Multiple Gravitational Wave Interferometry (MGWI): A Proposal for Ensemble Analysis of Post-Merger Residuals

John Reimer Morales

Independent Researcher

March 2026

1 Introduction

1.1 The Era of Gravitational Wave Catalogs

The detection of GW150914 by the LIGO-Virgo-KAGRA (LVK) collaboration inaugurated precision gravitational wave astronomy [1]. The collaboration has since compiled catalogs containing approximately 300 compact binary coalescences [2], providing novel constraints on stellar evolution, black hole demographics, the neutron star equation of state [3], and cosmology.

These detections were made possible by kilometer-scale laser interferometers whose operational principle is rooted in wave superposition [4]. In the weak-field limit appropriate to the signals reaching Earth—where the strain perturbation satisfies $|h_{\mu\nu}| \ll 1$—the Einstein field equations linearize to a standard wave equation [5, 6]. In this regime, gravitational waves obey the superposition principle: multiple weak waves passing through the same spacetime region combine by simple addition of their strains. This effective linearity is the cornerstone of interferometric detection, even though the underlying theory is fundamentally nonlinear [32].

Current data analysis techniques exploit this linearity to extract information from *individual* events. Matched filtering cross-correlates detector data with theoretical waveform templates to identify signals buried in noise [7]. Network coherence analysis validates candidates by requiring consistent arrival across multiple detectors [8]. While powerful, these methods treat each event as an independent entity; the information extracted pertains only to the properties of that event's source.

This paper proposes a complementary paradigm: the application of interferometric and correlation techniques to *separate, distinct* gravitational wave events drawn from the growing catalog. We term this framework **Multiple Gravitational Wave Interferometry (MGWI)**.

1.2 The Silence of the Echo: Why Phase-Coherent Searches Fail

In standard post-merger analysis, the ringdown is fitted to the quasinormal modes (QNMs) of a perturbed Kerr black hole. The consistency of these fits with the No-Hair Theorem is widely cited as evidence for the classical event horizon: a smooth, one-way null surface [43, 9]. Any residual energy remaining after template subtraction is assumed to be detector noise.

However, several theoretical programs have motivated the search for post-merger "echoes"—reflections caused by quantum-gravitational corrections near the horizon scale. Cardoso, Franzin, and Pani [34] first demonstrated that the early-time ringdown of compact objects with a light ring is indistinguishable from that of a black hole; only precision observations of the late-time signal can reveal exotic alternatives. Cardoso and Pani [10] subsequently showed that exotic compact objects (ECOs) with structure near the would-be horizon produce a universal train of modulated echo pulses. Abedi, Dykaar, and Afshordi [11] reported tentative evidence (2.5σ) for such echoes in LIGO data, subsequently assessed as statistically inconclusive [12]. Multiple independent analyses of the full GWTC-3 catalog using various echo templates have returned null results [13], though Abedi reports marginal Bayesian evidence ($\mathcal{B} \approx 9$) for stimulated Hawking

radiation in the high-mass event GW190521—suggestive but not decisive, and consistent with the interpretation that individual-event sensitivity remains insufficient.

We argue that these null results do not exclude horizon structure; rather, they may reflect a mismatch between the search strategy and the physical nature of the signal. The theoretical basis for this claim rests on a convergent body of results from quantum information theory and black hole physics spanning four decades.

The conceptual foundation traces to the *membrane paradigm* of Thorne, Price, and Macdonald [49], who showed that for an external observer, the classical event horizon behaves as a physical membrane endowed with viscosity, electrical resistivity, and thermodynamic entropy—a dissipative surface, not an abstract null boundary. Susskind, Thorlacius, and Uglum [15] elevated this to a quantum-mechanical framework: the *stretched horizon*, a Planck-thickness membrane carrying microscopic degrees of freedom that actively process infalling information. Hayden and Preskill [16] subsequently showed that if the internal dynamics of the black hole are unitary and rapidly mixing, information deposited in the black hole is returned in the Hawking radiation after a characteristic *scrambling time*—far faster than previously assumed. Building on this, Sekino and Susskind [14] conjectured that "black holes are the fastest scramblers in nature": they distribute incoming quantum information across their entire stretched horizon in a time scaling only logarithmically with entropy, $t_* \sim \beta \log S$. Maldacena, Shenker, and Stanford [17] placed this on rigorous footing by proving a universal upper bound on the rate of quantum chaos, $\lambda_L \leq 2\pi k_B T/\hbar$, and demonstrating that black holes saturate this bound—they are *maximally chaotic* systems.

Two further results sharpen the picture. Brown and Susskind [20] argued that the interior volume of a black hole grows linearly for exponential time as a consequence of the *Second Law of Quantum Complexity*, creating a severe geometric tension between interior volume and exterior surface area that any physical boundary must accommodate. Bouland, Fefferman, and Vazirani [18] demonstrated that as the black hole evolves, its internal state becomes *computationally pseudorandom*—indistinguishable from thermal noise by any polynomial-time computation.

The consequence for echo searches is significant. If the horizon's internal dynamics approach the fast-scrambling limit, the phase evolution $\phi(t)$ of any radiation interacting with the boundary is likely subject to stochastic delays and partial decoherence. Phase-coherent searches—designed to match identical, repeating wavepackets—may therefore be poorly suited to the scrambled or stochastic-response regime. MGWI is intended as a complementary probe, tailored to the case in which phase coherence is degraded or strongly model-dependent. We note that phase-marginalized coherent searches remain an active and viable approach [13]; MGWI does not replace them but addresses the regime in which even marginalized coherent templates may lose sensitivity.

This diagnosis has received independent support from recent work on stochastic horizon structure. Hu, Fang, and Guo [19] analyze black hole ringdown stability under quenched microscopic disorder at the horizon and demonstrate that random phase delays generated by geometric fluctuations lead to severe decoherence of echo signals during multiple round trips in the photon sphere cavity, smearing coherent echo trains into broadband noise. Their central finding

is that typical thermal states of the black hole microstate ensemble remain observationally indistinguishable from classical GR due to systematic phase cancellation of short-wavelength fluctuations.

However, while the *phase* of any returning radiation may be degraded by the internal dynamics of the black hole, the *resonant frequencies* at which a structured surface vibrates are determined by its macroscopic geometry and are expected to be approximately invariant across black holes of comparable mass and spin. A cathedral bell rings at the same fundamental pitch regardless of how chaotically it is struck. If the horizon possesses physical structure, its spectral properties—expressed in dimensionless units after remnant-mass normalization—should exhibit approximately universal features, even if the phase of the ringing is randomized.

This is the core insight motivating the MGWI protocol: *search for persistent dimensionless spectral structure in the power domain, where phase information is not required.*

The question MGWI poses is therefore elementary but fundamental: **is the event horizon a featureless classical absorber, or does it possess physical structure?** The protocol is model-agnostic by design—it does not presuppose which alternative to classical GR might be correct. It tests the binary distinction between a smooth vacuum boundary and anything else.

1.3 Prior Work: Coherent Mode Stacking

The idea of combining information from multiple merger events is not new. Yang et al. [21] proposed *coherent mode stacking* for black hole spectroscopy: rescaling BBH ringdown signals so that a target QNM frequency aligns across events, then summing waveforms constructively to boost subdominant mode SNR. This was extended to binary neutron star post-merger signals [22], where both coherent and "power stacking" methods were developed to detect the dominant post-merger oscillation mode.

These methods are powerful but address a different problem. Yang et al. stack the *signal* to find *known* modes louder. They require a priori knowledge of the target frequency and exploit phase information extracted from the inspiral to align waveforms coherently. Their "power stacking" variant relaxes phase alignment but still targets a specific, theoretically predicted oscillation frequency.

MGWI, as proposed here, stacks the *silence*—the post-template residual—to search for *unknown* structure. The target is not a known QNM mode but any universal spectral feature in the residual that survives across events when mapped onto a dimensionless frequency domain. The phaseless power stacking protocol requires no a priori frequency target and no phase alignment. It is designed precisely for the case where the phase is *effectively* scrambled and no coherent template exists.

The relationship is analogous to the distinction between standard amplitude interferometry and Hanbury Brown–Twiss (HBT) intensity interferometry [23]. In stellar observations, atmospheric turbulence destroys phase coherence (first-order coherence $g^{(1)}$), defeating standard interferometry. HBT bypasses this by correlating *intensities* (second-order coherence $g^{(2)}$), which remain

unaffected by random phase delays. In the gravitational domain, the horizon's scrambling plays the role of atmospheric turbulence; MGWI's power stacking plays the role of HBT.

1.4 Outline

Section 2 describes the general MGWI framework and its broader applications: identification of strongly lensed gravitational waves (§2.1), cross-correlation with large-scale structure (§2.2), and gravitational wave intensity interferometry (§2.3). Section 3 develops the specific phaseless power stacking protocol for horizon structure in detail. Section 4 presents projected sensitivity estimates from injection studies. Section 5 describes a three-phase observational roadmap and discusses broader implications. Section 6 concludes. Figure 1 provides a schematic overview of the method, its pipeline, and its controls.

