Referee Reports

Paper 41: "The Diagnostic in Action: Axiom Calibration of the AdS/CFT Correspondence"

February 15, 2026

Report 1: Constructive Reverse Mathematics Expert

Referee specialization: Constructive analysis, omniscience principles, Weihrauch degrees, reverse mathematics (classical and constructive).

Summary

This paper applies constructive reverse mathematics (CRM) to the holographic entanglement entropy formulae of AdS/CFT, classifying each computational step by the weakest non-constructive principle required. The central claim is that the holographic dictionary preserves axiom cost: bulk and boundary computations carry identical logical resources at every level examined. The paper is accompanied by a 901-line Lean 4 formalization.

Assessment

Strengths

S1. **Correct and non-trivial use of the CRM hierarchy.** The paper correctly identifies the strict inclusions BISH $\subset$ LLPO $\subset$ WLPO $\subset$ LPO and the independence of FT. The equivalence LLPO $\leftrightarrow$ totality of the real order ($\forall x\, y.\ x \leq y \vee y \leq x$) is correctly attributed and correctly applied to the phase-decision problem. The equivalence LPO $\leftrightarrow$ BMC is central to the QES infimum argument and is handled correctly.

S2. **The observable/decision distinction is genuine and well-motivated.** The separation between computing $\min(x, y)$ (which is BISH via the algebraic identity) and deciding the comparison $x \leq y \vee y \leq x$ (which is LLPO) is a real and important distinction in constructive analysis. This is sometimes overlooked even in the specialist literature, and the paper makes the point cleanly.

S3. **The scaffolding/infimum separation is the paper's strongest contribution.** The claim that computing $\inf_\gamma S_{\text{gen}}(\gamma)$ costs LPO while locating the minimizer γ^* costs FT is well-founded in constructive analysis. This is essentially the distinction between the completeness of $\mathbb{R}$ (every Cauchy sequence converges, which is BISH) and the compact existence principle (every continuous function on a compact set attains its extremum, which requires FT). The paper correctly identifies the Bounded Monotone Convergence principle as the mechanism for computing the infimum.

S4. **The introduction and background are excellent.** Section 1.1 provides an accurate and accessible summary of CRM, correctly citing Bishop (1967), Bishop–Bridges (1985), Bridges–Richman (1987), Ishihara (2006), Diener (2018), and the 2023 Handbook. The cellar/cathedral metaphor from Paper 12 is helpful for non-specialist readers.

S5. **The calibration table is well-organized.** Table 2 provides a clear summary of all calibrations. The mechanism column is a useful addition.

Weaknesses and Concerns

W1. **The FT formulation is non-standard.** The paper defines FT as the extreme value theorem for continuous functions on $[0, 1]$ (Definition in `Defs.lean`). While this is equivalent to the Fan Theorem in the constructive setting (via the work of Julian and Richman), it is not the standard statement of the Fan Theorem, which concerns uniform continuity of functions on the Cantor fan or, equivalently, bar induction on the binary tree. The paper should note this equivalence explicitly and cite the relevant result (e.g., Bridges–Richman 1987, or Diener 2018, Chapter 3). Without this, a CRM specialist may question whether the "FT" being used is genuinely Brouwer's Fan Theorem or merely the extreme value theorem.

W2. **The** LPO $\leftrightarrow$ BMC **equivalence needs more care.** The paper states this equivalence and cites Paper 29, but the standard reference is Ishihara (2006) for the result that LPO is equivalent to "every bounded monotone sequence of reals converges." The precise formulation matters: does BMC assert convergence of all bounded monotone sequences, or only those with a computable modulus? In the former case, the equivalence with LPO is well-established. The paper should clarify this point.

W3. **The** LLPO $\leftrightarrow$ **real comparison claim is stated without proof.** The paper axiomatizes this as a bridge axiom (`llpo_iff_real_comparison`), but this is actually a well-known result in constructive analysis—it is provable, not an axiom. It appears in Bridges–Richman (1987) and in Bishop–Bridges (1985, Chapter 2). Treating a provable theorem as a bridge axiom weakens the formalization unnecessarily. I would recommend that future versions prove this equivalence directly in Lean.

W4. **The bridge axiom methodology raises foundational questions.** 12 of the 18 results are proved using bridge axioms—axioms that encapsulate physics input. While the paper is transparent about this methodology, a CRM purist would note that the "calibration" is only as reliable as the bridge axioms. For instance, `camporesi_heat_kernel_bish` simply asserts that a real number exists (`∃ S : ℝ, True`). This does not encode any computational content—it does not assert that the heat kernel integral converges with a computable modulus, or that the Sommerfeld image sum has a BISH Cauchy modulus. The bridge axiom is computationally vacuous. The paper's claim that FLM is BISH is therefore a *claim about physics*, not a *theorem of constructive analysis*.

