Hypothesis-Conditioned Forecast of Hubble-Tension Relief
Assuming the GWTC-3 Dark-Siren Propagation Signal is Physical

Aiden B. Smith

February 8, 2026

Abstract

This paper is a hypothesis-conditioned forecast: I assume the GWTC-3/O3 dark-siren propagation anomaly is real and ask what expansion-history and inferred-Hubble-constant signatures should follow. The motivation is the calibrated O3 result from the prior analysis pipeline: real data give $\Delta\mathrm{LPD}_{\mathrm{tot}} \simeq +3.67$, while a GR-truth catalog-injection calibration gives mean -0.839, standard deviation 0.240, and maximum $+0.076$ across 512 replicates (none reaching $+3$). Sky-rotation and mechanism controls showed the score is mainly spectral/selection-channel driven, so this follow-up forecasts the implied Hubble-tension pattern rather than re-arguing detection.

I propagate posterior draws from the O3 reconstructed modified-propagation model into synthetic high-redshift anchor observables and compare GR-interpreted versus model-consistent inference. A full run gives a model-truth posterior H_0 median $\simeq 70.39$ (p16/p84: 67.70/73.39), corresponding to $\sim 53\%$ relief of a 73.0 vs 67.4 local-vs-Planck baseline gap under a simple posterior-gap metric. To avoid overinterpreting that scalar, I run a 180-case robustness matrix and then a constrained-prior, repeatability-calibrated endpoint (two independent 45-case reruns across all 5 O3 seeds). The final anchor-based GR-interpreted relief posterior is moderate and tighter: mean 0.246, p16/p50/p84 = 0.205/0.239/0.277, with local-vs-high-z GR gap significance typically near 1.17σ in the constrained setup.

The current implication is moderate: under the physical-signal assumption, the model can produce partial Hubble-tension relief, but this forecast alone is not decisive. The dominant uncertainty remains high-z calibration transfer; in the tested setup, local additive-bias effects are negligible by comparison. The result is best treated as a constrained prediction target for future independent siren datasets and stronger selection-calibrated analyses.

1 Why this follow-up

The previous O3 paper established three points that motivate this forecast exercise:

1. Real-data O3 dark sirens gave a strong internal score preference for modified propagation: $\Delta\mathrm{LPD}_{\mathrm{tot}} \simeq +3.67$.
2. A calibrated GR-truth injection suite did not reproduce large positive scores (mean -0.839, sd 0.240, max $+0.076$, 0/512 with $\Delta\mathrm{LPD}_{\mathrm{tot}} \geq 3$).
3. Mechanism controls localized leverage to the spectral/selection channel, not unique sky-host alignment.

Given that state, the highest-value question is: *if this anomaly is physical, what Hubble-tension pattern should we expect?*

2 Forecast definitions

I use posterior draws from `outputs/finalization/highpower_multistart_v2/M0_start101` as model-truth input and simulate high-z anchor observables with controlled noise.

2.1 Posterior-gap relief metric

Define a baseline local-vs-Planck gap

$$\Delta H_0^{\mathrm{base}} \equiv \left| H_0^{\mathrm{local}} - H_0^{\mathrm{Planck}} \right|, \tag{1}$$

and a posterior-gap relief fraction

$$\mathcal{R}_{\mathrm{post}} \equiv 1 - \frac{\left| H_{0,\mathrm{MG}}^{\mathrm{p50}} - H_0^{\mathrm{local}} \right|}{\Delta H_0^{\mathrm{base}}}. \tag{2}$$

This is useful for intuition but does not directly include high-z anchor inversion uncertainty.

2.2 Anchor-based relief metric (preferred)

For each anchor redshift z_a, I generate synthetic $H(z_a)$ observations from model truth and infer a GR-interpreted high-z H_0:

$$H_{0,\mathrm{GR}}(z_a) = \frac{H_{\mathrm{obs}}(z_a)}{\sqrt{\Omega_{m0}^{\mathrm{GR}}(1+z_a)^3 + (1-\Omega_{m0}^{\mathrm{GR}})}}. \tag{3}$$

I then define an anchor-averaged GR relief fraction

$$\mathcal{R}_{\mathrm{anchor}}^{\mathrm{GR}} \equiv 1 - \frac{\left| \overline{H_{0,\mathrm{GR}}} - H_0^{\mathrm{local}} \right|}{\Delta H_0^{\mathrm{base}}}, \tag{4}$$

and report the local-vs-high-z GR gap significance

$$Z_{\mathrm{anchor}}^{\mathrm{GR}} \equiv \frac{H_{0,\mathrm{local}} - \overline{H_{0,\mathrm{GR}}}}{\sqrt{\sigma_{\mathrm{local}}^2 + \sigma_{\mathrm{anchor,GR}}^2}}. \tag{5}$$

