

A search for boost-holonomy relics in the gravitational-wave catalog

Jonathan Schemoul

LibertAI, France

js@libertai.io

October 27, 2026

Abstract

A time machine’s unmanufacturable organ is a boost-holonomy identification: a locally flat region whose covering space is quotiented by a Lorentz boost. If such a relic exists, it leaves an exact, massless, frequency-multiplicative signature. A photon crossing the seam once emerges with its frequency multiplied by $r = \cosh \Lambda - n_z \sinh \Lambda$, exactly $e^{\pm \Lambda}$ for axial photons, giving a co-spatial spectral double image at a fixed, source-independent ratio and with no lens mass. For a gravitational wave the same holonomy scales the frequency of the whole waveform and therefore both inferred source-frame masses by one factor, leaving the mass ratio and χ_{eff} unchanged; for axial propagation the invariant transverse-traceless amplitude then forces the inferred luminosity distance to scale with the mass, so the invariant combination is M_c/D_L , not D_L itself, and the image is not magnified. Off axis the boost is not along the propagation direction, the amplitude is not invariant, and M_c/D_L is exact only at leading order. For the axial channel the two images are also co-spatial, while off axis the wound image is tilted by up to 27.5° . We execute the resulting cut chain—pair scaling, χ_{eff} (where available), the mass–distance scaling M_c/D_L (axial/leading-order), sky co-localization, posterior-level Λ , and the shared-base family test—on the public LVK data: the GWOSC catalog snapshot (391 events, 282 with source-frame masses), skymaps from GWTC-2.1, GWTC-3, GWTC-4.1, and GWTC-5.0 (281 of 282 events with a usable map after name matching), and full posterior samples for 16 events. Every pair-excess stage returns a null: the Stage-1 catalog count on 39,621 pairs at the one-degree-of-freedom 3σ cut is 32,434 against a label-permutation null median of 32,806 (two-sided $p \approx 0.10$, a mild deficit, not an excess); the corrected mass–distance scaling cut gives 30,343 candidates at 3σ against a matched independent-pair null median of 30,349 (two-sided $p \approx 0.96$), a null that recovers 5,688 pairs the archived equality- D_L cut discarded, and the 3– 6σ corrected sweep is likewise null; the corrected 51-event sky pilot gives $p = 0.75$ and the full volume $p = 0.95$ (largest candidate co-localization $C = 0.9513$). The corrected posterior-level analysis (16 events, 120 pairs) uses the same draw-matched pairing with the mass–distance scaling residual: 82 of 120 pairs survive, against 65 under the archived equality- D_L cut, and both named targets survive. The corrected cut is looser than equality, as it is at catalog level. The 120 include non-co-located pairs, and neither survival fraction is calibrated against a matched null. The two named targets are maximum candidates of their respective samples and are consistent with chance under the Stage-2 sky-permutation null, which prices the sample maximum; the remaining survivors do not have per-pair p -values. On the corrected co-located set, the catalog shared-base test returns a nominal $p = 0.005$ against the sky-permutation null, but a composition-matched null—holding the 3,669 co-located candidates’ $(\hat{\Lambda}, \sigma_\Lambda)$ composition and co-localization structure fixed—gives $p = 0.539$ (smooth composition models, 1.000; the assignment permutation is invariant, 1.000). The nominal excess is a broad low- $|\hat{\Lambda}|$ composition shift—75.0% of co-located candidates lie below $|\hat{\Lambda}| = 1.14$, against 71.9% of the candidate pool—not a shared-base family. The decisive methodological result is

that the family test is degenerate at current parameter-estimation precision: $\sigma_\Lambda = 0.08\text{--}0.5$, comparable to the Λ structure it must resolve, so the shared-base cluster over the corrected 120 pairs saturates (64 of 82 pairs, $p = 1.000$; 57 of 65 under the archived equality- D_L cut) and the winding-ladder diagnostic is exploratory and not part of the archived chain. We name two posterior-confirmed re-test targets, both maximum candidates consistent with chance, and state what the null does and does not establish: on the corrected co-located set, a sky-permutation-calibrated counting limit bounds the expected detectable-family yield to $\mu < 274$ at 90% CL at the operational co-localization threshold ($C > 0.1$; 205 under the Gamma-Poisson model, which is not adequate at $C > 0.1$ because the null is overdispersed), or $\mu < 4.50$ (4.95 conservatively) for the strong-co-localization $C > 0.9$ subset; the often-quoted $\mu < 2.3$ is only the ideal zero-observed case, not a data-driven bound. An injection/selection model gives $\varepsilon_{\text{family}} \simeq 1.8 \times 10^{-36}$, so the count limit implies $N_{\text{relic}} < 1.5 \times 10^{38}$ (1.1×10^{38} model-dependent) and no astrophysically meaningful density bound; the canonical metre-scale pocket is invisible to the pair channel because its 3.4 ns image delay is below the event-separation threshold, and the channel opens only for $\tau_{\text{wall}} \gtrsim 0.002\text{--}0.02$ AU. The survey strategy is the O5/full-O4 re-run, better mass precision, a hierarchical family fit, and the Gaia electromagnetic companion.

1 Introduction

Hawking’s chronology protection conjecture proposes that the laws of physics prevent the appearance of closed timelike curves [1]. The proposed mechanism is semiclassical: on a chronology horizon, null geodesics winding around a closed null geodesic pile up, virtual particles acquire a boost holonomy at each circuit, and the renormalized stress-energy tensor diverges. Explicit computations supported the picture on Taub–NUT space, wormhole time machines, and generic compactly generated horizons, and the Kay–Radzikowski–Wald (KRW) theorem put the compactly generated case on a rigorous footing [2, 3]. The theorem is nevertheless a statement about distributions on a fixed background, not about back-reaction, and it applies only to compactly generated horizons. Non-compactly generated machines—rotating cylinders, Krasnikov tubes, and the Ori–Soen family—are untouched by it [4].

The Ori program is the sharpest constructive challenge. Ori’s 2005 solution is a class of curved, asymptotically flat, topologically trivial vacuum spacetimes that develop closed timelike curves inside an empty toroidal region bounded by a non-compactly generated chronology horizon, with the surrounding matter satisfying the weak, dominant, and strong energy conditions [5]. The 2007 construction completes the initial data with a dust envelope [6], and the Xavier traveling-wave family gives a second explicit core [7]. What none of these constructions supplies is a classical manufacture [10] of the identification itself: the boost holonomy is the machine’s unmanufacturable organ. One can write the metric, but one cannot build the quotient by assembling ordinary matter.

That obstruction is also an opportunity. If the identification can be carried by a locally flat region—the $f = 0$ limit of the Ori core, or the $f \approx 0$ limit of the Xavier core—then a *bare relic* is a flat tube with a boost identification and nothing else. Such an object has no Newtonian mass, no local curvature, and no ordinary lens phenomenology, but it has an exact effect on any radiation that crosses its seam. The effect is multiplicative and source-independent: a frequency is multiplied by a fixed factor, and a whole spectrum is copied at that factor. This is the observational signature we search for. The question changes from “can the holonomy be manufactured?” to “is a relic already there?”.

This paper reports the gravitational-wave (GW) channel of that survey. The channel is attractive for three reasons. First, a GW’s frequency and its inferred source-frame masses are proportional, so the holonomy’s frequency scaling becomes a common scaling of both masses: two coalescences of one source, with the same mass ratio and spins, at distances that scale with the

masses (the invariant is M_c/D_L) and, for the axial channel, the same sky position, but with both masses shifted by $e^{\pm\Lambda}$. Second, a single pocket can write many such pairs, which is exactly what a family test needs. Third, the public LVK catalogs and skymap releases are large enough to execute the full cut chain today. The closest existing analyses are strong-lensing searches for multiple images, which use time-delay and magnification-ratio priors, including the interference-based search of [23] and the cluster-lens and waveform-systematics studies of [24, 25]; the holonomy pair is different in kind, because it is frequency-shifted and unmagnified. To our knowledge the shared-base family test across events has not been applied to the catalog.

The results are a null, and we state that plainly. At the one-degree-of-freedom 3σ catalog cut the pair test is saturated by chance: 32,434 candidates against a label-permutation null median of 32,806 (two-sided $p \approx 0.10$), a mild deficit rather than an excess. Replacing the archived equality-of- D_L cut by the correct mass–distance scaling cut gives 30,343 candidates at 3σ against a matched independent-pair null median of 30,349 (two-sided $p \approx 0.96$); the corrected cut recovers 5,688 pairs that the equality cut discarded and is still a null. The corrected 51-event sky co-localization pilot gives $p = 0.75$; the full-volume run with 281 mapped events gives a largest candidate $C = 0.9513$ and $p = 0.95$. The posterior-level analysis (16 events, 120 pairs) uses the corrected mass–distance scaling residual: 82 of 120 pairs survive it, against 65 under the archived equality- D_L cut, and both named targets survive both cuts; the 120 pairs include non-co-located pairs, and neither survival fraction is calibrated against a matched null. The two named targets are maximum candidates of their respective samples, and the Stage-2 sky-permutation null prices the sample maximum: both are consistent with chance. The null does not assign per-pair p -values to the other survivors. The paper’s most important result is not the null itself but the *measured* degeneracy of the family test: σ_Λ at catalog precision has a median of 0.18, so a 3σ window is ± 0.55 , comparable to the entire scan window $0.1 < |\Lambda| < 3$; at posterior precision σ_Λ is 0.08–0.5, still above the $\lesssim 0.05$ the test needs. The shared-base statistic is therefore dominated by the Λ distribution, and a lattice interpretation is not established. A composition-matched null confirms this: on the corrected co-located set, the catalog shared-base test’s nominal $p = 0.005$ falls to $p = 0.539$ when the co-located composition and co-localization structure are held fixed. We publish that as a methodological result, because it determines what the survey can and cannot claim.

The paper is organized as follows. Section 2 derives the signature in the flat relic limit, for photons and for GWs, and explains why it is not a lens. Section 3 defines the cut chain and the null controls. Section 4 describes the data. Section 5 presents the stage-by-stage results. Section 6 measures the family-test degeneracy. Section 7 records the two re-test targets. Section 8 states the limits honestly and lays out the survey strategy. Section 9 concludes. Appendix A lists the executable chain, and Appendix B gives the boost algebra.

2 The boost-holonomy signature

2.1 The relic as a flat boost quotient

The Ori 2005 core is a vacuum plane-wave geometry with a compact z -direction. In the $f = 0$ limit the (T, z) block has Riemann tensor identically zero; the core “is just the Misner space” [5, 8]. The same limit occurs in the Xavier core, where Riemann is zero exactly [7]. The covering is flat, and the identification $(T, z) \rightarrow (T, z + L)$ is a Lorentz boost of rapidity

$$\Lambda = \frac{L}{2} \tag{1}$$

in the Ori metric convention. The Milne dictionary gives $\Lambda = L/2$; no general uniqueness proof is claimed. An exact symplectic covering, a photon-transport check, and the Misner gate $\sigma_n = 4|T_0|(\cosh(nL/2) - 1)$ confirm it to 10^{-14} ; an earlier reading in which z itself was the rapidity ($\Lambda = L$) is wrong, because z is twice the covering boost angle up to a logarithm. The spectral cuts below use $\Lambda = L/2$.

A bare relic is this identification carried by a locally flat tube $\rho < \rho_0$, with no surrounding machine. It has no local curvature and no Newtonian mass. Its global holonomy is not a field that can be screened or shielded: a path that crosses the seam once returns with a Lorentz-transformed momentum. The wall matching to the exterior is supplied by the wrap in the full machine; in the relic survey we work in the flat ($a \rightarrow 0$, $f \approx 0$) limit and treat the seam as sharp. The canonical design used for the numerical examples below is $\rho_0 = 0.5$ m, $e = 0.1$, $a = 10^{-4}$, $\tau_0 = -1$ m², $L = 1$, hence $\Lambda = 0.5$.