W5. **No connection to Weihrauch degrees.** The paper cites Brattka–Gherardi (2012) but does not engage with the Weihrauch degree framework. The omniscience hierarchy has a precise Weihrauch-degree characterization, and the calibration results could in principle be stated as reductions between Weihrauch degrees. This would connect the results to the broader computable analysis community. This is not a requirement, but would strengthen the paper.

W6. **The "logical isomorphism" claim is slightly overstated.** The paper claims the holographic dictionary is a "logical isomorphism." Strictly, this means a bijection between logical theories that preserves provability. What the paper actually shows is that the axiom *cost* (as measured by the CRM hierarchy) is the same on both sides. This is weaker than a logical isomorphism in the usual sense. I suggest "axiom-cost equivalence" or "logical congruence" as more precise terminology.

Minor Comments

m1. The hierarchy diagram in the paper lists only strict inclusions. It would be helpful to note that all inclusions are known to be strict—e.g., there are Brouwerian models separating LLPO from WLPO.

m2. The FanTheorem definition uses $\leq$ in the conclusion, so it is the "maximum" version of EVT. For the infimum in the QES analysis, one needs the "minimum" version. The paper should note that these are equivalent by negation.

m3. The "BISH + LPO" notation is non-standard in CRM. Conventionally, BISH + LPO denotes the theory obtained by adding LPO to the axioms of BISH. The paper uses it this way, but the convention should be stated explicitly.

Verdict

The paper makes a genuine contribution by applying CRM to a new domain (holographic entanglement entropy) and identifying real structural distinctions (observable vs. decision, infimum vs. minimizer). The CRM content is largely correct, though the formulation of FT and the vacuity of bridge axioms are concerns. The "logical isomorphism" terminology should be softened.

Recommendation: Major revision. The CRM content is sound in its broad strokes, but the bridge axiom methodology needs a more careful epistemological discussion (what exactly is being proved vs. assumed?), the FT formulation needs clarification, and LLPO $\leftrightarrow$ real comparison should be proved rather than axiomatized.

Report 2: Theoretical Physics Expert

Referee specialization: Quantum gravity, holographic entanglement entropy, AdS/CFT correspondence, black hole information problem.

Summary

The paper examines the Ryu–Takayanagi formula and its quantum extensions (FLM, QES, island formula) through the lens of constructive reverse mathematics. It classifies each computation by the logical strength needed and argues that the holographic dictionary preserves this "axiom cost."

Assessment

Strengths

S1. **The physics pipeline is correctly described.** The paper accurately presents the progression RT → HRT → FLM → QES → island formula, with correct references. The introduction (Section 1.2) provides a competent summary of holographic entanglement entropy from Bekenstein through the replica wormhole programme. References are current and appropriately placed.

S2. **The Engelhardt–Wall framing is handled well.** The paper correctly treats the QES prescription as physics input to the calibration, not as a target for verification. The statement that "the physically meaningful part of the QES computation is the saddle-point competition, not the surface existence" is a genuinely interesting observation. It aligns with how practitioners use the QES formula: the entropy value is what appears in physical predictions, while the surface itself is an intermediate construct.

S3. **The perturbative/non-perturbative distinction is physically meaningful.** The observation that the perturbative QES (via the Jacobi equation) is BISH while the non-perturbative QES requires LPO is a real distinction. It corresponds to the breakdown of the semiclassical expansion—a transition that physicists care about. The paper correctly identifies the Jacobi geodesic deviation equation as a Lipschitz ODE, solvable by Picard–Lindelöf.

S4. **The Page curve analysis is correct.** The decomposition of the Page curve into continuous entropy (BISH) and discrete Page-time decision (LLPO) is a clean observation. The identification of the Page time with a weak omniscience principle is novel and physically interesting.

S5. **The discussion sections on replica wormholes and the information paradox are well-written.** The paper correctly identifies the open problem of calibrating the gravitational path integral sum over topologies.

Weaknesses and Concerns

W1. **The physical content is thin.** The paper does not compute any new physical quantity. All physics results (RT formula, Calabrese–Cardy, Brown–Henneaux, FLM, QES, island formula) are taken directly from the literature and repackaged with constructive labels. The paper is, at its core, a classification exercise. While classification can be valuable, the physical payoff is unclear: does knowing that a computation is "BISH" vs. "LPO" change any physical prediction? Does it constrain new physics? The paper does not address this.

