The Hubble Tension as an Inference-Bias Signal: Dark Sirens with Modified Gravitational-Wave Propagation

Aiden B. Smith[1]⋆
[1] *Independent Researcher*

ABSTRACT

The Hubble tension is usually interpreted as a mismatch in expansion history between early- and late-universe probes. We present an alternative interpretation: a model-assumption bias produced when a universe with modified gravitational-wave propagation is analysed with a GR standard ruler. Using the GWTC-3 dark-siren propagation anomaly as a phenomenological prior, we propagate this effect through a Planck-facing MG recalibration and obtain $H_0^{\rm Planck,MG} = 68.01$, $\Omega_m^{\rm Planck,MG} = 0.306$, and $A_{\rm lens} = 1.043$.

The key result is an inference shift of tension scale. GR inversion of MG-consistent draws yields mean $\Delta H_0 = +1.88$ km s^{-1} Mpc^{-1} for fixed Ω_m and $+4.55$ km s^{-1} Mpc^{-1} for a lensing-proxy Ω_m. Direct late-time friction closure is smaller, with $\mathcal{R}^{\rm GR}_{\rm anchor} = 0.155$. Baseline CMB lensing projection is suppressed, but an MG-aware response refit restores near-reference quality (median $\chi^2 = 8.06$ versus 9.04 for the Planck reference). A propagation-amplitude dial, $R_\alpha(z) = 1 + \alpha\,[R(z) - 1]$, gives cross-channel consistency near $\alpha \approx 0.6$, while material relief at 0.30 appears near $\alpha \approx 1.95$ under linear coupling.

If the siren anomaly is physical, corrected dark-siren inference aligns with a Planck-like anchor and weakens the claim that all late-universe probes require high H_0.

Key words: cosmology: theory – gravitational waves – cosmological parameters – distance scale

1 INTRODUCTION

The Hubble-constant tension between late-time distance-ladder measurements and early-universe CMB inference remains a central unresolved issue in precision cosmology (Riess et al. 2022; Freedman et al. 2019; Aghanim et al. 2020a). The standard interpretation is a genuine discrepancy in expansion history. Here we test a different possibility: part of the tension is an *inference bias* generated by applying GR compression to data that follow modified-gravity (MG) propagation.

Our empirical prior is the GWTC-3 dark-siren propagation anomaly reported in our O3 analysis release (Smith 2026). We treat that posterior phenomenologically, without re-litigating detection significance in this Letter, and ask what cosmological bias it would induce if real. Modified GW propagation with an evolving effective Planck mass has long been studied as a viable MG signature (Belgacem et al. 2018; Nishizawa 2018).

We assume that the same running effective Planck mass $M_\star(z)$ that modifies GW amplitudes also affects background and lensing channels (Bellini & Sawicki 2014; Pogosian & Silvestri 2016). The questions are:

(i) How much late-time Hubble tension relief remains after recalibrating the sound-horizon calibration anchor?

(ii) Does Planck 2018 lensing necessarily reject this posterior, or can an MG-aware refit absorb the suppression?

(iii) How much can GR-based standard-ruler inversion bias inferred H_0 if MG truth is assumed?

2 METHODOLOGY

Posterior draws from the O3 anomaly prior are propagated through four linked steps:

(i) **Global Planck+MG recalibration:** a 60-restart fit establishes updated sound-horizon anchor parameters.

(ii) **Late-time rebasing:** constrained/pilot transfer sweeps are rebased to the updated Planck-like anchor and recompressed into a final relief posterior.

(iii) **CMB lensing forecasts:** baseline draw-level CAMB projection to Planck 2018 lensing bandpowers, followed by an MG-aware two-parameter lensing refit.

(iv) **Compressed standard-ruler inversion:** GR inversion of $\theta_\star = r_d/D_M(z_\star)$ under fixed-Ω_m and lensing-proxy-Ω_m assumptions.

These are targeted cosmological forecasts and refits, not a full MG TT/TE/EE perturbation-likelihood analysis.