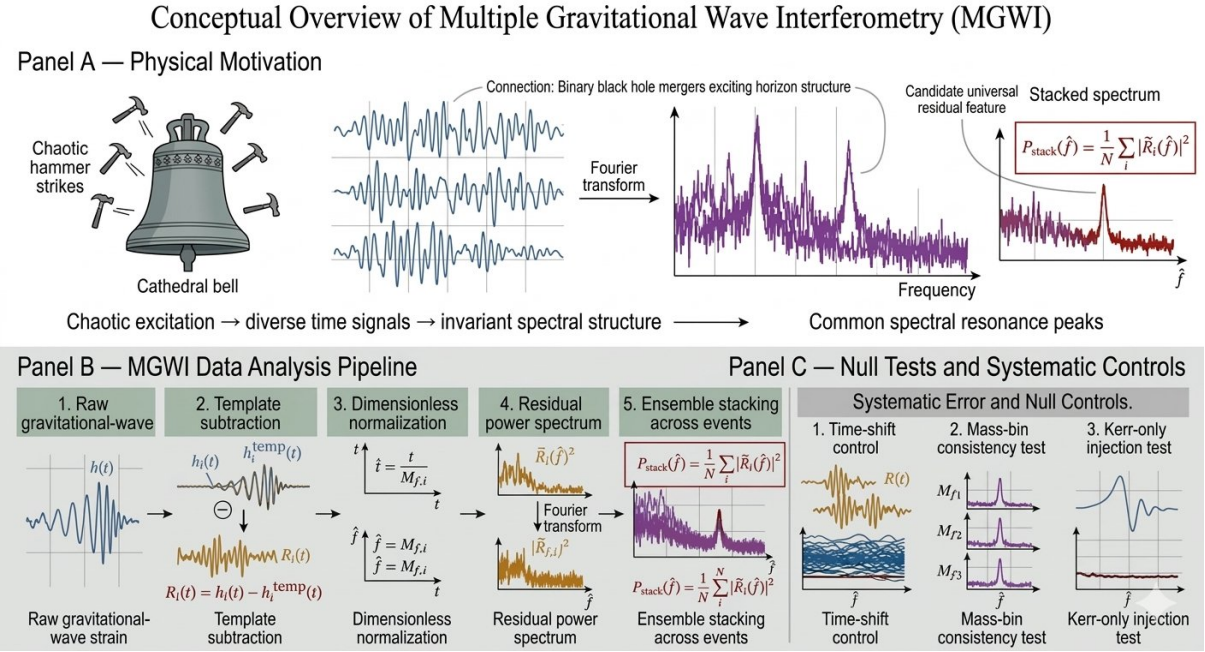

Figure 1: Conceptual overview of the Multiple Gravitational Wave Interferometry (MGWI) method. **(A)** Chaotic excitation of a resonant system can produce diverse time-domain responses while preserving characteristic spectral structure. The analogy illustrates how binary black hole mergers may excite horizon dynamics that generate event-dependent residual waveforms but share approximately universal spectral features. **(B)** MGWI data analysis pipeline. Gravitational-wave strain data $h_i(t)$ are template-subtracted to produce residual signals $R_i(t)$, normalized by remnant mass to define dimensionless variables $\hat{t}$ and $\hat{f}$, converted to residual power spectra $|\tilde{R}_i(\hat{f})|^2$, and stacked across events to produce the catalog-level statistic $P_{\rm stack}(\hat{f})$. **(C)** Null tests and systematic controls designed to identify structured false positives, including time-shift tests, mass-bin consistency checks, and Kerr-only injection tests.

2 The MGWI Framework: Cross-Catalog Interferometry

MGWI applies interferometric principles not within a single detector but across the entire catalog of detected events. The approach is analogous to aperture synthesis in radio astronomy

[27], where signals from multiple telescopes are combined to create an instrument with a much larger effective aperture. In MGWI, signals from multiple events—observed over time from different directions—are combined to construct a new probe of the intervening spacetime and the universal properties of compact objects.

We outline three applications beyond the primary phaseless power stacking protocol developed in Section 3.

2.1 Identification of Strongly Lensed Gravitational Waves

General Relativity predicts that gravitational waves, like light, can be gravitationally lensed by massive intervening objects. In strong lensing, multiple "images" of the same source event arrive at the detector at different times (minutes to months apart) and with different magnifications, but their intrinsic waveform shapes are identical up to a constant phase shift and overall amplitude rescaling [24].

Identifying lensed pairs from a large catalog is a significant challenge. Standard approaches compare inferred source parameters of all event pairs, but these are subject to parameter estimation uncertainties. MGWI offers a more direct method: systematic cross-correlation of reconstructed waveform data for every event pair in the catalog. A high correlation between two events detected at different times, with parameters consistent after accounting for magnification, would be strong evidence for lensing. In the wave optics regime, interference between lensed paths creates characteristic beat patterns in the waveform itself [24], and cross-correlation analysis is the natural method for identifying such features. As catalog sizes grow to 10^5 events with next-generation detectors, hierarchical search strategies [28] and machine learning feature extraction [29] will be essential to manage the N^2 scaling of pairwise comparisons.

2.2 Cross-Correlation with Large-Scale Structure

The superposition of all unresolved compact binary mergers produces a stochastic gravitational wave background (SGWB) predicted to be anisotropic, tracing the large-scale distribution of galaxies. Cross-correlating the gravitational wave sky map with galaxy surveys amplifies this faint signal against detector noise [25]. The LVK collaboration has already conducted directional searches for the SGWB [42], and the technique of cross-correlating gravitational wave data with electromagnetic surveys of large-scale structure represents an established form of cross-catalog interferometry. As detection rates increase, the angular resolution of the gravitational wave sky map will improve, enabling increasingly precise correlations with cosmological structure.

2.3 Gravitational Wave Intensity Interferometry

A forward-looking application involves applying HBT intensity interferometry directly to gravitational waves. Primordial gravitational waves from the early universe are predicted to be in a squeezed quantum state; detecting the non-classical intensity correlations predicted by this

state would provide a direct window into the quantum nature of gravity [26]. While this application requires sensitivity far beyond current detectors, it represents the conceptual endpoint of the MGWI program: extending the HBT analogy from a methodological tool (as in our power stacking protocol) to a direct physical measurement of the gravitational field's quantum statistics.

3 Phaseless Power Stacking for Horizon Structure

We now develop the primary methodological contribution: a protocol for detecting universal spectral structure in the post-merger residual of binary black hole coalescences. The protocol is agnostic to the specific theoretical model of horizon structure; it tests the general hypothesis that the post-template residual contains statistically significant spectral features that are invariant across events when expressed in dimensionless units scaled by remnant mass.

3.1 The Signal Model

Under the classical Kerr hypothesis, the event horizon is a perfect absorber with no internal structure. Template subtraction should leave only detector noise. Under any model predicting horizon structure—whether exotic compact objects [10], fuzzballs [30], firewalls [45], gravastars [48], membrane paradigm realizations [49], or other quantum-gravitational modifications—the residual should contain additional spectral content.

As argued in §1.2, the phase of this content may be degraded for any model involving a fast-scrambling boundary. However, the *resonant frequencies* at which a structured boundary vibrates are determined by its macroscopic geometry and are expected to be approximately universal across black holes of similar mass and spin, after normalization to dimensionless units. We note that individual QNM pole locations are known to be spectrally unstable under small perturbations of the effective potential; however, the *total spectral weight* integrated over frequency bands is a more robust observable, and it is this integrated power that the MGWI protocol targets rather than individual pole positions.

The MGWI protocol therefore searches for **persistent dimensionless spectral structure**: a feature in the power spectral density of the residual that appears at the same dimensionless frequency across events after remnant-mass normalization. We are not searching for a time-domain echo; we are searching for a spectral signature of the horizon's physical architecture.

Table 1 summarizes the competing horizon models and their distinct MGWI-discriminable signatures. All alternative hypotheses are evaluated under identical preprocessing, normalization, and stacking pipelines to prevent post hoc feature selection.

3.2 Parameterized Horizon-Response Model

To clarify the class of signals to which MGWI is sensitive, we introduce a parameterized model for the post-merger residual response. The purpose of this model is not to commit to a specific

Hypothesis	**Boundary Condition**	**Predicted Residual Signature**	**MGWI Discriminator**
H_0: Classical Kerr (Null)	Perfectly absorptive null surface	No structure beyond Kerr QNMs; residuals consistent with Gaussian noise	Stacked residual power consistent with zero
H_1: Partially Reflective Near-Horizon (Echo Models) [10, 11]	Small constant reflectivity near the would-be horizon	Time-domain echo sequence at intervals $\propto M_f \ln \epsilon$	Late-time excess power at consistent normalized delay
H_2: Exotic Compact Object with Surface (Gravastar-class) [48]	Hard surface replaces null horizon	Modified QNM spectrum; shifted overtone frequencies and damping	Systematic deviation in mode content after mass normalization
H_3: Frequency-Responsive Boundary (Membrane / Fuzzball-class) [49, 30]	Frequency-dependent impedance $R(\omega) \neq 0$	Banded excess or depletion of power in a specific normalized frequency range	Structured residual PSD in $\hat{\omega} = M_f\omega$ domain across events

Table 1: Competing models of horizon structure and their predicted gravitational wave ringdown signatures. MGWI treats the classical Kerr ringdown (H_0) as the null hypothesis. Each alternative predicts a distinct morphology in the stacked residual, enabling statistical discrimination via ensemble inference. Specific boundary-condition parameterizations (e.g., reflectivity amplitude, onset frequency, spectral slope) can be fit within the MGWI framework for any model admitting a frequency-dependent reflectivity $R(\omega)$.

microscopic description of horizon physics, but to demonstrate how a broad class of boundary-response scenarios produces a persistent structure in the dimensionless residual power spectrum that survives ensemble averaging.