W2. **The model is AdS_3/2d CFT, which is atypically simple.** The BTZ black hole is exactly solvable. The geodesic problem in AdS_3 reduces to elementary functions. The Calabrese–Cardy formula is an exact result. The paper acknowledges this (Section 13.1, "higher-dimensional holography" as an open question), but the generalizability of the calibration is untested. In AdS_5, the minimal surface problem is a genuine PDE (not an algebraic formula), and the geodesic problem requires solving elliptic integrals. Whether the BISH calibration of the vacuum RT survives in higher dimensions is a substantive open question that the paper does not resolve.

W3. **The FLM "calibration" is essentially vacuous.** The paper claims that FLM is BISH based on the Camporesi heat kernel. But the actual Lean axiom (`camporesi_heat_kernel_bish`) merely asserts that a real number exists. No actual heat kernel computation is formalized. No Sommerfeld image sum is implemented. No zeta-function regularization is carried out. The claim "FLM is BISH" is a physics assertion, not a machine-checked proof. The paper should be more transparent about this distinction.

W4. **No treatment of bulk reconstruction.** The paper focuses on entanglement entropy but does not address bulk reconstruction—the problem of recovering bulk operators from boundary data. Bulk reconstruction involves non-trivial analytic continuation (the HKLL prescription) and modular flow (the Jafferis–Lewkowycz–Maldacena–Suh construction), both of which could introduce logical costs not present in the entropy calculation. This is a significant gap if the paper aims to calibrate "the holographic dictionary" broadly.

W5. **The "duality consistent" result is tautological as formalized.** The Lean formalization defines a calibration table with hard-coded axiom costs, then checks that bulk and boundary entries match. This is checking internal consistency of the table, not deriving the axiom costs from first principles. The "duality consistent" theorem is thus a consistency check on the author's classifications, not an independent result.

W6. **Quantum error correction perspective is absent.** Modern understanding of AdS/CFT centrally involves the entanglement wedge and quantum error correction (Dong–Harlow–Wall, Hayden–Penington). The error correction interpretation of holography may introduce additional logical structure not captured by the entropy-focused analysis.

W7. **The Discussion overreaches in places.** The statement (Section 13.5) that "what survives the descent from cathedral to cellar is everything physically measurable" is a philosophical claim that goes beyond what the calibration table demonstrates. The paper has examined only entanglement entropy in 2+1 dimensions. Extrapolating to "everything physically measurable" is not warranted by the evidence presented.

Minor Comments

m1. The HRT covariant proposal is listed as an open question (dynamical holography, Section 13.1), but the paper does not discuss the subtleties of Lorentzian signature—e.g., the difference between spacelike extremal surfaces and quantum extremal surfaces in time-dependent backgrounds.

m2. The paper mentions the Cubitt–Perez-Garcia–Wolf undecidability result. This is for the spectral gap of a lattice Hamiltonian, which is a very different setting from holographic entanglement. The connection is tenuous and should be qualified.

m3. The replica trick (Section 4.1, boundary derivation) involves analytic continuation in the replica number n, which is a notoriously non-rigorous step. Labeling this as "BISH" without addressing the mathematical status of analytic continuation from integers to reals is optimistic.

Verdict

The paper is well-written and the physics is correctly described at the level of a review article. The classification exercise is carried out competently, and several observations (perturbative vs. non-perturbative QES, Page time $\sim$ LLPO, scaffolding mechanism) are genuinely interesting. However, the physical payoff of the classification is unclear, the formalization of the physics content is shallow (bridge axioms with trivial content), and the scope (AdS_3) limits generalizability.

Recommendation: Major revision. The paper should (1) clarify the physical payoff of the CRM classification, (2) be more transparent about the gap between the physics claims and the Lean formalization, (3) discuss the limitations of the AdS_3 setting more prominently, and (4) address the absence of bulk reconstruction and quantum error correction perspectives.

Report 3: Lean Formalization Expert

Referee specialization: Lean 4, Mathlib, interactive theorem proving, formalization methodology, verification of mathematical physics.

Summary

The paper presents a 901-line Lean 4/Mathlib formalization that classifies holographic entanglement entropy computations by their constructive strength. The formalization comprises 8 modules, 12 bridge axioms, and 6 genuine proofs. It compiles with 0 sorry and 0 warnings.

Assessment

Strengths

S1. **The genuine proofs are clean and correct.** The 6 genuine proofs—`lpo_implies_wlpo`, `wlpo_implies_llpo`, `lpo_implies_llpo`, `min_eq_algebraic`, `no_observable_exceeds_lpo`, `duality_consistent`—are verified by the Lean type checker and use appropriate Mathlib lemmas. The hierarchy proofs are standard but well-executed. The `min_eq_algebraic` proof using `le_total`, `abs_of_nonpos`/`abs_of_nonneg`, and `ring` is idiomatic Lean.

