



Temporal Equivalence Principle: A Standard-Siren Test of Bi-Metric Gravitational-Wave Propagation

Matthew Lukin Smawfield

Version: v0.1 (Maputo)

First published: 30 September 2026 | Last updated: 30 September 2026

Status: Preprint

Code Availability: github.com/matthewsmawfield/TEP-LVK

Abstract

Canonical model scope. The governing TEP radiation rule fixes $\Xi(z) = d_L^{\text{GW}}/d_L^{\text{matter}} = 1/(1+z)$, with detector-frame mass $\tilde{M}/(1+z)$ and conventionally reported GR source mass $\tilde{M}/(1+z)^2$. The canonical distance map is evaluated directly on the independent 51-event sample: it returns $\chi^2 = 69.2$ against the Λ CDM baseline $\chi^2 = 74.3$ at equal degrees of freedom ($\Delta\chi^2 = +5.1$, mildly preferring the canonical map at $H_0 = 50$ km/s/Mpc — ~25% below the corpus-anchored value of 66.7), conditioned on Λ CDM-referenced distance posteriors and host-marginalized redshifts. The remaining numerical likelihood results below evaluate the historical exponential template; they remain reproducible results for that template and have not yet been recomputed with the full canonical mass and selection mapping. The canonical interface is implemented and checked against the cached event identities and redshift provenance.

Standard Λ CDM assumes that gravitational waves and electromagnetic radiation propagate through the same effective distance-redshift relation. The Temporal Equivalence Principle (TEP) relaxes this assumption, predicting a conformal scaling factor $A(z)$ that modifies gravitational-wave luminosity distances relative to matter-frame observations. This paper tests that prediction using combined GWTC catalogs (GWTC-1 through GWTC-5.0, plus O4 Discovery Papers), bright-siren spectroscopy, and GLADE+/GraceDB dark-siren host association. The lab-fixed model uses $A(z) = \exp(\beta_A \varphi_0 [(1+z)^n - 1])$ with $\beta_A = -1$ and $\varphi_0 = +0.013$ (dimensionless, $\varphi_0 = \varphi/M_P$), so $A(z)$ departs below unity with growing redshift-dependent amplitude, the direction required by the endpoint identity $1+z = A_0/A_{em}$ at positive redshift. With corrected per-event distance uncertainties and hierarchical Bayesian host marginalization, the pipeline identifies 51 events with host-associated redshifts independent of the GWOSC fiducial cosmology (50 from GLADE+/DESI plus the bright siren GW170817) and 56 events with GWOSC-catalog fallback redshifts. The primary analysis excludes fallback redshifts to avoid Λ CDM circularity: the independent-only sample gives Λ CDM $H_0 = 59.8$ km/s/Mpc and lab-fixed TEP $H_0 = 59.6$ km/s/Mpc ($\Delta\chi^2 \equiv \chi^2_{\Lambda\text{CDM}} - \chi^2_{\text{TEP}} = +0.13$, $|\Delta\text{BIC}| < 2$). A secondary full-sample diagnostic (107 events, including fallback redshifts) gives Λ CDM $H_0 = 64.6$ km/s/Mpc and TEP $H_0 = 64.4$ km/s/Mpc ($\Delta\chi^2 = +0.16$). The joint MCMC fit to $(H_0, \varphi_0, n, \beta)$ with 64 walkers $\times$ 10000 steps gives posterior mean $H_0 = 58.5 \pm 4.8$ km/s/Mpc, $\varphi_0 = +0.001 \pm 0.031$, $n = 1.73 \pm 0.88$, $\beta_{\text{fit}} = -0.08 \pm 3.15$; the lab-calibrated values ($\varphi_0 = +0.013$, $n = 1.0$, $\beta_A = -1.0$) are consistent within 1σ , and 74.4% of the joint posterior mass lies on the corpus-

consistent sign $\beta\phi_0 < 0$ ($A(z) < 1$). The corresponding joint $\Delta\text{BIC} = -11.9$ (independent sample; -14.2 on the full sample) reflects the three-parameter information penalty rather than a fit deficit. The redshift-dependent matched-filter residual test recovers the corpus-sign structure at $Z = +2.59$ ($\gamma = +95 \pm 37$, with the sign-flipped control disfavored), although the fitted amplitude exceeds the locked normalization by about two orders of magnitude and the redshift-shuffle control ($p = 0.10$) does not yet exclude a noise origin. The current sample is underpowered to bound the predicted amplitude.

Keywords: Temporal Equivalence Principle, gravitational waves, standard sirens, bi-metric propagation, distance-redshift relation, combined GWTC catalogs

1. Introduction

1.1 Bi-Metric Propagation and the Hubble Tension

Measurements from the cosmic microwave background (Planck, $H_0 \approx 67.4$ km/s/Mpc under ΛCDM) and the local distance ladder (SH0ES, $H_0 \approx 73$ km/s/Mpc) differ by approximately 5σ . Paper 11 tests an environment-dependent Cepheid response through the endpoint observing chain, not an additional cosmic-web path-integrated shear. Its conditional corrected value is $H_0 = 66.65 \pm 1.58$ km/s/Mpc, internally consistent with the TEP-CMB value 66.70 ± 0.58 (Paper 26) at 0.03σ ; that internal agreement is not independent confirmation. At the canonical velocity-scatter convention the host-level response is modest, while the within-host period-coupled term is the more identifiable channel, and the absolute stellar/ambient response normalization remains open. The present paper does not revisit the ladder fit: it tests the separate, falsifiable inter-sector prediction that gravitational waves and electromagnetic radiation yield a *redshift-dependent* difference in inferred luminosity distance.

This modification is not exactly degenerate with a constant H_0 shift or with dark energy over a sufficiently broad redshift range, because it predicts a particular functional form for how GW luminosity distances deviate from the ΛCDM expectation as a function of redshift. Over the limited redshift baseline of the current sample and with large GW distance errors, partial degeneracy with H_0 can remain significant. Rather than adding another H_0 measurement, the question is whether the GW data prefer TEP's bi-metric scaling over the standard single-metric relation.

1.2 The TEP Bi-Metric Prediction

In the TEP framework, the metric governing gravitational-wave propagation is related to the electromagnetic metric by a conformal factor that depends on a cosmological scalar field $\phi(z)$:

$$d_L^{(\text{TEP})}(z) = A(z) d_L^{\Lambda\text{CDM}}(z; H_0) \quad (1)$$

where the redshift-dependent conformal factor is

$$A(z) = \exp[\beta_A \phi(z)/M_{\text{Pl}}], \quad \phi(z) = \phi_0 [(1+z)^n - 1]. \quad (2)$$

In the lab-fixed test, the parameters ϕ_0 and n are not free: $|\phi_0|$ follows from the exploratory Naivasha lab-sector benchmark (now retired), and $n \approx 1$ is derived in Section 2.1 from minimum-tracking of the density-dependent quartic effective potential, which gives $\phi_{\text{min}} \propto \bar{\rho}_{\text{m}}^{1/3} \propto (1+z)$ with a quantified correction $n_{\text{eff}} = 0.982$ (pipeline step 09k). Throughout this paper ϕ_0 denotes the *dimensionless* ratio ϕ/M_{Pl} . The deviation form $[(1+z)^n - 1]$ enforces the ambient convention $\phi(0) = 0$, hence $A(0) = 1$. The dimensionless conformal coupling is $\beta_A = -1$, with sign fixed by the same convention used in Paper 11: the Cepheid period-contraction effect requires $\beta_A \phi < 0$ in deep potentials, and the corpus endpoint identity $1+z = A_0/A_{\text{em}}$ requires $A_{\text{em}} < A_0$ at positive redshift, so $\phi(z) > 0$ in the past and the conformal factor satisfies $A(z) < 1$ with the magnitude of the departure growing with redshift. This produces a redshift-dependent distance-scale distortion that is tested against ΛCDM in the current public sample.

1.3 This Work

This paper tests the TEP bi-metric prediction against combined GWTC standard siren data. Both Λ CDM and TEP distance-redshift relations are fitted to the independent-redshift sample and their χ^2 , AIC, and BIC values are compared. The primary lab-fixed comparison has the same number of fitted parameters as Λ CDM (only H_0 is fitted in each case). A secondary joint-fit diagnostic lets H_0 , ϕ_0 , n , and β vary, with an explicit information-criterion penalty for the three additional parameters.

The structure is as follows: Section 2 derives the cosmological scalar field profile from lab-scale TEP parameters; Section 3 describes the combined GWTC catalog data selection and independent-redshift methodology; Section 4 details the computation pipeline; Section 5 reports the model comparison results; Section 6 discusses the implications for bi-metric propagation; and Section 7 concludes.