We write the measured strain for event i as

$$h_i(t) = h_i^{\mathrm{Kerr}}(t) + \delta h_i(t) + n_i(t), \tag{1}$$

where $h_i^{\mathrm{Kerr}}(t)$ denotes the classical Kerr ringdown, $\delta h_i(t)$ represents any boundary response not captured by the template model, and $n_i(t)$ is detector noise. After template subtraction, the residual is $R_i(t) \approx \delta h_i(t) + \epsilon_i(t) + n_i(t)$, where $\epsilon_i(t)$ accounts for residual mismatch from imperfect subtraction of known Kerr physics.

We parameterize the boundary response in the frequency domain as

$$\tilde{\delta h}_i(\omega) = A_i\, Q_i(\omega)\, e^{i\phi_i(\omega)}, \tag{2}$$

where A_i is an event-dependent amplitude, $Q_i(\omega)$ is the magnitude of the boundary-response transfer function, and $\phi_i(\omega)$ is an event-dependent phase. In scenarios motivated by chaotic horizon dynamics or stochastic boundary fluctuations, the phases $\phi_i(\omega)$ are not expected to remain coherent across different merger events. MGWI therefore focuses on the power spectrum,

which is insensitive to phase scrambling. Introducing the dimensionless frequency $\hat{\omega} = M_{f,i}\,\omega$, we assume

$$Q_i(\omega) \approx \alpha_i\,\mathcal{Q}(\hat{\omega}), \tag{3}$$

where α_i is a weak event-dependent scaling factor and $\mathcal{Q}(\hat{\omega})$ is a dimensionless response profile approximately shared across the event population to leading order. If the inter-event phases are effectively uncorrelated, the cross terms in the stacked power average down, and the MGWI statistic tends to

$$P_{\rm stack}(\hat{\omega}) \approx \langle w_i A_i^2 \alpha_i^2 \rangle\,|\mathcal{Q}(\hat{\omega})|^2 + P_{\rm KerrSys}(\hat{\omega}) + P_{\rm noise}(\hat{\omega}), \tag{4}$$

where $P_{\rm KerrSys}$ denotes the stacked contribution from residual known-physics systematics and $P_{\rm noise}$ the averaged detector noise floor.

This expression makes explicit what MGWI is sensitive to: not a phase-coherent waveform template, but a *persistent dimensionless residual power structure* $|\mathcal{Q}(\hat{\omega})|^2$ that survives ensemble averaging. The exact morphology of $\mathcal{Q}(\hat{\omega})$ is model-dependent. As illustrative examples, the response profile could take the form of a narrow resonance,

$$\mathcal{Q}(\hat{\omega}) = \frac{\beta}{(\hat{\omega} - \hat{\omega}_0)^2 + \Gamma^2}, \tag{5}$$

or a broader band-limited feature,

$$\mathcal{Q}(\hat{\omega}) = \beta \exp\left[-\frac{(\hat{\omega} - \hat{\omega}_0)^2}{2\sigma^2}\right], \tag{6}$$

where $\hat{\omega}_0$ defines the characteristic dimensionless resonance frequency, Γ or σ controls the feature width, and β sets its amplitude. These parameterizations are not intended as unique predictions of any particular microscopic model; rather, they define a generic class of structured boundary responses for which phaseless stacking is well matched.

In this framework, a null MGWI result constrains the amplitude and width of $|\mathcal{Q}(\hat{\omega})|^2$, while a statistically significant excess would indicate the presence of universal post-merger residual structure not captured by classical Kerr ringdown models. MGWI is therefore not a search for a specific waveform template but for persistent structure in the dimensionless residual power spectrum that survives ensemble averaging.

3.3 Data Selection: The Golden Set

The post-template residual is extraordinarily fragile. Any echo-like signal is theoretically predicted to appear at strain amplitudes well beneath the noise floor of individual events. Rigorous curation of the input data is essential.

Selection criteria:

- *High Fidelity (SNR ≥ 12):* Only binary black hole mergers with combined network SNR $\rho \geq 12$ are included. By the Fisher Information Matrix formalism [44], parameter variance

scales as $1/\rho^2$; this threshold ensures $< 5\%$ uncertainty in remnant mass M_f and spin a_f, which is critical for the temporal normalization step. Recent high-SNR events such as GW250114 ($\rho \approx 80$) [47] provide exceptional individual anchors for the stacking analysis.

- *Clean Ringdown:* Events dominated by the fundamental quadrupole mode ($\ell = 2$, $m = 2$) are prioritized, ensuring that template subtraction is maximally effective.

Exclusion criteria:

- *Spin-Precession ($\chi_p > 0.1$):* Precessing binaries exhibit complex amplitude modulations during ringdown that can mimic spectral structure.
- *Eccentricity and Higher Modes:* Events with residual eccentricity or strong higher-order mode content (e.g., mass ratio $q > 3$) are excluded.
- *Instrumental Glitches:* Events coincident with excess power in auxiliary monitoring channels, or classified as glitch types (Blip, Koi Fish, Scattered Light) by detector characterization pipelines such as GravitySpy [36], are discarded. Non-Gaussian artifacts violate the Central Limit Theorem assumptions required for noise cancellation.

Table 2 lists representative candidate events from GWTC-3 and O4 that satisfy the Golden Set criteria. This preliminary census yields approximately 15–25 qualifying BBH events from GWTC-3 alone, with additional candidates expected from the ongoing O4/O5 runs. The table demonstrates that the remnant mass uncertainties are typically 3–8% at the 90% credible level—adequate for the τ-normalization step, though events near the SNR threshold require careful treatment. The mass range spans ~ 18–$100\,M_\odot$ in remnant mass, providing substantial leverage for the mass-binning test.

Event	$M_f\ (M_\odot)$	q	χ_p	**SNR**	$\delta M_f/M_f$
GW150914	$62.3^{+3.7}_{-3.1}$	1.2	0.06	24	$\sim 5\%$
GW170608	$18.3^{+1.6}_{-0.8}$	1.6	0.05	15	$\sim 7\%$
GW170814	$53.2^{+3.2}_{-2.5}$	1.2	0.04	16	$\sim 5\%$
GW190408	$40.5^{+3.8}_{-2.7}$	1.3	0.07	16	$\sim 8\%$
GW190503	$60.6^{+5.1}_{-3.9}$	1.3	0.05	13	$\sim 7\%$
GW190519	$101.9^{+9.4}_{-7.3}$	1.6	0.08	14	$\sim 8\%$
GW190630	$56.2^{+4.3}_{-3.4}$	1.5	0.06	15	$\sim 7\%$
GW190708	$29.1^{+2.4}_{-1.7}$	1.3	0.05	13	$\sim 7\%$
GW190828	$55.6^{+3.3}_{-2.6}$	1.2	0.04	16	$\sim 5\%$
GW200112	$55.3^{+4.2}_{-3.3}$	1.4	0.06	16	$\sim 7\%$
GW200311	$59.0^{+4.0}_{-3.2}$	1.2	0.05	14	$\sim 6\%$
GW250114	$62.7^{+1.0}_{-1.1}$	1.04	0.03	80	$\sim 2\%$

Table 2: Representative Golden Set candidates from GWTC-3 and O4 satisfying the selection criteria of §3.3: network SNR ≥ 12, effective precession spin $\chi_p < 0.1$, mass ratio $q < 3$, and no identified instrumental glitches. Parameters are median values with 90% credible intervals from the LVK parameter estimation [2, 47]. The fractional remnant mass uncertainty $\delta M_f/M_f$ indicates the precision of the τ-normalization for each event. The O4 event GW250114, with $\rho \approx 80$ and near-unity mass ratio, represents the highest-precision anchor available.

3.4 Template Subtraction

For each event in the Golden Set, we subtract the maximum likelihood waveform predicted by a state-of-the-art phenomenological waveform model (e.g., IMRPhenomXHM [35], which includes higher harmonics and is calibrated to numerical relativity simulations [31]):

$$\delta h(t) = h_{\rm data}(t) - h_{\rm template}(t) \tag{7}$$

The resulting time-series $\delta h(t)$ is the *residual*. Under the null hypothesis, it is pure detector noise. Under the alternative hypothesis, it contains the spectral signature of horizon structure. We note that the quality of the residual depends on the fidelity of the waveform model; systematic errors in the template can introduce spurious spectral features. This is a known concern in ringdown analysis and motivates the strict selection criteria of §3.3, which ensure that the Golden Set events are well-modeled by the dominant quadrupole mode.