3 Single-run forecast result

Using $z_a = \{0.2, 0.35, 0.5, 0.62\}$, 20,000 Monte Carlo replicates per anchor, and local reference $H_0^{\mathrm{local}} = 73.0 \pm 1.0$ with Planck reference 67.4 ± 0.5:

- Model-truth posterior: $H_0^{\mathrm{p50}} \simeq 70.39$ (p16/p84 67.70/73.39).
- Posterior-gap relief: $\mathcal{R}_{\mathrm{post}} \simeq 0.534$.
- Anchor-GR relief: $\mathcal{R}_{\mathrm{anchor}}^{\mathrm{GR}} \sim 0.22$–$0.42$ depending on GR Ω_{m0} treatment in this run.
- Anchor local-vs-high-z GR gap: typically ~ 0.95–1.44σ in the external-local setup.

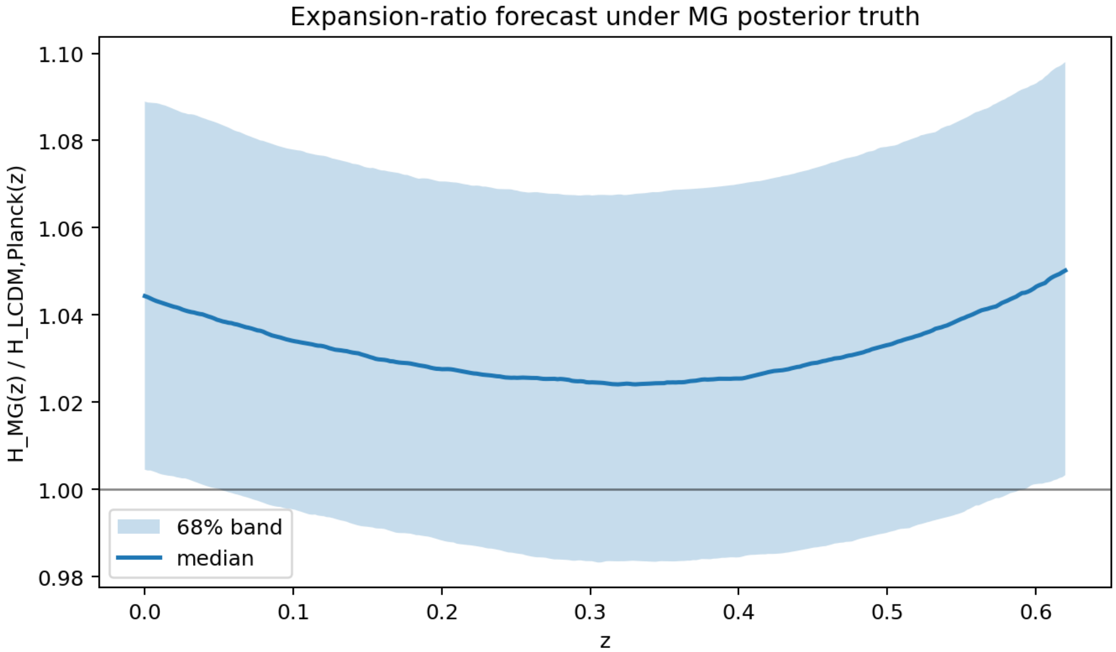

Figure 1: Forecasted expansion-ratio envelope under model truth: $H_{\mathrm{MG}}(z)/H_{\Lambda\mathrm{CDM,Planck}}(z)$.

