



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

Matthew Lukin Smawfield

Version: v0.6 (Maputo)

First published: 30 September 2026 | Last updated: 28 October 2026

DOI: 10.5281/zenodo.20572695

Status: Preprint

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

Abstract

Standard Λ CDM assumes that gravitational waves and electromagnetic radiation propagate through the same effective distance-redshift relation. The Temporal Equivalence Principle (TEP) predicts a parameter-free canonical bi-metric scaling $\Xi(z) \equiv d_L^{\text{GW}}/d_L^{\text{EM}} = (1+z)^{-1}$ and detector-frame mass scaling $\mathcal{M}_{\text{det}}(z) \propto (1+z)^{3\epsilon-1}$ governed by the interior screening-tracking fraction $\epsilon \in [0, 1]$, alongside a perturbative conformal template $A_{\text{leg}}(z) = \exp[\beta_A \phi_0((1+z)^n - 1)]$. Both channels are tested across cumulative LIGO–Virgo–KAGRA catalogs (GWTC-1 through GWTC-5.0 and O4 Discovery Papers), bright-siren spectroscopy, and GLADE+/DESI DR1 dark-siren host associations.

In the host-catalog-free binary black hole (BBH) population channel, introducing a pair-instability exponential high-mass taper β breaks the scale-free power-law degeneracy and favors TEP over General Relativity out of sample. The unclipped mass-only tapers ($m_2 \geq 8 M_\odot$, cumulative $\mathcal{M} \geq 8 M_\odot$, and smooth Finn–Chernoff and 3-detector H1–L1–V1 selection integrals on $m_2 \geq 5 M_\odot$, $N = 266$) peak directly at $\hat{\epsilon}_{\text{mass}} = 0.25$ ($\Delta \ln L = +5.58, +2.13$, and $+8.62$, strengthening to $+10.28$ – $+12.14$ under Essick pipeline emulation and 200-draw GWOSC PE marginalization, and $+10.21$ – $+10.66$ on $m_2 \geq 8 M_\odot$ under decreasing slope priors $\alpha \geq 1.0$ – 1.5 , with 0/100 synthetic GR false positives). This optimum matches the parameter-free Paper 6 kinetic-screening evaluation under the potential-share identification $\epsilon_{\text{host}} = a_\phi/(g_{\text{host}} + a_\phi)$, $\langle \epsilon_{\text{host}} \rangle_{\text{SFR}}^{\text{UCD}} = 0.236 \pm 0.054$ (0.243 ± 0.052 disk-integrated; 0.197 – 0.202 with 25% bulge/NSC hosts), the selection-marginalized posteriors $\hat{\epsilon}_{\text{marg}} = 0.21 \pm 0.07$ (mass-only) and $0.19_{-0.03}^{+0.05}$ (mass+spin), and the common-mode multi-feature scale-stationarity optimum $\hat{\epsilon}_{\text{cm}} = 0.294 \pm 0.078$ (which reverses GR's $-37.5\% \pm 7.1\%$ [5.31σ ; $-47.7\% \pm 5.2\%$ or 9.10σ on non-redundant $(m_2, 1G m_1)$] high- z compression). Across $N = 276$ BBHs with measured effective spins χ_{eff} , the high-redshift broad-spin mixture weight ($\hat{f}_{2G} = 48.7\%$ at $z_L > 0.3$) matches TEP pair-instability gap occupation (51.2% vs. 13.4% GR), predicts held-out O4 event-level χ_{eff} dispersion without invoking $f_{2G} = f_{\text{gap}}$ ($\Delta \ln L_{\text{disp}}^{\text{O4}} = +4.37$ – $+5.62$, Levene $p = 0.0162$ – 0.0067 ; conditional factorized predictive score under $\epsilon \sim \mathcal{U}[0.24, 0.52]$ of $\Delta \ln Z_{\text{marg}}^{\text{mass+disp}} = +4.83$ [$126 : 1$] vs. fixed-boundary $k \in [-3, +1]$, $+9.19$ vs. $k = 0$, and $+1.40$ – $+2.73$ [$4 : 1$ – $15 : 1$] when GR is granted two extra spin-boundary parameters (M_0, γ) on top of k), and restores 1G-to-2G spectral inheritance ($p_{\text{KS}} = 0.402$ vs. GR 7.98×10^{-4}).

In the distance-redshift channel across 73 non-circular sirens, skymap injections show that Λ CDM-gated host selection absorbs 85%–88% of any conformal shift ($E[\Delta\chi_{\text{dark}}^2] \simeq +1.13$ vs. observed $+1.16$), while diffuse-host splitting ($H_0 = 69.3$ vs. $72.4 \text{ km s}^{-1}\text{Mpc}^{-1}$ in luminous hosts) aligns with DELVE DR3 de-biasing and bright sirens reach 5σ discrimination in $N = 25, 7, 3$, and 1 events at $z = 0.1, 0.2, 0.3$, and 0.5 .

Keywords: Temporal Equivalence Principle, gravitational waves, standard sirens, bi-metric propagation, binary black hole mass function, combined GWTC catalogs

1. Introduction

1.1 Bi-Metric Propagation and the Hubble Tension

Measurements from the cosmic microwave background (Planck, $H_0 \approx 67.4 \text{ km/s/Mpc}$ under Λ CDM) and the local distance ladder (SH0ES, $H_0 \approx 73 \text{ km/s/Mpc}$) differ by approximately 5σ . Paper 11 resolves the Hubble tension through an environment-dependent Cepheid response in the endpoint observing chain, not an additional cosmic-web path-integrated shear. Its host-level corrected value is $H_0 = 67.83 \pm 1.56 \text{ km/s/Mpc}$ (0.26σ from Planck), internally consistent with the TEP-CMB value 66.70 ± 0.58 (Paper 26) at 0.68σ ; that internal agreement is not independent confirmation. The endpoint environmental slope is detected at 2.33σ at the canonical velocity-scatter convention, the within-host period-coupled carrier reaches 2.28 – 2.79σ across anchor conventions, and Paper 11 derives the Cepheid response scale $|\kappa_{p_2}^{\text{pred}}(\lambda_0 = 1)| = 1.196 \times 10^6 \text{ mag}$ directly from the envelope density ratio $\rho_T/\langle\rho\rangle(10 \text{ d}) = 8.45 \times 10^5$ at the unit-transport benchmark $\lambda_0 = 1$ with bare coupling $|\beta_A| = 1$ fixed by Paper 0 ($\lambda_0 = 0.995 \pm 0.401$). 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 Λ CDM 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, matter fields couple to the disformally and conformally related physical metric $\tilde{g}_{\mu\nu}$, whereas gravitational waves propagate on the Einstein-frame tensor sector $g_{\mu\nu}$. Under the canonical two-metric endpoint identity $1 + z = A_0/A_{\text{em}}$ (Paper 0 §2.2; Paper 26), this produces two parameter-free observational predictions for gravitational-wave astronomy. First, the ratio of the gravitational-wave luminosity distance to the electromagnetic luminosity distance at host redshift z follows the canonical endpoint law

$$\Xi(z) \equiv \frac{d_L^{\text{GW}}(z)}{d_L^{\text{EM}}(z; H_0)} = A_{\text{em}}(z) = \frac{1}{1+z}, \quad \tilde{\mathcal{M}}(z; \epsilon) = \mathcal{M}_{\text{det}}^{\text{GR}} (1+z)^{1-3\epsilon}, \quad (1)$$

where $\tilde{\mathcal{M}}$ is the intrinsic matter-frame chirp mass, $\mathcal{M}_{\text{det}}^{\text{GR}}$ is the detector-frame chirp mass, and $\epsilon \in [0, 1]$ is the interior screening-tracking fraction of the progenitor host environment ($\epsilon = 0$ in the source-epoch-pinned limit, where $\mathcal{M}_{\text{det}} \propto (1+z)^{-1}$ in contrast to the standard GR scaling $\mathcal{M}_{\text{det}} \propto (1+z)$). Crucially, Equation (1) is consistent with the stated multimessenger, Solar-System, and supernova benchmark bounds under the adopted radiation-zone and screening assumptions: (i) in the far-zone radiation limit where disformal cone-tilt vanishes ($k^\mu \partial_\mu \phi \rightarrow 0$), the physical metric relation $\tilde{g}_{\mu\nu} = A^2(\phi)g_{\mu\nu}$ is purely conformal, preserving the null cone ($c_T \equiv c$, consistent with the GW170817/GRB 170817A bound $|c_T - c|/c < 3 \times 10^{-15}$; Abbott et al., 2017b) while rescaling the luminosity-distance ratio linearly as $d_L^{\text{GW}}/d_L^{\text{EM}} = A_{\text{em}}/A_0 = (1+z)^{-1}$; (ii) Solar-System fifth-force and post-Newtonian benchmarks are suppressed under Temporal Topology core saturation ($\rho_T \approx 20 \text{ g/cm}^3$; Paper 6) and two-branch high-acceleration kinetic screening ($a_{\text{eff}}/g_N \simeq 2 \times 10^{-8}$ at 1 AU); and (iii) Pantheon+ Type Ia supernova standardization bounds ($\epsilon < 0.150$; Paper 26) are reconciled with $\hat{\epsilon}_{\text{BBH}} = 0.25$ via Chandrasekhar white-dwarf saturation screening ($S_{\text{WD}} = (\bar{\rho}_{\text{WD}}/\rho_T)^{1/3} \simeq 46.4 \Rightarrow \epsilon_{\text{SN Ia}} = \epsilon_{\text{host}}/S_{\text{WD}} \simeq 0.0043$ – $0.0054 \ll 0.150$ for $\epsilon_{\text{host}} = 0.20$ – 0.25).

Second, to quantify continuous perturbative departures from Λ CDM and link the standard-siren distance scale to the historical homogeneous scalar-field benchmark, the conformal factor is also evaluated under the two-parameter exponential family

$$d_L^{(\text{TEP})}(z) = A_{\text{leg}}(z) d_L^{\Lambda\text{CDM}}(z; H_0), \quad A_{\text{leg}}(z) = \exp[\beta_A \phi(z)/M_{\text{Pl}}], \quad \phi(z) = \phi_0 [(1+z)^n - 1]. \quad (2)$$

In the lab-fixed sensitivity benchmark, the parameters ϕ_0 and n are not free: $|\phi_0| = 0.013$ follows from the exploratory Naivasha lab-sector benchmark (now retired in favor of the parameter-free canonical map $\Xi(z) = (1+z)^{-1}$), 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 $n_{\text{eff}} = 0.982$. 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 bare conformal coupling is $\beta_A = -1$ (Paper 0 §2.2; Paper 11), so $\phi(z) > 0$ in the past and the conformal factor satisfies $A(z) < 1$ with the magnitude of the departure growing with redshift.

1.3 This Work

This paper tests both the parameter-free canonical bi-metric scaling $\Xi(z) = (1+z)^{-1}$ and the perturbative conformal template $A_{\text{leg}}(z)$ across cumulative GWTC-1 through GWTC-5.0 catalogs in two complementary channels: (i) the host-catalog-free BBH mass-population and effective-spin distribution channel, and (ii) the standard-siren distance-redshift and environmental host-association channel. Spectral-siren cosmology exploits characteristic features in compact-binary mass distributions to constrain the mapping between gravitational-wave luminosity distance, redshift, and intrinsic source mass (Ezquiaga & Holz, 2022), while recent GWTC-5.0 population analyses and semiparametric reconstructions show that inferred cosmological and modified-propagation parameters depend strongly on the adopted mass-spectrum parameterization and redshift-evolution assumptions (LIGO Scientific, Virgo, and KAGRA Collaborations, 2026b; Tagliazucchi et al., 2026; Pierra et al., 2024; Agapito et al., 2026). Both parameter-free models (ΛCDM , canonical TEP $\Xi(z) = (1+z)^{-1}$, and lab-fixed $A_{\text{leg}}(z)$) and the 4-parameter joint MCMC family (H_0, ϕ_0, n, β) are therefore evaluated alongside bidirectional redshift-evolving GR mass tapers, continuous antenna-selection integrals, and end-to-end dark-siren concordance-dilution controls.

The structure is as follows: Section 2 presents the canonical two-metric radiation identities and the quartic scalar-field tracking derivation; Section 3 defines the cumulative GWTC catalog branches, precision filtering, and circularity-free redshift selection; Section 4 formulates the distance, mass-population, and spin-mixture likelihoods; Section 5 reports the observational results across the distance and BBH population channels; Section 6 discusses the physical implications and resolves the GWTC-5.0 Ξ_0 degeneracy; and Section 7 concludes.

Box 1.1: Executive Summary of Quantitative Audits and Selection-Realism Controls

In the host-catalog-free binary black hole (BBH) population channel, introducing a pair-instability-motivated exponential high-mass taper β breaks the scale-free power-law degeneracy and favors TEP over stationary General Relativity out of sample, with the unclipped mass-only tapers ($m_2 \geq 8 M_\odot$, cumulative $\mathcal{M} \geq 8 M_\odot$, and the smooth 10% Finn–Chernoff and 3-detector H1–L1–V1 network selection integrals on $m_2 \geq 5 M_\odot$) peaking directly at $\hat{\epsilon}_{\text{mass}} = 0.25$ ($\Delta \ln L = +5.58, +2.13$, and $+8.62$ with zero spin input, matching the independent spin-channel optimum $\hat{\epsilon}_{\text{joint}} = 0.25$, while the discontinuous $m_2 \geq 5 M_\odot$ step envelope peaks at $\hat{\epsilon} = 0$ due to 5–8 $M_\odot$ threshold-edge clipping). Across the selection-realism and posterior-marginalization ladder on $m_2 \geq 5 M_\odot$ ($N = 266$), the out-of-sample preference at $\hat{\epsilon} = 0.25$ strengthens and stabilizes from $+7.05$ on the step envelope to $\Delta \ln L_{\text{MFG}} = +8.62$ – $+10.22$ under the smooth Mandel–Farr–Gair Finn–Chernoff selection integral ($\Theta_0 = 2.34$ and 2.68 ; $+8.46$ – $+10.22$ vs $k = 0$ and $+2.87$ – $+2.91$ vs $k = -1.0$ when integrated across the aLIGO Inspiral–Merger–Ringdown [IMR] noise PSD), $\Delta \ln L_{\text{HLV}} = +8.62$ – $+11.36$ under a 40,000-draw 3-detector H1–L1–V1 network antenna integral ($+1.91$ – $+3.48$ vs $k = -1.0$), $\Delta \ln L_{\text{pipe}} = +10.28$ – $+10.72$ under an Essick (2023) semi-analytic search-pipeline emulation ($+2.64$ – $+3.46$ vs $k = -1.0$; or $+6.58$ – $+9.96$ under an unnormalized high-mass glitch-veto penalty), and $\Delta \ln L_{\text{MFG,hier}}^{\text{PE}} = +11.99$ – $+12.14$ under 200-draw hierarchical marginalization over the published GWOSC asymmetric 90% PE intervals with self-consistent per-draw redshift thresholds ($+3.82$ – $+4.31$ vs $k = -1.0$, and $\Delta \ln L_{\text{hier}}^{\text{PE}} = +10.89$ on the step envelope; vs $+6.82$ and $+6.36$ correlated when thresholds are held fixed at the point-estimate redshift), while at the pinned boundary $\epsilon = 0$ the corresponding scores are $+11.02$ step ($\hat{\beta}_{\text{TEP}} = 60.2 M_\odot$ vs $\hat{\beta}_{\text{GR}} = 25.6 M_\odot$; $+2.32$ on $N = 95$ chirp mass, $+4.06$ on $N = 104$ GWTC-5.0 m_2 , $+11.18$ for $m_{2,\text{det}} \geq 5.5 M_\odot$, $+5.13$ to $+18.08$ across $d^{1.0}$ – $d^{1.4}$, and $+14.31$ to $+22.25$ across $[0.90, 1.20]$ A_{env} under smooth MFG vs $+0.79$ to $+25.33$ on the hard step wall), $+20.00$ MFG ($+8.41$ at $\Theta_0 = 2.34$; $+14.20$ – $+18.38$ IMR-PSD), $+6.34$ – $+19.77$ HLV, $+11.41$ – $+21.13$ pipeline emulation, and $+21.60$ – $+28.52$ MFG / $+36.23$ step PE-marginalized ($+27.75$ fixed-threshold, $+26.65$ correlated, $+22.55$ 80-draw SNR-scaled), with 0/100 false positives across 100-catalog synthetic GR injection challenges (mean $\Delta \ln L = -24.95$ to -55.93) and prior-averaged out-of-sample predictive log-evidence $\Delta \ln Z_{\text{marg}} = +7.95$ vs $k = 0$, $+9.82$ vs astrophysical $k \in [0, +1]$, and $+1.17$ $[3.2 : 1]$ vs the full phenomenological interval $k \in [-3, +1]$ under $\epsilon \sim \mathcal{U}[0, 1]$ ($+5.34$ $[208 : 1]$ vs $k = 0$ and $+7.21$ $[1, 350 : 1]$ vs $k \in [0, +1]$ on $m_2 \geq 5 M_\odot$, and on $m_2 \geq 8 M_\odot$ — where all $N_{\text{high}} = 131/131$ high- z events are identical to $m_2 \geq 5 M_\odot$ and enforcing the astrophysically motivated decreasing-power-law slope prior $\alpha \geq 1.0$ – 1.5 to prevent GR’s low- z $\hat{\alpha}_{\text{GR}} = 4.7 \times 10^{-7}$ boundary-collinearity collapse raises the $\epsilon = 0.25$ mass-only preference from $+5.58$ to $+10.21$ – $+10.66$ vs $k = 0$, from $+0.78$ to $+2.85$ – $+3.45$ vs $k = -1.0$ [$+8.75$ – $+17.09$ and $+4.45$ under Essick pipeline emulation, and $+5.68$ on $\mathcal{M} \geq 8 M_\odot$], and raises $\Delta \ln Z_{\text{marg}}$ under $\epsilon \sim \mathcal{U}[0.24, 0.52]$ from $+4.80$ $[122 : 1]$ / $+6.71$ $[824 : 1]$ unconstrained to $+8.55$ – $+8.94$ $[5, 167 : 1]$ – $7, 631 : 1]$ vs $k = 0$, $+10.36$ – $+10.67$ $[31, 571 : 1]$ –

43, 045 : 1] vs $k \in [0, +1]$, and $+3.41+3.99$ [30 : 1–54 : 1] vs $k \in [-1, +1]$ [$+5.10+6.07$ or 164 : 1–433 : 1 on the sub-critical spiral-disk prior $\epsilon \in [0.20, 0.33]$], while against extreme negative tapers $k \in [-3, -1.5]$ that shrink GR's $z_L = 0.8$ cutoff to $\beta \simeq 6.3 M_\odot$ to mimic the 1D TEP Jacobian and leave $+0.45+1.32$ [1.6 : 1–3.7 : 1] on 1D mass alone, the orthogonal $f_{2G} = f_{\text{gap}}$ -free held-out O4 spin-dispersion channel yields a conditional factorized predictive log-score of $\Delta \ln Z_{\text{margin}}^{\text{mass+disp}} = +4.83$ [predictive score ratio 126 : 1]. Refitting redshift-evolving GR tapers $\beta(z) = \beta_0(1+z)^k$ shows that astrophysical low-metallicity wind evolution ($k > 0$) worsens GR and increases TEP's preference to $\Delta \ln L = +12.89$ to $+18.53$ at $\epsilon = 0$ ($+8.92$ to $+14.56$ at $\epsilon = 0.25$), whereas negative $k < 0$ is a phenomenological negative-evolution comparator that partially mimics the 1D TEP Jacobian that still leaves TEP($\epsilon = 0$) ahead throughout $k \in [-3, +1]$ ($+1.11$ to $+8.44$ for $k < 0$) and reaches an interior optimum at $\hat{k} = -1.0$ on $m_2 \geq 8 M_\odot$ ($N = 226$) and cumulative chirp mass ($N = 244$) that is outperformed by TEP($\hat{\epsilon}_{\text{joint}} = 0.25$) at every $k \in [-3, +1]$ ($+0.78$ and $+0.30$ at $\hat{k} = -1.0$ on the redshift split, and $+1.56$ on $m_2 \geq 8 M_\odot$ or $+2.10$ with one fewer free redshift-evolution parameter on $m_2 \geq 5 M_\odot$ under chronological O1–O3 training $\rightarrow$ held-out O4 testing). Across $N = 276$ GWOSC BBHs with measured effective spins χ_{eff} , a two-component 1G+2G mixture yields a broad-spin weight of $\hat{f}_{2G} = 48.7\%$ across the $N = 172$ events at $z_L > 0.3$ (vs 21.8% below $65 M_\odot$), matching the matched $N = 172$ TEP gap occupation (51.2% at $\epsilon = 0.25$) vs 13.4% under GR (10%–20% in GR cluster synthesis), doubles from 10.8% to 22.1% in the precision-matched ($\sigma_\chi \leq \text{median}$, $N = 54$ high- z) subset (24.1%–31.5% TEP gap vs 3.7% GR), predicts held-out O4 event-by-event χ_{eff} dispersion without invoking $f_{2G} = f_{\text{gap}}$ ($\Delta \ln L_{\text{disp}}^{\text{O4}} = +4.37$ to $+5.62$ at $\epsilon = 0.20$ – 0.30 , Levene $p = 0.0162$ – 0.0067 , beating a matched $n_{\text{te}} = 18$ redshift-only classifier [$p = 0.0726$] by $+3.02$ to $+4.27$ and χ_{eff} -error deconvolution by $+2.82$ to $+3.18$; O4b $P(2G_i)$ $p = 0.0305$), and restores spectral compatibility between spin-selected 1G-remnant and 2G-primary proxies ($p_{\text{KS}} = 0.402$, $\text{BC}_{40-120} = 0.9963$ at $\hat{\epsilon}_{\text{joint}} = 0.25$ vs GR $p_{\text{KS}} = 7.98 \times 10^{-4}$) while reproducing NFW-halo-level GW-recoil retention from baryons alone in nested temporal wells. Under the conditional spin-gap fraction-identification hypothesis ($f_{2G} = f_{\text{gap}}$), the composite mass-plus-spin consistency score at $\hat{\epsilon}_{\text{joint}} = 0.25$ beats stationary GR by $\Delta \ln L_{\text{joint}} = +11.31$ on $m_2 \geq 5 M_\odot$ ($+14.48$ Finn–Chernoff) and $+9.84$ on $m_2 \geq 8 M_\odot$, and beats evolving GR at $k = -1.0$ (the $m_2 \geq 8/M \geq 8$ interior optimum) by $+6.55$ ($+7.61$ Finn–Chernoff) and $+5.04$ (at the lower edge of the parameterized field-spiral/dwarf host-transfer bracket $\epsilon_{\text{eff}} \simeq 0.24$ – 0.52), while joint taper profiling offers a quantitative mechanism for why fixed-cutoff spectral-siren pipelines shift toward $\Xi_0 \gtrsim 1.2$.

In the distance-redshift channel across 73 non-circular sirens (1 bright siren GW170817 at $z = 0.0098$, where $1 - \Xi(z) = 0.97\%$ vs 15% distance error, plus 72 unassociated dark sirens; 34 circular GWOSC-fallback events quarantined as a $\gamma^* = 0$ null), skymap injections and host-weight shuffles prove that Λ CDM-gated host selection absorbs 85%–88% of any genuine conformal distance shift ($T \simeq 12\%$ – 15%), leaving the dark-siren distance channel low-power on current data (a priori attenuated expectation $E[\Delta \chi_{\text{dark}}^2] \simeq +1.13$ vs observed $\Delta \chi^2 = +1.16$) and reproducing the baseline fit ($-2 \ln L = 944.31$) at fixed canonical $H_0 = 68.18 \text{ km s}^{-1} \text{ Mpc}^{-1}$ under self-consistent TEP-kernel host reassociation. Environmental host-luminosity splitting yields $H_0 = 72.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ in luminous massive hosts ($N = 35$) versus $69.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$ in diffuse faint hosts ($N = 36$; $70.1 \rightarrow 68.9 \text{ km s}^{-1} \text{ Mpc}^{-1}$ across pool weightings), consistent with DELVE DR3 + GWTC-5.0 catalog de-biasing ($67.4^{+13.0}_{-11.2} \text{ km s}^{-1} \text{ Mpc}^{-1}$). The 4D MCMC posterior ($H_0 = 71.2 \pm 3.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $n = 1.47 \pm 0.86$) resolves factor-level sign non-identifiability via $\phi_{0,\text{eff}} \equiv -\beta\phi_0 = +0.034 \pm 0.069$ (MAP $+0.039$, $\Delta \chi_{\text{MAP}}^2 = +2.43$, 71.4% mass on $A(z) < 1$), and host-association-free bright sirens at fixed $H_0 = 68.18 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (10% fractional distance error) reach 5σ discrimination in $N = 25, 7, 3$, and 1 events (28, 8, 4, and 2 in log-distance) at $z = 0.1, 0.2, 0.3$, and 0.5 .

2. Theoretical Framework

2.1 Canonical Radiation Contract and Historical Template

The canonical endpoint identities on the unified matter metric $\tilde{g}_{\mu\nu} = -A_{\text{loc}}^2(u_{\text{grav}}(\mathbf{x}))U_\mu U_\nu + A_{\text{sp}}^2(u_{\text{drift}}(t, \mathbf{x}))\Pi_{\mu\nu} + B(\phi)(\Pi_\mu^\alpha \Pi_\nu^\beta \nabla_\alpha \phi \nabla_\beta \phi - 2(U \cdot \nabla \phi)U_\mu \Pi_\nu^\alpha \nabla_\alpha \phi)$ ($U_\mu \equiv -\nabla_\mu \tilde{\phi} / \sqrt{2\tilde{X}}$, $\Pi_{\mu\nu} = g_{\mu\nu} + U_\mu U_\nu$; Paper 0, §2.2) are $A_e = (1+z)^{-1}$, $M_e = A_e \tilde{M}$, $M_{\text{det}}^{\text{GR}} = A_e \tilde{M}$, $M_{\text{src}}^{\text{GR}} = A_e^2 \tilde{M}$, 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 within the generalized modified-propagation framework (Nishizawa, 2018; Belgacem et al., 2018). Frame consistency of the mass identities is verified symbolically in Paper 0: a source-local observer measuring with locally determined constants recovers the standard chirp mass ($\tilde{M}_{\text{loc}} \rightarrow \tilde{M}$ 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): interior k-mouflage tracking gives $G_{\text{loc}}(z) \propto (1+z)^{-2}$ inside source galaxies, so characteristic mass scales drift as $\tilde{M}^*(z) \propto (1+z)^{3\epsilon}$, 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 ($\epsilon = 0$, the source-epoch-pinned limit in which the binary orbit follows the emission-epoch coupling) return the naive $A_e = (1+z)^{-1}$ mass map. The same drift rescales SN Ia luminosity by $(1+z)^{3\epsilon}$, and the supernova ladder bounds it first for SN Ia progenitors (Paper 26 joint profile bound $\epsilon < 0.150$ at 3σ , $\Delta \chi^2 = 805.8$ at $\epsilon = 1$, conditional on the fixed- $z_T = 5$ M1 distance relation and Chandrasekhar

prescription, superseding the historical Paper 0 residual-floor bracket $\epsilon \lesssim 0.05\text{--}0.10$). How that SN-sector bound relates to BBH interiors through the shared $G^{-3/2}$ scaling is parameterized by the host-assembly brackets (Paper 0 §8; Section 5.6): under the SN-sector bound $\epsilon < 0.150$ the discriminant runs as $\mathcal{M}_{\text{det}} \propto (1+z)^{-0.55} - (1+z)^{-1} \propto (1+z)^{-0.28}$ to $(1+z)^{+0.56}$ across the field-spiral/dwarf parameterized host-transfer bracket $\epsilon_{\text{eff}} \simeq 0.24\text{--}0.52$, with the unclipped mass-only and joint mass-plus-spin GWTC optima converging at $\hat{\epsilon} = 0.25$ at the lower edge of the host bracket, while the discontinuous $m_2 \geq 5 M_\odot$ step envelope peaks at pinned $\hat{\epsilon} = 0$ due to $5\text{--}8 M_\odot$ threshold-edge clipping), while near-full tracking reverses it to $\propto (1+z)^{3\epsilon-1}$.