A particularly insidious systematic arises from unmodeled subdominant Kerr modes. In high-SNR events, modes beyond the fundamental ($\ell = 2, m = 2$) quadrupole—such as the $(3, 3)$, $(4, 4)$, or overtone $(2, 2, 1)$—may carry detectable power that the template does not fully capture. Because these modes scale with $1/M$ like any horizon signature, their residual power would stack coherently under the MGWI protocol and could mimic a beyond-Kerr spectral feature. The Golden Set's restriction to quadrupole-dominated, non-precessing events mitigates this risk; moreover, for the near-equal-mass systems that dominate the Golden Set (e.g., GW150914, GW250114), the odd-m modes $(3, 3)$ and $(2, 1)$ are strongly suppressed by the approximate symmetry of the equal-mass limit, providing additional theoretical protection against subdominant mode contamination.

As a further diagnostic, we recommend repeating the full stacking analysis with at least two independent waveform families (e.g., IMRPhenomXHM and SEOBNRv5) to verify that any detected feature is robust to waveform modeling choices. A feature appearing in one family but not the other would indicate a template artifact rather than horizon structure. We note that these two families, while analytically distinct in construction (phenomenological frequency-domain vs. effective-one-body time-domain), share some common numerical relativity calibration data. A fully independent check would ultimately require comparison against new NR simulations produced with independent codes (e.g., SpEC vs. BAM vs. Einstein Toolkit), which is beyond the scope of this proposal but represents an important long-term validation target.

3.5 Dimensionless Normalization: The Universal Manifold

Black hole ringdown frequencies are mass-dependent, scaling as $f \propto 1/M$. A $10\,M_\odot$ remnant rings at $\sim 1.2\,\text{kHz}$; a $60\,M_\odot$ remnant at $\sim 200\,\text{Hz}$. To stack residuals from black holes of different masses, we map them onto a *dimensionless manifold* where the boundary physics of all black holes aligns.

Temporal mapping. We define the dimensionless time coordinate τ, anchored at peak strain:

$$\tau = \frac{t - t_{\text{merger}}}{GM_f/c^3} \tag{8}$$

This rescales the temporal axis so that ringdown phases of all black holes, regardless of mass, evolve on the same grid. The physical validity rests on the assumption of scale invariance: if horizon structure is universal, its dimensionless properties scale self-similarly with mass. Because the Fourier transform is subsequently computed in the τ domain, the conjugate frequency ω is automatically the dimensionless $\hat{\omega} = M_f\omega_{\text{physical}}$, and no additional Jacobian correction is required to preserve power spectral density under the coordinate transformation.

Amplitude normalization. We normalize the residual strain by the peak merger amplitude to obtain a dimensionless fractional residual:

$$\bar{h}(\tau) = \frac{\delta h(t)}{h_{\text{peak}}} \tag{9}$$

This renders the residual independent of the source's luminosity distance and absolute amplitude: a 1% echo appears as $\bar{h} = 0.01$ regardless of whether the event is nearby or distant. The relative statistical weight of each event—accounting for differences in detector noise at the time of observation—is handled by the inverse-variance weighting w_i in the stacking step (§3.6).

Effect of remnant-mass uncertainty. Because MGWI aligns spectra using $\hat{\omega} = M_{f,i}\,\omega$, uncertainty in the inferred remnant mass produces a corresponding uncertainty in the dimensionless frequency axis. If the inferred mass is $M_{f,i}^{\text{est}} = M_{f,i}(1 + \delta_i)$, a true spectral feature at $\hat{\omega}_0$ is shifted to $\hat{\omega}_{0,i}^{\text{est}} \approx \hat{\omega}_0(1 + \delta_i)$. For a population with RMS fractional mass uncertainty σ_M, the effective broadening of any stacked feature is

$$\sigma_{\text{obs}}^2 \approx \sigma_{\text{int}}^2 + (\hat{\omega}_0\,\sigma_M)^2, \tag{10}$$

where σ_{int} is the intrinsic width of the response profile. This broadening degrades sensitivity to narrow features but does not generate spurious spectral structure: remnant-mass uncertainty enters MGWI primarily as a conservative loss of alignment precision, not as a mechanism for producing catalog-wide artifacts. Events with poorly constrained remnant-mass posteriors can be downweighted in the stack to reduce alignment-induced broadening.

3.6 The Stacking Algorithm

Because the phase is scrambled, direct time-domain averaging would cause destructive interference, erasing any signal. Instead, we employ **phaseless power stacking**. We note that while this approach is conceptually analogous to Hanbury Brown–Twiss intensity interferometry—in that it bypasses phase information to access structural properties—it is technically an incoherent average of individual-event periodograms, not a correlation of simultaneous intensity fluctuations $\langle \Delta I_1 \Delta I_2 \rangle$ between detectors. The analogy illuminates the logic of the method; the statistics are those of averaged power spectra.

Step 1: Power Spectrogram. For each event i, compute the periodogram of the normalized residual in dimensionless frequency $\hat{f}$:

$$P_i(\hat{f}) = \left|\mathcal{F}\{\bar{h}_i(\tau)\}\right|^2 \tag{11}$$

A Tukey window (with $\alpha = 0.1$ to minimize spectral leakage while preserving the early post-merger signal) is applied prior to the Fourier transform. The analysis window extends from $\tau = 0$ (peak strain) to $\tau \approx 500$, corresponding to approximately 25–150 ms in physical time depending on remnant mass; this range encompasses the expected echo timescale for near-horizon modifications while remaining within the regime where detector noise is well-characterized. By taking the absolute square, we permanently discard the scrambled phase while preserving any underlying resonant structure.

Step 2: Weighted Stack. The stacked power density is the weighted mean:

$$P_{\text{stack}}(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N} w_i \cdot P_i(\hat{f}) \tag{12}$$

where w_i is an inverse-variance weight derived from the local detector noise power spectral density $S_n(f_{\text{ringdown}})^{-1}$ at the time of each event. This *optimal Wiener weighting* ensures events falling in the detector's sensitivity bucket contribute maximally, while those in noisy bands are down-weighted.

As N grows, the Central Limit Theorem forces incoherent Gaussian noise to average to a flat floor (scaling as $1/\sqrt{N}$), while any universal, mass-invariant spectral feature consistently reinforces. In analogy with Hanbury Brown–Twiss intensity interferometry, $P_{\text{stack}}(\hat{f})$ can be interpreted as the second-order correlation of residual "intensities" across events: because phases are effectively randomized between events, only the magnitude-squared structure survives, making the statistic sensitive to persistent spectral features even in the absence of phase coherence. Individual merger events thus act as independent realizations of the same underlying response, while the stacking procedure extracts the invariant statistical structure shared across the population.

This interpretation is well matched to the statistical properties of gravitational-wave detector noise. Because detector noise is typically modeled as approximately stationary and Gaussian over short time segments, the expectation value of cross-event correlations in the residual strain vanishes when events are independent. Under these conditions, $P_{\text{stack}}(\hat{f})$ behaves as an unbiased estimator of the ensemble residual power spectrum, with detector noise contributions averaging down as $1/\sqrt{N}$. Persistent features that survive this averaging therefore indicate correlated structure in the astrophysical residual component rather than random detector fluctuations.

The choice of a power-based statistic can also be justified from the perspective of detection theory. In standard matched filtering, the optimal statistic is linear in the data when the signal template—including its phase—is known. However, if the post-merger boundary response carries an event-dependent or effectively randomized phase, that phase becomes a nuisance parameter. After marginalizing over unknown phases, the likelihood depends primarily on the

magnitude-squared residual spectrum, so the relevant detection statistic becomes quadratic rather than linear in the data. $P_{\text{stack}}(\hat{f})$ therefore plays the role of a *noncoherent matched filter* in the residual power domain: it maximizes sensitivity to the shared residual power structure across the event ensemble, while uncorrelated noise and randomized phase contributions average down. This places MGWI in the same statistical class as stochastic gravitational-wave background cross-correlation searches, which extract signal from power correlations across independent detectors rather than from phase-coherent waveform matching.

