

Impact of detector calibration accuracy on black hole spectroscopy

Mallika Sinha

Supervisor: Dr. Ling Sun

Research School of Astronomy and Astrophysics
The Australian National University

Declaration

This is to certify that to the best of my knowledge, the content of this thesis is an account of research undertaken during my honours year with the Research School for Astronomy and Astrophysics at The Australian National University. This thesis has not been submitted for any other purpose and no work was done prior to the start of the program.

The intellectual content presented in this thesis, including the code, is a product of my own work, unless where otherwise acknowledged. All such information and materials are properly referenced and publicly available barring the examples of physical calibration errors shown in Fig. 10, which are obtained by private communication.

MS

Mallika Sinha
May 2024

Acknowledgements

I could not have successfully completed this journey without the help of my supervisor, Dr. Ling Sun, who has contributed endless time and patience towards this project. Thank you so much for guiding me through my honours year, I appreciate all the assistance you have provided! I would also like to extend my gratitude towards Dr. Sizheng Ma, without whom, I would definitely have spent much more time being lost. I am extremely grateful for all the help and discussions we had, me, and my code, thank you!

Many thanks to my fellow honours student, Ashley Hai Tung Tan, and all the tea we drank throughout the year that kept me sane and hydrated. I am also grateful to my friend, and roommate, Alice for all her support (sorry for all the rants in the kitchen). I would also like to highlight the importance of the OzSTAR Supercomputing facility at the Swinburne University of Technology in completing the simulations needed for this thesis. Thank you for all the time!

Abstract

Black hole (BH) spectroscopy methods aim to analyse the ringdown modes in the gravitational-waves (GWs) emitted by the remnant BH created in a compact binary coalescence to measure the properties of the remnant BH, and to test Einstein’s general theory of relativity in the extreme gravity regime. The outcome of BH spectroscopy heavily depends on the accuracy and robustness of the analysis, which, in turn, depends on a number of factors, including the calibration accuracy of the GW detector’s output data. Inaccurately calibrated data may lead to biased results in the downstream analysis. Investigating and quantifying the impact of calibration errors on BH ringdown analyses is essential in assessing the capability of current and future GW observatories for precision tests of general relativity. In this study, we employ numerical relativity waveforms of BH ringdown signals and simulate tunable, physically-motivated calibration errors to investigate the impact of miscalibrated data on BH spectroscopy using a powerful analysis tool, the rational quasinormal mode filter. We find that the current state-of-the-art calibration accuracy, with a magnitude error within 10% and a phase error within 10 deg in the most sensitive frequency band, does not bias BH ringdown analyses with current-generation observatories, but does not satisfy the requirements for BH spectroscopy with next-generation GW observatories. For analysing events with a relatively high signal-to-noise ratio, e.g., a source at 2 Gpc observed in a next-generation GW detector, calibration systematic errors cannot exceed $\sim 8\%$ in magnitude and ~ 8 deg in phase, in order to not bias the ringdown analysis. In a more extreme, high signal-to-noise ratio regime, more stringent requirements for calibration accuracy need to be satisfied. The results obtained in this study, albeit focused on a particular aspect of BH spectroscopy, may provide quantified information for calibration accuracy and precision requirements to allow future precision astrophysical studies in next-generation observatories.

Contents

| | |
|---|-----|
| Declaration | i |
| Acknowledgements | ii |
| Abstract | iii |
| 1 Introduction | 1 |
| 2 Gravitational waves and black holes | 2 |
| 2.1 Gravitational waves | 2 |
| 2.2 Compact binary coalescences | 5 |
| 2.3 Remnant black holes and quasinormal modes | 6 |
| 2.4 Black hole spectroscopy | 8 |
| 3 Gravitational-wave observatories and detector calibration | 9 |
| 3.1 Laser interferometer gravitational-wave detectors | 9 |
| 3.2 Current and next generation ground-based gravitational-wave observatories | 12 |
| 3.2.1 Current generation observatories | 12 |
| 3.2.2 Next generation observatories | 13 |
| 3.3 Detector calibration errors and uncertainty | 16 |
| 4 Impact of calibration errors on quasinormal mode analyses | 19 |
| 4.1 Rational quasinormal mode filter | 19 |
| 4.2 Calibration error profiles | 24 |
| 4.3 Impact of calibration errors on signal waveforms | 26 |
| 4.3.1 Miscalibrating the signal waveforms | 26 |
| 4.3.2 Applying the rational filter to miscalibrated waveforms | 28 |

| | | |
|----------|---|-----------|
| 4.3.3 | Quantifying the impact of calibration errors without additive noise | 34 |
| 4.4 | Simulations in white Gaussian noise | 37 |
| 5 | Prospects for current and next-generation observatories | 41 |
| 5.1 | Analyses with current-generation observatories | 41 |
| 5.2 | Calibration requirements for next-generation observatories | 45 |
| 5.3 | Events in extreme signal-to-noise regime | 48 |
| 6 | Conclusion | 50 |
| A | Appendix | 52 |
| A.1 | Estimates of (M, χ) with calibration errors at different frequencies | 52 |
| A.2 | Analysis of waveforms with two fundamental modes | 52 |

1 Introduction

Einstein’s general theory of relativity (GR) has long predicted the existence of black holes (BHs) as extremely compact celestial bodies, alongside their remarkable influence on the spacetime around them. The merging of two such massive bodies emits gravitational waves (GWs) before, during, and after the tremendous collision, providing a singular window into the extreme physics on display. These signals are encoded with information about the properties of spacetime in the high-velocity, high-energy, and strong gravitational field regime. The final stage of such events sees the formation of disturbed remnant BHs, which eventually ring down into stable states. These post-merger ringdown GW signals are predicted by BH perturbation theory to be described by a superposition of quasinormal modes (QNMs). The properties of these QNMs, namely their frequencies and damping times, are solely determined by the intrinsic BH properties of mass and spin, as dictated by the no-hair theorem. BH spectroscopy, through the analysis of these ringdown signals, offers a powerful means of probing linear and nonlinear BH dynamics, and testing the validity of GR in extreme environments, seen in no other situation in the Universe and none of which are fully understood [1].

GWs are predicted to span a broad spectrum of frequencies, extending past the kilohertz range from frequencies as small as $\sim 10^{-17}$ Hz, stemming from the cosmic microwave background polarisation. Ground-based GW observatories are complex laser interferometers that aim to measure the distortion of spacetime caused by the GWs passing through them, with frequencies in the audio band around 10 Hz to 10^3 Hz. The GWs resulting from, e.g., compact binary coalescences (CBCs), pulsars, and supernovae explosions, feature heavily in this band. The current generation of GW observatories have, over the span of their first three observing runs, detected 90 CBC merger events [2–5], over 80 of which are binary BH (BBH) mergers. The next generation of ground-based GW observatories, armed with major upgraded technologies and new facilities, aim to increase the overall sensitivity across the audio band and will allow for extremely precise tests of GR and probes of fundamental physics.

Ground-based GW observatories operate as null experiments with a differential arm length feedback control loop. They output a time series of dimensionless strain data, used for astrophysical studies, which are reconstructed by modelling the detector’s response in a calibration procedure. Due to the imperfect modelling of the detector’s response and statistical variations, systematic errors and statistical uncertainties present in calibrated strain data may impact the astrophysical analyses conducted using the data. Different analyses may be susceptible to different types, or levels, of calibration errors. Understanding what levels of calibration accuracy are required for robust astrophysical analyses is essential to provide a guideline of calibration techniques and requirements for the next-generation detectors.

This study aims to investigate the potential impact of detector calibration errors and quantify the calibration accuracy required in next-generation GW observatories for BH

spectroscopy studies, using a BH ringdown analysis tool called the rational QNM filter [6–8]. We simulate BH ringdown signals using the Simulating Extreme Spacetimes (SXS) catalogue of numerically calculated waveforms [9, 10] and evaluate the recovery of the BH properties via QNM analysis with and without the existence of calibration errors, and under different noise situations. A set of physically-motivated artificial calibration errors are generated and investigated in this study with reference to ‘typical’ errors witnessed in the Advanced LIGO (aLIGO) detectors during the third observing run (O3) of the LIGO-Virgo-KAGRA. Noise is an unavoidable characteristic in any observation and must be considered when simulating real signal analyses. This study step-by-step investigates the impact of calibration errors by analysing pure BH ringdown signal waveforms with zero noise, simulated signals in white Gaussian noise, and simulated signals in detector coloured Gaussian noise. Although focused on a specific area of astrophysical analyses, understanding calibration error tolerance in BH ringdown studies is essential for research and development towards desired calibration accuracy and precision in future GW observatories.

In chapter 2, we revisit the theoretical foundations of GWs and BHs and introduce QNMs and BH spectroscopy methods. In chapter 3, we describe the basics of GW detectors, the calibration procedure, and the calibration errors. In chapter 4, we investigate the impact of calibration errors on QNM analyses by reviewing the QNM filtering method and conducting simulated analyses in various noise setups and calibration error scenarios. The requirements for calibration accuracy with current and next-generation observatories are presented in chapter 5. Finally, we summarise the results and outline potential avenues for future work in chapter 6.

2 Gravitational waves and black holes

In this section, we describe the theoretical foundations of GWs in Sec. 2.1. We review the observations of compact binary coalescences in Sec. 2.2. We then describe BH ringdown and QNMs in Sec. 2.3, and introduce BH spectroscopy in Sec. 2.4.

2.1 Gravitational waves

GWs can be considered as wave-like solutions to the Einstein field equations¹ that relate the nature of spacetime with the matter fields that reside in it. The most general formulation

¹Please note we do not use SI or CGS units for most of this study, instead, mass-scaled units are used, often referred to as simulation units. Wherever necessary, the units are defined and the physical results are converted to SI units for context. For the following discussion on the Einstein field equations, we use units in which $G = 1 = c$, such that the mass, space, and time have the same units. When referring to tensor notations, the Greek alphabet letters range over spacetime indices, i.e., from 0 to 3 to include the time section of the coordinates while the Latin alphabet letters range only over space, i.e., from 1 to 3.

of the Einstein field equations is given by:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu}, \quad (1)$$

where $G_{\mu\nu}$ is the Einstein tensor that describes the curvature of spacetime, Λ is the cosmological constant, $g_{\mu\nu}$ is the metric tensor, and $T_{\mu\nu}$ is the energy-momentum tensor that describes continuous matter distributions and fields. Outside of the study of cosmology, the Λ term can be ignored [11]. The equations then become:

$$G_{\mu\nu} = 8\pi T_{\mu\nu}. \quad (2)$$

The component equations of Eq. (2) form a set of ten coupled equations, non-linear in the metric and its first and second derivatives. Expanding on the tensor notation and keeping in mind that $\nabla_a G^{ab} = 0$ and $\nabla_a T^{ab} = 0$ hold independent to the Einstein field equations, we can recover six independent non-linear partial differential equations to the second order in the metric,

$$R_{\mu\nu} = 8\pi \left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right). \quad (3)$$

Here, $R_{\mu\nu}$ is the Ricci tensor, and T is the trace of the energy-momentum tensor. The Ricci tensor is a lower-rank tensor defined from the Riemann curvature tensor, which is related to the relative acceleration of two close-by geodesics [11].

In a weak, linearised gravitational field where the metric $g_{\mu\nu}$ only slightly deviates from a flat Minkowski metric, $\eta_{\mu\nu}$, we can write

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad (4)$$

where $h_{\mu\nu}$ is a small perturbation, i.e., $\|h_{\mu\nu}\| \ll 1$. This condition necessitates the gravitational field as weak and the coordinate system as Cartesian-like [12]. Given the small nature of the perturbation, its higher-order terms and derivatives can be considered negligible. The non-linear dependence of the Einstein tensor on the metric tensor is reduced and the equation can be treated in a linearised regime.

We can further simplify the equation by carefully selecting an appropriate coordinate system, known as gauge transformations in GR. Here, we use a class of gauges that satisfy the Lorentz gauge condition [12]:

$$\partial^\mu \bar{h}_{\mu\nu} = 0, \quad (5)$$

where

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h^\alpha{}_\alpha. \quad (6)$$

This is called the trace-reversed perturbation, from the fact that $\bar{h}^\alpha{}_\alpha = -h^\alpha{}_\alpha$. Einstein's equations can then be reduced to:

$$\left(\nabla^2 - \frac{1}{c^2} \partial_{tt} \right) \bar{h}_{\mu\nu} = -16\pi T_{\mu\nu}, \quad (7)$$

which in a vacuum becomes

$$\left(\nabla^2 - \frac{1}{c^2} \partial_{tt} \right) \bar{h}_{\mu\nu} = 0. \quad (8)$$

Notice that Eq. (8) is akin to the classical wave equation, where $\bar{h}_{\mu\nu}$ describes the wave, a GW [11]. We then transform the metric perturbation into the traverse-traceless (TT) gauge, specialising to the asymptotically flat solutions for Eq. (8) where $h_{\mu\nu} \rightarrow 0$ as the distance approaches infinity. Imposing the Lorentz and TT gauge leaves us with only two free, physically meaningful components. In this gauge we consider the metric perturbation to be purely spatial and traceless [12]:

$$h_{tt} = h_{ti} = 0 \quad \text{and} \quad h = h_i^i = 0. \quad (9)$$

A plane-wave-like solution can then be found for the perturbation, giving us the following expression for the GW (where we choose the direction of travel to be in the $+z$ -direction):

$$h_{\mu\nu} = \bar{h}_{\mu\nu} = A_{\mu\nu} e^{i[k(z-ct)-\phi_0]}, \quad A_{\mu\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (10)$$

Here, $A_{\mu\nu}$ is an amplitude matrix, k is the positive (strictly non-zero) component of the wave vector, k^μ (the component of the wave vector in the direction of travel, in this case k^3), ϕ_0 is the phase of wave at $k_\mu x^\mu = 0$, and c is the speed of light. Given there are only two degrees of freedom in the amplitude matrix, we can factor $A_{\mu\nu}$ into two basis tensors, corresponding to the two polarisation states of GWs, the plus (“+”) and cross (“ \times ”) polarisations. Given the periodic nature of a GW source, these polarisations are typically 90 degrees out of phase temporally. We can use k to specify the gravitational wavelength, λ , and angular frequency, ω :

$$\lambda \equiv \frac{2\pi}{k}, \quad \omega \equiv ck. \quad (11)$$

The wave-like solution above has been found using the assumption that $T_{\mu\nu} = 0$, which gives solutions for GWs propagating in a vacuum. In the presence of a source, the equation would have a non-zero stress-energy tensor [12]. Given the assumption of a weak field and the TT gauge, a multipolar expansion of the time-varying mass distribution, or stress-energy tensor, can be used to find the perturbation of spacetime in this situation. A full derivation can be found in Tiec and Novak [11]. Monopolar and dipolar gravitational radiation is forbidden due to the mass and linear momentum conservation laws. The next contributing multipole is the quadrupole moment, given by²:

$$Q^{ij} = \int_{V'} dV' T^{00} x'^i x'^j, \quad (12)$$

where V' is the source volume defined such that within the volume, $T^{ij} \neq 0$, but $T^{ij} = 0$ at the boundary and outside the volume, and x' is the integration variable. The radiative, TT part of the full solution, $\bar{h}'^{\alpha\beta}$, is found to be:

$$h_{TT}^{ij} \equiv \frac{2G}{c^4 r} \partial_{tt} Q_{TT}^{ij}, \quad (13)$$

²Note that there is no consistent notation for the mass quadrupole moment in literature.

where r is the distance from the source, G is the gravitational constant, and ∂_{tt} represents the double partial time derivative. An order-of-magnitude estimate of the simplified quantities, such as mass and radius, gives us:

$$h_{TT} \sim \frac{GM}{c^2 r} \left(\frac{v}{c} \right)^2, \quad (14)$$

where M is the mass of the source, and v is the velocity of source matter. For a BBH system with a total mass of $M = 72M_\odot$, and an orbital velocity of $v = 0.4c$, at a distance of $r = 400$ Mpc, Eq. (14) gives a GW strain on the order of $h \sim 10^{-21}$.³

2.2 Compact binary coalescences

Compact binary coalescences (CBCs), the major GW source for current ground-based GW observatories, involve the inspiral and merger/collapse of two compact objects, most notably BHs and neutron stars (NSs). Contemporary research places a significant focus on CBCs and the extreme physics that govern them. Ground-based GW observatories are designed to be sensitive to stellar-mass CBC events due to limits on the sensitivity that can be achieved in the full frequency band. In Fig. 1a, the three major BBH merger stages (inspiral, merger, and ringdown) are depicted alongside the estimated GW strain amplitude from GW150914 (grey) and a numerically simulated model (red) of the signal. The sections of the signal before and after the merger are noticeably different, with the latter corresponding to the settling of a newly formed remnant BH. Fig. 1b depicts the GW150914 signal present in the Hanford (orange) and Livingston (blue) aLIGO observatories, accounting for the different arrival times and relative positions in the detectors by inverting the waveform. The difference between the template in Fig. 1a and observed signals in Fig. 1b can be attributed to noise in the detectors.

To date, three observing runs (O1, O2, O3) have been completed, with the fourth observing run (O4) currently ongoing. There have been a total of 90 confirmed detections,⁴ comprising of 83 BBH mergers, two binary neutron star (BNS) mergers, three neutron star-black hole (NSBH) mergers, and two mystery mergers⁵ [2–5]. O4 began on the 24th of May 2023, and while only one detection has been published [13], the rate of observing significant events is almost doubled compared to O3, one every couple of days. Our catalogue of detections will most likely be tripled by the end of O4, marking the start of a new GW astronomy era.

³Here we keep G and c in the equation, and use the SI units for an order of magnitude estimate of the strain.

⁴Note that results for O4 have not yet been published.

⁵Mystery mergers involve at least one celestial body with a mass approximately in-between $3\text{--}5M_\odot$, between the most massive known NSs and least massive known BHs.

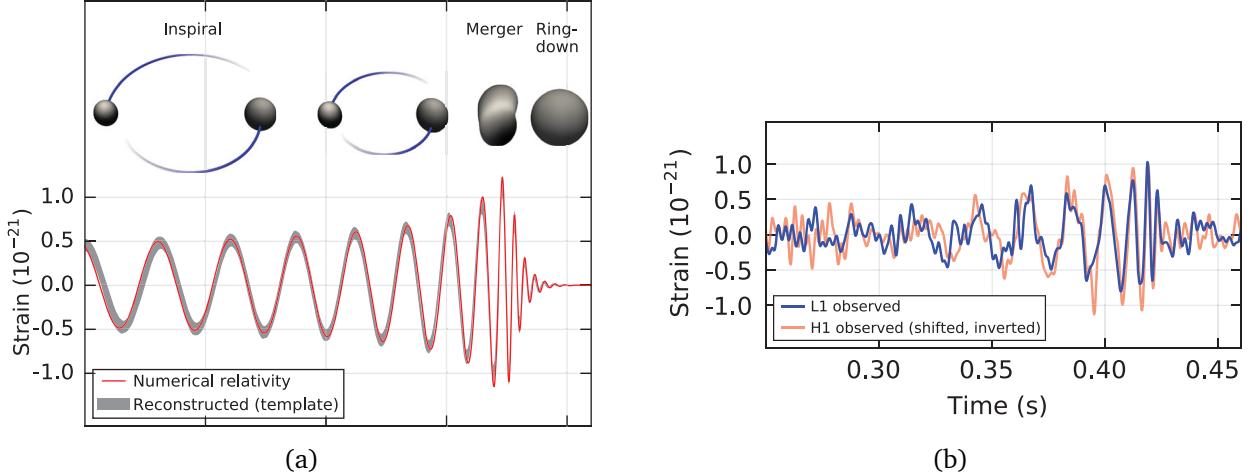


Figure 1: The first observed GW signal from two merging BHs, GW150914. Left: the inspiral, merger, and ringdown stages; the waveform computed from numerical relativity (red) and the reconstructed template (grey). Right: the signal as observed by the Hanford (orange) and Livingston (blue) aLIGO observatories. The arrival time and relative orientation of the two observatories are accounted for when aligning the signals in the two observatories on the right. Credit: Abbott et al. [1]

2.3 Remnant black holes and quasinormal modes

The existence of BHs can be regarded as a natural consequence of GR, and has been predicted for over a century as exact solutions to the Einstein field equations with certain necessary conditions. That is, the solution must describe a spacetime with a gravitational field such that particles (including light), within a boundary region, called the event horizon, cannot escape [14]. The region interior is thus causally disconnected from the exterior, rendering any information from the interior inaccessible [15]. The first prediction of BHs was made by Schwarzschild in 1916 with his solution to the Einstein field equations describing a regular, asymptotically-flat, static, and spherically-symmetric vacuum spacetime [16]. This solution describes any static, non-spinning BH existing in an approximate vacuum and relies on one variable, the total mass of the BH. Remarkably, BHs in four-dimensional, asymptotically flat spacetimes can be fully characterised by just three parameters, mass, charge, and angular momentum [17]. BHs with all three parameters can be described by the Kerr-Newman spacetime solution discovered in 1965 [18]. Many summarise the simplicity of the parameter space by describing BHs as possessing ‘no hair’, where hairs represent the potential complexity of the parameter space and resulting gravitational field [16].

The no-hair theorem states that Kerr geometry, a subset of the Kerr-Newman geometries describing BHs with only mass and angular momentum, is sufficient to describe any BH not subject to highly dynamic environments [16]. Kerr BHs, being axisymmetric solutions to the Einstein field equations, are commonly characterised by the total gravitational mass, M , and a dimensionless spin parameter taking values in a range of $[0, 1)$, χ [19, 20]. The

validity of the no-hair theorem is being actively tested with observations of GWs from BHs [16, 21]. The methods described in this study assume the no-hair theorem and consider only the Kerr family of BHs. Any perturbations of a Kerr BH must decay to a unique final Kerr BH state, determined by the initial conditions; that is, all hairs must be lost leaving behind only the mass and spin [17, 20]. This decay process, manifesting as gravitational radiation, is dominated by exponentially decaying radiation produced by the ringing down of the excited BH. This stage, colloquially known as the ‘ringdown’ stage, is dominated by QNMs determined by the final BH parameters, specifically the mass and spin of the stabilised BH. BH linear perturbation theory predicts that the ringdown can be described entirely as a linear superposition of QNMs. The quasinormal ‘ringing’ during this stage is primarily responsible for the loss of the BH’s hair [17]. As two BHs merge, they form a distorted remnant BH, beginning its existence in a perturbed state, and ultimately ringing down to a stable state.