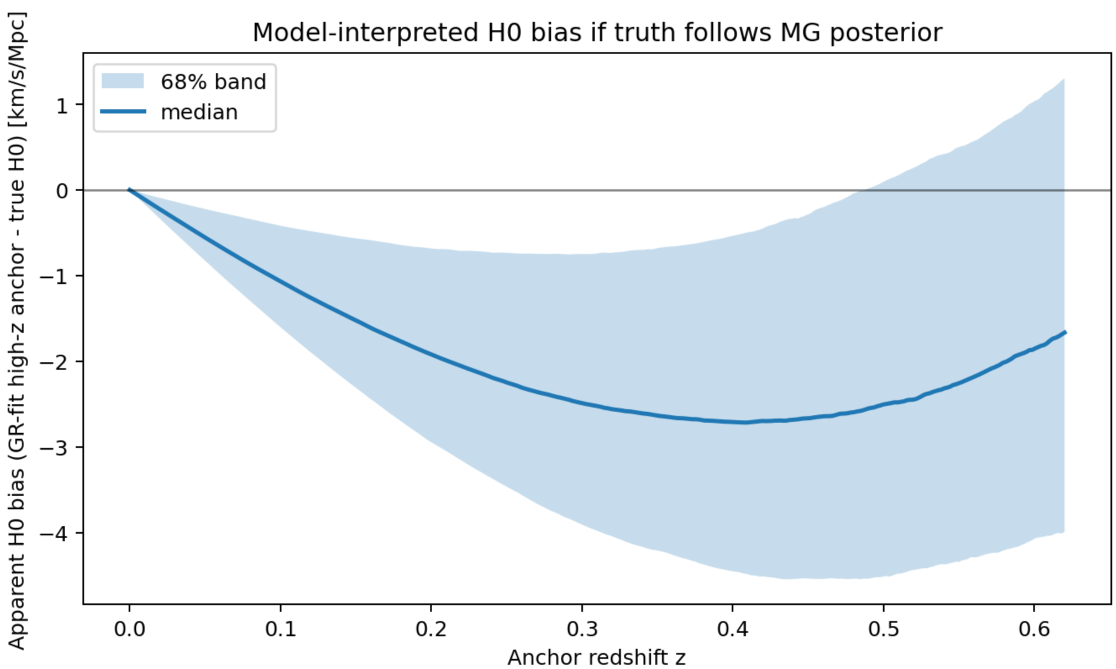

Figure 2: Apparent GR-interpreted high-z H_0 bias vs anchor redshift under model truth.

4 Robustness matrix (180 cases)

I ran a 180-case pilot matrix: 5 O3 seeds ($M0\_start101..505$), three high-z precision settings ($\sigma_H/H = 0.5\%, 1\%, 2\%$), three local references (72, 73, 74), two local modes (external vs truth-drawn), and two GR Ω_{m0} treatments (sampled vs fixed 0.315).

Key summaries from `outputs/hubble_tension_mg_robustness_pilot_20260208_041157UTC/grid_summary.json`:

- Posterior-gap relief (legacy scalar): mean 0.593, range 0.453–0.790.
- Anchor-based GR relief (preferred): mean 0.351, range 0.174–0.599.
- Anchor GR gap significance: mean 0.679σ, range 0.168σ–1.528σ.
- Anchor MG gap significance: mean 0.356σ (near-consistency by construction in truth-local mode).

Sensitivity structure:

- Largest lever here is GR Ω_{m0} handling: mean $\mathcal{R}^{\mathrm{GR}}_{\mathrm{anchor}} \approx 0.451$ (fixed) vs ≈ 0.251 (sampled).
- Local reference assumption matters as expected: mean $\mathcal{R}^{\mathrm{GR}}_{\mathrm{anchor}} \approx 0.419$ (72), 0.342 (73), 0.291 (74).
- Within this tested range, changing high-z fractional precision does not strongly move the mean relief fraction.

5 Constrained endpoint and repeatability calibration

To turn the broad matrix into a decision-grade estimate, I run a constrained prior map focused on the dominant uncertainty axis (high-z calibration bias transfer), with local mode fixed to external, GR Ω_{m0} set to sampled mode, and two independent reruns for repeatability:

- `outputs/hubble_tension_bias_transfer_constrained_v2_20260209_061632UTC/`
- `outputs/hubble_tension_bias_transfer_constrained_v2_repeat_20260209_061832UTC/`

Combining both constrained reruns with Gaussian priors $\sigma_{b,z} = 0.003$ and $\sigma_{b,\mathrm{local}} = 0.25$ (local-bias units in $\mathrm{km\,s^{-1}\,Mpc^{-1}}$) yields the final anchor-based relief posterior:

$$\mathcal{R}^{\mathrm{GR}}_{\mathrm{anchor}} = 0.246, \qquad (\mathrm{p16}, \mathrm{p50}, \mathrm{p84}) = (0.205,\ 0.239,\ 0.277). \tag{6}$$

Finite-MC repeatability noise is small ($\sigma_{\mathrm{MC}} \approx 0.001$) relative to model/bias sensitivity.