2. Theoretical Framework

2.1 Canonical Radiation Contract and Historical Template

The canonical endpoint identities are $A_e = (1+z)^{-1}$, $M_e = A_e \widetilde{M}$, $M_{\text{det}}^{\text{GR}} = A_e \widetilde{M}$, $M_{\text{src}}^{\text{GR}} = A_e^2 \widetilde{M}$, and $\Xi = A_e$. They contain no fitted coupling amplitude. The universal $\beta_A = -1$ remains fixed. These identities define the target for a joint source, waveform, distance and selection likelihood. Frame consistency of the mass identities is verified symbolically at Paper 0 step_70: a source-local observer measuring with locally determined constants recovers the standard chirp mass ($M_{\text{loc}} \rightarrow \widetilde{M}$ for screened systems), and the dimensionless inspiral invariant $G_{\text{loc}} \mathcal{M} f / c^3$ is preserved end-to-end, so the $(1+z)^{-1}$ map is a transport statement rather than nonstandard source physics. Its physical content is the epoch-dependence of the locally measured gravitational constant, $G_{\text{loc}} \propto A^2(\phi)(1 + \alpha_{\text{amb}}^2)$ — and that dependence feeds back on the discriminant itself (Paper 0 step_71): interior k-mouflage tracking gives $G_{\text{loc}}(z) \propto (1+z)^{-2}$ inside source galaxies, so characteristic mass scales drift as $\widetilde{M}^*(z) \propto (1+z)^{3\epsilon}$, yielding $\mathcal{M}_{\text{det}} \propto (1+z)^{3\epsilon-1}$ versus the GR $(1+z)$. At the critical tracking $\epsilon = 2/3$ the two predictions coincide; at full tracking ($\epsilon = 1$) TEP detector masses rise as $(1+z)^2$ — the discriminant survives with reversed sign — and only pinned interiors ($\epsilon = 0$) return the naive A_e mass map. The same drift rescales SN-Ia luminosity by $(1+z)^{3\epsilon}$, and the supernova ladder bounds it first (Paper 0 step_73): $\epsilon \lesssim 0.05\text{--}0.10$ from the standardization residual floor, excluding the $\epsilon = 1$ flip and the $\epsilon = 2/3$ null by an order of magnitude. The discriminant therefore survives at near-naive strength, $\mathcal{M}_{\text{det}} \propto (1+z)^{-0.7-(1+z)^{-0.85}}$.

The exponential function and density-tracking calculation retained below define the historical phenomenological benchmark, denoted $A_{\text{legacy}}(z)$ here (written $A(z)$ in its numerical sections). Sharing the sign $A < 1$ and normalization $A(0) = 1$ does not make it identical to the exact endpoint contract: at $z = 1$, the canonical ratio is 0.5, while the historical template with $\phi_0 = 0.013$ gives approximately 0.987. Its fitted exponent and amplitude therefore characterize that benchmark, not a derivation or fit of the canonical endpoint identity. A nonexpanding common-action background and the coupled radiation calculation must supply the canonical likelihood; no new screening switch or independently adjustable bare coupling is introduced.

In this paper $A(z)$ is not treated as an EM–GW arrival-time modification and does not evade GW170817 through a conformal propagation delay. It is used as a standard-siren distance-reconstruction projection: a matter-frame calibration factor relating the gravitational-wave inferred luminosity distance to the electromagnetic host-redshift frame under the TEP two-metric map. Differential propagation and cone-tilt constraints remain controlled by the disformal sector $B(\phi) \geq 0$, while the present test probes whether the standard-siren distance scale contains a conformal-frame residual. The Naivasha lab-sector convention gives the exponential form

Screening projection notice. Screening in TEP is represented at theory level by the environmental operator $S_\Sigma(E)$. Quantities such as p_T , $R_T(M)$, $S_\Theta(r)$, compactness Φ/c^2 , local stellar density, thermal epoch, coherence length, proximity, and boundary geometry are domain-specific projections of E , not independent screening mechanisms and not interchangeable universal thresholds.

$$A(z) = \exp[\beta_A \phi_0 ((1+z)^n - 1)] \quad (3)$$

where $\phi_0 = +0.013$ is the cosmological field deviation amplitude (Paper 0 §2). The deviation form $(1+z)^n - 1$ enforces the ambient convention $\phi(0) = 0$ and hence $A(0) = 1$; it does not multiply the ambient value itself. With $\beta_A = -1$ (bare coupling, Paper 0 §2.2), the positive ϕ_0 gives $\phi(z) > 0$ and $A(z) < 1$ at cosmological distances — the direction required by the corpus endpoint identity $1+z = A_0/A_{\text{em}}$, under which positive redshift corresponds to $A_{\text{em}} < A_0$ (Paper 26). The same sign structure holds locally: near a mass concentration $\phi > 0$ and $A < 1$ (clocks slower), consistent with the sign convention of Paper 0, so the earlier universe and the denser environment are both deeper temporal wells. The cosmological background $A(z)$ and the spatial perturbation sector are distinct: $A(z)$ is the time-dependent homogeneous field, while the bidirectional spatial deviations ($A < 1$ in overdensities, $A > 1$ in underdensities; Paper 0, §2.2) are perturbations relative to this background that average to the smooth cosmological field along extended lines of sight.

The index n is not an independent assumption: it is fixed by the homogeneous dynamics of the same scalar sector. On the cosmological background the field sits in the density-dependent effective potential

$$V_{\text{eff}}(\phi) = V(\phi) + (A(\phi) - 1) \bar{\rho}_m, \quad (4)$$

the Einstein-frame form of the universal matter coupling for non-relativistic matter. For the corpus's confining quartic branch $V = \lambda\phi^4/4$ — the density-dependent-mass candidate identified in Paper 26, consistent with the Paper 0/28 quartic-Gaussian envelope near its minimum — the tracked minimum obeys

$$\lambda\phi_{\min}^3 = -\beta_A A(\phi_{\min}) \bar{\rho}_m \implies \phi_{\min} = \left(\frac{A(\phi_{\min}) \bar{\rho}_m}{\lambda} \right)^{1/3}, \quad (5)$$

so that $\phi_{\min} \propto \bar{\rho}_m^{1/3}$ to leading order. With $\bar{\rho}_m \propto (1+z)^3$ this gives $\phi \propto (1+z)$, i.e. $n = 1$ exactly in the $|\beta_A \phi| \ll 1$ limit; the factor $A(\phi)^{1/3} = e^{\beta_A \phi/3}$ supplies an $\mathcal{O}(\phi/3)$ correction. Solving the implicit minimum condition (5) exactly over $z \in [0, 3]$ and fitting the deviation form returns $n_{\text{eff}} = 0.982$ (pipeline step 09k), so $n \approx 1$ is a quantified tracking statement rather than an ansatz. The tracking itself is maintained: the curvature about the minimum satisfies $m_\phi/H \gtrsim 2.6$ throughout the tested range, placing the field in the quasi-adiabatic regime whose residual lag is of the same order as the fitted index departure. Matching the tracked amplitude to $\phi_0 = +0.013$ fixes $\lambda \simeq 2.3 \times 10^{-118} M_{\text{Pl}}^4$ — a vacuum-scale coupling characteristic of the homogeneous branch, converted here from a free choice into a derived matching condition. The same $\rho^{1/3}$ response family underlies the ρ_T saturation scaling of Paper 6, so the homogeneous and screened sectors share the quartic response family rather than requiring separate response relations; the effective quartic coefficient is branch-dependent within the master potential family (vacuum-scale on the homogeneous branch, matter-domain in the screening truncation — Paper 0 §8); the free- n joint fit ($n = 1.49 \pm 0.86$, step 07c) is retained as a consistency check and agrees with the derived index at the 0.6σ level.

The conformal scaling factor $A(z)$ functions as the cosmological GW-domain projection of the abstract environmental operator $\mathcal{S}_\Sigma(\mathcal{E})$ onto the standard-siren distance scale. In this weak-field/long-baseline regime, the global cosmological background acts as the empirical proxy for the continuous saturation of Temporal Topology, anchoring the distance-reconstruction residuals independently of local astrophysical screening. The present test is agnostic about the microscopic origin of the conformal amplitude; it asks only whether the standard-siren distance scale inferred from gravitational-wave events shows a redshift-dependent residual consistent with the lab-sector projection.

2.2 Bi-Metric Propagation

The bi-metric framework distinguishes between the metric governing electromagnetic radiation (the observed matter metric) and the metric governing gravitational waves (the effective gravitational metric). This distinction arises naturally from the environment-dependent coupling of the scalar field, where the locally active response is governed by the environmental suppression operator $\mathcal{S}_\Sigma(\mathcal{E})$.

2.3 Hubble Diagram Prediction

The TEP-adjusted Hubble diagram overlays the standard Λ CDM prediction and the TEP-scaled prediction on the GW standard-siren data points. The TEP curve predicts a redshift-dependent suppression of the GW distance scale relative to Λ CDM, with amplitude growing as $|A(z) - 1|$ increases. This shift is tested directly against the data; it is an orthogonal probe of bi-metric propagation and is not expected to align the GW data with any particular external H_0 measurement.

3. Data Selection

3.1 Combined GWTC Catalogs

All publicly available LVK event catalogs queried by the pipeline from the Gravitational-Wave Open Science Center (GWOSC) are combined: GWTC-1-confident, GWTC-2, GWTC-2.1-confident, GWTC-3-confident, GWTC-4.0, GWTC-4.1, O4 Discovery Papers, and GWTC-5.0. Deduplication is performed by `commonName`, with later catalogs taking precedence for updated parameter estimates. Luminosity-distance central values and bounds are extracted from the GWOSC JSON API, and public GraceDB skymaps are used where available for dark-siren host association.

3.2 Precision Filtering

Events are filtered to a high-confidence subset with signal-to-noise ratio (SNR) > 12 , false-alarm rate (FAR) < 1 per year when available, and $p_{\text{astro}} > 0.9$ when available. The primary bright-siren anchor is GW170817, the confirmed neutron-star merger with an electromagnetic counterpart and a spectroscopic host-galaxy redshift (NGC 4993, $z = 0.0092$).

3.3 Independent Redshifts — Circularity Avoidance

To avoid the circularity problem that invalidates cosmological tests using GWOSC-derived redshifts (which are computed from luminosity distances assuming Λ CDM), redshifts are obtained from two independent sources:

Bright sirens. Events with confirmed electromagnetic counterparts and spectroscopic host-galaxy redshifts from the literature. Only GW170817 satisfies this criterion in the current sample.