While the amplitudes of QNMs are contingent on the excitation source,⁶ the complex-valued frequency component is exclusively decided by the intrinsic parameters (i.e., mass and spin) of the remnant BH. Consistent with their dissipative nature, these QNMs exhibit an imaginary component correlated with the decay of the radiation [17], resulting in a damped sinusoidal time-domain evolution. The QNMs are labeled by two angular index numbers, (l, m) , and an overtone index, n . For example, the fundamental (zeroth overtone) $l = 2, m = 2$ mode is referred to as the 220 QNM. When the overtone number is excluded, e.g., the 22 QNMs, the $l = 2, m = 2$ angular mode with all overtones is considered. GWs emitted during the ringdown phase manifest as a linear superposition of a set of QNMs, each oscillating at a characteristic frequency and decaying at a distinct rate. The existence of QNMs as a consequence of BH perturbation has been a subject of consideration for decades, with QNMs being calculated for BHs before the 2000s [22–25]. The QNM frequencies for Kerr BHs, due to the complexity, have only been calculated more recently [17, 19]. The linear superposition of the complex-valued BH ringdown signal can be written as [26]:

$$\begin{aligned} h(\iota, \beta, t) &= (h_+ - i h_\times)(\iota, \beta, t) \\ &= \frac{M}{r} \sum_{l=2}^{l_{\max}} \sum_{m=-l}^{m=l} \sum_{n=0}^{n_{\max}} \left[A_{lmn}^+ e^{-i\omega_{lmn}(t-t_0)} {}_{-2}S_{lmn}(\gamma_{lmn}, \iota, \beta) \right. \\ &\quad \left. + A_{lmn}^- e^{i\omega_{lmn}^*(t-t_0)} {}_{-2}S_{lmn}^*(\gamma_{lmn}, \pi - \iota, \beta) \right], \end{aligned} \quad (15)$$

where ι is the inclination angle, β is the azimuth angle, M is the remnant BH mass, r is the distance of the source from the Earth, $A_{lmn}^{+/-}$ are the amplitudes of the mode emitted in the positive or negative (north or south) direction, ω_{lmn} is the complex-valued QNM frequency (with the real and imaginary parts being the angular frequency and decay rate of the lmn mode, respectively), t_0 is the ringdown starting time, and ${}_{-2}S_{lmn}$ are the spin-weighted spheroidal harmonic functions. The dimensionless spheroidicity parameter is given by

$$\gamma_{lmn} = \chi M \omega_{lmn}, \quad (16)$$

⁶In this case, we would consider the initial BBH system to be the excitation source.

where χ is the dimensionless spin of the remnant BH. The quadrupole $l = m = 2$ QNMs are the dominant components of the ringdown signal. Excitation of other QNMs depends on the intrinsic characteristics of the BBH system, like the mass ratio or the spin directions of the two BHs before the merger. The angular distribution of each mode and extrinsic parameters like the orientation of the system determine the emission strength in each direction.

2.4 Black hole spectroscopy

Analysing the QNMs in the ringdown phase of the GW signal and identifying the characteristic frequencies and decay rates can yield information about the parameters of the BH. This method of determining BH parameters is called BH spectroscopy [7]. BH spectroscopy studies as a whole aim to analyse QNMs in the ringdown regime to glean information from GWs about the source, but the ideal method in order to achieve the aim is contentious [27]. There is much literature discussing methods of analysis and potential research outcomes⁷ with the goal of extracting the maximum amount of information possible [7, 27–30]. There are, however, many complexities that can drastically reduce the sensitivity of the analysis and the scope of possible conclusions. For example, investigating solely the ringdown portion of the signal means identifying a given analysis start time, t_i , where the ringdown signal (and the QNMs in question) is dominant, inherently introducing a potential uncertainty in the time domain. Given the basis-like nature of QNMs, it is also possible to overfit the ringdown with too many QNMs [7]. Then, the analysis method may be sensitive to potential systematics and thus may bias the observational results [30]. Identifying a robust method of BH spectroscopy that achieves sufficient sensitivity in accurately determining the mass and spin of the remnant BH is not a trivial task.

Detections of GWs have allowed the use of BH spectroscopy on observational data. Namely establishing whether QNMs could be detected in the ringdown and the confidence of the inference obtained from the observational data. While the identification of the dominant fundamental 22 mode in signals like GW150914 is fairly well accepted, the presence, or ability to identify the presence of, overtones is contested [31–33]. This reinforces the case for a robust and effective ringdown analysis method which can help identify these modes with greater certainty. The susceptibility of BH spectroscopy methods to GW detector calibration errors is an important factor in determining the effectiveness of the method and is not fully studied. It also remains to be studied how the improved sensitivity of future observatories will impact the effect introduced by detector calibration errors. In worst-case scenarios, the calibration error could distort the QNMs, leading to biased BH parameter estimates, or the calibration error could be inconsequential. Not knowing the extent to which the spectroscopic results could be biased is dangerous when trying to use the results for astrophysical analyses and inferences, like validating GR or the no-hair theorem.

⁷Potential outcomes include testing the no-hair theorem, GR, and potentially the quantum nature of compact objects [7].

3 Gravitational-wave observatories and detector calibration

Einstein presented his linearised theory of GWs in the late 1910s. Many theorists, such as Feynman, Kerr, and Schwarzschild, sought to resolve doubts surrounding the nature and physical effects of GWs. These studies provided insight into the potential of finding physically compelling evidence for the existence of GWs. The first claim of a direct detection was made by J. Weber in 1969, but the sensitivity of $10^{-16} \text{ Hz}^{-1/2}$, ultimately determined by the bar's thermal fluctuation noise limit, indicated a false detection. J. H. Taylor and J. M. Weisberg made the first significant indirect observation of GWs in 1982 using pulsar arrival times from the binary pulsar PSR 1913+16, also known as the Hulse-Taylor pulsar [34]. Laser interferometry as a means of detecting GWs was conceived in 1962 by Michael Gertsenshtein and Vladislav Pustovoit [35], and independently in 1966 by Rainer Weiss [36], but would take decades of research before it was viable. The first direct observations of GWs were made in 2015 by the advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO) [1, 37], a pair of modern ground-based GW observatories.

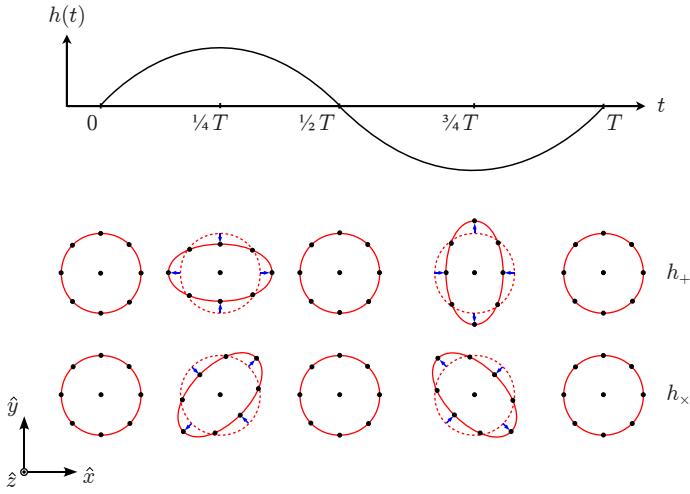
In this section, we discuss laser interferometry as a means of detecting GWs (Sec. 3.1) and introduce the current and next-generation ground-based observatories (Sec. 3.2). Finally, we describe the scheme of GW detector calibration and the definition of calibration errors (Sec. 3.3).

3.1 Laser interferometer gravitational-wave detectors

As outlined in Sec. 2.1, GR predicts that GWs, propagating at the speed of light, are transverse, possess a quadrupolar nature, and oscillate perpendicular to the direction of travel with two polarised states. The quadrupolar nature of GWs induces a cyclic pattern of the alternating stretching and squeezing of space along two mutually perpendicular axes to the direction of travel. The two polarisations are a quarter of a cycle out of phase and one is spatially rotated by 45 deg with respect to the other, as illustrated in Fig. 2a. Laser interferometric detectors measure the displacement between reference points by detecting modulations in light travel times caused by the passage of GWs through the detector. This study centres its focus on ground-based GW observatories, currently the only category of observatories with confirmed observations. It is noteworthy that pulsar timing arrays and space-based observatories also leverage the quadrupolar nature of the oscillations of GWs for their detection prospects.

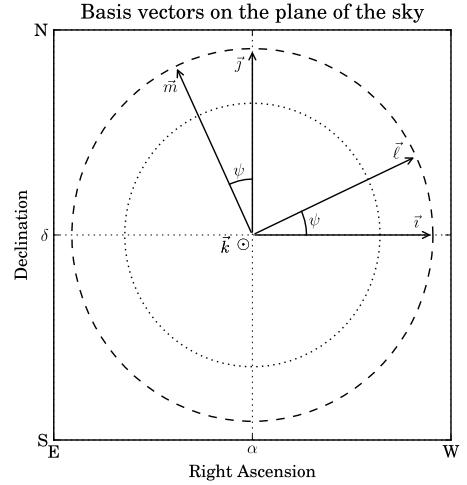
Ground-based detector designs are founded on the principles of the two-arm Michelson laser interferometer, which outputs a time series of strain that measures the response of the detector to the “+” and “ \times ” polarised components of GWs. The time series of strain caused by a GW signal in the detector is given by:

$$h_{\text{GW}} \equiv F_+ h_+ + F_\times h_\times, \quad (17)$$



(a) An example of the induced + and \times polarisations on a ring of free-falling particles in a local inertial frame as a result of a monochromatic GW propagating in the \hat{z} direction.

Credit: Tiec and Novak [11]



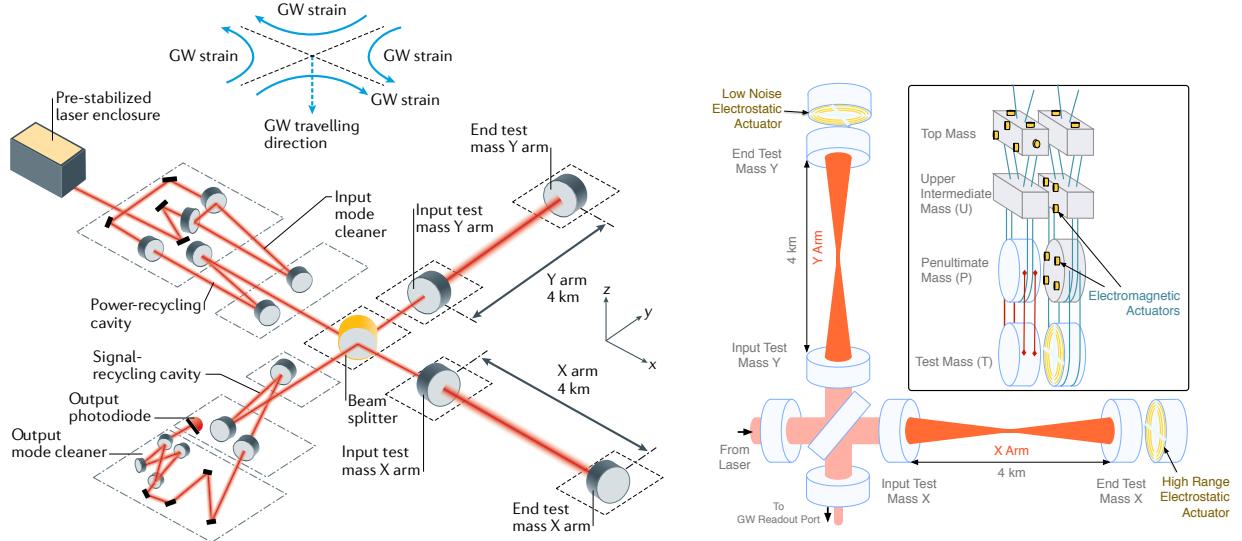
(b) GW frame with respect to the wave polarisation basis tensors. Here, \vec{k} is the wave propagation direction, \vec{i} and \vec{j} are the reference wave polarisation unit vectors, \vec{m} and \vec{l} are the unit vectors for the detector arms, and ψ is the polarisation angle. Credit: Whelan [38]

Figure 2: The + and \times polarisations of GWs. Given the perpendicular arms in a ground-based GW detector, the measured + and \times polarisations depend on the orientation of the wave with respect to the detector arms.

where h_+ and h_\times are the + and \times polarised components caused by the GW, respectively, and F_+ and F_\times are the corresponding detector antenna pattern functions. The antenna pattern of a detector is contingent on various factors related to both the detector configuration and the source properties. The physical configuration of the detector, with respect to the GW frame and/or basis chosen to factor the + and \times polarisation tensors, influences the response pattern in the detector [38]. As seen in Fig. 2b, the detector response is, in part, determined by the sky position of the source, given as right ascension (RA) and declination (DEC), and the polarisation angle, ψ , determined by the relative position of the polarisation vectors and detector arms [38].

The basic configuration of the Michelson interferometer comprises the laser, a beam splitter, two mirrors, and a photodetector that records the interference pattern. Laser interferometric GW detectors are essentially Michelson interferometers, with the lasers destructively interfering at the photodetector. Thus, in the absence of a GW traversing the detector, the photodetector should have a null response. Fig. 3 depicts a simplified schematic of aLIGO. As GWs traverse the detector, the two polarisation states induce distinct variations in each arm due to the local change in spacetime (illustrated at the top of Fig. 3a). The interaction between the GW and detector generates non-zero fluctuations in laser power, measured by the output photodiodes. While the core of a GW detector is

the Michelson interferometer, a myriad of advanced technologies are vital in producing a substantial sensitivity to the GW strain, by enhancing the GW signals and limiting potential noise sources. The arms are equipped with input test masses to function as Fabry-Pérot optical cavities, which significantly increase the total distance travelled by each laser beam and the total arm cavity power such that the effect from GW strain is both enlarged and enhanced (shown in both Figs. 3a and 3b). Other key optics form the power recycling and signal recycling cavities, which further enhance the signal, as shown in Fig. 3a.⁸



(a) Simplified schematic of the ground-based GW detector aLIGO. The power recycling cavity placed in the upper left recycles the laser and puts it back into the beam splitter, increasing the power stored up to 10^2 kW. The signal recycling mirror placed in the bottom left after the laser leaves the beam splitter alters the frequency response of the detector to fluctuations in the differential arm length and increases the power in a certain frequency band.

Credit: Bailes et al. [40]

(b) Simplified schematic of the aLIGO arm cavities. The two pairs of highly reflective test masses form Fabry-Pérot cavities which increase the laser power and photon storage time in each arm, resulting in increased signal amplitudes. The full four-stage dual-chain suspension system is shown in the inset. The electromagnetic and electrostatic actuators allow for the precise controlled displacement of the test masses in the differential arm length feedback control, described in Sec. 3.3. Credit: Abbott et al. [41]

Figure 3: Simplified optical configuration of the aLIGO interferometer, including the two 4-km-long arm cavities, dual recycling (power and signal) cavities (left), and a schematic of the quadruple pendulum suspension systems, which keep the test masses in free-fall, suppress seismic noise and provide feedback controls (right).

The sensitivity that a GW detector can achieve is essentially limited by the total contribution of a variety of noise sources, originating from the instrument itself, or the environment.⁹ Noise sources predominantly exhibit a frequency dependence, and combined, determine the amplitude spectral density (ASD) of the detector (in units of $\text{Hz}^{-1/2}$), referred to as the noise budget. The differential arm length noise budget of aLIGO Hanford

⁸See Ref. [39] for a comprehensive description of the optical systems and ‘locking’ (steps taken for the detector to arrive in an observation-ready state).

⁹See Ref. [42] for a comprehensive description of noise sources and mitigation techniques.

is shown in Fig. 4, in units of $\text{m Hz}^{-1/2}$, which equals the ASD multiplied by the total arm length, 4 km for aLIGO. The black curve represents the total of known noise contributions. The measured total noise budget in O3 is shown in blue. The difference between the blue and black curves indicates there exist unknown noise contributions. The smooth solid curves represent theoretically estimated noise contributions, and the dots and bumpy curves are the measured noise contributions from various noise sources (see legend). A lower ASD indicates that the detector is more sensitive to GW signals.

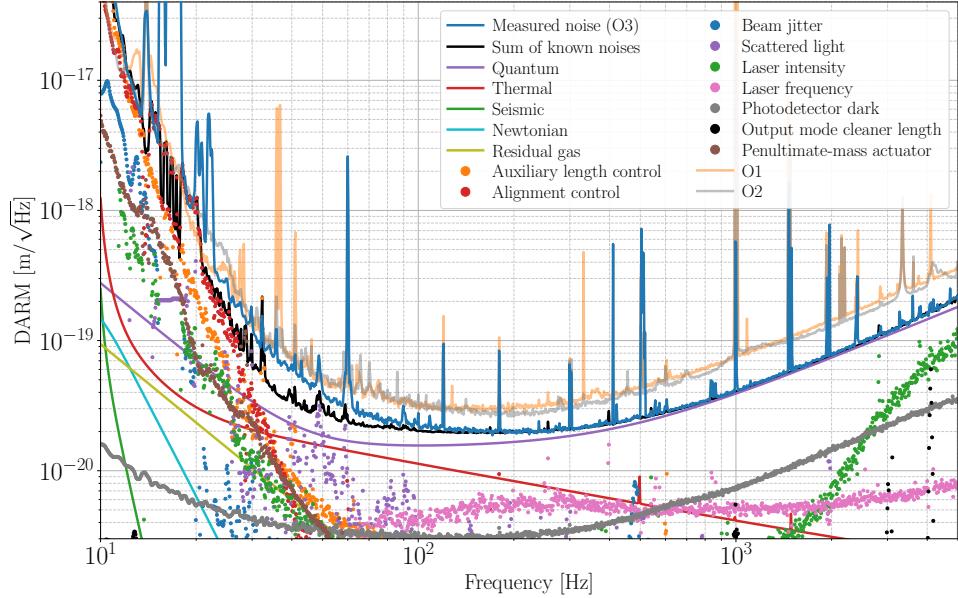


Figure 4: Differential arm length (DARM) noise budget at aLIGO Hanford as a function of frequency. Smooth solid curves represent theoretically calculated noise contributions. Dots and bumpy curves are the measured noise contributions. The total noise contribution and, thus, the sensitivity improves across the three observing runs (from O1 to O3) due to more efficient noise management strategies. Credit: Buikema et al. [42]

3.2 Current and next generation ground-based gravitational-wave observatories

3.2.1 Current generation observatories

Current ground-based observatories effectively probe audio-band GWs in a frequency range from roughly 10 Hz to 5 kHz, which predominantly features the merging stellar-mass compact binaries, alongside thus far unobserved burst signals from supernovae, continuous waves from spinning NSs, and the stochastic background from unresolved sources. The first generation of detectors (1G) was comprised of the initial LIGO and Virgo. LIGO, encompassing two ‘L’-shaped detectors situated in Hanford and Livingston, USA, with arms

extending 4 km in length, was completed in 1999. LIGO’s inaugural observation campaign ran from 2002 to 2010 with an initial BNS ‘range’ of 15 Mpc [43].¹⁰ Stationed near Pisa, Italy, the Virgo detector with 3-km-long arms was completed in 2003 and joined LIGO for the initial observing run in 2007 [45]. The establishment of the LIGO and Virgo Collaboration was a pivotal milestone in GW detection, with an agreement to synchronise observation runs, effectively running all three detectors as a network. After the initial runs, the LIGO and Virgo detectors underwent significant instrumental upgrading and transitioned into the second-generation (2G) observatories, Advanced LIGO (aLIGO) and Advanced Virgo (aVirgo). Major enhancements include the overhaul of the test-mass suspension, transitioning from a single pendulum to a quadruple suspension set-up. The aLIGO started the first observing run (O1) in 2015 and, in September, made the first detection of a GW signal generated by a BBH merger [46]. Virgo joined the network in 2017, partway through O2, significantly bolstering the sky localisation capabilities and contributing to the discovery of the first multi-messenger event, GW170817 [44]. The detector sensitivities improved between observing runs as the detectors were upgraded further; the sensitivity improvements for the Hanford aLIGO observatory are shown in Fig. 4 (see legend O1–O3). The Kamioka Gravitational Wave Detector (KAGRA), located in Gifu Prefecture, Japan, is the first observatory built underground and uses cryogenic test masses. The construction was completed in 2019 and observations first began in 2020, joining aLIGO and aVirgo for O3 as part of the LIGO-Virgo-KAGRA network [47]. The collection of 2G observatories will be complete in the next decade with the construction of LIGO-India located in Maharashtra, India, with a network of five observatories [40].

After O4 finishes in 2025, the detector network will undergo the next set of major upgrades leading towards the fifth observing run (O5), with the target sensitivity for LIGO and Virgo denoted as Advanced LIGO A+ (LIGO A+) and Advanced Virgo Plus. Both aim to increase the sensitivity by a factor of two compared to O4 by decreasing key noise contributions, such as the quantum, or coating thermal noise. LIGO-India plans to come online in the same configuration as LIGO A+ [40, 59, 60]. Post-O5 upgrades for LIGO, denoted as LIGO A#, are currently in the research and development phase, and aim to improve the strain sensitivity by another factor of two by reducing the noise further. LIGO A# will not only increase the detection rate by a factor of ~ 5 from LIGO A+, but will also provide a unique test bed for technologies required in next-generation observatories [50, 60]. The proposed design sensitivities for these upgrades are shown in Fig. 5. Improvements beyond the sensitivity of LIGO A# are unlikely due to the physical constraints imposed by the 4-km-long arms [50, 60].

3.2.2 Next generation observatories

The 90 observed GW events from compact binary mergers over O1–O3 have revolutionised our understanding of the extreme physics that governs these compact bodies, both solving some and creating new mysteries. Observations of celestial bodies in the mass gap between

¹⁰The closest observed GW event, the BNS event GW170817, occurred at a distance of 40^{+8}_{-14} Mpc [44].