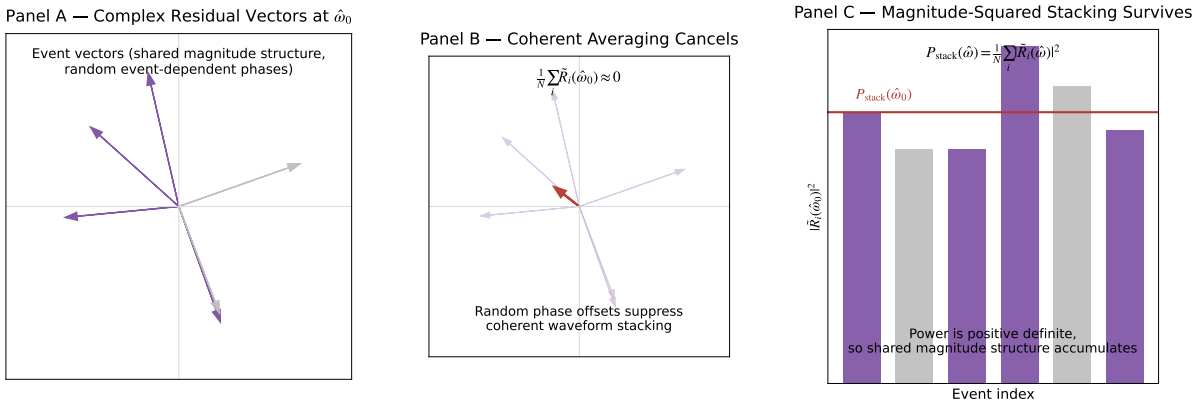

Figure 2: Vector geometry of MGWI stacking. **(A)** At fixed dimensionless frequency $\hat{\omega}_0$, each merger residual can be represented as a complex vector whose magnitude reflects the underlying boundary-response spectrum and whose phase is event-dependent. **(B)** Coherent averaging of the complex residuals suppresses the signal when phases are randomized. **(C)** The magnitude-squared quantities $|\tilde{R}_i(\hat{\omega})|^2$ remain positive and therefore accumulate under stacking. MGWI exploits this geometry by searching for persistent structure in $P_{\text{stack}}(\hat{\omega})$, the ensemble-averaged residual power spectrum.

Figure 2 illustrates this geometry: at a given dimensionless frequency, the complex residual vectors from different events share a common magnitude structure but carry randomized phases. Coherent waveform averaging cancels; magnitude-squared stacking accumulates. Figure 3 demonstrates the resulting sensitivity gain with a synthetic example.

3.7 The Null Control: Time-Shifted Double-Blind

To distinguish genuine spectral structure from instrumental artifacts or algorithmic biases, we define a rigorous falsification mechanism: the **time-shifted null control**.

We repeat the identical stacking algorithm on "off-source" data segments extracted 5–10 seconds before or after the merger events, processed using the same M_f and a_f parameters. This generates a background stack that preserves the spectral noise shape of the detectors but entirely lacks the astrophysical signal. Because the off-source data contains no aligned merger, it lacks the temporal anchor $\tau = 0$; any true horizon signature is destroyed by the misalignment.

Any spectral feature in the on-source stack must be *absent* in the off-source control to be claimed as a horizon signature. If a feature appears in the empty sky control, the methodology is immediately falsified as an instrumental or algorithmic artifact. As a further guard

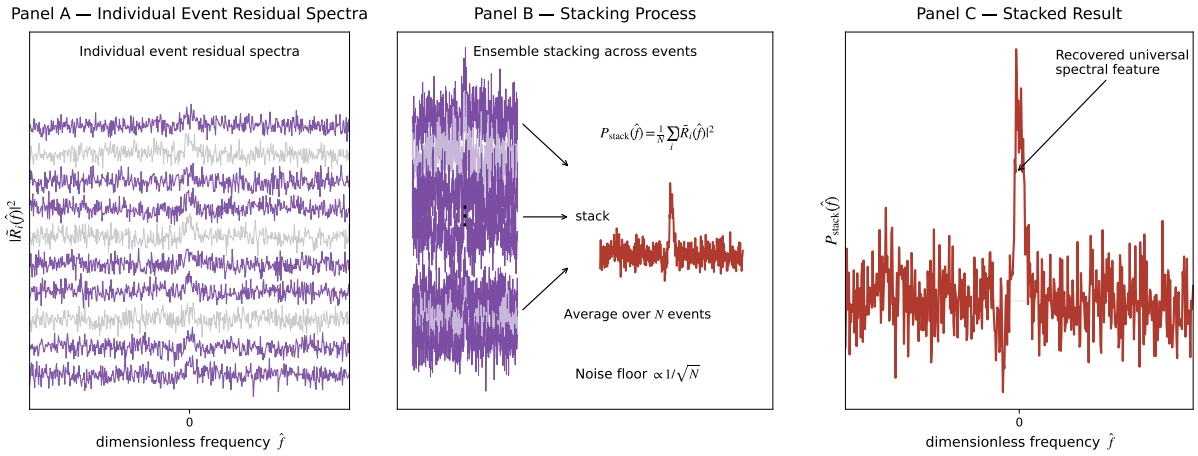

Figure 3: Synthetic demonstration of MGWI stacking sensitivity. **(A)** Simulated residual power spectra for individual gravitational-wave events. Each spectrum contains a weak shared spectral feature that is not clearly visible in any single event due to noise fluctuations. **(B)** Ensemble stacking of residual spectra using the MGWI statistic $P_{\rm stack}(\hat{f}) = \frac{1}{N}\sum_i |\tilde{R}_i(\hat{f})|^2$. Sensitivity improves as the noise floor decreases proportionally to $1/\sqrt{N}$ with each additional event averaged. **(C)** The stacked spectrum reveals the underlying universal spectral feature, which emerges above the noise floor only after combining multiple events.

against non-Gaussian noise transients that may survive the GravitySpy vetting process, a complementary control can be constructed by phase-randomizing the on-source residuals (replacing each event's spectral phase with uniform random draws while preserving the amplitude) and re-stacking; any feature surviving phase randomization is by construction an artifact of the amplitude distribution rather than a coherent spectral structure.

3.8 Mass-Binning as a Trial-Factor Constraint

A key advantage of the dimensionless normalization is its power to constrain the look-elsewhere effect. Terrestrial instrumental noise lines (e.g., the 60 Hz power line, suspension violin modes) are fixed in absolute frequency; they do not scale with the mass of a distant merging black hole. A genuine horizon signature *must* scale as $1/M$ with the remnant mass, appearing at different absolute frequencies for different events but at the *same* dimensionless frequency $\hat{\omega}$ after normalization. The growing mass range of the LVK catalog—recently extended by the detection of a $\sim 240\,M_\odot$ remnant from the most massive binary black hole merger observed to date [46]—strengthens this test by providing wider leverage in absolute frequency space.

By binning the Golden Set by remnant mass and verifying that the spectral feature moves across the absolute frequency band (in Hz) from bin to bin—but remains stationary in $\hat{\omega}$—we impose a kinematic constraint that random terrestrial noise cannot mimic. A Kolmogorov-Smirnov test applied to the distribution of spectral peaks across mass bins can quantify whether they are consistent with a single universal population in the $\hat{\omega}$-domain (rejecting the null at $p < 0.01$) while being inconsistent with the stationary noise background, which cannot scale itself to match the remnant mass of distant astrophysical events.

To quantify the trials penalty: the analysis window of $\tau \approx 500$ yields a frequency resolution of $\Delta\hat{\omega} \approx 2\pi/500 \approx 0.013$. Over a search range of $\hat{\omega} \in [0.1, 2.0]$ (encompassing the fundamental QNM through several overtones), this gives $N_{\rm freq} \approx 150$ independent frequency bins. In a blind single-bin search, the probability of a $\geq 3\sigma$ fluctuation in at least one bin is $p_{\rm false} \approx N_{\rm freq} \times 0.0013 \approx 0.20$—rendering a single-bin 3σ result meaningless. However, if the Golden Set is divided into $k = 3$ independent mass bins and the feature is required to appear at the *same* $\hat{\omega}$ in all bins, the false alarm probability drops to $p_{\rm false} \approx N_{\rm freq} \times (0.0013)^k \approx 3 \times 10^{-7}$, equivalent to a $\sim 5\sigma$ significance. This demonstrates that the mass-binning consistency requirement reduces the effective trial factor by orders of magnitude, converting a statistically weak wide-band search into a highly constrained test.

3.9 Bayesian Model Comparison

To quantify evidence, we employ Bayesian model comparison rather than frequentist p-values, which are susceptible to the look-elsewhere effect [37]. We calculate the Bayes Factor:

$$\mathcal{B}_{10} = \frac{P(\text{Data} \,|\, H_1 : \text{Structured Horizon})}{P(\text{Data} \,|\, H_0 : \text{Vacuum GR})} \tag{13}$$

This integrates over the prior volumes of both models, inherently penalizing H_1 for its additional parameters (location, width, and depth of any spectral feature). On the Jeffreys scale [38], $\mathcal{B}_{10} > 100$ constitutes decisive evidence. Critically, although MGWI is model-agnostic in that it does not presuppose a specific frequency target, the mass-scaling requirement of §3.8 effectively restricts the H_1 prior volume to spectral features that track $1/M$—excluding arbitrary spectral structure and substantially reducing the Occam penalty that would otherwise apply to an unconstrained wide-band search.