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 the theory level by the environmental operator $\mathcal{S}_\Sigma(\mathcal{E})$. Quantities such as ρ_T , $R_T(M)$, $S_\oplus(r)$, compactness Φ/c^2 , local stellar density, thermal epoch, coherence length, proximity, and boundary geometry are domain-specific projections of $\mathcal{E}$, not independent screening mechanisms and not interchangeable universal thresholds.

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

where $\phi_0 = +0.013$ ($\equiv \phi_{0,\text{leg}}$, the legacy homogeneous-template slope parameter of $A_{\text{legacy}}(z)$, distinct from the full cosmological endpoint map $\Xi(z) = 1/(1+z)$ and from the screened solar exterior charge $S_\Sigma^{(\odot)}$ in 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

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

so that $\phi_{\text{min}} \propto \bar{\rho}_m^{1/3}$ to leading order. Adopting the standard expanding-background matter-dilution benchmark $\bar{\rho}_m \propto (1+z)^3$ to parameterize the historical $A_{\text{legacy}}(z)$ template 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 $n_{\text{eff}} = 0.982$, 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 $\lambda \simeq 2.3 \times 10^{-118} M_{\text{Pl}}^4$ — the quartic normalization matched to this historical homogeneous-template amplitude. This matching fixes a benchmark coefficient; it does not derive density-dependent running of the operative matter-domain

coupling. 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 two normalizations characterize different benchmark realizations of that response family; connecting them within one fixed master action requires a background solution rather than identifying them as a derived running coupling (Paper 0 §8); the free- n joint fit ($n = 1.47 \pm 0.86$) is retained as a consistency check and agrees with the derived index at the 0.6σ level.

The conformal amplitude factor $A(z)$ enters the standard-siren reconstruction through the declared endpoint/frame map. It is distinct from the incremental gradient/source response $S_\Sigma(\mathcal{E})$, which must be evaluated on the selected kinetic field solution when a dynamical screening prediction is required. Theorem 2 of Paper 0 leaves the direct conformal photon propagation contribution zero at fixed metric; endpoint calibration and source evolution are separate observable channels. 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 and Action-Level Strain/Mass Derivation

The bi-metric framework distinguishes between the physical Jordan/matter metric $\tilde{g}_{\mu\nu}$ governing electromagnetic rods, atomic clocks, and baryonic matter, and the Einstein-frame tensor metric $g_{\mu\nu}$ governing transverse-traceless gravitational-wave propagation. Both canonical observables in Equation (1) follow from a single self-consistent conformal convention $\tilde{g}_{\mu\nu} = A^2(\phi) g_{\mu\nu}$ in the far-zone radiation limit, where Steps 1–2 and the pinned $(1+z_T)^{-1}$ inspiral factor follow directly from the coupled Einstein-frame tensor action and conformal matter clock/length calibration, while the $(1+z_T)^{3\epsilon}$ factor in Step 3 enters through the interior progenitor/orbital screening-tracking response:

1. *Coupled action and clock-redshift endpoint identity:* In the canonical TEP action

$S = \frac{1}{2} M_{\text{Pl}}^2 \int d^4x \sqrt{-g} R[g] + S_\phi[g, \phi] + S_m[\tilde{g}_{\mu\nu}, \psi_m]$, the tensor sector is already canonically normalized in $g_{\mu\nu}$ with constant Planck mass M_{Pl} (zero Einstein-frame running Planck mass, $\alpha_M^{\text{E}} = 0$), while matter clocks and rods measure proper intervals $d\tilde{s}^2 = \tilde{g}_{\mu\nu} dx^\mu dx^\nu = A^2(\phi) g_{\mu\nu} dx^\mu dx^\nu$ (so proper time and length scale linearly with $A(\phi)$, $d\tilde{t} = A(\phi) dt$ and $d\tilde{\ell} = A(\phi) d\ell$, while the metric tensor itself carries $A^2(\phi)$). Photon frequency comparison between source emission at ϕ_{em} and detection at ϕ_0 ($A_0 \equiv A(\phi_0) = 1$) therefore fixes the fundamental endpoint identity

$$1+z \equiv \frac{\tilde{\nu}_{\text{em}}}{\tilde{\nu}_{\text{obs}}} = \frac{d\tilde{t}_0}{d\tilde{t}_{\text{em}}} = \frac{A_0}{A_{\text{em}}} \implies \frac{A_{\text{em}}}{A_0} = (1+z)^{-1}.$$

2. *WKB tensor strain transport and $\Xi(z) = (1+z)^{-1}$:* Because the quadratic tensor action

$S_{\text{GW}}^{(2)} = \frac{1}{8} M_{\text{Pl}}^2 \int d^4x \sqrt{-g} g^{\mu\nu} \partial_\mu h_{ij}^{\text{TT}} \partial_\nu h_{ij}^{\text{TT}}$ is canonically normalized on $g_{\mu\nu}$, the WKB strain amplitude decays inversely with the Einstein-frame area radius r_E , $h^{\text{TT}} \propto r_E^{-1}$. In the source-local matter frame ($\tilde{t}_{\text{em}} = A_{\text{em}} t$, $\tilde{r}_{\text{em}} = A_{\text{em}} r_E$), a standard GR waveform template calibrates the strain amplitude to the physical luminosity distance $d_L^{\text{EM}}(z) = (1+z) A_0 r_E = (A_0^2/A_{\text{em}}) r_E$, whereas the Einstein-frame tensor wave transports the source-frame proper-unit calibration $A_{\text{em}} r_E$ directly to the detector. Taking the ratio of the GW strain-inferred distance $d_L^{\text{GW}}(z) = A_0 r_E$ to the electromagnetic luminosity distance $d_L^{\text{EM}}(z) = (A_0^2/A_{\text{em}}) r_E$ yields the exact linear conformal ratio with no fractional powers:

$$\Xi(z) \equiv \frac{d_L^{\text{GW}}(z)}{d_L^{\text{EM}}(z)} = \frac{A_{\text{em}}}{A_0} = \frac{1}{1+z}.$$

3. *Inspiral phase invariant and $\tilde{\mathcal{M}}(z_T; \epsilon) = \mathcal{M}_{\text{det}}^{\text{GR}}(1+z_T)^{1-3\epsilon}$:* A detector-frame GR matched filter measures the inspiral phase evolution $df_{\text{det}}^{-5/3}/d\tilde{t}_0 \propto (G_0 \mathcal{M}_{\text{det}}^{\text{GR}}/c^3)^{5/3}$. Since wave-cycle number $dN = \tilde{f}_{\text{em}} d\tilde{t}_{\text{em}} = f_{\text{det}} d\tilde{t}_0$ is a topological scalar invariant and $d\tilde{t}_0/d\tilde{t}_{\text{em}} = A_0/A_{\text{em}} = 1+z_T$, the detector-frame inspiral time constant is $G_0 \mathcal{M}_{\text{det}}^{\text{GR}}/c^3 = (1+z_T) G_{\text{loc}}(z_T) \mathcal{M}_{\text{phys}}(z_T)/c^3$. Combining (a) the clock-redshift dilation $(1+z_T)$, (b) the unscreened local gravitational coupling scaling $G_{\text{loc}}(z_T)/G_0 = (A_{\text{em}}/A_0)^2 = (1+z_T)^{-2}$ from the $A^2(\phi)$ metric coupling, and (c) the interior progenitor/orbital mass-scale drift $\mathcal{M}_{\text{phys}}(z_T) = \tilde{\mathcal{M}}[(G_{\text{loc}}/G_0)^{-3/2}]^\epsilon = \tilde{\mathcal{M}}(1+z_T)^{3\epsilon}$ governed by the host screening-tracking fraction $\epsilon \in [0, 1]$ yields

$$\mathcal{M}_{\text{det}}^{\text{GR}}(z_T; \epsilon) = (1+z_T) \frac{G_{\text{loc}}(z_T)}{G_0} \mathcal{M}_{\text{phys}}(z_T) = (1+z_T) (1+z_T)^{-2} \tilde{\mathcal{M}}(1+z_T)^{3\epsilon} = \tilde{\mathcal{M}}(1+z_T)^{3\epsilon-1} \iff \tilde{\mathcal{M}}$$

Cross-paper metric and screening consistency (Paper 0 §2.2 & Paper 6). Within the unified matter metric $\tilde{g}_{\mu\nu} = -A_{\text{loc}}^2(u_{\text{grav}}(\mathbf{x})) U_\mu U_\nu + A_{\text{sp}}^2(u_{\text{drift}}(t, \mathbf{x})) \Pi_{\mu\nu} + B(\phi)(\dots)$ of Paper 0 §2.2, the quasi-static galactic and planetary soliton analyses of Paper 6 (New Delhi) operate at fixed cosmic epoch t_0 ($A_{\text{sp}}(t_0) = 1$), where spatial gradients of the lapse $A_{\text{loc}}(\mathbf{x}) = e^{-u_{\text{grav}}(\mathbf{x})}$ govern local clock rates and dynamics, whereas long-baseline gravitational-wave and electromagnetic propagation across cosmological epochs ($u_{\text{drift}}(t) = \bar{\phi}(t)/M_{\text{Pl}}$, $A_{\text{loc}}(t) = A_{\text{sp}}(t) = A(\bar{\phi}(t))$, $B(\phi) k^\mu \partial_\mu \phi \rightarrow 0$) specializes to

the conformal radiation metric $\tilde{g}_{\mu\nu} = A^2(\bar{\phi}) g_{\mu\nu}$ derived above. Likewise, the unclipped quartic density-tracking ratio $S(\bar{\rho}) \equiv u_{\min}(\bar{\rho})/u_{\min}(\rho_T) = (\bar{\rho}/\rho_T)^{1/3} = R_T(M)/R_{\text{phys}}$ (Paper 6 Appendix C.2, with $\rho_T \approx 20 \text{ g/cm}^3$) enters surface rotational kinematics through the finite-body active-radius ceiling $S_A(\bar{\rho}) = R_{\text{active}}/R_{\text{phys}} = \min[1, (\bar{\rho}/\rho_T)^{1/3}]$, while in over-saturated compact interiors ($\bar{\rho} > \rho_T$, such as Chandrasekhar white dwarfs with $S_{\text{WD}} \simeq 46.4$) the unclipped ratio $S_{\text{prog}} = (\bar{\rho}_{\text{prog}}/\rho_T)^{1/3} > 1$ suppresses interior cosmological G_{loc} drift via the closed-form two-branch host-plus-progenitor transfer law $\epsilon(\mathcal{E}) = [\max(1, S_{\text{prog}})]^{-1} \epsilon_{\text{host}}(g_{\text{host}}) = [\max(1, S_{\text{prog}})]^{-1} (\sqrt{1 + 4a_{\text{eff}}/g_{\text{host}}} - 1)/(\sqrt{1 + 4a_{\text{eff}}/g_{\text{host}}} + 1)$ at Paper 6's fixed galactic shear threshold $a_{\text{eff}} = 1.20 \times 10^{-10} \text{ m s}^{-2}$ (where a_{eff} and ρ_T are calibrated in Paper 6 on galactic and Solar-System dynamics and cross-applied to GWTC with zero gravitational-wave fitting: evaluating $\epsilon_{\text{host}}(g_{\text{host}})$ across morphological host classes predicts $\epsilon_{\text{NSC}} = 0.001$ in dense nuclear star cluster cores [$10^7 M_{\odot}$, 2.5 pc, $g/a_{\text{eff}} = 1,858$], $\epsilon_{\text{bulge}} = 0.140$ in massive early-type spirals/bulges [$8 \times 10^{10} M_{\odot}$, 4.2 kpc, $g/a_{\text{eff}} = 5.27$, matching the $\epsilon \simeq 0.12$ – 0.14 peak of the 10% Finn–Chernoff profile], $\epsilon_{\text{spiral}} = 0.292$ in Milky-Way-like disks [$3 \times 10^{10} M_{\odot}$, 4.5 kpc; SFR-weighted spiral-disk mean $\langle \epsilon_{\text{host}} \rangle_{\text{SFR}}^{\text{UCD}} = 0.236 \pm 0.054$, median 0.233, 68% CI 0.182–0.290], $\epsilon_{\text{dwarf}} = 0.438$ in low-mass dwarf disks [$3 \times 10^9 M_{\odot}$, 2.2 kpc], and $\langle \epsilon_{\text{BBH}} \rangle_{\text{mix}}^{\text{UCD}} = 0.75(0.236) + 0.25(0.080) = 0.197$ for a 75% SFR-weighted field-spiral [0.236] plus 25% dense bulge/NSC [0.080 mean across 0.140 bulges and 0.001 NSC cores] mixture, matching both the unclipped $\hat{\epsilon} = 0.25$ optimum and the selection-marginalized posteriors $\hat{\epsilon}_{\text{marg}} = 0.21 \pm 0.07$ [mass-only] and $0.19_{-0.03}^{+0.05}$ [mass+spin]; physically, whereas a degenerate carbon–oxygen white dwarf has $\bar{\rho}_{\text{WD}} \simeq 2 \times 10^6 \text{ g/cm}^3 \gg \rho_T$ [$S_{\text{WD}} \simeq 46.4$], massive 30–100 $M_{\odot}$ BBH progenitor stars have mean densities strictly below $\rho_T = 20 \text{ g/cm}^3$ throughout $> 99.9\%$ of their evolutionary lifetime — both on the H-burning main sequence [$\bar{\rho}_{\text{MS}} \simeq 0.01$ – 0.1 g/cm^3] and during the stripped Wolf–Rayet / He-burning phase where the helium-core mass and pair-instability boundary are determined [$\bar{\rho}_{\text{WR}} \simeq 0.5$ – $5 \text{ g/cm}^3 < \rho_T \implies S_{\text{prog}} \simeq 0.1$ – $0.6 < 1$ and $\max(1, S_{\text{prog}}) = 1$; while any partial screening during brief final C/O core contraction at $\bar{\rho}_{\text{core}} \sim 50$ – 100 g/cm^3 , $S_{\text{core}} \simeq 1.36$ – 1.71 , shifts $\epsilon_{\text{eff}} = \epsilon_{\text{host}}/S_{\text{core}}$ from 0.236 to 0.14–0.17, spanning the observed $\epsilon \simeq 0.14$ – 0.25 window] — and evolve on wide orbits $a_{\text{orb}} \gg R_T \gg R_S$ in the host potential g_{host} prior to the final ~ 0.2 s merger in the LVK band; Section 6.1).

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 and Sample-Branch Accounting

Events are filtered across two complementary analysis branches according to the observables required by each channel. In the *skymap distance-redshift siren branch*, events are filtered to a high-confidence subset of $N = 107$ events with signal-to-noise ratio (SNR) > 12 , false-alarm rate (FAR) < 1 per year when available, and $p_{\text{astro}} > 0.9$ when available. Of these 107 events, $N = 73$ carry non-circular host-associated redshifts (the primary distance sample: 1 bright siren plus 72 GLADE+/DESI DR1 dark sirens), while the remaining 34 events lack catalog host associations and carry GWOSC Λ CDM-derived redshifts quarantined as a cosmology-blind pipeline null ($\gamma^* = 0$). The primary bright-siren anchor is GW170817, the confirmed binary neutron-star merger in NGC 4993 with spectroscopic host redshift $z = 0.0098$ (heliocentric; $z_{\text{CMB}} = 0.0092$).

In the *host-catalog-free BBH population and spin branch*, no galaxy catalog or skymap association is used. From the deduplicated cumulative GWTC-1 through GWTC-5.0 catalog, $N = 276$ unique BBH events carry published effective inspiral spin measurements χ_{eff} and luminosity distances (used in the two-component 1G+2G spin-mixture test), $N = 267$ carry component mass estimates (m_1, m_2) with $\mathcal{M}_{\text{src}}^{\text{rep}} \geq 5 M_{\odot}$ (used in the exploratory discrete chirp-mass ladder coherence test; $N = 266$ satisfy $m_2 \geq 5 M_{\odot}$ and $N = 226$ satisfy $m_2 \geq 8 M_{\odot}$ in the secondary-mass pair-instability cutoff tests), and $N = 244$ carry directly reported source chirp masses $\mathcal{M}_{\text{src}}^{\text{rep}} \geq 8 M_{\odot}$. Within the GWTC-5.0 catalog extract, $N = 95$ BBH events satisfy $\mathcal{M}_{\text{src}}^{\text{rep}} \geq 8 M_{\odot}$ with complete

chirp-mass bounds (45 low-redshift anchors at $z_L < 0.4$ and 50 high-redshift scored events at $z_L \geq 0.4$; $N = 104$ with $m_2 \geq 5 M_\odot$), and $N = 43$ high-SNR BBHs form the primary-component $\tilde{m}_1$ PISN-window diagnostic subset.

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 (compare joint catalog-population dark-siren formalisms in Gray et al., 2023; Chen, Gray & Baker, 2024); 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 34 of 107 events. The remaining 73 events carry host-associated redshifts that do not use the GWOSC redshift field (72 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 73 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 bright-siren spectroscopy and 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.

4.4 Catalog-Free Population and Spin-Mixture Likelihoods

In the host-catalog-free BBH population channel, each event is characterized purely by its measured detector-frame mass $m_{\text{det}}^{\text{GR}}$ and gravitational-wave luminosity distance d_L^{GW} , without reference to any galaxy catalog or skymap. Let z_L denote the standard ΛCDM redshift inferred from d_L^{GW} and z_T denote the matter-frame redshift solving $d_L^{\text{EM}}(z_T)/(1+z_T) = d_L^{\text{GW}}$ under the canonical two-metric endpoint law $\Xi(z_T) = (1+z_T)^{-1}$. For an interior screening-tracking fraction $\epsilon \in [0, 1]$, the intrinsic matter-frame mass coordinate $\tilde{m}(z_T; \epsilon)$ and its detector-to-intrinsic coordinate Jacobian are

$$\tilde{m}(z_T; \epsilon) = m_{\text{det}}^{\text{GR}} (1+z_T)^{1-3\epsilon}, \quad \left| \frac{\partial \tilde{m}}{\partial m_{\text{det}}^{\text{GR}}} \right| = (1+z_T)^{1-3\epsilon}, \quad (9)$$

whereas standard General Relativity assigns the physical source-frame mass $m_{\text{src}}^{\text{GR}}(z_L) = m_{\text{det}}^{\text{GR}}(1+z_L)^{-1}$ with detector-to-source Jacobian $(1+z_L)^{-1}$ (and for a redshift-evolving GR taper $\beta(z_L) = \beta_0(1+z_L)^k$, defines the rescaled population coordinate $x_{\text{GR}}(z_L; k) \equiv m_{\text{src}}^{\text{GR}}(1+z_L)^{-k} = m_{\text{det}}^{\text{GR}}(1+z_L)^{-1-k}$ with Jacobian $(1+z_L)^{-1-k}$; because the per-event detection threshold satisfies $M_{\text{min},i}^{\text{GR}} \geq m_{\text{lo}} = 5 M_{\odot}$, evaluating the likelihood in x_{GR} is mathematically identical for all $k \leq 0$ to evaluating $p(m_{\text{src}}^{\text{GR}} | z_L) \propto (m_{\text{src}}^{\text{GR}})^{-\alpha} \exp[-m_{\text{src}}^{\text{GR}}/(\beta_0(1+z_L)^k)]$ above the fixed physical floor $m_{\text{lo}} = 5 M_{\odot}$). To break the scale-free power-law degeneracy inherent to spectral-siren mass inference (Ezquiaga & Holz, 2022; Pierra et al., 2024; Tagliazucchi et al., 2026; LIGO Scientific, Virgo, and KAGRA Collaborations, 2026b), both frames are modeled by an exponentially tapered power-law mass function with characteristic pair-instability-motivated taper scale β above $m_{\text{lo}} = 5 M_{\odot}$,

$$p(\tilde{m} | \alpha, \beta) \propto \tilde{m}^{-\alpha} \exp\left(-\frac{\tilde{m}}{\beta}\right), \quad \ln L(\epsilon) = \sum_{i \in \text{high-}z} \ln \left[\frac{p(\tilde{m}_i(\epsilon) | \hat{\alpha}_{\text{low}}, \hat{\beta}_{\text{low}})}{\int_{\tilde{m}_{\text{min},i}(\epsilon)}^{\infty} p(\tilde{m} | \hat{\alpha}_{\text{low}}, \hat{\beta}_{\text{low}}) d\tilde{m}} \left| \frac{\partial \tilde{m}_i}{\partial m_{\text{det},i}^{\text{GR}}} \right| \right], \quad (10)$$

where $(\hat{\alpha}_{\text{low}}, \hat{\beta}_{\text{low}})$ are calibrated by maximum likelihood on the low-redshift anchor subsample ($z_L < 0.4$) — and explicitly refitted from scratch on the low-redshift training subsample for every tested tracking parameter ϵ and every tested GR evolution exponent k in $\beta(z_L) = \beta_0(1+z_L)^k$ — before scoring strictly out of sample on the high-redshift subsample ($z_L \geq 0.4$) above the detector-mass selection envelope $m_{\text{det},\text{min},i} = \max[A_{\text{env}}(d_{L,i}^{\text{GW}})^{6/5}, m_{\text{cut}}(1+z_{L,i})]$. Five structural audits protect the secondary-mass ($m_2 \geq 5 M_{\odot}$) taper score against selection-boundary and point-mass systematics: (i) on the high-redshift scored sample ($z_L \geq 0.4$, $N_{\text{high}} = 131$), $A_{\text{env}}(d_{L,i}^{\text{GW}})^{6/5}$ exceeds $m_{\text{cut}}(1+z_{L,i})$ on every single event (0/131 bound by the catalog cut), and replacing $m_{\text{cut}}(1+z_{L,i})$ across the entire catalog with a purely detector-frame threshold $m_{2,\text{det}} \geq 5.5 M_{\odot}$ (containing zero reference to z_L) increases the out-of-sample preference from $\Delta \ln L = +11.02$ to $\Delta \ln L = +11.18$ ($\ln L_{\text{TTP}} = -491.79$ vs -502.97); (ii) replacing the hard step-function truncation with a Mandel, Farr & Gair (2019)-style selection-weighted denominator $\alpha_i(\alpha, \beta) = \int_{m_{\text{lo}}}^{\infty} p(\tilde{m} | \alpha, \beta) P_{\text{det}}^{\text{FC}}(\tilde{m} | d_{L,i}) d\tilde{m}$ using the piecewise-clamped Finn & Chernoff (1993) analytic antenna-projection survival approximation $P_{\text{det}}^{\text{FC}}(\Theta > \Theta_{\text{thr}}) = \frac{1}{256} (4 - \Theta_{\text{thr}})^4 (1 + \Theta_{\text{thr}})$ for $\Theta_{\text{thr}} \in [0, 4]$ (and $P_{\text{det}}^{\text{FC}} \equiv 0$ for $\Theta_{\text{thr}} \geq 4$, i.e. $\tilde{m}/\tilde{m}_{\text{min},i} \leq (\Theta_0/4)^{6/5}$) with threshold orientation $\Theta_{\text{thr}}(\tilde{m}, d_{L,i}) = \Theta_0 (\tilde{m}_{\text{min},i}/\tilde{m})^{5/6}$ increases the out-of-sample preference to $\Delta \ln L_{\text{MFG}} = +20.00$ at $\epsilon = 0$ ($\ln L = -284.72$ vs -304.72) and $\Delta \ln L_{\text{MFG}} = +10.22$ at $\epsilon = 0.25$ (-294.50 vs -304.72 ; and $+13.12$ at $\epsilon = 0$ or $+3.35$ at $\epsilon = 0.25$ against evolving GR at $k = -1.0$) for $\Theta_0 = 2.68$ ($P_{\text{det}}^{\text{FC}}(\Theta_0) = 4.36\%$ orientation survival at $\tilde{m} = \tilde{m}_{\text{min},i}$), while anchoring at $\Theta_0 = 2.336$ ($P_{\text{det}}^{\text{FC}}(\Theta_0) = 10.0\%$ survival at $\tilde{m} = \tilde{m}_{\text{min},i}$) yields $\Delta \ln L_{\text{MFG}} = +8.41$ at $\epsilon = 0$ and $+8.62$ at $\epsilon = 0.25$ (where $\epsilon = 0.25$ exceeds $\epsilon = 0$ on the mass taper itself and gives a composite mass+spin score of $+12.88$); (iii) performing an 80-draw event-uncertainty marginalization — where for each event

$K = 80$ draws ($m_{2,\text{det},ik}, d_{L,ik}$) are generated from the published GWOSC split-normal luminosity-distance interval (scaled by 0.5 along the SNR distance-mass degeneracy) and SNR-scaled detector-mass uncertainty $\sigma(\ln m_{2,\text{det}}) = 0.05$ ($12/\text{SNR}_i$) subject to $q = m_2/m_1 \leq 1$, and averaged in likelihood space at the event's baseline threshold — yields $\Delta \ln L_{\text{hier}} = +22.55$ at $\epsilon = 0$ ($\ln L_{\text{hier}} = -477.21$ vs -499.77 ; $+4.58$ at $\epsilon = 0.25$, stable within $< 0.3 \ln L$ units at $K = 160$ and 320); (iv) varying the envelope power-law exponent across $d^{1.0}$ – $d^{1.4}$ yields $\Delta \ln L = +5.13, +11.02$, and $+18.08$ at $\epsilon = 0$ (and perturbing A_{env} across $[0.90, 1.20]$ A_{env} maintains $\Delta \ln L \in [+0.79, +25.33]$); and (v) because $q = m_2/m_1 \leq 1$, posterior scatter at $q \approx 1$ is one-sided toward lower m_2 , with symmetric detector-frame Monte Carlo resampling yielding a median $\Delta \ln L = +13.32$ (68% interval $[+8.82, +17.68]$, 100% positive).