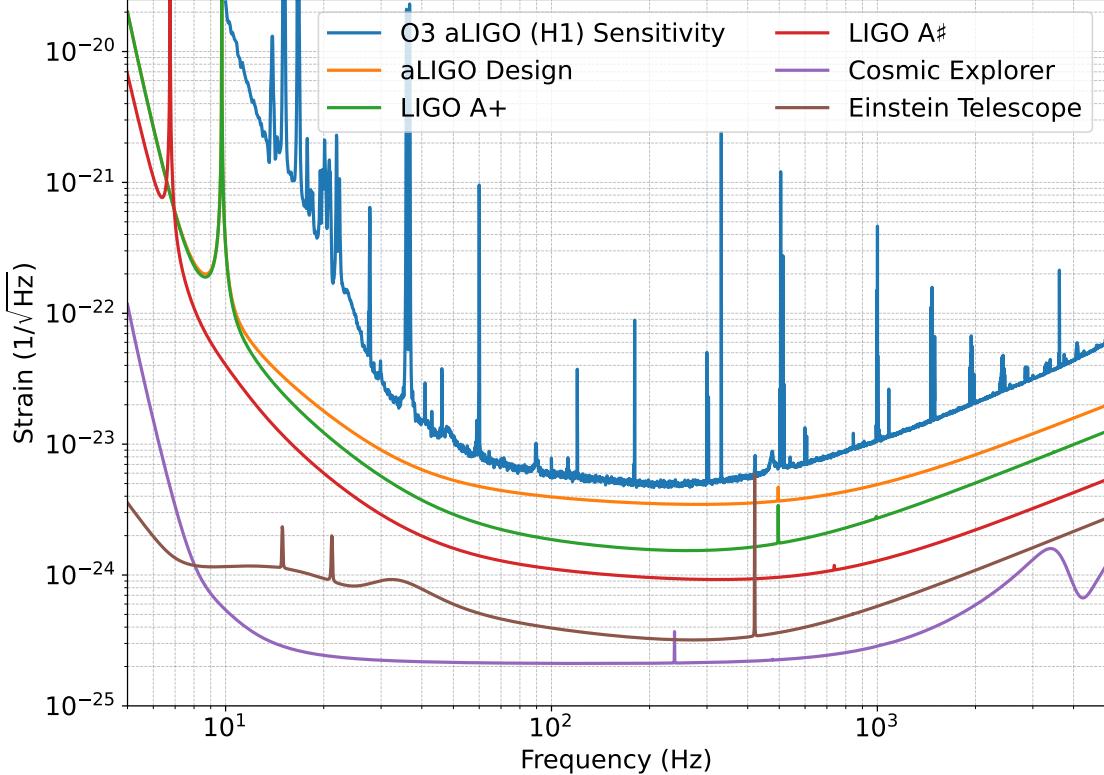


Figure 5: Amplitude spectral densities (ASDs) for aLIGO [37, 48, 49], LIGO A+ [50], LIGO A# [50, 51], Cosmic Explorer [52–54], Einstein Telescope [55–58], at their design sensitivities, and for aLIGO Hanford measured during O3 [42]. Lower ASD levels represent better sensitivities.

NSs and BHs have pushed the boundaries of our knowledge concerning the nature of these compact objects. The landmark multi-messenger event, GW170817, confirmed that BNS mergers are a source of short gamma-ray bursts and heavy elements [44, 61]. It also further corroborated that GWs travel at the speed of light, constrained NS radii, and provided valuable insight into the NS equation of state. These are but a fraction of the profound insights GW detections have given us into the elusive realm of compact objects.

While current observatories strive for optimised sensitivities, there is a limit within the current infrastructure. The predictions of the compact binary merger population and other GW sources reveal the necessity for a new generation of detectors to fully explore the gravitational-radiation-filled Universe [62]. Fig. 5 includes the design sensitivities for current 2G observatories (aLIGO design), the upgrading targets within existing infrastructures (aLIGO A+ and A#), and next-generation observatories (Cosmic Explorer and Einstein Telescope). A nearly tenfold increase in sensitivity and a much broader frequency band with the desired best sensitivity can be achieved in the next-generation observatories [63]. These next-generation observatories will be able to probe millions of sources closer to the edge of the Universe, providing a massive catalogue of precise detections that will inform never-before-observed parts of the Universe, the evolution of massive objects,

and cosmology. The unprecedented sensitivity will allow the resolution of GW signals to extreme precision, open the window to probe dark matter, dark energy, and new physics, observe the early Universe through a stochastic GW background, and investigate the densest matter in the cosmos in a way that could never be achieved in a laboratory [52, 53]. Fig. 6 shows the astrophysical horizon distances that the current and next-generation observatories can reach, with the Cosmic Explorer and Einstein Telescope covering the whole population of stellar-mass BBH mergers and almost all BNS mergers in the Universe.

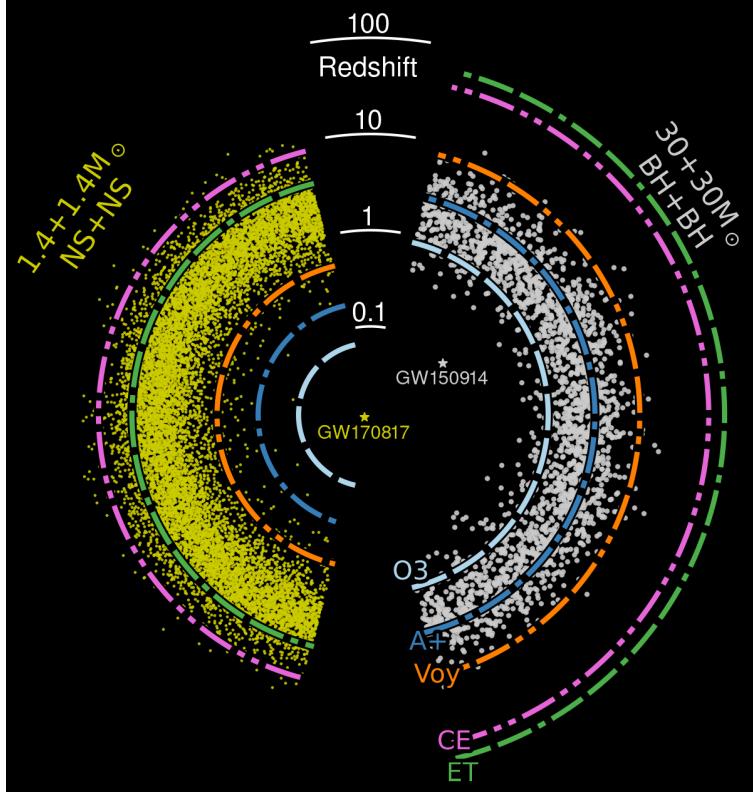


Figure 6: Astrophysical horizon of current and next-generation ground-based detectors for BBH (BH + BH) and BNS (NS + NS) mergers. The yellow and white dots represent a simulated population of BNS and BBH mergers, respectively, with a characteristic merger time of 10^8 years. The coloured curves indicate the maximum redshift at which a detection with an SNR of 8 could be made, assuming a $1.4 + 1.4 M_{\odot}$ BNS or $30 + 30 M_{\odot}$ BBH merger, with the detector in the label. The first BBH event, GW150914, and the first BNS event, GW170817, detected are marked by the white and yellow stars, respectively. The observatories or configurations depicted here are aLIGO in the third observing run (O3) [42], LIGO A+ (A+) [50], Voyager (Voy) [64], Cosmic Explorer (CE) [52], and Einstein Telescope (ET) [56]. Credit: Evans et al. [52]

The Cosmic Explorer (CE), led by the US, and the Einstein Telescope (ET), led by Europe, are the two major proposed next-generation observatories. Their lifetimes will be longer lasting, on the order of 50 years¹¹, facilitating substantial upgrades without the

¹¹The LIGO observatories had a planned 25-year operational lifetime, and are currently seeing the effects

construction of new facilities [63]. ET will likely be built, like KAGRA, underground at about a depth of 100–200 m. It will consist of three sets of interlaced detectors in a triangle formation, each with two arms of 10 km, spanning an angle of 60 deg [58]. Each detector will consist of two interferometers with one dedicated to achieving the best sensitivity at low frequencies (ET-LF) around 2–40 Hz, and the other to high frequencies (ET-HF) above ~ 40 Hz, forming a ‘xylophone configuration’ of six interferometers in total. CE will be in the classic ‘L’-shaped, above-ground configuration, notably with 40-km-long arms (configurations of 20-km-long arms are also under consideration) [52]. The long arm lengths of these future observatories naturally promise an order-of-magnitude improvement in strain sensitivity.

3.3 Detector calibration errors and uncertainty

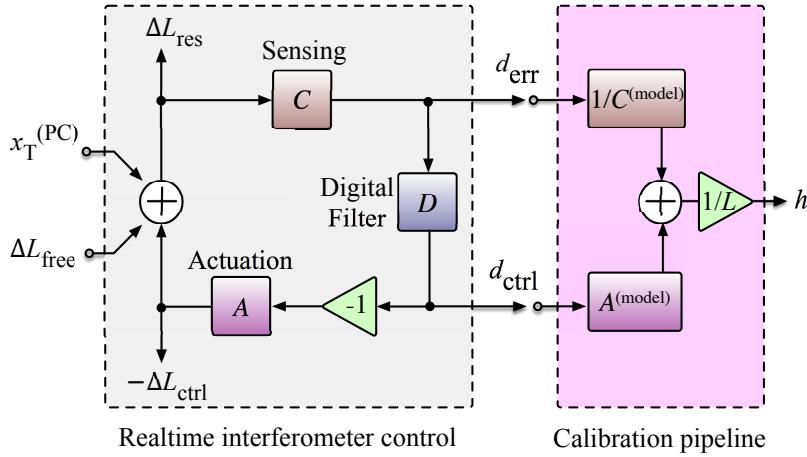


Figure 7: Advanced LIGO differential arm length feedback control loop (grey box) and the calibration pipeline that reconstructs the strain data (pink box). The sensing function, C , converts the residual differential arm displacement, ΔL_{res} , to a digital error signal, d_{err} , which is processed through a digital filter, D , to generate the digital control signal, d_{ctrl} . The actuation function, A , then converts d_{ctrl} to an actuation force acting on the test masses and produces a control displacement, $-\Delta L_{\text{ctrl}}$, to suppress the differential arm displacement, ΔL_{free} . The excitation $x_T^{(\text{PC})}$ marked in the diagram is added via the photon radiation pressure actuator system for loop characterisation. In the calibration pipeline (pink box), the digital signals, d_{err} and d_{ctrl} , are processed through the modelled sensing and actuation functions, $C^{(\text{model})}$, and $A^{(\text{model})}$, respectively, to reconstruct the strain data, $h = \Delta L_{\text{free}}/L$. Credit: Abbott et al. [41]

Ground-based GW observatories are controlled in a feedback loop to operate as a null instrument and output a digital voltage error signal from the photodetector, d_{err} . Dimensionless strain time-series data is reconstructed from the raw digital error signal using a of weathering [60].

complex-valued frequency-dependent transfer function, found via a model of the detector response, or a calibration model [65]. In the frequency domain, the strain data, h , defined by the differential changes in the arm length, ΔL_{free} , divided by the arm length, L , can be written as¹²

$$h \equiv \frac{\Delta L_{\text{free}}}{L} = \frac{R d_{\text{err}}}{L}, \quad (18)$$

where R is the detector's response function. The strain is constructed with both the voltage readout and a signal used to control the feedback loop [41]. Note here that the strain data h is not the GW strain but the measured strain in the detector, which contains both noise and potentially, GW signals. A diagram of the differential arm control loop is shown in Fig. 7. The feedback loop suppresses changes in the differential arm length, keeps the detector in a null range, and leaves only a residual differential arm displacement, ΔL_{res} . In order to reconstruct ΔL_{free} , and thus the strain, $\Delta L_{\text{free}}/L$, the effect of the feedback control loop must be compensated for.

The feedback control operates in real time, and consists of a sensing path, a set of digital filters, and an actuation path. First, the measured analogue residual displacement, ΔL_{res} , is converted to the digital output error signal, d_{err} , by the sensing function, C :

$$d_{\text{err}} = C \Delta L_{\text{res}}. \quad (19)$$

A set of digital filters, D , are then applied to d_{err} to produce the digital control signal:

$$d_{\text{ctrl}} = D d_{\text{err}}, \quad (20)$$

which is fed back through the actuation path, producing an analog control displacement, ΔL_{ctrl} , on the test masses to suppress ΔL_{free} (also see Fig. 3b):

$$\Delta L_{\text{ctrl}} = A d_{\text{ctrl}}, \quad (21)$$

where A denotes the actuation transfer function. With ΔL_{ctrl} subtracted from ΔL_{free} , all that remains is the residual ΔL_{res} . Combining the above functions, we find:

$$\begin{aligned} \Delta L_{\text{free}} &= \Delta L_{\text{res}} + \Delta L_{\text{ctrl}} \\ &= \frac{d_{\text{err}}}{C} + A d_{\text{ctrl}} \\ &= \left[\frac{1}{C} + A D \right] d_{\text{err}}. \end{aligned} \quad (22)$$

Thus, ΔL_{free} can be reconstructed from the raw output digital signal d_{err} via the response function:

$$R \equiv \frac{1}{C} + A D. \quad (23)$$

The modeled response function, $R^{(\text{model})}$, is constructed using models of the sensing and actuation functions, $C^{(\text{model})}$ and $A^{(\text{model})}$ [41]. The physical parameters in the sensing and actuation functions, e.g., the optical gain, differential coupled-cavity pole frequency, time

¹²Here, we describe the derivation fully in the frequency domain.

delays, etc., have a slow time dependence (on a time scale of minutes for the sensing and days to weeks for the actuation functions).¹³ Thus, the detector response function is both time-dependent and frequency-dependent.

Calibration systematic errors result from imperfections in the modelled response function, $R^{(\text{model})}$, due to features in the true response function, R , that are missing in the model, the imperfect estimates of model parameters, or uncompensated time dependence. We can write the frequency-dependent and time-dependent calibration errors as a ratio, $\eta(f, t)$, between the true response and modelled response functions:

$$\eta(f, t) = [1 + \delta\mathcal{A}(f, t)]e^{i\delta\phi(f, t)} = \frac{R(f, t)}{R^{(\text{model})}(f, t)}, \quad (24)$$

where $\delta\mathcal{A}(f, t)$ and $\delta\phi(f, t)$ denote the magnitude and phase errors, respectively. A perfect model corresponds to $\eta = 1$, i.e., $\delta\mathcal{A} = 0$ and $\delta\phi = 0$. However, due to inevitable imperfections in the model, frequency-dependent and time-dependent magnitude and phase errors with their associated statistical uncertainties exist in the calibration model, and thus the reconstructed strain data throughout observing runs [65]. This study focuses on the effects of calibration errors on individual CBC signals, which occur over a duration of less than a second. As such, we assume that the slow-varying time dependence has little effect, and approximate these calibration errors as constant in time, but frequency-dependent, viz.

$$R(f) = [1 + \delta\mathcal{A}(f)]e^{i\delta\phi(f)} R^{(\text{model})}(f). \quad (25)$$

The same frequency-dependent magnitude and phase errors, $\delta\mathcal{A}(f)$ and $\delta\phi(f)$, exist in the reconstructed strain data and impact the GW signals [see Eq. (18)].

The inevitability of calibration errors prompts the need to understand exactly to what extent they can bias the estimates of astrophysical source parameters and, in turn, affect the astrophysical inferences. It is generally accepted that calibration errors do not tend to bias GW analyses with current-generation detectors [66–69], and the uncertainty in astrophysical parameter estimation is instead dominated by the limited signal-to-noise ratio (SNR) that can be achieved with current detector sensitivities. The situation is likely to change when significantly increased SNRs can be achieved in the next-generation of observatories. The effect of calibration errors on future GW data analyses has yet to be determined, with Evans et al. [52] predicting that a limit of 1% in magnitude, i.e., $\delta\mathcal{A} < 1\%$ is needed. For reference, in the current observing runs with current-generation detectors, calibration errors with $\delta\mathcal{A} \lesssim 10\%$ and $\delta\phi \lesssim 10$ deg across the most sensitive band of 20–2 kHz are reported. This study aims to quantify the extent to which calibration errors can bias BH ringdown analyses using rational QNM filtering (explained in detail in Sec. 4.1) for both current and next-generation GW observatories. We derive the accuracy and precision of detector calibration required to conduct robust ringdown analyses and BH spectroscopy from NR simulations.

¹³The time dependence is due to physical evolution during the observing, e.g., variations of optical alignment in the arm cavities, charges accumulated around the test masses, etc. [65]

4 Impact of calibration errors on quasinormal mode analyses

In this section, we introduce the analysis tool, the rational QNM filter, in Sec. 4.1 and describe the calibration error profiles that we investigate in this study in Sec. 4.2. We then carry out a stage-by-stage investigation by adding calibration errors to pure signal waveforms, or waveforms injected into the detector frame (without adding noise), in Sec. 4.3. Finally, we quantify the impact of calibration errors in white Gaussian noise simulations in Sec. 4.4.

4.1 Rational quasinormal mode filter

In this study, we leverage a QNM analysis tool, namely the rational QNM filter [6–8], to study the impact of calibration errors on BH ringdown analyses. The filter, designed to reveal sub-dominant modes by cleaning (i.e., subtracting) dominant modes from the ringdown, provides a powerful tool for BH spectroscopy.

When we start analysing the signal at a given time after the GW strain reaches the peak amplitude in a CBC merger, the ringdown part of the wave can be described by a linear superposition of a set of QNMs (also see Sec. 2.3). We use t_i to denote the analysis starting time offset from the time when the GW strain reaches the peak amplitude. Note that the dominant QNMs in the wave after t_i are different when starting the analysis at different times in the ringdown stage. For example, at earlier times, right after the merger, higher-order overtones may contribute, but at later times overtones have mostly decayed, and only the fundamental mode dominates. To demonstrate how the rational QNM filter works, we start with a simplified toy model. Given the fundamental 220 mode is the most dominant one in ringdown signals, we consider an example with only the $(l, m, n) = (2, 2, 0)$ mode turned on at $t = 0$ and construct a strain signal:

$$h(t) = e^{-i\omega_{220}(t)} \Theta(t), \quad (26)$$

where ω_{220} is the 220 QNM complex-valued frequency, and Θ is the Heaviside step function. The real part of the waveform is shown in blue in Fig. 8. Here, in the ringdown regime after $t = 0$, only a single QNM, the 220 mode, exists. We can remove this mode from the ringdown regime using a rational QNM filter applied in the frequency domain with an analysis starting time $t_i = 0$, aligned with the signal turn-on time.¹⁴ The filter, \mathcal{F}_{lmn} , for filtering out the ω_{lmn} component can be written as [7]:

$$\mathcal{F}_{lmn} = \frac{\omega - \omega_{lmn}}{\omega - \omega_{lmn}^*}, \quad (27)$$

¹⁴Applying the filter in the time domain can lead to numerical noise [7].

where $*$ denotes the complex conjugate. To apply the filter in the frequency domain, we first transform $h(t)$ to the Fourier domain:

$$\begin{aligned} H(\omega) &= \frac{1}{\sqrt{2\pi}} \int h(t) e^{i\omega t} dt \\ &= \frac{i}{\sqrt{2\pi}} \frac{1}{\omega - \omega_{220}}. \end{aligned} \quad (28)$$

The filtered strain data in the Fourier domain is given by:

$$\begin{aligned} H_{\text{filtered}}(\omega) &= \mathcal{F}_{220} \frac{i}{\sqrt{2\pi}} \frac{1}{\omega - \omega_{220}} \\ &= \frac{i}{\sqrt{2\pi}} \frac{1}{\omega - \omega_{220}^*}, \end{aligned} \quad (29)$$

which can be transformed back to the time domain to yield the filtered time-series data:

$$\begin{aligned} h_{\text{filtered}}(t) &= \frac{1}{\sqrt{2\pi}} \int H_{\text{filtered}}(\omega) e^{-i\omega t} d\omega \\ &= -e^{-i\omega_{220}^* t} \Theta(-t). \end{aligned} \quad (30)$$

This filtered signal is shown by the orange dashed curve in Fig. 8. The filtered strain shows no existence of the 220 QNM in the ringdown portion of the signal, and instead has a flipped 220 QNM decaying backward in time prior to the ringdown at $t < 0$, which is discarded when investigating the ringdown. In other words, the ringdown part at $t > 0$ is fully cleaned by the filter.

In the toy model above, we only consider one direction of the emission. Real GW signals consist of components emitting towards the north and the south directions with respect to the system. As a consequence, the final form of the filter for one (l, m, n) mode is [8]:¹⁵

$$\mathcal{F}_{lmn} = \frac{\omega - \omega_{lmn}}{\omega - \omega_{lmn}^*} \frac{\omega + \omega_{lmn}^*}{\omega + \omega_{lmn}}. \quad (31)$$

Note that \mathcal{F}_{lmn} is determined by the QNM frequency ω_{lmn} which is, in turn, determined by the remnant BH's mass and spin due to the no-hair theorem. For a ringdown signal that consists of multiple angular modes or overtones, we can use a total stacked filter to annihilate the multiple modes:

$$\mathcal{F}_{\text{tot}} = \prod_{lmn} \mathcal{F}_{lmn}. \quad (32)$$

It is also important to note that the filter, \mathcal{F}_{lmn} , also impacts other QNMs with $(l', m', n') \neq (l, m, n)$. The amplitude of the (l', m', n') QNM is reduced by a factor of $B_{lmn}^{l'm'n'}$ depending

¹⁵In this study, we assume $m > 0$ and use the convention that m and $-m$ stands for emission toward the north and south with respect to the system, respectively. We do not use $-m$ to denote the retrograde modes, which rotate in the opposite direction and at different frequencies compared to the prograde modes depicted here. Retrograde modes have negligible contributions to the simulations conducted in this study. See, e.g., Ref. [26], for details.