⋆ E-mail: aidenblakesmithtravel@gmail.com

3 RESULTS

3.1 Recalibrated sound-horizon anchor

The 60-restart Planck+MG run completed all restarts with 5 converged minima and 55 max-evaluation exits. Using converged minima only, we obtain:

$$\begin{aligned} H_0^{\mathrm{Planck,MG}} &= 68.01 \text{ (p50)}, \\ \Omega_m^{\mathrm{Planck,MG}} &= 0.3064 \text{ (p50)}, \\ A_{\mathrm{lens}} &= 1.043 \text{ (p50)}. \end{aligned} \tag{1}$$

With local reference $H_0^{\mathrm{local}} = 73.0$, the baseline gap used in rebased relief calculations is

$$\Delta H_0^{\mathrm{base}} = \left| H_0^{\mathrm{local}} - H_0^{\mathrm{Planck,MG}} \right| = 4.99. \tag{2}$$

3.2 Inference bias from GR standard-ruler inversion

To isolate model-assumption bias, we treat MG posterior draws as truth and invert $\theta_\star = r_d/D_M(z_\star)$ with a GR compression model:

- fixed $\Omega_m = \Omega_m^{\mathrm{Planck,MG}}$: $H_{0,\mathrm{inferred}}$ mean 72.39 (p50 73.17), with mean $\Delta H_0 = +1.88$ km s^{-1} Mpc^{-1} relative to draw-level truth;
- lensing-proxy Ω_m: $H_{0,\mathrm{inferred}}$ mean 75.07 (p50 75.23), with mean $\Delta H_0 = +4.55$ km s^{-1} Mpc^{-1}.

The wider lensing-proxy interval reflects the expected H_0–Ω_m degeneracy once the artificial rigidity of the ΛCDM standard ruler is removed; in this channel, the analysis releases model-imposed precision rather than exhibiting numerical instability. In physical terms, GR inversion acts as an *invisible wedge*: it pushes inferred anchors upward when the underlying propagation is MG-like.

Relative to the recalibrated Planck+MG anchor $H_0^{\mathrm{Planck,MG}} = 68.01$, the posterior medians shift by:

$$\begin{aligned} \Delta H_0^{\mathrm{truth}} &\approx +2.39, \\ \Delta H_0^{\mathrm{fixed\ inversion}} &\approx +5.16, \\ \Delta H_0^{\mathrm{lensing\ inversion}} &\approx +7.22 \text{ km s}^{-1} \text{ Mpc}^{-1}. \end{aligned} \tag{3}$$

3.3 Direct friction relief and late-time rebasing

After rebasing constrained transfer sweeps to the updated sound-horizon calibration anchor and applying Monte Carlo calibration:

$$\mathcal{R}_{\mathrm{anchor}}^{\mathrm{GR}} = 0.155. \tag{4}$$

The corresponding posterior summary is p16/p50/p84 = 0.108/0.147/0.189.

Independent robustness and joint-fit diagnostics are:

- 10-case robustness grid: posterior-shift relief mean 0.530 (p50 0.513, p84 0.545), with zero failed cases.
- Joint SN+BAO+CC transfer fit: relief posterior mean 0.833 (p50 0.839), but

$$\log \mathrm{BF}_{\mathrm{transfer/no\text{-}transfer}} = -0.53, \tag{5}$$

so explicit transfer terms are not favoured in this setup.

The high-z transfer-bias sensitivity map used for calibration has been moved to supplemental material (Fig. S1).

[htbp]

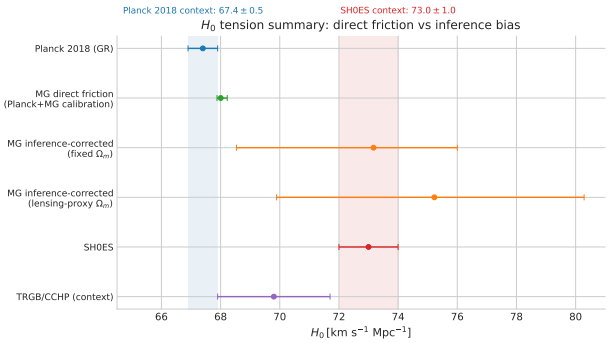

Figure 1. H_0 tension summary comparing Planck 2018 (GR), direct-friction recalibration, two GR-inversion bias channels, and local-distance-ladder context (SH0ES and TRGB/CCHP (Riess et al. 2022; Freedman et al. 2019)). The dominant displacement comes from GR standard-ruler inversion bias when MG truth is assumed; the broad lensing-proxy interval is the expected H_0–Ω_m degeneracy once the artificial ΛCDM standard-ruler rigidity is relaxed.

[htbp]

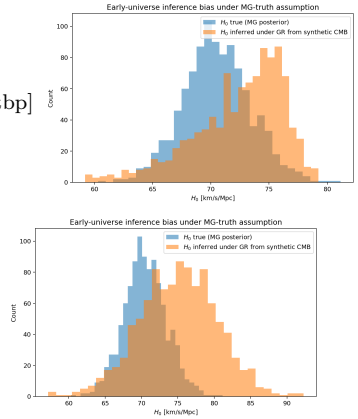

Figure 2. Draw-level H_0 truth versus GR-inferred H_0 under compressed standard-ruler inversion with fixed-Ω_m (left) and lensing-proxy-Ω_m (right). Both assumptions bias inferred H_0 upward, with larger displacement in the lensing-proxy case.