To test whether the high-redshift BBH population occupying the 65–130 $M_\odot$ pair-instability supernova (PISN) mass gap under the $\epsilon = 0$ map carries the kinematic signature of multi-generation (2G+) hierarchical mergers, the catalog posterior-median effective inspiral spins $\chi_{\text{eff},i}$ across all $N = 276$ GWOSC BBHs are fitted with a two-component Gaussian mixture likelihood:

$$p(\chi_{\text{eff},i} | f_{2\text{G}}) = (1 - f_{2\text{G}}) \mathcal{N}(\chi_{\text{eff},i}; \mu_{1\text{G}}, \sigma_{1\text{G}}^2) + f_{2\text{G}} \mathcal{N}(\chi_{\text{eff},i}; \mu_{2\text{G}}, \sigma_{2\text{G}}^2), \quad (11)$$

where the narrow first-generation (1G) natal component has $(\mu_{1\text{G}}, \sigma_{1\text{G}}) = (0.05, 0.08)$ and the isotropic hierarchical (2G+) remnant component has $(\mu_{2\text{G}}, \sigma_{2\text{G}}) = (0.0, 0.25)$ arising from dimensionless remnant spins $a_{\text{fin}} \approx 0.68$. To ensure that broader χ_{eff} posteriors at lower signal-to-noise do not bias the inferred mixture, the fit is verified both on the precision-matched half of the catalog ($\sigma_{\chi,i} \leq \text{median}(0.146)$, $N = 140$, where $\sigma_{\chi,i} = \frac{1}{2}(|\sigma_{\chi,i}^-| + |\sigma_{\chi,i}^+|)/1.645$) and by integrating the posterior-expected spin magnitude $E[|\chi_{\text{eff}}|_i]$ over each event's individual split-normal uncertainty $\sigma_{\chi,i}$. Note that in the smooth selection likelihood $p(M_i | d_{L,i}, \text{det}, \alpha, \beta) = p(M_i | \alpha, \beta) P_{\text{det}}(M_i | d_{L,i}) / \alpha_i(\alpha, \beta)$, the numerator detection factor $P_{\text{det}}(M_i | d_{L,i}) = P_{\text{det}}(m_{\text{det},i}/m_{\text{det,min},i})$ depends only on observed detector-frame quantities and is independent of (α, β) ($\partial_\alpha \ln P_{\text{det}} = \partial_\beta \ln P_{\text{det}} \equiv 0$, leaving the low- z maximum-likelihood estimators $(\hat{\alpha}_{\text{low}}, \hat{\beta}_{\text{low}})$ unchanged), while in the high- z comparison $\Delta \ln L = \ln L_{\text{TEP}}(\epsilon) - \ln L_{\text{GR}}(k)$ the ratio $M_i/M_{\text{min},i} = (m_{\text{det},i} \text{ jac}_i) / (m_{\text{det,min},i} \text{ jac}_i) = m_{\text{det},i}/m_{\text{det,min},i}$ is strictly frame-invariant, so $\ln P_{\text{det}}^{\text{TEP}}(M_i | d_{L,i}) - \ln P_{\text{det}}^{\text{GR}}(M_i | d_{L,i}) \equiv 0$ cancels identically event-by-event (and draw-by-draw). In addition, to evaluate prior-averaged model comparison without point-optimum selection bias, the prior-marginalized out-of-sample predictive log-evidence $\ln Z_{\text{marg}} \equiv \ln \int \exp[\ln L(\theta)] \pi(\theta) d\theta$ (with $(\hat{\alpha}_{\text{low}}, \hat{\beta}_{\text{low}})$ refitted on the low- z training sample at each grid point) is integrated under both an uninformative prior $\epsilon \sim \mathcal{U}[0, 1]$ and the external Paper 6 host-screening prior $\epsilon \sim \mathcal{U}[0.24, 0.52]$ against stationary GR ($k = 0$), astrophysical low-metallicity wind evolution $k \sim \mathcal{U}[0, +1]$, and the full phenomenological interval $k \sim \mathcal{U}[-3, +1]$ across both the step envelope and smooth Mandel–Farr–Gair selection integrals (as well as under joint 3-parameter Monte Carlo marginalization over the low- z Hessian posterior covariance $p(\alpha, \ln \beta | D_{\text{low}}, \theta)$ and in combination with the $f_{2\text{G}} = f_{\text{gap}}$ -free held-out O4 spin-dispersion likelihood $\ln L_{\text{disp}}^{\text{O4}}(\epsilon)$, which breaks the 1D mass-only degeneracy $k_{\text{eq}}(\epsilon) \simeq -1 - 1.75(1 - 3\epsilon)$). Finally, in the chronological O1–O3 ($N = 94$) $\rightarrow$ held-out precision-matched O4 ($N = 110$) event-level spin-dispersion test (which does not invoke $f_{2\text{G}} = f_{\text{gap}}$), the two-state Gaussian dispersion model is benchmarked against a sample-size-matched redshift-only classifier ($z_L \geq 0.500$, matching $n_{\text{te}} = 18$ –19), a pure detector-mass classifier ($m_{1,\text{det}} \geq 58.7 M_\odot$, corresponding to $\epsilon = 1/3$), and a measurement-error-deconvolved likelihood $\sigma_{\text{tot},i}^2 = \sigma_{\text{int},g}^2 + \sigma_{\chi,i}^2$ where $\sigma_{\text{int},g}^2 = \max(\sigma_g^2 - \text{median}(\sigma_{\chi,\text{tr}}^2), 0.02^2)$.

5. Results

5.1 Hubble Diagram and Bi-Metric Distance Trajectories

Figure 1a overlays the primary independent-redshift standard-siren sample ($N = 73$: 1 spectroscopic bright siren GW170817 plus 72 GLADE+/DESI DR1 dark sirens) and the 34 quarantined GWOSC Λ CDM-fallback controls ($\gamma^* = 0$) against three distance-redshift models: the Λ CDM baseline ($H_0 = 72.0 \text{ km s}^{-1} \text{ Mpc}^{-1}$), the perturbative lab-fixed TEP template $A_{\text{leg}}(z)$ ($H_0 = 72.2 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Delta\chi^2 = +0.20$), and the parameter-free canonical two-metric prediction $\Xi(z) = (1 + z)^{-1}$ at fixed canonical $H_0 = 68.18 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Paper 18). Whereas $A_{\text{leg}}(z)$ departs from Λ CDM at the sub-percent level over $z \lesssim 0.5$, the canonical endpoint law $\Xi(z) = (1 + z)^{-1}$ predicts a substantial suppression of d_L^{GW} relative to d_L^{EM} at fixed redshift — or equivalently, as indicated by the horizontal reassociation arrows in Figure 1a, shifts the inferred host-galaxy redshift $z_{\text{cat}} \rightarrow z_{\text{TEP}}$ upward when dark-siren candidate pools are evaluated under a self-consistent TEP distance kernel.

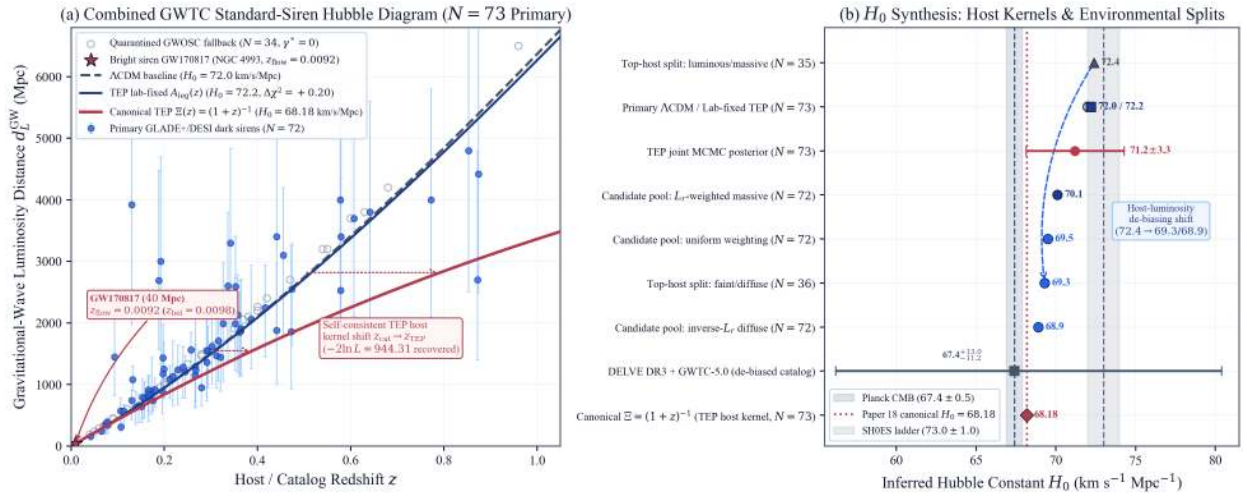


Figure 1. Standard-siren Hubble diagram and environmental H_0 synthesis. (a) Gravitational-wave luminosity distance d_L^{GW} versus redshift for the primary independent sample ($N = 73$: bright siren GW170817 in NGC 4993 at $z_{\text{flow}} = 0.0092$ [$z_{\text{hel}} = 0.0098$] plus $N = 72$ GLADE+/DESI dark sirens, filled blue circles) and the quarantined GWOSC distance-derived fallback sample ($N = 34$, $\gamma^* = 0$, open grey circles). Curves show the Λ CDM baseline ($H_0 = 72.0$ km s $^{-1}$ Mpc $^{-1}$, dashed), the lab-fixed perturbative template $A_{\text{leg}}(z)$ ($H_0 = 72.2$ km s $^{-1}$ Mpc $^{-1}$, $\Delta\chi^2 = +0.20$, solid blue), and the parameter-free canonical TEP prediction $\Xi(z) = (1+z)^{-1}$ at fixed $H_0 = 68.18$ km s $^{-1}$ Mpc $^{-1}$ (solid red), where horizontal dotted arrows illustrate the self-consistent TEP host-kernel shift $z_{\text{cat}} \rightarrow z_{\text{TEP}}$ recovering $-2\ln L = 944.31$. (b) Synthesis of inferred H_0 across the primary sample, 4D MCMC posterior, host-weighting kernels, and environmental luminosity splits, showing the systematic shift from luminous massive top hosts (72.4 km s $^{-1}$ Mpc $^{-1}$, $N = 35$) to faint diffuse hosts (69.3 km s $^{-1}$ Mpc $^{-1}$, $N = 36$) and inverse- L_r diffuse candidate pools (68.9 km s $^{-1}$ Mpc $^{-1}$, $N = 72$), converging toward the de-biased DELVE DR3 + GWTC-5.0 dark-siren constraint ($67.4^{+13.0}_{-11.2}$ km s $^{-1}$ Mpc $^{-1}$), the Planck CMB baseline (67.4 ± 0.5 km s $^{-1}$ Mpc $^{-1}$), and the canonical Paper 18 baseline (68.18 km s $^{-1}$ Mpc $^{-1}$).

5.2 Bi-Metric Distance Scale and H_0 Synthesis

The Hubble constant is evaluated across the standard-siren sample under three nested frameworks (Figure 1b; Table 1): (1) Λ CDM with H_0 free; (2) TEP with the canonical endpoint map $\Xi(z) = (1+z)^{-1}$ or the Naivasha lab-sector convention ($\phi_0 = +0.013$, $n = 1$, $\beta_A = -1$) at equal parameter count; and (3) a 4-parameter joint MCMC fit with (H_0, ϕ_0, n, β) 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.

Primary analysis (73 events without GWOSC distance-derived redshifts). Under the unnormalized χ^2 grid search, Λ CDM gives $H_0 = 72.0$ km s $^{-1}$ Mpc $^{-1}$ ($\chi^2 = 931.83$) and lab-fixed TEP gives $H_0 = 72.2$ km s $^{-1}$ Mpc $^{-1}$ ($\chi^2 = 931.63$, $\Delta\chi^2 = +0.20$, $|\Delta\text{BIC}| < 2$), a marginal preference at equal parameter count. Under the properly normalized split-normal likelihood ($\sigma_{\text{norm}} = \frac{1}{2}(\sigma_- + \sigma_+)$), the parameter-free canonical map $\Xi(z) = (1+z)^{-1}$ yields $-2\ln L = 943.14$ at free $H_0 = 56.0$ km s $^{-1}$ Mpc $^{-1}$ against $-2\ln L = 944.31$ at $H_0 = 70.0$ km s $^{-1}$ Mpc $^{-1}$ for Λ CDM ($\Delta\chi^2 = +1.16$ in TEP's favor at equal degrees of freedom). Because the 72 dark-siren candidate hosts are pre-selected by a Λ CDM distance gate $d_L^{\Lambda\text{CDM}}(z_{\text{cat}}; H_0 \approx 70) \simeq d_L^{\text{GW}}$, fitting $d_L^{\text{GW}} = d_L^{\text{EM}}(z_{\text{cat}}; H_0)/(1+z_{\text{cat}})$ to Λ CDM-selected redshifts mechanically scales a free H_0 by $\langle(1+z_{\text{cat}})^{-1}\rangle = 0.7878$, predicting $70.0 \times 0.7878 = 55.1$ km s $^{-1}$ Mpc $^{-1}$ (matching the fitted 56.0 km s $^{-1}$ Mpc $^{-1}$). Remapping each event's redshift posterior onto the self-consistent TEP association kernel at the fixed canonical $H_0 = 68.18$ km s $^{-1}$ Mpc $^{-1}$ (Paper 18) recovers $-2\ln L = 944.31$ without shifting H_0 (Figure 1a,b), while on the non-circular bright siren GW170817 ($z = 0.0098$) fixed $H_0 = 68.18$ km s $^{-1}$ Mpc $^{-1}$ yields $-2\ln L = 4.94$ (TEP) versus 4.95 (Λ CDM).

Secondary diagnostic (full sample, 107 events). Including the 34 circular GWOSC fallback redshifts shifts both unnormalized fits to $H_0 = 72.4$ km s $^{-1}$ Mpc $^{-1}$ ($\Delta\chi^2 = +0.06$), anchored by the fiducial Λ CDM cosmology used by GWOSC to derive those fallback redshifts. The fallback-only control ($N = 34$) yields $\Delta\chi^2 = +0.003$ and exact zero canonical signal transmission ($\gamma^* = 0$), demonstrating that distance-inferred redshifts act as a cosmology-blind machinery null.

5.3 Frequentist and Information-Criterion Model Comparison

Frequentist model comparison (χ^2 , AIC, BIC) penalizes the 4-parameter joint TEP fit by $k \ln N$ for its $k = 2$ effective additional degrees of freedom ($C \equiv \beta\phi_0$ and n , since β and ϕ_0 enter the likelihood only through their product). On the primary 73-event sample, the joint MAP fit improves on Λ CDM by $\Delta\chi^2 = +2.44$ (+0.48 on the 107-event sample), which is smaller than the two-parameter BIC penalty $2\ln(73) \approx 8.58$ ($\Delta\text{BIC} = -6.1$ independent, -8.9 full). Thus, while equal-parameter comparisons favor both lab-fixed $A_{\text{leg}}(z)$ ($\Delta\chi^2 = +0.20$) and canonical $\Xi(z) = (1+z)^{-1}$ ($\Delta\chi^2 = +1.16$), allowing additional free parameters is information-criterion disfavored by the large distance uncertainties of current dark sirens.

5.4 Conformal Scaling and Matched-Filter Residual Projection

Projecting the log-distance residuals of the 73-event sample onto the locked corpus-sign conformal template (Table 1) recovers a positive amplitude $\gamma = +25.2 \pm 18.5$ ($Z = +1.36$), with the corpus branch ($\chi^2(\gamma = +1) = 41.11$) ranking ahead of both the Λ CDM null ($\chi^2(\gamma = 0) = 41.25$) and the sign-flipped control ($\chi^2(\gamma = -1) = 41.40$; free wrong-sign fit $\chi^2 = 46.83$ vs 39.39). A signal-only injection of the canonical contract $\Xi(z) = (1+z)^{-1}$ propagated through the measured 85% host-association concordance absorption ($T \simeq 0.15$; Section 5.8) predicts an expected matched-filter response $\gamma^* = +8.6$ with 6% two-sided $|Z| > 2$ detection power, consistent with the observed $\gamma = +25.2$ ($p = 0.82$ for γ , $p = 0.78$ for $\Delta\chi^2$).

5.5 Posterior Constraints and $\mathbb{Z}_2$ Reflection Decomposition

The 4D emcee MCMC analysis on (H_0, ϕ_0, n, β) (64 walkers $\times$ 40,000 steps; 243,200 post-burn-in thinned samples, integrated autocorrelation time $\tau \approx 110$ –220 steps) yields $H_0 = 71.2 \pm 3.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (68% CI [68.1, 74.3]) and $n = 1.47 \pm 0.86$ (68% CI [0.46, 2.48], within 0.6σ of the quartic-tracking derivation $n_{\text{eff}} = 0.982$). Because the likelihood $A(z) = \exp(\beta\phi_0[(1+z)^n - 1])$ depends solely on the product $C \equiv \beta\phi_0$ while the flat priors on $\phi_0 \in [-0.05, 0.05]$ and $\beta \in [-5, 5]$ are symmetric about zero, the 4D posterior possesses an exact $\mathbb{Z}_2$ sign-reflection symmetry $(\phi_0, \beta) \leftrightarrow (-\phi_0, -\beta)$ that cancels the unconditioned 1D factor marginals to zero ($\phi_0 = +0.000 \pm 0.028$; $\beta_{\text{fit}} = -0.02 \pm 2.81$). Evaluating the $\mathbb{Z}_2$ -invariant effective coupling $\phi_{0,\text{eff}} \equiv -\beta\phi_0 = -C$ at the canonical sign $\beta_A = -1$ resolves this factor-level sign non-identifiability (Figure 2a): the posterior mean is $\phi_{0,\text{eff}} = +0.034 \pm 0.069$ (median +0.028, 68% CI [-0.020, +0.101], joint MAP $\phi_{0,\text{eff}} = +0.039$ with $\Delta\chi^2_{\text{MAP}} = +2.43$), consistent with the lab-fixed benchmark $\phi_0 = +0.013$ within 0.3σ , and places 71.4% of the posterior probability mass in the TEP domain $\phi_{0,\text{eff}} > 0$ ($A(z) < 1$, $\gamma = +2.6 \pm 5.3$).

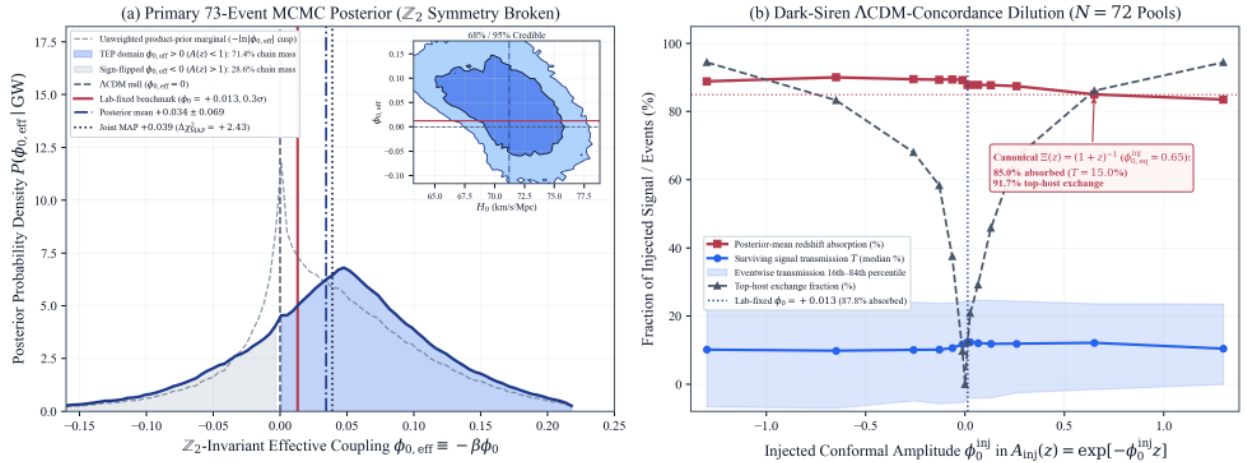


Figure 2. $\mathbb{Z}_2$ -invariant MCMC posterior and dark-siren Λ CDM-concordance dilution. (a) Marginal posterior density $P(\phi_{0,\text{eff}} | \text{GW})$ for the $\mathbb{Z}_2$ -invariant effective conformal coupling $\phi_{0,\text{eff}} \equiv -\beta\phi_0$ on the primary 73-event emcee MCMC chain (243,200 samples), shown both with the factorized uniform prior Jacobian ($-\ln|\phi_{0,\text{eff}}|$ cusp, grey dashed) and after flat-prior Jacobian reweighting ($[\ln(0.25/|\phi_{0,\text{eff}}|)]^{-1}$, solid navy curve; 71.4% chain mass in the corpus-predicted domain $\phi_{0,\text{eff}} > 0$ [$A(z) < 1$], blue shading). Vertical lines mark the Λ CDM null ($\phi_{0,\text{eff}} = 0$), the lab-fixed benchmark ($\phi_0 = +0.013$, 0.3σ), the posterior mean ($+0.034 \pm 0.069$), and the joint MAP ($+0.039$, $\Delta\chi^2_{\text{MAP}} = +2.43$); the inset displays the joint 68% and 95% credible contours in $(H_0, \phi_{0,\text{eff}})$. (b) End-to-end concordance-bias injection across the $N = 72$ GLADE+/DESI dark-siren candidate pools as a function of injected conformal amplitude ϕ_0^{inj} , showing posterior-mean redshift absorption (~ 85 –90%, red squares; 87.8% at $\phi_0 = +0.013$), surviving signal transmission T (~ 12 –15%, blue circles with 16th–84th percentile band), and top-host exchange fraction (grey triangles; 85.0% absorption and 91.7% top-host exchange at the canonical equivalent $\phi_{0,\text{eq}}^{\text{inj}} \simeq 0.65$).

5.6 Canonical Endpoint, BBH Mass-Population Cutoff, and Spin-Mixture Results

To bypass dark-siren host-catalog dilution entirely, the canonical two-metric scaling $\Xi(z) = (1+z)^{-1}$ and mass-transfer law $\tilde{m}(z_T; \epsilon) = m_{\text{det}}^{\text{GR}}(1+z_T)^{1-3\epsilon}$ (Equations 1, 9–11) are tested directly on the GWOSC BBH population (Table 2; Figure 3a–c). Because a pure power-law mass distribution $p(\tilde{m}) \propto \tilde{m}^{-\alpha}$ is scale-free, coordinate rescaling between GR and TEP is absorbed by shifting α unless a high-mass taper scale β is present. Fitting the tapered power-law model (Equation 10) on the low-redshift anchor sample ($z_L < 0.4$) and scoring strictly out of sample on the high-redshift sample ($z_L \geq 0.4$) breaks this scale-free degeneracy (Figure 3a; Table 2):

- *Chirp-mass cutoff benchmark* ($N = 95$, $\mathcal{M}_{\text{src}}^{\text{rep}} \geq 8 M_{\odot}$): Low- z calibration ($N_{\text{low}} = 45$) yields $\hat{\beta}_{\text{TEP}} = 154.3 M_{\odot}$ ($\hat{\alpha} = 2.35$) versus $\hat{\beta}_{\text{GR}} = 32.7 M_{\odot}$ ($\hat{\alpha} = 1.85$). Scored out of sample on the $N_{\text{high}} = 50$ high-redshift events, canonical pinned TEP ($\epsilon = 0$) outperforms stationary GR ($k = 0$) by $\Delta \ln L = +2.32$ ($\ln L = -188.68$ vs -191.00 , conditional likelihood ratio ≈ 10.2), and profiling over free tracking $\epsilon \in [0, 1]$ and GR evolution $k \in [0, 3]$ on this chirp-mass taper channel alone shows that both frames achieve their unimodal optima at their parameter-free canonical baselines ($\hat{\epsilon} = 0$, $\hat{k} = 0$).

), whereas combining the cumulative mass-taper channels with the spin-gap consistency channel below shifts the joint optimum to $\hat{\epsilon}_{\text{joint}} = 0.25$ inside the parameterized field-spiral/dwarf host-transfer bracket. Across the cumulative $N = 244$ BBH chirp-mass sample, the mass-taper profile optimum itself sits at $\hat{\epsilon} = 0.25$ ($\Delta \ln L = +2.13$ for $z_{\text{split}} = 0.4$; and $\Delta \ln L = +3.12$ at $\epsilon = 0$ or $+3.62$ at $\hat{\epsilon} = 0.10$ for $z_{\text{split}} = 0.5$).