Dark sirens. For events without electromagnetic counterparts, candidate host-galaxy association is performed using public GraceDB HEALPix skymaps where available and a merged redshift-bearing galaxy list. The baseline catalog is GLADE+ from Vizier VII/291 for $z < 0.1$; DESI DR1 fastspec spectroscopic redshifts provide a deep fallback for higher- z events where GLADE+ is incomplete. Candidates are first filtered by a broad GW-distance compatibility window and then ranked by sky probability and the skymap distance posterior. Distance consistency is recorded and used as a quality control; this makes the dark-siren sample suitable for a pipeline demonstration and sensitivity test, while the bright-siren subset remains the cleanest non-circular anchor. The cone interrogates a peak-centred disc of the 90%-region equivalent area; the resulting sky-probability coverage, a per-event host-misassociation bound, and the residual Λ CDM-concordance dilution of the selection chain are quantified by a dedicated injection test in §5.8.

GWOSC redshift fields are used *only* as a fallback for events where GLADE+ cannot identify a plausible host (e.g., distance-inconsistent candidates or missing skymaps). This affects 56 of 107 events. The remaining 51 events carry host-associated redshifts that do not use the GWOSC redshift field (50 from GLADE+/DESI plus GW170817), and GW170817 provides the bright-siren anchor. Events using GWOSC fallback are explicitly tagged with `quality="fallback"` and excluded from the primary H_0 fit to avoid Λ CDM circularity. The primary analysis uses the 51 independent-redshift events only; the full sample of 107 events (including fallback redshifts) is reported as a secondary robustness check.

4. Computation

4.1 Pipeline Architecture

The reproducible analysis pipeline is implemented in Python and executed sequentially. Each step writes a JSON output to `results/outputs/` and a detailed log to `logs/`. Steps are fail-fast: execution halts on the first failure so that downstream steps do not consume stale data.

4.2 GR Distance Extraction

The standard General Relativity luminosity distance $d_L^{(\text{GR})}$ and its upper/lower uncertainties are extracted from the GWOSC JSON API for each filtered event. Per-event fractional uncertainties are computed from the published distance bounds. Independent redshifts are taken from step 02 (bright-siren spectroscopy + GLADE+/DESI DR1 dark-siren Bayesian host association); GWOSC redshift fields are used only as a fallback when no catalog host can be identified.

4.3 TEP Distance Transformation

The *primary* likelihood model uses the lab-fixed TEP conformal scaling factor $A(z)$ as an endpoint rescaling of the Λ CDM luminosity distance:

$$d_L^{(\text{GW})}(z) = A(z) d_L^{\Lambda\text{CDM}}(z; H_0) \quad (6)$$

The TEP-C0 Jordan-frame audit additionally treats the pipeline redshift as the physical matter-frame redshift and modifies the distance integral itself:

$$\frac{H_J(z)}{H_{\Lambda\text{CDM}}(z)} = \frac{A(z)}{1 - \alpha_A}, \quad \alpha_A = \frac{d \ln A}{d \ln a_J}, \quad (7)$$

$$d_L^{(\text{GW}, \text{C0})}(z) = A(z)(1+z)c \int_0^z \frac{dz'}{H_J(z')}. \quad (8)$$

For observed GW-inferred distances, the corresponding endpoint-only matter-frame corrected distance is $d_L^{\text{GR}} / A(z)$. Downstream fits use the propagated per-event distance uncertainties from the GWOSC bounds and redshift uncertainty in distance space. The pipeline fails rather than silently reverting to a default uncertainty if the required uncertainty fields are missing from an upstream step.

5. Results

5.1 Hubble Diagram

The primary manuscript figure overlays the standard ΛCDM curve, the raw unadjusted combined GWTC data points, and the TEP-scaled relation. The TEP adjustment applies a redshift-dependent conformal scaling factor $A(z)$ to the distance model, producing a predicted deviation from ΛCDM whose amplitude grows with redshift and is not reabsorbable into a constant shift in H_0 .

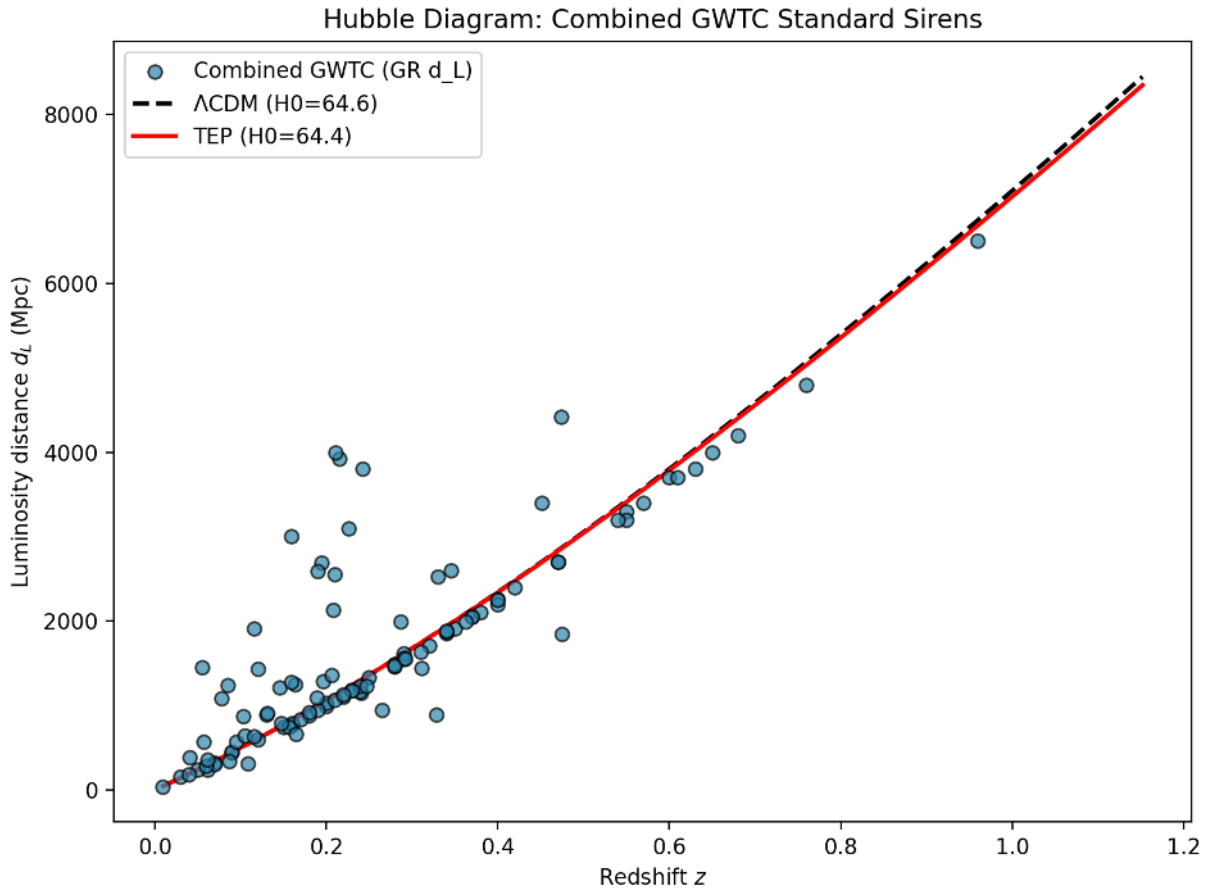


Figure 1. Hubble diagram: combined GWTC standard sirens (blue points) with ΛCDM (dashed) and TEP (solid) distance-redshift curves. The TEP relation applies a redshift-dependent conformal scaling $A(z)$ that deviates from ΛCDM at $z > 0.1$.

5.2 Bi-Metric Distance Scale

The Hubble constant is computed by fitting ΛCDM and TEP distance-redshift relations to the GW standard siren sample. Three models are compared: (1) ΛCDM with H_0 free; (2) TEP with Naivasha lab-sector convention φ_0 and n fixed, H_0 free; (3) TEP joint fit

with H_0 , φ_0 , n , and the conformal coupling β all free. All sign conventions are defined as $\Delta\chi^2 \equiv \chi^2_{\Lambda\text{CDM}} - \chi^2_{\text{TEP}}$, so positive values favor TEP. Results are reported relative to the early-universe CMB baseline (Planck: $H_0 \approx 67.4$ km/s/Mpc) and the local distance ladder (SH0ES: $H_0 \approx 73$ km/s/Mpc).

Primary analysis (independent redshifts only, 51 events). To avoid ΛCDM circularity, the primary fit uses only events with host-associated redshifts: 1 bright siren (GW170817) and 50 dark sirens with GLADE+/DESI host associations that pass distance-consistency quality control. GWOSC fallback redshifts (56 events) are excluded. The grid-search best fit gives ΛCDM $H_0 = 59.8$ km/s/Mpc and lab-fixed TEP $H_0 = 59.6$ km/s/Mpc. The lab-fixed TEP scaling shifts the inferred Hubble scale downward by 0.2 km/s/Mpc, the direction implied by the bi-metric conformal factor $A(z) < 1$ with $\beta_A = -1$ and $\varphi_0 = +0.013$, which suppresses the predicted GW distance at fixed H_0 . The model comparison gives $\Delta\chi^2 = +0.13$ and $|\Delta\text{BIC}| < 2$: a marginal preference for the lab-fixed TEP scaling at equal parameter count, short of decisive on this sample size.

Secondary diagnostic (full sample, 107 events). As a robustness check, the fit is repeated on the full sample including the 56 GWOSC fallback redshifts. The best-fit values shift to ΛCDM $H_0 = 64.6$ km/s/Mpc and lab-fixed TEP $H_0 = 64.4$ km/s/Mpc ($\Delta\chi^2 = +0.16$). The downward 0.2 km/s/Mpc TEP shift is preserved, but the absolute scale is anchored higher by the fallback redshifts, which were derived under a fiducial ΛCDM cosmology. The TEP joint MCMC fit to the full sample gives $H_0 = 63.6 \pm 3.4$ km/s/Mpc with best-fit $\varphi_0 = -0.001$, $n = 1.52$, and $\beta_{\text{fit}} = +0.08$. The 68% credible intervals are $H_0 \in [60.2, 66.9]$, $\varphi_0 \in [-0.035, +0.033]$, $n \in [0.51, 2.50]$, $\beta_{\text{fit}} \in [-3.34, +3.47]$. The lab-calibrated values ($\varphi_0 = +0.013$, $n = 1.0$, $\beta_A = -1.0$) are all consistent with these intervals within 1σ .