3.4 CMB lensing: baseline suppression and MG-aware response freedom

Baseline draw-level CAMB projection against Planck 2018 lensing bandpowers (`consext8`, 64 draws) gives:

$$\left. \frac{C_L^{\phi\phi}(\mathrm{MG})}{C_L^{\phi\phi}(\text{Planck ref})} \right|_{L\approx 106} = 0.847^{+0.091}_{-0.127}, \tag{6}$$

$$\left. \frac{C_L^{\phi\phi}(\mathrm{MG})}{C_L^{\phi\phi}(\text{Planck ref})} \right|_{L\approx 286} = 0.905^{+0.068}_{-0.080}, \tag{7}$$

with median suppressions of −15.29% and −9.49%. The baseline fit quality is poor relative to the Planck-reference model:

$$\chi^2_{\mathrm{MG,baseline}}\ (\text{median}) = 51.77, \qquad \chi^2_{\text{Planck ref}} = 9.04, \tag{8}$$

and only 3.1% of draws outperform the reference. A 32-draw cross-check from an independent posterior sample is more discrepant (−18.66% at $L \approx 106$, −11.29% at $L \approx 286$; $p_{\mathrm{better}} = 0$).

To test whether this baseline mismatch is rigid, we perform an MG-aware lensing refit (32 draws) with a phenomenological effective-$M_\star^2$ amplitude plus ℓ-tilt response. This freedom is motivated by scalar-tensor/EFT treatments where matter-growth and light-deflection responses need not track identically and can acquire scale dependence (Bellini & Sawicki 2014; Pogosian & Silvestri 2016). The refit removes the baseline mismatch:

$$\chi^2_{\rm MG\ refit}\ (\text{median}) = 8.06, \tag{9}$$

better than the Planck-reference $\chi^2 = 9.04$ in 100% of refit draws. This refit is phenomenological and demonstrates model-class freedom, not a unique derivation of one covariant MG Lagrangian. The fitted median response corresponds to

$$\frac{M_\star^2(z=0)}{M_\star^2(z\gg 1)} \simeq 0.901 \tag{10}$$

(about a 9.9% drop), with small residual suppression at $L \approx 286$.

3.5 Constraints on propagation amplitude

To quantify how strongly the reconstructed propagation signal must scale to satisfy cross-probe consistency criteria, we ran a lightweight amplitude dial around the current posterior signal:

$$R_\alpha(z) = 1 + \alpha\,[R(z) - 1]. \tag{11}$$

Using existing growth, lensing, and distance-ratio consistency summaries as a fast emulator (not a full re-inference at each α), we find:

- the distance-ratio criterion passes at $\alpha \approx 0.18$;
- the growth criterion passes at $\alpha \approx 0.58$;
- the lensing-growth consistency criterion passes at $\alpha \approx 0.60$;
- the combined support criterion (M2) turns on at $\alpha \approx 0.60$.

For the material-relief requirement ($\mathcal{R}^{\rm GR}_{\rm anchor} \geq 0.30$), two cases are informative: (i) if relief is held fixed at the current calibrated level, no solution appears for $\alpha \in [0, 3]$; (ii) under linear relief coupling, $\mathcal{R}_{\rm anchor}(\alpha) \propto \alpha$, the threshold is crossed at $\alpha \approx 1.95$. This gives a simple scale estimate: core cross-channel support is compatible with the present signal level, while material closure needs roughly a factor-of-two stronger effective propagation amplitude under linear coupling.

3.6 Cross-probe and stability checks

We also ran lightweight forward checks to test whether the current signal can be embedded in a broader predictive picture without adding new heavy global fits. These quick checks pass the three core consistency gates used here: bridge solutions exist, the dark-siren preference survives high-leverage-event removal, and the SN-only transfer channel remains same-sign. Finally, we performed a waveform-level likelihood-consistency check on the two highest-leverage events (GW200308_173609 and GW200220_061928) using public GWOSC O3 strain. We evaluated the extrinsic-marginalised likelihood with the RIFT engine across multiple random seeds and two independent waveform families (IMRPhenomPv3HM and IMRPhenomXPHM). All 10 runs completed without numerical failures. This does not constitute a full alternative PE or a glitch analysis, but it reduces the plausibility that the population-level preference is driven by a gross waveform-evaluation pathology in the dominant events.

