma = 1/24$) produced full-simulation residuals differing by a factor of two, which the ledger opened into named channels and closed—via a pre-registered falsifier that removed 96.4% of the discrepancy—onto one timestep-policy mechanism. The companion paper supplies the exact mechanism; this paper supplies the account.

1 Introduction

Consider a bounded region of a physical field and ask where its energy went. The answer depends on choices that are usually left implicit: which surface counts as the boundary, which quantity is being tracked, over what interval, and whether a decrease means the quantity crossed out, was stored internally, was dissipated, or was simply mismeasured. In a finite computation these choices are not philosophical niceties—they are the difference between a trustworthy number and a plausible artifact. A reduced model and a full simulation can agree on a dimensionless coordinate yet disagree on the measured observable, and without a declared account the disagreement has nowhere to go.

This paper proposes that a finite-observer account is incomplete unless every claimed transfer is defined relative to a declared boundary and separated into named channels, a quantified closure test, a deterministic receipt, a certification state, and explicit nonclaims. We call the resulting per-claim object a *ledger*, and we argue that its natural home is an executable notebook, where the fields are not documentation written after the fact but structure that the computation is required to fill. The companion paper [1] develops an exactly solved boundary-scattering model; here that model is the executable lead example, and the discipline is the subject.

2 The finite-observer accounting problem

Fix a region, a boundary, a quantity, and an interval. The quantity may leave through a boundary face (crossing), return through a face (return), remain in the region (storage), or be removed by a dissipative term (dissipation); an observer may also pay a cost to measure it (observer cost); and every such statement carries a level of evidential standing (certification). The failure mode this paper targets is the collapse of these distinct accounts into a single word. A measured decrease in interior energy is not, by itself, "absorption"; it could be export, storage in a mode the diagnostic did not weight, numerical dissipation, or estimator bias. The accounting problem is to keep

$$\text{crossing} \neq \text{export} \neq \text{absorption} \neq \text{erasure} \tag{1}$$

separate, with each claim tied to the boundary that generated it and the receipt that certifies it. The remainder of the paper makes this operational.

3 The six-field ledger

The ledger assigns to each claim six required fields, in order: a declared *boundary* (which face or interface generated the measurement, and what does and does not count as crossing it); a *channel* decomposition (named transfers that sum to a measured total, each with a sign convention and a receipt pointer); a *closure* test (a quantitative balance identity with a residual, tolerance, and observed order); a deterministic *receipt* (configuration-stamped files with hashes and a reuse status); a *certification* state (an explicit epistemic grade, defined in Section 9); and explicit *nonclaims* (what the evidence does not establish, recorded as a first-class field rather than omitted). Figure 1 shows the chain.

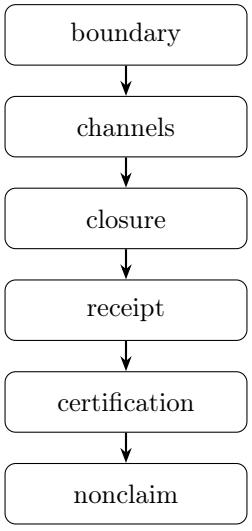

Figure 1: The six-field ledger. Each claim declares all six fields as first-class structure.

The matched-pair investigation of Section 5 required five additions to this core, which we retain: `reduced_coordinates` (the coordinates in which a claim is invariant, and the channels known to break that invariance); `estimator` (the fitting procedure and its known bias model, with a same-model calibration reference); `discrepancy_ledger` (a decomposition of a mismatch into channels with verdicts); `relations` (dependency and supersession links between claims); and

`status_history` (the timestamped record of state transitions). These are not decorative metadata. They were required to explain why the same reduced coordinate did not imply the same full experiment, and a schema that could not represent that distinction could not have caught it. The estimator field must declare the map from physical or numerical state to reported observable, together with any apparatus correction, because access to a representation does not by itself certify the state or mechanism that generated it.

We do not present this schema as a new primitive. Each field has a mature ancestor—validation domains and credibility scales in verification and validation [2, 3, 4]; flux and budget decompositions in computational fluid dynamics and climate modeling; provenance graphs and research-object packaging [5, 6]; claim-to-provenance linkage in nanopublications [7]; and explicit limitations in documentation cards [8, 9]. The contribution is the integration and operationalization of these into a single boundary-aware executable object, addressed in Section 8.