2.2 Photons: the co-spatial spectral double image

Consider a photon that crosses the seam once. Its four-momentum is acted on by the inverse boost. For a boost along z with rapidity Λ and a photon direction cosine n_z along the boost axis, the frequency ratio between the wound and direct images is

$$r \equiv \frac{\nu_{\text{wound}}}{\nu_{\text{direct}}} = \cosh \Lambda - n_z \sinh \Lambda. \quad (2)$$

Because $n_z \in [-1, 1]$, the ratio lies in the interval $(e^{-\Lambda}, e^{+\Lambda})$. For photons running along the boost axis the boost does not redirect the light and the ratio is exact:

$$r(n_z = +1) = e^{-\Lambda}, \quad r(n_z = -1) = e^{+\Lambda}. \quad (3)$$

The source therefore shows a *co-spatial spectral double image*: one position on the sky, two copies of one spectrum at a fixed ratio $e^{\pm\Lambda}$. There is no mass, no deflection, and no source dependence; for the axial channel the same Λ applies to every pair from the same pocket. For $\Lambda = L/2 = 0.5$ ($L = 1$) the ratios at $n_z = -1, -\frac{1}{2}, 0, +\frac{1}{2}, +1$ are

$$r = 1.649, 1.388, 1.128, 0.867, 0.607, \quad (4)$$

so each pair comes in a blue ($e^{+\Lambda}$) and a red ($e^{-\Lambda}$) copy. These numbers are reproduced in Table 1.

Off-axis photons are additionally tilted, by up to $\arctan(\sinh \Lambda) = 27.5^\circ$ for $\Lambda = 0.5$; on-axis photons are not tilted at all. The observable is therefore chiefly spectroscopic, not astrometric. A multi-wound exit gives ratios $e^{\pm n\Lambda}$: integer powers of one fixed base, still a fixed-family signature. The in-tube Misner-phase sweep sets the winding number, $\Delta\psi \approx 2w(1 + n_z)/(n_x\tau_c)$ and $n \approx 2w/(\Lambda\tau_c)$ per unit axial alignment, where τ_c is the pocket clock at the crossing and w the chord half-width.

There is also a time echo. The wound copy's extra light path is of order the boost displacement

$$\Delta t_n \simeq 2\tau_{\text{wall}} \sinh\left(\frac{n\Lambda}{2}\right). \quad (5)$$

For the canonical metre pocket the wall Rindler radius is $r = \sqrt{e\rho_0^2 - t_0} = 1.01$ m, the Milne covering radius is $\tau_{\text{wall}} = 2r = 2.02$ m, the circuit length is $\Lambda\tau_{\text{wall}} = 1.01$ m $= L\sqrt{e\rho_0^2 - t_0}$, and the echo is 3.4 ns per circuit. An earlier pass fed r where the covering radius belongs and obtained 1.7 ns; the factor of two is the r -versus- $2r$ correction. For a large pocket, τ_{wall} can be astronomical and the echo can be days.

n_z	$r = \cosh \Lambda - n_z \sinh \Lambda$
-1	1.649 = $e^{+0.5}$
-1/2	1.388
0	1.128
+1/2	0.867
+1	0.607 = $e^{-0.5}$

Table 1: Photon frequency ratios for $\Lambda = L/2 = 0.5$ ($L = 1$). The axial values are exact; the intermediate values are the off-axis continuum. The wound copy is blue for $n_z = -1$ and red for $n_z = +1$, so a pair always contains one of each.

2.3 Gravitational waves: the mass-shifted pair

A gravitational wave crossing the seam is acted on by the same boost. For the axial channel the frequency of the whole waveform is multiplied by $r = e^{\pm\Lambda_{\text{hol}}}$; off axis it is multiplied by the same $r(n_z)$ of Eq. (2). The waveform’s frequency evolution is the clock from which the detector-frame chirp mass is inferred. A common scaling of all frequencies, $f \rightarrow rf$, $\dot{f} \rightarrow r^2 \dot{f}$, becomes a common scaling of the inferred masses, $M_c \rightarrow M_c/r$, so both masses are multiplied by the same factor. Write the common log-mass-ratio as

$$\Lambda = \ln \frac{m'_1}{m_1} = \ln \frac{m'_2}{m_2}, \quad m'_1 = e^\Lambda m_1, \quad m'_2 = e^\Lambda m_2, \quad (6)$$

so that $r = e^{-\Lambda}$; for the axial channel $\Lambda = \mp\Lambda_{\text{hol}}$. The mass ratio $q = m_2/m_1$ is invariant, and the dimensionless spins and χ_{eff} are invariant.

The transverse-traceless amplitude is invariant for axial propagation, so the axial image is not magnified. But $h \propto M_c^{5/3} f^{2/3}/D_L$: with h , q and χ_{eff} fixed, the frequency scaling $f \rightarrow rf$ and the mass scaling $M_c \rightarrow M_c/r$ force

$$\frac{D'_L}{D_L} = \frac{1}{r} = e^\Lambda = \frac{M'_c}{M_c}. \quad (7)$$

The two images are one source at one true distance, but the frequency scaling biases the inferred distance: a true holonomy pair lies on a mass–distance scaling degeneracy, not at a common D_L . The invariant combination is M_c/D_L (equivalently h) on axis, not D_L by itself. Off axis the boost is not along the propagation direction: the image aberrates by up to 27.5° and the amplitude acquires a direction-dependent factor, so Eq. (7) and M_c/D_L are exact only on axis and at leading order. An ordinary lens preserves the source frequencies and magnifies the flux instead; the archived GW cut that required equality of D_L therefore rejects true pairs and is a null on that cut, not on the scaling signature.

For the axial channel the two events are co-located on the sky, and the delay between them is the geometric echo of Eq. (5), not a lens time delay. Off axis the wound image is tilted by up to $\arctan(\sinh \Lambda) = 27.5^\circ$, which can exceed the better GW sky localizations; the sky co-localization cut is therefore sharpest for the axial channel and can miss strongly off-axis pairs.

The catalog reports source-frame masses, inferred with each event’s own redshift. The exact invariant is the detector-frame chirp mass over D_L , $M_{c,\text{det}}/D_L$ with $M_{c,\text{det}} = M_{c,\text{source}}(1+z)$. Using source-frame masses and the leading relation $D'_L \propto M'_c$ is a small $(1+z)$ -dependent approximation; the catalog-level analysis below uses it and the exact detector-frame invariant as a cross-check (Sec. 3).

The pair-internal test therefore does not require axial alignment: any crossing scales both masses by the same factor, so the statement $\Lambda_1 = \Lambda_2$ is exact for every pair. The shared-base family test

Observable	Electromagnetic	Gravitational wave
Frequency shift	$r = \cosh \Lambda - n_z \sinh \Lambda$	same r ; waveform frequency scaled
Mass shift	—	$m'_1 = e^\Lambda m_1, m'_2 = e^\Lambda m_2$
Mass ratio	—	q invariant
Spin	—	χ_{eff} invariant
Distance	same source, no magnification	$D'_L \propto M'_c$; M_c/D_L invariant
Sky position	co-spatial (axial), tilt $\leq \arctan(\sinh \Lambda)$	co-localized within the GW sky map
Time separation	echo $2\tau_{\text{wall}} \sinh(n\Lambda/2)$	same, geometric not lens-like
Family	$e^{\pm n\Lambda}$, one base	$\Lambda_1 = \Lambda_2$ per pair; shared base across pairs (axial)

Table 2: The signature dictionary. The exact content is the common multiplicative scaling; the absence of a lens mass and of a magnification anomaly is what separates the holonomy pair from a conventional lens.

across different pairs from the same pocket is sharper: it is exact for the axial channel, where the mass log-ratio has one magnitude $|\Lambda|$ (equivalently $r = e^{\pm\Lambda}$ with one holonomy rapidity Λ_{hol}), and it is broadened off axis, where r varies continuously between $e^{-\Lambda_{\text{hol}}}$ and $e^{+\Lambda_{\text{hol}}}$. This distinction matters in Sec. 6: the pair test is robust, the family test is a clean axial-channel statement.

Table 2 summarizes the dictionary between the electromagnetic and GW observables.

2.4 What the signature is not: no lens, no mass

The relic has no Newtonian mass. The direct image’s deflection, integrated through the exact metric, is independent of the coupling a across the whole design window $a \in (2 \times 10^{-17}, 0.1)$; for a small-drift test photon it is about 3×10^{-5} deg, and the e -coupling $\Gamma^x_{zz} = -ex$ shows that the effect is cone geometry, not wave and not mass. Any mass fitted to the pair’s astrometry is fake: a 27.5° pair at 1 kpc would fake $M \sim 5 \times 10^{45} \text{ kg} = 2.4 \times 10^{15} M_\odot$. The discriminator that no mass lens can fake is the fixed frequency ratio $e^{\pm\Lambda}$, and its integer-power ladder $e^{n\Lambda}$.

A conventional lens produces multiple images with the same frequency (the source redshift is the same), a magnification ratio, and a time-delay geometry tied to the lens mass. The holonomy pair has a frequency ratio, no magnification, and a delay set by the pocket’s wall radius. A frequency-shifted pair with no magnification anomaly therefore does not match any conventional lens model. At catalog level the magnification term is not available, so the survey’s mass–distance scaling cut, $D'_L \propto M'_c$, is the closest stand-in; at posterior level the absence of a magnification anomaly is a target-set expectation, not a fitted parameter. It is not applied as a separate Stage-3 cut because the catalog does not provide a per-event magnification posterior. The current LVK lensing searches test pair consistency in time delay, magnification and mass ratio [21]; the boost-holonomy pair is frequency-shifted and unmagnified, so the family test used here is a different cut.

The finite- a wave does not spoil the clean channel. Per-circuit transverse amplification is $x_{\text{end}}/x_0 = 1 + 2 \times 10^{-5}$ at $a = 10^{-4}$, 1.0043 at $a = 0.02$, and 1.0214 at $a = 0.1$: the once-wound channel survives the wave, while higher windings are progressively ejected [11, 12]. The clean signal is the $n = \pm 1$ pair, consistent with the a -floor logic of the machine design [12]. The integer ladder is a search channel, not a guarantee.

Astrophysical and waveform false positives are real: precessional modulation can mimic lensing fringes [26], and waveform systematics can mimic lensing support [25]. They do not, however, produce a shared base across independent pairs, so the family test is their falsifier in principle; at current precision the family test is degenerate (Sec. 6), so the current null is a pair-excess null.

Pocket ρ_0	Shell stars
0.5 m	3.8×10^{-51}
1 AU	1.6×10^{-17}
1000 AU	1.6×10^{-8}

Table 3: Stars in the entry shell of a bare relic. The OFF relic is invisible to a generic stellar survey; the channel opens only for large pockets or at the trigger epoch.

2.5 Causal cloistering and the two open windows

Why is the survey not dominated by nearby relics? The pre-chronology-horizon pocket world-tube is the Rindler band $0 < \tau < \tau_{\text{wall}}$. A null chord from a generic background star crosses the tube’s spatial location at covering radius $\tau \sim \tau_{\text{now}}$ (the covering age), far outside the band, and never enters. Entry requires the chord to graze the birth-cone null surfaces $W = \pm Z$ inside the tube, a shell of thickness $\sim r_{\text{wall}}$ (the Milne radius $2r_{\text{wall}}$ is the same order). With a volume $\sim \pi \rho_0^2 r_{\text{wall}}$ in the disk shadow and a stellar density of 0.004 stars/ly³, the shell star count is Table 3.