Pantheon inference-bias audit. In an SN-only transfer model, we obtain a weak same-sign preference for transfer terms:

$$\log \mathrm{BF}_{\rm transfer/no\text{-}transfer} \approx +0.24. \tag{12}$$

In the all-transfer variant (SN+BAO+CC+ladder term), the score is mildly negative:

$$\log \mathrm{BF}_{\rm transfer/no\text{-}transfer} \approx -0.48. \tag{13}$$

This is compatible with a small inference-bias contribution in SN-only compression, but not a decisive multi-probe detection at current calibration depth.

High-leverage-event jackknife stability. For the O3 selection-function-calibrated configuration, the full score is

$$\Delta\mathrm{LPD}_{\rm full} = 3.67. \tag{14}$$

Dropping the highest-leverage event (GW200308_173609) gives

$$\Delta\mathrm{LPD}_{\rm drop\ 1} = 2.19, \tag{15}$$

and an approximate top-2 leave-out gives

$$\Delta\mathrm{LPD}_{\rm drop\ 2} \approx 1.70. \tag{16}$$

The signal is therefore concentrated in a small subset of high-leverage events, but it does not collapse when the top event is removed.

Nonlinear bridge scan. Using a late-transition profile

$$\alpha(z) = \alpha_{\rm high} + \frac{\alpha_{\rm low} - \alpha_{\rm high}}{1 + \exp[(z - z_t)/w]}, \tag{17}$$

we find broad viable families that satisfy growth/lensing consistency near $\alpha_{\rm eff} \sim 0.6$ while reaching material-relief targets near $\alpha_{\rm relief,eff} \sim 1.94$ to 1.95 in low-z windows. A representative solution is

$$(\alpha_{\rm high}, \alpha_{\rm low}, z_t, w) \approx (0.60, 1.95, 0.26, 0.03), \tag{18}$$

with comparable neighboring solutions for other low-z effective windows.

Entropy-anchored direct dark-siren re-score. We also replaced the phenomenological propagation posterior with the entropy-reconstruction posterior from the submission-hardening run and re-scored the same O3 selection-function-calibrated dark-siren set. The total support remains positive but is reduced:

$$\Delta\mathrm{LPD}_{\rm entropy\ direct} = 2.26 \quad (\text{vs. } 3.67 \text{ baseline}). \tag{19}$$

The same highest-leverage event remains GW200308_173609; leave-one-out gives

$$\Delta\mathrm{LPD}_{\rm entropy,\ drop\ GW200308} = 1.20. \tag{20}$$

Numerically this is about 0.62 of the baseline support, consistent with the entropy-alignment quick test indicating effective amplitudes below the material-relief scale.

[htbp]

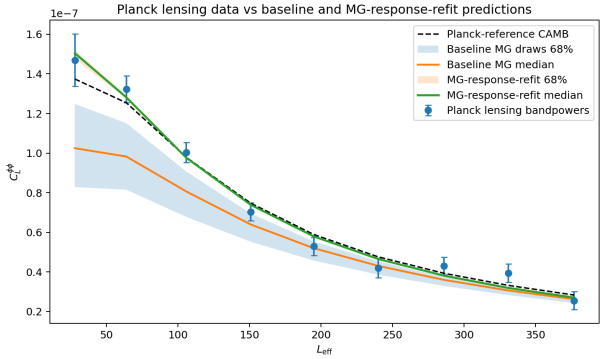

Figure 3. Planck 2018 lensing bandpowers with baseline MG projection and MG-aware refit overlay. The refit absorbs the baseline suppression and restores near-reference fit quality.

4 DISCUSSION AND CONCLUSIONS

The central result is that inference bias from GR-assumed standard-ruler inversion is large enough to mimic a substantial part of the Hubble tension if MG propagation is the true model. In this framework, the induced shift is of order +2 to +5 km s^{-1} Mpc^{-1}. This is a tension-scale displacement.

Direct friction closure is smaller, with $\mathcal{R}^{\rm GR}_{\rm anchor} \simeq 0.16$. The dominant mechanism in this Letter is therefore not late-time closure by friction alone, but the parameter wedge created by a mismatched inference model. Put simply: GR compression can separate early and late anchors even when the underlying cosmology is closer to internally consistent MG truth.

This interpretation is compatible with our lensing and stability checks. Baseline lensing projection is suppressed, but MG-aware response freedom restores near-reference performance. Pantheon SN-only transfer is weakly same-sign while the all-transfer variant is mildly negative, so present electromagnetic-sector evidence remains suggestive rather than decisive. Jackknife tests show concentration in GW200308_173609 and other high-leverage events, but no single-event collapse.