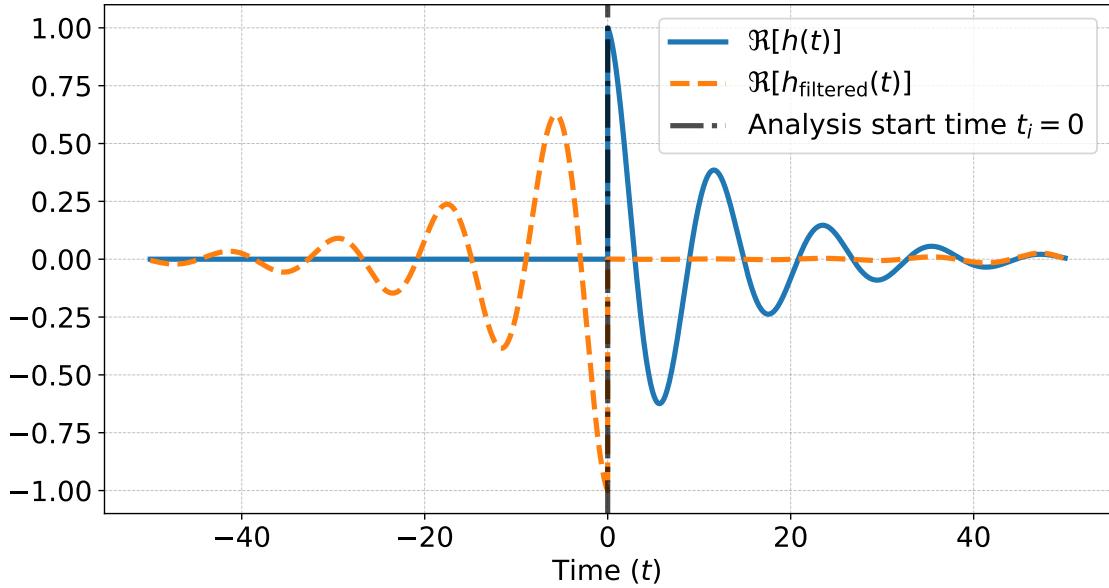


Figure 8: Effect of the rational QNM filter on a signal defined in Eq. (26) with only the 220 QNM starting at $t = 0$. The real parts of the original and filtered signals are shown in the figure. The filter removes the original QNM (blue) from the ringdown regime and introduces a flipped QNM at $t < 0$ (orange) that decays backward in time. The filter fully cleaned the ringdown portion of the data at $t > 0$, and the altered data at $t < 0$ are not used in ringdown analysis. The analysis starting time (when the filter starts to take effect), $t_i = 0$, is marked by the vertical dash-dotted line.

on the other mode's frequency, $\omega_{l'm'n'}$, and the phase between the two modes. (See Ref. [8] for the full derivation of $B_{lmn}^{l'm'n'}$.) For example, when studying an overtone mode, such as the 221 QNM, by filtering out the fundamental mode, such as the 220 QNM, from a signal, the overtone will present as weaker than it does in the original signal, due to the amplitude reduction introduced by the filter.

The effect of a stacked filter for two QNMs is shown in Fig. 9. The top and bottom panels show a signal in the time domain (both pre-ringdown and ringdown) and the frequency domain (ringdown only), respectively. The blue curves depict the 220 and 330 QNMs in a ringdown signal from a BBH merger with system parameters similar to the first detection event, GW150914 [1]. The true parameters of the BH's mass and spin are used to calculate the QNM frequencies and thus the filter(s), which are then applied to the signal. The orange curves are the residual signal after removing the 220 QNM using \mathcal{F}_{220} . The red curves are the final results after removing both the 220 and 330 QNMs using $\mathcal{F}_{220}\mathcal{F}_{330}$. In the frequency-domain plot, the amplitude reduction of one mode can be seen by comparing the 330 QNM peaks before and after applying the 220 QNM filter, \mathcal{F}_{220} (blue and orange peaks around the brown dash-dotted line).

The ringdown signal, if all the QNMs are filtered out, should essentially disappear. The final flat red curves shown in Fig. 9 demonstrate the ‘cleaning’ of the ringdown signal

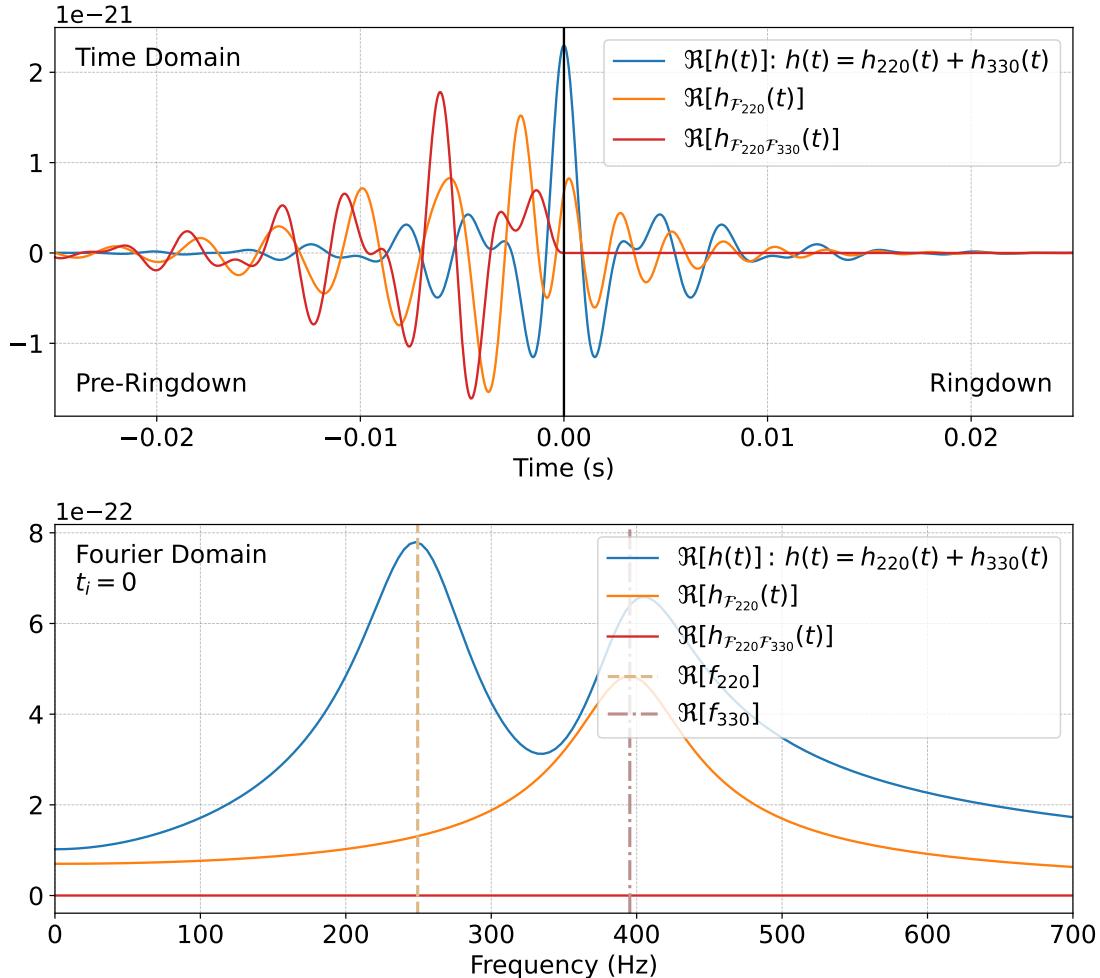


Figure 9: Ringdown waveforms with two QNMs (220 and 330), for a GW150914-like system before and after applying the filters in the time domain (top), and frequency domain (bottom; only for $t > 0$). The blue curves indicate the original signal. The orange and red curves correspond to the residuals after applying \mathcal{F}_{220} and $\mathcal{F}_{220}\mathcal{F}_{330}$, respectively. The red curves demonstrate that the ringdown portion of the signal is completely eliminated after applying a total filter of $\mathcal{F}_{220}\mathcal{F}_{330}$. The vertical lines denote the real part of the 220 and 330 mode frequencies.

as the correct filters are applied. In practice, due to the presence of other sub-dominant modes and noise, and the imperfect knowledge of the BH's true mass and spin, residual power remains in the data after applying the filter(s). Applying filters built from a set of parameters closest to the BH's true final mass and spin leads to the best subtraction, leaving behind minimal residual signal power. By evaluating the quality of mode cleaning in a Bayesian framework (applying a set of filters built from the true values of BH mass and spin leads to residuals consistent with pure noise), we are able to estimate the mass and spin of the final BH, denoted by M and χ , and do BH spectroscopy.¹⁶ This is markedly

¹⁶Throughout this study, we evaluate the BH's mass and spin in the detector's frame.

different from the conventional analysis of ringdown signals to estimate M and χ , where ringdown templates are created from a set of QNMs (with mode frequencies determined by different M and χ values) associated with mode amplitudes and phases. The templates are then used for likelihood calculations, and the estimates of M and χ , as well as the mode amplitudes and phases, are obtained in a Bayesian framework via Markov chain Monte Carlo [27]. By contrast, we can estimate M and χ of the BH by removing the relevant QNM components using the rational filters without the need to fit for any amplitude or phase values, thus reducing the risk of overfitting [8]. The downside is that the mode amplitude and phase information cannot be extracted using the filtering method.

Here, we describe the Bayesian framework that we adopt in this study. For a time-series of observational data, d_t , which includes a portion of the ringdown signal turned on at a time t_0 , we can apply a total filter $\mathcal{F}_{\text{tot}}(M, \chi)$ built from a set of parameters (M, χ) , that removes all the QNMs; the filtered time-series data, d_t^F , is expected to be consistent with pure noise in the ringdown regime. The likelihood function can be written as [8]:

$$\ln P(d_t | M, \chi, t_0, \mathcal{F}_{\text{tot}}) = -\frac{1}{2} \sum_{i,j > I_0} d_i^F C_{ij}^{-1} d_j^F, \quad (33)$$

where d_i^F are samples of the filtered data after t_0 , I_0 is the index for time t_0 , and C_{ij} is the noise covariance matrix. The likelihood function can be converted into a joint posterior in terms of M and χ [8]:

$$\ln P(M, \chi | d_t, t_0, \mathcal{F}_{\text{tot}}) = \ln P(d_t | M, \chi, t_0, \mathcal{F}_{\text{tot}}) + \ln \Pi(M, \chi) + \text{constant}, \quad (34)$$

where $\ln \Pi(M, \chi)$ is the prior. The two-dimensional joint posterior can be directly and efficiently computed through the likelihood function on a grid of (M, χ) without using a full Markov chain Monte Carlo analysis [8]. The rational QNM filter and the Bayesian framework have been applied to the first detection event, GW150914, demonstrating evidence of the existence of the first overtone 221 mode in the ringdown signal by cleaning the fundamental 220 mode [6, 8].

This study uses the Python package `qnm_filter` [6–8] to perform the QNM analysis for BH ringdown signals. The general procedures involves the following steps. First, calculate the likelihood values in Eq. (33) at each grid point on a (M, χ) plane. Second, calculate the joint posterior in Eq. (34) and find the 90% credible region by integrating the joint posterior in the two-dimensional plane. The maximum a posteriori (MAP) values of (M, χ) (found where the likelihood achieves a maximum) provides the best estimate of the BH parameters, denoted by $(M_{\text{MAP}}, \chi_{\text{MAP}})$. Third, calculate the joint posterior quantile, or credible interval, referred to as $p(M_t, \chi_t)$, corresponding to the credible region contour on which the true BH parameters, (M_t, χ_t) , lie. For example, $p(M_t, \chi_t) = 0.9$ indicates that (M_t, χ_t) lies on the 90% credible region contour. A lower $p(M_t, \chi_t)$ value corresponds to a more accurate recovery of (M_t, χ_t) , i.e., the true mass and spin values are consistent with tighter constraints of (M, χ) . In addition, we can evaluate the parameter distance, ϵ , between the MAP values and the true BH parameters, given by [8]:

$$\epsilon = \sqrt{\left(\frac{M_{\text{MAP}} - M_t}{M_t}\right)^2 + (\chi_{\text{MAP}} - \chi_t)^2}. \quad (35)$$

Smaller ϵ values indicate better agreement between the best estimates of the BH parameters, $(M_{\text{MAP}}, \chi_{\text{MAP}})$, and the true values, (M_t, χ_t) .

4.2 Calibration error profiles

Now, having set up the analysis tools in Sec. 4.1, we continue to describe the calibration errors we investigate in this study. Calibration errors can affect the amplitude and/or phase of the observed waveform [see Eq. (25) in Sec. 3.3] as a multiplicative effect on the ‘true’ strain in the Fourier, or frequency domain. Investigations into systematic calibration errors have been conducted across calibrated data in all existing observing runs with aLIGO [41, 65, 70, 71] and Virgo [72]. We use the latest aLIGO O3 studies [65, 71] to primarily inform the calibration error characteristics chosen for this study. In Fig. 10, we select an example time in O3 and plot five out of 10^4 realistic samples of calibration errors estimated at that time for the aLIGO Livingston detector [65, 71]. Calibration errors may peak at some frequencies (like the peaks in the examples in Fig. 10), span a wide band, or appear as a sharp spike in a narrow frequency band, depending on the physical causes of the error, e.g., an incorrect estimate of the optical gain, or an imperfect compensation of the electronics. In addition, the significance of the errors can vary in different situations. During O3, the upper limits on the magnitude and phase errors for the two aLIGO detectors are $\lesssim 10\%$ and $\lesssim 10$ deg, respectively, at a 1-sigma confidence interval in the most sensitive frequency band of 20–2000 Hz [65, 71]. Evans et al. [52] predicts tighter constraints, below 1%, on the calibration accuracy are necessary for next-generation observatories, such as Cosmic Explorer, in order to allow robust scientific analyses.

We aim to investigate the effects of calibration errors in terms of four varying factors: the frequency at which the frequency-domain error is at a maximum, f_{peak} , the characteristic width of the frequency-domain error, f_{width} , the peak of the magnitude error, δA_{peak} , and the peak of the phase error, $\delta \phi_{\text{peak}}$. Thus, we create an artificial calibration error model based on the commonly seen physical errors but leave the four factors tunable so that we can control the simulated error and evaluate the impact. Most realistic physical errors, as shown in Refs. [65, 71] and the examples in Fig. 10, have multiple features across the full frequency band with wildly varied shapes. Here, we only approximate one feature with the artificial error since our analysis is focused on a relatively narrow band spanning ~ 10 –100 Hz, where the relevant QNM frequencies lie; calibration errors far from the QNM peak frequencies do not have an impact on the analysis (see details in Sec. 4.3.3). The shape of any single commonly seen feature can be approximated by a Gaussian function, which has smooth edges and thus naturally prevents the introduction of artefacts to the time-domain data.¹⁷

¹⁷The modelled errors do not mimic some of the realistic features such as crossing unity magnitude and zero phase, e.g. the errors around 40 Hz for the examples shown in Fig. 10, but such features can be regarded as a combination of two modelled errors towards opposite directions. Given this study aims to quantify the impact of calibration errors in general situations, we do not consider specific combinations of features here.

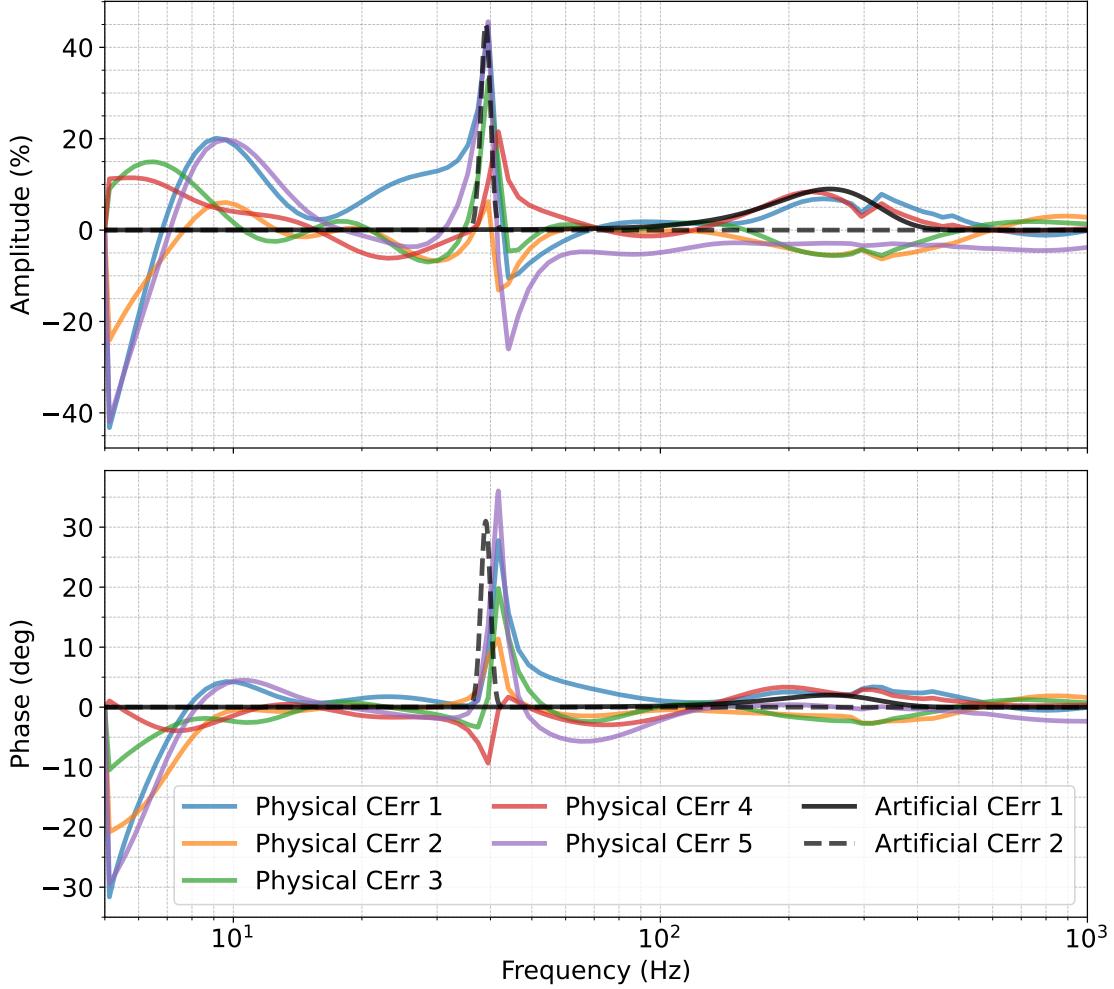


Figure 10: Examples of physical calibration errors (labelled as ‘Physical CErr’) estimated during O3 at aLIGO Livingston (colored) [65, 71] and examples of artificial calibration errors (labelled as ‘Artificial CErr’) generated in this study (solid and dashed black; see text for detailed parameters).

We define the width of the Gaussian-shaped artificial error as 4σ , i.e., the frequency range of $\pm 2\sigma$ from the centre. We construct the frequency-domain error

$$\eta(f) = [1 + \delta\mathcal{A}(f)]e^{i\delta\phi(f)}, \quad (36)$$

using

$$\begin{aligned} \delta\mathcal{A}(f) &= \delta\mathcal{A}_{\text{peak}} \exp \left[-\frac{1}{2} \left(\frac{f - f_{\text{peak}}}{f_{\text{width}}/4} \right)^2 \right], \\ \delta\phi(f) &= \delta\phi_{\text{peak}} \exp \left[-\frac{1}{2} \left(\frac{f - f_{\text{peak}}}{f_{\text{width}}/4} \right)^2 \right]. \end{aligned} \quad (37)$$

We plot two examples of the artificial errors in Fig. 10, one with $f_{\text{peak}} = 39$ Hz, $f_{\text{width}} = 4$ Hz, $\delta\mathcal{A}_{\text{peak}} = 45\%$, and $\delta\phi_{\text{peak}} = 31$ deg (dashed black curve); and the other with $f_{\text{peak}} =$

250 Hz, $f_{\text{width}} = 280$ Hz, $\delta A_{\text{peak}} = 9\%$, and $\delta \phi_{\text{peak}} = 8$ deg (solid black curve). It is physically motivated to set the magnitude and phase error peaks at the same frequency because any imperfect modelling in the calibration procedure is likely to produce both magnitude and phase errors. It is demonstrated in Fig. 10 that these artificial errors can mimic the realistic physical errors in shape in the frequency domain.

4.3 Impact of calibration errors on signal waveforms

In this section, we study the miscalibrated signal waveforms (Sec. 4.3.1), use the QNM filter to analyse the miscalibrated waveforms in the detector frame with different choices of parameters (Sec. 4.3.2), and quantify the impact of calibration errors on simulated detector data containing pure signals without additive noise (Sec. 4.3.3).

4.3.1 Miscalibrating the signal waveforms

Calibration errors present in the calibrated data affect both the noise and any existing signals. We start by investigating the effect on pure signals. The effect can be instantiated by intentionally ‘miscalibrating’ the signal, i.e., applying a calibration error to the signal waveform. The distortion of the signal waveform due to miscalibration is visible if the error is significant enough. In Fig. 11, panel (a) shows a Gaussian-shaped artificial calibration error (blue) applied to a pure ringdown signal with only the 220 QNM (brown). See the figure caption for the parameters of the artificial calibration error. We intentionally make the error large enough to demonstrate the change in the miscalibrated signal. Panel (b) shows the original (blue) and miscalibrated signal (orange) in the time domain (upper) and the frequency domain (lower). The effect of the phase error can be seen clearer in the time domain, with the oscillations of the miscalibrated signal mismatching the original signal. The effect of the magnitude error is seen clearly in both time and frequency domains as an increase in signal amplitude. In the frequency domain, the peak error appears around the 220 QNM frequency as expected, where the calibration error with $f_{\text{peak}} = 249.43$ Hz is applied.