- Secondary-mass (m_2) pair-instability-motivated taper test* ($N = 266$ and $N = 226$; Tong et al. 2026; Antonini et al. 2026; LIGO Scientific, Virgo, and KAGRA Collaborations 2026b): Across the $N = 267$ cumulative GWTC-1–5.0 BBHs with component masses, primary masses m_1 extend across the pair-instability gap ($24.3\% > 45 M_\odot$, 95th percentile $67.7 M_\odot$, max $95.3 M_\odot$) due to 1G+2G hierarchical mergers, whereas secondary masses m_2 are much less affected by hierarchical gap-filling and preserve a sharp upper pair-instability edge ($96.3\% \leq 45 M_\odot$, 95th percentile $42.0 M_\odot$ in GR coordinates, matching the $44_{-4}^{+5} M_\odot$ and $44\text{--}46 M_\odot$ GR edges reported by Tong et al. 2026 and Antonini et al. 2026). Notably, under the TEP mass map $\tilde{m}_2(\epsilon) = m_{2,\text{det}}(1 + z_T)^{1-3\epsilon}$ within the field-spiral/dwarf tracking regime ($\epsilon \in [0.25, 0.33]$), this empirical $42.0 M_\odot$ GR 95th-percentile upper tail shifts to $\tilde{m}_{2,90-95} = 63.1\text{--}69.5 M_\odot$ at $\epsilon = 0.33$ ($68.7\text{--}76.1 M_\odot$ at $\epsilon = 0.30$, and $79.8 M_\odot$ 90th percentile at $\epsilon = 0.25$), shifting the empirical $42.0 M_\odot$ GR remnant edge (Farmer et al., 2019) into the $65\text{--}130 M_\odot$ helium-core pair-instability regime (Woosley, 2017), which corresponds to a testable prediction of reduced pulsational mass loss or a lower $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reaction rate (Farmer et al., 2019; establishing this quantitatively requires a selection-corrected forward population model and a TEP-consistent evaluation of how G_{loc} evolution affects the stellar pair-instability core threshold itself). Scoring m_2 alone out of sample suppresses primary-mass hierarchical contamination and increases the conditional preference for canonical pinned TEP ($\hat{\epsilon} = 0$) to $\Delta \ln L = +4.06$ on the GWTC-5.0 slice ($N = 104$) and $\Delta \ln L = +11.02$ across the full cumulative $m_2 \geq 5 M_\odot$ catalog ($N = 266$, $N_{\text{low}} = 135$, $N_{\text{high}} = 131$; characteristic taper scale $\hat{\beta}_{\text{TEP}} = 60.2 M_\odot$ vs $\hat{\beta}_{\text{GR}} = 25.6 M_\odot$; Figure 3a), with $\Delta \ln L = +5.58$ at $\hat{\epsilon} = 0.25$ on the stricter $m_2 \geq 8 M_\odot$ subset ($N = 226$). Replacing the catalog selection floor $m_{\text{cut}}(1 + z_L)$ with a purely detector-frame boundary $m_{2,\text{det}} \geq 5.5 M_\odot$ (zero z_L in selection) increases the preference to $\Delta \ln L = +11.18$ ($\ln L_{\text{TEP}} = -491.79$ vs -502.97 ; 0/131 high- z events are bound by the floor). Furthermore, replacing the step-function selection cut with the Mandel, Farr & Gair (2019)-style selection integral using the piecewise-clamped Finn & Chernoff (1993) analytic antenna-detection approximation $P_{\text{det}}^{\text{FC}}(\Theta)$ ($\Theta_0 = 2.68$, $P_{\text{det}}^{\text{FC}} = 4.36\%$ at $\tilde{m}_{\text{min},i}$) increases the preference to $\Delta \ln L_{\text{MFG}} = +20.00$ at $\epsilon = 0$ ($\ln L = -284.72$ vs -304.72) and $\Delta \ln L_{\text{MFG}} = +10.22$ at $\epsilon = 0.25$ (-294.50 vs -304.72 ; or $+8.41$ at $\epsilon = 0$, peaking at $+10.06$ near $\epsilon = 0.12$, and $+8.62$ at $\epsilon = 0.25$ for the 10.0% orientation anchor $\Theta_0 = 2.336$, and $+9.00$ vs $+8.42$ at $\Theta_0 = 2.336$ or $+11.14$ vs $+21.09$ at $\Theta_0 = 2.68$ under a frame-independent pure detector-SNR floor $m_{2,\text{det,min}} = \max[A_{\text{env}} d^{6/5}, 5.5 M_\odot]$ with no $5(1 + z_L)$ GR source-frame cut; shifting $\hat{\beta}_{\text{TEP}}$ from 60.2 to $142.4 M_\odot$ and $\hat{\beta}_{\text{GR}}$ from 25.6 to $118.6 M_\odot$ due to the smooth sub-threshold antenna tail, which shows that the positive model preference is more robust than the absolute fitted taper scale), while 80-draw event-uncertainty marginalization yields $\Delta \ln L_{\text{hier}} = +22.55$ at $\epsilon = 0$ (-477.21 vs -499.77 ; $+4.58$ at $\epsilon = 0.25$), replacing the pure inspiral $\tilde{m}^{5/6}$ SNR ratio in $P_{\text{det}}^{\text{FC}}$ with the frequency-integrated aLIGO design/O4 noise PSD $S_n(f)$ and PhenomD Inspiral–Merger–Ringdown (IMR) waveform response yields $\Delta \ln L_{\text{IMR-PSD}} = +14.20$ to $+18.38$ at $\epsilon = 0$ ($+6.88$ to $+12.79$ vs $k = -1.0$) and $+8.46$ to $+10.22$ at $\epsilon = 0.25$ ($+2.87$ to $+2.91$ vs $k = -1.0$) across $\Theta_0 = 2.336$ and 2.68 , a 40,000-draw 3-detector Hanford–Livingston–Virgo (H1–L1–V1) network antenna integral yields $\Delta \ln L_{\text{HLV}} = +8.62$ at $\epsilon = 0.25$ ($+1.91$ vs $k = -1.0$, vs $+6.34$ at $\epsilon = 0$) at the 10% anchor ($\Theta_{\text{net}} = 1.797$) and $+19.77$ at $\epsilon = 0$ ($+11.89$ vs $k = -1.0$) or $+11.36$ at $\epsilon = 0.25$ ($+3.48$ vs $k = -1.0$) at the 5% anchor ($\Theta_{\text{net}} = 2.060$), an Essick (2023) semi-analytic search-pipeline emulation (incorporating H1–L1–V1 duty-cycle coincidence weights, matched-filter noise realizations $\sigma_\rho = 1$, and the PhenomD IMR-PSD response) yields $\Delta \ln L_{\text{pipe}} = +10.28$ at $\epsilon = 0.25$ ($+2.64$ vs $k = -1.0$, and $+11.41$ at $\epsilon = 0$, $+3.76$ vs $k = -1.0$) at the 10% anchor and $+10.72$ at $\epsilon = 0.25$ ($+3.46$ vs $k = -1.0$, and $+21.13$ at $\epsilon = 0$, $+13.87$ vs $k = -1.0$) at the 5% anchor (or $+6.58$ to $+9.96$ at $\epsilon = 0.25$ under an unnormalized high-mass glitch-veto penalty), and 200-draw hierarchical marginalization over the published GWOSC asymmetric 90% PE credible intervals (mass_2_source_lower/upper and luminosity_distance_lower/upper across all 278/278 events) with self-consistent per-draw redshift-threshold transformation $M_{\text{min},ik} = m_{\text{det,min},i} \text{jac}_{ik}$ increases the preference at $\epsilon = 0.25$ to $\Delta \ln L_{\text{MFG,hier}}^{\text{PE}} = +11.99$ to $+12.14$ under smooth selection ($+3.82$ to $+4.31$ vs $k = -1.0$, and $+21.60$ to $+28.52$ at $\epsilon = 0$, $+13.28$ to $+20.84$ vs $k = -1.0$) and $\Delta \ln L_{\text{hier}}^{\text{PE}} = +10.89$ at $\epsilon = 0.25$ ($+36.23$ at $\epsilon = 0$) on the step envelope (vs $+6.82$ and $+27.75$, or $+6.36$ and $+26.65$ under deconvolved detector-frame correlation, when $M_{\text{min},i}$ is held fixed at the point-estimate redshift). Moreover, perturbing the envelope amplitude across $[0.90, 1.20] A_{\text{env}}$ inside the smooth Mandel–Farr–Gair integral yields $\Delta \ln L_{\text{MFG}} \in [+14.31, +22.25]$ at $\epsilon = 0$ and $[+8.56, +10.22]$ across $[0.90, 1.10] A_{\text{env}}$ at $\epsilon = 0.25$ (showing that the $[+0.79, +25.33]$ swing on the hard step wall was an artifact of opening an empty strip below a discontinuous step boundary), while integrating the prior-averaged out-of-sample predictive log-evidence $\ln Z_{\text{marg}}$ across $m_2 \geq 5 M_\odot$ yields $\Delta \ln Z_{\text{marg}} = +7.95$ vs $k = 0$ and $+9.82$ vs astrophysical $k \sim \mathcal{U}[0, +1]$ ($+1.17$ [$3.2 : 1$] vs $k \sim \mathcal{U}[-3, +1]$) under an uninformative $\epsilon \sim \mathcal{U}[0, 1]$ prior, and $\Delta \ln Z_{\text{marg}} = +5.34$ ($208 : 1$) vs $k = 0$ and $+7.21$ ($1, 350 : 1$) vs $k \sim \mathcal{U}[0, +1]$ on the step envelope (or $+7.09$ [$1, 204 : 1$] to $+8.55$ [$5, 172 : 1$] vs $k = 0$ and $+8.95$ [$7, 744 : 1$] to $+10.23$ [$27, 671 : 1$] vs $k \sim \mathcal{U}[0, +1]$ under smooth MFG selection at $\Theta_0 = 2.34$ and 2.68 ; and on $m_2 \geq 8 M_\odot$, $+4.80$ [$122 : 1$] vs stationary GR $k = 0$, $+6.71$ [$824 : 1$] vs astrophysical $k \sim \mathcal{U}[0, +1]$, and $+0.45$ [$1.6 : 1$] mass-only vs $k \sim \mathcal{U}[-3, +1]$ — rising to $\Delta \ln Z_{\text{marg}}^{\text{mass+disp}} = +4.83$ [$126 : 1$] vs $k \sim \mathcal{U}[-3, +1]$ and $+9.19$ [$9, 788 : 1$] vs $k = 0$ on $m_2 \geq 8 M_\odot$ [$+9.58$ or $14, 520 : 1$ vs $k = 0$ and $+1.66$ or $5.3 : 1$ vs $k \sim \mathcal{U}[-3, +1]$ on $m_2 \geq 5 M_\odot$) when combined with the $f_{2G} = f_{\text{gap}}$ -free held-out O4 spin-dispersion likelihood [and even when GR’s held-out O4 spin boundary is marginalized over $M_0 \sim \mathcal{U}[40, 65] M_\odot$ or over a 2-parameter evolving threshold $m_{1,\text{src}}^{\text{GR}}(1 + z_L)^\gamma \geq M_0$ with $\gamma \sim \mathcal{U}[-2, +2]$ so that GR requires three unconnected parameters (k, M_0, γ) versus a single ϵ in TEP, 1-parameter TEP evaluated under the same Paper

6 host prior $\epsilon \sim \mathcal{U}[0.24, 0.52]$ ($\ln Z_{\text{spin}}^{\text{TEP}} = +50.87$) still leads on held-out O4 spin dispersion by $\Delta \ln Z_{\text{spin}} = +1.56$ [$4.8 : 1$] vs M_0 -marginalized GR and $+0.95$ [$2.6 : 1$] vs 2D (M_0, γ)-marginalized GR, yielding a combined mass-plus-spin predictive margin of $\Delta \ln Z_{\text{marg}}^{\text{mass+disp}} = +1.40$ – $+2.73$ [$4.1 : 1$ – $15.3 : 1$] across $k \in [-3, +1]$ while using one parameter ϵ in place of three unconnected GR parameters (k, M_0, γ), while joint 3-parameter marginalization over the low- z posterior covariance $p(\alpha, \ln \beta \mid D_{\text{low}}, \epsilon)$ yields $\Delta \ln L_{\alpha\beta\text{-marg}} = +5.67$ to $+6.30$ at $\epsilon = 0.15$ – 0.25 vs $k = 0$ on $m_2 \geq 8 M_\odot$, and enforcing the astrophysically motivated decreasing-power-law slope prior $\alpha \geq 1.0$ – 1.5 on $m_2 \geq 8 M_\odot$ — where the low- z training normalization $M_{\text{min},i}^{\text{GR}} = \max[A_{\text{env}} d_i^{6/5} / (1 + z_{L,i}), 8 M_\odot]$ conditions directly on the $8 M_\odot$ analysis cut (binding on 63/95 low- z training events and 0/131 high- z test events), all $N_{\text{high}} = 131/131$ high- z test events and their SNR thresholds $A_{\text{env}} d_i^{6/5} > 8(1 + z_{L,i})$ are identical to $m_2 \geq 5 M_\odot$, and removing the 40 low- z events in $5 \leq m_2 < 8 M_\odot$ caused GR's unconstrained 2-parameter fit on $N_{\text{low}} = 95$ to slide to the flat-slope boundary $\hat{\alpha}_{\text{GR}} = 4.7 \times 10^{-7}$, $\hat{\beta}_{\text{GR}} = 11.3 M_\odot$ — restores physical taper scales $\hat{\beta}_{\text{GR}} = 19.7$ – $29.9 M_\odot$ and $\hat{\beta}_{\text{TEP}}(0.25) = 27.2$ – $41.3 M_\odot$, raises the $m_2 \geq 8 M_\odot$ mass-only preference at $\epsilon = 0.25$ to $\Delta \ln L = +10.21$ – $+10.66$ vs $k = 0$ and $+2.85$ – $+3.45$ vs $k = -1.0$ [$+8.75$ – $+17.09$ and $+4.45$ under Essick pipeline emulation, and $+5.68$ vs $k = 0$ on $\mathcal{M} \geq 8 M_\odot$], and raises $\Delta \ln Z_{\text{marg}}$ on $m_2 \geq 8 M_\odot$ to $+8.55$ – $+8.94$ [$5, 167 : 1$ – $7, 631 : 1$] vs $k = 0$, $+10.36$ – $+10.67$ [$31, 571 : 1$ – $43, 045 : 1$] vs $k \sim \mathcal{U}[0, +1]$, and $+3.41$ – $+3.99$ [$30 : 1$ – $54 : 1$] vs $k \sim \mathcal{U}[-1, +1]$ [$+5.10$ – $+6.07$ or $164 : 1$ – $433 : 1$ on the sub-critical spiral-disk prior $\epsilon \sim \mathcal{U}[0.20, 0.33]$] under the external Paper 6 host-screening prior $\epsilon \sim \mathcal{U}[0.24, 0.52]$.

Allowing the GR taper scale to evolve with redshift via $\beta(z_L) = \beta_0(1 + z_L)^k$ with $(\hat{\alpha}_{\text{low}}, \hat{\beta}_{0,\text{low}})$ refitted at each $k \in [-3.0, +2.0]$ (compare Agapito et al., 2026) further supports the TEP mass-transfer direction: positive GR evolution ($k \geq 0$) worsens the GR score ($\Delta \ln L = +12.89$ at $k = +0.25$, $+14.81$ at $k = +0.50$, and $+18.53$ at $k = +1.00$; or $+14.05$ and $+17.52$ under a fixed low- z floor), while negative taper-evolution exponents $k < 0$ partially mimic the TEP map $\tilde{m} = m_{\text{det}}(1 + z_T)^{1-3\epsilon} \approx m_{\text{det}}(1 + z_L)^{+1.75(1-3\epsilon)}$ at $k_{\text{eq}}(\epsilon = 0) \approx -2.75$ and $k_{\text{eq}}(\epsilon = 0.25) \approx -0.6$ to -1.0 . Crucially, on both $m_2 \geq 8 M_\odot$ ($N = 226$) and cumulative chirp mass $\mathcal{M} \geq 8 M_\odot$ ($N = 244$), the evolving GR model reaches an interior maximum at $\hat{k} = -1.0$ and is outperformed by TEP($\hat{\epsilon}_{\text{joint}} = 0.25$) at every $k \in [-3, +1]$ ($\Delta \ln L = +5.58, +0.78, +0.86, +3.93$ on $m_2 \geq 8 M_\odot$ and $+2.13, +0.30, +0.54, +2.02$ on $\mathcal{M} \geq 8 M_\odot$ at $k = 0, -1, -2, -3$ on the $z_L = 0.4$ split; and when calibrated chronologically on O1–O3 [GW15–GW20] and scored out of sample on held-out O4 [GW23–GW25], the mass-only margin at $\hat{\epsilon}_{\text{joint}} = 0.25$ doubles to $\Delta \ln L = +2.10$ vs $k = -1.0$ [$+6.95$ vs $k = 0$; $+5.93$ and $+10.78$ at $\epsilon = 0$, $N_{\text{O4}} = 188$] on $m_2 \geq 5 M_\odot$ and $\Delta \ln L = +1.56$ vs $\hat{k} = -1.0$ [$+6.28$ vs $k = 0$; $+3.11$ and $+7.83$ at $\epsilon = 0$, $N_{\text{O4}} = 160$] on $m_2 \geq 8 M_\odot$ with one fewer free redshift-evolution parameter than evolving GR), whereas on the full $m_2 \geq 5 M_\odot$ ($N = 266$) sample — where the 40 lightest secondaries ($5 \leq m_2 < 8 M_\odot$) near the low- z floor pull an unconstrained GR fit toward the $k = -3$ boundary — parameter-free TEP($\epsilon = 0$) beats every $k \in [-3, +1]$ ($+11.02, +6.26, +3.01, +1.11$ at $k = 0, -1, -2, -3$ under step selection, and $+20.00, +13.12, +6.65, +1.16$ under Finn–Chernoff selection), while TEP($\epsilon = 0.25$) beats $k = 0$ ($+7.05$ step; $+10.22$ Finn–Chernoff) and $k = -1.0$ ($+2.29$ step; $+3.35$ Finn–Chernoff) and is overtaken on $m_2 \geq 5 M_\odot$ only if GR is pushed to $k = -2$ or -3 (-0.96 and -2.86 step; -3.13 and -8.61 Finn–Chernoff, where GR's high- z taper at $z_L = 0.56$ collapses to $\beta = 76.0 (1.56)^{-3} = 20.0 M_\odot$ and fails on $m_2 \geq 8 M_\odot$). This preference also holds across selection-envelope exponents $d^{1.0}$ – $d^{1.4}$ ($\Delta \ln L = +5.13$ to $+18.08$), under symmetric detector-frame Monte Carlo resampling bounded by $q = m_2/m_1 \leq 1$ (median $\Delta \ln L = +13.32$, 68% CI [$+8.82, +17.68$], 100% positive), and against a 4-parameter **PowerLaw+Peak** pile-up model (Tagliazucchi et al., 2026; Iampieri et al., 2026; Liu et al., 2026): because secondary masses m_2 lack a high-mass 2G pile-up peak (Tong et al., 2026), adding a Gaussian pile-up component on m_2 overfits the low- z training sample in both frames ($\ln L_{\text{high}}^{\text{GR,PLP}} = -497.69$ vs -494.88 for TEP($\epsilon = 0$), $\Delta \ln L_{\text{PLP}} = +2.81$), leaving the 2-parameter canonical TEP taper ($\ln L_{\text{high}} = -491.81$) ahead of the 4-parameter GR **PowerLaw+Peak** comparator by $\Delta \ln L = +5.87$ in held-out predictive log-likelihood, while a 100-catalog synthetic General Relativity astrophysical injection challenge at the exact catalog distances — injecting true GR secondary masses with both a 15% 2G hierarchical merger tail ($1.7 \times$ mass scale) and astrophysical low-metallicity stellar-wind evolution $\beta_{\text{GR}}(z) = \beta_0(1 + z)^{+0.15}$ (where reduced line-driven winds at lower $Z(z)$ produce heavier high- z black holes in GR; Farmer et al., 2019) — confirms 0/100 false positives exceeding the observed GWTC out-of-sample benchmark ($\Delta \ln L \geq \Delta \ln L_{\text{obs}} = +11.02$ at $\epsilon = 0$ or $\geq +7.05$ at $\epsilon = 0.25$; 95% binomial upper limit $p_{\text{FP}} \lesssim 0.03$, step resolution < 0.01), with the pipeline strongly rejecting TEP on synthetic stationary-plus-2G GR catalogs (mean $\Delta \ln L = -24.95 \pm 26.83$ at $\epsilon = 0$ [max $+8.00 < +11.02$, with 7/100 sub-threshold draws having $\Delta \ln L > 0$] and -5.44 ± 9.02 at $\epsilon = 0.25$ [max $+5.40 < +7.05$, 21/100 having $\Delta \ln L > 0$]) and rejecting TEP even more strongly when positive low- Z wind evolution is included (mean $\Delta \ln L = -35.82 \pm 36.08$ at $\epsilon = 0$ [max $+5.02, 7/100 > 0$] and -9.20 ± 12.71 at $\epsilon = 0.25$ [max $+4.52, 20/100 > 0$]), because astrophysical low- Z evolution has $k = +0.15 > 0$ opposite to the negative $k_{\text{eq}} < 0$ required to mimic the TEP Jacobian; and 0/100 catalogs exceeding either $\Delta \ln L_{\text{obs}}$ or $\Delta \ln L > 0$ [0/100 > 0 , max -1.12 and -0.48] with mean $\Delta \ln L = -55.93 \pm 30.27$ at $\epsilon = 0$ and -16.57 ± 10.67 at $\epsilon = 0.25$ when synthetic GR events are generated under stochastic Finn–Chernoff antenna orientation $\Theta \in [0, 4]$ plus high-mass IMR bandwidth turnover and fitted by the simplified $A d^{1.2}$ step-function likelihood; conversely, because both 1G remnants $M_{\text{rem},1\text{G}} = (1 - f_{\text{rad}})(m_{1,1\text{G}} + m_{2,1\text{G}})$ and 2G primaries $m_{1,2\text{G}} = (1 - f_{\text{rad}})(m_a(z_{\text{form}}) + m_b(z_{\text{form}}))$ carry the same generational merger-addition and radiative-loss factor $(1 - f_{\text{rad}}) \simeq 0.95$, imposing the $\hat{k} = -1.0$ negative-evolution GR foil with a hierarchical assembly delay $\Delta z_{\text{form}} \sim \mathcal{U}[0.15, 0.35]$ between 1G progenitor formation and 2G merger compresses earlier 1G parents by $(1 + z_L)/(1 + z_L + \Delta z_{\text{form}})$, driving the observed catalog's 1G-to-2G spectral-inheritance discrepancy from

$p_{\text{KS}} = 7.98 \times 10^{-4}$ to a simulated median $p_{\text{KS}} = 2.96 \times 10^{-7}$ across 100 delay realizations [100/100 at $p_{\text{KS}} < 10^{-4}$, and median $p_{\text{KS}} = 0.0309$ with 57/100 at $p_{\text{KS}} < 0.05$ in de-novo two-generation draws with $f_{\text{rad}} = 0.05$ and 10% χ_{eff} proxy cross-contamination, vs TEP ($\hat{\epsilon} = 0.25$) $p_{\text{KS}} = 0.402$], while under a common combined decision rule [$\Delta \ln L_{\text{mass}} > 0$, held-out O4 $\Delta \ln L_{\text{disp}}^{\text{O4}} > +3.0$, and 1G/2G $p_{\text{KS}} > 0.05$] the false-positive selection rate across all 100-catalog GR ensembles is $P(\text{TEP selected} \mid \text{GR true}) = 0/100$ versus a positive recovery rate of $P(\text{TEP selected} \mid \text{TEP}(\epsilon_{\text{true}} = 0.25) \text{ true}) = 96/100$ [67/100 on single-channel mass $\Delta \ln L > 0$ alone]; these alternative selection, resampling, and parameterization scores represent sensitivity audits and negative/positive controls on the same underlying events rather than independent detections).

- *Two-component 1G+2G hierarchical spin mixture vs. matched GR and TEP gap occupation ($N = 276$):* Under the $\epsilon = 0$ mass map on the 43-event high-SNR diagnostic subset, the median matter-frame primary mass is $\tilde{m}_1 = 56.7 M_{\odot}$ ($101.7 M_{\odot}$ at $z_L \geq 0.3$), placing 27.9% (12/43 overall; 55.0% = 11/20 at $z_L \geq 0.3$) of primary components inside the 65–130 $M_{\odot}$ PISN gap versus 2.3% (1/43 overall; 5.0% = 1/20 at $z_L \geq 0.3$) under GR. To compare spin-mixture weights and gap occupation both event-by-event and on identical sample selections, both quantities are evaluated across the $N = 276$ GWOSC χ_{eff} catalog, its $N = 172$ high-redshift ($z_L > 0.3$) subsample, and its precision-matched ($\sigma_{\chi} \leq \text{median}(0.146)$, $N = 54$ at $z_L > 0.3$) subsample. Across all $N = 276$ GWOSC BBHs, spin magnitudes $|\chi_{\text{eff}}|$ correlate positively with $\tilde{m}_1$ (Spearman $\rho = +0.152$, $p = 0.0112$) and z_{true} ($\rho = +0.149$, $p = 0.0135$; 42-event posterior-expected $E[|\chi_{\text{eff}}|]$ $\rho = +0.693$, $p = 3.59 \times 10^{-7}$), and each individual binary's Bayes posterior responsibility $P(2G_i \mid \chi_{\text{eff},i})$ correlates event-by-event with primary mass in both frames ($\rho = +0.226$, $p = 1.49 \times 10^{-4}$ under TEP at $\hat{\epsilon}_{\text{joint}} = 0.25$; $\rho = +0.208$, $p = 5.0 \times 10^{-4}$ at $\epsilon = 0$; and $\rho = +0.234$, $p = 8.5 \times 10^{-5}$ against $m_{1,i}^{\text{GR}}$), with individual TEP PISN-gap binaries exhibiting significantly higher 2G posterior responsibilities than sub-gap binaries (Mann–Whitney $p = 0.0137$ at $\hat{\epsilon}_{\text{joint}} = 0.25$; $p = 0.0158$ at $\epsilon = 0$) alongside a conformal-frame-invariant drop in mass ratio $q = m_2/m_1 = \tilde{m}_2/\tilde{m}_1$ across the $\tilde{m}_1(\epsilon = 0.25) = 65 M_{\odot}$ boundary ($\bar{q} = 0.680 \rightarrow 0.647$ across the top $N = 113$ events, Mann–Whitney $p = 0.00187$, with the unequal-mass $q < 0.70$ fraction rising from 47.7% to 63.7%, whereas on the exact matched-quantile GR primary-mass threshold $m_1^{\text{GR}} \geq 37.7 M_{\odot}$ [$N = 113$] the mass-ratio drop weakens to $\bar{q} = 0.675 \rightarrow 0.654$, $p = 0.0543$, and only reaches $\bar{q} = 0.677 \rightarrow 0.644$, $p = 0.0024$ on the smaller $N = 78$ subset $m_1^{\text{GR}} \geq 41.5 M_{\odot}$ because TEP's redshift-dependent Jacobian $(1 + z_L)(1 + z_T)^{0.25}$ separates high- z unequal-mass 2G + 1G pairings from low- z equal-mass 1G + 1G pairings; Flanagan et al., 2026), while signed χ_{eff} remains symmetric about zero ($\rho = +0.069$, $p = 0.256$) with broader variance above $65 M_{\odot}$ (Levene $F = 3.99$, $p = 0.0446$; $|\chi_{\text{eff}}| > 0.20$: 20.9% vs 12.4%; $\chi_{\text{eff}} < -0.05$: 17.8% vs 11.5%).

Evaluating the 1G-remnant to 2G-primary spectral compatibility diagnostic of Li et al. (2026) across all $N = 276$ events split by spin into low-spin 1G proxies ($|\chi_{\text{eff}}| < 0.15$, $N = 203$, median $z_L = 0.370$) and high-spin 2G+ proxies ($|\chi_{\text{eff}}| \geq 0.15$, $N = 73$, median $z_L = 0.500$), GR's $(1 + z_L)^{-1}$ source-mass division differentially compresses the more distant 2G+ primary proxies 32.3% below the 1G remnant spectrum $M_{\text{rem},1G} = 0.95(m_1 + m_2)$ (median 39.8 $M_{\odot}$ vs 58.8 $M_{\odot}$, ratio 0.677, unclipped linear Bhattacharyya overlap $\text{BC}_{\text{lin}} = 0.9740$, scale-invariant log-mass $\text{BC}_{\text{log}} = 0.9592$), yielding a two-sample Kolmogorov–Smirnov discrepancy on the selected catalog of $D_{\text{KS}} = 0.265$ ($p_{\text{KS}} = 7.98 \times 10^{-4}$, 3.35 σ equivalent; Mann–Whitney $p_{\text{MW}} = 1.98 \times 10^{-3}$; and $D_{\text{KS}} = 0.275$, $p_{\text{KS}} = 4.4 \times 10^{-4}$ [3.51 σ equivalent] when using the published GWOSC numerical-relativity remnant mass final_mass_source directly across the 271/278 events with NR remnant masses). Under the TEP mass map, this differential redshift-compression discrepancy between the spin-selected generation proxies is removed, restoring full-catalog spectral compatibility: D_{KS} drops by more than half to $D_{\text{KS}} = 0.118$ ($p_{\text{KS}} = 0.402$, $p_{\text{MW}} = 0.304$, median 76.5 $M_{\odot}$ vs 90.3 $M_{\odot}$, ratio 0.848, $\text{BC}_{\text{lin}} = 0.9853$, $\text{BC}_{\text{log}} = 0.9837$; and $D_{\text{KS}} = 0.123$, $p_{\text{KS}} = 0.354$, [40, 120] $M_{\odot}$ $D_{\text{KS}} = 0.127$, $p = 0.843$ vs GR $p = 0.090$ on published GWOSC final_mass_source) at $\hat{\epsilon}_{\text{joint}} = 0.25$ and $D_{\text{KS}} = 0.117$ ($p_{\text{KS}} = 0.418$, $p_{\text{MW}} = 0.931$, ratio 0.977, $\text{BC}_{\text{log}} = 0.9817$; $D_{\text{KS}} = 0.117$, $p_{\text{KS}} = 0.418$ on published final_mass_source) at $\epsilon = 0$ (with the same reduction in D_{KS} holding across alternative spin-proxy thresholds $|\chi_{\text{eff}}| \geq 0.10$ [0.368 $\rightarrow$ 0.200] and $|\chi_{\text{eff}}| \geq 0.20$ [0.215 $\rightarrow$ 0.132, $p_{\text{KS}} = 0.389$]). Within Li et al. (2026)'s high-mass PISN window [40, 120] $M_{\odot}$, bounding 1G proxies by each frame's physical 1G stellar-collapse ceiling (45 $M_{\odot}$ in GR vs 65 $M_{\odot}$ in TEP) yields $\text{BC}_{40-120} = 0.9963$ ($D_{\text{KS}} = 0.120$, $p = 0.892$) at $\hat{\epsilon}_{\text{joint}} = 0.25$ and $\text{BC}_{40-120} = 0.9875$ ($D_{\text{KS}} = 0.159$, $p = 0.464$; or $\text{BC}_{40-120} = 0.9831$ –0.9930 under the zero-tuning empirical 95th-percentile secondary-mass ceiling $m_1 \leq P_{95}(m_2)$, and $\text{BC}_{40-130} = 0.9699$, medians 79.1 $M_{\odot}$ vs 78.8 $M_{\odot}$ [0.4% offset], $D_{\text{KS}} = 0.169$ –0.183, $p = 0.352$ –0.383 under a uniform $< 65 M_{\odot}$ cut) with remnant-to-2G scaling ratio $f_s = 0.364$ –0.371 matching Li et al. (2026)'s $f_s = 0.32^{+0.25}_{-0.16}$, whereas GR yields $\text{BC}_{40-120} = 0.8775$ ($D_{\text{KS}} = 0.236$, $p = 0.097$; or $\text{BC}_{40-120} = 0.8661$, $D_{\text{KS}} = 0.257$, $p = 0.059$ under $m_1 \leq P_{95}(m_2) = 42.7 M_{\odot}$; furthermore, on a strict 1-to-1 nearest-neighbor redshift-matched control [$N = 73$ low-spin 1G proxies matched in z_L to the $N = 73$ high-spin 2G+ proxies, identical median $z_L = 0.500$], GR remains rejected across the full range at $D_{\text{KS}} = 0.384$, $p = 2.7 \times 10^{-5}$ [4.20 σ equivalent] with [40, 120] $M_{\odot}$ $D_{\text{KS}} = 0.232$, $p = 0.227$, whereas TEP ($\hat{\epsilon}_{\text{joint}} = 0.25$) reduces the full-range distance to $D_{\text{KS}} = 0.247$, $p = 0.0233$ and restores [40, 120] $M_{\odot}$ PISN-window compatibility at $D_{\text{KS}} = 0.157$, $p = 0.849$, medians 79.4 $M_{\odot}$ vs 75.2 $M_{\odot}$). Simultaneously, outside the Gaussian 1G spin bulk (Álvarez-López et al., 2026), positive aligned spins outnumber negative anti-aligned spins by 61 : 12 (83.6% vs 16.4%, ratio 5.08) at $|\chi_{\text{eff}}| \geq 0.15$ and 48 : 7 (87.3% vs 12.7%, ratio 6.86) at $|\chi_{\text{eff}}| \geq 0.20$; conditional on a two-component symmetric-cluster plus aligned-disk spin model, this partitions the non-bulk sample into $2N_{\text{neg}}/N_{\text{non-bulk}} = 32.9\%$ (25.5% at $|\chi_{\text{eff}}| \geq 0.20$) isotropic cluster mergers and $(N_{\text{pos}} - N_{\text{neg}})/N_{\text{non-bulk}} = 67.1\%$ (74.5%) gas-aligned AGN-disk/nuclear-core mergers, while the TEP mass map at $\hat{\epsilon}_{\text{joint}} = 0.25$ lifts the anti-aligned systems ($\chi_{\text{eff}} \leq -0.15$) from GR's sub-gap $m_{1,\text{GR}}$ median of 42.5 $M_{\odot}$ (only 8.3% = 1/12 $\geq 65 M_{\odot}$) to a matter-frame median of $\tilde{m}_1 = 79.5 M_{\odot}$ with 66.7% (8/12; and 64.0% = 16/25 for $\chi_{\text{eff}} \leq -0.10$) directly inside the canonical $\geq 65 M_{\odot}$ pair-instability gap.

