Testing the Bekenstein-Hawking Area Law at Cosmological Scales: A Data-Driven Reconstruction of the Horizon Entropy Current

Aiden B. Smith[1]

[1] *Independent Researcher*

The Bekenstein-Hawking area law implies a constant horizon entropy slope, $\mu(A) \propto dS/dA =$ const. We test this assumption directly at cosmological scales using a data-driven reconstruction of the apparent-horizon entropy current from late-time background geometry (Pantheon+, BAO, and cosmic chronometers), in a framework designed to avoid GR-anchored circularity. In the baseline mapping, the reconstructed slope statistic is negative, with posterior mean $\langle d\log\mu/d\log A\rangle = -0.492$ and posterior weight $P(d\log\mu/d\log A < 0) = 0.904$. Across mapping variants, all converged runs keep a negative mean slope. However, decision-grade calibration under the Bekenstein-Hawking null ($\mu = 1$) does not yet support a smoking-gun rejection: with 2000 null simulations, the calibrated exceedance is $p_{\rm one} = 0.296$ (95% Monte Carlo CI $[0.276, 0.316]$), and low-false-alarm thresholds tuned to $\alpha = 0.05$ and 0.01 have weak power for plausible injected slopes. The current evidence is therefore a suggestive preference for negative slope, not a calibrated exclusion of the strict area law.

I. INTRODUCTION

The thermodynamic interpretation of gravity links horizon geometry to entropy. In standard General Relativity with Bekenstein-Hawking entropy, the horizon entropy is proportional to area, $S \propto A$, so dS/dA is constant. A central question for late-time cosmology is whether this area law remains valid on the apparent horizon traced by the observed expansion history.

This paper is framed as a direct test of that assumption. Rather than starting from a fixed gravitational field equation and fitting parameters, we reconstruct the entropy-current proxy $\mu(A)$ from observed background geometry and then ask whether $\mu(A) = 1$ is supported. In this sense, the analysis is not only a reconstruction exercise; it is a fundamental consistency test of space-time thermodynamics at cosmological scales.

The same thermodynamic degree of freedom is also the ingredient that controls modified gravitational-wave propagation in dark-siren sectors. The present work therefore serves as the theory-facing companion to dark-siren propagation tests: here we infer the background thermodynamic current, while dark sirens test its propagation consequence.

II. THERMODYNAMIC LINK AND INFERENCE STRATEGY

We define the entropy-slope modification

$$\mu(A) \equiv \frac{(dS/dA)_{\rm BH}}{dS/dA}, \tag{1}$$

so that $\mu = 1$ corresponds to strict Bekenstein-Hawking behavior. In the apparent-horizon mapping used in this pipeline, the late-time background obeys

$$\frac{dH^2}{dz} = 3H_0^2\Omega_{m0}(1+z)^2\mu(A), \qquad A(z) = 4\pi\left(\frac{c}{H(z)}\right)^2 \tag{2}$$

for the flat mapping variant.

The inference is performed by forward modeling: we sample a spline representation of $\log\mu$ and solve for $H(z)$, then evaluate the likelihood in data space. The baseline geometry stack is Pantheon+, BAO, and cosmic chronometers. Inference products are stress-tested with synthetic closure, simulation-based calibration (SBC), and explicit Bekenstein-Hawking null exceedance experiments using the same production settings as the real-data run.

III. RESULTS

A. Area-law test from reconstructed entropy current

The reconstructed posterior favors a negative entropy-current slope over a constant-slope area law. A representative summary is

$$\left\langle \frac{d\log\mu}{d\log A}\right\rangle_{\rm V1,free} = -0.492, \qquad P\left(\frac{d\log\mu}{d\log A} < 0\right)_{\rm V1,free} = 0.904. \tag{3}$$

Interpreted physically, this corresponds to a decreasing $\mu(A)$ as the horizon area grows, i.e., an evolving entropy density on cosmological horizons. Sign stability across mapping variants is preserved in this run: V0_fixedOm (-0.483, $P_{<0} = 0.903$), V1_curved (-0.453, $P_{<0} = 0.881$), V1_free (-0.492, $P_{<0} = 0.904$), and V2_residual (-0.285, $P_{<0} = 0.764$).

B. Decision-grade calibration and BH-null exceedance

We calibrate thresholds directly under the Bekenstein-Hawking null using full end-to-end simulations. For each target false-alarm level α, a threshold is selected from 2000 null simulations and then evaluated for injected-slope power (500 simulations per injected level).

At $\alpha = 0.05$, the calibrated threshold is -1.4475, with achieved FPR $= 0.050$ (95% CI $[0.041, 0.060]$), but power is low: TPR $= 0.028, 0.032, 0.052$ for injected slopes $-0.3, -0.2, -0.1$ respectively. At $\alpha = 0.01$, the threshold is -1.9846, with achieved FPR $= 0.010$ (95% CI $[0.0065, 0.0154]$) and TPR $= 0.006, 0.012, 0.010$ for the same injections.