To guard against the possibility that imperfect template subtraction of known Kerr modes produces a coherent residual mimicking horizon structure, the Bayesian comparison should include a third hypothesis:

$$H_{\rm KerrSys} : \text{Residual from imperfect multi-mode Kerr subtraction} \tag{14}$$

Under $H_{\rm KerrSys}$, the stacked residual is modeled not as Gaussian noise (H_0) nor as beyond-Kerr structure (H_1), but as the expected residual power from known QNM overtones (particularly the $(2, 2, 1)$ mode) that the template may not perfectly capture. The "Kerr floor" established by the injection study (§4.1) provides the empirical calibration for this hypothesis. The final evidence for horizon structure must satisfy $\mathcal{B}_{1,\rm KerrSys} > 100$—that is, the beyond-Kerr hypothesis must decisively outperform not just pure noise but also the known-physics systematic floor. This three-way comparison (H_0 vs. $H_{\rm KerrSys}$ vs. H_1) is essential for a credible claim of detection.

4 Projected Sensitivity

4.1 Injection Study Design

To estimate the statistical power of the MGWI protocol, we outline a sensitivity study using synthetic data. We emphasize that these are *projected* detection thresholds, not claims of detection from real data.

The study proceeds as follows:

1. Generate $N_{\text{sim}} = 1000$ realizations of colored Gaussian noise modeled on the actual power spectral density curves of the Advanced LIGO and Virgo detectors during O3, capturing the optimal sensitivity band between approximately 50–300 Hz and the seismic and shot-noise walls.
2. Into each realization, inject a synthetic signal: a pseudorandomly phase-scrambled residual containing a localized spectral feature (either an excess or a depletion of power) at a target dimensionless frequency $\hat{\omega}_{\text{target}}$. The phase $\phi(\tau)$ of each injected component is sampled uniformly from $[0, 2\pi]$, simulating the scrambling effect. The amplitude is scaled to $h_{\text{echo}} \sim 10^{-2}\, h_{\text{peak}}$, placing it entirely beneath the single-event noise floor (individual event SNR < 1).
3. Apply standard matched filtering to individual synthetic events. Confirm that the injected signal is undetectable in any single event.
4. Apply the full MGWI phaseless power stacking protocol to ensembles of $N = 10, 20, 50, 100$ events.
5. Quantify recovery significance as a function of N and compare to the null (no injection) distribution. Crucially, the Kerr-only injections (without any beyond-Kerr signal) establish a "Kerr floor"—the residual spectral power attributable to imperfect template subtraction of known features such as the $(2, 2, 1)$ overtone. Any claimed beyond-Kerr detection must exceed this floor with statistical significance; features at or below the Kerr floor are attributed to waveform modeling limitations rather than horizon structure.
6. As a systematic error diagnostic, repeat the full pipeline using an alternative waveform family for the template subtraction step (e.g., substituting SEOBNRv5 for IMRPhenomXHM). Any spectral feature that appears in one waveform family's residuals but vanishes in the other is identified as a template artifact and excluded from the final analysis.

4.2 Expected Scaling

For coherent stacking (where phases are known), the stacked SNR scales as $\sqrt{N}$ relative to a single event. However, for power stacking where phases are unknown, the scaling is less favorable. Yang et al. [22] show that power stacking SNR scales as $N^{1/4}$ rather than $N^{1/2}$ for individually sub-threshold events. The critical difference is that power stacking does not require *any* phase information, making it the only viable approach when the phase is effectively scrambled.

The minimum number of events required for detection depends critically on the (unknown) amplitude of the spectral feature relative to the noise floor. For orientation: assuming a feature at $\sim 1\%$ of peak strain amplitude and worst-case $N^{1/4}$ scaling, order-of-magnitude estimates suggest that $N \sim 50$–100 high-quality BBH events may be required to reach the threshold of statistical significance ($\sim 3\sigma$), with several hundred events needed for a decisive 5σ detection. The GWTC-3 catalog provides a first viable sample for preliminary analysis; the O4/O5 observing runs will substantially expand the available sample. These estimates do not account for potential improvements from optimized spectral kernels or hierarchical population modeling [37].

An important caveat applies. Hu et al. [19] establish a geometric selection rule: a detectable deviation from GR requires both macroscopic spatial coherence ($L_c \sim M$) and classical-level intensity ($\epsilon \gtrsim 10^{-4}$). If horizon modifications are confined to the Planck scale with no long-range correlations, no stacking protocol—including MGWI—can recover the signal. The MGWI protocol is therefore most sensitive to models predicting *macroscopic* boundary structure (e.g., gravastars, long-range fuzzball geometries, or membrane paradigm realizations with $L_c \sim M$), and least sensitive to purely microscopic Planck-scale corrections. A null result from MGWI constrains the former class of models while remaining agnostic about the latter.

A further physical constraint applies to spinning remnants: if the boundary is partially reflective in the superradiant frequency band ($\omega < m\Omega_H$, where Ω_H is the horizon angular velocity), ergoregion instabilities can develop [10]. The absence of observed instabilities in astrophysical black holes already places independent upper bounds on the reflectivity in this band, complementary to the constraints MGWI would provide.

4.3 Distinguishing Signal from Artifact

The injection study also validates the null control. Running the stacking algorithm on time-shifted off-source data with identical parameters must yield a flat, featureless background. The spectral feature must appear exclusively in the on-source data and scale correctly with remnant mass across mass bins. These requirements establish a binary pass/fail criterion: the protocol either detects a feature that survives all controls, or it returns a null result that constrains the amplitude of any universal horizon structure.

4.4 Planned Proof-of-Concept on Public Data

While the synthetic demonstration in Figure 3 validates the stacking logic under controlled conditions, a critical near-term milestone is the application of the full MGWI pipeline to real gravitational-wave data. We outline the planned proof-of-concept analysis, which requires no proprietary data or specialized computing infrastructure.

The test proceeds as follows:

1. Download calibrated strain segments for the 12 Golden Set candidates in Table 2 from GWOSC (`gwosc.org`), together with published posterior samples for remnant mass M_f

and spin a_f from the associated Zenodo data releases.

2. For each event, generate the maximum-likelihood IMRPhenomXHM waveform using PyCBC, subtract it from the on-source strain, and extract the post-merger residual $R_i(t)$.
3. Apply the dimensionless normalization (§3.5), Tukey windowing, and periodogram computation to obtain $P_i(\hat{f})$ for each event.
4. Stack the individual power spectra with inverse-variance weighting to produce $P_{\text{stack}}(\hat{f})$.
5. Repeat identically on time-shifted off-source segments (5–10 s displaced) to produce the null control $P_{\text{null}}(\hat{f})$.

The expected outcome of this first-pass analysis is a *null result*: the on-source stacked spectrum should be statistically consistent with the off-source control at $N \sim 12$, given the $N^{1/4}$ scaling of power-stacking sensitivity. This null result is itself scientifically informative, as it would establish the empirical noise floor and Kerr residual baseline for future analyses with larger catalogs. Any unexpected spectral structure in the on-source stack that is absent in the null control would warrant immediate follow-up with the mass-binning and multi-waveform cross-checks described in §3.8 and §3.4. The proof-of-concept pipeline will be released as open-source analysis scripts to facilitate independent verification and reuse by the community.

5 Discussion and Future Directions

5.1 Operational Roadmap

We propose a three-phase program:

Phase I: Current LVK data (O4/O5). Apply the full MGWI protocol to the expanding BBH catalog. Implement Bayesian hierarchical modeling to constrain the population-level parameters of any spectral feature. Execute strict time-shifted null controls. The detection threshold is $\mathcal{B}_{10} > 100$.

Phase II: Multi-band synthesis (LISA and Einstein Telescope). Extend the protocol to the millihertz band (LISA [39]) and ultra-high-sensitivity band (ET [40] and CE [41]). Supermassive black holes detected by LISA ring at radically lower frequencies but offer months-long integration times. If horizon structure is universal, the dimensionless spectral parameters must be *scale-invariant* across more than six orders of magnitude in mass. This cross-band consistency test would be powerful evidence for (or against) universal boundary physics.

Phase III: Multi-messenger cross-check. If a spectral feature is detected, correlate its inferred boundary properties with electromagnetic observations. Models predicting a physical boundary (e.g., gravastars, membrane paradigm) make specific predictions for the interaction of accreting matter with the horizon surface; the thermodynamic properties inferred from gravitational wave spectroscopy (acoustic impedance, reflectivity) should be consistent with those inferred from Tidal Disruption Events and Fast Blue Optical Transients (emission geometry, temperature distributions). A mismatch would falsify a unified boundary model; consistency would constitute independent confirmation across orthogonal observational channels.

5.2 Interpretation of Results

A *positive* detection—a spectral feature surviving all controls, with $\mathcal{B}_{1,\mathrm{KerrSys}} > 100$—should be interpreted according to an explicit evidential ladder:

1. *First:* The result constitutes evidence for **unresolved universal post-merger residual structure** not fully explained by the current subtraction and noise model. This is the strongest immediate claim the data can support.
2. *Second:* If the feature persists across independent waveform families (IMRPhenomXHM, SEOBNRv5) and independent NR calibration sets, it constitutes evidence for **astrophysical residual structure beyond the current Kerr modeling frontier**.
3. *Third:* If the feature's morphology (frequency localization, bandwidth, spin dependence) is inconsistent with any known Kerr overtone or mode-mixing artifact, it becomes a candidate for **non-Kerr boundary response**—horizon structure in the sense of the competing models in Table 1.