To study the impact of calibration errors, we need to analyse the miscalibrated signals in the detector frame and compare the results to those from the correctly calibrated data. We take the BBH merger numerical relativity (NR) waveforms from the Simulating eXtreme Spacetimes (SXS) catalogue [9, 10] (referred to as NR waveforms onwards) and ‘inject’ the NR waveforms into the detector. These NR waveforms, unlike the ringdown QNM decomposition defined in Eq. (15) in spin-weighted spheroidal harmonics, decompose the numerically calculated strain using the spin-weighted spherical harmonic basis, given by:

$$\begin{aligned} h(\iota, \beta, t) &= (h_+ - i h_\times)(\iota, \beta, t) \\ &= \sum_{lm} {}_{-2}Y_{lm}(\iota, \beta) h_{lm}(t), \end{aligned} \tag{38}$$

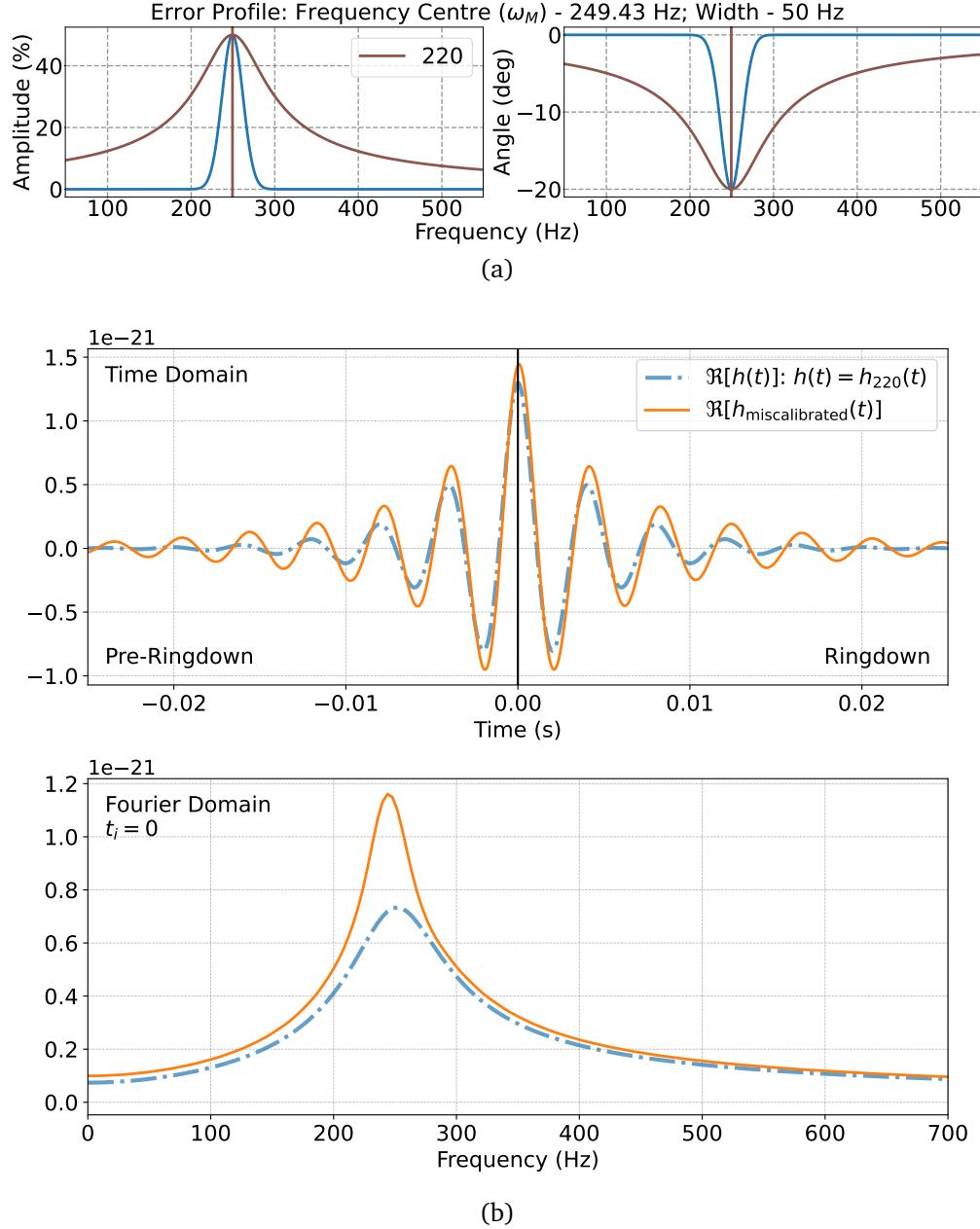


Figure 11: Example of a miscalibrated QNM signal waveform. (a) A Gaussian-shaped artificial calibration error (blue) with $\delta\mathcal{A}_{\text{peak}} = 50\%$, $\delta\phi_{\text{peak}} = -20 \text{ deg}$, $f_{\text{width}} = 50 \text{ Hz}$ and $f_{\text{peak}} = 249.4 \text{ Hz}$, around the 220 QNM frequency (brown) of a GW150914-like ringdown signal (the vertical line marks the QNM peak frequency). (b) The miscalibrated waveform in the time domain (upper; pre-ringdown and ringdown) and frequency domain (lower; ringdown only). The blue and orange curves indicate the original and miscalibrated waveforms, respectively. The magnitude error is more apparent in the frequency domain around the 220 QNM frequency, and the phase error displays as a time shift of the wave in the time domain.

where ι is the inclination angle, β is the azimuth angle, and ${}_{-2}Y_{lm}$ are the spin-weighted spherical harmonic functions. The spin-weighted *spherical* harmonics are orthogonal while the spin-weighted *spheroidal* harmonics are not. Thus, an lm mode may contribute to Y_{lm} and $Y_{l'm}$ components, where $l' \neq l$. The mode mixing due to different basis functions is a subtle and subdominant effect when taking a particular Y_{lm} component from the NR waveforms, but we do not expect to see an impact from mode mixing on our results in this study. (See detailed discussion in, e.g., Ref. [26].)

We perform three types of injection in this study: (1) a zero-noise injection, which simply produces a time-series with the pure signal waveform in the detector frame without additive noise (Secs. 4.3.2 and 4.3.3), (2) a white Gaussian noise injection, which produces a time-series with the signal in the detector frame combined with simulated additive white Gaussian noise (Sec. 4.4), and (3) a coloured-noise injection, which produces a time-series with the signal added to the coloured Gaussian noise in the detector frame (the coloured noise level is determined by the ASD of the detector’s design sensitivity) (Sec. 5).

4.3.2 Applying the rational filter to miscalibrated waveforms

As described in Sec. 4.1, the residual power remaining after certain QNM components have been filtered out is compared to pure noise to evaluate if the mode subtraction is effective. Fig. 12 shows the residual of the original (blue) and miscalibrated signals (orange), as seen in Fig. 11, after the correct 220 QNM has been cleaned, i.e., the \mathcal{F}_{220} filter has been applied. The upper and lower panels are in the time domain and frequency domain, respectively. The original, accurately calibrated signal is cleaned perfectly with no residual power left, but this is not the case for the miscalibrated signal where a ‘mark’ from the calibration error remains. Albeit the calibration error in this example is unrealistically large for demonstration purposes, in realistic scenarios, smaller residuals due to the existence of calibration errors may affect the final step of the BH spectroscopy process to determine the (M, χ) values that return minimal residual power in the filtered signal, and thus bias the (M, χ) estimates.

In order to perform a ringdown analysis, certain hyper-parameters must be defined; this includes the initial time of the analysis, i.e., the analysis starting time t_i , and the set of QNMs to be filtered. Here, the analysis starting time t_i is chosen such that all the QNMs considered have sufficient amplitudes and contribute to the parameter estimation. In this study, we focus on analysing the fundamental 220 and the first overtone 221 QNMs in the ringdown wave from a GW150914-like remnant BH. It is important to note that overtones decay quicker than the fundamental 220 mode, and the higher the overtone (i.e. the higher the n number), the quicker it decays. Setting the analysis starting time t_i too early may lead to biased results as higher-order overtones ($n \geq 2$) are present in the data in addition to the 220 and 221 QNMs; at a much later t_i the data may only contain the 220 QNM as the 221 QNM has already decayed.

In addition to the BH intrinsic properties, the final mass and spin, the injection process

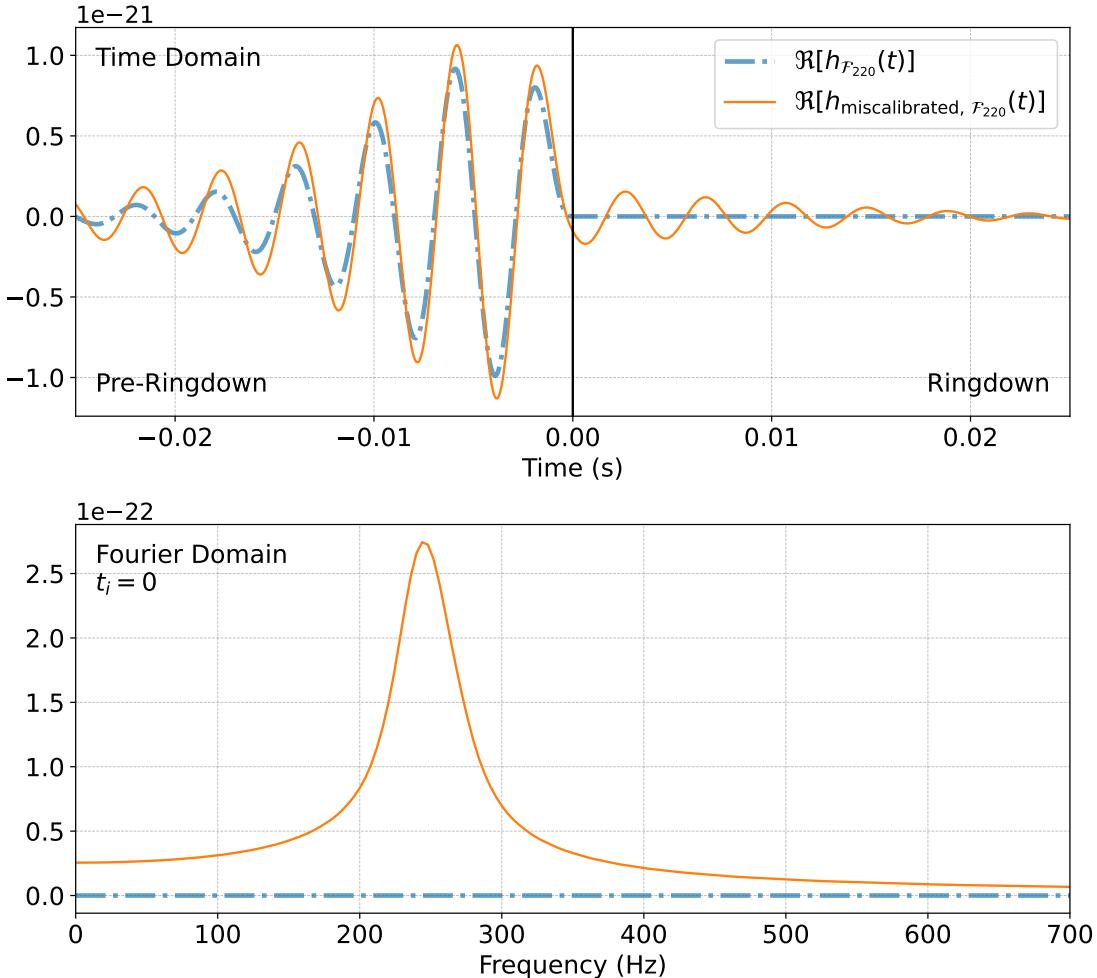


Figure 12: Example of applying the QNM filter to a miscalibrated QNM signal in time domain (upper) and frequency domain (lower). The calibration error and pure signal waveform are depicted in Fig. 11. The blue and orange curves indicate, respectively, the results from applying the filter, \mathcal{F}_{220} , to the original and miscalibrated waveforms. The calibration error results in a non-zero residual, seen in both time and frequency domains (exaggerated in this example), which may potentially lead to biases in the BH parameter estimation.

to produce a signal in the detector is also determined by other extrinsic parameters, e.g., the orientation, sky position, and luminosity distance of the source, which impact the amplitude and polarisation of the signal seen in the detector. We fix some parameters to the values of the real GW150914 event: a total binary mass of $70.6M_\odot$ and a luminosity distance of $D_L = 410$ Mpc [46]. We vary other parameters, including the source sky location, right ascension (RA) and declination (DEC), polarisation angle, ψ , azimuth angle, β , and inclination angle, ι , listed in Table 1. The first three are parameters associated with the relative position of the Earth to the GW source, and the last two are associated with the orientation of the source system itself.

| Parameter | Symbol | Range |
|--------------------|---------|-------------------------------------|
| Right ascension | RA | $[0, 2\pi]$ |
| Declination | DEC | $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$ |
| Polarisation angle | ψ | $[0, \pi]$ |
| Azimuth angle | β | $[0, \pi]$ |
| Inclination angle | ι | $[0, \pi]$ |

Table 1: Varied parameters for signal injections (samples selected from the range listed in the right column). These parameters affect the observed strain in the detector.

We can compute the matched-filter signal-to-noise ratio (SNR) via

$$\text{SNR} = \frac{\langle h_t | x \rangle}{\sqrt{\langle h_t | h_t \rangle}}, \quad (39)$$

where h_t is the ringdown waveform template (consists of one or multiple QNMs), and x is the observed data in the detector containing the injected signal, and

$$\langle h_t | x \rangle = \sum_{i,j > I_0} h_i C_{ij}^{-1} x_j. \quad (40)$$

Here, C_{ij} is the covariance matrix, as discussed in Sec. 4.1, the subscript indicates the discrete time index, and I_0 is the index for time t_0 when the ringdown signal is turned on. Here we determine t_0 as the time when the strain amplitude reaches its maximum,¹⁸ and the ringdown SNR is computed for the full waveform after t_0 regardless of when we start the analysis. The parameters in Table 1 can impact the SNR of each QNM in the detector data due to the non-isotropic angular emission pattern of QNMs in 3D space, and could potentially lead to different levels of impact on the BH parameter estimation, even when the same level of calibration error occurs.

We now use a zero-noise injection of a GW150914-like NR waveform ($Y_{2\pm 2}$ components only) to investigate the impact of the analysis starting time, t_i , and the source parameters listed in Table 1. We consider a calibration error of $\delta\mathcal{A}_{\text{peak}} = 10\%$, $\delta\phi_{\text{peak}} = -10$ deg, $f_{\text{width}} = 50$ Hz, and $f_{\text{peak}} = 249.4$ Hz (at the 220 QNM frequency), shown as blue curves in the top panel of Fig. 13. The QNM filter analysis method is applied and the results between analysing the original injection data and intentionally miscalibrated data are compared. The 90% credible regions and maximum a posteriori (MAP) values for the (M, χ) joint posterior are shown in red and blue in the bottom three panels, for the correctly calibrated data and miscalibrated data, respectively.¹⁹ The green stars indicate the true value of (M_t, χ_t) . Each vertical column corresponds to a different analysis starting time, offset from $t_0 = 0$ in units of the true BH mass, as labeled in the top left corner of each subplot. Each row corresponds to the parameter estimation using a different set of QNMs. In the

¹⁸Note that the exact ringdown start time, when the waveform can be treated as the superposition of the QNMs, is debatable [73].

¹⁹This study places uniform priors in ranges $M \in [34M_\odot, 100M_\odot]$ and $\chi \in [0, 0.95]$.

first row, the 220 and 221 modes are fitted together. In the second (third) row, the 221 (220) mode is removed before fitting for the 220 (221) mode only. At earlier times, the joint posteriors do not match the true value because other higher-order overtones, i.e., $22n$ ($n \geq 2$) modes, are still present, while we only apply the QNM filter with 220 and 221 modes.

Fig. 14 shows the joint posterior quantile values that the true BH parameters correspond to, $p(M_t, \chi_t)$ (see Sec. 4.1), for both the correctly calibrated and miscalibrated signals, when changing the hyper-parameter t_i , and the injection parameters RA, DEC, ψ , β and ι . In panel (a), we set RA = DEC = ψ = β = ι = 0, but analyse the injections at different starting times, t_i . To effectively evaluate the impact of calibration errors by studying 220 and 221 QNMs, we must have a controlled set of analysis in the correctly calibrated data as a reference. We need to find a starting t_i at which the 220 and 221 QNMs dominate. Given the injection is generated without additive noise, we expect to recover the true value of the BH mass and spin using the $\mathcal{F}_{220}\mathcal{F}_{221}$, i.e., $p(M_t, \chi_t) = 0$, when these two modes dominate. This is achieved when $t_i \geq 14M_t$ [see red dots in Panel (a)]. We can also look into the 90% credible region contours to cross-check. Examples are shown in Fig. 13, with the t_i of interest being $t_i = 14, 16, 18M_t$. At earlier times, other higher-order overtones are contributing, and thus a two-QNM filter is not good enough, while at later times, the 221 component has significantly decayed. We select $t_i = 16M_t$ as the starting time for analysing 220 and 221 modes in the GW150914-like NR waveform without additive noise.

In panel (b), we always fix $t_i = 16M_t$ and vary the parameter on the horizontal axis of each subplot while holding the rest at 0. In the top right corner of each subplot, the difference between the maximum and minimum SNRs in the correctly calibrated data across the eight sample injections, $\Delta\text{SNR} = \max(\text{SNR}) - \min(\text{SNR})$, is displayed. The $p(M_t, \chi_t)$ values for the correctly calibrated injections all align with 0 (a perfect recovery) and do not depend on the varying parameters, which is as expected.²⁰ The results from the miscalibrated data are always worse than the correctly calibrated results. The $p(M_t, \chi_t)$ values for the miscalibrated injections show a relation with the SNR, being larger (a worse recovery) when the parameters lead to a higher SNR in the detector. This is due to the multiplicative nature of calibration errors, as seen in Eq.(25), which results in a larger departure from the correctly calibrated injection when the SNR is higher. This SNR dependence is also seen in panel (a). As the analysis starting time, t_i , increases, the SNR left for $t > t_i$ is lower, and thus the difference between the miscalibrated results and the accurate results gets smaller. In other words, in lower SNR regimes, the bias caused by the same level of calibration errors is smaller compared to in the higher SNR regime. Overall, changing the parameters RA, DEC, ψ , β and ι , in some sense, degenerates with changing the SNR, which can be better controlled by changing the luminosity distance of the source or the noise level. Thus, for the rest of the thesis, without loss of generality, we set injection parameters to

²⁰Note that the SNRs quoted here do not represent the ratio between the signal and the noise present in the data (as there is no additive noise), but instead are calculated from the signal and a number we input as the autocovariance function for the likelihood calculation in the case when there is no noise. The ‘SNR’ still indicates the significance of the signal injection.

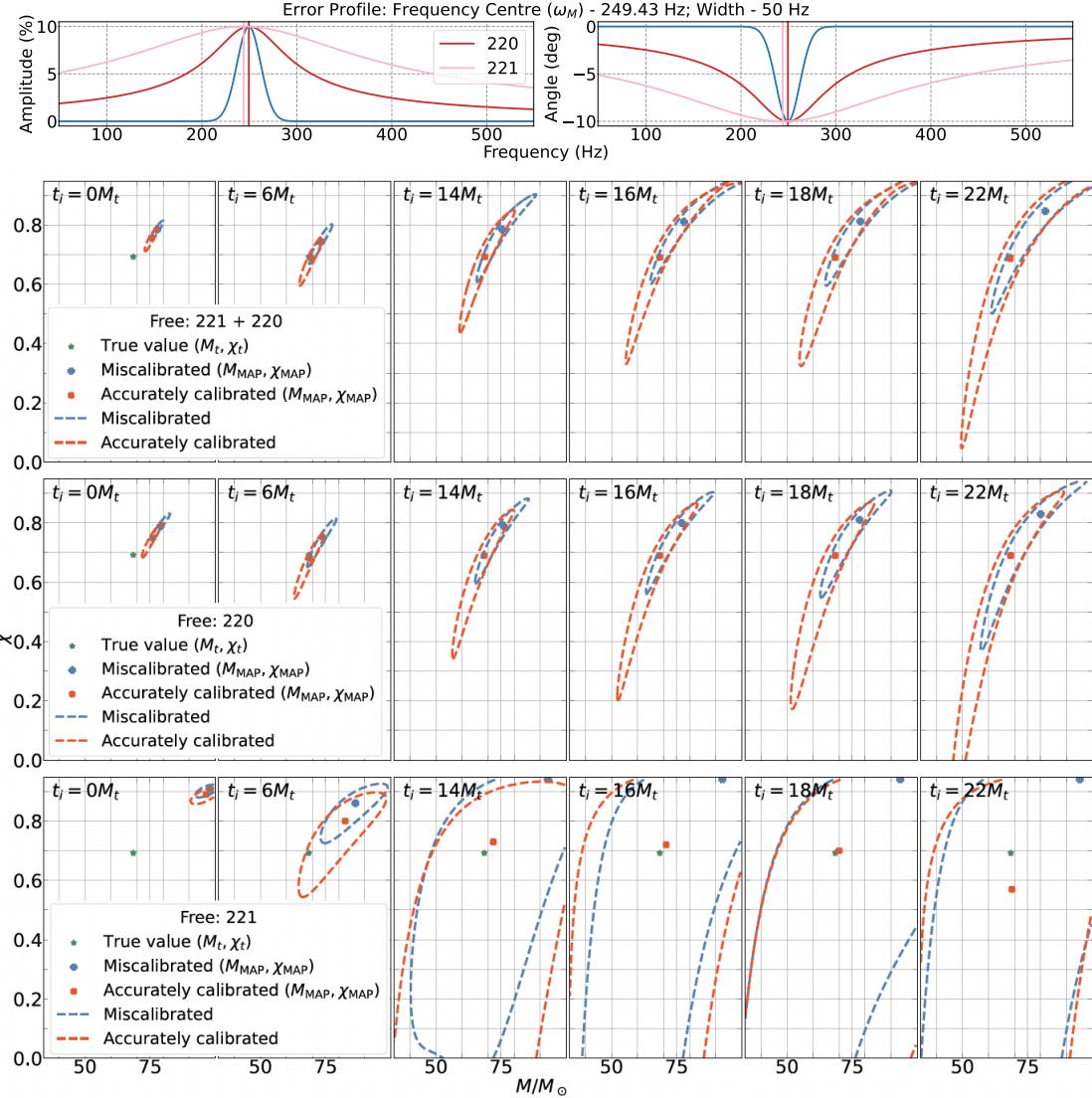
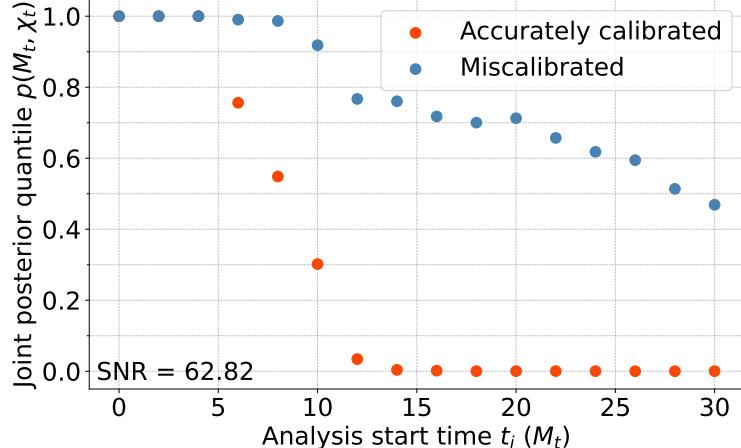
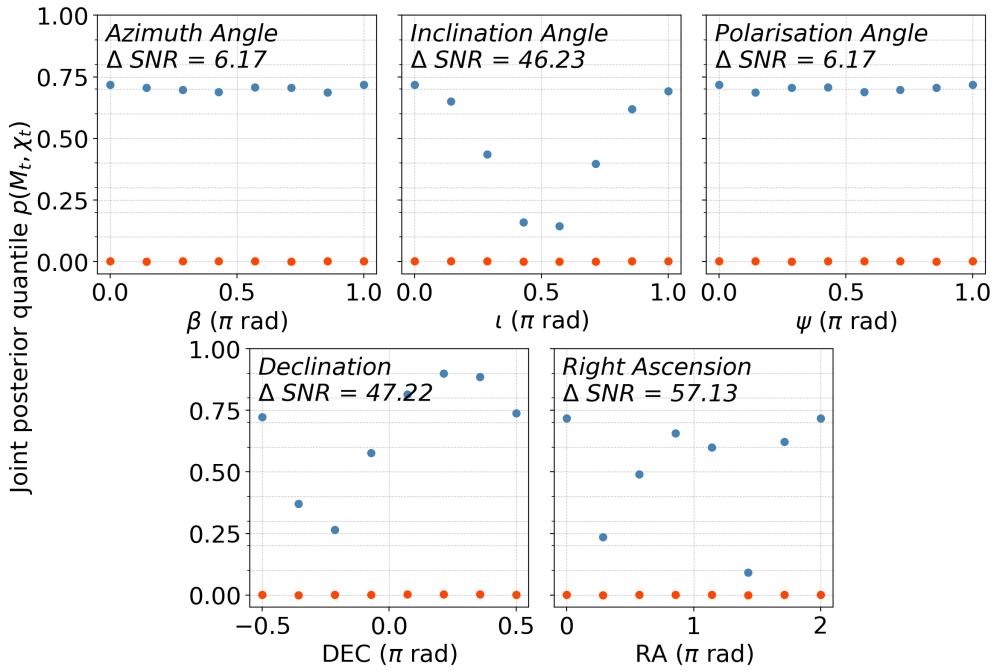