4 Three error classes caught by the executable ledger

The discipline is validated by what it disqualified. An ordinary reproducible notebook preserves a computation faithfully, including a faithfully wrong one; the ledger's value is that its closure and channel fields reject faithfully preserved but non-conjugate accounts. Three classes arose in the companion model.

Spatial conjugacy. A locally plausible centered flux, the "obvious" discretization of the boundary energy transfer, is not conjugate to the discrete energy being evolved; substituting it degrades the energy ledger's closure order from second to first. The closure test caught this: a flux that looked consistent failed the balance identity at the wrong order.

Temporal integration. A published stored-energy fraction was shown, under timestep attribution, to be more than 99.98% operator-splitting artifact, collapsing by a factor of several thousand as the split step was resolved. The channel decomposition caught this: a quantity attributed to physical storage was re-attributed to a named integrator channel. Figures effig:B-channel-attribution and effig:B-ledger-residual show representative executed evidence: named energy channels under fixed-grid timestep attribution and the corresponding relative ledger residual across timestep.

Boundary impedance. A ten-digit-constant reflection-error kernel was traced to a closed-form per-mode impedance defect of order $(kh)^2/8$, a property of the discretization's boundary symbol. The boundary declaration and closure test caught this: the error belonged to the declared face, not to the physics.

We keep three classes. The timestep-policy episode of Section 5—a second temporal mechanism, in which integrator amplitude error is routed through a boundary-parameter-dependent timestep policy—is a second *instance* of the temporal-integration class, not a fourth class. Both temporal instances share one failure mode: the time discretization is not conjugate to the comparison being drawn (operator splitting in the first, policy-induced branch asymmetry in the second). Two instances of one class strengthen the taxonomy; a fourth class named for one example would dilute it.

5 Flagship: same reduced coordinate, different full experiment

The clearest demonstration is a deliberately matched pair of configurations, $(n, \sigma) = (801, 3)$ and $(1601, 1.5)$, sharing the reduced coordinate $h/\sigma = 1/24$. In the infinite-domain exact-symbol model of the companion paper, all packet observables depend only on h/σ, so the two are identical—and their negative-branch solver outputs agree to a maximum absolute difference of 6.0×10^{-14}. Yet

their fully discrete estimator residuals differ by a factor of two: 4.815×10^{-8} against 9.631×10^{-8}. The reduced coordinate matched; the full experiment did not.

The ledger's discrepancy field opens this mismatch into named channels,

$$\Delta_{\mathrm{PDE-symbol}} = \Delta_{\mathrm{domain}} + \Delta_{\mathrm{sampling}} + \Delta_{\mathrm{init}} + \Delta_{\mathrm{timestep}} + \Delta_{\mathrm{estimator}} + \Delta_{\mathrm{floor}} + \Delta_{\mathrm{unresolved}}, \qquad (2)$$

and derived scalings exclude every channel but one. The survivor is the timestep-policy mechanism (Figure 2): the solver's stiffness-aware policy runs the $+\mathcal{R}$ and $-\mathcal{R}$ branches at different absolute timesteps, so classical Runge–Kutta dissipation enters the odd-in-$\mathcal{R}$ reflected-power channel with a scaling that is exactly the observed factor of two at fixed h/σ. A parameter-free model reproduces the residual curves to a fraction of a percent and was independently reproduced across implementations and agents on hash-pinned inputs.

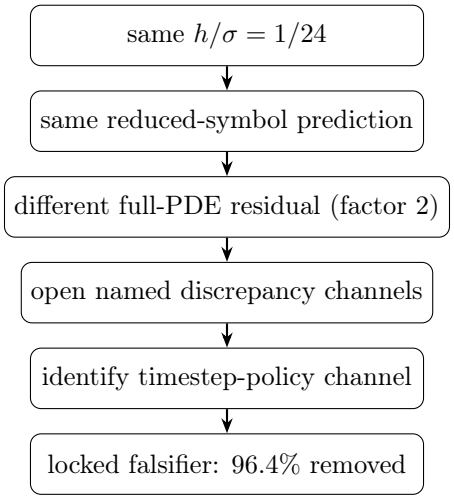

Figure 2: The flagship discrepancy ledger. A matched reduced coordinate, a divergent full experiment, and a closure onto one named channel confirmed by a pre-registered falsifier.