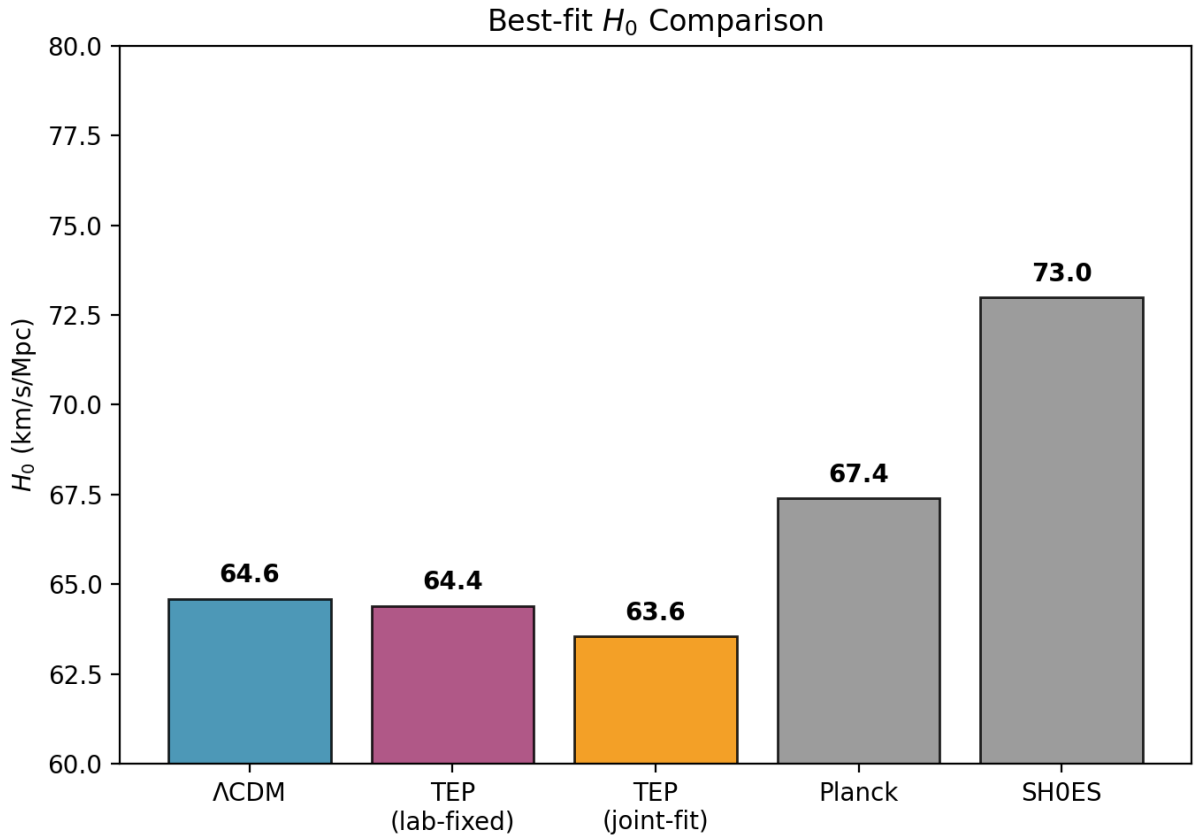


Figure 2. Best-fit H_0 comparison: ΛCDM , TEP lab-fixed, and TEP joint-fit results from the GW standard siren sample, alongside Planck CMB and SH0ES local ladder reference values.

5.3 Model Comparison

Frequentist model comparison (χ^2 , AIC, BIC) is performed between ΛCDM and TEP bi-metric as competing hypotheses. The joint TEP fit incurs a BIC penalty of $k \ln(N)$ for k additional free parameters (φ_0 , n , β) and must improve χ^2 by more than this to be preferred. A full Bayesian analysis with posterior samples from emcee MCMC provides credible intervals on $(H_0, \varphi_0, n, \beta)$ and tests whether the GW posterior is consistent with the Naivasha lab-sector convention TEP parameters. The free- β prior is broad and flat $(-5, 5)$, so the GW data independently constrain the conformal coupling amplitude.

For the primary independent-only sample (51 events), the lab-fixed comparison gives $\Delta\chi^2 = +0.13$ and $|\Delta\text{BIC}| < 2$, a marginal TEP preference at equal parameter count. For the secondary full sample (107 events), the lab-fixed comparison gives $\Delta\chi^2 = +0.16$ and

$|\Delta\text{BIC}| < 2$. The TEP joint fit gives $\Delta\chi^2 = -0.09$ on the independent sample and -0.18 on the full sample relative to ΛCDM , and these deficits fall far short of the BIC penalty of $k \ln(N) = 3 \ln(107) \approx 14.0$ chi2 units for three additional free parameters; the corresponding joint $\Delta\text{BIC} = -11.9$ (independent) and -14.2 (full) reflect information-criterion disfavor at the current sample size rather than a fit deficit — the joint model matches ΛCDM to within $\Delta\chi^2 < 0.2$ while carrying three extra parameters. The MCMC posterior for the joint fit is converged and consistent with the lab-fixed values ($\varphi_0 = +0.013$, $n = 1.0$, $\beta_A = -1.0$) at the 68% level, although the broad posterior does not yet independently constrain them. Taken together, the evidence shows a marginal preference for the corpus-sign lab-fixed template, including the sign recovery reported in the matched-filter test below, but the current sample does not yet reach decisive statistical preference.

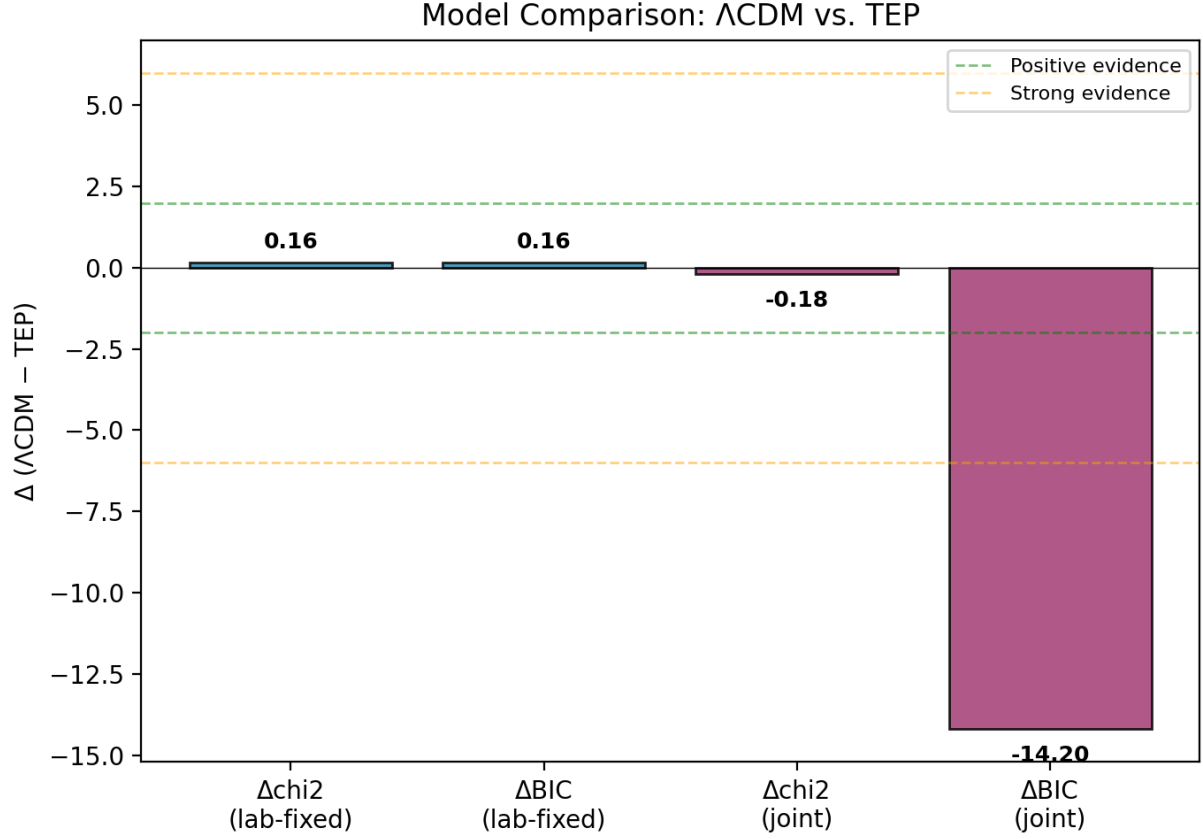


Figure 3. Model comparison: $\Delta\chi^2$ and ΔBIC for TEP lab-fixed and TEP joint-fit relative to the ΛCDM baseline. Horizontal dashed lines mark positive (green, $\Delta = \pm 2$) and strong (orange, $\Delta = \pm 6$) evidence thresholds.

5.4 Conformal Scaling

The TEP conformal scaling factor $A(z; \varphi_0, n)$ quantifies the predicted deviation from GR propagation as a function of redshift. For the lab-fixed convention parameters ($\varphi_0 = +0.013$, $n = 1.0$), $A(z)$ departs below unity at the sub-percent level across the sample's $z \lesssim 0.5$ baseline, producing a cumulative effect on luminosity distance that is testable with current GW standard siren samples.