A pre-horizon bare relic is thus causally cloistered: it hides its holonomy until the chronology horizon opens it. This is the covering-space form of the program’s topological-censorship statement [9]. There are two windows in which light does enter. The first is *large pockets*, with τ_{wall} astronomically large, where the pair channel opens with the entry-shell cross-section. The second is *the trigger*: at $t > e\rho^2$ the z -circles turn timelike, winding is free of a phase budget, and the classical Misner pile-up opens—many images of one source at geometrically compressed frequencies crowding the horizon direction.

The event-rate accounting is sobering. The disk cross-section $\pi \rho_0^2$ at 1 kpc subtends 8×10^{-40} sr, so a metre-scale pocket covers about 10^{-31} Gaia-like stars even ignoring the entry shell; metre-scale relics are individually undetectable by stellar surveys. The channel becomes statistical only for $\gtrsim 1000$ AU pockets, and even the wrapped 10^{22} – 10^{26} kg dust at 1 kpc has an Einstein angle of only 1.4×10^{-4} – 1.4×10^{-2} mas (0.14–14 μas), far below the resolution of current optical astrometry. The wrapped machine is found by its pair, not by its lens. The GW channel is the only one with a large existing catalogue of individually parameter-estimated events in which a single pocket can produce repeated signature pairs, because one armed or large pocket can write many events. The electromagnetic large-pocket window is real but sparse (Sec. 2.5); it is the companion survey’s channel and does not yet supply a comparable catalogued pair sample. That is why the GW channel is the survey’s first front.

3 Cuts and statistics

3.1 The pair statistic

For an unordered pair of events (i, j) define the two log-mass ratios

$$\Lambda_1 = \ln \frac{m_{1j}}{m_{1i}}, \quad \Lambda_2 = \ln \frac{m_{2j}}{m_{2i}}. \quad (8)$$

A holonomy pair has $\Lambda_1 = \Lambda_2 = \Lambda$. The per-component mass uncertainties are taken from the catalog’s 90% credible intervals as $\sigma_m = (\text{halfwidth})/1.64$, and the log-ratio variances are

$$\sigma_{\Lambda_a}^2 = \left(\frac{\sigma_{m_{a,i}}}{m_{a,i}} \right)^2 + \left(\frac{\sigma_{m_{a,j}}}{m_{a,j}} \right)^2, \quad a = 1, 2. \quad (9)$$

The weighted mean and its uncertainty are

$$\hat{\Lambda} = \frac{w_1\Lambda_1 + w_2\Lambda_2}{w_1 + w_2}, \quad \sigma_\Lambda = (w_1 + w_2)^{-1/2}, \quad w_a = \sigma_{\Lambda_a}^{-2}, \quad (10)$$

and the consistency statistic is

$$\chi^2 = w_1(\Lambda_1 - \hat{\Lambda})^2 + w_2(\Lambda_2 - \hat{\Lambda})^2, \quad (11)$$

with one degree of freedom because one parameter Λ is fitted to two measurements. A pair is a candidate when

$$\chi^2 < \sigma_{\text{cut}}^2, \quad |\chi_{\text{eff}}^{(i)} - \chi_{\text{eff}}^{(j)}| < \sigma_{\text{cut}} \sqrt{\sigma_{\chi_{\text{eff}}^{(i)}}^2 + \sigma_{\chi_{\text{eff}}^{(j)}}^2}, \quad (12)$$

and, where both distances are defined,

$$\left[\ln \frac{D_{Lj}}{D_{Li}} - \hat{\Lambda} \right]^2 < \sigma_{\text{cut}}^2 \left[\sigma_\Lambda^2 + \left(\frac{\sigma_{D_{Lj}}}{D_{Lj}} \right)^2 + \left(\frac{\sigma_{D_{Li}}}{D_{Li}} \right)^2 \right], \quad (13)$$

with the scan window $0.1 < |\hat{\Lambda}| < 3$. The distance term is the corrected mass–distance scaling: with $\hat{\Lambda} = \ln(m_j/m_i)$, the D_L ratio must equal $e^{\hat{\Lambda}}$, equivalently M_c/D_L must be equal. For the frequency convention $f_j/f_i = e^{\Lambda_f}$ the same relation reads $\ln(D_{Lj}/D_{Li}) = -\Lambda_f$. The archived equality cut replaced Eq. (13) by $|\ln(D_{Lj}/D_{Li})| < \sigma_{\text{cut}}\sigma_{\ln D_L}$; it rejects true pairs and is retained below only as a comparison. The sweep uses $\sigma_{\text{cut}} = 3, 4, 5, 6$; the sky stages use 3. The mass χ^2 is a one-degree-of-freedom statistic, and Eq. (13) is a second one-degree-of-freedom constraint, so both use the quantile $\chi^2 < \sigma_{\text{cut}}^2$, i.e. $\chi^2 < 9.0$ at 3σ ; the two-degree-of-freedom value 13.8 is not used.

The 90% intervals are treated as symmetric halfwidth/1.64. This approximation is imperfect for skewed posteriors, but the identical treatment is applied in the null, so it calibrates out of the significance statement. The catalog-level scan is blind for $|\Lambda| < 0.1$ —small-holonomy pockets are hidden by mass noise—and the scan stops at $|\Lambda| = 3$, outside the natural machine window. The window is deliberately the machine window, not a generic prior.

For the sky stages the pair is scanned in both orientations and the orientation with the smaller σ_Λ is kept. The catalog stage uses $i < j$ and folds to $|\hat{\Lambda}|$, because the unordered pair has no preferred sign.

3.2 χ_{eff} and luminosity distance

The dimensionless effective spin $\chi_{\text{eff}} = (m_1\chi_1 + m_2\chi_2)/(m_1 + m_2)$ is invariant when both masses are scaled by the same factor and the spins are unchanged. A holonomy pair must therefore have consistent χ_{eff} . The catalog provides χ_{eff} and its 90% interval for most events; where it is missing, the cut is not applied for that pair. A mass lens does not change χ_{eff} either, so this cut is a consistency check, not a discriminator by itself.

The luminosity distance is the catalog’s last unused piece of holonomy information, but it is not a consistency variable: it scales with the inferred mass. A holonomy pair is one source seen twice; the pocket produces no magnification, so h is invariant, and $h \propto M_c^{5/3} f^{2/3}/D_L$ forces $D'_L \propto M'_c$. The catalog-level cut is therefore the scaling relation of Eq. (13), $\ln(D_{Lj}/D_{Li}) = \hat{\Lambda}$, equivalently M_c/D_L equal, at σ_{cut} . The archived equality cut $|\ln(D_{Lj}/D_{Li})| < \sigma_{\text{cut}}\sigma_{\ln D_L}$ asks for a common D_L and rejects true pairs; it is not the signature. Redshift is derived from D_L in the catalog, so using source-frame masses introduces a small $(1+z)$ correction; the exact detector-frame invariant $M_{c,\text{det}}/D_L$ is used as a cross-check. Pairs missing D_L are not tested on distance. At posterior level

the full D_L posterior is used. The absence of a magnification anomaly is a target-set expectation, but it is not one of the four implemented Stage-3 cuts (the catalog does not supply a per-event magnification posterior).

3.3 Sky co-localization

For two events with normalized sky posterior probability maps P_i , P_j , define the cosine co-localization

$$C_{ij} = \frac{\int P_i(\hat{n})P_j(\hat{n})d\Omega}{\sqrt{\int P_i^2 d\Omega \int P_j^2 d\Omega}} \in [0, 1]. \quad (14)$$

$C \rightarrow 1$ for a true pair (one source, one pocket position); $C \sim 0$ for chance pairs. GW 90% sky regions are 10^2 – 10^4 deg² against 41,253 deg² of sky, so chance co-location is suppressed by roughly $A/4\pi$ per pair. This is the discriminating power the catalog-level test lacks, and it is the reason the sky cut is Stage 2.

The maps are MOC/UNIQ HEALPix probability density maps. The 51-event pilot used dense NESTED maps at nside 512; the full-volume stage uses nside 256, a choice validated against nside 512 (C agrees to five decimals). The preferred map per event is the Mixed (combined) posterior; when only a single-model map is available, that map is used, because any PE sky posterior is valid for co-localization. This fallback matters for GWTC-5.0, which is mostly single-model maps.

3.4 Posterior-level Λ

At posterior level we draw-matched the two events' samples and form $\Lambda_1 = \ln(m_{1j}/m_{1i})$, $\Lambda_2 = \ln(m_{2j}/m_{2i})$, and $\Delta = \Lambda_1 - \Lambda_2$. A pair survives when

1. 0 lies in the 90% credible interval of Δ ;
2. 0 lies in the 90% credible interval of $\ln(D_{Lj}/D_{Li}) - \Lambda$ (the mass–distance scaling residual);
3. 0 lies in the 90% credible interval of $\chi_{\text{eff}}^{(j)} - \chi_{\text{eff}}^{(i)}$;
4. $0.1 < |\hat{\Lambda}| < 3$.

Draw matching avoids assuming Gaussian posteriors and avoids pairing the median of one event with the median of another as if the two were independent. We use 2×10^5 draws per pair. The surviving $\hat{\Lambda}$ and σ_Λ are the posterior-level measurements used in the family tests. The corrected Stage-3 run (`relic_survey_gw4_dlcorr.py`) applies item 2 to the reconstructed 16-event PE cache and reproduces the archived equality-cut run on the identical draws for comparison: 82 of 120 pairs survive the corrected scaling residual, 65 survive equality, 63 survive both. Using the exact chirp-mass ratio $\ln(M_{c,j}/M_{c,i})$ in place of Λ gives 81, and the detector-frame invariant $M_{c,\text{det}}/D_L$ gives 78; all three are looser than equality, consistently with the catalog stage. The archived equality quantity, 0 in the interval of $\ln(D_{Lj}/D_{Li})$, is retained only as the comparison.

3.5 The family tests

The pair test is necessary but not sufficient: a chance pair can pass it. The decisive discriminator in principle is the family structure; at current precision it is degenerate (Sec. 6), so the null reported here is a pair-excess null. A pocket writes many pairs with one shared base, or an integer ladder of bases. We implement two versions.

Shared base. For the surviving candidates, cluster $|\hat{\Lambda}|$ values within 3σ of a common base. The statistic is the maximum number of pairs in one cluster. The finite- a ejection program predicts that the once-wound channel dominates, so the sharp version uses $n = \pm 1$ only—one shared base, not a free integer lattice. A naive free-base integer-lattice version is degenerate by construction: with enough free integers and a tolerance, any set of ratios fits some pattern. It returns chance $p = 1.000$ by construction and is not used as evidence. This is not a technicality; it is the reason the sharp version must be specified before looking at the data.

Winding ladder. Three images at $n = 0, 1, 2$ give pairwise log-ratios $\Lambda, 2\Lambda, \Lambda$; in general a source seen at several windings gives an integer pattern in the pairwise log-ratios. The ladder test looks for such patterns within a tolerance. It is a search channel, not a guarantee, because the wave ejects higher windings and because a set of random ratios can fit several small rational patterns at the achievable precision.

3.6 Null controls

The nulls are matched to the stage they calibrate.

Label permutation. For the catalog-level count, permute the $(m_1, m_2, \chi_{\text{eff}}, D_L)$ labels among events. This preserves the marginal distributions and destroys the source pairing; the archived implementation holds the per-component mass errors at their event positions. We use 200 permutations for the archived equality-cut comparison. Because such a permutation keeps each event’s mass- D_L block together, it is not a valid null for the corrected mass-distance scaling cut; that cut uses the matched independent-pair resampling null described below.