The account was then closed by a status transition recorded in advance. A falsifier was pre-registered: halving the offending branch's timestep should collapse the mechanism's leading contribution by a factor of 32, driving the estimator offset from 1.227×10^{-7} to a locked branch-aware target of 2.9776×10^{-8}. The realized offset was 2.9897×10^{-8}: within 0.41% of the locked target, removing 96.4% of the prior residual, with the small remainder recorded rather than absorbed. The ledger opened the accounts, and then emptied them into a named one.

6 The notebook as an executable ledger

In the notebook these fields are not prose. A claim record co-locates the boundary, the quantity and its normalization, the channels with their receipt pointers, the closure identity with its residual and observed order, the prediction with its lock timestamp and hash, the receipt files with their SHA-256 and solver hash, the certification state with its basis, and the nonclaims—and, for claims that are matched or that decompose a mismatch, the reduced coordinates, the estimator and its bias model, and the discrepancy ledger. Predictions are locked before the corresponding outputs exist; receipts are byte-reproducible; negative results and retired interpretations are kept with their status transitions rather than deleted. The diagnostic quantities that were demoted during the investigation—for example, a cross-observable ratio that compares two different observables and is

therefore not a closure metric—remain in the provenance under a diagnostic-only status: retired from claims is not deleted from evidence. Figure effig:B-flat-calibration shows the same discipline at the estimator level: the full dynamic range is retained and the measured minimum is declared rather than hidden by a local crop. The discipline predates the present model: an earlier executable phase-change cell already reported `ACCOUNTED_BUT_UNRESOLVED` when its energy and entropy balances closed but its information measure remained explicitly incomplete, keeping a fraction of a bit visible rather than burying it behind a zero residual.

Figure 3 shows the reading order a reviewer follows through one record. The same record drives three renderings: the formal papers, the machine-readable atlas, and the visual evidence companion, so that a claim, its equation, its figure, its receipt, its state, and its nonclaim never diverge across surfaces.

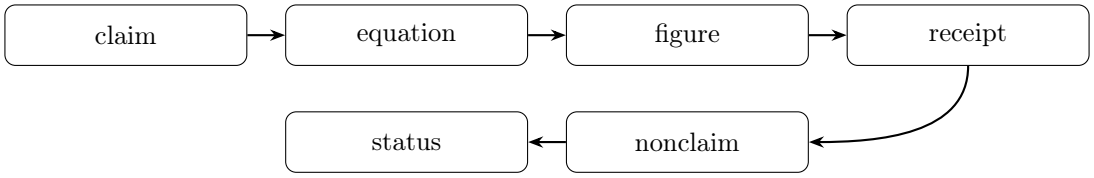

Figure 3: The claim-to-evidence chain a reviewer traverses per record, identical across paper, atlas, and evidence companion.

7 Case studies across physical systems

The ledger is applied to compact-object boundary accounting only at the certification actually earned, and the nonclaims are as load-bearing as the claims. Finite-radius export at the left face is a declared proxy, not an event-horizon flux; finite-radius export at the right face is a declared proxy, not a future-null-infinity flux. The barrier transmission and reflection bins are packet-spectrum-weighted fractions of the declared packet, not universal barrier constants. An executed characteristic control family is a finite-domain wave control problem, not a realistic neutron-star or stellar-surface model; a design-only control that was specified but not executed does not enter the evidence table as executed science. Across all cases, ledger closure is not physical boundary certification: the ledger certifies internal consistency, reproducibility, and stated-model closure, and leaves external validation and physical certification open by construction.

8 Related work and contribution boundary

The constituents of the ledger discipline are well established. Literate programming [10] and the reproducible-research tradition [11, 12, 13, 14] established the executable, re-runnable record; large-scale studies of notebook quality [15, 16] documented the gap between preservation and accounting. Verification and validation frameworks [2, 3, 17, 4] formalize credibility assessment and the decomposition of numerical, parametric, and model-form uncertainty. Provenance and research-object standards [5, 6] anchor computations to reproducible receipts; nanopublications [7] tie individual claims to provenance; artifact-evaluation badging [18] audits availability and functionality at the paper level; documentation cards and registered reports [8, 9] institutionalize explicit limitations. And physics-native closure tests—constraint-violation monitoring in numerical relativity, energy-budget closure in climate and fluid modeling—are routine accuracy diagnostics. Contemporaneous

claim-ledger systems for AI-assisted research [19, 20] implement per-claim epistemic states, preserved negative results, and evidence gates, but include neither a conservation-closure residual nor a boundary/channel decomposition, and are not specialized to physical accounting.