This ladder ensures that the interpretation remains commensurate with the evidence at each stage. A premature claim of "quantum horizon microstructure" from a first-generation MGWI excess would not be warranted; the correct initial framing is the discovery of persistent, universal, post-merger residual structure requiring explanation.

A *null* result would place upper bounds on the amplitude of any universal spectral structure in the horizon, constraining the reflectivity and stiffness parameters of modified horizon models. This is a scientifically valuable outcome: it would establish that if horizon structure exists, it is weaker than a specific threshold, tightening the parameter space for quantum gravity phenomenology.

To make this concrete: for models parameterized by a frequency-dependent reflectivity $|R(\omega)|^2$, the amplitude of the stacked residual excess scales approximately as $|R|^2$ relative to the Kerr baseline. The existing ergoregion instability bounds [10] constrain $|R|^2 \lesssim 0.01$ in the superradiant band for astrophysically stable spinning remnants. The Abedi (2025) meta-analysis [13] places a 90% confidence upper limit on the universal echo amplitude at $A < 0.42$, corresponding roughly to $|R|^2 \lesssim 0.18$. MGWI targets the intermediate regime: reflectivities too small to produce individually detectable echoes ($|R|^2 \lesssim 10^{-2}$) but large enough to emerge from the noise floor after population stacking ($|R|^2 \gtrsim 10^{-4}$, consistent with the Hu et al. [19] geometric selection rule $\epsilon \gtrsim 10^{-4}$). A null result from MGWI with $N \sim 100$ events would push the upper bound on $|R|^2$ below 10^{-3}—the most stringent constraint to date on the acoustic impedance of the event horizon, directly excluding gravastar and hard-surface ECO models with reflectivities above this threshold.

5.3 Connection to Broader Frameworks

The MGWI protocol is motivated by the general principle that stability in dissipative systems often requires a mediating boundary with specific structural properties. The *Singularity-Duality-Triality* (S-D-T) framework [33] proposes that this pattern—a high-symmetry canvas,

a symmetry-breaking tension, and a mediating rule that resolves the tension into a new stable attractor—appears across disparate physical domains, from topological superconductors (Spin(8) triality) to anyonic statistics to the AdS/CFT correspondence. In this context, the event horizon would represent the "Triality"—the stiff boundary mediating the duality between the interior (singularity) and the exterior (vacuum). MGWI provides a concrete, falsifiable empirical test of whether this structural role is realized in nature: does the horizon possess the physical architecture that such a mediating function requires?

This connection is offered as philosophical context, not as a theoretical prediction constraining the analysis. The MGWI protocol is model-agnostic by design; it tests the existence of spectral structure without presupposing its origin.

5.4 Data, Software, and Computational Requirements

The MGWI protocol is designed to be executable with publicly available data and open-source software.

Data. Calibrated strain data for all GWTC-3 events are publicly available through the Gravitational Wave Open Science Center (GWOSC, `gwosc.org`). Parameter estimation posterior samples—including remnant mass M_f, spin a_f, and sky location—are released on Zenodo as part of each catalog publication [2]. The O4 data release policy continues this practice. No proprietary LVK data products are required for the core stacking analysis; all strain segments and posterior samples for the Golden Set events in Table 2 are accessible without collaboration membership.

Software. Template subtraction uses standard open-source tools: PyCBC [50] for matched filtering and waveform generation, and Bilby [51] for Bayesian parameter estimation. The stacking algorithm itself requires only standard numerical libraries (NumPy, SciPy) for Fourier transforms, periodogram estimation, and weighted averaging. No custom software development is required beyond a straightforward scripting layer connecting these existing tools.

Computational cost. The stacking pipeline is computationally inexpensive: for $N \sim 50$–100 events, the Fourier transforms, normalization, and weighted power averaging require minutes on a standard workstation. The injection study (§4.1) is more demanding—$N_{\rm sim} = 1000$ realizations $\times$ multiple waveform families—but remains tractable on a modest computing cluster or cloud resource (estimated $\sim 10^3$–10^4 CPU-hours). The pairwise lensing cross-correlations (§2.1), which scale as N^2, become computationally expensive only in the next-generation detector era ($N > 10^4$) and are not required for the primary Phase I analysis.

Timeline. The proof-of-concept analysis described in §4.4—applying the full pipeline to ~ 12 Golden Set events from GWOSC—is executable within weeks using existing public data and open-source tools. The full injection study (§4.1) and systematic cross-checks across waveform families would require approximately 3–6 months of dedicated analysis effort.

6 Conclusion

The null results of current echo searches may reflect not the absence of horizon structure but the limitations of phase-coherent search strategies in a regime where phase coherence is degraded. The central contribution of this paper is the observation that the correct search domain is the *power spectrum*, not the *waveform*, and that the correct unit of analysis is the *population*, not the individual event.

Multiple Gravitational Wave Interferometry extends the principle of aperture synthesis into the gravitational wave domain. By stacking the phaseless power spectral densities of post-merger residuals across the catalog, normalized onto a dimensionless manifold scaled by remnant mass, we convert the "silence" following each merger into a high-SNR spectroscopic probe of the horizon's structure—or, in the event of a null result, a rigorous upper bound on its reflectivity. The protocol includes three independent controls against false positives: a time-shifted null control for instrumental artifacts, a phase-randomized on-source control for non-Gaussian amplitude artifacts, and a multi-waveform cross-check for template systematic errors. A three-way Bayesian comparison (H_0: noise, H_{KerrSys}: template systematics, H_1: horizon structure) ensures that any claimed detection must outperform not just the noise floor but also the known-physics systematic floor. The mass-binning consistency requirement further constrains the look-elsewhere effect by demanding that any candidate feature scale kinematically with remnant mass.

The framework additionally enables cross-catalog applications including lensed event identification, stochastic background mapping, and intensity interferometry sensitive to the quantum state of primordial radiation.

The MGWI test has two clear outcomes. A null stacked spectrum would place quantitative upper bounds on horizon reflectivity and constrain models predicting horizon-scale structure. A statistically significant residual spectral feature, by contrast, would indicate the presence of a universal boundary response not captured by classical Kerr ringdown models.

Because MGWI tests the binary distinction between a classical absorber and *any* structured boundary, its results bear on a wide spectrum of active research programs. A positive detection would provide empirical fuel for models predicting horizon-scale structure, including: the fuzzball program in string theory [30]; the firewall and complementarity debate [45, 15]; the gravastar hypothesis [48]; wormhole mimicker scenarios [52]; the soft quantum structure and nonviolent nonlocality proposals [53]; partially reflective ECO models [10, 34]; and membrane paradigm realizations with non-trivial horizon impedance [49]. Conversely, a null result would constrain all of these simultaneously, establishing that the event horizon's reflectivity falls below a quantitative threshold across the full range of proposed modifications. MGWI therefore provides a common empirical test relevant to multiple theoretical proposals concerning horizon structure.

With the GWTC-3 catalog providing the first viable sample for a preliminary analysis, and the O4/O5 observing runs projected to expand the high-quality BBH catalog substantially, MGWI provides a near-term, falsifiable test of the most basic open question about the event horizon: is

it the featureless classical boundary predicted by General Relativity, or does it possess physical structure?

References

[1] Abbott, B. P., et al. (LIGO Scientific Collaboration and Virgo Collaboration). (2016). Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6), 061102.

[2] Abbott, R., et al. (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration). (2023). GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run. *Phys. Rev. X*, 13(4), 041039.

[3] Abbott, B. P., et al. (LIGO Scientific Collaboration and Virgo Collaboration). (2017). GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16), 161101.

[4] Aasi, J., et al. (LIGO Scientific Collaboration). (2015). Advanced LIGO. *Class. Quantum Grav.*, 32(7), 074001.

[5] Carroll, S. M. (2019). *Spacetime and Geometry: An Introduction to General Relativity.* Cambridge University Press.

[6] Blanchet, L. (2014). Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries. *Living Rev. Relativ.*, 17, 2.

[7] Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A., & Creighton, J. D. E. (2012). FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries. *Phys. Rev. D*, 85(12), 122006.

[8] Klimenko, S., et al. (2005). Coherent method for detection of gravitational wave bursts. *Phys. Rev. D*, 72(12), 122002.

[9] Giesler, M., et al. (2019). Black Hole Ringdown: The Importance of Overtones. *Phys. Rev. X*, 9(4), 041060.

[10] Cardoso, V., & Pani, P. (2017). Tests for the existence of black holes through gravitational wave echoes. *Nature Astronomy*, 1(9), 586–591.

[11] Abedi, J., Dykaar, H., & Afshordi, N. (2017). Echoes from the Abyss: Tentative evidence for Planck-scale structure at black hole horizons. *Phys. Rev. D*, 96(8), 082004.

[12] Westerweck, J., et al. (2018). Low significance of evidence for black hole echoes in gravitational wave data. *Phys. Rev. D*, 97(12), 124037.

[13] Abedi, J. (2025). Search for echoes on the edge of quantum black holes. *Class. Quantum Grav.*, 42(20), 205004.