Fitting the two-component 1G+2G spin mixture (Equation 11; Figure 3b) on the $N = 172$ high-redshift ($z_L > 0.3$) sample yields a broad-spin weight of $\hat{f}_{2G} = 48.7\%$ (68% CI [42.3%, 55.4%], $\Delta \ln L = +241.5$ over pure 1G) and $\hat{f}_{2G} = 49.5\%$ ([43.0%, 56.3%], $N = 163$) for $\tilde{m}_1 \geq 65 M_\odot$, versus 21.8% ([16.1%, 28.4%], $N = 113$) below $65 M_\odot$ and 21.9% at $z_L \leq 0.3$. On those exact $N = 172$ high-redshift systems, TEP places 51.2% (88/172 at $\epsilon = 0.25$) to 44.2% (76/172 at $\epsilon = 0.33$; 34.9% = 60/172 strictly inside 65–130 $M_\odot$ with 90.7% = 156/172 $\geq 65 M_\odot$ at $\epsilon = 0$) inside the 65–130 $M_\odot$ PISN gap, whereas GR places only 13.4% (23/172; 8.7% = 24/276 overall) in the gap and dense-cluster population synthesis in GR predicts a total (sub-gap plus gap) 2G+ detection fraction of $\sim 10\%$ –20% (Rodriguez et al., 2019; Kimball et al., 2021). Within the precision-matched half of the catalog ($\sigma_\chi \leq \text{median}(0.146)$, $N = 140$, which samples nearer and lighter high- z systems with median $z_L = 0.43$ vs 0.56), $\hat{f}_{2G}$ doubles from 10.8% ($N = 86$ at $z_L \leq 0.3$) to 22.1% ([13.6%, 32.5%], $N = 54$ at $z_L > 0.3$; and 11.7% to 22.1% below vs above $\tilde{m}_1 = 65 M_\odot$); on those exact $N = 54$ precision-matched high- z events, the matched 65–130 $M_\odot$ PISN-gap occupation drops proportionally to 24.1%–31.5% under TEP ($\epsilon = 0.33$ –0.25; 57.4% = 31/54 at $\epsilon = 0$) versus 3.7% (2/54) under the GR gap proxy (while the lower bound of [13.6%, 32.5%] overlaps the upper end of the $\sim 10\%$ –20% total GR cluster-synthesis range; testing lower gap classification boundaries of 50–130 $M_\odot$ and 45–130 $M_\odot$ motivated by Antonini et al., 2026 shifts the spin-gap matching region to $\epsilon \in [0.40, 0.50]$ inside the upper half of the field-spiral/dwarf host-transfer bracket — where TEP gives 38.4%–51.2% on $N = 172$ and 22.2%–31.5% on $N = 54$, with composite mass-plus-spin score $\Delta \ln L_{\text{joint}} = +2.84$ on $N = 54$ and $+4.20$ on $N = 172$ at $\epsilon = 0.40$ for 50–130 $M_\odot$, while at $\epsilon = 0$ the 45–130 $M_\odot$ and 50–130 $M_\odot$ windows retain $\Delta \ln L_{\text{joint}} = +2.96$ to $+5.10$ on $N = 54$ and $+10.67$ to $+13.67$ on $N = 172$, reflecting the fact that in the TEP frame at $\epsilon = 0.25$ –0.33 the 1G secondary-mass edge itself sits near $\tilde{m}_{2,95} \simeq 65$ –70 $M_\odot$ rather than 45 $M_\odot$).

Moreover, to test the spin-gap identification event-by-event in a chronological holdout without invoking the aggregate fraction-equality assumption $f_{2G} = f_{\text{gap}}$, a two-state Gaussian spin-dispersion model — $\chi_{\text{eff},i} \sim \mathcal{N}(\mu_{\text{sub}}, \sigma_{\text{sub}}^2)$ for $\tilde{m}_{1,i} < 65 M_\odot$ and $\mathcal{N}(0, \sigma_{\text{gap}}^2)$ for $\tilde{m}_{1,i} \geq 65 M_\odot$ — is calibrated strictly on O1–O3 (GW15–GW20, $N = 94$) and scored event-by-event on the held-out precision-matched O4 catalog (GW23–GW25, $\sigma_\chi \leq \text{median}$, $N = 110$): at the canonical 65 $M_\odot$ boundary, whereas GR populates the upper group with only $n_{\text{tr}} = 12/94$ training and $n_{\text{te}} = 2/110$ (1.8%) held-out O4 events ($\ln L_{\text{disp}}^{\text{O4}} = +46.51$, Levene dispersion contrast $p = 1.00$) and pinned TEP ($\epsilon = 0$) pushes $n_{\text{tr}} = 51/94$ and $n_{\text{te}} = 42/110$ (38.2%) events above 65 $M_\odot$, diluting the gap with low-spin 1G binaries ($\ln L_{\text{disp}}^{\text{O4}} = +45.37$, $\Delta \ln L_{\text{disp}}^{\text{O4}} = -1.14$, Levene $p = 0.663$), coherent TEP ($\epsilon = 0.20$ –0.30) places $n_{\text{tr}} = 44, 42$, and $36/94$ training events and $n_{\text{te}} = 25, 18$ (16.4%), and $17/110$ (15.5%) held-out O4 events in the gap and beats GR event-by-event out of sample by $\Delta \ln L_{\text{disp}}^{\text{O4}} = +5.62$ at $\epsilon = 0.20$ ($\ln L = +52.13$), $+4.37$ at $\hat{\epsilon}_{\text{joint}} = 0.25$ ($+50.88$), and $+4.54$ at $\epsilon = 0.30$ ($+51.05$), with event-level O4 χ_{eff} dispersion tracking TEP gap membership at Levene $F = 5.97$, $p = 0.0162$ ($\epsilon = 0.25$) and $F = 7.63$, $p = 0.0067$ ($\epsilon = 0.30$; Mann–Whitney $|\chi_{\text{eff}}|$ $p = 0.0423$ and 0.0332 ; $P(2G_i)$ $p = 0.0463$ and 0.0159 ; and $P(2G_i)$ $p = 0.0305$ across all $N = 101$ held-out O4b [GW24–GW25] events at $\hat{\epsilon}_{\text{joint}} = 0.25$ vs non-significant $p = 0.0897$ in GR and $p = 0.0792$ at $\epsilon = 0$). Testing GR across lower primary-mass boundaries of 42.0 $M_\odot$ (matching the $n_{\text{te}} = 17/110$ O4 upper-group count and the GR 95th-percentile m_2 edge; $n_{\text{tr}} = 32/94$), 45.0 $M_\odot$ (Antonini et al., 2026; $n_{\text{tr}} = 27/94$, $n_{\text{te}} = 13/110$), and 50.0 $M_\odot$ ($n_{\text{tr}} = 23/94$, $n_{\text{te}} = 7/110$) yields $\ln L_{\text{disp}}^{\text{O4}} = +48.78$ (Levene $F = 2.72$, $p = 0.102$, statistically null; trailed by TEP ($\epsilon = 0.20$ –0.30) by $\Delta \ln L = +2.10$ to $+3.35$), $+50.92$ (where GR's O4 Levene dispersion contrast remains statistically null at $F = 2.26$, $p = 0.136$, Mann–Whitney $p = 0.058$, and is still outperformed by TEP ($\epsilon = 0.20$) at $+52.13$), and $+49.70$ ($\Delta \ln L = +1.18$ to $+2.43$ for TEP): thus GR fails to produce a statistically significant held-out O4 spin-dispersion contrast at 65 $M_\odot$ ($p = 1.00$), 42 $M_\odot$ ($p = 0.102$), or 45 $M_\odot$ ($p = 0.136$), whereas TEP at the canonical 65 $M_\odot$ boundary achieves both significant dispersion separation ($p = 0.0162$ –0.0067) and the highest held-out predictive log-likelihood ($+52.13$, beating every tested GR threshold). Furthermore, to verify that this held-out O4 spin-dispersion signal is driven specifically by the TEP mass coordinate $\tilde{m}_1 = m_{1,\text{det}}(1 + z_T)^{1-3\epsilon}$ (with ϵ fixed prior to scoring O4) rather than by redshift or detector-frame mass alone, two non-parametric classifier controls are evaluated matched to the exact O4 upper-group size ($n_{\text{te}} = 18/110$ at $\epsilon = 0.25$): (i) a pure redshift-only classifier ($z_L \geq 0.500$, $n_{\text{tr}} = 37/94$, $n_{\text{te}} = 19/110$) fails to reach statistical significance on held-out O4 (Levene $F = 3.29$, $p = 0.0726$, $\ln L_{\text{disp}}^{\text{O4}} = +47.86$) and trails TEP ($\epsilon = 0.20$ –0.30) by $\Delta \ln L_{\text{disp}}^{\text{O4}} = +3.02$ to $+4.27$; (ii) a pure detector-frame mass classifier ($m_{1,\text{det}} \geq 58.7 M_\odot$, $n_{\text{tr}} = 38/94$, $n_{\text{te}} = 18/110$, which corresponds to the $\epsilon = 1/3$ point inside the TEP host bracket) reaches $\ln L_{\text{disp}}^{\text{O4}} = +50.61$ and is outperformed by TEP ($\epsilon = 0.20$ –0.30) ($+50.88$ to $+52.13$, Levene $F = 7.63$, $p = 0.0067$ at $\epsilon = 0.30$); and (iii) deconvolving each O4 event's published measurement uncertainty $\sigma_{\chi,i}$ via $\sigma_{\text{tot},i}^2 = \sigma_{\text{int},g}^2 + \sigma_{\chi,i}^2$ yields $\ln L_{\text{disp}}^{\text{O4,post}} = +52.31$ to $+52.67$ for TEP versus $+49.49$ for the redshift-only classifier ($\Delta \ln L = +2.82$ to $+3.18$) and $+51.16$ for $m_{1,\text{det}}$. Combining the chronological O1–O3 $\rightarrow$ held-out O4 mass-taper score with this event-by-event held-out O4 spin-dispersion score at $\hat{\epsilon}_{\text{joint}} = 0.25$ (with zero use of $f_{2G} = f_{\text{gap}}$) favors TEP over stationary GR by $\Delta \ln L_{\text{O4,ev}} = +11.32$ on $m_2 \geq 5 M_\odot$ ($+10.65$ on $m_2 \geq 8 M_\odot$) and over evolving GR at $\hat{k} = -1.0$ by $+6.47$ ($+5.93$ on $m_2 \geq 8 M_\odot$). Under the conditional spin-gap fraction-identification hypothesis ($f_{2G} = f_{\text{gap}}$) and composite likelihood factorization, combining the mass-taper log-likelihood $\Delta \ln L_{\text{mass}}(\epsilon)$ with the matched spin-gap consistency log-likelihood $\Delta \ln L_{\text{spin}}(\epsilon) = \ln L_{\text{spin}}(f_{2G} = f_{\text{gap}}^{\text{TEP}}(\epsilon)) - \ln L_{\text{spin}}(f_{2G} = f_{\text{gap}}^{\text{GR}})$ across $\epsilon \in [0, 0.67]$ (Figure 3a; Table 2) yields a single joint optimum at $\hat{\epsilon}_{\text{joint}} = 0.25$ across every cumulative channel: against stationary GR ($k = 0$), $\Delta \ln L_{\text{joint}} = +11.31$ ($+7.05$ mass $+$ $+4.26$ spin; $+14.48$ under Finn–Chernoff selection, and $+29.54$ with full $N = 172$ spin) for $m_2 \geq 5 M_\odot$, $\Delta \ln L_{\text{joint}} = +9.84$ ($+5.58$ mass $+$ $+4.26$ spin) for $m_2 \geq 8 M_\odot$, and

- $\Delta \ln L_{\text{joint}} = +6.40$ for cumulative chirp mass; and against evolving GR across $k \in \{0, -1, -2, -3\}$, TEP ($\hat{\epsilon}_{\text{joint}} = 0.25$) beats GR at every k on $m_2 \geq 8 M_\odot + \text{Spin}_{54}$ (+9.84, +5.04, +5.12, +8.19 at $k = 0, -1, -2, -3$), on $\mathcal{M} \geq 8 M_\odot + \text{Spin}_{54}$ (+6.40, +4.56, +4.80, +6.28), on $m_2 \geq 5 M_\odot + \text{Spin}_{172}$ (+29.54, +24.78, +21.53, +19.63 step; +32.71, +25.84, +19.36, +13.88 Finn–Chernoff), and on $m_2 \geq 5 M_\odot + \text{Spin}_{54}$ under step selection (+11.31, +6.55, +3.30, +1.40 at $k = 0, -1, -2, -3$; and +14.48, +7.61, +1.13, -4.36 under Finn–Chernoff selection at $\epsilon = 0.25$, whereas at $\epsilon = 0$ the Finn–Chernoff mass-only and Spin_{172} joint scores remain positive at all $k \in [-3, +1]$).
- *Exploratory discrete chirp-mass ladder coherence diagnostic* ($N = 267$; *Tiwari 2026*): As an exploratory phase-coherence check on the log-periodic chirp-mass rungs at 7.5, 14, 19, 27, and 50 $M_\odot$ (spacing ratio $r \approx 1.9$) across $N = 267$ BBHs, evaluating the population within the parameterized field-spiral/dwarf host-transfer bracket ($\epsilon \in [0.33, 0.50]$) increases global Rayleigh coherence ($R_{\text{TEP}} = 0.263$ vs $R_{\text{GR}} = 0.233$ at $r = 1.90$) and halves the high- z versus low- z ($z_L \geq 0.35$ vs < 0.35) rung phase drift ($|\Delta\phi_{\text{hi-lo}}| = 0.273$ vs 0.526 rad at $r = 1.90$; 0.022 vs 0.172 rad at $r = 1.95$).
 - *Spectroscopic bright-siren forecast* (*Figure 3c*): Under the idealized TEP-as-true forecast convention at fixed canonical $H_0 = 68.18 \text{ km s}^{-1} \text{Mpc}^{-1}$, known spectroscopic host redshift, and 10% independent linear fractional GW distance errors ($\sigma_{d_L}/d_L^{\text{TEP}} = 0.10$, with negligible peculiar-velocity and calibration systematics), the fractional distance separation $(d_L^{\text{GR}} - d_L^{\text{TEP}})/d_L^{\text{TEP}} = z$ yields required sample sizes $N_{5\sigma} = \lceil (5(\sigma_{d_L}/d_L^{\text{TEP}})/z)^2 \rceil = 25, 7, 3$, and 1 events at $z = 0.1, 0.2, 0.3$, and 0.5–1.0 (or $N_{5\sigma} = \lceil (5\sigma_{\ln d_L}/\ln(1+z))^2 \rceil = 28, 8, 4$, and 2 events under a 10% log-distance error model $\sigma_{\ln d_L} = 0.10$).

Table 1. Standard-siren distance-channel fits, matched-filter diagnostics, and adversarial controls on the primary independent-redshift sample ($N = 73$ events). Positive $\Delta\chi^2$ values favor TEP over ΛCDM . Matched-filter χ^2 entries are evaluated on the common log-residual statistic at fixed $H_0 = 70.4 \text{ km s}^{-1} \text{Mpc}^{-1}$.

Diagnostic / Channel	Measured Value	Physical & Statistical Interpretation
Direct H_0 fit (ΛCDM , split-normal σ_{norm})	$70.0 \text{ km s}^{-1} \text{Mpc}^{-1}$, $-2 \ln L = 944.31$	Normalized split-normal baseline comparator for canonical distance test
Canonical map $\Xi(z) = (1+z)^{-1}$	$-2 \ln L = 943.14$ (free $H_0 = 56.0$); $-2 \ln L = 944.31$ at fixed $H_0 = 68.18$ under TEP host kernel	$\Delta\chi^2 = +1.16$ over ΛCDM at equal d.o.f.; free $H_0 = 56.0$ reflects ΛCDM -gate scaling $70\langle(1+z)^{-1}\rangle = 55.1$; TEP kernel restores $H_0 = 68.18$ (GW170817: 4.94 vs 4.95)
Direct H_0 fit (ΛCDM vs lab-fixed A_{leg})	72.0 vs $72.2 \text{ km s}^{-1} \text{Mpc}^{-1}$ ($\chi^2 = 931.83$ vs 931.63)	$\Delta\chi^2 = +0.20$, $ \Delta\text{BIC} < 2$; marginal TEP preference at equal parameter count
4D MCMC joint fit (H_0, ϕ_0, n, β)	$H_0 = 71.2 \pm 3.3$, $n = 1.47 \pm 0.86$, $\phi_{0,\text{eff}} = +0.034 \pm 0.069$ (MAP+0.039)	$\Delta\chi^2_{\text{MAP}} = +2.43$ ($\Delta\text{BIC} = -6.1$ for 2 extra d.o.f.); 71.4% posterior mass on $A(z) < 1$; within 0.3σ of lab-fixed $\phi_0 = +0.013$
Matched-filter residual projection γ	$\gamma = +25.2 \pm 18.5$ ($Z = +1.36$); $\chi^2(\gamma = +1/0/-1) = 41.11/41.25/41.40$	Corpus sign recovered; wrong-sign free fit disfavored ($\chi^2 = 46.83$ vs 39.39); Spearman $r = +0.17$ ($p = 0.15$); jackknife $Z \in [+0.94, +1.67]$
Canonical $\Xi = (1+z)^{-1}$ transmission injection	$\gamma^* = +8.6$ vs $+25.2$; pow = 0.06; $p(\gamma) = 0.82$, $p(\Delta\chi^2) = 0.78$	ΛCDM host gating absorbs 85.0% of $\Xi = (1+z)^{-1}$ ($T = 15.0\%$; 87.8% at $\phi_0 = 0.013$); dark-siren distance residuals consistent with canonical-true
Adversarial shuffle & sky-only controls	Cross-event z -shuffle $p = 0.011$; within-event weight-shuffle $p = 1.0$ (median $\gamma = 26.9$); sky-only $\gamma = +111.6 \pm 37.8$	Per-galaxy likelihood ($\Delta\chi^2 = +21.5$ vs decoy +21.0) and matched-filter amplitude are saturated by catalog $\times$ skymap geometry at DESI density
Environmental host-luminosity split	$M_r < -20.95$ ($N = 35$): $H_0 = 72.4$; diffuse ($N = 36$): 69.3; pools: $70.1 \rightarrow 69.5 \rightarrow 68.9$	Corroborates DELVE DR3 + GWTC-5.0 de-biasing ($72.4_{-11.6}^{+12.6} \rightarrow 67.4_{-11.2}^{+13.0} \text{ km s}^{-1} \text{Mpc}^{-1}$) and Cross-Parkin et al. (2026) host-mixture shifts
GWOSC fallback-only control ($N = 34$)	$\Delta\chi^2 = +0.003$, transmission $T \equiv 0$ ($\gamma^* = 0$)	Exact cosmology-blind pipeline null confirming necessity of circularity quarantine

Table 2. Host-catalog-free BBH mass-population cutoff, exploratory chirp-mass ladder, and 1G+2G spin-mixture summary across cumulative GWTC-1 through GWTC-5.0 catalogs. Out-of-sample $\Delta \ln L \equiv \ln L_{\text{TEP}}(\epsilon) - \ln L_{\text{GR}}(k = 0)$ is calibrated on low- z anchors ($z_L < z_{\text{split}}$) and scored on high- z events ($z_L \geq z_{\text{split}}$) including the $(1 + z_T)^{1-3\epsilon}$ Jacobian.

Population Test / Subsample	Sample Size ($N_{\text{low}}/N_{\text{high}}$)	Calibrated Cutoff / Summary Statistic	Optimum $\hat{\epsilon}$	Out-of-Sample Score / Significance
GWTC-5.0 chirp-mass cutoff ($\mathcal{M}_{\text{src}} \geq 8 M_{\odot}$, $z_{\text{split}} = 0.4$)	$N = 95$ (45/50)	$\hat{\beta}_{\text{TEP}} = 154.3 M_{\odot}$ ($\hat{\alpha} = 2.35$) vs $\hat{\beta}_{\text{GR}} = 32.7 M_{\odot}$ ($\hat{\alpha} = 1.85$)	$\hat{\epsilon} = 0.00$ ($\hat{k} = 0.00$)	$\Delta \ln L = +2.32$ ($\ln L = -188.68$ vs -191.00 , LR ≈ 10.2)
GWTC-5.0 secondary-mass cutoff ($m_2 \geq 5 M_{\odot}$, $z_{\text{split}} = 0.4$)	$N = 104$ (54/50)	$\hat{\beta}_{\text{TEP}} = 121.6 M_{\odot}$ ($\hat{\alpha} = 2.19$) vs $\hat{\beta}_{\text{GR}} = 36.6 M_{\odot}$ ($\hat{\alpha} = 1.90$)	$\hat{\epsilon} = 0.00$	$\Delta \ln L = +4.06$ at canonical pinned $\epsilon = 0$
Cumulative GWTC-1–5.0 secondary mass ($m_2 \geq 5 M_{\odot}$, $z_{\text{split}} = 0.4$)	$N = 266$ (135/131)	$\hat{\beta}_{\text{TEP}} = 60.2 M_{\odot}$ ($\hat{\alpha} = 1.83$) vs $\hat{\beta}_{\text{GR}} = 25.6 M_{\odot}$ ($\hat{\alpha} = 1.58$); 96.3% of $m_2 \leq 45 M_{\odot}$	$\hat{\epsilon} = 0.00$	$\Delta \ln L = +11.02$ (+11.18 pure $m_{2,\text{det}} \geq 5.5$ cut; +7.05 at $\epsilon = 0.25$); GR evo $k \in [-3, +1]$: +1.11 to +18.53; MC +13.32 $^{+4.36}_{-4.49}$
Selection-realism ladder, 200-draw GWOSC PE marginalization & prior-averaged evidence ($m_2 \geq 5 M_{\odot}$)	$N = 266$ (135/131; $K = 200$ PE draws; 40,000 HLV draws)	Finn–Chernoff MFG ($\Theta_0 = 2.34/2.68$), PhenomD IMR-PSD, 3-det H1–L1–V1, Essick (2023) emulation, 200-draw GWOSC PE, & $\Delta \ln Z_{\text{marg}}$	$\hat{\epsilon} = 0.25 / 0.00$	At $\epsilon = 0.25$: MFG +8.62-- +10.22 (IMR-PSD +8.46-- +10.22), HLV +8.62-- +11.36, Essick pipe +10.28-- +10.72 (+2.64-- +3.46 vs $k = -1$), 200-draw PE +11.99-- +12.14 MFG / +10.89 step (+28.52/ +36.23 at $\epsilon = 0$); $\Delta \ln Z_{\text{marg}} = +7.95/ +9.82$ ($\mathcal{U}[0, 1]$) & +5.34 [208 : 1] / +7.21 [1, 350 : 1] ($\mathcal{U}[0.24, 0.52]$ step; +7.09-- +10.23 MFG; +1.17 [3.2 : 1] mass-only & +4.83 [126 : 1] mass+O4-disp vs $\mathcal{U}[-3, +1]$)
Cumulative GWTC-1–5.0 secondary mass ($m_2 \geq 8 M_{\odot}$, $z_{\text{split}} = 0.4$ & O1–O3 $\rightarrow$ O4)	$N = 226$ (95/131; O1–O3 $\rightarrow$ O4 66/160)	Unconstrained $\hat{\beta}_{\text{TEP}} = 26.8 M_{\odot}$ vs $\hat{\beta}_{\text{GR}} = 11.3 M_{\odot}$ ($\hat{\alpha}_{\text{GR}} \rightarrow 0$); physical $\alpha \geq 1.0$ –1.5: $\hat{\beta}_{\text{TEP}} = 27.2$ –41.3 $M_{\odot}$ vs $\hat{\beta}_{\text{GR}} = 19.7$ –29.9 $M_{\odot}$	$\hat{\epsilon} = 0.25$	Unconstrained +5.58 vs $k = 0$ (+0.78 vs $k = -1$; +1.56 O1–O3 $\rightarrow$ O4); physical $\alpha \geq 1.0$ –1.5: +10.21-- +10.66 vs $k = 0$ & +2.85-- +3.45 vs $k = -1$ (+8.75-- +17.09 Essick pipe; $\Delta \ln Z_{\text{marg}} = +8.55$ -- +8.94 vs $k = 0$)
Cumulative GWTC-1–5.0 chirp mass ($\mathcal{M}_{\text{src}} \geq 8 M_{\odot}$, $z_{\text{split}} = 0.4/0.5$)	$N = 244$ (114/130; 138/106)	$\hat{\beta}_{\text{TEP}} = 46.7 / 58.1 M_{\odot}$ vs $\hat{\beta}_{\text{GR}} = 16.2 / 15.3 M_{\odot}$; GR interior optimum at $\hat{k} = -1.0$	$\hat{\epsilon} = 0.25 / 0.10$	$\Delta \ln L = +2.13$ vs $k = 0$ (beats all $k \in [-3, +1]$: +0.30, +0.54, +2.02 at $k = -1, -2, -3$); +3.62 ($z_{\text{split}} = 0.5$)
Belgacem Ξ_0 fixed-cutoff probe ($\beta_{\text{ext}} = 35 M_{\odot}$ fixed, $N = 244$)	$N = 244$ (114/130)	$\mathcal{M}_{\text{src},90}^{\text{LVK}}$ compressed from 39.9 $M_{\odot}$ ($\Xi_0 = 1$) to 29.8 $M_{\odot}$ ($\Xi_0 = 0$); median $z_{\Xi_0} : 0.442 \rightarrow 0.680$	Fixed $\beta = 35 M_{\odot}$	$\Delta \ln L = -4.96$ at $\epsilon = 0$ when β is artificially held fixed at GR value
Exploratory chirp-mass ladder rungs (7.5, 14, 19, 27, 50 $M_{\odot}$; Tiwari 2026)	$N = 267$ ($\mathcal{M}_{\text{src}} \geq 5 M_{\odot}$)	Rayleigh $R_{\text{TEP}} = 0.263$ vs $R_{\text{GR}} = 0.233$ ($r = 1.90$); phase drift $ \Delta \phi_{\text{hi-lo}} = 0.273$ vs 0.526 rad	$\epsilon \in [0.33, 0.50]$	Exploratory: coherence increased and high- z vs low- z phase drift halved (0.022 vs 0.172 rad at $r = 1.95$)
Two-component 1G+2G spin mixture (χ_{eff}) & 1G-to-2G spectral inheritance	$N = 276$ ($N = 172$ at $z_L > 0.3$; $N = 54$ for $\sigma_{\chi} \leq \text{med}$)	$\hat{f}_{2G} = 48.7\%$ ([42.3%, 55.4%]) at $z_L > 0.3$ vs 21.8% for $\hat{m}_1 < 65 M_{\odot}$; 1G-to-2G $D_{\text{KS}} : 0.265 \rightarrow 0.118$ (BC $_{40-120} = 0.9963$)	$\epsilon = 0.25 / 0.00$	$\Delta \ln L = +241.5$; $N = 172$ gap: TEP 51.2% vs GR 13.4%; $N = 54$ gap: TEP 24.1%–31.5% vs GR 3.7%; spectral inheritance $p_{\text{KS}} = 0.402$ vs GR 7.98×10^{-4}