Our bounded contribution: we did not find, in the literature examined, an executable research object that co-locates all six ledger fields as first-class per-claim structure, nor one that uses them to force a reduced-model-versus-full-experiment discrepancy into declared physical accounting channels. We do not claim to have invented provenance, claim-to-evidence linkage, closure testing, certification states, or explicit nonclaims; each is prior art. We claim their integration and operationalization, specialized to boundary-aware physical accounting.

9 Certification states and nonclaims

The certification field takes a graded value, and the grades do not collapse into one another. A result may be *derived* (established by proof or symbolic computation); *numerically reproduced* (matched by an independent numerical run); *cross-implemented* (matched by a second, independent implementation); *cross-agent reproduced* (matched by an independent reasoning agent on hash-pinned inputs); *externally independently validated* (confirmed outside the project); or *physically certified* (validated against physical measurement). Two walls are enforced as schema rules. Internal cross-code or cross-agent agreement is not external validation: reproduction $\neq$ external validation. And a closed ledger is not a certified boundary: ledger closure $\neq$ physical certification. The companion model reaches cross-implementation and cross-agent reproduction on its attribution chain and derivation-plus-reproduction on its exact results; it does not reach external validation or physical certification, and the ledger records that ceiling explicitly rather than letting a strong internal result read as a physical one.

10 Package architecture

The publication package is four renderings of one claim spine. The formal papers carry the mathematical argument: the companion paper the exact mechanism, this paper the accounting architecture. The executable notebook is the reference implementation of the ledger, the layer in which predictions are locked and receipts are generated. The atlas is the machine-readable claim-state layer, one record per claim with its fields, receipts, and certification state. The visual evidence companion is the navigable map, generated from the same records so that a reader can move from a claim to its equation, figure, receipt, state, and nonclaim. The invariant across all four is the claim identifier: a single spine, rendered formally, executably, machine-readably, and visually, with no surface permitted to diverge from the others.

11 Discussion

The architecture solves a specific problem: it prevents a decrease from being called a loss, a reproduction from being called a validation, and a reduced coordinate from being mistaken for a full experiment. It does not certify physics, and it does not replace external replication; a closed ledger is a well-kept book, not a law of nature. Its conceptual ancestry is a prior essay [21] that distinguished accessible count from withheld order and insisted that internal satisfaction is not standing—that a claim earns shared status only when an independent path can reconstruct it. That essay supplied the gate; the notebook turned the gate into a schema. The lineage is genuine and bounded: the essay's distinctions (count $\neq$ order, address $\neq$ meaning, capacity $\neq$ content, structure $\neq$ identity)

become this paper's operational walls (crossing ≠ export ≠ absorption ≠ erasure; receipt ≠ interpretation), while the essay's black-hole, thermodynamic, and observer-independence claims are not imported as evidence for anything here.

The ledger also has a direct *operational* ancestor in the earlier Phase-Change Ledger [22], which treated a single finite sensing medium as sensor, memory, and operating reservoir at once. That construction already forced the separations this paper formalizes: persistent record versus indefinitely reusable capacity, explicit heat and control cost, and storage capacity versus recoverable temporal order, with readout treated as apparatus-corrected rather than a closed cosmic ledger and finite-radius observation fenced from exact asymptotic export. Its role here is architectural, not evidentiary: no phase-change result is used to support the wave-model mathematics.

12 Conclusion

A bounded claim needs a bounded account. When every transfer declares its boundary, names its channels, closes its balance, carries its receipt, states its certification, and records what it does not establish, an unexplained discrepancy has somewhere to go: it is opened into channels, tested, and either closed onto a named mechanism or left honestly in an unresolved account. The companion paper shows what the mechanism was; this paper shows how the evidence was kept.

A Claim-to-receipt seed table

Table 1: Paper B claim seed. Numbering is provisional and tied to labels, not frozen.