S2. **The code is well-documented.** Every definition and theorem has a docstring explaining its purpose and mathematical context. The module header comments are informative. The code is readable by someone familiar with Lean 4 but not with the physics.

S3. **The module architecture is sensible.** The dependency graph (Figure 1) reflects the logical structure of the calibration. `Defs.lean` provides a clean interface of types and bridge axioms. The leaf modules prove section-level results without cross-dependencies. `CalibrationTable.lean` assembles results into a verifiable table. This is a reasonable design for a project of this size.

S4. **The 0-sorry, 0-warning build is correctly reported.** This is a genuine achievement for a Lean project. The paper is transparent about the Lean/Mathlib version and build instructions.

S5. **The `#print axioms` audit is correctly interpreted.** The paper correctly notes that `Classical.choice` in the axiom printout is an artifact of Mathlib's construction of $\mathbb{R}$ as a Cauchy completion, not a use of classical logic in the proofs. This is a common source of confusion in Lean formalization, and the paper handles it well.

Weaknesses and Concerns

W1. **The bridge axioms carry no computational content.** This is the most serious concern. Consider the bridge axiom:

```
axiom camporesi_heat_kernel_bish :
    exists (_S_bulk : R), True
```

This asserts only that a real number exists. It does not encode: (a) that the heat kernel has a specific functional form, (b) that the image sum converges with a computable modulus, (c) that the regularized entropy equals a specific value, or (d) that the computation is "BISH" in any formal sense. The axiom is logically equivalent to $\top$ (trivially true) and carries zero information.

Similarly, `zeta_reg_finite_bish` asserts $\exists x : \mathbb{R}, \mathsf{True}$, which is provable without any axiom at all ($\langle$`0, trivial`$\rangle$). These bridge axioms do not constrain the formalization in any way.

Impact: The FLM calibration theorem (`FLM_correction_bish`) follows trivially from a trivially true axiom. The "0 sorry" claim is technically correct but misleading: the bridge axioms do the work that `sorry` would have done, but without the red flag in the build output. A `sorry` would be more honest, as it explicitly signals incompleteness.

W2. **The `BISHComputable` structure is a placeholder.** The definition:

```
structure BISHComputable (f : R -> R) : Prop where
  computable : True
```

is logically vacuous. Every function is trivially `BISHComputable` by $\langle$`trivial`$\rangle$. In a formalization that claims to distinguish BISH from LPO, the inability to formally characterize BISH-computability is a fundamental limitation. The paper acknowledges this ("Mathlib $\mathbb{R}$ uses `Classical.choice` anyway"), but the acknowledgment does not resolve the problem.

Suggestion: The paper could use a more structured approach—e.g., tagging functions with a `CRMLevel` inductive type and verifying that compositions respect the hierarchy. This would not capture true computability, but it would provide a non-trivial type-level classification.

W3. **The calibration table is hard-coded, not derived.** The `CalibrationTable.lean` module defines a list of `CalibrationEntry` values with explicit axiom costs:

```
{ name := "Vacuum RT", bulk_cost := .BISH,
  boundary_cost := .BISH, duality_preserves := true }
```

The theorems `no_observable_exceeds_lpo` and `duality_consistent` check properties of this hard-coded list. They do not derive the axiom costs from the proofs themselves. A more ambitious formalization would compute axiom costs from proof terms—but this is admittedly beyond current Lean capabilities.

W4. **Several "genuine proofs" are trivial assemblies.** The `BTZ_entropy_bish` theorem:

```
theorem BTZ_entropy_bish ... := by
  obtain <L1, L2, hc1, hc2, _hform> :=
    BTZ_geodesic_lengths p
  exact <L1, L2, hc1, hc2,
    fun t => by rw [min_eq_algebraic]>
```

This destructures a bridge axiom and applies a genuine lemma. It is not a genuine proof—it is an assembly of a bridge axiom with a proved lemma. The paper's CRM audit correctly distinguishes "genuine proofs" from "bridge axiom assemblies," but the distinction could be clearer in the code itself (e.g., via attributes or naming conventions).

W5. **The omniscience hierarchy does not prove strict separation.** The paper proves LPO $\Rightarrow$ WLPO $\Rightarrow$ LLPO but does not prove (or even state) that these implications are strict. In Lean, proving strictness would require constructing models—e.g., a realizability model where LLPO holds but WLPO does not. This is a well-known limitation (model construction in Lean is hard), but the paper should acknowledge it.