Matched independent-pair null. For the corrected catalog count, draw as many unordered pairs as the catalog, with replacement, from the observed event sample. Each draw carries the event’s full $(m_1, m_2, \sigma_{m_1}, \sigma_{m_2}, \chi_{\text{eff}}, D_L, \sigma_{D_L})$ block, so the joint mass- D_L selection correlation is preserved exactly and only the physical pairing between two events is destroyed. This is the matched null for the corrected signature; we use 500 draws.

Sky-map permutation. For the co-localization stages, permute the sky maps among events and re-derive which candidates are co-located. This holds the full 30,311-candidate corrected pool’s $(\Lambda, \sigma_\Lambda)$ marginal fixed and randomizes the sky assignment; it does not hold the *selected* co-located subset’s composition fixed. We use 5,000 permutations for the corrected max- C statistic and 5,000 for the corrected family sky null; the archived equality-set family value was $p = 0.00065$ (13/20,000), and the corrected set gives $p = 0.005$.

Composition-matched bootstrap. For the catalog family statistic, hold the 3,669 corrected co-located candidates ($C > 0.1$), their C values, and their σ_Λ values fixed, resample the $(|\Lambda|, \sigma_\Lambda)$ pairs jointly with replacement (20,000 draws), and recompute the maximum shared-base multiplicity. This preserves the co-located composition and the co-localization structure and destroys only the Λ -to-edge assignment and any narrow shared-base spike not already part of the observed marginal. It is the primary matched null for the catalog family test; smooth KDE/Gamma composition models and an assignment permutation are sensitivity checks (Sec. 6).

Λ permutation. For the posterior family test, permute the survivors’ Λ values among the survivor slots and recompute the cluster multiplicity. We use 5,000 permutations.

An earlier family calibration used a Λ -marginal draw from the full candidate pool. That null was misspecified for the same reason the sky permutation is not composition-matched: on the archived set it reproduced the nominal excess ($p < 2 \times 10^{-4}$) from the composition difference alone; on the corrected set it gives $p = 0.119$ but is still not the matched null. Neither is used as evidence for a family. The composition-matched bootstrap gives $p = 0.539$ (Sec. 6).

4 Data

4.1 The catalog

The cached catalog is a GWOSC event API snapshot. It contains 391 events, of which 282 have source-frame masses. The catalog lineage is GWTC-1 [14], GWTC-2 [15] and its extended reanalysis GWTC-2.1 [16], GWTC-3 [17], GWTC-4.1 [18], and GWTC-5.0 [20, 19]. Of the 282 mass events, 1 is from GWTC-1, 54 from GWTC-2.1, 35 from GWTC-3, 88 from GWTC-4.1, and 104 from GWTC-5.0. The GWTC-5.0 catalog is introduced in [20]; its transient-identification methods are documented by the LVK Collaboration in [22]. The catalog provides source-frame masses and 90% intervals, χ_{eff} and its interval, luminosity distance and its interval, GPS time, network matched-filter SNR, and the catalog name. The 282 events define exactly 39,621 unordered pairs ($282 \times 281/2$); the archived code reproduces this count, and all catalog-level counts below use that set.

The catalog-level summaries are the only place where the mass, spin, and distance cuts can be applied to all 282 events. They carry no sky localization, which is why the sky stage needs the per-event posterior releases.

4.2 Skymaps

Four public skymap releases were ingested:

- GWTC-2.1: 54 event maps (51 Mixed plus 3 single-model fallbacks) from the IGWN GWTC-2.1 PE skymap release (IGWN-GWTC2p1-v2-PESkyMaps.tar.gz, 87 MB, Zenodo 6513631);
- GWTC-3: 36 event maps from the GWTC-3 PE sky-localization release (Zenodo 8177023), of which 35 match the 282-event catalog;
- GWTC-4.1: 88 event maps from the GWTC-4.1 archived skymaps (Zenodo 20275769);
- GWTC-5.0: 104 event maps from the GWTC-5.0 archived skymaps (Zenodo 20348005).

Skymaps were ingested from the four public releases. The releases contain 54, 36, 88, and 104 unique event tokens (282 total, with no token appearing in two releases). Name matching to the 282-event catalog gives 54, 35, 88, and 104 (281 total; the GWTC-3 token GW200105_162426 is not in the catalog snapshot), with zero pairwise overlap, so the executed union is 281. The single catalog event without a map is GW170817 (GWTC-1, absent from all four releases). Every sky-stage count in this paper uses the 281-event union. The GWTC-2.1 pilot used the 51 Mixed maps; the full-volume preferred-map rule adds the three events with only single-model maps, so its matched count is 54. The maps are MOC/UNIQU HEALPix with a PROBDENSITY layer and a line-of-sight distance layer; they are converted once to dense nside-256 maps and cached. The GWTC-5.0 release is mostly single-model maps; the fallback to any PE map is what allows the full volume to be used for co-localization.

Data set	Size	Use
GWOSC snapshot (GWTC-1–5.0)	391 events, 282 with masses	Stage 1, 1.5 pair scan
GWTC-2.1 skymaps	54 maps (51 Mixed + 3 fallback)	Stage 2 pilot/full volume
GWTC-3 skymaps	36 tokens (35 matched)	Stage 2 full volume
GWTC-4.1 skymaps	88 maps	Stage 2 full volume
GWTC-5.0 skymaps	104 maps	Stage 2 full volume
Name-matched union (executed)	281 of 282 events	Stage 2 full volume (source of record)
Full posteriors	16 events, 1.40 GB	Stage 3

Table 4: Data used. The four per-release rows are unique event tokens: $54+36+88+104=282$, with no token repeated across releases. After name matching to the 282-event mass catalog the counts are $54+35+88+104=281$ (the GWTC-3 token GW200105_162426 is absent from the catalog snapshot); the executed union is therefore 281, not 279. The single catalog event without a map is GW170817. The GWTC-2.1 pilot used the 51 Mixed maps; the full-volume preferred-map rule adds the three events with only single-model maps, giving 54. Every sky-stage count in this paper uses 281.

4.3 Posterior samples

Stage 3 used the full public posterior samples for 16 events covering the top ~ 30 co-located candidates: the two named re-test targets, clique A (GW231231/GW240104/GW230726/GW230814), clique B, and the secondary hubs. The 16 public LVK PE event files total 1.40 GB (Appendix A); the reconstruction is scripted by `pe_extract.py`. The exact Mixed group was used per event; two events required the SEOBNRv5PHM fallback family. GW200322 has only $n = 69$ draws in its extracted set, which is noted as a precision limitation. The extracted draw sets contain source-frame masses, χ_{eff} , and luminosity distance.

5 Results

5.1 Stage 1: the catalog pair scan

The catalog pair scan uses the 39,621 unordered pairs of the 282 events with source-frame masses (Sec. 4.1). The statistic minimizes over Λ , so it has one degree of freedom and the 3σ cut is $\chi^2 < 9.0$. It finds 32,434 candidate pairs. The label-permutation null (10,000 permutations) gives a median of 32,806 and a 90% range of [32,427, 33,135]; the one-sided excess probability is $p_{\text{excess}} = 0.948$ and the deficit probability is $p_{\text{deficit}} = 0.053$, so the two-sided p is ≈ 0.10 . The observed count is mildly low—about the 5th percentile of the null—not an excess. The largest Λ -family is one chained block of 32,434 pairs, which is the signature of a saturated statistic, not of a family.

The reason is quantified. With 10–20% mass precision per component, the per-pair Λ precision is 0.15–0.3. With 39.6k pairs, about 32k pass even a one-degree-of-freedom 3σ q -invariance test, distributed essentially uniformly across the window. Cut 1 alone is powerless against 32k chance pairs. The catalog summary carries no sky localization, so the no-deflector, no-ordinary-lens, and co-location cuts cannot be applied at this level. This recount converts the design lesson from a surmise into a measured calibration: at GWTC-5.0 volume, the pair test without co-localization carries zero excess significance.

σ_{cut}	archived equality D_L		corrected M_c/D_L scaling	
	observed	null median	observed	null median
3	24,655	24,364	30,343	30,349
4	29,311	29,206	32,513	32,513
5	31,508	31,614	33,276	33,276
6	32,860	33,070	33,753	33,756

Table 5: Stage 1.5 at 3–6 σ . The left null is the archived 200-permutation label permutation (central values permuted, per-component mass errors held at their event positions); it reproduces the archived equality-cut medians. The right null is the matched independent-pair resampling null (500 trials), which preserves the event-level mass– D_L selection correlation and destroys only the physical pairing. The corrected observed count is at the matched null median at every threshold; the two-sided p is 0.96 at 3 σ and 0.99–1.00 at 4–6 σ .

5.2 Stage 1.5: the mass–distance scaling cut

The catalog’s last unused holonomy variable is the luminosity distance, but it does not stay equal. A boost scales the frequency, hence the inferred mass, and h invariance forces $D_L \propto M_c$ (Eq. 7). The corrected cut is therefore $\ln(D_{Lj}/D_{Li}) = \hat{\Lambda}$, equivalently M_c/D_L equal, at σ_{cut} (Eq. 13). The archived equality cut asked for a common D_L ; it rejects true pairs and is a null on a weaker, wrong subset. The corrected cut is an axial/leading-order filter: off axis the boost is not along the propagation direction, the TT amplitude is not invariant, and the M_c/D_L relation is exact only at leading order; strongly off-axis pairs (tilt up to 27.5°) are a known loss. Table 5 gives both.

The corrected cut finds 30,343 candidates at 3 σ , 5,688 more than the equality cut. The matched independent-pair null gives a median of 30,349 (90% range [30,212, 30,491]) and a two-sided $p \approx 0.96$: the corrected cut is a null, and its count is not an excess. The old equality count 24,655 is also a null under the same matched null, but the archived equality-cut sweep (smallest one-sided $p = 0.025$ at 3 σ , two-sided ≈ 0.05 , 8-look-corrected ≈ 0.4) is a null on that cut, not on the scaling signature. The archived permutation null only varies the mass-error assignment while holding each event’s mass– D_L block together; it is not the null for the corrected signature.

The naive alternative is a trap. Permuting D_L (and its error) among events while keeping the masses fixed, or permuting the masses while keeping D_L fixed, destroys the astrophysical mass– D_L selection correlation (more massive binaries are detectable farther away). The corrected statistic then reads that correlation as a signal: the 3 σ null median drops to about 21,500 and the observed 30,343 becomes a nominal 40% excess. That is misspecification, not evidence; the matched independent-pair null of Table 5 is the appropriate calibration. The invariant cross-check agrees with the residual cut: M_c/D_L equal in source-frame form gives 30,377, and the exact detector-frame invariant $M_{c,\text{det}}/D_L$ gives 30,999 (2.2% higher from the $(1+z)$ factor); neither changes the conclusion.

The power caveat remains decisive. About 30k chance pairs pass the corrected cut, and even 100 true pairs move the total by 0.3%; catalog-level information is exhausted (masses, χ_{eff} , and the mass–distance scaling are all used). The corrected cut is a null with a measured, not supposed, calibration.

5.3 Stage 2 pilot: sky co-localization on GWTC-2.1

The pilot used the 51 GWTC-2.1 events with public MOC skymaps. Among them are 1,275 pairs, of which 959 pass the corrected Stage-1.5 scaling cut (741 passed the archived equality cut). The all-pairs co-localization has median $C = 0.0008$ and maximum $C = 0.6358$. The maximum- C pair

of the entire sample is also holonomy-consistent: GW190706_222641 / GW190719_215514, with masses $74.0 + 39.4 M_\odot \rightarrow 36.6 + 19.9 M_\odot$ (ratio $\approx e^{0.692} = 2.0$ for both components), $D_L = 3630$ vs 3730 Mpc, 13 days apart, $C = 0.636$, and $\Lambda = -0.692 \pm 0.302$. The corrected scaling residual is $+0.72$ (1.11σ): the pair passes only because its D_L errors are 40–50%, so this pilot pair does not distinguish the scaling relation from equality.