ID	Claim	Anchor
FO-B-001	finite-observer ledger thesis	Sec. 1
FO-B-002	six-field claim schema	Sec. 3, Fig. 1
FO-B-003	spatial-conjugacy error class	Sec. 4
FO-B-004	temporal-integration error class	Sec. 4; Figs. 4–5
FO-B-005	boundary-impedance error class	Sec. 4
FO-B-006	matched-pair discrepancy ledger	Sec. 5, Fig. 2
FO-B-007	locked-falsifier status transition	Sec. 5
FO-B-008	notebook-as-executable-ledger	Sec. 6, Figs. 3, 6
FO-B-009	certification wall	Sec. 9
FO-B-010	package-as-evidence-system	Sec. 10

B Figure provenance

Figures 1, 2, and 3 are LaTeX-native TikZ diagrams; they carry no external image provenance. The three selected executed plots included below are copied byte-for-byte from the frozen notebook image export and bind to `img/image_manifest.csv` by SHA-256:

- Fig. 4: notebook cell 61, sha `f9e34e13ea28512d1dc1c359841c0395 74c9b497d310167193f18d606f9b34a5`.
- Fig. 5: notebook cell 63, sha `a772132792f8ed273c7294cff7be3b4e 7abbf08ee7fcd9e8a08ece729daed1dd`.

- Fig. 6: notebook cell 120, sha `7f869cc603a2316d84b8478899421b18800af4ae0feedaa069158580796f49b9`.

The full gallery remains in the archive and visual evidence companion; these three are included in the paper because they show the ledger's practice directly.

C Selected visual evidence

The plots below are not decorative illustrations. They are selected examples of how the ledger makes channels, closure, and estimator calibration visible to a human reader while retaining the receipt-bound source image.

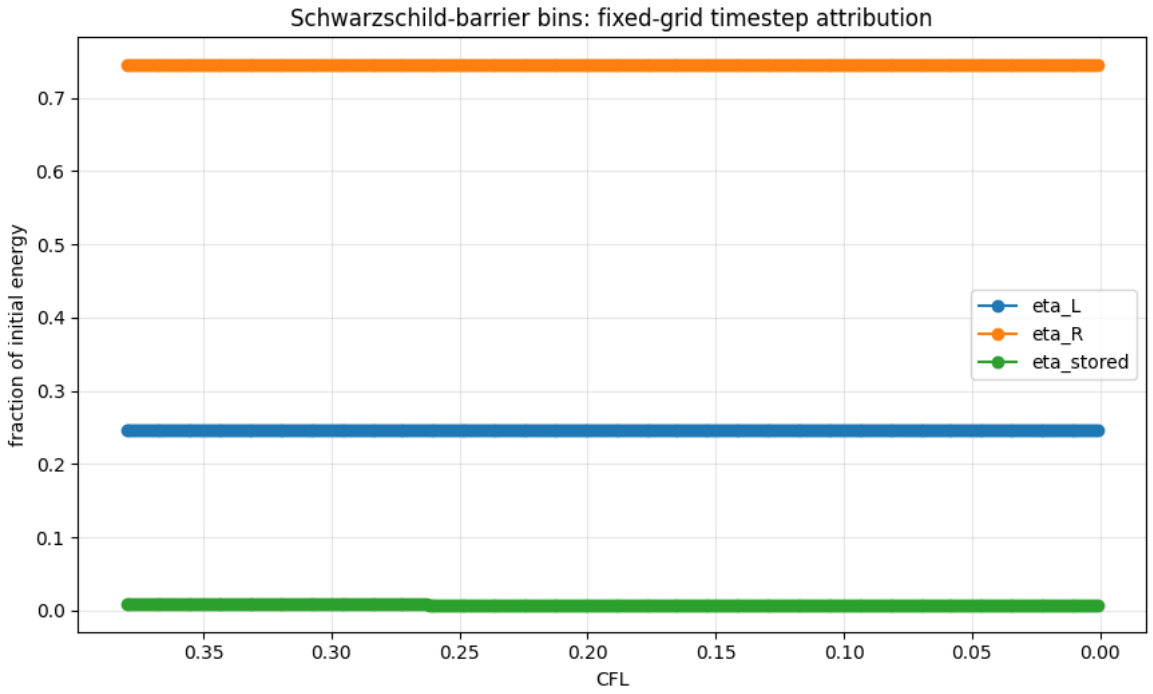

Figure 4: Named Schwarzschild-barrier energy channels under fixed-grid timestep attribution. The separate left-export, right-export, and stored-energy fractions show the channel decomposition before interpretation.