Figure 13: The 90% credible region (dashed contours) of the (M, χ) joint posterior for a GW150914-like NR waveform injection ($Y_{2\pm 2}$ components only, no additive noise (SNR = 63), RA = DEC = $\psi = \beta = \iota = 0$). The red and blue contours stand for the results from the accurately calibrated and miscalibrated data, respectively. The red cross and blue plus markers are the maximum a posteriori (MAP) values from the data without and with calibration errors, respectively. The green star denotes the true BH parameter values. The calibration error is shown in blue in the top panel in the frequency domain: $\delta\mathcal{A}_{\text{peak}} = 10\%$, $\delta\phi_{\text{peak}} = -10 \text{ deg}$, $f_{\text{width}} = 50 \text{ Hz}$, and $f_{\text{peak}} = 249.4 \text{ Hz}$, along with the 220 (dark red) and 221 (light red) QNMs. In the bottom three panels, each vertical column corresponds to a different analysis starting time, labelled in the top left corner of each subplot. Each row corresponds to the parameter estimation using a different set of QNMs. In the first row, the 220 and 221 modes are fitted together. In the second (third) row, the 221 (220) mode is removed before fitting for the 220 (221) mode only. The mode(s) used in the fitting are denoted by the “free” mode(s) in the legend.



(a)



(b)

Figure 14: Joint posterior quantile values, $p(M_t, \chi_t)$, for a GW150914-like NR waveform injections ($Y_{2\pm 2}$ components only) when varying (a) t_i and (b) RA, DEC, ψ , β , ι . The red and blue dots are from the correctly calibrated data and miscalibrated data, respectively. The calibration errors applied are the same as described in Fig. 13. In panel (a), different starting times are tested by fixing injection parameters to $\text{RA} = \text{DEC} = \psi = \beta = \iota = 0$. In panel (b), t_i is fixed to $16M_t$, and all subplots vary the single one parameter on the horizontal axis and hold the rest at 0. The difference between the maximum and minimum SNRs in the eight correctly calibrated injections, $\Delta \text{SNR} = \max(\text{SNR}) - \min(\text{SNR})$, is shown in the top right corner of each subplot. The results from the correctly calibrated data sets all align with 0 (perfect recovery). The results from the miscalibrated data are always worse than the correctly calibrated results. The variation in the miscalibrated results among different injections is mainly due to the SNR determined by the injection parameters.

$\text{RA} = \text{DEC} = \psi = \beta = \iota = 0$ and tune the SNR only by changing the luminosity distance or the noise level.

4.3.3 Quantifying the impact of calibration errors without additive noise

After investigating the calibration error using a single error profile (Sec. 4.3.2), in this section, we seek to understand the impact of different calibration errors on the signal itself without any potential additional bias introduced by additive noise. We begin with studies of a manually constructed ringdown signal with only two QNMs, 220 and 221 (the mode frequencies are determined by a GW150914-like remnant BH, and the mode amplitudes and phases are arbitrarily chosen), before analysing a more realistic GW150914-like NR waveform ($Y_{2\pm 2}$ components only), so we can verify the analysis procedure with a set of clean, controlled injections that are free of impacts from additional mode components, higher-order overtones, numerical noise, etc. contained in the NR waveform. By comparing the results from the controlled and NR injections, we verify the analysis method and investigate the features unique to NR waveforms that may affect the results.

In Fig. 15, we show the joint posterior quantile values that the true BH parameters correspond to, $p(M_t, \chi_t)$, for a manually-constructed, controlled 220 and 221 signal (with mode frequencies determined by a GW150914-like system) injected into aLIGO Hanford detector without additive noise. Note that the lower the $p(M_t, \chi_t)$ value, the better the recovery of the BH parameters. We vary the four properties of the calibration error, the offset between the peak frequency of the error and the 220 QNM frequency $\Delta f = f_{\text{peak}} - f_{220}$, error width f_{width} , peak magnitude error δA_{peak} , and peak phase error $\delta \phi_{\text{peak}}$, one at a time, with other parameters fixed. In panel (a), we vary Δf and fix $f_{\text{width}} = 50$ Hz, $\delta A_{\text{peak}} = 10\%$, and $\delta \phi_{\text{peak}} = -10$ deg. In panel (b), we vary f_{width} and fix $\Delta f = 0$, $\delta A_{\text{peak}} = 10\%$, and $\delta \phi_{\text{peak}} = -10$ deg. In panels (c) and (d), we vary δA_{peak} and $\delta \phi_{\text{peak}}$, respectively, and hold the other fixed at given values (see legend), with $\Delta f = 0$ and $f_{\text{width}} = 50$ Hz. The red horizontal lines indicate $p(M_t, \chi_t) = 0$ from correctly calibrated data, which demonstrates that without calibration errors and noise, the BH parameters can be perfectly recovered. Fig. 16 shows the same set of tests, but using the NR waveform of a GW150914-like system ($Y_{2\pm 2}$ components) for injections.

The general trends between Figs. 15 and 16 are similar. Overall, we see that when the calibration error occurs around the QNM peak frequency, and especially with a $f_{\text{width}} \sim 10\text{--}100$ Hz (comparable to the characteristic width of the QNM in the frequency domain; see the top row of Fig. 13), the impact on the BH parameter estimation is maximised. Recall that higher $p(M_t, \chi_t)$ values indicate worse recoveries. This is as expected because the calibration error simply mixes with the QNM (e.g., see Fig. 11) and leads to biased results. Also as expected, larger δA_{peak} and/or larger $\delta \phi_{\text{peak}}$ lead to a more significant bias in the analysis. These trends are clearly seen in both Figs. 15 and 16.²¹ Interestingly, a phase

²¹As a cross-check, we also look at the signals in both the time and frequency domains and compare the results with other studies to validate the method.

error seems to introduce a relatively larger bias than a magnitude error, as seen in the steeper gradient in panel (d) of both figures.

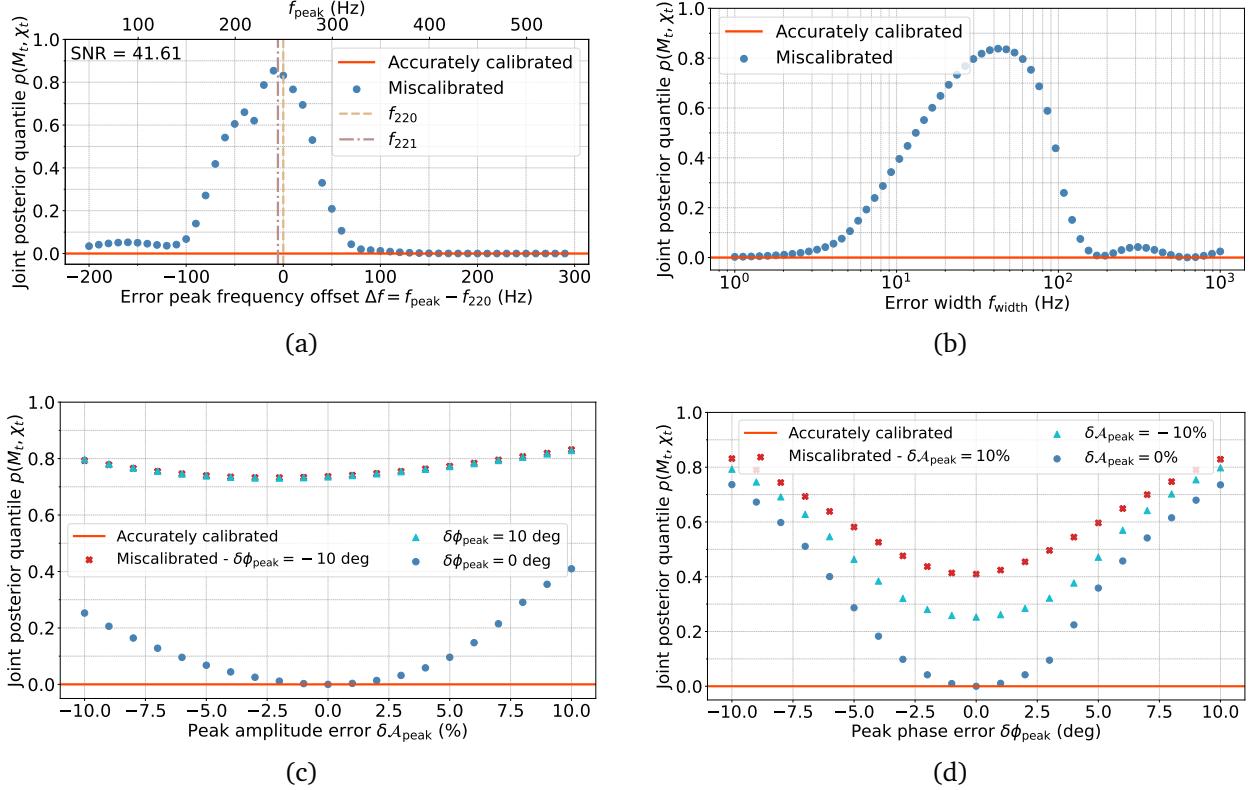


Figure 15: Joint posterior quantile values, $p(M_t, \chi_t)$, for a manually constructed 220 and 221 signal from a GW150914-like system (not from an NR waveform) in aLIGO Hanford without additive noise as a function of the calibration error properties: (a) the offset between the peak frequency of the error and the 220 QNM frequency $\Delta f = f_{\text{peak}} - f_{220}$ ($f_{\text{width}} = 50$ Hz, $\delta A_{\text{peak}} = 10\%$, and $\delta \phi_{\text{peak}} = -10$ deg), (b) error width f_{width} ($\Delta f = 0$, $\delta A_{\text{peak}} = 10\%$, and $\delta \phi_{\text{peak}} = -10$ deg), (c) peak magnitude error δA_{peak} ($\Delta f = 0$, $f_{\text{width}} = 50$ Hz), and (d) peak phase error $\delta \phi_{\text{peak}}$ ($\Delta f = 0$, $f_{\text{width}} = 50$ Hz). In the second row, the coloured markers indicate different fixed values of $\delta \phi_{\text{peak}}$ and δA_{peak} on the left and right, respectively. The red horizontal lines indicate the perfect recovery of (M, χ) from correctly calibrated data, with all $p(M_t, \chi_t)$ aligned with 0.

The differences between Figs. 15 and 16 come from the impact of additional mode components and the numerical noise in the NR waveform. The most obvious difference is in panel (a), showing $p(M_t, \chi_t)$ as a function of Δf . For the NR waveform, the results are less symmetric when the calibration error peak moves away from the QNM frequency. When the error is at lower frequencies, it seems to cause a larger bias in the NR waveform analysis than the analysis of the manually constructed, controlled signal. The biggest bias for the NR waveform occurs slightly on the left of the 220 and 221 QNM frequencies. Such differences can partly be explained by the superposition of the positive and negative frequency components [see Eq. (15)], which involves a phase factor determined by the

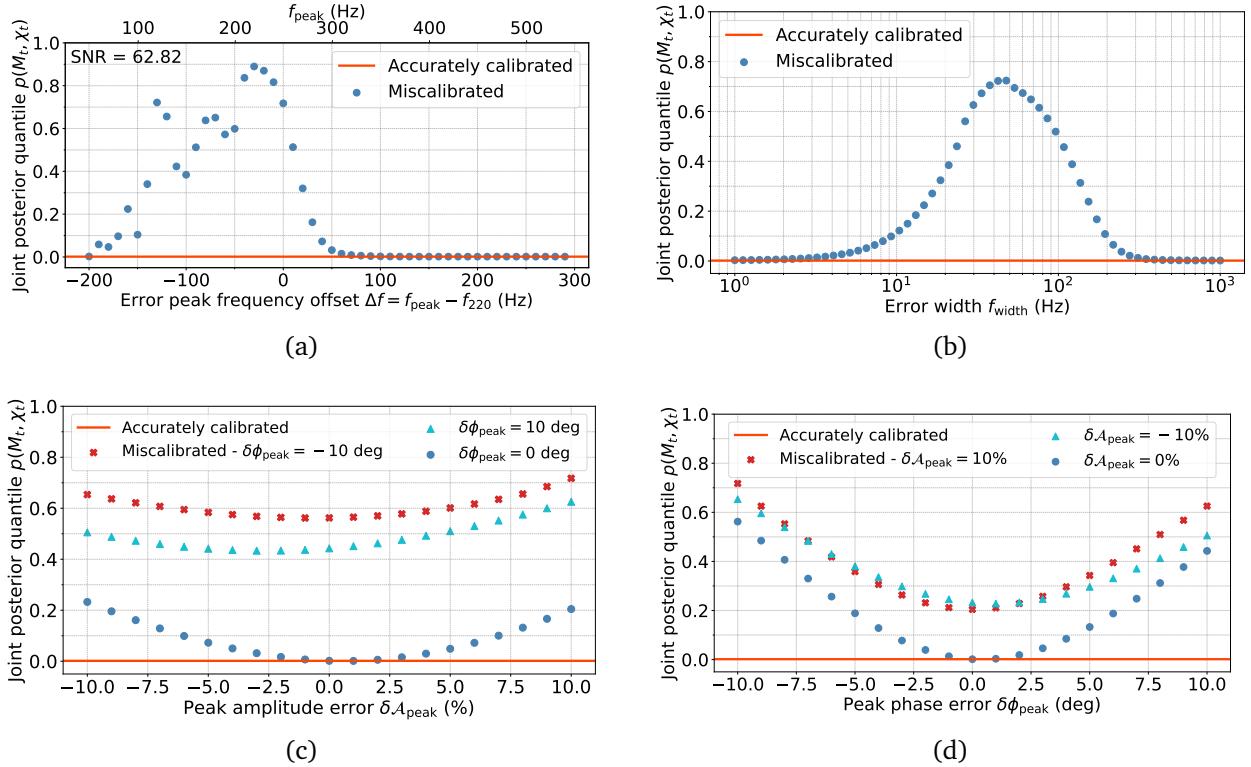


Figure 16: (Similar to Fig. 15) Joint posterior quantile values, $p(M_t, \chi_t)$, for a GW150914-like NR waveform ($Y_{2\pm 2}$ components only) injected in aLIGO Hanford without additive noise as a function of the calibration error properties. (See details in Fig. 15 caption.)

source system. The phase factor is different between the NR waveform and the manually constructed signal. In addition, the NR waveform contains higher-order overtones, retrograde modes, other angular modes mixed into $Y_{2\pm 2}$ directions, and numerical noise. Minor differences in the trends between Figs. 15 and 16 are also due to differences in the shape of the waveform and the way different mode components are mixed. For instance, positive magnitude errors lead to a larger bias than negative magnitude errors in the manually constructed, controlled signals, but not in the NR waveform [compare Fig. 15 (d) and Fig. 16 (d)].

The scenarios investigated here are realistic in the calibrated data products in the existing runs. Calibration errors with $f_{\text{width}} \sim 10\text{--}100$ Hz, $\delta A_{\text{peak}} \sim \pm 10\%$, and $\delta\phi \sim \pm 10$ deg are commonly seen in the existing observing runs [65, 71]. Note that the errors reported in Refs. [65, 71] are identified and quantified ones, which in principle can be accounted for or incorporated in the analysis procedure to reduce the potential impact. However, similar unidentified calibration errors may exist in the data production procedure, which could inevitably lead to biased estimates of astrophysical parameters.

4.4 Simulations in white Gaussian noise

In this section, we introduce white Gaussian noise to the injection process and investigate the impact of calibration errors on QNM analysis in the presence of noise. When analysing the 220 and 221 modes in a GW150914-like waveform, the QNM frequencies are around 200–300 Hz (see top panel of Fig. 13), where the GW detectors are most sensitive and have a relatively flat ASD (see Fig. 5). As such, it is a good approximation to simulate the signals in white noise. Fig. 17 shows the injected signal from a GW150914-like NR waveform ($Y_{2\pm 2}$ components only) without additive noise (studied in Sec. 4.3) in red and with additive white Gaussian noise (studied in this section) in blue.

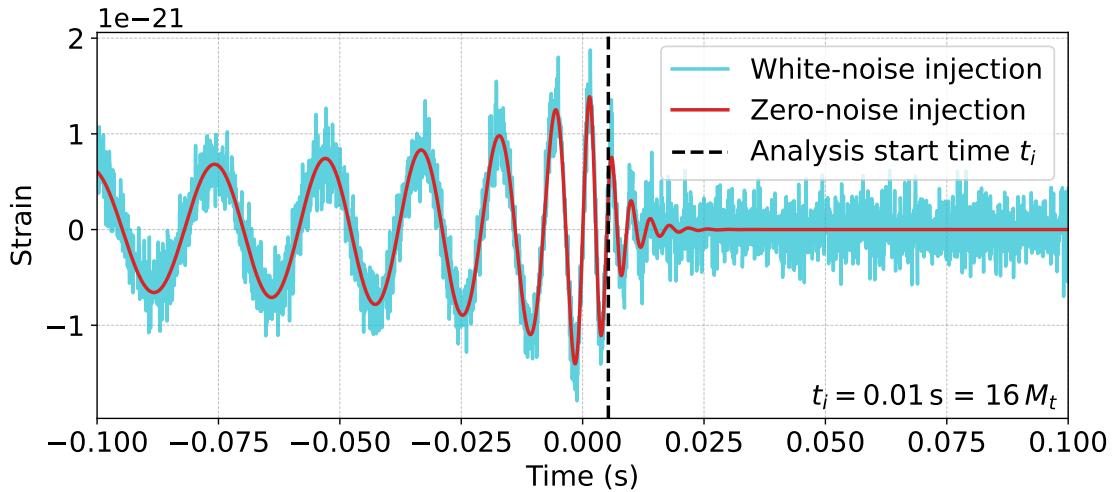


Figure 17: Time-domain NR waveform of a GW150914-like system ($Y_{2\pm 2}$ components only) injected into a detector frame with zero noise (red) and random white Gaussian noise (blue). The analysis starting time, $t_i = 16M_t$, as determined in Sec. 4.3.2, is marked by the black dashed line. Time at 0 is when the signal strain reaches its maximum.

Analysing the signal in additive noise leads to a bias in the parameter estimation and results in a non-zero $p(M_t, \chi_t)$, even without any calibration error. Different random noise realisations lead to statistical uncertainties in the $p(M_t, \chi_t)$ values. In order to determine the number of realisations, N , needed to produce a converged distribution of $p(M_t, \chi_t)$, i.e. to ensure the median and $\pm 1\sigma$ bounds of $p(M_t, \chi_t)$ do not vary with more realisations, we plot $p(M_t, \chi_t)$ as a function of the number of trials in Fig. 18 for analysing a manually constructed, controlled 220 signal from a GW150914-like system injected into white Gaussian noise (without any calibration error). Note that the analysis starting time, $t_i = 3M_t$, is used (earlier than $t_i = 16M_t$ used for NR waveform injections because the manually-constructed, controlled signals are cleaner and do not contain higher-order overtones at earlier times). The total SNR of the full ringdown part in each noise realisation differs, so we give the median and $\pm 1\sigma$ of the SNRs from all the trials in the bottom left corner of the plot. The distribution of $p(M_t, \chi_t)$ converges with $N > 50$, we select a more conservative number of $N = 100$ for the rest of the thesis.