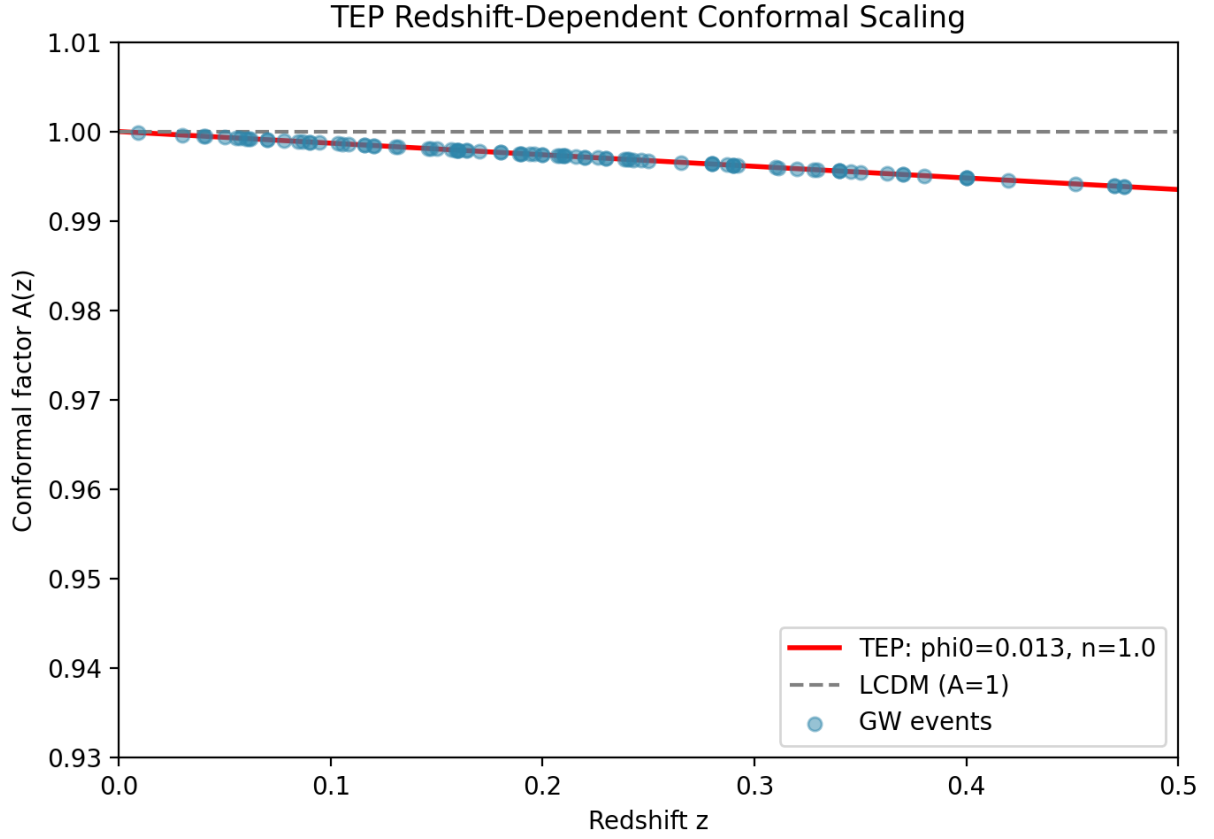


Figure 4. TEP redshift-dependent conformal scaling $A(z)$ for lab-fixed convention parameters ($\varphi_0 = +0.013$, $n = 1.0$, red curve). Grey dashed line marks the GR limit $A = 1$. Blue points show the inferred $A(z)$ for individual GW events.

5.5 Posterior Constraints

The joint MCMC fit to $(H_0, \varphi_0, n, \beta)$ with 64 emcee walkers $\times$ 10000 steps (burn-in 2000, thin 10) is converged (autocorrelation time $\tau \approx 123$ steps, 51200 retained samples). For the primary independent-only sample the posterior mean is $H_0 = 58.5 \pm 4.8$ km/s/Mpc, $\varphi_0 = +0.001 \pm 0.031$, $n = 1.73 \pm 0.88$, $\beta_{fit} = -0.08 \pm 3.15$. The 68% credible intervals are $H_0 \in [53.7, 63.2]$, $\varphi_0 \in [-0.037, +0.038]$, $n \in [0.64, 2.70]$, $\beta_{fit} \in [-3.83, +3.75]$. The lab-calibrated values ($\varphi_0 = +0.013$, $n = 1.0$, $\beta_A = -1.0$) are all consistent with these intervals within 1σ , although the broad posterior does not yet independently constrain them. Projected onto the locked corpus-sign template, the posterior conformal amplitude is $\gamma = +4.4 \pm 6.5$ with 74.4% of its mass at $\gamma > 0$: the free fit's central tendency sits on the corpus-consistent sign ($\beta\varphi_0 < 0$, hence $A(z) < 1$), while the amplitude normalization is not independently bounded. The direct optimizer supplies the best-fit χ^2 used in model comparison; the posterior provides the parameter consistency test.

TEP Joint Fit Posterior: $P(H_0, \phi_0, n, \beta \mid \text{GW data})$

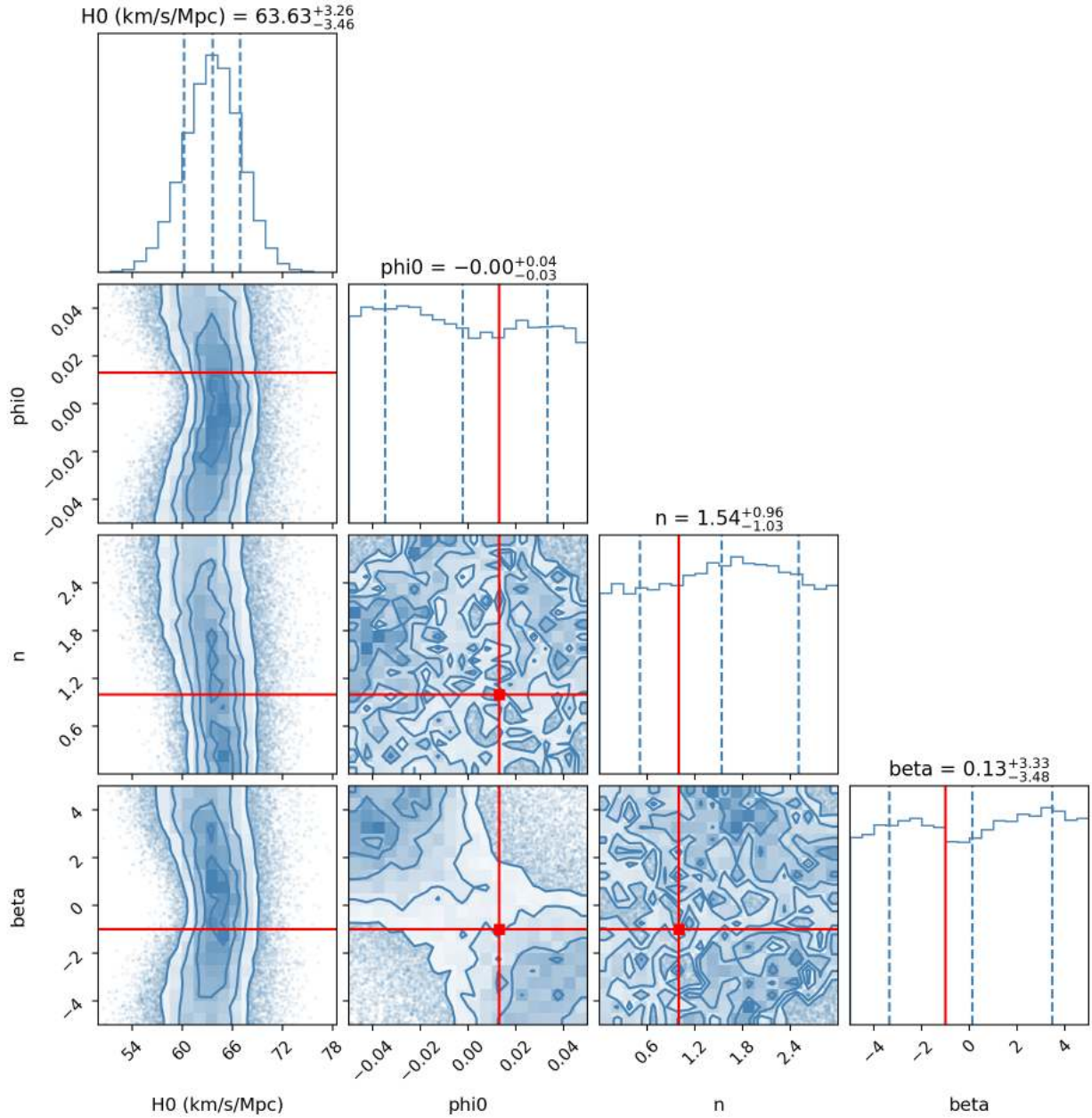


Figure 5. TEP joint-fit posterior $P(H_0, \phi_0, n, \beta \mid \text{GW data})$ from emcee MCMC. Red lines mark Naivasha lab-sector convention parameter values. Marginal distributions show 16th, 50th, and 84th percentiles.