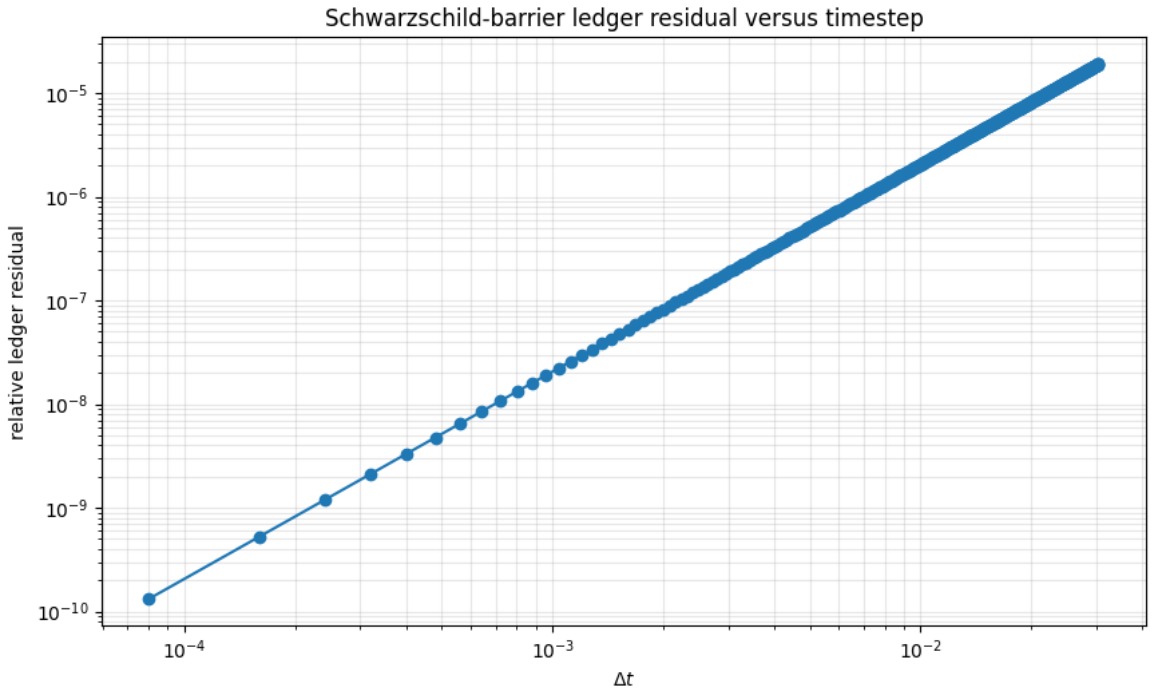

Figure 5: Relative Schwarzschild-barrier ledger residual versus timestep over a multi-decade range. The residual is shown directly rather than hidden behind a single summary number, illustrating closure as a quantitative field of the ledger.