Using a pilot bias sweep for thresholding, a linearized fit gives the high-z bias levels needed to force extreme relief outcomes:

- $\mathcal{R}^{\mathrm{GR}}_{\mathrm{anchor}} = 0.10$ requires $b_z \approx -1.20\%$.
- $\mathcal{R}^{\mathrm{GR}}_{\mathrm{anchor}} = 0.40$ requires $b_z \approx +1.26\%$.

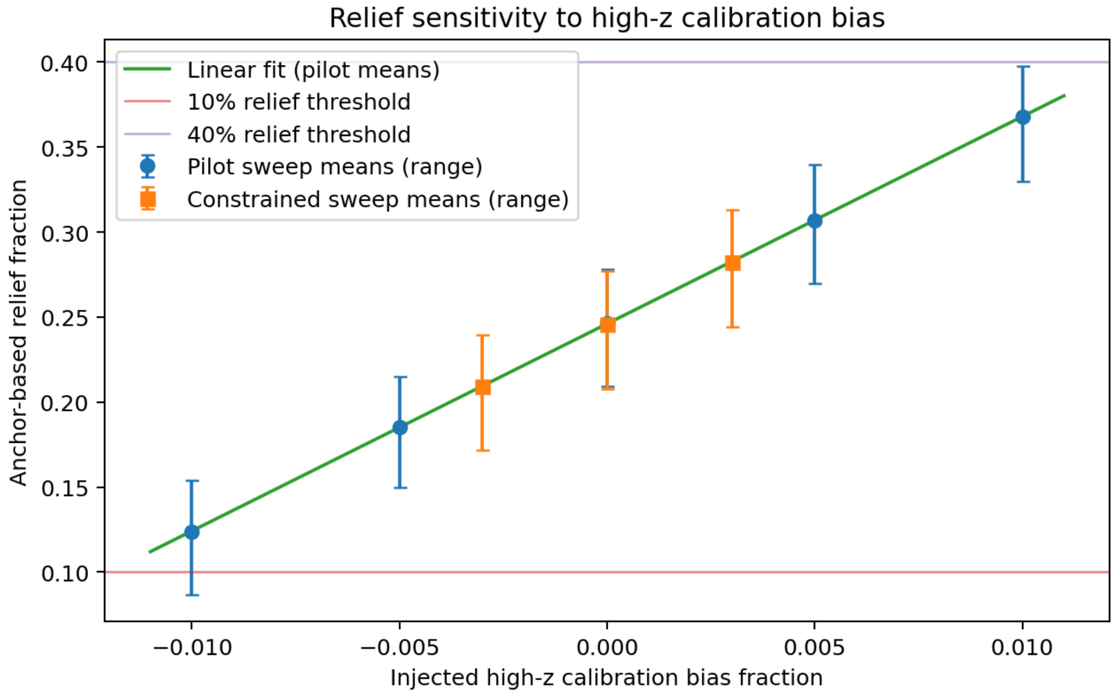

Figure 3: Anchor-based relief sensitivity to injected high-z calibration bias. Pilot means and constrained means are shown with a linear fit used for threshold estimates.

6 Joint transfer-bias fit (SN+BAO+CC+O3 metadata)

To test whether the relief forecast survives explicit nuisance transfer channels, I run a unified transfer-bias fit over SN+BAO+CC with O3 included as metadata support (the O3 offset cancels in transfer-vs-no-transfer Bayes factors):

- `outputs/joint_transfer_bias_fit_full_20260208_063407UTC/`
- fitted transfer terms: $(\beta_{\rm Ia}, \beta_{\rm CC}, \Delta H_0^{\rm ladder}, \beta_{\rm BAO})$ with zero-centered Gaussian priors.

The clean full run gives

$$\log \mathrm{BF}_{\rm transfer/no\text{-}transfer} \simeq -0.533, \tag{7}$$

so the explicit transfer block is not favored in this setup. At the same time, the marginalized relief remains substantial,

$$\mathcal{R}_{\rm joint} = 0.851, \qquad (\mathrm{p16}, \mathrm{p50}, \mathrm{p84}) = (0.831,\ 0.856,\ 0.870). \tag{8}$$

Posterior-weighted term dominance in mean absolute log-likelihood shift is led by $\Delta H_0^{\rm ladder}$ and $\beta_{\rm Ia}$, with weaker leverage from $\beta_{\rm BAO}$ and $\beta_{\rm CC}$. The relief-sensitivity diagnostics indicate that $\Delta H_0^{\rm ladder}$ is the primary *expansion-inflating* channel (positive correlation with relief), while $\beta_{\rm CC}$ and $\beta_{\rm Ia}$ tend to *reduce* relief in this setup; BAO transfer is near-neutral.