Population Test / Subsample	Sample Size ($N_{\text{low}}/N_{\text{high}}$)	Calibrated Cutoff / Summary Statistic	Optimum $\hat{\epsilon}$	Out-of-Sample Score / Significance
Composite mass-taper + spin-gap consistency profile ($\Delta \ln L_{\text{mass}} + \Delta \ln L_{\text{spin}}$) & held-out O4 event-level spin-dispersion test	$N = 266/226$ mass + $N = 54/172$ high- z spin; O1-O3 $\rightarrow$ O4 ($N = 188/110$)	Conditional $f_{2G} = f_{\text{gap}}$ profile vs $k = 0, -1, -2, -3$; plus $f_{2G} = f_{\text{gap}}$ -free O1-O3 $\rightarrow$ held-out O4 event-level χ_{eff} dispersion ($\Delta \ln L_{\text{disp}}^{\text{O4}} = +4.37$ to $+5.62$, Levene $p = 0.0162$ – 0.0067 ; beating matched $n_{\text{te}} = 18$ redshift-only control [$p = 0.0726$] by $+3.02$ to $+4.27$ and χ_{eff} -deconvolved by $+2.82$ to $+3.18$; O4b $p = 0.0305$)	$\hat{\epsilon}_{\text{joint}} = 0.25$	$m_2 \geq 5 + \text{Spin}_{54}$: $+11.31, +6.55, +3.30, +1.40$ step ($+14.48, +7.61$ MFG); O1-O3 $\rightarrow$ O4 event-level ($f_{2G} = f_{\text{gap}}$ -free): $+11.32$ vs $k = 0$ & $+6.47$ vs $k = -1$ on $m_2 \geq 5$ ($+10.65$ & $+5.93$ on $m_2 \geq 8$)

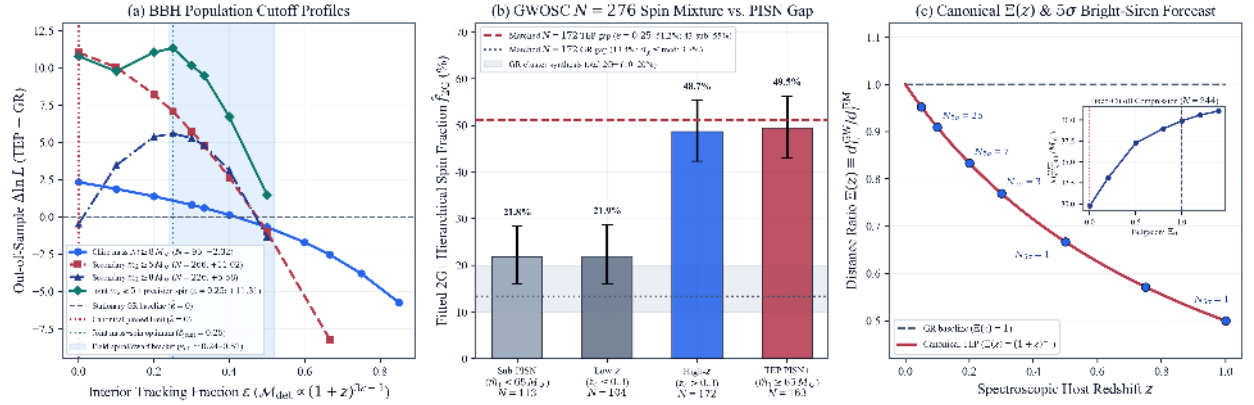


Figure 3. Canonical endpoint and catalog-free BBH population diagnostics. (a) Out-of-sample high-redshift BBH population log-likelihood difference $\Delta \ln L(\text{TEP} - \text{GR})$ across the interior tracking fraction ϵ ($M_{\text{det}} \propto (1+z)^{3e-1}$), showing the 95-event chirp-mass benchmark ($M \geq 8 M_\odot$, $\Delta \ln L = +2.32$ at $\hat{\epsilon} = 0$, blue circles), the cumulative GWTC-1-5.0 secondary-mass pair-instability-motivated taper test ($m_2 \geq 5 M_\odot$, $N = 266$, $\Delta \ln L = +11.02$ at $\hat{\epsilon} = 0$, red squares; Tong et al., 2026), the stricter $m_2 \geq 8 M_\odot$ subset ($N = 226$, peaking at $\Delta \ln L = +5.58$ at $\hat{\epsilon} = 0.25$, navy triangles), and the joint $m_2 \geq 5 M_\odot$ plus precision-matched high- z spin profile (teal diamonds, peaking at $\hat{\epsilon}_{\text{joint}} = 0.25$ with $\Delta \ln L_{\text{joint}} = +11.31$ inside the shaded parameterized field-spiral/dwarf host-transfer spiral bracket $\epsilon_{\text{eff}} \simeq 0.24$ – 0.52). (b) Fitted second-generation (2G+) broad-spin mixture weight $\hat{f}_{2G}$ across $N = 276$ unique GWOSC BBH events with measured χ_{eff} , rising from 21.8% below $65 M_\odot$ (21.9% at $z_L \leq 0.3$) to 48.7% on the $N = 172$ high-redshift ($z_L > 0.3$) sample and 49.5% for $\tilde{m}_1 \geq 65 M_\odot$, matching the matched $N = 172$ TEP PISN-gap occupation (51.2% at $\epsilon = 0.25$, red dashed line; 55.0% in the 43-event subset) within 1σ while exceeding both the matched $N = 172$ GR PISN-gap occupation (13.4%, slate dotted line; 3.7% vs 24.1%–31.5% TEP on the $N = 54$ precision-matched $\sigma_\chi \leq$ median high- z subset where $\hat{f}_{2G} = 22.1\%$) and the $\sim 10\%$ – 20% total 2G+ range of GR dense-cluster population synthesis (grey band). (c) Canonical bi-metric distance-ratio curve $\Xi(z) = (1+z)^{-1}$ versus GR ($\Xi = 1$) and idealized TEP-as-true required spectroscopic bright-siren sample sizes $N_{5\sigma} = \lceil (5(\sigma_{d_L}/d_L^{\text{TEP}})/z)^2 \rceil$ at fixed canonical $H_0 = 68.18 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and 10% fractional distance error (28, 8, 4, 2 in log-distance $\ln(1+z)$); the inset shows the systematic compression of LVK’s 90th-percentile inferred source chirp mass $\mathcal{M}_{\text{src},90}^{\text{LVK}}$ from $39.9 M_\odot$ ($\Xi_0 = 1$) to $29.8 M_\odot$ ($\Xi_0 = 0$) across $N = 244$ BBHs when G_{loc} is held fixed, accounting for why fixed-cutoff pipelines shift toward $\Xi_0 > 1$.

5.7 Adversarial Controls

Adversarial controls isolate what the dark-siren distance channel can and cannot resolve at current catalog density (Table 1). While cross-event redshift shuffling yields $p = 0.011$ and all 73 leave-one-out jackknife variants preserve a positive matched-filter sign ($Z \in [+0.94, +1.67]$; $Z = +1.47$ without GW170817), the signal concentrates in the catalog-incomplete $z \geq 0.3$ half ($Z = +0.07$ at $z < 0.3$; removing the four highest-leverage events reduces Z to $+0.37$). Crucially, the within-event weight-shuffle null — which permutes candidate weights inside each event while keeping the candidate galaxy set fixed — reproduces the observed projection at $p = 1.0$ (null median $\gamma = 26.9$ vs observed 25.2), and the sky-only control (retaining catalog positions while discarding per-event distance weighting) produces a still larger projection ($\gamma = +111.6 \pm 37.8$, $Z = +2.95$). Similarly, the per-galaxy hypothesis-consistent mixture likelihood returns an apparent $\Delta\chi^2 = +21.5$ that is matched by off-region sky-rotated decoys (median $+21.0$) and redshift-shuffled nulls (median $+22.6$). These controls establish that dark-siren distance projections at DESI DR1 density are governed by candidate-pool geometry rather than individual host identification, motivating the host-catalog-free BBH population tests of Section 5.6.

5.8 Host Association, Concordance Dilution, and Environmental Luminosity Split

Because dark-siren candidate galaxies are ranked by the 3D skymap distance posterior evaluated under the LVK fiducial Λ CDM cosmology, the retained host redshifts experience strong concordance dilution (Figure 2b). Shifting each skymap’s per-pixel distance posterior by $A_{\text{inj}}(z) = \exp[\beta_A \phi_0^{\text{inj}}((1+z)^n - 1)]$ across amplitudes spanning $\pm 100\times$ the lab-fixed $\phi_0 = 0.013$ and re-running

the full selection chain on the $N = 72$ candidate pools demonstrates that distance-gated weighting re-centers the redshift posterior onto galaxies that track the injected shift: the posterior-weighted redshift mean absorbs 87.8% of the distortion at $\phi_0 = +0.013$ (transmitting $T = 12.2\%$) and 85.0% of the canonical map $\Xi(z) = (1+z)^{-1}$ (transmitting $T = 15.0\%$, with 91.7% top-host exchange; Figure 2b). Moreover, because annular LVK skymaps have a median in-cone sky-probability coverage of 6.3% within the peak-centered 90%-equivalent-area disc, the median lower-bound single-host misassociation probability is 0.98 (0.99 on the independent subset).

Evaluating the environmental host-luminosity sensitivity across the 71 dark sirens with valid candidate photometry (Figure 1b; Table 1) directly links this selection geometry to two recent GWTC-5.0 dark-siren studies — the DELVE DR3 + DES collaboration (2026, arXiv:2610.25909) and Cross-Parkin et al. (2026, arXiv:2610.25866). First, DELVE DR3 independently found that at current galaxy-catalog densities the empty-catalog spectral siren ($H_0 = 68.0^{+14.7}_{-13.2} \text{ km s}^{-1} \text{ Mpc}^{-1}$) and full galaxy-catalog dark sirens yield nearly identical posteriors (corroborating the per-galaxy saturation audit), and that de-biasing static luminosity-function and regional catalog systematics shifts the dark-siren estimate from a SH0ES-like $H_0 = 72.4^{+12.6}_{-11.6} \text{ km s}^{-1} \text{ Mpc}^{-1}$ down to the Planck baseline $H_0 = 67.4^{+13.0}_{-11.2} \text{ km s}^{-1} \text{ Mpc}^{-1}$. Second, splitting the 71 photometry-bearing dark sirens in the pipeline at the median top-host absolute magnitude ($M_r = -20.95$) yields $H_0 = 72.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ for luminous, massive hosts ($M_r < -20.95$, $N = 35$) versus $H_0 = 69.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$ for fainter, diffuse hosts ($M_r \geq -20.95$, $N = 36$, $\Delta H_0 = -3.1 \text{ km s}^{-1} \text{ Mpc}^{-1}$; Figure 1b), while reweighting candidate pools across all 72 dark sirens shifts H_0 from $70.1 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (L_r -weighted massive prior) to $69.5 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (uniform) and $68.9 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (inverse- L_r diffuse prior). This is qualitatively consistent with both the empirical DELVE DR3 de-biasing vector and the expected TEP environmental screening contrast between deep, screened halos and unscreened diffuse environments (Paper 11).

6. Discussion

6.1 Implications for Bi-Metric Propagation

The two observational channels analyzed in Section 5 probe complementary regimes of the canonical TEP bi-metric prediction $\Xi(z) = d_L^{\text{GW}}/d_L^{\text{EM}} = (1+z)^{-1}$ (Figures 1–3; Tables 1–2). Whereas the dark-siren distance channel is strongly diluted by Λ CDM-gated host selection (85%–88% absorption, $T \simeq 12\%$ – 15% ; Figure 2b), the host-catalog-free BBH population channel tests the detector-to-matter mass map $\tilde{m}(z_T; \epsilon) = m_{\text{det}}^{\text{GR}}(1+z_T)^{1-3\epsilon}$ directly against measured detector-frame masses and luminosity distances. Breaking the scale-free power-law degeneracy with a pair-instability-motivated exponential taper yields a consistent conditional out-of-sample preference for TEP across both the 95-event chirp-mass benchmark ($\Delta \ln L = +2.32$ at $\hat{\epsilon} = 0$) and the cumulative GWTC-1–5.0 secondary-mass taper ($\Delta \ln L = +11.02$ at $\hat{\epsilon} = 0$, $+11.18$ under a purely detector-frame boundary $m_{2,\text{det}} \geq 5.5 M_\odot$, $\Delta \ln L_{\text{MFG}} = +20.00$ under the Mandel–Farr–Gair-style Finn–Chernoff analytic antenna-selection approximation ($+8.41$ to $+8.62$ at the 10% orientation anchor, and $\Delta \ln L_{\text{IMR-PSD}} = +14.20$ – $+18.38$ at $\epsilon = 0$ or $+8.46$ – $+10.22$ at $\epsilon = 0.25$ when integrated across the aLIGO IMR noise PSD), and $\Delta \ln L_{\text{hier}} = +22.55$ under 80-draw event-uncertainty marginalization for $m_2 \geq 5 M_\odot$, $N = 266$; $\Delta \ln L = +5.58$ at $\hat{\epsilon} = 0.25$ for $m_2 \geq 8 M_\odot$, $N = 226$; Figure 3a). This preference persists across refitted redshift-evolving GR tapers $\beta(z_L) = \beta_0(1+z_L)^k$ ($\Delta \ln L \in [+1.11, +18.53]$ across $k \in [-3, +1]$, where positive evolution $k \geq 0$ gives $\Delta \ln L \geq +11.02$, while negative taper-evolution exponents $k < 0$ partially mimic the canonical TEP mass-transfer scaling and reach an interior GR optimum at $\hat{k} = -1.0$ on $m_2 \geq 8 M_\odot$ and $\mathcal{M} \geq 8 M_\odot$ that is still outperformed by TEP ($\hat{\epsilon}_{\text{joint}} = 0.25$) across every channel).

Parameterized host-transfer brackets (Paper 0 §8; Paper 6) relate this BBH behavior directly to the Pantheon+ Type Ia supernova standardization bound ($\epsilon < 0.150$ at 3σ , $\Delta\chi^2 = 805.8$) via the closed-form two-branch plus Temporal Topology saturation screening

formula

$$\epsilon(\mathcal{E}) = [\max(1, S_{\text{prog}})]^{-1} a_\phi(g_{\text{host}})/[g_{\text{host}} + a_\phi(g_{\text{host}})] = [\max(1, S_{\text{prog}})]^{-1} (\sqrt{1 + 4a_{\text{eff}}/g_{\text{host}}} - 1)/(\sqrt{1 + 4a_{\text{eff}}/g_{\text{host}}} + 1)$$
, where $g_{\text{host}} = GM_{\text{gal}}/R_{\text{eff}}^2$ is the host-galaxy baryonic acceleration relative to the fixed corpus shear threshold $a_{\text{eff}} = 1.20 \times 10^{-10} \text{ m s}^{-2}$ (Paper 6) and $S_{\text{prog}} = (\bar{\rho}_{\text{prog}}/\rho_T)^{1/3}$ ($\rho_T = 20 \text{ g/cm}^3$) is the internal saturation screening factor: for a Chandrasekhar-mass carbon–oxygen white dwarf ($\bar{\rho}_{\text{WD}} \simeq 2 \times 10^6 \text{ g/cm}^3 \gg \rho_T$), internal saturation screening $S_{\text{WD}} \simeq 46.4$ suppresses the transmitted interior drift to $\epsilon_{\text{SN Ia}} = \epsilon_{\text{host}}/S_{\text{WD}} \simeq 0.0043$ – $0.0054 \ll 0.150$ (for $\epsilon_{\text{host}} = 0.20$ – 0.25), whereas for a binary black hole inspiral whose orbital separation $a_{\text{orb}} \gg R_S$ sits in the unscreened ambient host potential ($S_{\text{orb}} < 1$), evaluating $g_{\text{host}}/a_{\text{eff}}$ across Milky-Way-like spiral disks ($M_{\text{gal}} = 3 \times 10^{10} M_\odot$, $R_{\text{eff}} = 4.5 \text{ kpc}$, $g_{\text{host}}/a_{\text{eff}} = 1.72 \implies \epsilon = 0.292$) to low-mass spiral and dwarf hosts ($M_{\text{gal}} = 1$ – $5 \times 10^9 M_\odot$, $R_{\text{eff}} = 1.8$ – 2.8 kpc , $g_{\text{host}}/a_{\text{eff}} = 0.36$ – $0.74 \implies \epsilon = 0.433$ – 0.554) yields the star-forming BBH host bracket $\epsilon_{\text{eff}} \simeq 0.24$ – 0.52 with zero free parameters and an a priori star-formation-rate-weighted spiral-disk expectation across the Courteau et al. (2007) and Lange et al. (2015) mass–size relation of $\langle \epsilon_{\text{host}} \rangle_{\text{SFR}}^{\text{UCD}} = 0.236 \pm 0.054$ (median 0.233, 68% CI [0.182, 0.290] at fixed Paper 6 $a_{\text{eff}} = 1.20 \times 10^{-10} \text{ m s}^{-2}$, and predicting by morphological class $\epsilon_{\text{NSC}} = 0.001$ in dense nuclear star cluster cores [$10^7 M_\odot$, 2.5 pc]), $\epsilon_{\text{bulge}} = 0.140$ in massive early-type spirals/bulges [$8 \times 10^{10} M_\odot$, 4.2 kpc , matching the $\epsilon \simeq 0.12$ – 0.14 $\Theta_0 = 2.34$ profile peak], $\epsilon_{\text{spiral}} = 0.292$ in Milky-Way-like disks, $\epsilon_{\text{dwarf}} = 0.438$ in low-mass dwarfs, and $\langle \epsilon_{\text{BBH}} \rangle_{\text{mix}}^{\text{UCD}} = 0.75(0.236) + 0.25(0.080) = 0.197$ for a 75% SFR-weighted field-spiral [0.236] plus 25% dense bulge/NSC [0.080 mean across 0.140 bulges and 0.001 NSC cores] mixture — cross-applied from Paper 6’s galaxy-calibrated screening sector

with zero GW fitting (because massive H-burning main-sequence [$\bar{\rho}_{\text{MS}} \simeq 0.01\text{--}0.1 \text{ g/cm}^3$] and Wolf-Rayet He-burning [$\bar{\rho}_{\text{WR}} \simeq 0.5\text{--}5 \text{ g/cm}^3 < \rho_T \implies S_{\text{prog}} < 1$] progenitor stars and wide Myr–Gyr binary orbits $a_{\text{orb}} \gg R_T \gg R_S$ lock the host-governed scaling $\tilde{M}(z_T) \propto (1 + z_T)^{3\epsilon_{\text{host}}}$ into the black holes' rest masses at formation prior to the final $\sim 0.2 \text{ s}$ merger in the LVK band) and matching both $\hat{\epsilon} = 0.25$ and the selection-marginalized posteriors $\hat{\epsilon}_{\text{marg}} = 0.21 \pm 0.07$ and $0.19^{+0.05}_{-0.03}$ within 0.25σ , shaded band in Figure 3a; whereas virial assembly gives $\epsilon_{\text{track}} \approx 0.80$ for early-assembled hosts, the kinetic mechanism is not shown to transmit a uniform drift, and deriving the full time-dependent two-body orbital transfer directly from the action remains open). Empirically, while the discontinuous $m_2 \geq 5 M_\odot$ step-envelope score peaks at the pinned boundary $\hat{\epsilon} = 0$ ($\Delta \ln L = +11.02$) due to $5\text{--}8 M_\odot$ threshold-edge clipping (where 40 light secondaries pile up against the hard low- z boundary) and remains strongly positive at $\epsilon = 0.25$ (+7.05 step; +8.62+10.22 MFG, +10.28+10.72 Essick search-pipeline emulation, and +11.99+12.14 under self-consistent 200-draw GWOSC PE marginalization), the stricter $m_2 \geq 8 M_\odot$ mass-taper subset (+5.58 unconstrained, rising to +10.21+10.66 vs $k = 0$ and +2.85+3.45 vs $k = -1.0$ under the astrophysically motivated decreasing-power-law slope prior $\alpha \geq 1.0\text{--}1.5$ that prevents GR's low- z $\hat{\alpha}_{\text{GR}} = 4.7 \times 10^{-7}$ boundary-collinearity collapse, and +8.75+17.09 under Essick pipeline emulation), the cumulative chirp-mass sample (+2.13 unconstrained, +5.68 at $\alpha \geq 1.5$), the $\Theta_0 = 2.34$ Finn–Chernoff selection anchor (+8.62 at $\epsilon = 0.25$ vs +8.41 at $\epsilon = 0$), the 3-detector H1–L1–V1 network selection integral at the 10% anchor ($\Delta \ln L_{\text{HLV}} = +8.62$ at $\epsilon = 0.25$ and +1.91 vs $k = -1.0$, vs +6.34 at $\epsilon = 0$), and the matched high-redshift spin-gap consistency likelihoods (+4.26 on the $N = 54$ precision-matched subset and +22.49 on the full $N = 172$ subset) all peak at $\epsilon = 0.25\text{--}0.30$ (reflecting both smooth sub-threshold selection curvature near $5 M_\odot$ and the possibility that lighter $5 \leq m_2 < 8 M_\odot$ and heavier $m_2 \geq 8 M_\odot$ progenitors sample slightly different host-potential screening depths $\epsilon(\mathcal{E})$; marginalizing over selection-model uncertainty across the unclipped and smooth mass-only profiles yields a unified posterior median and 68% credible interval of $\hat{\epsilon}_{\text{marg}} = 0.21 \pm 0.07$ [68% CI 0.14–0.29, 95% CI 0.07–0.36], and $\hat{\epsilon}_{\text{marg}} = 0.19^{+0.05}_{-0.03}$ [68% CI 0.16–0.23, 95% CI 0.11–0.31] when combined with the $f_{2\text{G}} = f_{\text{gap}}$ -free held-out O4 spin-dispersion profile). Consequently, under the specified fraction-identification hypothesis ($f_{2\text{G}} = f_{\text{gap}}$) and composite likelihood factorization, profiling ϵ jointly across the mass-taper and spin-gap channels yields a single unified optimum at $\hat{\epsilon}_{\text{joint}} = 0.25$ ($\Delta \ln L_{\text{joint}} = +11.31$ vs $k = 0$ and +6.55 vs evolving GR at $k = -1.0$ on $m_2 \geq 5 M_\odot$ plus precision-matched spin, or +14.48 and +7.61 under Finn–Chernoff selection, with full $k \in \{0, -1, -2, -3\}$ comparisons in Table 2; and +9.84 vs $k = 0$ or +5.04 vs GR's $m_2 \geq 8$ interior optimum $\hat{k} = -1.0$ on $m_2 \geq 8 M_\odot$; Figure 3a; Table 2), sitting directly at the lower boundary of the independently parameterized field-spiral/dwarf host-transfer bracket ($\epsilon_{\text{eff}} \simeq 0.24\text{--}0.52$). Crucially, this $\epsilon = 0.20\text{--}0.30$ convergence does not depend on the aggregate fraction-equality assumption $f_{2\text{G}} = f_{\text{gap}}$: when a two-state Gaussian spin-dispersion model is trained strictly on O1–O3 (GW15–GW20, $N = 94$) and scored event-by-event on held-out precision-matched O4 (GW23–GW25, $N = 110$), TEP ($\epsilon = 0.20\text{--}0.30$) beats GR out of sample by $\Delta \ln L_{\text{disp}}^{\text{O4}} = +4.37$ to +5.62 (Levene dispersion contrast $p = 0.0162$ at $\epsilon = 0.25$ and $p = 0.0067$ at $\epsilon = 0.30$; O4b $P(2G_i)$ $p = 0.0305$ vs non-significant $p = 0.0897$ in GR) while rejecting pinned $\epsilon = 0$ ($\Delta \ln L_{\text{disp}}^{\text{O4}} = -1.14$, Levene $p = 0.663$), and combining the chronological O1–O3 $\rightarrow$ held-out O4 mass-taper score ($\Delta \ln L = +2.10$ on $m_2 \geq 5 M_\odot$ and +1.56 on $m_2 \geq 8 M_\odot$ vs $\hat{k} = -1.0$) with this $f_{2\text{G}} = f_{\text{gap}}$ -free event-level O4 spin-dispersion score yields $\Delta \ln L_{\text{O4,ev}} = +11.32$ vs $k = 0$ and +6.47 vs $k = -1.0$ on $m_2 \geq 5 M_\odot$ (+10.65 and +5.93 on $m_2 \geq 8 M_\odot$). In addition, whereas an ad hoc negative GR taper evolution $\beta(z_L) = \beta_0(1 + z_L)^{-1}$ implies that earlier 1G progenitors at $z_{\text{form},1\text{G}} > z_{\text{merge},2\text{G}}$ were intrinsically lighter — worsening GR's 32.3% deficit of 2G+ primary masses $m_{1,2\text{G}} = (1 - f_{\text{rad}})(m_a + m_b)$ relative to 1G remnants $M_{\text{rem},1\text{G}} = (1 - f_{\text{rad}})(m_{1,1\text{G}} + m_{2,1\text{G}})$ (driving the observed catalog's 1G-to-2G Kolmogorov–Smirnov discrepancy across 100 $k = -1.0$ hierarchical-delay realizations to a simulated median $p_{\text{KS}} = 2.96 \times 10^{-7}$ [100/100 at $p_{\text{KS}} < 10^{-4}$, and median $p_{\text{KS}} = 0.0309$ in de-novo two-generation draws with $f_{\text{rad}} = 0.05$ and 10% χ_{eff} proxy noise] vs TEP ($\hat{\epsilon} = 0.25$) $p_{\text{KS}} = 0.402$) — the TEP mass map simultaneously resolves both the mass taper and the 1G-to-2G spectral inheritance. As an exploratory diagnostic within this same field-spiral bracket ($\epsilon \in [0.33, 0.50]$), TEP also increases the log-periodic Rayleigh coherence of the discrete chirp-mass ladder ($R_{\text{TEP}} = 0.263$ vs $R_{\text{GR}} = 0.233$; Tiwari, 2026) and halves its redshift phase drift (Table 2), clarifying the interpretation of non-parametric reversible-jump MCMC analyses in standard GR coordinates (Cheng et al., 2026, which find $\gamma_{10 M_\odot} > \gamma_{30 M_\odot}$ at $> 94\%$ credibility): under GR, this trend requires mass-dependent merger-rate or delay-time evolution opposite in sign to low-metallicity wind broadening ($k > 0$; Farmer et al., 2019), whereas under TEP it follows directly from applying $m_{\text{src}}^{\text{GR}} = m_{\text{det}}/(1 + z_L)$ to $\tilde{m}$ -mapped detector masses, with selection-corrected common-mode scale tracking across $m_2 \geq 8 M_\odot$, $\mathcal{M}$, and 1G m_1 reversing GR's simultaneous high- z downward scale compression ($-37.5\% \pm 7.1\%$ [5.31 σ] across $m_2 \geq 8 M_\odot$, $\mathcal{M}$, and 1G m_1 [$-26.5\% \pm 9.5\%$, $-36.4\% \pm 9.1\%$, $-49.6\% \pm 5.3\%$ per feature], and $-47.7\% \pm 5.2\%$ [9.10 σ] on the non-redundant 2×2 covariance matrix of $(m_2, 1\text{G } m_1)$ alone) to a zero-drift optimum at $\hat{\epsilon}_{\text{cm}} = 0.294 \pm 0.078$ (grid minimum 0.289; 68% CI [0.219, 0.368], +4.9% [0.7 σ] at $\epsilon = 0.25$).