The corrected joint null (5,000 sky-map permutations) gives $p(\max C \geq \text{observed}) = 0.75$; the archived equality-cut pilot gave $p = 0.594$. The top coincidence is exactly what chance produces with 1,275 pairs on the table. The pair is recorded as a re-test target, not a detection.

5.4 Stage 2 at full volume

All four public skymap releases were ingested, and 281 of the 282 catalog events with masses are mapped. Among the mapped events there are 30,311 corrected candidate pairs (24,655 under the archived equality cut); the 32-pair difference from the catalog-level 30,343 is the set involving the unmapped GW170817. The all-pairs co-localization has median $C = 0.0061$ and maximum $C = 0.9658$ (not a candidate). The maximum candidate co-localization is

$$C = 0.9513 \quad (15)$$

for GW231231_154016 / GW240104_164932: masses $22.5 + 17.2 M_\odot \rightarrow 42.3 + 32.2 M_\odot$ (ratio $e^{0.629} = 1.88$), $D_L = 1070$ vs 1910 Mpc, observed ratio 1.79, 4 days apart. The corrected scaling residual is -0.05 (0.10σ): this pair is a strong corrected candidate, not merely a pair consistent with equal distance. If the pair were real, the delay of Eq. (5) would imply $\tau_{\text{wall}} \approx 6$ light-days, a 1000-AU-class pocket—again the large-pocket window that cloistering leaves open.

The corrected sky-permutation joint null gives $p(\max C \geq \text{observed}) = 0.95$. The archived equality-cut value was 0.860. Both are null. The top five candidates still show a hub structure, with GW231231_154016 and GW230726_002940 recurring as partners—the classic chance geometry. This is the full-volume version of the pilot result, and it is the null that the Stage-3 posterior analysis then tests pair by pair.

5.5 Stage 3: posterior-level Λ

The Stage-3 sample covers the top ~ 30 co-located candidates with 16 events. The Stage-3 script evaluates all $\binom{16}{2} = 120$ pairs among those events, not only the co-located subset: the 16 events were selected on co-localization, but the pair test itself is unrestricted. The corrected posterior-level pair test (draw-matched, survival = 0 in the 90% CI of $\Delta\Lambda$, χ_{eff} , and the scaling residual $\ln(D_{Lj}/D_{Li}) - \Lambda$) finds that 82 of the 120 pairs survive, against 65 under the archived equality- D_L cut applied to the identical draws. The corrected cut is looser than equality, as it is at catalog level: 63 pairs pass both, 19 only the corrected cut, and 2 only equality. Using the exact chirp-mass ratio in place of Λ gives 81 survivors and the detector-frame invariant 78 (Sec. 3.4). This survival fraction is not calibrated against a matched null: the family tests below are calibrated by permutation, but the count itself is not, so it is not by itself a statement about chance. Because the 120 pairs include non-co-located pairs, the survivors and the cluster below are not a co-located family sample. Both named re-test targets survive the corrected cut:

$$\begin{aligned} \text{GW190706_222641/GW190719_215514 : } \Lambda &= -0.682 \pm 0.214, \\ &\text{residual 90\% CI } [-0.450, +1.893], \end{aligned} \quad (16)$$

$$\begin{aligned} \text{GW231231_154016/GW240104_164932 : } \Lambda &= +0.620 \pm 0.101, \\ &\text{residual 90\% CI } [-1.011, +0.906]. \end{aligned} \quad (17)$$

Stage	Test	Result	Verdict
1	masses $+\chi_{\text{eff}}$ (where available), 39,621 pairs	32,434 vs null 32,806; two-sided $p \approx 0.10$	null (mild deficit)
1.5a	archived equality- D_L cut, $3-6\sigma$	24,655 vs 24,364 at 3σ ; smallest one-sided $p = 0.025$	null on the wrong cut
1.5b	corrected M_c/D_L scaling, $3-6\sigma$	30,343 vs matched null 30,349; two-sided $p \approx 0.96$	null (corrected)
2 pilot	sky C , 51 events	959/1,275 consistent; $p = 0.75$	null
2 full	sky C , 281 events	30,311 candidates; max $C = 0.9513$; $p = 0.95$	null
3	posterior Λ , 16 events (corrected scaling)	82/120 survive (65 under equality); family 64/82, $p = 1.000$	null (corrected); family degenerate

Table 6: Stage-by-stage summary. The Stage-1 row is the one-degree-of-freedom 3σ recount on the 39,621-pair catalog (Sec. 5.1); it is a mild deficit, not an excess. The archived equality- D_L row is a null on that cut, not on the stated signature; the corrected M_c/D_L row is the corrected result and is also null. All pair-excess stages are null; the Stage-3 row is now the corrected posterior scaling residual, with the equality count as a labelled comparison. The Stage-3 family test is degenerate rather than informative, neither survival fraction is calibrated against a matched null, and the 120 pairs are not restricted to co-located pairs.

The full-volume pair’s posterior residual is centred near zero (median -0.04), consistent with its catalog-level -0.05 (0.10σ); the pilot pair’s median is $+0.71$, consistent with its catalog-level $+0.72$ (1.11σ) but with errors too large to use the distance as a discriminator. The two named targets are maximum candidates of their respective samples, and the Stage-2 sky-permutation null prices the sample maximum: both are consistent with chance. The Stage-2 null does not assign per-pair p -values, so we do not claim that each of the other survivors is individually consistent with chance.

The family tests at posterior precision are the subject of Sec. 6. In brief, the shared-base cluster over the corrected 120 pairs puts 64 of 82 survivors in one cluster at $|\Lambda_0| = 0.467$ with a Λ -permutation chance $p = 1.000$ (57 of 65 under the archived equality cut), and a winding-ladder diagnostic was run in an exploratory session but is not implemented in the archived `gw4.py`, so no quantitative ladder result is used here. Both are saturated. The corrected chain is complete and null at every pair-excess stage—catalog, sky, and posterior. The family test is precision-limited.

6 The family-test degeneracy

6.1 What the test was supposed to do

The pair test can be faked by chance: at GWTC volume, tens of thousands of pairs pass a common-scaling cut. The family test is supposed to be the discriminator. A pocket writes many pairs with one shared base e^{Λ_0} (or an integer ladder), while chance pairs have no reason to share a base. If σ_Λ were much smaller than the spacing between unrelated Λ values, the shared-base multiplicity would be a sharp statistic. The question is whether the data have that precision. They do not, and we measured it rather than assuming it.

6.2 At catalog precision

The catalog-level σ_Λ has a median of 0.18. A 3σ window is therefore ± 0.55 , comparable to the entire scan window $0.1 < |\Lambda| < 3$. The candidate Λ distribution is concentrated at small $|\Lambda|$, so “shared-base multiplicity” is strongly shaped by the Λ CDF; a lattice interpretation is fragile. Two calibrations make this explicit.

First, the naive free-integer-lattice version returns chance $p = 1.000$ by construction: with free integers and a tolerance, any set of ratios fits some pattern. It is not evidence and is not used.

Second, the sharp single-shared-base version (per the finite- a ejection program, $n = \pm 1$ only) returns a nominal $p = 0.005$ against the sky-permutation null on the corrected co-located set (the archived equality-cut set gave $p = 0.00065$). That null holds the full 30,311-candidate pool’s $(\hat{\Lambda}, \sigma_\Lambda)$ marginal fixed, not the composition of the 3,669 *co-located* candidates it selects, so it asks whether the co-located composition is unusual relative to the pool, not whether the cluster multiplicity is unusual given that composition. We therefore computed a composition-matched null: the co-located candidates are held fixed, with their sky co-localization structure and σ_Λ values, and their $|\hat{\Lambda}|$ values are resampled jointly with σ_Λ (20,000 draws). The observed maximum, $S = 2,697$ pairs at $|\hat{\Lambda}_0| = 0.587$, is at the median of this null ($p = 0.539$; null median 2,699, 90% range 2,655–2,743). Smooth KDE and Gamma composition models give $p = 1.000$, and the assignment permutation that holds the $|\hat{\Lambda}|$ marginal and C structure fixed is invariant ($p = 1.000$). The full-pool draw gives $p = 0.119$, not the old $p < 2 \times 10^{-4}$; on the corrected set it no longer reproduces a strong nominal excess. The reason is composition: 75.0% of co-located candidates have $|\hat{\Lambda}| < 1.14$, against 71.9% of the corrected candidate pool. The shift is broad, and the 3σ window at $\sigma_\Lambda \approx 0.18$ spans roughly $0.1 < |\hat{\Lambda}| < 1.14$, so the statistic reads the shift as multiplicity. The sharp catalog test therefore does not survive a composition-matched null: it is a measured degeneracy, not a family detection. The matched null is a pair-excess null, not a no-family theorem—it cannot separate a family whose pairs are already part of the co-located $|\hat{\Lambda}|$ marginal.

The statistic also has little power at catalog precision. On the archived set, a synthetic shared-base family of 200 pairs at $\Lambda_0 = 0.587$ is detected 33% of the time and 500 pairs 100%, while a family at $\Lambda_0 = 2$ is invisible up to 1,000 pairs because the low- $|\hat{\Lambda}|$ CDF cluster dominates the maximum. A targeted base known a priori does much better—the oracle local test detects 50 pairs at $\Lambda_0 = 2$ with 100% power—but no such base is available. The matched null is calibrated on real non-candidate sets: on the archived set it gives $p = 0.534$ for non-co-located candidates and $p = 0.591$ for pairs failing the holonomy cut, so it does not manufacture a family on real non-candidate data.

6.3 At posterior precision

Stage 3 was specified to fix this: joint-posterior Λ of the top co-located candidates, with $\sigma_\Lambda \lesssim 0.05$. The measured posterior precision is $\sigma_\Lambda = 0.08\text{--}0.5$, not $\lesssim 0.05$. The shared-base cluster again saturates: under the corrected scaling residual, 64 of 82 survivors—over all 120 pairs among the 16 events, not only the co-located subset—fall in one cluster at $|\Lambda_0| = 0.467$, and the Λ -permutation chance calibration gives $p = 1.000$. The archived equality cut gives 57 of 65 at the same base and $p = 1.000$. A winding-ladder diagnostic was run in an exploratory session but is not implemented in the archived `gw4.py`; no quantitative ladder result is used here. Both family statistics are saturated. The corrected posterior cut is looser than equality (82 vs 65 survivors), so it enlarges the saturated cluster rather than resolving it; the degeneracy is a property of the precision, not of the cut.

6.4 The root cause

The root cause is parameter-estimation mass precision. The public PE masses are accurate at roughly 10–15% per component, which gives $\sigma_\Lambda \approx 0.1$ or larger per pair. The holonomy structure to be resolved is of order $\Lambda \sim 0.1$ –1. The measurement error is comparable to the signal. This is not a flaw in the survey design; it is a property of the current data. The family test becomes sharp only if

1. a pocket writes many pairs, so the multiplicity grows faster than the chance background; or
2. mass precision improves beyond current PE, so $\sigma_\Lambda \ll \Lambda$; or
3. an independent channel (an electromagnetic counterpart at the same position with the corresponding spectral ratio) supplies the missing discrimination.

The degeneracy is therefore a measured closure condition, not a null result about relics. We state it plainly because it determines what the survey can claim: the null is a pair-excess null, not a no-family theorem.