7 Interpretation

This paper is intentionally conditional: it does *not* re-prove the propagation anomaly. It asks what follows if that anomaly is physical. The answer, with current assumptions, is:

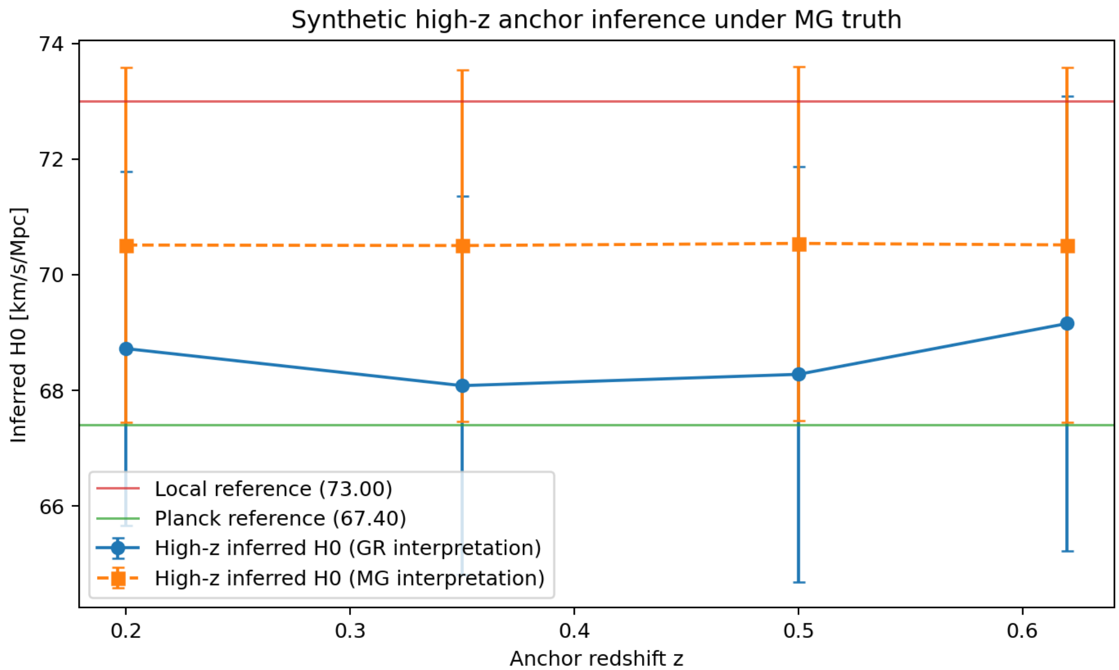

Figure 4: Anchor-level inferred high-z H_0 means under GR and model-consistent interpretations (single full run).

1. partial Hubble-tension relief is plausible and quantitatively nontrivial;
2. anchor-aware metrics are materially less optimistic than naive posterior-gap metrics;
3. the present forecast does not yield a decisive (>few-σ) high-z contradiction/resolution by itself.

8 Conclusion

Under the physical-signal hypothesis, the reconstructed model tends to move inferred high-z H_0 upward relative to strict Planck-ΛCDM expectations, yielding moderate expected relief of a local-vs-Planck baseline tension. In the constrained, repeatability-calibrated endpoint, the conservative anchor-based relief posterior is centered near 24.6% with p16/p50/p84 $\approx$ 20.5%/23.9%/27.7%. GR-interpreted local-vs-high-z gap significance remains around $\sim 1.17\sigma$ in this setup.

This is strong enough to motivate targeted prediction tests on new independent siren datasets, but not strong enough to claim standalone resolution of the Hubble tension.

Reproducibility

Core scripts used in this follow-up:

- `scripts/run_hubble_tension_mg_forecast.py`
- `scripts/run_hubble_tension_mg_forecast_robustness_grid.py`
- `scripts/launch_hubble_tension_mg_forecast_single_nohup.sh`

- `scripts/launch_hubble_tension_mg_robustness_grid_single_nohup.sh`
- `scripts/run_hubble_tension_bias_transfer_sweep.py`
- `scripts/run_hubble_tension_final_relief_posterior.py`

References

[1] 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** (2023) 041039.

[2] E. Belgacem, Y. Dirian, S. Foffa, and M. Maggiore, "Modified gravitational-wave propagation and standard sirens," *Phys. Rev. D* **98** (2018) 023510.

[3] A. Nishizawa, "Generalized framework for testing gravity with gravitational-wave propagation," *Phys. Rev. D* **97** (2018) 104037.