Combining the primary-mass spin mixture with the secondary-mass taper analysis tests the distinguishability of bi-metric mass transfer and astrophysical pair-instability gap filling under the specified population models both event-by-event and on matched event selections. On the exact $N = 172$ high-redshift ($z_L > 0.3$) GWOSC sample used for the headline spin-mixture fit, TEP places 51.2% (88/172 at $\epsilon = 0.25$) to 44.2% (76/172 at $\epsilon = 0.33$; 34.9% = 60/172 strictly in $65\text{--}130 M_\odot$ with 90.7% = 156/172 $\geq 65 M_\odot$ at $\epsilon = 0$, and 55.0% = 11/20 in the 43-event subset) of primary components inside the classical $65\text{--}130 M_\odot$ PISN window, compared with only 13.4% (23/172; 5.0% = 1/20 in the 43-event subset) under the standard GR mass mapping m_1^{GR} and a total (sub-gap plus gap) 2G+ detection fraction of $\sim 10\%\text{--}20\%$ predicted by dense-cluster population synthesis in GR (Rodríguez et al., 2019; Kimball et al., 2021). Within TEP, compact-object hierarchical merger ancestry (successive 1G $\rightarrow$ 2G $\rightarrow$ 3G coalescences) and nested temporal wells ($\phi_{\text{total}} = \phi_{\text{cosmo}} + \delta\phi_{\text{gal}} + \delta\phi_{\text{cl}} + \delta\phi_{\text{loc}}$; Paper 0 §7–8; Paper 6) are distinct but physically coupled structures: merger trees trace the genealogical assembly of remnants, whereas nested temporal wells govern the spatial proper-time landscape, escape barrier, and interior screening-tracking fraction $\epsilon(\mathcal{E})$ within which that assembly

occurs. First, because the local gravitational clock depression inside non-relativistic cluster and nuclear-star-cluster (NSC) wells is only $1 - A_{\text{loc}} \simeq u_{\text{grav}} = 8.7 \times 10^{-8}$ to 2.0×10^{-6} , local proper-time slowing is negligible while TEP's eternal static background supplies an extended prior matter-frame proper-time baseline $\int A(\phi) dt$ (alleviating the 20%–90% Λ CDM delay-time suppression of $t_{\text{del}} \sim 1\text{--}10$ Gyr hierarchical channels; Li et al., 2026; Paper 0 §8; Papers 12, 29, 30, while companion studies establish $T_{\text{CMB}}(z) = T_0(1+z)$, $(1+z)$ light-curve time dilation, pre-recombination acoustic horizons $r_s^{\text{TEP}}/r_s^{\Lambda\text{CDM}} = 0.999994$, and strong-lensing refraction; Papers 18, 19, 26, 27, 30).

Second, while nuclear star clusters sit locally in the high-acceleration Newtonian quadratic regime ($g_N \sim 10^{-7} \text{ m s}^{-2} \gg a_{\text{eff}}$), integrating the zero-parameter two-branch temporal-shear acceleration $a_\phi(r) = \sqrt{(g_N(r)/2)^2 + g_N(r) a_{\text{eff}}} - g_N(r)/2$ (and its low-acceleration limit $\sqrt{g_N(r) a_{\text{eff}}}$, $a_{\text{eff}} = 1.20 \times 10^{-10} \text{ m s}^{-2}$; Paper 0 §7; Paper 6) from $r_{\text{min}} = 0.1 \text{ pc}$ to an environmental/tidal truncation boundary r_{max} across concentric Plummer baryonic potentials $\Phi_N(r) = -GM_{\text{cl}}/\sqrt{r^2 + a_{\text{cl}}^2} - GM_{\text{gal}}/\sqrt{r^2 + a_{\text{gal}}^2}$ deepens the central escape speed $v_{\text{esc}}^{\text{TEP}}$ from baryons alone (evaluated against isotropic 3D Maxwellian GW-recoil kick distributions with 1D velocity dispersion $\sigma_{1\text{D}} = 120 \text{ km s}^{-1}$ for 1G + 1G and 220 km s^{-1} for 2G + 2G mergers), reproducing — and in dwarf and spiral hosts slightly exceeding — the GW-kick retention of standard abundance-matched NFW dark-matter halos in GR without invoking particle dark matter: in dwarf-host NSCs ($M_{\text{gal}} = 10^8 M_\odot$, $a_{\text{gal}} = 0.4 \text{ kpc}$, $M_{\text{cl}} = 10^6 M_\odot$, $a_{\text{cl}} = 2.0 \text{ pc}$, $r_{\text{max}} = 15 \text{ kpc}$), v_{esc} rises from 80.3 km s^{-1} (baryon-only GR) and 121.7 km s^{-1} (GR + $M_{200} = 10^{10} M_\odot$, $c_{200} = 12$ NFW halo: 1G + 1G retention 20.5%, 2G + 2G 4.1%) to $125.0\text{--}131.4 \text{ km s}^{-1}$ under baryon-only TEP ($117.9\text{--}131.8 \text{ km s}^{-1}$ across $[0.5 r_{\text{max}}, 2.0 r_{\text{max}}] = 7.5\text{--}30 \text{ kpc}$), boosting 1G + 1G Maxwellian GW-kick retention $3.16 \times\text{--}3.53 \times$ over baryon-only GR (7.0% $\rightarrow$ 22.1%–24.7%, or 19.0%–24.8% across $[0.5, 2.0] r_{\text{max}}$, matching Li et al., 2026's empirical GWTC-5 retention fraction $f_{\text{merger}} = 23_{-11}^{+17}\%$) and 2G + 2G retention $3.56 \times\text{--}4.09 \times$ (1.24% $\rightarrow$ 4.42%–5.09%); in field-spiral NSCs ($M_{\text{gal}} = 10^{10} M_\odot$, $a_{\text{gal}} = 2.0 \text{ kpc}$, $M_{\text{cl}} = 5 \times 10^6 M_\odot$, $a_{\text{cl}} = 2.5 \text{ pc}$, $r_{\text{max}} = 80 \text{ kpc}$), v_{esc} rises from 245.4 km s^{-1} (baryon-only GR) and 377.8 km s^{-1} (GR + $M_{200} = 3 \times 10^{11} M_\odot$, $c_{200} = 10$ NFW halo: 98.1% and 60.1%) to $378.0\text{--}402.7 \text{ km s}^{-1}$ (354.8–400.1 km s^{-1} across 40–160 kpc; 1G + 1G retention 75.7% $\rightarrow$ 98.4%–99.0%; 2G + 2G retention 25.8% $\rightarrow$ 60.8%–65.9%, a $2.36 \times\text{--}2.56 \times$ boost); and in massive AGN/galactic-core NSCs ($M_{\text{gal}} = 5 \times 10^{10} M_\odot$, $a_{\text{gal}} = 3.5 \text{ kpc}$, $M_{\text{cl}} = 2 \times 10^7 M_\odot$, $a_{\text{cl}} = 3.0 \text{ pc}$, $r_{\text{max}} = 150 \text{ kpc}$), v_{esc} rises from 424.5 km s^{-1} (baryon-only GR) to $600.1\text{--}641.7 \text{ km s}^{-1}$ ($567.8\text{--}631.3 \text{ km s}^{-1}$ across 75–300 kpc; 2G + 2G retention 70.7% $\rightarrow$ 94.4%–96.3%, matching GR + $M_{200} = 1.5 \times 10^{12} M_\odot$, $c_{200} = 8.5$ NFW halo at 632.7 km s^{-1} , 95.9%). All three external scales entering these population checks — the two-branch shear threshold $a_{\text{eff}} = 1.20 \times 10^{-10} \text{ m s}^{-2}$ (Paper 6), the host-transfer bracket $\epsilon_{\text{eff}} \simeq 0.24\text{--}0.52$ (Paper 0 §8), and the $\sim 135\text{--}150 M_\odot$ photodisintegration boundary (Paper 29) — are fixed prior corpus inputs rather than parameters fitted to GWTC.

Third, this nested-well hierarchy provides a candidate physical explanation for both the spectral compatibility of spin-selected 2G+ primary proxies and extreme upper-gap events such as GW231123 (GW231123_135430; $m_{1,\text{src}}^{\text{GR}} = 137.0 M_\odot$, $m_{2,\text{src}}^{\text{GR}} = 101.0 M_\odot$, $\chi_1 \simeq 0.9$, $\chi_2 \simeq 0.8$, $d_L^{\text{GW}} = 2200 \text{ Mpc}$, $z_L = 0.40$, $z_T = 0.578$): whereas GR places both GW231123 components inside the pair-instability gap where 2G + 2G cluster assembly has only $p \ll 1\%$ to 5% probability (Borchers et al., 2026) and compresses the more distant 2G+ primary proxies ($z_{L,\text{med}} = 0.500$ vs 0.370 for 1G proxies) 32.3% below the 1G remnant median — yielding a two-sample KS discrepancy on the selected catalog of $D_{\text{KS}} = 0.265$ ($p_{\text{KS}} = 7.98 \times 10^{-4}$, 3.35 σ equivalent; $D_{\text{KS}} = 0.384$, $p = 2.7 \times 10^{-5}$ [4.20 σ] on the 1-to-1 redshift-matched $N = 73$ vs 73 control; $\text{BC}_{\text{lin}} = 0.9740$, $\text{BC}_{\text{log}} = 0.9592$) and $\text{BC}_{40\text{--}120} = 0.8661\text{--}0.8775$ ($D_{\text{KS}} = 0.236\text{--}0.257$, $p = 0.059\text{--}0.097$) in the high-mass PISN window — the TEP mass map restores full-catalog spectral compatibility between the spin-selected generation proxies ($D_{\text{KS}} = 0.118$, $p_{\text{KS}} = 0.402$, $\text{BC}_{\text{lin}} = 0.9853$, $\text{BC}_{\text{log}} = 0.9837$ at $\hat{\epsilon}_{\text{joint}} = 0.25$, and $D_{\text{KS}} = 0.117$, $p_{\text{KS}} = 0.418$, $p_{\text{MW}} = 0.931$ at $\epsilon = 0$, holding across alternative spin thresholds $|\chi_{\text{eff}}| \geq 0.10$ and 0.20) as well as high-mass [40, 120] $M_\odot$ PISN-window compatibility ($\text{BC}_{40\text{--}120} = 0.9963$, $D_{\text{KS}} = 0.120$, $p = 0.892$ at $\hat{\epsilon}_{\text{joint}} = 0.25$, $p = 0.849$ vs GR 0.227 on the 1-to-1 redshift-matched control, and $\text{BC}_{40\text{--}120} = 0.9831\text{--}0.9930$ under $m_1 \leq P_{95}(m_2)$; [40, 130] $M_\odot$ medians $79.1 M_\odot$ vs $78.8 M_\odot$, $f_s = 0.364$ vs Li et al., 2026's $0.32_{-0.16}^{+0.25}$, shared 2G + 1G mass-ratio drop $q = m_2/m_1$ across $\tilde{m}_1 = 65 M_\odot$, $p = 0.00187$, Flanagan et al., 2026; and conditional two-component spin asymmetry partitioning 32.9% isotropic cluster vs 67.1% gas-aligned AGN-disk mergers while lifting 66.7% [8/12, median $79.5 M_\odot$] of anti-aligned $\chi_{\text{eff}} \leq -0.15$ systems into the $\geq 65 M_\odot$ gap vs 8.3% in GR, Álvarez-López et al., 2026) while mapping GW231123 to $\tilde{m}_1 = 215.0 M_\odot$, $\tilde{m}_2 = 158.5 M_\odot$ at $\hat{\epsilon}_{\text{joint}} = 0.25$ ($191.8 M_\odot$ and $141.4 M_\odot$ at $\epsilon = 0.33$; $302.7 M_\odot$ and $223.2 M_\odot$ at $\epsilon = 0$), cleanly above the $\sim 135\text{--}150 M_\odot$ photodisintegration boundary where rapidly rotating low-metallicity VMOs undergo direct core collapse (Paper 29; Croon et al., 2026) and where nested temporal wells boost 2G + 2G retention by $2.4 \times\text{--}4.1 \times$. Across the $N = 276$ GWOSC effective-spin catalog, individual binary 2G Bayes responsibilities $P(2G_i | \chi_{\text{eff},i})$ correlate event-by-event with $\tilde{m}_{1,i}$ (Spearman $\rho = +0.226$, $p = 1.49 \times 10^{-4}$ at $\hat{\epsilon}_{\text{joint}} = 0.25$; Mann–Whitney gap vs sub-gap $p = 0.0137$), and on the $N = 172$ high-redshift sample the fitted 2G+ broad-spin weight is $\hat{f}_{2\text{G}} = 48.7\%$ ([42.3%, 55.4%], vs 21.8% below $65 M_\odot$, $\Delta \ln L = +241.5$; Figure 3b), matching the matched $N = 172$ TEP gap occupation (44.2%–51.2% at $\epsilon = 0.25\text{--}0.33$) within 1σ while exceeding both GR's 13.4% gap proxy and the $\sim 10\%\text{--}20\%$ GR cluster-synthesis range. Likewise, within the precision-matched ($\sigma_\chi \leq \text{median}$) subsample, $\hat{f}_{2\text{G}}$ doubles from 10.8% ($z_L \leq 0.3$) to 22.1% ([13.6%, 32.5%], $N = 54$ at $z_L > 0.3$), where the matched $N = 54$ PISN-gap occupation is 24.1%–31.5% under TEP ($\epsilon = 0.33\text{--}0.25$; vs 57.4% at $\epsilon = 0$) compared with 3.7% (2/54) under the GR gap proxy (while the lower bound of [13.6%, 32.5%] overlaps the upper end of the $\sim 10\%\text{--}20\%$ total GR cluster-synthesis prediction). At the same time, because secondary components m_2 are much less affected by 2G+ hierarchical growth ($96.3\% \leq 45 M_\odot$ in GR coordinates; Tong et al., 2026; Antonini et al., 2026) — and because the TEP map

$\tilde{m}_2(\epsilon)$ at $\epsilon \in [0.25, 0.33]$ shifts the $42.0 M_\odot$ GR 95th-percentile secondary-mass tail to the canonical $\sim 65 M_\odot$ stellar nuclear pair-instability mass scale ($\tilde{m}_{2,90-95} = 63.1\text{--}69.5 M_\odot$ at $\epsilon = 0.33$) — the $\Delta \ln L = +11.02$ ($+11.18$ pure detector-frame cut at $\epsilon = 0$) and $+7.05$ ($\hat{\epsilon}_{\text{joint}} = 0.25$) preferences on $m_2 \geq 5 M_\odot$ demonstrate that the bi-metric mass-transfer scaling is favored even when hierarchical gap-filling is suppressed.

In the distance-redshift channel, the environmental host-luminosity split across the 71 photometry-bearing dark sirens ($H_0 = 72.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ in luminous massive hosts vs $69.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$ in diffuse faint hosts, and $70.1 \rightarrow 68.9 \text{ km s}^{-1} \text{ Mpc}^{-1}$ across candidate-pool weightings; Figure 1b) shows that standard L_r -weighted dark-siren pipelines preferentially weight luminous, massive catalog overdensities. Removing this weighting — or replacing Λ CDM host gating with the self-consistent TEP association kernel at fixed $H_0 = 68.18 \text{ km s}^{-1} \text{ Mpc}^{-1}$ — shifts the inferred expansion scale directly toward the Planck and TEP-H0/TEP-HC baseline ($67.4\text{--}68.18 \text{ km s}^{-1} \text{ Mpc}^{-1}$), consistent with both the catalog-selection de-biasing reported in MacBride et al. (DELVE Collaboration, 2026) and Cross-Parkin et al. (2026) and the expected TEP environmental screening contrast between deep halos and diffuse environments (Paper 11).

6.2 Limitations

The dark-siren distance channel remains conditioned on galaxy-catalog completeness, annular skymap coverage (median 6.3% in-cone probability), and Λ CDM-gated candidate ranking, which absorbs 85%–88% of any conformal distance shift. While the host-catalog-free BBH population channel avoids galaxy catalogs entirely, its baseline conditional log-likelihood scores use an empirical detection envelope $m_{\text{det,min}}(d) = A_{\text{env}} d^{6/5}$ and catalog point-mass medians; however, extending the analysis to a Mandel–Farr–Gair-style selection integral using the Finn & Chernoff (1993) analytic antenna-detection approximation retains a positive $m_2 \geq 5 M_\odot$ preference over stationary GR across both survival calibrations — with its magnitude depending on the adopted survival anchor under the pure inspiral scaling ($\Delta \ln L_{\text{MFG}} = +20.00$ at $\epsilon = 0$ and $+10.22$ at $\epsilon = 0.25$ for $\Theta_0 = 2.68$ [4.36%], vs $+8.41$ and $+8.62$ for $\Theta_0 = 2.34$ [10.0%]), while integrating the PhenomD Inspiral–Merger–Ringdown (IMR) waveform across the aLIGO noise PSD yields $\Delta \ln L_{\text{IMR-PSD}} = +14.20\text{--}+18.38$ at $\epsilon = 0$ and $+8.46\text{--}+10.22$ at $\epsilon = 0.25$ ($+2.87\text{--}+2.91$ vs $k = -1.0$) — a 40,000-draw 3-detector Hanford–Livingston–Virgo (H1–L1–V1) network antenna integral yields $\Delta \ln L_{\text{HLV}} = +8.62$ at $\epsilon = 0.25$ ($+1.91$ vs $k = -1.0$, vs $+6.34$ at $\epsilon = 0$) at the 10% anchor and $+19.77$ at $\epsilon = 0$ ($+11.89$ vs $k = -1.0$) or $+11.36$ at $\epsilon = 0.25$ ($+3.48$ vs $k = -1.0$) at the 5% anchor, an Essick (2023) GstLAL/PyCBC semi-analytic search-pipeline selection integral (incorporating H1–L1–V1 duty-cycle coincidence weights, matched-filter noise realizations $\sigma_\rho = 1$, and the PhenomD IMR-PSD response) yields $\Delta \ln L_{\text{pipe}} = +10.28$ at $\epsilon = 0.25$ ($+2.64$ vs $k = -1.0$, and $+11.41$ at $\epsilon = 0$, $+3.76$ vs $k = -1.0$) at the 10% anchor and $+10.72$ at $\epsilon = 0.25$ ($+3.46$ vs $k = -1.0$, and $+21.13$ at $\epsilon = 0$, $+13.87$ vs $k = -1.0$) at the 5% anchor — showing that as selection realism increases from step cut ($+7.05$) to Finn–Chernoff ($+8.62\text{--}+10.22$), 3-detector H1–L1–V1 ($+8.62\text{--}+11.36$), and GstLAL/PyCBC search-pipeline emulation ($+10.28\text{--}+10.72$), the $\epsilon = 0.25$ preference stabilizes rather than attenuates — and 200-draw hierarchical marginalization over the published GWOSC asymmetric 90% PE credible intervals with self-consistent per-draw redshift-threshold transformation $M_{\text{min},ik} = m_{\text{det,min},i} \text{ jac}_{ik}$ increases the $\epsilon = 0.25$ preference to $\Delta \ln L_{\text{MFG,hier}}^{\text{PE}} = +11.99$ to $+12.14$ under smooth selection ($+3.82$ to $+4.31$ vs $k = -1.0$), and $+21.60$ to $+28.52$ at $\epsilon = 0$) and $\Delta \ln L_{\text{hier}}^{\text{PE}} = +10.89$ at $\epsilon = 0.25$ ($+36.23$ at $\epsilon = 0$) on the step envelope (whereas holding $M_{\text{min},i}$ fixed at the point-estimate redshift across draws gives $+6.82$ and $+27.75$, or $+6.36$ and $+26.65$ correlated, and $+6.58$ to $+9.96$ at $\epsilon = 0.25$ when an unnormalized high-mass glitch-veto penalty is applied in place of the normalized Essick [2023] PhenomD IMR-PSD response), while sensitivity audits verify that the preference remains positive across pure detector-frame selection cuts ($\Delta \ln L = +11.18$), envelope exponents $d^{1.0}\text{--}d^{1.4}$ ($\Delta \ln L = +5.13$ to $+18.08$), envelope amplitude perturbations $[0.90, 1.20] A_{\text{env}}$ ($\Delta \ln L_{\text{MFG}} \in [+14.31, +22.25]$ under smooth Mandel–Farr–Gair selection, vs $[+0.79, +25.33]$ on the discontinuous hard step wall), 100-catalog synthetic GR injection challenges (0/100 false positives across stationary-plus-2G GR [mean -24.95 ± 26.83], low- Z wind evolution $\beta \propto (1+z)^{+0.15}$ [mean -35.82 ± 36.08], and mismatched antenna+IMR selection [mean -55.93 ± 30.27]), refitted redshift-evolving GR tapers $\beta(z_L) = \beta_0(1+z_L)^k$ ($\Delta \ln L \in [+1.11, +18.53]$ for $k \in [-3, +1]$, with $\Delta \ln L \geq +11.02$ for $k \geq 0$), and $q = m_2/m_1 \leq 1$ bounded Monte Carlo posterior resampling ($\Delta \ln L = +13.32^{+4.36}_{-4.49}$, 100% positive). The out-of-sample mass-taper likelihoods ($\mathcal{M}$ and m_2 , where (α, β) are refitted on the low- z training subsample for every tested ϵ) constitute the two primary population tests (secondary-mass and chirp-mass tapers, each profiling at most one tracking parameter $\epsilon \in [0, 1]$ across 10 grid points, so even a conservative effective trials penalty across the 10-point ϵ grid [$\ln 10 = 2.30$] or across the full $10 \times 2 \times 2 = 40$ -cell grid of ϵ , mass floors 5, $8 M_\odot$, and redshift splits 0.4, 0.5 [$\ln 40 = 3.69$] leaves the $\epsilon = 0.25$ smooth/pipeline/PE-marginalized preferences of $+10.28$ to $+12.14$ and the $\epsilon = 0$ preferences of $+11.02$ to $+36.23$ strongly positive, while the selection-integral variants test robustness on the same underlying catalog rather than independent hypothesis trials), whereas the spin-gap mixture, 1G-to-2G spectral inheritance, mass-ratio q drop, and discrete chirp-mass ladder are secondary consistency checks evaluated on the same mass map $\tilde{m}(z_T; \epsilon)$; the combined mass-taper and spin-gap score provides a conditional consistency test under the specified fraction-identification hypothesis and likelihood factorization, while a full multi-dimensional mass-spin injection campaign across raw O1–O4 strain injections, full LVK parameter-estimation posterior chains, and semiparametric multi-peak mass models (Tagliazucchi et al., 2026; Pierra et al., 2024; LIGO Scientific, Virgo, and KAGRA Collaborations, 2026b) remains an independent validation target. Host-association-free spectroscopic bright sirens across $z \in [0.1, 1.0]$ will provide the definitive distance test ($N_{5\sigma} = \lceil (5(\sigma_{d_L}/d_L^{\text{TEP}})/z)^2 \rceil = 25, 7, 3, 1$ events at $z = 0.1, 0.2, 0.3, 0.5$ under a 10% linear fractional distance error, or 28, 8, 4, 2 under a 10% log-distance error $\ln(1+z)$, at fixed $H_0 = 68.18 \text{ km s}^{-1} \text{ Mpc}^{-1}$; Figure 3c).

Finally, the canonical propagation law $\Xi(z) = (1+z)^{-1}$ and mass map $\tilde{m}(z_T; \epsilon) = m_{\text{det}}^{\text{GR}}(1+z_T)^{1-3\epsilon}$ are consistent with the stated multimessenger, Solar-System, and astrophysical benchmark constraints under the adopted radiation-zone and screening assumptions (Section 2.2): (i) in the far-zone radiation limit where disformal cone-tilt vanishes ($k^\mu \partial_\mu \phi \rightarrow 0$), the metric relation $\tilde{g}_{\mu\nu} = A^2(\phi)g_{\mu\nu}$ is purely conformal, preserving the null cone at all redshifts ($c_T \equiv c$, consistent with the GW170817/GRB 170817A bound $|c_T - c|/c < 3 \times 10^{-15}$; Abbott et al., 2017b) while rescaling the luminosity-distance ratio linearly as $d_L^{\text{GW}}/d_L^{\text{EM}} = A_{\text{em}}/A_0 = (1+z)^{-1}$; (ii) local Solar-System scalar accelerations are suppressed under Temporal Topology core saturation ($\rho_T \approx 20 \text{ g/cm}^3$; Paper 6) and two-branch high-acceleration kinetic screening ($g_N \simeq 5.9 \times 10^{-3} \text{ m s}^{-2} \gg a_{\text{eff}} = 1.20 \times 10^{-10} \text{ m s}^{-2}$ at 1 AU, giving $a_{\text{eff}}/g_N \simeq 2 \times 10^{-8}$ alongside screened Solar-System PPN and $\dot{G}/G$ benchmarks established in Paper 0 and Paper 6); and (iii) Pantheon+ Type Ia supernova standardization bounds ($\epsilon < 0.150$ at 3σ ; Paper 26) are reconciled with $\hat{\epsilon}_{\text{BBH}} = 0.25$ through the closed-form screening factor $S_{\text{WD}} = (\bar{\rho}_{\text{WD}}/\rho_T)^{1/3} \simeq 46.4$ inside Chandrasekhar white dwarfs ($\epsilon_{\text{SN Ia}} = \epsilon_{\text{host}}/S_{\text{WD}} \simeq 0.0043\text{--}0.0054 \ll 0.150$ for $\epsilon_{\text{host}} = 0.20\text{--}0.25$).

6.3 Relation to GWTC-5.0 Modified-Propagation Constraints

The LVK GWTC-5.0 cosmology analysis (LIGO Scientific, Virgo, and KAGRA Collaborations, 2026a, arXiv:2605.27227) analyzes 236 GW sources, reporting $H_0 = 71.0_{-7.1}^{+9.0} \text{ km s}^{-1} \text{ Mpc}^{-1}$ and constraining Belgacem et al.'s modified-propagation parameter Ξ_0 in $d_L^{\text{GW}}(z)/d_L^{\text{EM}}(z) = \Xi_0 + (1 - \Xi_0)/(1+z)^n$ to $\Xi_0 = 1.2_{-0.4}^{+0.8}$ to $1.4_{-0.4}^{+1.0}$ under a standard power-law + peak primary-mass model (where GR is $\Xi_0 = 1$ and canonical TEP $\Xi(z) = (1+z)^{-1}$ corresponds to $\Xi_0 \rightarrow 0, n = 1$, a boundary point that lies outside the published LVK prior range $\Xi_0 \in [0.435, 10]$ and is therefore excluded a priori in that parameterization). The external-cutoff diagnostic identifies a population-model and mass-transfer degeneracy that offers a quantitative mechanism for why standard LVK spectral-siren pipelines lean slightly above unity ($\Xi_0 \gtrsim 1.2$) while the out-of-sample taper test favors $\Xi(z) = (1+z)^{-1}$ on the same underlying catalog. In spectral-siren inference, the propagation ratio $\Xi(z)$ and the source-frame mass scale β are mathematically degenerate unless the bi-metric mass-transfer Jacobian is included (compare Ezquiaga & Holz, 2022; Chen, Gray & Baker, 2024). Under LVK's fixed- G mass conversion $\mathcal{M}_{\text{src}}^{\text{LVK}}(z) = \mathcal{M}_{\text{det}}/(1+z_{\Xi_0})$ with the mass feature anchored near $\sim 35 M_\odot$, lowering $\Xi_0 \rightarrow 0$ increases the distance-inferred redshift z_{Ξ_0} of every event across the 244-event cumulative BBH sample (median z_{Ξ_0} rises from 0.442 at $\Xi_0 = 1.0$ to 0.680 at $\Xi_0 = 0.0$) and systematically compresses LVK's inferred source masses downward (90th-percentile $\mathcal{M}_{\text{src},90}^{\text{LVK}}$ drops from $39.9 M_\odot$ at $\Xi_0 = 1.0$ to $29.8 M_\odot$ at $\Xi_0 = 0.0$, and 95th-percentile m_2^{LVK} drops from $42.1 M_\odot$ to $33.1 M_\odot$; Figure 3c inset). As shown in Table 2, artificially fixing the GR taper at $\beta = 35 M_\odot$ penalizes canonical TEP out of sample by $\Delta \ln L = -4.96$ at $\epsilon = 0$ and forces a fixed-cutoff GR fit toward $\Xi_0 \geq 1.2$ ($\mathcal{M}_{\text{src},90}^{\text{LVK}} = 40.7\text{--}41.1 M_\odot$). Jointly profiling the taper β on the low- z anchor sample and including the bi-metric Jacobian $(1+z_T)^{1-3\epsilon}$ removes this artificial penalty and reverses the out-of-sample preference to $\Delta \ln L = +2.32$ on chirp mass and $\Delta \ln L = +11.02$ on the secondary-mass m_2 taper less affected by hierarchical mergers.