6.5 The methodological lesson

Two lessons are worth recording. First, a family statistic must be specified with its precision in mind: a test that looks decisive ($p = 0.005$) can be a composition artifact, and a matched null is needed to say which. Here the composition-matched bootstrap removes the nominal excess ($p = 0.539$), and the residual effect is a broad $|\hat{\Lambda}|$ CDF shift, not a shared base. Second, the null must be matched to the *selected* candidate set. The sky-map permutation holds the full candidate pool’s $(\Lambda, \sigma_\Lambda)$ marginal and randomizes the sky; the Λ -marginal pool draw does not preserve the co-located composition either and gives $p = 0.119$, not the matched value. Only the composition-matched bootstrap conditions on the co-located set, and it returns $p = 0.539$. Both lessons are general to population searches for exact multiplicative signatures.

7 The two re-test targets

Table 7 records the two posterior-confirmed targets. Both are maximum candidates and are consistent with chance at the Stage-2 null; neither is a detection. They are the pairs to re-test when the O5 catalog, better mass precision, or an electromagnetic counterpart arrives.

What would elevate a target? A second, independent pair at the same sky position sharing the same base (or an integer multiple) would be the family signature. A posterior $\sigma_\Lambda \lesssim 0.05$, with a measured base shared across pairs, would make the family test sharp. An electromagnetic counterpart at the same position with the corresponding spectral ratio would be decisive, because the EM and GW channels have independent systematics. A time delay and magnification pattern inconsistent with any lens mass would also help; the corrected posterior scaling residual is now applied and both targets pass it. At catalog level the full-volume pair is the one whose distance actually tracks its mass; the pilot pair’s D_L errors are too large to use the distance alone as a discriminator. The honest status is that both targets are plausible chance coincidences in the current data.

	GW190706_222641 / GW190719_215514	GW231231_154016 / GW240104_164932
Masses ($M_\odot$)	$74.0 + 39.4 \rightarrow 36.6 + 19.9$	$22.5 + 17.2 \rightarrow 42.3 + 32.2$
Ratio	$\approx e^{0.692}$	$e^{0.629}$
D_L (Mpc)	3630 / 3730	1070 / 1910
Separation	13 days	4 days
C	0.636	0.9513
Catalog Λ	-0.692 ± 0.302	$\approx +0.629$
Corrected residual	$+0.72$ (1.11σ)	-0.05 (0.10σ)
Posterior Λ	-0.682 ± 0.214	$+0.620 \pm 0.101$
Posterior residual	$+0.71$; 90% CI $[-0.45, +1.89]$	-0.04 ; 90% CI $[-1.01, +0.91]$
Stage-2 max- C pricing	sample $p = 0.75$ (pilot)	sample $p = 0.95$ (full)

Table 7: The two posterior-confirmed re-test targets. Both pass the corrected posterior consistency cuts, including the mass–distance scaling residual; the archived equality- D_L cut is retained only as a comparison (both also pass it). At catalog level the full-volume pair’s distance actually tracks its mass (residual -0.05 , 0.10σ), while the pilot pair passes only because its D_L errors are large (residual $+0.72$, 1.11σ). Both are consistent with chance under the sky-permutation null for the sample maximum. The full-volume pair’s 4-day separation, if real, would imply $\tau_{\text{wall}} \approx 6$ light-days, a 1000-AU-class pocket.

8 Limits and survey strategy

8.1 What the null establishes

The null is a null at every pair-excess stage: no excess of common-scaling pairs at catalog level, no co-localized excess at full volume, and no anomalous posterior maximum candidate. The catalog shared-base test’s nominal $p = 0.005$ is not a detection either: a composition-matched null gives $p = 0.539$, and the nominal excess is a broad low- $|\hat{\Lambda}|$ composition shift. It is not a theorem that relics do not exist. What it establishes is conditional and useful: a relic population that would write a large number of detectable, posterior-confirmed, co-located, shared-base pairs would have produced an excess over the sky-permutation null. Under the calibrated counting assumptions of Sec. 8.2, the quantitative statement is $\mu < 274$ at 90% CL (the robust empirical value; 205 under the Gamma-Poisson model, which is not adequate at $C > 0.1$) on the expected number of detectable families at the operational $C > 0.1$ threshold, or $\mu < 4.50$ (4.95 conservatively) for the $C > 0.9$ strong-co-localization subset. This disfavors populations whose expected detectable-family yield exceeds 274. A one-pair-per-pocket population is precisely the family case constrained by this bound: it is excluded if its expected detectable yield exceeds 274. What is not excluded is a population whose expected detectable yield lies below the bound—for example because the per-relic efficiency is small or the holonomy is too small to resolve. The sample maximum candidates are consistent with chance at $p = 0.75$ (pilot) and $p = 0.95$ (full volume); these p -values price the maximum- C statistic, not the observed counts. The family statistics are saturated or null (the catalog shared-base matched null gives $p = 0.539$; the posterior cluster gives $p = 1.000$), so a population whose family structure is washed out by the measured precision is not excluded by the family test alone; the counting bound still applies to its detectable co-located pairs.

The catalog-level count-statistic bound is weak by construction. About 30k chance pairs pass the corrected cut; a population of 100 true pairs changes the total by 0.3%, which is invisible. The co-localization cut is what gives the survey its power, and the family test is what would make it decisive in principle. The latter is limited by mass precision and by the co-located composition at

fixed precision; the composition-matched null gives $p = 0.539$, so the nominal catalog excess is not a detection (Sec. 6).

8.2 The calibrated counting limit and its assumptions

A full hierarchical Bayesian population limit is not available. We therefore calibrate the counting limit against the measured sky-permutation background and then fold in an explicit injection/selection model for the family efficiency.

The ideal zero-observed case is the reference point. Suppose the survey had an expected yield μ of detectable relic families—one family per relic, counted if at least one of its catalogued pair images passes the full cut chain—and observed zero with negligible background. The Poisson upper limit would be

$$\mu < -\ln(0.1) = 2.3 \quad \text{at 90\% CL.} \quad (18)$$

This is the placeholder used in earlier versions. The measured data have nonzero counts at every operational cut and a measured background, so Eq. (18) is not a data-driven bound.

To replace this placeholder, we calibrate the chance background of the co-localization statistic with 5,000 sky-map permutations of the 281-event mapped catalog. Of the 30,311 corrected Stage-1.5 candidates, 3,669 have $C > 0.1$ (permutation background 3,544), 96 have $C > 0.5$ (96.4), 4 have $C > 0.9$ (4.6), and 1 has $C > 0.95$ (1.5); no cut shows a significant excess (the smallest one-sided p is 0.062 at $C > 0.1$, 0.58 at $C > 0.5$, 0.82 at $C > 0.9$; the six-cut scan correction gives about 0.37). Writing the count as Poisson($b + s$) with b constrained by the permutation null and a flat prior on the true pair yield s , the 90% upper limits are $s_{90} = 205$ under the Gamma-Poisson model and 274 under the conservative empirical envelope at $C > 0.1$; the $C > 0.1$ null is overdispersed (var/mean = 2.17), so the Gamma-Poisson model is not adequate there and we take 274 as the primary robust value (205 is model-dependent). At the stronger cuts the null is sub-Poisson (underdispersed; var/mean = 0.48, 0.32, and 0.23 at $C > 0.5$, 0.9, 0.95), so the Poisson/Gamma-Poisson limits there are conservative, and the limits are 17.3 (20.9 conservatively) at $C > 0.5$, 4.50 (4.95) at $C > 0.9$, and 3.10 (3.23) at $C > 0.95$. Because every full-chain family has at least one pair above the operational Stage-2 threshold $C > 0.1$, the first value bounds the expected detectable-family yield: $\mu < 274$ at 90% CL (205 model-dependent), not 2.3. The value 2.3 is recovered only in the ideal zero-observed limit (with negligible background assumed) and is not a data-driven bound. The stricter cuts bound only the subset of families with a pair above the cut; under the idealization that every relic produces at least one counted pair ($\epsilon = 1$), $N_{\text{relic}} < 274$ ($C > 0.1$; 205 model-dependent) or < 4.50 ($C > 0.9$; 4.95 conservatively), while $\epsilon = 0.1$ weakens these to $< 2,740$ (2,050) and < 45 (49.5). The relic bound is dominated by the unmeasured efficiency; a stronger all-family bound requires the injection model of Sec. 8.3 or a sharper family statistic. The result is a count in the surveyed catalog volume, not a density.

Write $\mu = \epsilon_{\text{family}} N_{\text{relic}}$, where N_{relic} is the number of relics in the surveyed volume and ϵ_{family} is the probability that a relic produces at least one catalogued pair family that passes the full cut chain and is counted. The number k of pairs a relic writes does not enter the yield of families; it enters the multiplicity and the chance background. The calibration replaces the crude assumptions of zero observed families and negligible background with a measured count and a permutation-calibrated background. What remains assumed is: (i) one relic contributes at most one detected family, and the counted families are posterior-confirmed, co-located, and shared-base; (ii) the pair-count model Poisson($b + s$) with the Gamma-Poisson background; and (iii) ϵ_{family} is a single effective efficiency. The Stage-3 82/120 corrected survivor fraction (65/120 under equality) and the family-test degeneracy are not efficiencies.

A one-pair-per-pocket population is precisely the family case constrained by the counting bound: it is excluded if its expected detectable yield exceeds $\mu \approx 274$. What is not excluded is a population whose expected yield lies below the bound, for example because $\varepsilon_{\text{family}}$ is small or the holonomy is too small to resolve. The stricter-cut limits (17.3 at $C > 0.5$, 4.50 at $C > 0.9$) apply only to the subset of families that produce a pair above those cuts.

8.3 The injection/selection model: no relic-density bound

We model $\varepsilon_{\text{family}}$ explicitly as an archived-cut order-of-magnitude model: the efficiency was computed for the archived equality- D_L cut and is not re-derived for the corrected cut. A relic of proper radius ρ_0 at angular-diameter distance D_A subtends the solid-angle fraction $\rho_0^2/(4D_A^2)$; only sources behind that shadow can be paired. The entry-shell cloistering of Sec. 2.5 suppresses the wound channel by $\tau_{\text{wall}}/\tau_{\text{cov}}$, and the image delay $2\tau_{\text{wall}} \sinh(\Lambda/2)/c$ must exceed the event-separation threshold for the two images to be catalogued separately. Averaging over a log-uniform $\rho_0 \in [0.5 \text{ m}, 1000 \text{ AU}]$, a comoving-volume distance prior to $z = 1.17$, a log-uniform $|\Lambda| \in [0.1, 3]$, and the 281-map A_{90} distribution, the fiducial efficiency is $\varepsilon_{\text{family}} = 1.8 \times 10^{-36}$ (expected counted pairs per relic $k = 1.8 \times 10^{-36}$), dominated by the geometric shadow and the cloistering. The corrected 90% counting limit then gives $N_{\text{relic}} < 1.5 \times 10^{38}$ (the robust 274) or $< 1.1 \times 10^{38}$ (205, model-dependent), and $n_{\text{relic}} < 6.5 \times 10^{35} \text{ Gpc}^{-3}$ (4.9×10^{35}) at $C > 0.1$: no astrophysically meaningful density bound. Conditional on a catalog source lying behind the pocket, $\varepsilon = 3.5 \times 10^{-14}$ and $N_{\text{relic}} < 5.9 \times 10^{15}$ (205; 7.8×10^{15} for 274); even with cloistering switched off, $\varepsilon = 1.7 \times 10^{-24}$ and the bound is only $N_{\text{relic}} \sim 1.6 \times 10^{26}$ (1.2×10^{26} for the 205 Gamma-Poisson value). The bound is dominated by the unmeasured pocket-size/entry-shell prior, which spans ~ 24 orders of magnitude; the Λ window, the co-localization probability, and the mass function are $O(0.1\text{--}1)$ and are not dominant. The canonical metre-scale pocket is invisible to the GW pair channel because its image delay ($\sim 3.4 \text{ ns}$) is far below the event-separation threshold; the channel opens only for $\tau_{\text{wall}} \gtrsim 0.002\text{--}0.02 \text{ AU}$. The GW pair channel therefore cannot constrain the relic density; a meaningful bound requires a measured pocket-abundance prior or a channel with a larger per-relic cross-section.