Using the observed real-data scalar statistic $s_{\rm obs} = -0.4922$ (baseline V1_free), the calibrated null exceedance from 2000 Bekenstein-Hawking simulations is

$$p_{\rm one} = P(s_{\rm sim} \leq s_{\rm obs}) = 0.296, \qquad p_{\rm two} = 0.5465. \quad (4)$$

With the pre-defined stop rule, these results classify the current run as *suggestive preference* rather than a calibrated rejection of $\mu = 1$.

C. Hubble-scale anchors and robustness context

For transfer-map stress tests, we track the conventional anchor scales $H_0^{\rm Planck} = 67.4 \pm 0.5$ and $H_0^{\rm local} = 73.0 \pm 1.0$ $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ as external references. These anchors are used for sensitivity mapping, not as hard priors in the core area-law test, to preserve modified-gravity neutrality in the reconstruction stage.

D. Directional null test (Pantheon+ hemispheres and Planck lensing)

Because the entropy-current reconstruction is built from low-redshift geometry, it is important to check that the recovered trend is not dominated by a particular sky region or survey footprint. We therefore perform an explicit statistical-isotropy diagnostic using a hemisphere-split scan on the Pantheon+ sky: for each axis on a coarse HEALPix grid, we compute the slope-scar statistic s separately in opposite hemispheres and form Δs and its uncertainty.

To suppress look-elsewhere effects, we use a five-fold crossfit procedure in which the "best" axis is chosen on a training split (maximizing $|\Delta s|/\sigma_{\Delta s}$) and then evaluated only once on held-out data. The held-out confirmations are consistent with noise, with fold-level $z_{\rm test} \in \{-0.354, -0.254, -0.028, +0.488, +0.065\}$ and a convenient aggregate diagnostic $z_{\rm Stouffer} \approx -0.037$.

As an external cross-check, we evaluate a hemispherical Planck 2018 CMB lensing convergence variance statistic along the strongest training axis, finding $\Delta \log \mathrm{Var}(\kappa) \approx 0.032$ with a random-axis calibrated fixed-axis p-value $p \approx 0.22$ ($z \approx 1.23$). Within current sensitivity, these diagnostics provide no evidence for a robust dipolar modulation of the inferred slope statistic.

IV. DISCUSSION

The key physical output is two-part: the posterior trend is directionally negative, but the calibrated decision layer does not yet reject the strict area law at low false-alarm settings. A decreasing $\mu(A)$ remains a plausible interpretation of the reconstructed geometry, yet the BH-null exceedance rate and low power under stringent FPR control imply that the present dataset and pipeline settings do not constitute smoking-gun evidence.

The calibration and closure layers are included specifically to separate posterior preference from decision-grade rejection. In this run, mapping variants agree on slope sign and converge, but the low-FPR power and null-tail probability fail the pre-specified decision threshold.

V. CONCLUSION

We recast late-time entropy-slope reconstruction as a fundamental cosmological test of the Bekenstein-Hawking area law. In the reconstructed posterior, the horizon entropy current is biased toward a negative slope and this sign preference is stable across tested mapping variants.

At the same time, decision-grade calibration with BH-null simulations shows that the present evidence is *suggestive* rather than decisive: calibrated null exceedance is not small and low-FPR operation has weak power at plausible injected effect sizes. The immediate implication is methodological: stronger data, more informative cross-probe predictive tests, or higher-power test statistics are required before claiming a smoking-gun violation of $\mu(A) = \text{const}$.

DATA AND SOFTWARE AVAILABILITY

Software archive for this pipeline (Zenodo title: *Negative Entropy Slope*): doi:10.5281/zenodo.18604922.

External data/product DOIs used by the underlying stack include:

- GWTC-3 products: doi:10.1103/PhysRevX.13.041039.
- GLADE+: doi:10.1093/mnras/stac1443.
- Pantheon+ papers and dataset: doi:10.3847/1538-4357/ac8b7a, doi:10.3847/1538-4357/ac8e04, doi:10.5281/zenodo.16365279.
- Cosmic-chronometer compilations: doi:10.1088/1475-7516/2012/08/006, doi:10.1088/1475-7516/2016/05/014.
- BOSS DR12 BAO+FS: doi:10.1093/mnras/stx721.
- eBOSS DR16 LRG BAO+RSD: doi:10.1093/mnras/staa2455.

- DESI 2024 BAO analyses: doi:10.1088/1475-7516/2025/04/012, doi:10.1088/1475-7516/2025/02/021.
- Planck 2018 lensing: doi:10.1051/0004-6361/201833886.

AI-USE STATEMENT

AI systems were used for drafting support and editing during manuscript preparation.

[1] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), Phys. Rev. X **13**, 041039 (2023), doi:10.1103/PhysRevX.13.041039.

[2] G. Dálya *et al.*, Mon. Not. R. Astron. Soc. **514**, 1403 (2022), doi:10.1093/mnras/stac1443.