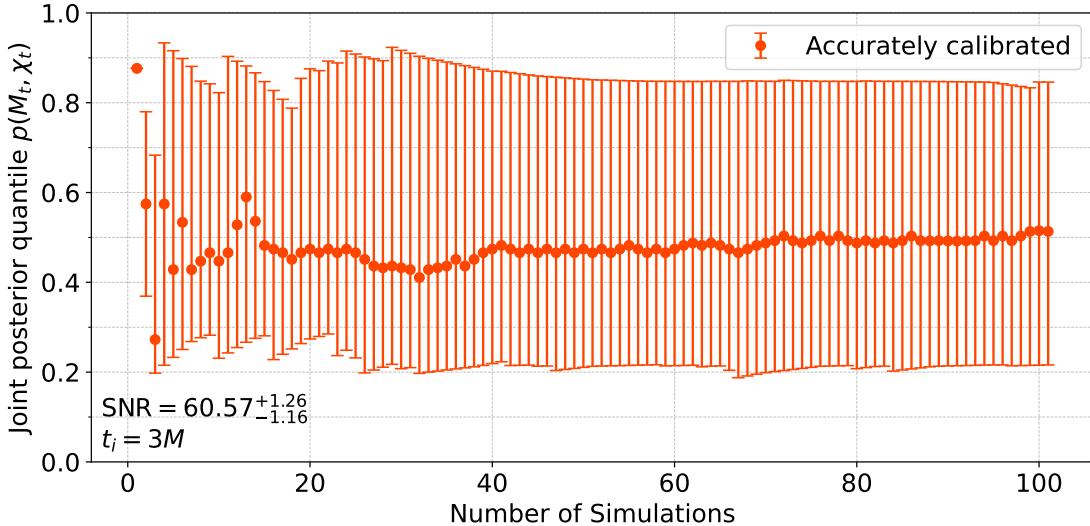
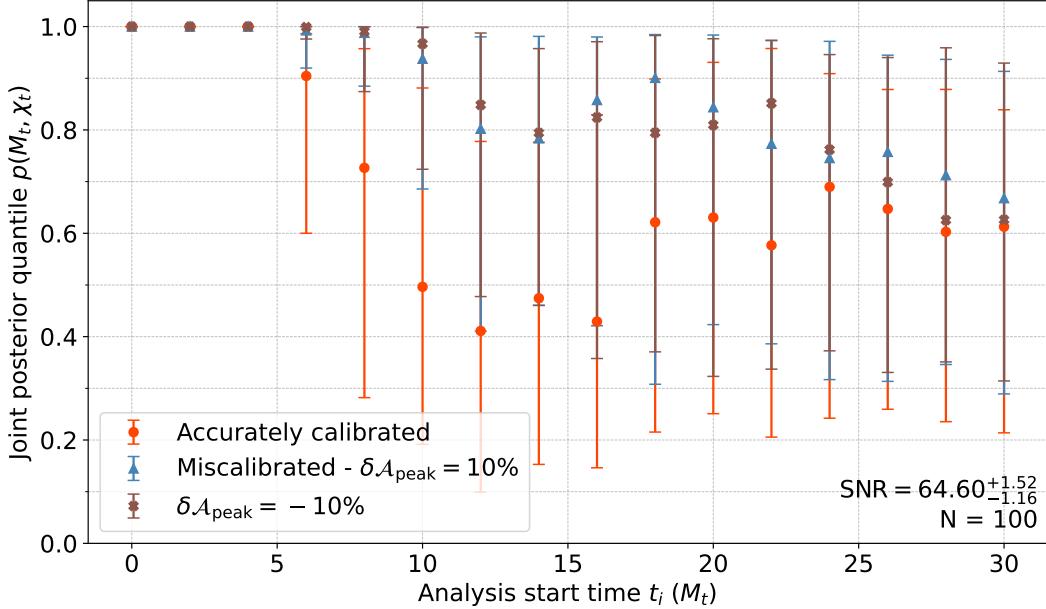
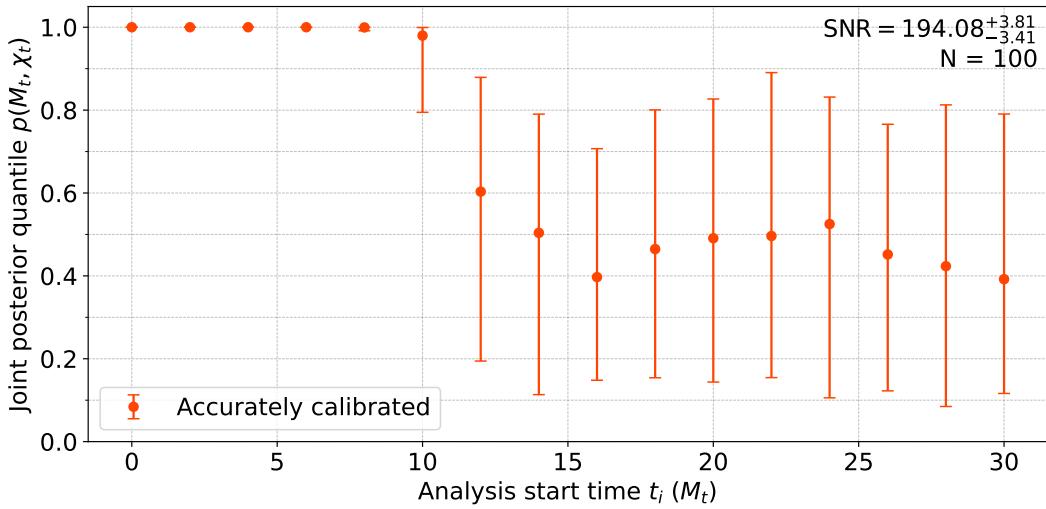


Figure 18: Joint posterior quantile values, $p(M_t, \chi_t)$, for a manually-constructed, controlled 220 signal from a GW150914-like system injected into white Gaussian noise as a function of the total number of simulations. The variation of the $p(M_t, \chi_t)$ values is from different trials in random Gaussian noise. The dots stand for the median of the $p(M_t, \chi_t)$, and the error bars indicate the $\pm 1\sigma$ bounds. The distribution converges around 50 trials. We select a more conservative number, 100 trials, for the studies in the simulations with the presence of noise. No calibration error is added to the simulations. The $p(M_t, \chi_t)$ values deviate from zero (perfect recovery) due to the impact of noise.

Fig. 19 depicts the joint posterior quantile $p(M_t, \chi_t)$ as a function of analysis starting time t_i for a GW150914-like NR waveform ($Y_{2\pm 2}$ components only) injected into white Gaussian noise in (a) a lower SNR regime and (b) a higher SNR regime. In Fig. 19a (total ringdown SNR ~ 65), the results are shown for accurately calibrated data in red and miscalibrated data with $f_{\text{peak}} = f_{220}$, $f_{\text{width}} = 50$ Hz, $\delta\phi = -10$ deg, and $\delta\mathcal{A}_{\text{peak}} = 10\%$ (-10%) in blue (brown). Fig. 19b shows the results for accurately calibrated data with a higher SNR of ~ 194 . In both Figs. 19a and 19b, a minimum of $p(M_t, \chi_t)$ (best recovery) is found around $t_i = 16M_t$, which is consistent with the same analysis for an injection without additive noise in Fig. 14a. However, Fig. 19 demonstrates a more realistic scenario with the presence of noise. The $p(M_t, \chi_t)$ values in the lower SNR regime increases after $t_i = 16M_t$, unlike in Fig. 14a where $p(M_t, \chi_t)$ remains at zero. This is because the later in time the analysis starts (with larger t_i), the lower the SNR left in the data (the QNMs have decayed). The SNR values we quote in the plots are the total SNR after $t_0 = 0$. For example, in Fig. 19a, the total SNR of ~ 65 reduces to ~ 29 when the analysis starting time is $t_i = 16M_t$, and further reduces to ~ 14 when $t_i = 25M_t$. At a low SNR of ~ 14 , the additive noise starts to dominate and the BH parameter recovery degrades. This is not the case in a higher SNR regime, as shown in Fig. 19b. Here, the SNR is still at the level of ~ 40 when starting at $t_i = 25M_t$, sufficient for decent BH parameter recovery. We can see that $p(M_t, \chi_t)$ plateaus in a manner similar to Fig. 14a where no noise is added. However, notice that the 1σ bounds slightly expand as the SNR decreases (t_i increases), indicating



(a)



(b)

Figure 19: Joint posterior quantile values, $p(M_t, \chi_t)$, for a GW150914-like NR waveform ($Y_{2\pm 2}$ components only) injected into white Gaussian noise as a function of the analysis starting time t_i . Panel (a) shows results for accurately calibrated (red) and miscalibrated data in a lower SNR regime with a total ringdown SNR of ~ 65 . Two calibration magnitude errors are tested, $\delta A_{\text{peak}} = 10\%$ (blue) and -10% (brown), with $\delta\phi_{\text{peak}} = -10$ deg, $f_{\text{peak}} = f_{220}$, and $f_{\text{width}} = 50$ Hz. The $p(M_t, \chi_t)$ values show that the recovery of the BH parameters is always worse in miscalibrated data. Panel (b) shows the results for accurately calibrated data in a higher SNR regime with a total ringdown SNR of ~ 194 . The markers are the median $p(M_t, \chi_t)$ from 100 trials, and the error bars indicate $\pm 1\sigma$ bounds. The error bars around the median SNRs indicate the $\pm 1\sigma$ variation among 100 trials. The best recovery can be achieved after $t_i \sim 16M_t$ in NR waveform injections in white Gaussian noise, consistent with the results obtained in injections without additive noise (see Fig. 14a).

a larger uncertainty in the recovery. In addition, Fig. 19a shows a weak trend at $t_i \sim 16-20M_t$, a positive magnitude error ($\delta\mathcal{A}_{\text{peak}} = 10\%$) leads to slightly worse $p(M_t, \chi_t)$ than a negative magnitude error ($\delta\mathcal{A}_{\text{peak}} = -10\%$), which is consistent with the results obtained without additive noise in Fig. 16d (when $\delta\phi_{\text{peak}} = -10$ deg).

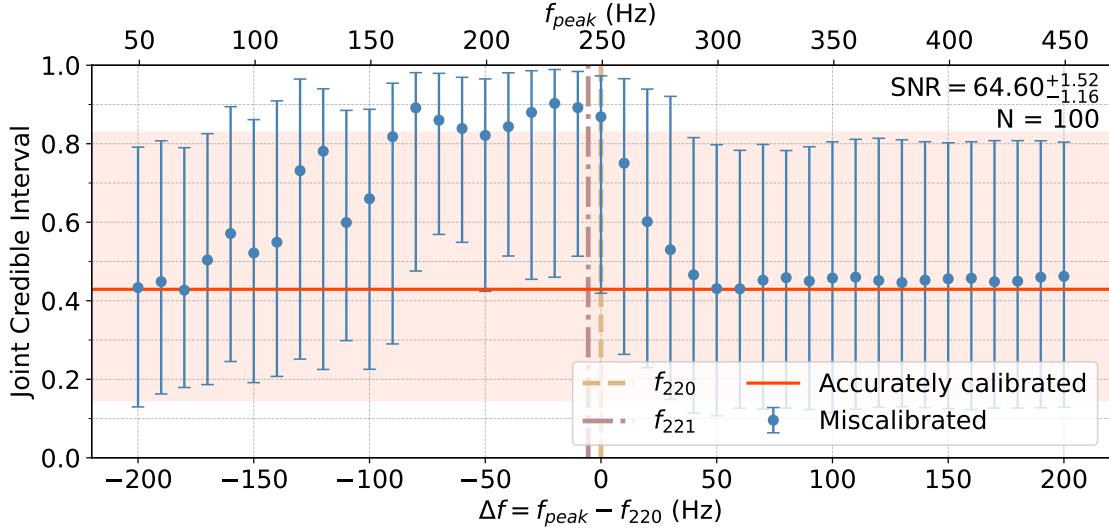


Figure 20: Joint posterior quantile values, $p(M_t, \chi_t)$ (shown in blue), for a GW150914-like NR waveform ($Y_{2\pm 2}$ components only), injected into white Gaussian noise, as a function of the offset between the peak frequency of the calibration error and the 220 QNM frequency, $\Delta f = f_{\text{peak}} - f_{220}$ (other error parameters: $f_{\text{width}} = 50$ Hz, $\delta\mathcal{A}_{\text{peak}} = 10\%$, $\delta\phi_{\text{peak}} = -10$ deg). The dots are the median $p(M_t, \chi_t)$ from 100 trials, and the error bars indicate $\pm 1\sigma$ bounds. The results from analysing the accurately calibrated data are shown in red (the line indicates the median, and the shade is the $\pm 1\sigma$ bounds).

We then carry out the same analysis as shown in Fig. 16a (no additive noise) but with white Gaussian noise added to the injection. The results are shown in Fig. 20. We plot $p(M_t, \chi_t)$ (blue dots with error bars indicating $\pm 1\sigma$ bounds) as a function of $\Delta f = f_{\text{peak}} - f_{220}$, the distance between the peak frequencies of the calibration error and the 220 QNM. The other calibration error parameters are fixed at $f_{\text{width}} = 50$ Hz, $\delta\mathcal{A}_{\text{peak}} = 10\%$, and $\delta\phi_{\text{peak}} = -10$ deg. The red line and shade indicate the median $p(M_t, \chi_t)$ value and the $\pm 1\sigma$ bounds from 100 trials in the accurately calibrated data. The Δf values that yield blue dots lying outside the red shaded region indicate that the results obtained from the miscalibrated data disagree (to a level beyond 1σ) with the results when there is no calibration error. Note that with the presence of noise, the analysis in the accurately calibrated data does not recover (M_t, χ_t) perfectly, i.e., the red line does not align with 0, and the shaded band is due to the statistical variation in random Gaussian noise. When f_{peak} moves away from f_{220} , e.g., $\Delta f > 50$ Hz, there is negligible effect on the analysis, and the results from the miscalibrated data largely align with those obtained from the data without calibration errors, with a maximum difference of 0.03 in the median $p(M_t, \chi_t)$ values. The trend matches the results in Fig. 16a, when there is no noise, i.e., calibration errors no longer have an impact with $\Delta f > 50$ Hz. When $\Delta f < 0$, i.e., the f_{peak} moves

towards lower frequencies, the calibration error still impacts the results due to the mixing of the positive and negative QNM frequency components, unless f_{peak} is sufficiently far away, e.g., $\Delta f \sim -200$ Hz. The feature of a worse recovery at $\Delta f = -120$ Hz and -130 Hz is also seen in the no-noise case in Fig. 16a, which is again, likely to be an effect of the mixing between positive and negative QNM frequency components.

5 Prospects for current and next-generation observatories

Thus far, we have investigated how the four calibration error parameters, f_{peak} , f_{width} , δA_{peak} , and $\delta \phi_{\text{peak}}$, affect the bias introduced in the QNM analysis of BH ringdown signals, when the data is miscalibrated. We have performed analyses for a GW150914-like NR waveform injected into the detector frame without additive noise, and identified key trends in BH parameter recovery as a function of the four calibration error characteristics using a metric of $p(M_t, \chi_t)$. We have then conducted studies in more realistic scenarios by adding white Gaussian noise. In this section, we investigate the calibration accuracy required in order to perform robust ringdown analyses in coloured detector noise. Given that the current state-of-the-art calibration accuracy is $|\delta A| \lesssim 10\%$ and $|\delta \phi| \lesssim 10$ deg in aLIGO [65, 71], we first investigate whether the current level of calibration error impacts the ringdown analyses with current-generation detectors. Evans et al. [52] predicts that a subpercent level of calibration accuracy will be required for next-generation detectors. We quantify the detailed calibration accuracy requirements for BH spectroscopy with next-generation observatories through simulations using the framework and procedure developed in this study. Given the QNMs studied here are within a frequency band of ~ 200 – 300 Hz, where the ASD of the ground-based detectors is generally flat, the overall trends of the impacts from calibration errors are expected to be similar to the results we obtained in white Gaussian noise. As such, we fix $f_{\text{peak}} = f_{220}$ and $f_{\text{width}} = 50$ Hz for the rest of the study, which leads the largest bias in the analysis, as shown in Fig. 16 [panels (a) and (b)], and vary δA_{peak} and $\delta \phi_{\text{peak}}$.

We start the studies in the relatively low SNR regime with aLIGO at its design sensitivity (Sec. 5.1). Then, we simulate relatively high SNR signals in Cosmic Explorer (Sec. 5.2), and discuss future work to investigate, in particular, extreme high SNR events in the next-generation observatories (Sec. 5.3).

5.1 Analyses with current-generation observatories

We study the impact of calibration errors on ringdown analysis with current-generation observatories, using the aLIGO Hanford detector as an example in terms of the calculation in the detector frame and noise budget (the ASD of the detector). We use the GW150914-like

NR waveform from the SXS catalog with a total binary mass of $70.6M_{\odot}$ and a luminosity distance of $D_L = 410$ Mpc (similar to the real GW150914 event) [46].

To begin with, we conduct the same test as shown in Fig. 19a (with white Gaussian noise) at different analysis starting times. Here, instead of white Gaussian noise, we add coloured Gaussian noise to the injected signal using the ASD of the aLIGO Hanford detector at design sensitivity (the orange curve in Fig. 5). The results are plotted in Fig. 21. The blue and red dots show the median joint posterior quantile values $p(M_t, \chi_t)$ among 100 noise realisations, for miscalibrated and accurately calibrated data, respectively, with the error bars indicating the $\pm 1\sigma$ bounds. The calibration error is fixed at $f_{\text{peak}} = f_{220}$, $f_{\text{width}} = 50$ Hz, $\delta\mathcal{A}_{\text{peak}} = 10\%$, and $\delta\phi_{\text{peak}} = -10$ deg. Compared to the results in white Gaussian noise (Fig. 19a), the trend is generally the same, with the miscalibrated results being always worse than the accurately calibrated ones. Overall, the recovery of the BH parameters is slightly worse [larger $p(M_t, \chi_t)$], even in the accurately calibrated data, due to the lower SNR ~ 33 (about half compared to the SNR in Fig. 19a). The impact of the calibration errors is also weaker at a given time compared to Fig. 19a, i.e., the difference between the blue and red dots is smaller. We fix $t_i = 16M_t$ for the rest of the study (consistent with other similar tests in this thesis).

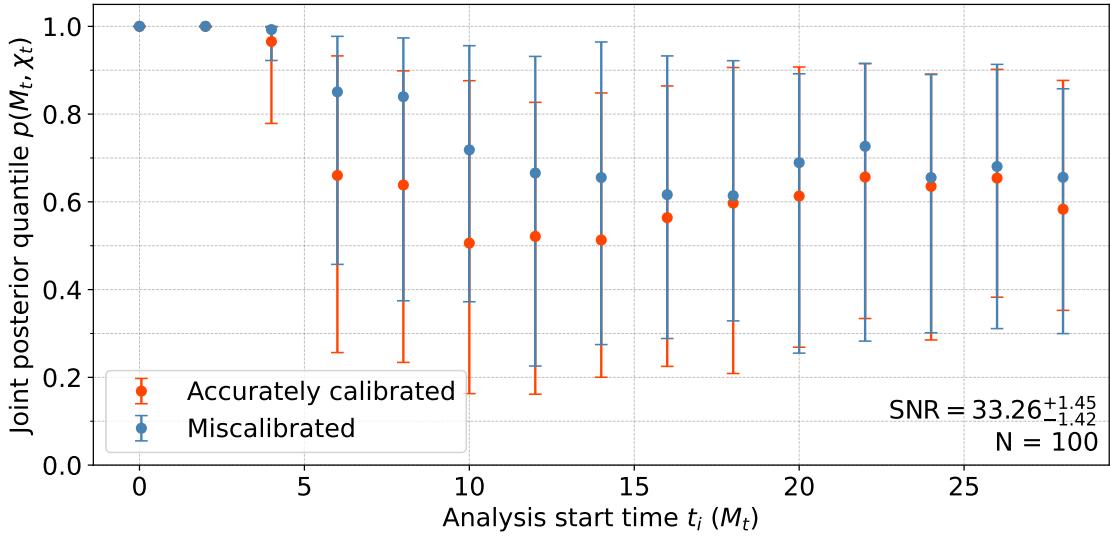


Figure 21: Joint posterior quantile values, $p(M_t, \chi_t)$, for a GW150914-like NR waveform ($Y_{2\pm 2}$ components only) injected into aLIGO Hanford with ASD at the design sensitivity, as a function of the analysis starting time t_i . The signal is injected at a luminosity distance of $D_L = 410$ Mpc (similar to GW150914), equivalent to a total ringdown SNR of ~ 33 (with error bars around the SNRs indicating the $\pm 1\sigma$ variation among the 100 trials). The blue and red colours stand for the results from miscalibrated data and accurately calibrated data, respectively, with the error bars indicating the $\pm 1\sigma$ bounds. The calibration error parameters are $f_{\text{peak}} = f_{220}$, $f_{\text{width}} = 50$ Hz, $\delta\mathcal{A}_{\text{peak}} = 10\%$ and $\delta\phi_{\text{peak}} = -10$ deg. The $p(M_t, \chi_t)$ values show that the recovery of the BH properties is always worse in miscalibrated data.