8.4 Power and the multi-pair requirement

The survey's sensitivity scales with the number of pairs per pocket. One pair is indistinguishable from chance at these volumes: even the $C = 0.9513$ full-volume pair is consistent with chance at $p = 0.95$ because the sky maps are broad and the candidate set is large. A pocket that writes several co-located pairs with one base is a different object: the family multiplicity grows with k , while the chance background grows more slowly. This is the physical reason the GW channel is attractive despite the null: a single armed or large pocket can write many events, and the catalog is large enough that a sufficiently high-multiplicity family would stand out. This is a qualitative expectation about multiplicity; the selection model of Sec. 8.3 computes the per-relic detectable-pair yield, not the family-multiplicity distribution. The current data do not contain such a family at the achievable precision.

8.5 The strategy

O5 and the full O4 releases. The mapped sample will grow from 281 events toward the full O4/O5 catalog, and the sky maps will improve. The pilot's $p = 0.75$ and the full-volume $p = 0.95$ must be re-run at the new volume. The Stage-2 machinery is data-limited, not compute-limited: the dense-map conversion is cached during a run and the pair scan is vectorized. A larger catalog

increases both the chance background and the sensitivity to a multi-pair pocket; the family test is the statistic to watch.

Better mass precision. The family test needs $\sigma_\Lambda \ll \Lambda$. Current PE gives $\sigma_\Lambda \approx 0.1\text{--}0.5$; the target is $\lesssim 0.05$. Longer signals, higher SNR, better waveform models, and joint PE of candidate pairs all help. Next-generation detectors would help more. Until then, the family test should be reported with its measured σ_Λ , not as a decisive null.

A hierarchical population fit. The counting limit is now calibrated, and Sec. 8.3 supplies a first injection/selection model; the next analysis should replace both with a hierarchical likelihood: a population of relics with a Λ distribution and a pair-multiplicity distribution, folded through the selection function, fitted to the candidate set. The model’s dominant inputs are the pocket-size and entry-shell priors, which span ~ 24 orders of magnitude and are not measured. A hierarchical fit over those priors is the only way to turn the count bound into a density bound, and it is the paper’s main open computational item.

The Gaia electromagnetic channel. The EM channel is a companion survey. Gaia DR3 [13] has 33.8 million RV sources; a 1-arcsec RV–RV self-join on tiles is tractable at the 60 s sync cap, and the full-sky census is a many-hour background run. The chance co-located pair yield is about 3,400 full-sky, before the astrometry match; the astrometry cut (parallax and proper motion consistent at 3σ) is what no line-of-sight superposition gives. The EM channel probes the small- Λ window $\Lambda \sim 10^{-4}\text{--}1.6 \times 10^{-3}$, complementary to the GW window $0.1 < |\Lambda| < 3$. The full-sky census is now complete: 16,358 pairs, 7,217 astrometry-matched (44.1%), 939 candidates (matched, $|\Delta\text{RV}| > 30$ km/s and $> 3\sigma$), a matched-pair ΔRV distribution with median 7.4, 90% 41.6, 99% 126.8 and maximum 489.5 km/s, and a family statistic whose 173-of-939 maximum is the expected smooth density (the tolerance is a half-width, not a cluster), the same CDF-dominated degeneracy measured here. No holonomy signal is claimed. The decisive EM test is the epoch-RVS spectrum: a holonomy shift is constant across epochs and exactly multiplicative across features, while a binary’s ΔRV varies with orbital phase. That test is data-limited on Gaia DR3, which exposes mean RVS spectra but no epoch RVS spectra: only 9 of the 1,132 candidate sources in the EM-2 snapshot (the final census has 939 candidates) have a mean RVS, none has a published epoch series, and no candidate pair has both spectra. The follow-up is targeted multi-epoch spectroscopy of those candidates, or a future Gaia epoch-spectrum release. The companion paper reports the full census and the targeted follow-up. We mention it here because the two channels are independent hardware and independent systematics, and a coincident candidate would be much harder to explain than either alone.

Other channels. Rubin/LSST and OGLE/Gaia alert streams can search for transient pairs and repeated flares at fixed positions with fixed flux ratios and no microlensing light curve. VLBI pair imaging can target the large-pocket window (two compact components, identical structure, no mass surface density). GW microlensing wave-optics programs can snag an $e^{n\Lambda}$ ladder if it appears in the detector band [27]. These are secondary to the catalog and Gaia channels but they share the same falsifier in principle: the fixed ratio and the family base.

8.6 Falsifiers

The survey’s logic is falsification-first, and the falsifiers are worth stating explicitly.

1. *Family falsifier.* A candidate pair whose ratio is not an exact integer power of a base shared with other pairs from the same field would kill the holonomy interpretation for that candidate. This is the GW channel’s decisive test in principle; at current precision it is degenerate (Sec. 6), so the current null is a pair-excess null, not a family-exclusion.
2. *Mass falsifier.* A candidate pair whose astrometry fits a point-mass lens and lacks the spectral offset is an ordinary lens. The bare relic has no Newtonian mass, so a fitted lens mass is evidence against the relic interpretation for that pair.
3. *Population falsifier.* A full Gaia/Rubin accounting with zero valid pairs bounds the local number density of OFF pockets with $\tau_{\text{wall}} \gtrsim 0.1$ ly (the only shell-populated bare window); metre-scale bare relics are unfalsifiable by this channel, so the survey must target the trigger epoch or large pockets.
4. *Trigger falsifier.* Absence of the pile-up cascade signature in transient and GW surveys bounds the armed-natural-machine density.

The GW channel executes the first falsifier in principle; at current precision the family test is degenerate, so the execution is a pair-excess null. The EM companion will execute the mass and population falsifiers at stellar-survey scale.

8.7 What a detection would look like

A detection would require all of: (i) a co-located pair with posterior-confirmed $\Delta\Lambda = 0$, the mass–distance scaling M_c/D_L , and χ_{eff} ; (ii) a second pair at the same sky position sharing the same base or an integer multiple; (iii) no lens mass or magnification pattern that fits the data; (iv) a time delay consistent with a pocket wall radius, not a lens model; and ideally (v) an electromagnetic counterpart with the corresponding spectral ratio. The family condition is the decisive one in principle; at current precision it is degenerate. A single pair, however clean, is not enough at current precision.

9 Conclusion

The boost-holonomy identification is the part of a time machine that cannot be manufactured classically [10]. If a primordial relic carries it, the signature is exact and simple: a fixed multiplicative frequency ratio, a common scaling of both masses in a GW pair, a distance that scales with the inferred mass (M_c/D_L invariant), no lens mass, and no magnification (with exact sky co-location only for the axial channel; off axis the image is tilted by up to 27.5°). We executed the search for that signature on the public GW data. The cut chain—masses, χ_{eff} where available, the mass–distance scaling M_c/D_L , sky co-localization, posterior Λ , family structure—runs end to end on 281 mapped events and 16 posterior sets. The catalog-level pair test is saturated; the corrected mass–distance cut gives 30,343 candidates against a matched independent-pair null median of 30,349 (two-sided $p \approx 0.96$), and the corrected sky pilot and full-volume run give $p = 0.75$ and $p = 0.95$. The archived equality- D_L chain is null too, but it is a null on that cut, not on the stated signature. The posterior maximum candidates are consistent with chance, while the corrected 82/120 survival fraction (65/120 under equality; over all 120 pairs, not only co-located ones) is not calibrated against a matched null.

The paper’s central methodological result is the measured degeneracy of the family test. At catalog precision σ_Λ has a median of 0.18, so the 3σ window is comparable to the scan window; at

posterior precision $\sigma_\Lambda = 0.08\text{--}0.5$, still above the $\lesssim 0.05$ the test needs. The shared-base cluster is therefore saturated, the winding-ladder diagnostic is not archived, and the sharp catalog test’s nominal $p = 0.005$ does not survive a composition-matched null ($p = 0.539$; smooth composition models 1.000; the assignment permutation is invariant, 1.000). The nominal excess is a broad low- $|\hat{\Lambda}|$ composition shift, not a family detection. We publish the degeneracy because it is the honest limit of the current data and the design target for the next analysis.

The null is not a verdict on relics. The calibrated counting limit of Sec. 8.2, with a sky-permutation background, bounds the expected detectable-family yield to $\mu < 274$ at 90% CL at the operational co-localization threshold $C > 0.1$ (205 model-dependent), or $\mu < 4.50$ (4.95 conservatively) for the strong-co-localization $C > 0.9$ subset; the ideal zero-observed value $\mu < 2.3$ is not a data-driven bound. The injection/selection model of Sec. 8.3 gives $\varepsilon_{\text{family}} \simeq 1.8 \times 10^{-36}$, hence $N_{\text{relic}} < 1.5 \times 10^{38}$ (1.1×10^{38} model-dependent) and no astrophysically meaningful density bound. A one-pair-per-pocket population is the family case constrained by this bound: it is excluded if its expected detectable yield exceeds 274, while a population below the bound—or one whose holonomy is too small to resolve—is not excluded. The canonical metre pocket is invisible to the pair channel because its ~ 3.4 ns image delay is below the event-separation threshold. The next steps are the O5/full-O4 re-run, better mass precision, a hierarchical family fit over the pocket-size and cloistering priors, and the Gaia electromagnetic companion. The survey has moved from “can the holonomy be built?” to “is a relic already there?”; the first GW answer is no at the current sensitivity, with the limits and the degeneracy now measured rather than supposed.

A Reproducibility and the executable chain

The analysis chain is in `simulations/`:

Stages 1 and 1.5 reproduce from the cached catalog alone with `relic_survey_stage1.py` and `relic_survey_gw.py`. The skymap stages, the counting limit, and the selection model require the public releases listed in Sec. 4; the dense-map conversion is cached as `.npy` files when the releases are present. The composition-matched family null of Sec. 6, on the corrected M_c/D_L set, is the `--family` run of `relic_survey_gw_dlcrr.py`; the older `relic_survey_family_null.py` is the historical equality- D_L implementation, and the companion provenance note records both, with the corrected values in its header. The script caches the candidate set under a gitignored skymap cache and writes a gitignored results file; the archived sky null is reproduced with seed 43 and 500 trials before the 20,000-trial matched run. The cached catalog has 282 mass events and exactly 39,621 unordered pairs; the Stage-1 recount finds 32,434 at $\chi^2 < 9.0$. The corrected Stage-1.5 scaling cut, its matched independent-pair null, the naive-variable diagnostic, and the exact detector-frame cross-check are `relic_survey_gw_dlcrr.py`; the companion provenance note documents their algebra, sign conventions, counts, null definitions, sky/family/limit re-runs, and caveats, and the result is written to a gitignored JSON cache. The archived equality-of- D_L sweep is retained only through `relic_survey_gw.py` as the comparison. The corrected family null and counting limit in Secs. 6 and 8.2 are the `--family` and `--limit` runs of the same script; the older `relic_survey_family_null.py` remains the provenance for the archived-set family null. The winding-ladder diagnostic reported in Sec. 6 was exploratory output: it is not implemented in the archived `gw4.py` and no quantitative ladder result is used in this paper. Stage 3 is reproducible from the repository. The extractor `simulations/pe_extract.py` queries the GWOSC event API, downloads the 16 public PE files (1.40 GB) into the gitignored `data/gw_pe/raw/`, and writes compact draw sets to `data/gw_pe/`. Per event it takes the exact Mixed posterior group (C00:Mixed in GWTC-4.1/GWTC-5.0, C01:Mixed in GWTC-2.1/GWTC-3), with the SEOBNRv5PHM fallback

for the two events that lack it (GW231026_130704 and GW240420_175625). GW190706_222641 and GW190719_215514 use the GWTC-2.1-confident release, and GW200322_091133 carries the archived $n = 69$ Mixed draws. The reconstruction is:

```
python simulations/pe_extract.py --fetch
python simulations/relic_survey_gw4.py
python simulations/relic_survey_gw4_dlcorr.py --save
```

The corrected script reproduces the archived equality-cut run on the identical draws and applies the corrected scaling residual: it finds $\binom{16}{2} = 120$ pairs, 82 surviving the corrected cut (65 under equality; 63 both, 19 corrected-only, 2 equality-only), the same $\Lambda = -0.682 \pm 0.214$ and $+0.620 \pm 0.101$ for the two named targets, a 64-of-82 shared-base cluster at $|\Lambda_0| = 0.467$ with $p = 1.000$ (57 of 65 under equality), and the exact chirp-mass and detector-frame cross-checks at 81 and 78 survivors. The draw-matching seed (42) and 2×10^5 draws per pair are fixed in both scripts. The archived `gw4.py` implements only the equality-of- D_L posterior pair test; the corrected `gw4_dlcorr.py` implements the signature of Eq. (13) at posterior level. The winding-ladder diagnostic reported in Sec. 6 remains exploratory: it is not implemented in either script, so its integer-pattern count is not reproduced here.