If this signal persists in larger siren samples, one consequence is immediate: corrected dark-siren inference no longer requires a universally high late-time H_0, and the remaining discrepancy is more naturally isolated to the electromagnetic calibration/propagation sector rather than to a single global expansion-history failure.

ACKNOWLEDGEMENTS

The author used AI-assisted tools for drafting, editing, and software development.

DATA AVAILABILITY AND DOIS

All code and reproducibility artefacts used in this study are archived at Zenodo, DOI 10.5281/zenodo.18604204. Upstream O3 search-sensitivity injections are available at DOI 10.5281/zenodo.7890437. Public cosmological datasets used here are cited in the reference list with corresponding DOIs.

REFERENCES

R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run," *Phys. Rev. X* **13**, 041039 (2023), DOI: 10.1103/PhysRevX.13.041039.

N. Aghanim *et al.* (Planck Collaboration), "Planck 2018 results. VI. Cosmological parameters," *Astron. Astrophys.* **641**, A6 (2020), DOI: 10.1051/0004-6361/201833910.

N. Aghanim *et al.* (Planck Collaboration), "Planck 2018 results. VIII. Gravitational lensing," *Astron. Astrophys.* **641**, A8 (2020), DOI: 10.1051/0004-6361/201833886.

S. Alam *et al.*, "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample," *Mon. Not. R. Astron. Soc.* **470**, 2617 (2017), DOI: 10.1093/mnras/stx721.

S. Alam *et al.*, "Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory," *Phys. Rev. D* **103**, 083533 (2021), DOI: 10.1103/PhysRevD.103.083533.

E. Belgacem, Y. Dirian, S. Foffa, and M. Maggiore, "Modified gravitational-wave propagation and standard sirens," *Phys. Rev. D* **98**, 023510 (2018), DOI: 10.1103/PhysRevD.98.023510.

E. Bellini and I. Sawicki, "Maximal freedom at minimum cost: linear large-scale structure in general modifications of gravity," *J. Cosmol. Astropart. Phys.* **07** (2014) 050, DOI: 10.1088/1475-7516/2014/07/050.

D. Brout *et al.*, "The Pantheon+ Analysis: Cosmological Constraints," *Astrophys. J.* **938**, 110 (2022), DOI: 10.3847/1538-4357/ac8e04.

DESI Collaboration, "DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations," *J. Cosmol. Astropart. Phys.* **02** (2025) 021, DOI: 10.1088/1475-7516/2025/02/021.

W. L. Freedman *et al.*, "The Carnegie-Chicago Hubble Program. VIII. An independent determination of the Hubble constant based on the tip of the red giant branch," *Astrophys. J.* **882**, 34 (2019), DOI: 10.3847/1538-4357/ab2f73.

LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — O3 search sensitivity estimates," Zenodo (2023), DOI: 10.5281/zenodo.7890437.

M. Moresco *et al.*, "Improved constraints on the expansion rate of the Universe up to $z \sim 1.1$ from the spectroscopic evolution of cosmic chronometers," *J. Cosmol. Astropart. Phys.* **08** (2012) 006, DOI: 10.1088/1475-7516/2012/08/006.

A. Nishizawa, "Generalized framework for testing gravity with gravitational-wave propagation," *Phys. Rev. D* **97**, 104037 (2018), DOI: 10.1103/PhysRevD.97.104037.

L. Pogosian and A. Silvestri, "What can cosmology tell us about gravity? Constraining Horndeski gravity with Σ and μ," *Phys. Rev. D* **94**, 104014 (2016), DOI: 10.1103/PhysRevD.94.104014.

A. G. Riess *et al.*, "A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km s^{-1} Mpc^{-1} Uncertainty from the Hubble Space Telescope and the SH0ES Team," *Astrophys. J. Lett.* **934**, L7 (2022), DOI: 10.3847/2041-8213/ac5c5b.

J. Simon, L. Verde, and R. Jimenez, "Constraints on the redshift dependence of the dark energy potential," *Phys. Rev. D* **71**, 123001 (2005), DOI: 10.1103/PhysRevD.71.123001.

A. B. Smith, "O3 Modified Gravity Tension Replication," Zenodo (2026), DOI: 10.5281/zenodo.18604204.

D. Stern *et al.*, "Cosmic chronometers: constraining the equation of state of dark energy. I: $H(z)$ measurements," *J. Cosmol. Astropart. Phys.* **02** (2010) 008, DOI: 10.1088/1475-

7516/2010/02/008.