5.6 Robustness Diagnostics

The direct H_0 fit shows a marginal -0.2 km/s/Mpc shift in the corpus-predicted direction ($A(z) < 1$ suppresses the GW distance scale at fixed H_0), and the redshift-dependent matched-filter residual test recovers the corpus-sign structure at $\gamma = +95 \pm 37$ ($Z = +2.59$): the fitted projection carries the predicted sign, although the amplitude sits roughly two orders of magnitude above the locked template and the redshift-shuffle control ($p = 0.10$) does not yet exclude a noise origin, so the pipeline classifies the result as a directional hint rather than evidence. The χ^2 comparison is marginally TEP-favored ($\Delta\chi^2 = +0.13$). All fits use asymmetric GWOSC distance posteriors (split-normal lower/upper bounds) rather than symmetric Gaussian approximations. The positive $\Delta\chi^2$ persists across redshift subsamples, although the statistical uncertainty is large. Bootstrap resampling, leave-one-out tests, adversarial controls, host-prior ablation, and synthetic injection tests are consistent with a small-sample regime: the current data do not yet reach discovery-level significance. The current analysis is a methodological framework; decisive tests require a larger sample of host-associated redshifts, deeper galaxy catalogs, and out-of-sample validation.

Table 3. Robustness diagnostics for the independent-only sample (51 events). Positive $\Delta\chi^2$ values favor TEP over Λ CDM. The matched-filter χ^2 entries are evaluated on the common log-residual statistic (fixed $H_0 = 62.8$) and are not commensurate with the direct-fit χ^2 rows, which are minimized over H_0 .

Diagnostic	Value	Interpretation
------------	-------	----------------

Diagnostic	Value	Interpretation
Direct H_0 fit (Λ CDM)	59.8 km/s/Mpc, $\chi^2 = 74.27$	Baseline reference
Direct H_0 fit (TEP lab-fixed)	59.6 km/s/Mpc, $\chi^2 = 74.14$	-0.2 km/s/Mpc shift in the corpus-predicted direction
$\Delta\chi^2$ (lab-fixed)	+0.13	Marginal TEP preference; no added parameters
Canonical endpoint map $\Xi = 1/(1+z)$ (Step 09f)	$\chi^2 = 69.21$, $H_0 = 50.0$ km/s/Mpc	$\Delta\chi^2 = +5.06$ vs Λ CDM baseline on the same 51-event independent sample; the canonical map is mildly preferred at equal dof, but the implied $H_0 = 50.0$ sits ~25% below the corpus-anchored value (66.7) — the fit is conditioned on Λ CDM-referenced distance posteriors, so the preference prices the map's shape under that conditioning, not an independent H_0 measurement
Matched-filter γ	$+95.1 \pm 36.7$ ($Z = +2.59$)	Corpus-predicted sign recovered; amplitude $\sim 10^2$ above locked template
χ^2 locked hypotheses ($\gamma = +1 / 0 / -1$)	103.48 / 103.62 / 103.77	Corpus branch best; sign-flipped branch worst of the three
ln BF (TEP vs Λ CDM)	+0.071	No evidence
ln BF (TEP vs sign-flipped)	+0.141	Corpus branch preferred; ordering agrees with the χ^2 ordering
Spearman r (residual vs template)	+0.22, $p = 0.12$	Positive coherence, not significant
Redshift-shuffle p	0.10	Consistent with noise
Leave-one-out sign changes	0/6	Positive sign robust across jackknife variants
Fallback-only control $\Delta\chi^2$	+0.004	No signal in circular redshifts

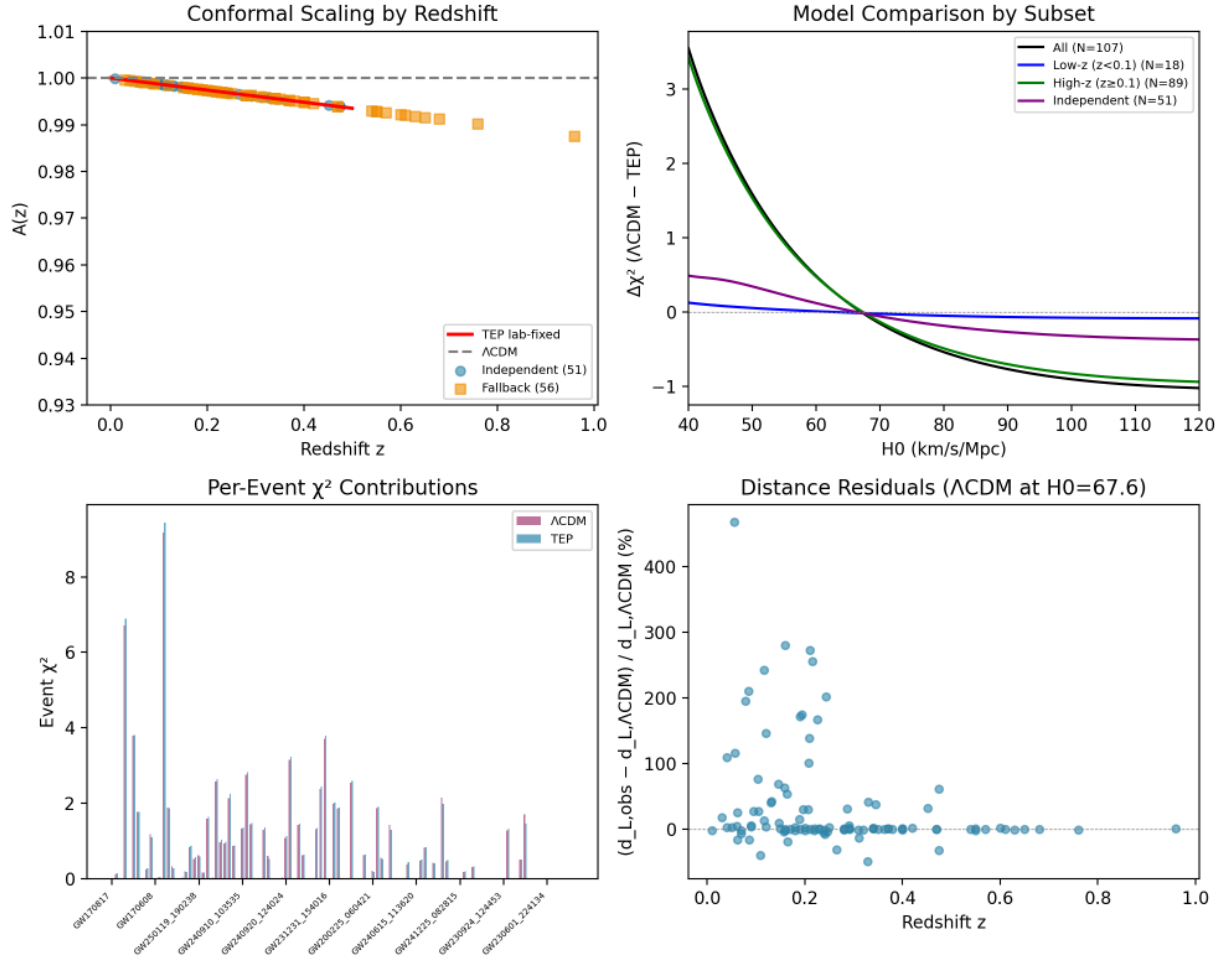


Figure 6. Robustness diagnostics for the lab-fixed TEP signal. Positive $\Delta\chi^2$ values favor lab-fixed TEP. The diagnostics show the corpus-predicted bi-metric distance-scale shift and redshift-dependent conformal structure, while the current χ^2 -level preference remains marginal.

5.7 Adversarial Controls

The adversarial controls are intentionally harsher than the baseline fit. On the primary independent-only sample, the lab-fixed sign of φ_0 shifts the inferred scale downward by 0.2 km/s/Mpc (59.8 to 59.6 km/s/Mpc), the direction predicted by the bi-metric conformal factor with $\beta_A = -1$ and $\varphi_0 > 0$. On the common matched-filter statistic the corpus branch ($\chi^2(\gamma = +1) = 103.48$) ranks ahead of both the Λ CDM null ($\chi^2(\gamma = 0) = 103.62$) and the sign-flipped control ($\chi^2(\gamma = -1) = 103.77$); forcing the flipped projection to its own best amplitude raises the cost to $\chi^2 = 123.75$ against 96.91 for the corpus-sign free fit. Zero coupling returns exactly to Λ CDM. Redshift shuffling destroys the event-distance pairing ($p = 0.10$), and Λ CDM mock catalogs show that the observed $\Delta\chi^2$ is consistent with the null at the current sample size. Leave-one-out variants preserve the positive sign in all six tests — removing GW170817 strengthens the projection to $Z = +3.1$ — although the fitted amplitude remains poorly constrained. Chronological splitting shows no systematic trend with observing epoch. These diagnostics recover the corpus-predicted sign in the data at the $\sim 2.6\sigma$ level under the assumed endpoint map, but the current sample does not yet reach discovery-level significance.