The skymap inventory is reproducible without dense conversion by running `relic_survey_gw3.py` with the `--count-only` option and the `--map-root` option pointed at the skymap cache; it reports the four release token sets, the per-release name-matched counts, the pairwise overlaps, and the union. The project provenance notes record the four Zenodo download URLs, the file counts, the SHA-256 hashes, and the token-level reconciliation. The dense maps are not committed; the skymap cache is gitignored.

B The boost algebra

Let the seam identification be a boost along z with rapidity Λ . Here Λ is the boost rapidity; in the pair statistic of Sec. 3.1 it denotes the mass log-ratio, $\Lambda = \ln(m'/m) = -\Lambda_{\text{hol}}$. In the covering space a photon has four-momentum $k^\mu = (E, k_x, k_y, k_z) = E(1, n_x, n_y, n_z)$. The inverse boost acts as

$$E' = E(\cosh \Lambda - n_z \sinh \Lambda), \quad (19)$$

$$k'_z = E(n_z \cosh \Lambda - \sinh \Lambda), \quad (20)$$

while $k'_x = k_x$ and $k'_y = k_y$. The frequency ratio is therefore Eq. (2). For $n_z = \pm 1$ the transverse momentum is zero and the boost maps the light direction to itself, so the image is not deflected and the ratio is exactly $e^{\mp \Lambda}$; for $|n_z| < 1$ the direction is aberrated by $\tan \theta' = \sin \theta / (\cos \theta \cosh \Lambda - \sinh \Lambda)$ in the z -transverse plane, which gives the $\arctan(\sinh \Lambda)$ maximum tilt quoted in the text.

For a gravitational wave, the transverse-traceless amplitude is invariant under a boost along the propagation direction, while the frequency is multiplied by r . A common scaling $f \rightarrow rf$, $\dot{f} \rightarrow r^2 \dot{f}$ makes the inferred chirp mass scale as $M_c \rightarrow M_c/r$, so both components scale by the same factor; the mass ratio and the dimensionless spins are unchanged. With h invariant and $h \propto M_c^{5/3} f^{2/3}/D_L$, the inferred distance scales as $D_L \rightarrow D_L/r$, i.e. $D'_L \propto M'_c$; the invariant is M_c/D_L , not D_L by itself. The amplitude argument is exact on axis; off axis the boost is not along the propagation direction, and the survey uses the mass-distance scaling cut as the operational statement of no magnification. This is the algebra behind Eq. (7) and Table 2.

References

- [1] S. W. Hawking, “Chronology protection conjecture,” *Phys. Rev. D* **46**, 603 (1992).
- [2] B. S. Kay, M. J. Radzikowski, and R. M. Wald, “Quantum field theory on space-times with a compactly generated Cauchy horizon,” *Commun. Math. Phys.* **183**, 533 (1997); arXiv:gr-qc/9603012.
- [3] G. Klinkhammer, “Vacuum polarization of scalar and spinor fields near closed null geodesics,” *Phys. Rev. D* **46**, 3388 (1992).
- [4] M. Visser, “The quantum physics of chronology protection,” in *The Future of Theoretical Physics and Cosmology* (Cambridge University Press, Cambridge, 2003), pp. 161–176; arXiv:gr-qc/0204022.
- [5] A. Ori, “A class of time-machine solutions with a compact vacuum core,” *Phys. Rev. Lett.* **95**, 021101 (2005); arXiv:gr-qc/0503077.
- [6] A. Ori, “Formation of closed timelike curves in a composite vacuum/dust asymptotically flat spacetime,” *Phys. Rev. D* **76**, 044002 (2007); arXiv:gr-qc/0701024.
- [7] S. Xavier, “Closed Timelike Curves from a Vacuum Traveling Wave,” arXiv:2607.00788 (2026).
- [8] C. W. Misner, “Taub-NUT space as a counterexample to almost anything,” in *Lectures in Applied Mathematics*, Vol. 8 (American Mathematical Society, Providence, 1967), pp. 160–169.
- [9] J. L. Friedman, K. Schleich, and D. M. Witt, “Topological censorship,” *Phys. Rev. Lett.* **71**, 1486 (1993).
- [10] J. Schemoul, “Initial data for a time machine from dust,” companion construction paper, in preparation (2026).
- [11] J. Schemoul, “Tensor-mode stability of the Ori time machine,” companion stability paper, in preparation (2026).
- [12] J. Schemoul, “What a working time machine takes,” companion construction paper, in preparation (2026).
- [13] Gaia Collaboration, A. Vallenari *et al.*, “Gaia Data Release 3: Summary of the content and survey properties,” *Astron. Astrophys.* **674**, A1 (2023); arXiv:2208.00211.
- [14] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), “GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs,” *Phys. Rev. X* **9**, 031040 (2019); arXiv:1811.12907.
- [15] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), “GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo during the First Half of the Third Observing Run,” *Phys. Rev. X* **11**, 021053 (2021); arXiv:2010.14527.
- [16] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), “GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo during the First Half of the Third Observing Run,” *Phys. Rev. D* **109**, 022001 (2024); arXiv:2108.01045.

- [17] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run,” *Phys. Rev. X* **13**, 041039 (2023); arXiv:2111.03606.
- [18] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, “IGWN-GWTC4p1-18965dda8_5-Archived_Skymaps.tar.gz” data release, Zenodo record 20275769, doi:10.5281/zenodo.20275769.
- [19] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, “IGWN-GWTC5p0-29ebe06b7_25-Archived_Skymaps.tar.gz” data release, Zenodo record 20348005, doi:10.5281/zenodo.20348005.
- [20] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration, “GWTC-5.0: An Introduction to Version 5.0 of the Gravitational-Wave Transient Catalog,” arXiv:2605.27223 (2026).
- [21] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration, “GWTC-4.0: Searches for Gravitational-Wave Lensing Signatures,” arXiv:2512.16347 (2025).
- [22] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration, “GWTC-5.0: Methods for Identifying and Characterizing Gravitational-wave Transients,” arXiv:2605.27224 (2026).
- [23] X. Shan, B. Hu, X. Chen, and R.-G. Cai, “An interference-based method for the detection of strongly lensed gravitational waves,” *Nature Astronomy* **9**, 916 (2025); arXiv:2301.06117.
- [24] X. Shan, Y.-F. Wang, S. Mao, H. Yang, and M. Zumalacárregui, “Gravitational lensing of gravitational waves by galaxy clusters,” arXiv:2609.16891 (2026).
- [25] D. Hegde and J. Janquart, “On the interplay between waveform systematics and lensing signatures in gravitational-wave signals,” arXiv:2609.03935 (2026).
- [26] T. N. Nguyen-Vo, T. Singh, B. McKallip, M. Kesden, and L. King, “Distinguishing lensing and precessional modulation in binary black-hole inspiral waveforms,” arXiv:2609.06347 (2026).
- [27] U. Deka, A. Ganguly, S. Bhattacharyya, and P. Ajith, “Impact of the host galaxy on the search for gravitational-wave microlensing by compact objects,” arXiv:2609.16990 (2026).

Script	Function
<code>relic_survey_stage1.py</code>	Stage-1 recount: one-degree-of-freedom 3σ pair scan with masses and χ_{eff} , 10,000 label permutations, on the 39,621-pair cached catalog.
<code>relic_survey_gw.py</code>	Archived Stage-1.5: catalog load, all-pairs $\Lambda_1 = \Lambda_2$ scan, χ_{eff} and equality-of- D_L consistency, 3–6 σ sweep, label-permutation null. Kept as the comparison; its distance cut is not the signature.
<code>relic_survey_gw_dlcrr.py</code>	Corrected Stage-1.5 (W62): loads the catalog, reproduces the archived Stage-1/1.5 counts, applies the mass-distance scaling cut M_c/D_L (or $\ln(D_{Lj}/D_{Li}) = \hat{\Lambda}$), computes the matched independent-pair null and the naive-variable null, and cross-checks the exact detector-frame invariant. Options <code>--sky</code> , <code>--family</code> , and <code>--limit</code> re-run the corrected sky, family, and counting-limit stages. Writes a gitignored result cache.
<code>relic_survey_gw2.py</code>	Stage-2 pilot: MOC/UNIQ decode, dense nside-512 maps, cosine co-localization, 51-event pair set, 2,000 sky permutations.
<code>relic_survey_gw3.py</code>	Full-volume Stage 2: all four skymap releases, dense nside-256 maps, 281 mapped events, max- C null, shared-base family test with sky-permutation calibration.
<code>relic_survey_family_null.py</code>	Archived equality- D_L composition-matched family null (historical; the corrected family null is the <code>gw_dlcrr.py --family</code> run): reproduces the archived sky null, then bootstraps the co-located $(\Lambda, \sigma_\Lambda)$ composition, smooth KDE/Gamma models, assignment permutation, pool-draw check, synthetic power, and non-candidate controls.
<code>pe_extract.py</code>	Stage-3 cache builder: queries the GWOSC event API, downloads the public Zenodo PE files, takes the exact Mixed group (SEOB-NRv5PHM fallback where absent), and writes source-frame masses, χ_{eff} , and D_L to the gitignored <code>data/gw_pe/</code> .
<code>relic_survey_gw4.py</code>	Archived Stage 3: draw-matched pair test with the equality- D_L cut and family test; reproduces 65/120.
<code>relic_survey_gw4_dlcrr.py</code>	Corrected Stage 3 (W73): same pair test with the scaling residual $\ln(D_{Lj}/D_{Li}) - \Lambda$ in place of equality; chirp-mass and detector-frame cross-checks; gitignored cache.
<code>relic_survey_limit.py</code>	Calibrated counting limit from 5,000 sky-map permutations; Gamma-Poisson and empirical 90% limits.
<code>relic_survey_selection.py</code>	Monte Carlo injection/selection model for $\varepsilon_{\text{family}}$; relic-number and density limits.

Table 8: The analysis chain. The scripts cache the catalog; the skymap cache and the posterior cache are gitignored, so neither is part of the repository or redistributed. Stages 1 and 1.5 reproduce from the cached catalog; the skymap, family-null, limit, and selection stages require the public skymap releases, and Stage 3 requires the PE draw sets.