[14] Sekino, Y., & Susskind, L. (2008). Fast Scramblers. *J. High Energy Phys.*, 2008(10), 065.

[15] Susskind, L., Thorlacius, L., & Uglum, J. (1993). The Stretched Horizon and Black Hole Complementarity. *Phys. Rev. D*, 48(8), 3743–3761.

[16] Hayden, P., & Preskill, J. (2007). Black holes as mirrors: quantum information in random subsystems. *J. High Energy Phys.*, 2007(09), 120.

[17] Maldacena, J., Shenker, S. H., & Stanford, D. (2016). A bound on chaos. *J. High Energy Phys.*, 2016(08), 106.

[18] Bouland, A., Fefferman, B., & Vazirani, U. (2019). Computational pseudorandomness, the wormhole growth paradox, and black hole holography. arXiv:1910.14646.

[19] Hu, H.-W., Fang, C.-J., & Guo, Z.-K. (2026). Waveform stability of black hole ringdown with stochastic horizon structure. arXiv:2602.08034.

[20] Brown, A. R., & Susskind, L. (2018). Second law of quantum complexity. *Phys. Rev. D*, 97(8), 086015.

[21] Yang, H., Yagi, K., Blackman, J., Lehner, L., Paschalidis, V., Pretorius, F., & Yunes, N. (2017). Black Hole Spectroscopy with Coherent Mode Stacking. *Phys. Rev. Lett.*, 118(16), 161101.

[22] Yang, H., et al. (2018). Gravitational wave spectroscopy of binary neutron star merger remnants with mode stacking. *Phys. Rev. D*, 97(2), 024049.

[23] Hanbury Brown, R., & Twiss, R. Q. (1954). A new type of interferometer for use in radio astronomy. *Philos. Mag.*, 45, 663–682.

[24] Takahashi, R., & Nakamura, T. (2003). Wave Effects in Gravitational Lensing of Gravitational Waves from Chirping Binaries. *Astrophys. J.*, 595, 1039–1051.

[25] Cusin, G., Pitrou, C., & Uzan, J.-P. (2017). Anisotropies of the astrophysical gravitational-wave background. *Phys. Rev. D*, 96(10), 103019.

[26] Kanno, S., Soda, J., & Watanabe, J. (2019). Hanbury-Brown–Twiss interferometry of primordial gravitational waves. *J. Cosmol. Astropart. Phys.*, 2019, 023.

[27] Ryle, M., & Hewish, A. (1960). The synthesis of large radio telescopes. *Mon. Not. R. Astron. Soc.*, 120, 220–230.

[28] Cannon, K., et al. (2012). Toward real-time detection of gravitational waves from compact binary coalescence. *Astrophys. J.*, 748, 136.

[29] Gabbard, H., Williams, M., Astone, F., & Meacher, D. (2018). Matching matched filtering with deep networks for gravitational-wave astronomy. *Phys. Rev. Lett.*, 120(14), 141103.

[30] Bena, I., Martinec, E. J., Mathur, S. D., & Warner, N. P. (2022). Fuzzballs and Microstate Geometries: Black-Hole Structure in String Theory. arXiv:2204.13113.

[31] Baumgarte, T. W., & Shapiro, S. L. (2010). *Numerical Relativity: Solving Einstein's Equations on the Computer*. Cambridge University Press.

[32] Christodoulou, D. (1991). Nonlinear nature of gravitation and gravitational-wave experiments. *Phys. Rev. Lett.*, 67, 1486–1489.

[33] Morales, J. R. (2026). A Foundational Framework for Natural Philosophy: Singularity, Duality, and Triality—A Falsifiable Heuristic Model of Emergent Stability.

[34] Cardoso, V., Franzin, E., & Pani, P. (2016). Is the Gravitational-Wave Ringdown a Probe of the Event Horizon? *Phys. Rev. Lett.*, 116(17), 171101.

[35] García-Quirós, C., et al. (2020). Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries. *Phys. Rev. D*, 102(6), 064002.

[36] Zevin, M., et al. (2017). Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science. *Class. Quantum Grav.*, 34(6), 064003.

[37] Thrane, E., & Talbot, C. (2019). An introduction to Bayesian inference in gravitational-wave astronomy. *Publ. Astron. Soc. Aust.*, 36, e010.

[38] Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford University Press.

[39] Amaro-Seoane, P., et al. (2017). Laser Interferometer Space Antenna. arXiv:1702.00786.

[40] Maggiore, M., et al. (2020). Science case for the Einstein Telescope. *J. Cosmol. Astropart. Phys.*, 2020, 050.

[41] Reitze, D., et al. (2019). Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. *Bull. Am. Astron. Soc.*, 51, 035.

[42] Abbott, R., et al. (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration). (2021). Search for anisotropic gravitational-wave backgrounds using data from Advanced LIGO's and Advanced Virgo's first three observing runs. *Phys. Rev. D*, 104(2), 022005.

[43] Abbott, R., et al. (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration). (2021). Tests of General Relativity with Binary Black Holes from the second LIGO-Virgo Gravitational-Wave Transient Catalog. *Phys. Rev. D*, 103(12), 122002.

[44] Cutler, C., & Flanagan, É. E. (1994). Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform? *Phys. Rev. D*, 49(6), 2658–2697.

[45] Almheiri, A., Marolf, D., Polchinski, J., & Sully, J. (2013). Black Holes: Complementarity vs. Firewalls. *J. High Energy Phys.*, 2013(02), 062.

[46] Abbott, R., et al. (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration). (2025). GW231123: Detection of the most massive binary black hole merger. Presented at GR-Amaldi 2025, Glasgow.

[47] Abac, A. G., et al. (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration). (2025). Black Hole Spectroscopy and Tests of General Relativity with GW250114. *Phys. Rev. Lett.*, 135(11), 111403. arXiv:2509.08099.

[48] Mazur, P. O., & Mottola, E. (2004). Gravitational vacuum condensate stars. *Proc. Natl. Acad. Sci.*, 101(26), 9545–9550.

[49] Thorne, K. S., Price, R. H., & Macdonald, D. A. (1986). *Black Holes: The Membrane Paradigm*. Yale University Press.

[50] Usman, S. A., Nitz, A. H., Harry, I. W., et al. (2016). The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quantum Grav.*, 33(21), 215004.

[51] Ashton, G., Hübner, M., Lasky, P. D., Talbot, C., et al. (2019). Bilby: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy. *Astrophys. J. Suppl.*, 241(2), 27.

[52] Damour, T., & Solodukhin, S. N. (2007). Wormholes as Black Hole Foils. *Phys. Rev. D*, 76(2), 024016.

[53] Giddings, S. B. (2013). Nonviolent nonlocality. *Phys. Rev. D*, 88(6), 064023.

A Reproducible Code for Figure 3

The synthetic demonstration in Figure 3 was generated using the following Python script (NumPy, Matplotlib). The random seed is fixed for exact reproducibility. The full styling code is available as a Google Colab notebook.[1]

```
import numpy as np
import matplotlib.pyplot as plt

# Reproducibility
rng = np.random.default_rng(7)

# --- Synthetic data ---
n_events = 10
x = np.linspace(-1.2, 1.2, 500)

# Weak universal feature near fhat = 0
sigma = 0.03
feature = 0.22 * np.exp(-(x / sigma) ** 2 / 2)

spectra = []
for i in range(n_events):
    noise = rng.normal(0, 0.07, size=x.size)
    baseline = 0.02 * np.sin(
        2 * np.pi * (i + 1) * (x + 1.2) / 6.0
```

[1] https://colab.research.google.com/drive/1DXAC8Y5Idud17g2rCJTFDBeBp8K7JrJI?usp=sharing

```
    )
    amp = 0.6 + 0.25 * rng.random()
    y = 0.42 + baseline + noise + amp * feature
    y = np.clip(y, 0.02, None)
    spectra.append(y)

spectra = np.array(spectra)
stacked = spectra.mean(axis=0)

# --- Plot (simplified; full styling in supplementary Colab) ---
fig, axes = plt.subplots(1, 3, figsize=(15, 5))

# Panel A: individual spectra (offset waterfall)
for i, y in enumerate(spectra):
    axes[0].plot(x, y + i * 0.34, lw=0.9)
axes[0].set_xlabel(r"$\hat{f}$")
axes[0].set_ylabel(r"$|\tilde{R}_i(\hat{f})|^2$")
axes[0].set_title("Individual Residual Spectra")

# Panel B: overlay
for y in spectra:
    axes[1].plot(x, y, alpha=0.3, lw=0.7)
axes[1].plot(x, stacked, color="red", lw=2, label="stacked")
axes[1].set_xlabel(r"$\hat{f}$")
axes[1].set_title("Stacking Process")
axes[1].legend()

# Panel C: stacked result
axes[2].plot(x, stacked, color="red", lw=2)
axes[2].set_xlabel(r"$\hat{f}$")
axes[2].set_ylabel(r"$P_{\mathrm{stack}}(\hat{f})$")
axes[2].set_title("Stacked Result")

plt.tight_layout()
plt.savefig("mgwi_figure2.pdf")
```