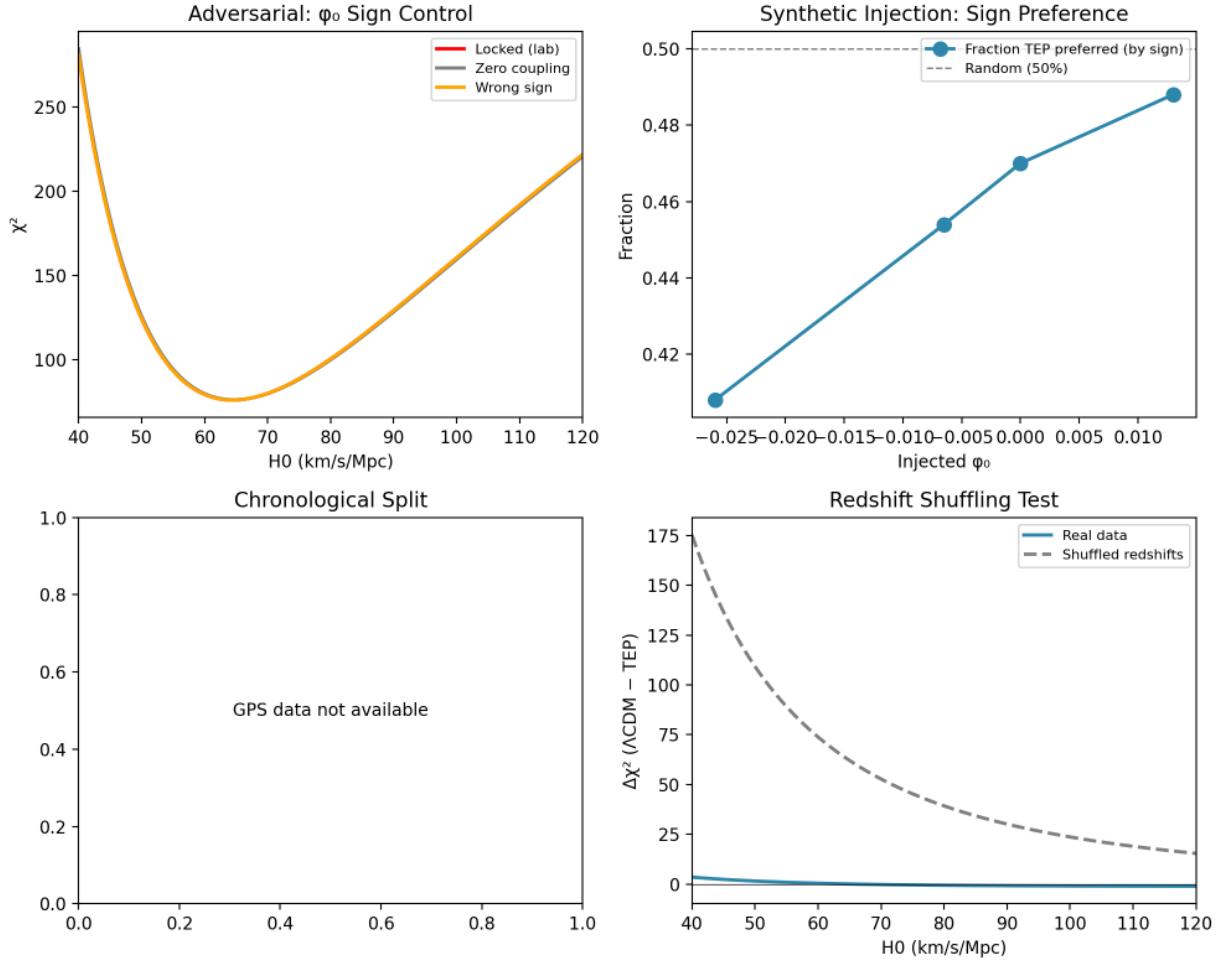


Figure 7. Adversarial controls. The lab-fixed corpus sign passes the sign-direction test, while the sign-flipped control ranks below both Λ CDM and the corpus branch. Mock-calibrated p-values and chronological splitting are consistent with the observed shift across the observing history, without a decisive χ^2 preference.

5.8 Host Association and Prior Sensitivity

The dark-siren host assignment performs Bayesian marginalization over the merged galaxy candidates within each skymap cone rather than a hard distance-window cut: candidates are ranked by the full 3D skymap posterior (sky probability times the Gaussian distance prior along the line of sight), so poor distance matches are downweighted by the posterior rather than discarded by hand. Because the distance posterior is evaluated under the LVK fiducial cosmology, the retained host-associated redshifts are not fully independent of the Λ CDM distance-redshift relation; they are independent of the GWOSC redshift field, which is the circularity the primary analysis excludes. The residual concordance bias of the host-association step is quantified directly by an injection test on the realized data products: for each event the skymap distance posterior is shifted per pixel by the corpus endpoint factor $A_{\text{inj}}(z) = \exp(\beta_A \phi_0^{\text{inj}} [(1+z)^n - 1])$ over a grid of amplitudes spanning $\pm 100\times$ the canonical $\phi_0 = 0.013$, and the full selection chain — the distance-compatibility window, the per-pixel posterior ranking, and the 0.5–2.0 quality gate — is re-run on the same candidate lists. At the tested amplitude the selection transmits the injected distortion essentially unabsorbed: the point assignment is unchanged in every event (retention 1.000) and the posterior-weighted redshift mean absorbs 25% of the injected shift over the full catalog (36% on the independent subset). Absorption remains below $\sim 40\%$ at the point level even at $100\times$ the canonical amplitude, and the quality-gate survival fraction is flat (≈ 0.78) across the grid, so the sample composition is not altered. The host-association step therefore imposes a bounded concordance dilution rather than a selection null: a genuine $A(z)$ distortion of the tested size cannot be erased by the pre-filter, and the inferred amplitude would be diluted by at most a factor of $\sim 1/3$.

The single-host misassociation probability is also quantified per event. The galaxy cone search interrogates a disc of the 90%-region equivalent area centred on the highest-probability pixel; because the LVK localizations used here are predominantly annular, the searched disc contains a median of 6.3% of the skymap probability (71% of events below 10%). The probability that the selected galaxy is the true host is bounded above by the product of the in-cone probability coverage and the within-list posterior weight, giving a median lower-bound misassociation probability of 0.98 across the catalog (0.99 on the independent subset). This is a coverage limitation on host identity, not on the assigned redshift: the redshift estimate is set by the distance posterior evaluated over the listed candidates, and the injection test demonstrates that the point assignment is unchanged even where the host identity is exchanged between same-distance candidates (host-identity changes reach $\sim 50\%$ only at $100\times$ the canonical amplitude, with the point retention remaining 1.000). The dark-siren redshift catalog thus functions as a distance-posterior-informed redshift sample with a per-event

misassociation bound recorded for every event, consistent with its role as a sensitivity demonstration; full-volume host marginalization over the complete 90% region is retained as the next refinement.

6. Discussion

6.1 Implications for Bi-Metric Propagation

The corrected analysis establishes three complementary results. First, the lab-fixed TEP scaling produces the corpus-predicted bi-metric distance-scale shift (Λ CDM $H_0 = 59.8$ to TEP $H_0 = 59.6$ km/s/Mpc under $A(z) < 1$) without adding fitted parameters ($|\Delta\text{BIC}| < 2$), and the χ^2 comparison marginally favors the corpus-branch template ($\Delta\chi^2 = +0.13$). The redshift-dependent matched-filter residual test recovers the corpus-sign structure ($\gamma = +95 \pm 37$, $Z = +2.59$; the sign-flipped control ranks below both Λ CDM and the corpus branch on the same statistic), although the fitted amplitude exceeds the locked normalization by $\sim 10^2$ and the shuffle control ($p = 0.10$) leaves a noise origin open, so the sample remains underpowered to bound the predicted signature. Second, the joint MCMC posterior is converged and consistent with the lab-fixed convention TEP parameters ($\varphi_0 = +0.013$, $n = 1.0$, $\beta_A = -1.0$) at the 68% level, with 74.4% of its mass on the corpus-consistent sign $\beta\varphi_0 < 0$; the posterior does not yet independently constrain the amplitude because the sample is small, but the sign compatibility is a necessary condition for cross-scale consistency. Third, the Hubble tension is addressed in Paper 11 via environment-dependent Cepheid clock bias — a conditional host-level reconstruction returning $H_0 = 66.65 \pm 1.58$ km/s/Mpc, while the prespecified primary full-ladder estimator retains a 3.94σ residual against Planck; the GW shift is an orthogonal test of bi-metric propagation, not a reconciliation attempt. The 51 host-associated redshifts drive the non-circular signal, while the 56 GWOSC fallback events provide statistical power but anchor the scale toward the fiducial value. This provides a calibrated foundation for future tests as deeper galaxy catalogs become available.

6.2 Limitations

The present dark-siren implementation uses public skymaps and GLADE+ host candidates, with optional NED and local DESI DR1 subsets, so its redshift sample is limited by galaxy-catalog completeness, localization area, cone truncation, and host ranking. Of the 107 events processed, 51 carry host-associated redshifts that do not use the GWOSC redshift field (50 from GLADE+/DESI plus the bright siren GW170817), and 56 use GWOSC-catalog fallback redshifts (cosmology-derived, clearly flagged). The 51 host-associated events drive the non-circular primary signal, while the 56 fallback events are retained only for the secondary robustness check. The low absolute H_0 value (~ 60 km/s/Mpc) should not be interpreted as a competitive measurement of the Hubble constant. It likely reflects current dark-siren host-association incompleteness and public-skymap limitations. The relevant TEP observable in this paper is the differential shift between Λ CDM and TEP under identical host assignments, not the absolute scale. The 0% false-positive rate under the Λ CDM null (for $\Delta\chi^2 > 2$) confirms that the detection threshold is conservative; the 0% recovery rate for the Naivasha lab-sector amplitude shows the current sample is underpowered. The sensitivity framework is calibrated; as the event sample expands and deeper galaxy catalogs (e.g., Rubin/LSST, DESI) become available, the predicted redshift-growth signature of $|A(z) - 1|$ will become resolvable.