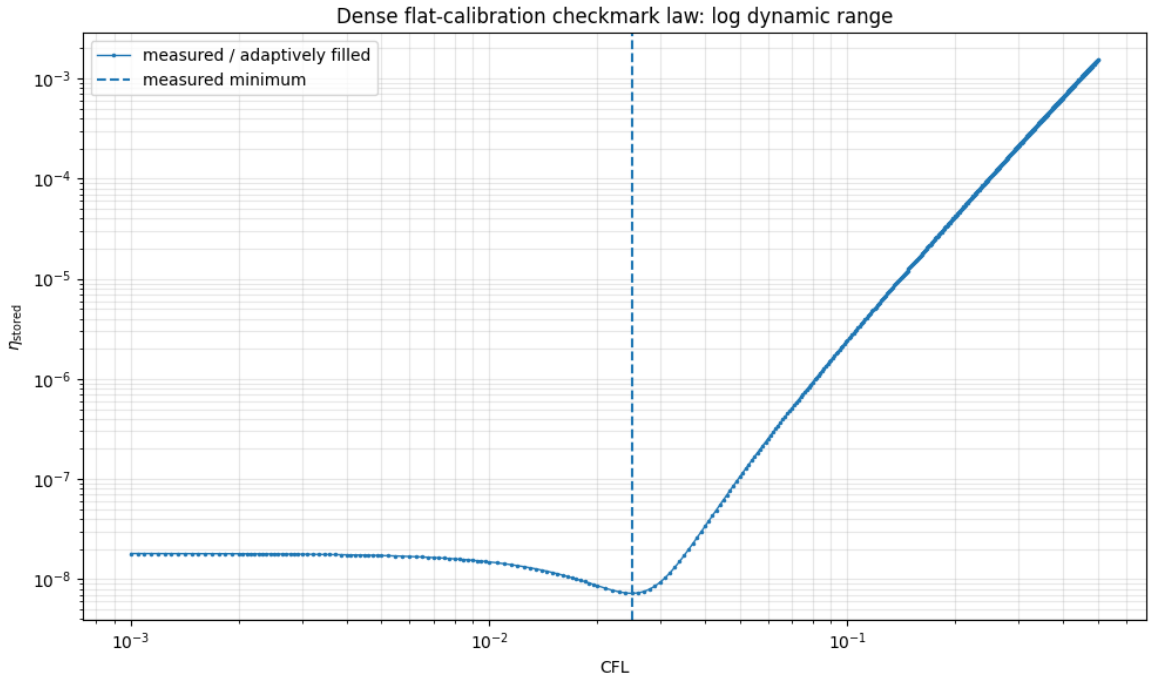

Figure 6: Dense flat-calibration checkmark law across the full log dynamic range, with the measured minimum declared explicitly. The plot preserves both the low-CFL floor and the high-CFL rise, preventing a local crop from becoming the whole story.

References

[1] Jeffery Huckstead. From measure to flux: Energy-conjugate bookkeeping and exact semidiscrete scattering in finite radial wave models. *companion manuscript, this package*, 2026. v0.3 production draft.

[2] William L. Oberkampf and Christopher J. Roy. *Verification and Validation in Scientific Computing*. Cambridge University Press, 2010.

[3] ASME. V&v 10: Standard for verification and validation in computational solid mechanics. Technical report, American Society of Mechanical Engineers, 2019. Document-level citation; clause text not independently verified.

[4] NASA. Nasa-std-7009: Standard for models and simulations. Technical report, National Aeronautics and Space Administration, 2008. Document-level citation; clause text not independently verified.

[5] Timothy Lebo, Satya Sahoo, and Deborah McGuinness. PROV-O: The PROV ontology. W3C Recommendation, 2013.

[6] Stian Soiland-Reyes et al. Recording provenance of workflow runs with RO-Crate. *PLoS ONE*, 19(9):e0309210, 2024.

[7] Paul Groth, Andrew Gibson, and Jan Velterop. The anatomy of a nanopublication. *Information Services & Use*, 30(1-2):51–56, 2010.

[8] Margaret Mitchell et al. Model cards for model reporting. In *Proceedings of FAccT 2019*, 2019.

[9] Timnit Gebru et al. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021.

[10] Donald E. Knuth. Literate programming. *The Computer Journal*, 27(2):97–111, 1984.

[11] Jonathan B. Buckheit and David L. Donoho. Wavelab and reproducible research. In *Wavelets and Statistics*, volume 103 of *Lecture Notes in Statistics*, pages 55–81. Springer, 1995.

[12] Roger D. Peng. Reproducible research in computational science. *Science*, 334(6060):1226–1227, 2011.

[13] Geir Kjetil Sandve, Anton Nekrutenko, James Taylor, and Eivind Hovig. Ten simple rules for reproducible computational research. *PLoS Computational Biology*, 9(10):e1003285, 2013.

[14] Mark D. Wilkinson et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018, 2016.

[15] Adam Rule, Aurélien Tabard, and James D. Hollan. Exploration and explanation in computational notebooks. In *Proceedings of CHI 2018*, 2018.

[16] João Felipe Pimentel, Leonardo Murta, Vanessa Braganholo, and Juliana Freire. A large-scale study about quality and reproducibility of Jupyter notebooks. In *Proceedings of MSR 2019*, pages 507–517, 2019.

[17] ASME. V&v 20: Standard for verification and validation in computational fluid dynamics and heat transfer. Technical report, American Society of Mechanical Engineers, 2009. Document-level citation; clause text not independently verified.

[18] ACM. Artifact review and badging, version 1.1, 2020.

[19] Y. Xia and T. Wang. Researchloop: An evidence-gated control plane for AI-assisted research, 2026. Preprint.

[20] M. Koren. Theorist toolbox: Tools for agent-based LLM-assisted economic theory research, 2026. Preprint.

[21] Jeffery Huckstead. One ledger, and the withholding of the second, 2026. Cerebral Graphix; conceptual predecessor essay.

[22] Jeffery Huckstead. The phase-change ledger, 2026. Cerebral Graphix; operational predecessor to the finite-observer ledger.