W6. **The LLPO $\leftrightarrow$ real comparison is axiomatized rather than proved.** As the CRM referee also notes, this is a standard theorem of constructive analysis (see Bishop–Bridges, Chapter 2). In Lean/Mathlib, the forward direction (real comparison $\rightarrow$ LLPO) would require constructing a specific binary sequence from real number data, which is feasible using Mathlib's `Real` API. This is a missed opportunity for a genuine proof.

W7. **No testing or CI pipeline is mentioned.** The paper states "`lake build`" as the build command, but does not mention continuous integration, regression testing, or Mathlib version pinning strategy. For reproducibility, the `lean-toolchain` and `lake-manifest.json` should be committed and the build should be tested in CI.

Minor Comments

m1. The Lean code listings in the paper use `«<` and `»>` for angle brackets. This is a LaTeX rendering choice, not a Lean syntax issue, but it may confuse readers unfamiliar with the convention.

m2. Several bridge axioms use the pattern `∃ x : ℝ, True`. This is logically equivalent to `True` and should be noted in the paper as a minimal stub rather than a meaningful specification.

m3. The `noncomputable section` at the top of each file is necessary because of Mathlib's `Real`, but it means that none of the definitions are computationally executable. This is ironic for a paper about constructive computability, and should be discussed more prominently.

m4. The paper's claim of "901 lines" includes blank lines, comments, and module headers. A more informative metric would be "lines of proof" (e.g., counting only tactic blocks and term-mode proofs).

m5. The `CalibrationPair` structure does not include a field for the proof witness linking the entry to the actual theorem. Adding a field like `witness : Prop` with a proof obligation would make the table non-trivially connected to the formalization.

Verdict

The formalization is a competent Lean 4 project with clean code, good documentation, and a genuine 0-sorry build. The genuine proofs (hierarchy, min identity, table consistency) are well-executed. However, the bridge axioms are computationally vacuous, the `BISHComputable` structure is a placeholder, and the calibration table is hard-coded rather than derived. The formalization verifies the *structure* of the argument ("if these bridge axioms hold, then the table is consistent") but does not verify the *content* ("these computations are genuinely BISH-computable").

Recommendation: Minor revision. The formalization is honest about its limitations (the paper's CRM audit table clearly distinguishes genuine proofs from bridge axiom assemblies). The main revisions needed are: (1) discuss the vacuity of the bridge axiom stubs more prominently, (2) prove LLPO $\leftrightarrow$ real comparison rather than axiomatizing it, (3) add a discussion of the `BISHComputable` placeholder and the fundamental tension between Mathlib's classical $\mathbb{R}$ and the paper's constructive claims, and (4) consider replacing the most trivially true bridge axioms with more structured type signatures that encode meaningful specifications.

Editorial Summary

Points of Agreement Across Reviewers

1. The observable/decision distinction and the infimum/minimizer (scaffolding) separation are the paper's strongest contributions and are recognized as genuine by all three reviewers.
2. The bridge axiom methodology is the paper's most significant weakness. All reviewers note that the bridge axioms (especially `camporesi_heat_kernel_bish` and `zeta_reg_finite_bish`) carry no computational content, making the FLM calibration a physics claim rather than a formal result.
3. The paper is well-written and well-organized. The introduction provides adequate background for non-specialists. The code is clean and well-documented.
4. The scope limitation (AdS_3/2d CFT) is noted by both the physics and CRM reviewers. The generalizability claims should be tempered.

Combined Recommendation

CRM Expert:	Major Revision
Physics Expert:	Major Revision
Lean Expert:	Minor Revision
Editorial Decision:	**Major Revision**

Key Revisions Required

1. **Epistemological transparency.** Clearly delineate what is *proved* (hierarchy, min identity, table structure) from what is *claimed* (FLM is BISH, QES infimum is LPO) vs. what is *axiomatized* (bridge axioms). The current paper makes this distinction in the CRM audit table, but the main text sometimes blurs the boundary.
2. **Prove rather than axiomatize where feasible.** LLPO $\leftrightarrow$ real comparison is provable in Lean/Mathlib and should be a genuine proof.
3. **Clarify physical payoff.** Does the CRM classification predict anything? Constrain anything? Suggest new computations?
4. **Strengthen bridge axiom specifications.** Replace trivially-true bridge axioms with more informative type signatures encoding the computational structure being claimed.
5. **Temper scope claims.** The analysis covers entanglement entropy in AdS_3. Claims about "the holographic dictionary" broadly should be qualified.