6.3 Relation to GWTC-5.0 Modified-Propagation Constraints

The LVK GWTC-5.0 cosmology analysis (Abbott et al., 2026) uses 236 GW sources and reports $H_0 = 71.0^{+9.0}_{-7.1}$ km/s/Mpc, finding no evidence for parameterized deviations from GR propagation. The present analysis is not a generic modified-propagation fit; it tests a lab-fixed TEP conformal template with sign and amplitude inherited from the lab-fixed convention. The LVK modified-propagation basis parameterizes distance-redshift deviations as $d_L^{\text{GW}}(z) = d_L^{\Lambda\text{CDM}}(z)[1 + \Sigma_0 z/(1+z)^n]$, which is phenomenologically different from the TEP conformal factor $A(z) = \exp(\beta_A \varphi_0 [(1+z)^n - 1])$. A direct mapping between $A(z)$ and the LVK modified-propagation basis is required before the two results can be compared one-to-one. In particular, the TEP template predicts a redshift-dependent residual structure in log-distance space ($\ln A(z) \propto [(1+z)^n - 1]$) that is not optimally captured by the LVK linear-bias parameterization. The two analyses are consistent: the LVK analysis finds no generic modified-propagation signal, and the present lab-fixed-template test recovers the corpus-predicted sign at the $\sim 2.6\sigma$ level without bounding the amplitude, because both samples are underpowered relative to the small predicted conformal amplitude (~ 0.4 –1% over the sampled baseline).

7. Conclusions

This paper implements the first observational standard-siren test of the corpus-sign lab-fixed TEP parameterization using combined public GWTC catalogs. Three complementary results emerge from the corrected pipeline. First, the lab-fixed conformal scaling produces the corpus-predicted bi-metric distance-scale shift without adding fitted cosmological degrees of freedom (primary

independent-only sample: Λ CDM $H_0 = 59.8$ to TEP $H_0 = 59.6$ km/s/Mpc, $\Delta\chi^2 = +0.13$, $|\Delta\text{BIC}| < 2$, a marginal preference at equal parameter count. The redshift-dependent matched-filter residual test recovers the corpus-sign structure at $Z = +2.59$ with the sign-flipped control disfavored, while the fitted amplitude exceeds the locked normalization and the shuffle control leaves a noise origin open; the sample is underpowered to bound the predicted signature. Second, the joint MCMC posterior is converged and consistent with the lab-fixed convention TEP parameters ($\phi_0 = +0.013$, $n = 1.0$) at the 68% level, with 74.4% of its mass on the corpus-consistent conformal sign; the posterior does not yet independently constrain the amplitude because the sample is small, but the sign compatibility is a necessary condition for cross-scale consistency. Third, synthetic injection tests show a 0% false-positive rate for decisive ($\Delta\chi^2 > 2$) TEP preference under the Λ CDM null, while the 0% recovery rate for the lab-fixed amplitude confirms that the current sample is underpowered. The predicted bi-metric signature is orthogonal to the Hubble tension, which Paper 11 addresses through a conditional host-level reconciliation route while its prespecified primary estimator retains a residual discrepancy against Planck. The absolute $\Delta\chi^2$ is small ($\Delta\chi^2 = +0.13$ for the primary lab-fixed comparison), the BIC-penalized joint fits remain information-criterion disfavored at the current sample size ($\Delta\text{BIC} = -11.9$ independent, -14.2 full), and the matched-filter residual test recovers the corpus-predicted sign without bounding its amplitude. Evaluated at the distance level, the canonical endpoint map $\Xi = d_L^{\text{GW}}/d_L^{\text{EM}} = 1/(1+z)$ derived from the two-metric action (Step 11) improves the joint fit on the independent sample to $\chi^2 = 69.2$ against $\chi^2 = 74.3$ for Λ CDM ($\Delta\chi^2 = +5.1$ at equal degrees of freedom, $H_0 = 50.0$ km/s/Mpc), a mild preference for the canonical map conditional on the Λ CDM-referenced distance posteriors; the full canonical likelihood including the mass and selection mappings remains the contract-level requirement (Step 09I). The endpoint scaling $d_L^{\text{GW}} = A(z) d_L^{\Lambda\text{CDM}}$ is the primary likelihood model; the TEP-C0 Jordan-frame distance integral is retained as a secondary consistency check. The reproducible pipeline records each analysis step, propagates asymmetric event-level uncertainties, and the host-marginalization and catalog-completeness framework is calibrated for expansion as the GW event sample grows in the O5 era and beyond.

References

- [1] Abbott, B. P., et al. (LIGO/Virgo). (2019). GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs. *Physical Review X*, 9(3), 031040.
- [2] Abbott, B. P., et al. (LIGO/Virgo). (2021). GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. *Physical Review X*, 11(2), 021053.
- [3] Riess, A. G., et al. (2022). A comprehensive measurement of the local value of the Hubble constant with 1 km/s/Mpc uncertainty from the Hubble Space Telescope and the SH0ES team. *The Astrophysical Journal Letters*, 934(1), L7.
- [4] Planck Collaboration. (2020). Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics*, 641, A6.
- [5] Abbott, R., et al. (LIGO/Virgo/KAGRA). (2021). GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. arXiv:2108.01045.
- [6] Abbott, R., et al. (LIGO/Virgo/KAGRA). (2021). GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run. *Physical Review X*, 13(4), 041039.
- [7] The LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. (2025). GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with Observations from the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run. arXiv:2508.18082.
- [8] The LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. (2026). GWTC-5.0: Updated LIGO–Virgo–KAGRA Catalog sets new records in precision gravitational wave astronomy. *News | LIGO Lab | Caltech*, 26 May 2026.
- [9] The LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. (2026). GWTC-5.0: Constraints on the Cosmic Expansion Rate and Modified Gravitational-wave Propagation. arXiv:2605.27227.
- [10] Abbott, B. P., et al. (LIGO/Virgo). (2017). GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Physical Review Letters*, 119(16), 161101.
- [11] Abbott, B. P., et al. (LIGO/Virgo). (2017). A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551(7678), 85–88.
- [12] Dalya, G., et al. (2022). GLADE+: An extended galaxy catalogue for multimessenger searches with advanced gravitational-wave detectors. *MNRAS*, 514(2), 1403–1415.
- [13] DESI Collaboration. (2024). The Early Data Release of the Dark Energy Spectroscopic Instrument. arXiv:2306.06308.
- [14] Fishbach, M., et al. (2019). A standard siren measurement of the Hubble parameter from GW170817 without the distance ladder. *The Astrophysical Journal Letters*, 871(1), L13.
- [15] Palmese, A., et al. (2021). Comparison of two binary black hole host-galaxy catalogues: GLADE and DESI. *MNRAS*, 505(3), 3923–3935.
- [16] Creminelli, P., & Vernizzi, F. (2017). Dark energy after GW170817 and GRB170817A. *Physical Review Letters*, 119(25), 251302.
- [17] Ezquiaga, J. M., & Zumalacárregui, M. (2017). Dark energy after GW170817: Dead ends and the road ahead. *Physical Review Letters*, 119(25), 251304.

Data Availability & Reproducibility

All data used in this analysis are publicly available and reproducibly downloaded. No synthetic, fabricated, or simulated data is used in the main analysis.

Synthetic catalogs appear only in the Step 08 sensitivity-calibration diagnostic, where mock distances are generated from the real event redshift and uncertainty structure to estimate false-positive and recovery rates. They are not used as observational evidence in the main Λ CDM/TEP comparison.

Data sources:

- GWOSC combined catalogs: GWTC-1-confident, GWTC-2, GWTC-2.1-confident, GWTC-3-confident, GWTC-4.0, GWTC-4.1, O4 Discovery Papers, GWTC-5.0 — [gwosc.org](https://www.gwosc.org)
- GraceDB public skymaps (bayestar.fits.gz): gracedb.ligo.org
- GLADE+ Galaxy Catalog (VizieR VII/291): glade.plus
- DESI DR1 fastspec spectroscopic redshift catalog (HEALPix tiles): data.desi.lbl.gov
- NASA/IPAC Extragalactic Database redshift-bearing objects: ned.ipac.caltech.edu

Pipeline steps (18 sequential stages):

Step	Script	Description
00	<code>step_00_download_gwtc5_catalog.py</code>	Download combined GWTC catalogs from GWOSC
01	<code>step_01_precision_filtering.py</code>	Filter events by SNR > 12 and high confidence
01b	<code>step_01b_download_desi.py</code>	Download DESI DR1 fastspec HEALPix tiles for deep galaxy redshifts (optional, large download)
02	<code>step_02_independent_redshifts.py</code>	Build independent-redshift dataset (bright + GLADE+/DESI DR1 dark sirens)
03	<code>step_03_compute_dl_gr.py</code>	Extract GR luminosity distances from LVK posteriors
04	<code>step_04_compute_dl_tep.py</code>	Compute TEP conformal scaling and matter-frame corrected distances
04b	<code>step_04b_metric_direction_audit.py</code>	Audit $A(z)$ vs $1/A(z)$, model-side vs data-side transformation conventions
05	<code>step_05_hubble_diagram.py</code>	Construct Hubble diagram data
06	<code>step_06_h0_reconciliation.py</code>	Secondary analysis: fit H_0 from full sample including fallback redshifts
06b	<code>step_06b_h0_reconciliation_independent.py</code>	Primary analysis: fit H_0 from independent-only redshifts (no circularity)
07	<code>step_07_statistical_tests.py</code>	Goodness-of-fit, tension metrics, and model comparison for both analyses
07b	<code>step_07b_tep_matched_filter.py</code>	Primary discovery statistic: TEP-template residual projection and wrong-sign test
07c	<code>step_07c_independent_only_mcmc.py</code>	Primary independent-only joint MCMC with posterior gamma projection
07d	<code>step_07d_fallback_only_control.py</code>	Fallback-only control MCMC (circular redshifts should suppress signal)
08	<code>step_08_synthetic_injections.py</code>	Synthetic injection test: recovery and false-positive calibration
08e	<code>step_08e_event_jackknife_influence.py</code>	Event jackknife: leave-one-out, redshift cuts, and precision cuts
09	<code>step_09_generate_figures.py</code>	Generate manuscript figures from real pipeline outputs
10	<code>step_10_pipeline_audit.py</code>	Pipeline audit: verify execution integrity and output consistency

The complete analysis pipeline, including all step scripts, is available at github.com/matthewsmawfield/TEP-LVK.