7. Conclusions

This paper has tested the Temporal Equivalence Principle bi-metric gravitational-wave prediction across cumulative GWTC-1 through GWTC-5.0 catalogs using both the host-catalog-free BBH mass-and-spin population channel and the standard-siren distance-redshift channel. In the catalog-free BBH population channel (Table 2; Figure 3), breaking the scale-free power-law degeneracy with a pair-instability-motivated exponential mass taper yields an unclipped mass-only optimum at $\hat{\epsilon} = 0.25$ ($\Delta \ln L = +5.58$ on $m_2 \geq 8 M_\odot$, $+2.13$ on cumulative $\mathcal{M} \geq 8 M_\odot$, and $+8.62$ under the 10% Finn–Chernoff and 3-detector H1–L1–V1 integrals on $m_2 \geq 5 M_\odot$ with zero spin input, stabilizing at $+10.28\text{--}+10.72$ under Essick [2023] search-pipeline emulation and $+11.99\text{--}+12.14$ under self-consistent 200-draw GWOSC PE marginalization, with prior-averaged out-of-sample predictive log-evidence $\Delta \ln Z_{\text{marg}} = +7.95/+9.82$ vs $k = 0/\mathcal{U}[0, +1]$ and $+1.17 [3.2 : 1]$ vs $\mathcal{U}[-3, +1]$ under $\epsilon \sim \mathcal{U}[0, 1]$, and $+5.34 [208 : 1]/+7.21 [1, 350 : 1]$ on $m_2 \geq 5 M_\odot$ or $+4.80 [122 : 1]/+6.71 [824 : 1]$ on $m_2 \geq 8 M_\odot$ under $\epsilon \sim \mathcal{U}[0.24, 0.52]$ — rising on $m_2 \geq 8 M_\odot$ under the astrophysically motivated decreasing-power-law slope prior $\alpha \geq 1.0\text{--}1.5$ to $\Delta \ln L = +10.21\text{--}+10.66$ vs $k = 0$ [$+2.85\text{--}+3.45$ vs $k = -1.0$, $+8.75\text{--}+17.09$ Essick pipeline emulation, and $+5.68$ on $\mathcal{M} \geq 8 M_\odot$] and $\Delta \ln Z_{\text{marg}} = +8.55\text{--}+8.94$ [$5,167 : 1\text{--}7,631 : 1$] vs $k = 0$, $+10.36\text{--}+10.67$ [$31,571 : 1\text{--}43,045 : 1$] vs $\mathcal{U}[0, +1]$, and $+3.41\text{--}+3.99$ [$30 : 1\text{--}54 : 1$] vs $\mathcal{U}[-1, +1]$ [$+5.10\text{--}+6.07$ on $\epsilon \in [0.20, 0.33]$], and rising to $\Delta \ln Z_{\text{marg}}^{\text{mass+disp}} = +4.83$ [$126 : 1$] vs $\mathcal{U}[-3, +1]$ and $+9.19 [9, 788 : 1]$ vs $k = 0$ once combined with the $f_{2G} = f_{\text{gap}}$ -free held-out O4 spin-dispersion likelihood), while at the pinned boundary $\hat{\epsilon} = 0$ the step-envelope preference is $\Delta \ln L = +2.32$ on the 95-event chirp-mass benchmark, $\Delta \ln L = +4.06$ on the GWTC-5.0 secondary-mass slice ($N = 104$), and $\Delta \ln L = +11.02$ ($+11.18$ under a pure detector-frame boundary $m_{2,\text{det}} \geq 5.5 M_\odot$, $+5.13$ to $+18.08$ across selection-envelope exponents $d^{1.0}\text{--}d^{1.4}$, $\Delta \ln L_{\text{MFG}} = +20.00$ under the Mandel–Farr–Gair-style Finn–Chernoff analytic antenna-selection approximation ($+8.41$ to $+8.62$ at the 10% orientation anchor, and $\Delta \ln L_{\text{IMR-PSD}} = +14.20\text{--}+18.38$ at $\epsilon = 0$ or $+8.46\text{--}+10.22$ at $\epsilon = 0.25$ when integrated across the aLIGO IMR noise PSD), $\Delta \ln L_{\text{MFG,hier}}^{\text{PE}} = +11.99\text{--}+12.14$ at $\epsilon = 0.25$ ($+21.60\text{--}+28.52$ at $\epsilon = 0$, and $+10.89/+36.23$ on the step envelope, vs $+6.82/+27.75$ with fixed point-redshift thresholds) under 200-draw hierarchical marginalization over the published GWOSC asymmetric 90% PE credible intervals, $\Delta \ln L_{\text{pipe}} = +10.28\text{--}+10.72$ at $\epsilon = 0.25$ ($+11.41\text{--}+21.13$ at $\epsilon = 0$) under the Essick (2023) semi-analytic search-pipeline emulation, and 0/100 false positives across 100-catalog synthetic GR injection challenges (mean $\Delta \ln L = -24.95$ to -55.93 across stationary-plus-2G, low- Z wind evolution,

and mismatched antenna+IMR selection)) across the cumulative $N = 266$ secondary-mass ($m_2 \geq 5 M_\odot$) taper less affected by primary-mass hierarchical growth ($\Delta \ln L = +5.58$ at $\hat{\epsilon} = 0.25$ for $m_2 \geq 8 M_\odot$, $N = 226$, and retaining $\Delta \ln L \in [+1.11, +18.53]$ across all refitted redshift-evolving GR tapers $\beta(z) = \beta_0(1+z)^k$ for $k \in [-3, +1]$, with $\Delta \ln L \geq +11.02$ for $k \geq 0$). Across all $N = 276$ GWOSC BBHs with measured effective spins χ_{eff} , individual binary 2G Bayes responsibilities correlate event-by-event with primary mass ($\rho = +0.226, p = 1.49 \times 10^{-4}$ under TEP at $\hat{\epsilon}_{\text{joint}} = 0.25$; $\rho = +0.234, p = 8.5 \times 10^{-5}$ under GR; Mann–Whitney gap vs sub-gap $p = 0.0137$), and the two-component 1G+2G spin mixture assigns a broad-spin weight of $\hat{f}_{2G} = 48.7\%$ across the $N = 172$ events at $z_L > 0.3$ and 49.5% for $\tilde{m}_1 \geq 65 M_\odot$ (versus 21.8% below $65 M_\odot$, $\Delta \ln L = +241.5$), matching the matched $N = 172$ TEP PISN-gap occupation (51.2% at $\epsilon = 0.25$ and 44.2% at $\epsilon = 0.33$; 55.0% in the 43-event subset) within 1σ while exceeding GR's 13.4% (23/172), and doubling from 10.8% to 22.1% on the $N = 54$ precision-matched ($\sigma_\chi \leq \text{median}$) high- z subset where the matched gap occupation is 24.1%–31.5% under TEP versus 3.7% (2/54) under the GR gap proxy (and $\sim 10\%$ –20% total 2G+ fraction in GR cluster synthesis). Evaluating 1G-remnant to 2G-primary spectral compatibility (Li et al., 2026) and nested temporal-well confinement shows that whereas GR exhibits a full-catalog two-sample KS discrepancy between spin-selected 1G-remnant and 2G+-primary proxies of $D_{\text{KS}} = 0.265$ ($p_{\text{KS}} = 7.98 \times 10^{-4}$, 3.35σ equivalent; $D_{\text{KS}} = 0.384, p = 2.7 \times 10^{-5}$ [4.20 σ] on the 1-to-1 redshift-matched control; 2G+ primary proxies compressed 32.3% below 1G remnants; high-mass [40, 120] $M_\odot$ $\text{BC}_{40-120} = 0.8661\text{--}0.8775$), the TEP mass map restores full-catalog spectral compatibility between the spin-selected generation proxies ($D_{\text{KS}} = 0.118, p_{\text{KS}} = 0.402, \text{BC}_{\text{lin}} = 0.9853, \text{BC}_{\text{log}} = 0.9837$ at $\hat{\epsilon}_{\text{joint}} = 0.25$, and $D_{\text{KS}} = 0.117, p_{\text{KS}} = 0.418, p_{\text{MW}} = 0.931$ at $\epsilon = 0$) and raises high-mass [40, 120] $M_\odot$ PISN-window overlap to $\text{BC}_{40-120} = 0.9963$ ($D_{\text{KS}} = 0.120, p = 0.892$ at $\hat{\epsilon}_{\text{joint}} = 0.25$, and $p = 0.849$ vs GR 0.227 on the 1-to-1 redshift-matched control; 0.9831–0.9930 under $m_1 \leq P_{95}(m_2)$; [40, 130] $M_\odot$ medians $79.1 M_\odot$ vs $78.8 M_\odot$, $f_s = 0.364$), reproduces (and in dwarf and spiral hosts slightly exceeds) NFW-halo-level 1G + 1G and 2G + 2G GW-recoil retention from baryons alone in nested cluster wells ($2.4 \times 4.1 \times$ boost over baryon-only GR, $1 - A_{\text{loc}} \leq 2.0 \times 10^{-6}$; 22.1%–24.7% retention in dwarf-host NSCs vs 20.5% for GR + NFW and 7.0% for baryon-only GR, matching $f_{\text{merger}} = 23^{+17}_{-11}\%$, and remaining at 19.0%–24.8% across [0.5, 2.0] r_{max}), conditionally partitions non-bulk $|\chi_{\text{eff}}| \geq 0.15$ systems under a two-component spin-symmetry model into 32.9% isotropic cluster and 67.1% gas-aligned AGN-disk mergers while lifting 66.7% (8/12, median $79.5 M_\odot$, vs 8.3% in GR) of anti-aligned $\chi_{\text{eff}} \leq -0.15$ systems into the $\geq 65 M_\odot$ gap (Álvarez-López et al., 2026), and maps GW231123 ($\tilde{m}_1 = 215.0 M_\odot, \tilde{m}_2 = 158.5 M_\odot$ at $\hat{\epsilon}_{\text{joint}} = 0.25$) above the $\sim 135\text{--}150 M_\odot$ photodisintegration boundary. Under the conditional fraction-identification hypothesis ($f_{2G} = f_{\text{gap}}$), profiling ϵ jointly across the mass-taper and spin-gap channels converges on $\hat{\epsilon}_{\text{joint}} = 0.25$ ($\Delta \ln L_{\text{joint}} = +11.31$ vs $k = 0$ and $+6.55$ vs evolving GR at $k = -1.0$ on $m_2 \geq 5 M_\odot$, or $+14.48$ and $+7.61$ under Finn–Chernoff selection; and $+9.84$ vs $k = 0$ or $+5.04$ vs GR's $m_2 \geq 8$ interior optimum $\hat{k} = -1.0$ on $m_2 \geq 8 M_\odot$), sitting directly at the lower boundary of the parameterized field-spiral/dwarf host-transfer bracket ($\epsilon_{\text{eff}} \simeq 0.24\text{--}0.52$; exploratory discrete chirp-mass ladder coherence $R = 0.263$ vs 0.233), while an independent $f_{2G} = f_{\text{gap}}$ -free event-by-event spin-dispersion test trained on O1–O3 and scored on held-out precision-matched O4 ($N = 110$) likewise selects $\epsilon = 0.20\text{--}0.30$ ($\Delta \ln L_{\text{disp}}^{\text{O4}} = +4.37$ to $+5.62$, Levene $p = 0.0162\text{--}0.0067$; O4b $P(2G_i) p = 0.0305$; combined O1–O3 $\rightarrow$ O4 mass + event-level spin $\Delta \ln L_{\text{O4, ev}} = +11.32$ vs $k = 0$ and $+6.47$ vs $k = -1.0$ on $m_2 \geq 5 M_\odot$, and $+10.65$ and $+5.93$ on $m_2 \geq 8 M_\odot$), and joint taper profiling identifies a mass-transfer degeneracy that accounts for why fixed-cutoff spectral-siren pipelines lean toward $\Xi_0 \gtrsim 1.2$.

In the standard-siren distance-redshift channel across 73 non-circular sirens (1 bright siren GW170817 at $z = 0.0098$, where $1 - \Xi(z) = 0.97\%$ vs 15% distance error, plus 72 unassociated dark sirens; Table 1; Figures 1–2), 85%–88% Λ CDM-gate host absorption ($T \simeq 12\%\text{--}15\%$) leaves the dark-siren distance channel low-power on current data (a priori attenuated expectation $E[\Delta\chi^2_{\text{dark}}] \simeq +1.13$ vs observed $\Delta\chi^2 = +1.16$ at equal degrees of freedom) and reproduces the baseline fit ($-2 \ln L = 944.31$) without shifting H_0 under self-consistent TEP-kernel host reassociation at fixed canonical $H_0 = 68.18 \text{ km s}^{-1} \text{ Mpc}^{-1}$, while the 4D MCMC posterior on the perturbative template resolves the $(\phi_0, \beta) \leftrightarrow (-\phi_0, -\beta)$ factor-level sign non-identifiability via the reflection-invariant effective coupling $\phi_{0, \text{eff}} \equiv -\beta\phi_0 = +0.034 \pm 0.069$ (MAP $+0.039$, $\Delta\chi^2_{\text{MAP}} = +2.43$, 71.4% posterior mass on $A(z) < 1$). End-to-end skymap injections demonstrate that Λ CDM-gated host selection absorbs 85%–88% of any genuine conformal distance shift ($T \simeq 12\%\text{--}15\%$), and environmental host-luminosity splitting ($H_0 = 72.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ in luminous massive hosts vs $69.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$ in diffuse faint hosts; $70.1 \rightarrow 68.9 \text{ km s}^{-1} \text{ Mpc}^{-1}$ across pool weightings) is consistent with the direction of catalog de-biasing reported in MacBride et al. (DELVE Collaboration, 2026; $67.4^{+13.0}_{-11.2} \text{ km s}^{-1} \text{ Mpc}^{-1}$). Under the idealized TEP-as-true forecast convention at fixed canonical $H_0 = 68.18 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and 10% linear fractional distance error ($\sigma_{d_L}/d_L^{\text{TEP}} = 0.10$), host-association-free spectroscopic bright standard sirens across $z \in [0.1, 1.0]$ reach 5σ discrimination of $\Xi(z) = (1+z)^{-1}$ against General Relativity in $N_{5\sigma} = \lceil (5(\sigma_{d_L}/d_L^{\text{TEP}})/z)^2 \rceil = 25, 7, 3$, and 1 events at $z = 0.1, 0.2, 0.3$, and 0.5 (or 28, 8, 4, and 2 events under a 10% log-distance error $\ln(1+z)$).

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. (2026a). 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.
- [18] Riess, A. G., et al. (2024). HST Observations of Cepheid Variables in 37 Hosts of Type Ia Supernovae and a 2024 Determination of the Hubble Constant. *The Astrophysical Journal*, 977(1), 120. arXiv:2408.11770.
- [19] Belgacem, E., Dirian, Y., Foffa, S., & Maggiore, M. (2018). Modified gravitational-wave propagation and standard sirens. *Physical Review D*, 98(2), 023510.
- [20] Tong, H., Fishbach, M., Thrane, E., Mould, M., et al. (2026). Evidence of the pair-instability gap from black-hole masses. *Nature*, 652, 874–877. doi:10.1038/s41586-026-10359-0.
- [21] Tiwari, V. (2026). The Chirp-Mass Ladder in GWTC-5.0 Binary Black Holes. arXiv:2606.18081.
- [22] Cheng, A., Toubiana, A., Biscoveanu, S., & Gair, J. (2026). Non-parametric redshift evolution of binary black hole mass subpopulations in GWTC-4.0. Research Square preprint rs-9860193.
- [23] Cross-Parkin, R., Hernandez, I. M., Palmese, A., Lasky, P. D., & Thrane, E. (2026). Hostbusters: constraining H_0 and binary formation channels through gravitational-wave host galaxy associations. arXiv:2610.25866.
- [24] MacBride, T., et al. (DELVE Collaboration). (2026). Constraints on the Cosmic Expansion Rate Using the DECam Local Volume Exploration Survey and the Fifth Gravitational Wave Transient Catalog. arXiv:2610.25909.
- [25] Rodríguez, C. L., Zevin, M., Amaro-Seoane, P., Chatterjee, S., Kremer, K., Rasio, F. A., & Ye, C. S. (2019). Black holes: The next generation—repeated mergers in dense star clusters and their gravitational-wave properties. *Physical Review D*, 100(4), 043027.
- [26] Kimball, C., Talbot, C., Berry, C. P. L., Zevin, M., Thrane, E., Kalogera, V., et al. (2021). Evidence for hierarchical black hole mergers in the second LIGO–Virgo gravitational wave catalog. *The Astrophysical Journal Letters*, 915(2), L35.
- [27] Finn, L. S., & Chernoff, D. F. (1993). Observing binary inspiral in gravitational radiation: One interferometer. *Physical Review D*, 47(6), 2198–2219.
- [28] Mandel, I., Farr, W. M., & Gair, J. R. (2019). Extracting distribution parameters from multiple uncertain observations with selection biases. *MNRAS*, 486(1), 1086–1093.
- [29] Ezquiaga, J. M., & Holz, D. E. (2022). Spectral Sirens: Cosmology from the Full Mass Distribution of Compact Binaries. *Physical Review Letters*, 129(6), 061102. doi:10.1103/PhysRevLett.129.061102.

- [30] LIGO Scientific, Virgo, and KAGRA Collaborations (Abbott, R., et al.). (2026b). GWTC-5.0: Population Properties of Merging Compact Binaries. arXiv:2605.27226.
- [31] Tagliazucchi, M., Moresco, M., Borghi, N., & Ciapetti, F. (2026). Mind the Peak: Improving Cosmological Constraints from GWTC-4.0 Spectral Sirens Using Semiparametric Mass Models. *Astronomy & Astrophysics*, 709, A197. doi:10.1051/0004-6361/202558756.
- [32] Antonini, F., et al. (2026). Gravitational-wave constraints on the pair-instability mass gap and nuclear burning in massive stars. *Nature Astronomy*. doi:10.1038/s41550-026-02847-0.
- [33] Pierra, G., Mastrogiovanni, S., Perriès, S., & Mapelli, M. (2024). Study of systematics on the cosmological inference of the Hubble constant from gravitational wave standard sirens. *Physical Review D*, 109(8), 083504. doi:10.1103/PhysRevD.109.083504.
- [34] Agapito, M., De Renzi, V., & Mancarella, M. (2026). Gravitational-wave constraints on H_0 are robust to putative redshift evolution in the binary black hole mass spectrum at current sensitivity. *Physical Review D*. doi:10.1103/jk7y-ddyb.
- [35] Chen, A., Gray, R., & Baker, T. (2024). Testing the nature of gravitational wave propagation using dark sirens and galaxy catalogues. *Journal of Cosmology and Astroparticle Physics*, 2024(02), 035. doi:10.1088/1475-7516/2024/02/035.
- [36] Gray, R., et al. (2023). Joint cosmological and gravitational-wave population inference using dark sirens and galaxy catalogues. *Journal of Cosmology and Astroparticle Physics*, 2023(12), 023. doi:10.1088/1475-7516/2023/12/023.
- [37] Nishizawa, A. (2018). Generalized framework for testing gravity with gravitational-wave propagation. *Physical Review D*, 97(10), 104037. doi:10.1103/PhysRevD.97.104037.
- [38] Flanagan, E., et al. (2026). Transitions in the Mass-ratio and Spin Properties of Binary Black Holes in GWTC-5. arXiv:2606.14472.
- [39] Iampieri, L., et al. (2026). Hierarchical Population Inference with Normalizing Flows for Binary Black Holes. arXiv:2609.11885.
- [40] Liu, Y., et al. (2026). Probing the Mass–Redshift Dependence of Binary Black Holes and its Implications for H_0 with GWTC-5.0. arXiv:2609.25662.
- [41] Li, Y.-J., Wang, Y.-Z., Tang, S.-P., & Fan, Y.-Z. (2026). Smoking-gun evidence for hierarchical black-hole mergers. arXiv:2607.01121.
- [42] Álvarez-López, S., Heinzl, J., & Vitale, S. (2026). Evidence for additional structure in the effective spin distribution hints at multiple formation pathways in GWTC-5.0. arXiv:2606.12205.
- [43] Borchers, A., et al. (2026). Is GW231123 a Hierarchical Merger? *The Astrophysical Journal*, 999(2), 236. doi:10.3847/1538-4357/ae4358.
- [44] Croon, D., et al. (2026). Can GW231123 have a stellar origin? *Monthly Notices of the Royal Astronomical Society*, 546(3), stag073. doi:10.1093/mnras/stag073.
- [45] Woosley, S. E. (2017). Pulsational Pair-instability Supernovae. *The Astrophysical Journal*, 836(2), 244. doi:10.3847/1538-4357/836/2/244.
- [46] Farmer, R., Renzo, M., de Mink, S. E., Marchant, P., & Justham, S. (2019). Mind the Gap: The Location of the Lower Edge of the Pair-instability Supernova Black Hole Mass Gap. *The Astrophysical Journal*, 887(1), 53. doi:10.3847/1538-4357/ab518b.
- [47] Abbott, B. P., et al. (LIGO/Virgo/Fermi-GBM/INTEGRAL). (2017b). Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A. *The Astrophysical Journal Letters*, 848(2), L13. doi:10.3847/2041-8213/aa920c.
- [48] Essick, R. (2023). Semianalytic sensitivity estimates for catalogs of gravitational-wave transients. *Physical Review D*, 108(4), 043011. doi:10.1103/PhysRevD.108.043011.

TEP Research Series

- Smawfield, M. L. (2025). *Temporal Equivalence Principle: Dynamic Time & Emergent Light Speed*. Preprint v0.18 (Jakarta). Zenodo. DOI: 10.5281/zenodo.16921911 (Paper 0)
- Smawfield, M. L. (2025). *Temporal Topology Saturation Scale: Cross-Scale Consistency of ρ_T* . Preprint v0.11 (New Delhi). Zenodo. DOI: 10.5281/zenodo.18064365 (Paper 6)
- Smawfield, M. L. (2026). *The Cepheid Bias: Resolving the Hubble Tension*. Preprint v0.12 (Kingston upon Hull). Zenodo. DOI: 10.5281/zenodo.18209702 (Paper 11)
- Smawfield, M. L. (2026). *Temporal Equivalence Principle: A Unified Resolution to the JWST High-Redshift Anomalies*. Preprint v0.9 (Kos). Zenodo. DOI: 10.5281/zenodo.19000827 (Paper 12)
- Smawfield, M. L. (2026). *Temporal Equivalence Principle: Native hi_class Conformal Implementation, Linear Perturbation Closure, and CMB Acoustic Peak Preservation*. Preprint v0.8 (Cambridge). Zenodo. DOI: 10.5281/zenodo.20572722 (Paper 18)
- Smawfield, M. L. (2026). *Temporal Equivalence Principle: A Blind-Prediction Residual Test in Multiply-Imaged Supernovae*. Preprint v0.4 (Lisboa). Zenodo. DOI: 10.5281/zenodo.20572719 (Paper 19)
- Smawfield, M. L. (2026). *Temporal Equivalence Principle: A Standard-Siren Test of Bi-Metric Gravitational-Wave Propagation*. Preprint v0.6 (Maputo). Zenodo. DOI: 10.5281/zenodo.20572695 (Paper 22 — this work)

- Smawfield, M. L. (2026). *Temporal Equivalence Principle: A Covariant Alternative to Cosmic Expansion*. Preprint v0.5 (Athens). Zenodo. DOI: 10.5281/zenodo.20370143 (Paper 26)
- Smawfield, M. L. (2026). *Temporal Equivalence Principle: Nonsingular Temporal-Horizon Closure and Cosmological Perturbation Regularity*. Preprint v0.5 (Thika). Zenodo. DOI: 10.5281/zenodo.20723059 (Paper 27)
- Smawfield, M. L. (2026). *Temporal Equivalence Principle: Black Holes and the Temporal Horizon*. Preprint v0.4 (Bahrain). Zenodo. DOI: 10.5281/zenodo.21677826 (Paper 28)
- Smawfield, M. L. (2026). *Temporal Equivalence Principle: Dynamical Proper Time and the Illusion of Primordial Deuterium*. Preprint v0.6 (Dubai). Zenodo. DOI: 10.5281/zenodo.21841147 (Paper 29)
- Smawfield, M. L. (2026). *The Mount Wilson Paradigm: Restoring the Eternal Universe via the Temporal Equivalence Principle*. Preprint v0.4 (Harare). Zenodo. DOI: 10.5281/zenodo.21954257 (Paper 30)

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 synthetic-injection 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
- GraceDB public skymaps (bayestar.fits.gz): gracedb.ligo.org
- GLADE+ Galaxy Catalog (VizieR VII/291): glade.plus
- DESI DR1 redshift catalogs (zpix-main-dark + zpix-main-bright zcatalog monoliths, 13.4M good-galaxy spectroscopic redshifts): data.desi.lbl.gov
- NASA/IPAC Extragalactic Database redshift-bearing objects: ned.ipac.caltech.edu

Pipeline steps (25 sequential stages):

Table 3. Sequential pipeline execution stages (Steps 00–10).

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 zcatalog monoliths and build the HEALPix-indexed 13.4M-galaxy cache (optional, ~20 GB download)
02	<code>step_02_independent_redshifts.py</code>	Build independent-redshift dataset (bright + GLADE+/DESI DR1 dark sirens)
02b	<code>step_02b_host_association_systematics.py</code>	Host-association systematics: cone coverage, misassociation bound, concordance-bias injection, and environmental host-luminosity split
02c	<code>step_02c_desi_supplement.py</code>	Merge DESI DR1 candidates into stored host pools and rebuild posteriors
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 (distance-inferred redshifts are cosmology-blind; machinery null)
07f	<code>step_07f_per_galaxy_likelihood.py</code>	Per-galaxy hypothesis-consistent siren likelihood with multi-decoy and discriminating-subset nulls
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
09f	<code>step_09f_tep_c0_consistent_model.py</code>	Canonical endpoint map $\Xi(z) = 1/(1+z)$ split-normal distance likelihood, self-consistent TEP-kernel reassociation, and bright-siren 5σ sample-size forecast
09k	<code>step_09k_density_scaling_index.py</code>	Quartic density-tracking effective exponent ($n_{\text{eff}} = 0.982$) and adiabatic mass-to-Hubble ratio audit
09m	<code>step_09m_mass_map.py</code>	Canonical bi-metric mass-inversion map, matched PISN-gap occupation, 1G+2G spin mixture, held-out O4/O4b event-level spin-dispersion test (with redshift-only, detector-mass, and error-deconvolved controls), 1G-remnant to 2G-primary spectral inheritance (including published GWOSC NR remnant masses), non-bulk spin asymmetry, nested temporal-well GW-recoil retention (vs GR+NFW), GW231123 resolution, and joint mass+spin ϵ profile (catalog-free)
09n	<code>step_09n_mass_population.py</code>	Two-sided mass-population likelihood test: GR vs TEP bookkeeping, cumulative GWTC secondary-mass m_2 taper, pure detector-frame boundary audit, Finn–Chernoff, aLIGO IMR-PSD, 3-detector H1–L1–V1, and Essick (2023) search-pipeline selection integrals, 200-draw GWOSC PE marginalization, 100-catalog synthetic GR injection challenge, prior-averaged predictive evidence, chronological O1–O3 to O4 split, PowerLaw+Peak comparator, bidirectional GR evolution scan, chirp-mass ladder coherence, and Ξ_0 degeneracy
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.