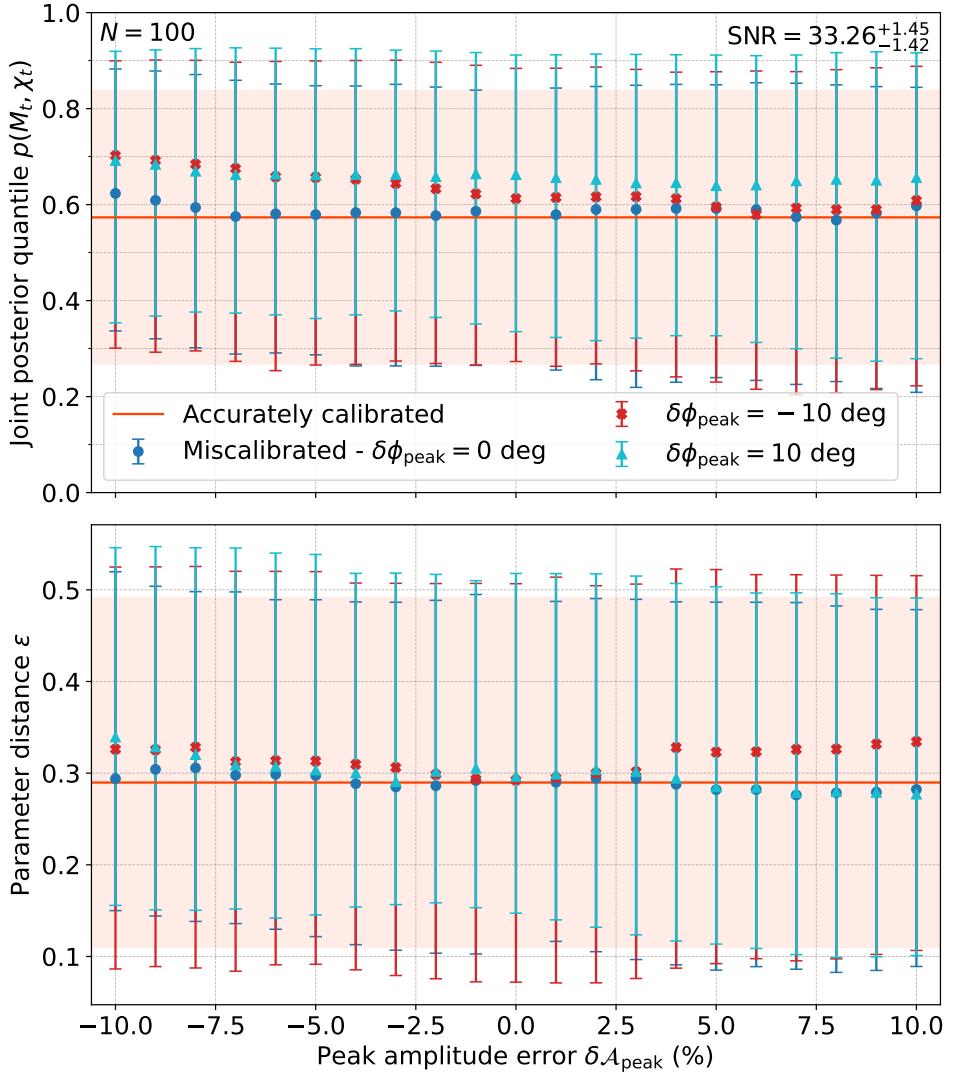


Figure 22: Joint posterior quantile values, $p(M_t, \chi_t)$ (top), and parameter distance, ϵ (bottom), for a GW150914-like NR waveform ($Y_{2\pm 2}$ components only) injected into aLIGO Hanford with ASD at the design sensitivity ($D_L = 410$ Mpc), as a function of the peak magnitude error in calibration, $\delta\mathcal{A}_{\text{peak}}$, with three choices of peak phase error shown in the legend (other error parameters: $f_{\text{peak}} = f_{220}$, $f_{\text{width}} = 50$ Hz). The red line and shade represent the median results from the accurately calibrated data, and the $\pm 1\sigma$ bounds, respectively. Overall, the estimated parameters are not biased in the miscalibrated data, with the median $p(M_t, \chi_t)$ and ϵ falling within the red shade (the 1σ uncertainty bounds of the results from accurately calibrated data).

Next, we show that the state-of-the-art calibration accuracy in the existing observing runs generally does not impact ringdown QNM analyses with current-generation observatories. In addition to the joint posterior quantile values, $p(M_t, \chi_t)$, we adopt an additional metric, the distance between the MAP values and the true BH parameters, ϵ , defined in Eq. (35), to provide a more comprehensive evaluation. Smaller values of both $p(M_t, \chi_t)$

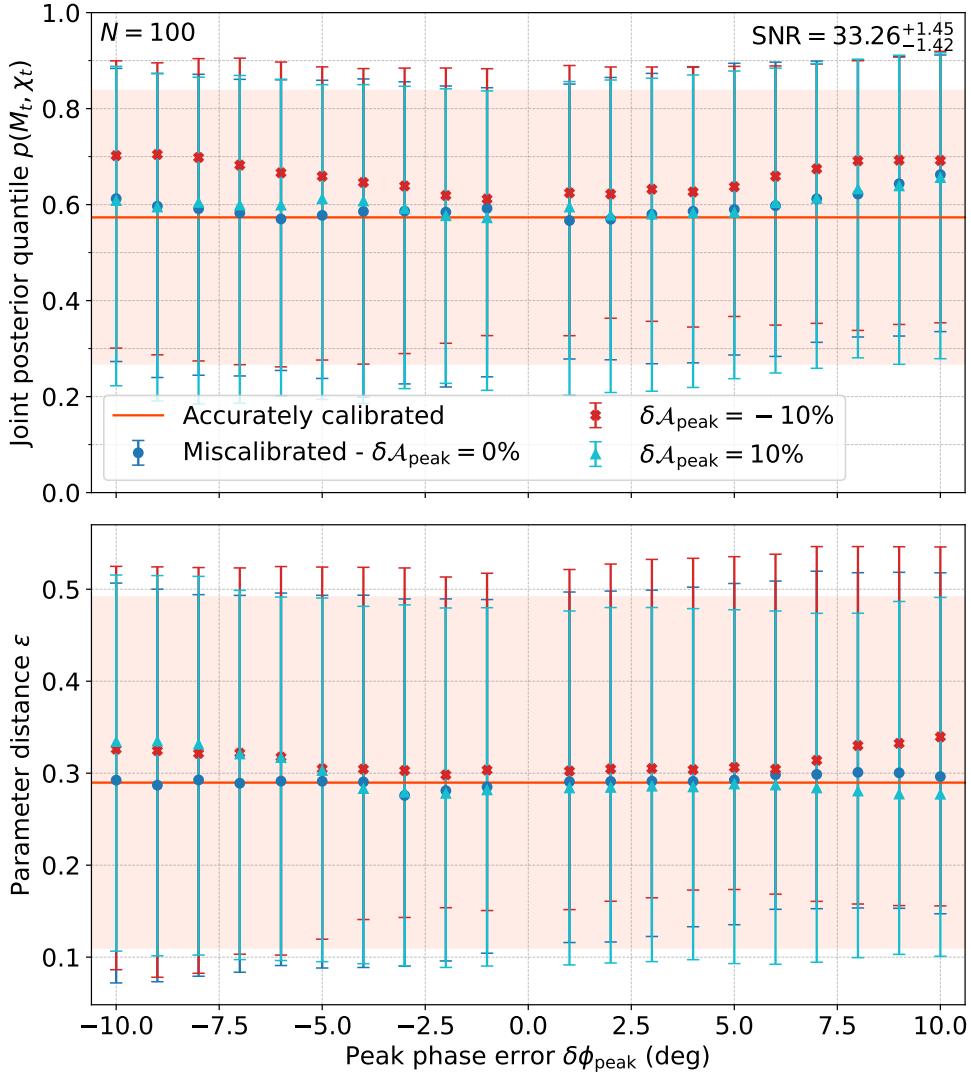


Figure 23: Joint posterior quantile values, $p(M_t, \chi_t)$ (top), and parameter distance, ϵ (bottom), for a GW150914-like NR waveform ($Y_{2\pm 2}$ components only) injected into aLIGO Hanford with ASD at the design sensitivity ($D_L = 410$ Mpc) as a function of the peak phase error in calibration, $\delta\phi_{\text{peak}}$, with three choices of magnitude error shown in the legend (other error parameters: $f_{\text{peak}} = f_{220}$, $f_{\text{width}} = 50$ Hz). The red line and shade represent the median results from the accurately calibrated data and the $\pm 1\sigma$ bounds, respectively. Overall, the estimated parameters are not biased in the miscalibrated data, with the median $p(M_t, \chi_t)$ and ϵ falling within the red shade (the 1σ uncertainty bounds of the results from accurately calibrated data).

and ϵ indicate better recovery of the BH parameters.

In Figs. 22 and 23, we show $p(M_t, \chi_t)$ and ϵ for miscalibrated and accurately calibrated data with an injected GW150914-like NR waveform ($Y_{2\pm 2}$ components only) at 410 Mpc into aLIGO Hanford at design sensitivity. In Fig. 22 and Fig. 23, we vary the peak magni-

tude error $\delta\mathcal{A}_{\text{peak}}$ and peak phase error $\delta\phi_{\text{peak}}$, respectively, and hold the other parameter to three choices of values as shown in the legend. The red line and shade represent the median and $\pm 1\sigma$ bounds when there is no calibration error. Comparing the top panels of $p(M_t, \chi_t)$ values to the similar tests conducted in earlier sections without additive noise, or with white Gaussian noise, we see a similar ‘U’-shaped trend (but much gentler), i.e., larger calibration errors lead to slightly worse recovery. Interestingly, in this case, having a negative magnitude error leads to a slightly worse recovery than having a positive one (e.g., compare the red and cyan markers in Fig. 23), possibly because a positive magnitude error elevates the strain amplitude at the mode frequency, and slightly benefits the fitting at a low SNR. All of the $p(M_t, \chi_t)$ and ϵ values are generally consistent with the results obtained in the accurately calibrated data, regardless of what the $\delta\mathcal{A}_{\text{peak}}$ or $\delta\phi_{\text{peak}}$ values are. The largest difference between the median values of $p(M_t, \chi_t)$ and ϵ in miscalibrated and accurately calibrated data is 0.17 and 0.06, respectively. All median values of $p(M_t, \chi_t)$ and ϵ are well within the 1σ uncertainty bounds of the accurately calibrated results, indicating that the bias introduced by the calibration error is negligible. In the low SNR regime, noise obscures the signal, increases the uncertainty in parameter estimation, and thus weakens the impact of calibration errors. We inject a signal from 410 Mpc, the distance of the closest BBH event observed in the first three observing runs [2–5], into aLIGO at its design sensitivity. We also adopt the worst-case calibration error scenario seen in the existing runs. Thus, we do not expect the current level of calibration accuracy to cause biases in BH ringdown studies with current-generation observatories.

5.2 Calibration requirements for next-generation observatories

With next-generation observatories, e.g., Cosmic Explorer and Einstein Telescope, we will be able to do BH spectroscopy with high SNR events, revealing more subdominant modes and carrying out precision tests of GR. In this section, we select a reasonably high SNR level, e.g., a GW150914-like event at a luminosity distance of $D_L = 2$ Gpc, which gives a total ringdown SNR of ~ 125 . For extremely high SNR events, e.g., a GW150914-like signal at $D_L = 410$ Mpc, which yields a total ringdown SNR of ~ 600 , we require more careful scrutiny by zooming into a narrower parameter space and increasing the resolution of the (M, χ) grid in the likelihood calculation procedure. We defer the study in extremely high SNR regimes to future work and discuss the preliminary studies in Sec. 5.3.

In Figs. 24 and 25 (similar to Figs. 22 and 23 for the current-generation observatories discussed in Sec. 5.1), we plot $p(M_t, \chi_t)$ and ϵ for miscalibrated and accurately calibrated data with an injected GW150914-like NR waveform ($Y_{2\pm 2}$ components only) at 2 Gpc into Cosmic Explorer (sensitivity indicated by the purple curve in Fig. 5). We vary $\delta\mathcal{A}_{\text{peak}}$ and $\delta\phi_{\text{peak}}$ in Figs. 24 and 25, respectively, and hold the other at 0. In the top panels for $p(M_t, \chi_t)$, the ‘U’-shaped trends, seen in the study without additive noise, become more significant compared to those in the lower-SNR regime in Sec. 5.1. The larger magnitude or phase calibration errors still lead to a more significant bias in recovering the BH parameters. Note that the recovery in accurately calibrated data is improved due to the

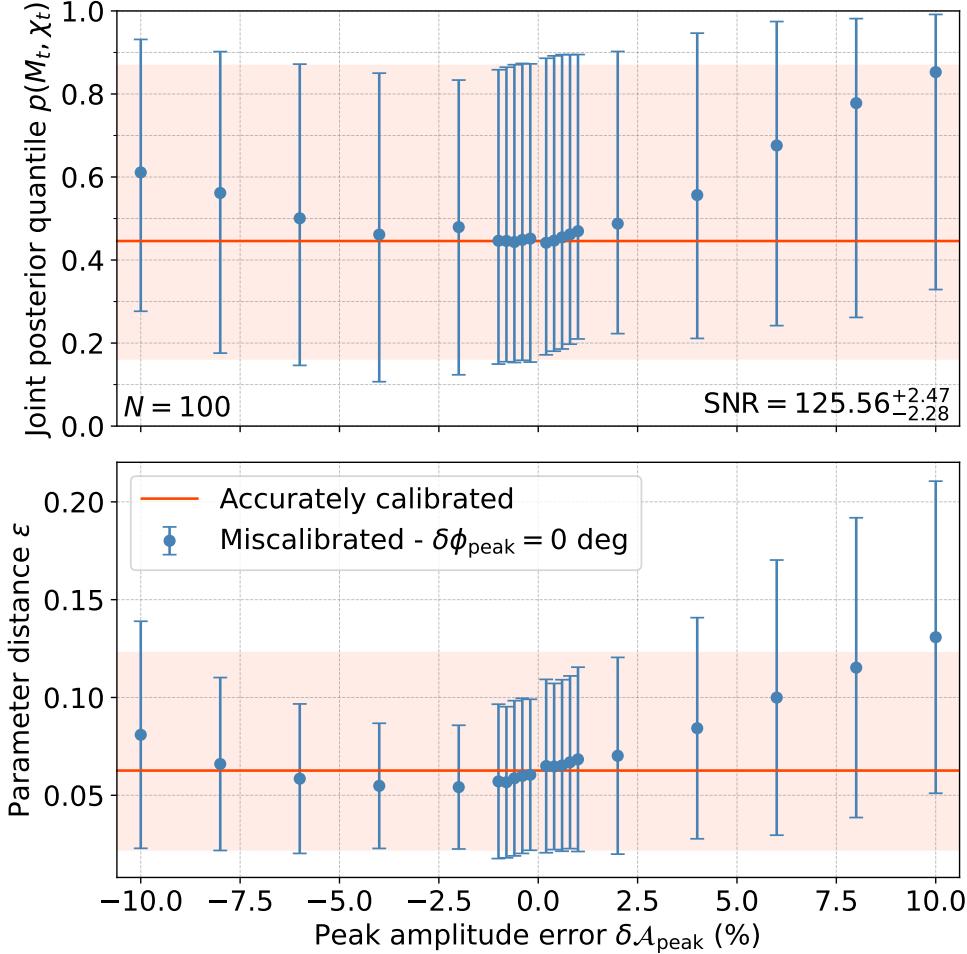


Figure 24: Joint posterior quantile values, $p(M_t, \chi_t)$ (top), and parameter distance, ϵ (bottom), for a GW150914-like NR waveform ($Y_{2\pm 2}$ components only) injected into Cosmic Explorer ($D_L = 2$ Gpc, equivalent to a total ringdown SNR of ~ 125) as a function of the peak magnitude error in calibration, $\delta\mathcal{A}_{\text{peak}}$, with $\delta\phi_{\text{peak}} = 0$, $f_{\text{peak}} = f_{220}$, and $f_{\text{width}} = 50$ Hz. The blue dots with error bars (red line with surrounding shade) represent the median and $\pm 1\sigma$ bounds, respectively, obtained from the miscalibrated data (accurately calibrated data).

higher SNR, with $p(M_t, \chi_t) \sim 0.45$ and $\epsilon \sim 0.06$, compared to the results in Sec. 5.1 with $p(M_t, \chi_t) \sim 0.58$ and $\epsilon \sim 0.28$. The most significant bias in the recovery of the BH parameters is caused by a phase error with $\delta\phi_{\text{peak}} = -10$ deg, pushing the median values of both $p(M_t, \chi_t)$ and ϵ outside the 1σ bounds in the case that the data are accurately calibrated. Generally, a calibration error of $|\delta\mathcal{A}_{\text{peak}}| \gtrsim 8\%$ or $|\delta\phi_{\text{peak}}| \gtrsim 8$ deg can lead to significant bias in recovering the BH parameters. Thus, the state-of-the-art calibration accuracy of the past observing runs with current-generation observatories does not satisfy the requirement for precision tests of GR and BH spectroscopy with next-generation observatories. New methodologies and techniques need to be developed to achieve a higher level of accuracy in calibration in the next-generation observatories.

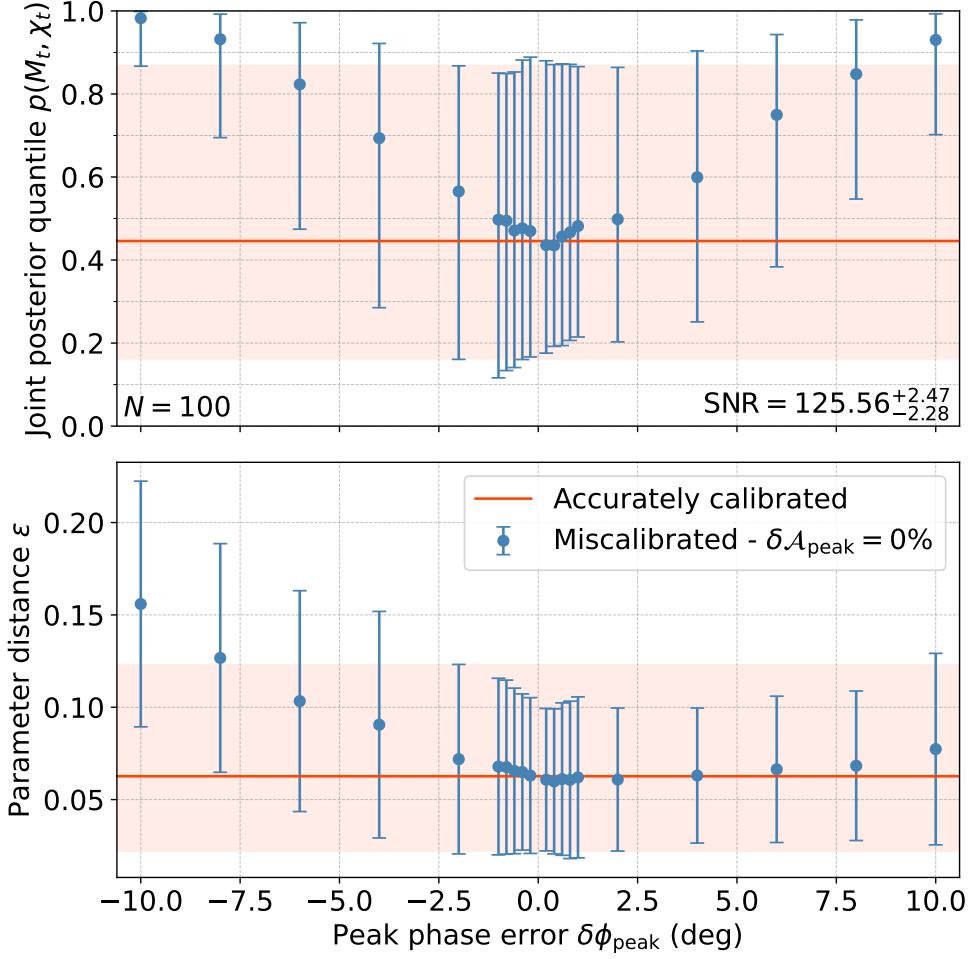


Figure 25: Joint posterior quantile values, $p(M_t, \chi_t)$ (top), and parameter distance, ϵ (bottom), for a GW150914-like NR waveform ($Y_{2\pm 2}$ components only) injected into Cosmic Explorer ($D_L = 2$ Gpc, equivalent to a total ringdown SNR of ~ 125) as a function of the peak phase error in calibration, $\delta\phi_{\text{peak}}$, with $\delta\mathcal{A}_{\text{peak}} = 0$, $f_{\text{peak}} = f_{220}$, and $f_{\text{width}} = 50$ Hz. The blue dots with error bars (red line with surrounding shade) represent the median and $\pm 1\sigma$ bounds, respectively, obtained from the miscalibrated data (accurately calibrated data).

Here, we only miscalibrate the data by applying either a magnitude error or a phase error (unlike Figs. 22 and 23). Certain combinations of both magnitude and phase errors may lead to larger biases with next-generation observatories. We leave the study of other configurations of calibration errors to future work.

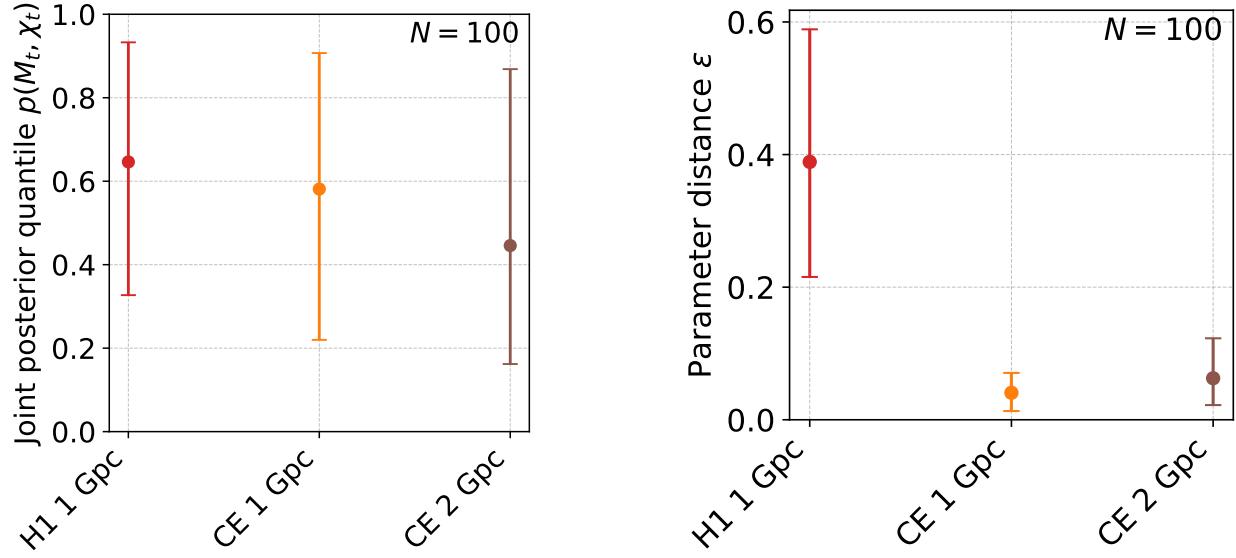


Figure 26: Joint posterior quantile values, $p(M_t, \chi_t)$ (left), and parameter distance, ϵ (right), for an injected GW150914-like NR waveform (without calibration errors). The detector used and the luminosity distance of the source are listed on the horizontal axis. For conciseness, aLIGO Hanford at design sensitivity is denoted by ‘H1’, and Cosmic Explorer is denoted by ‘CE’ in the axis label. The dots with error bars indicate the median and $\pm 1\sigma$ bounds in 100 coloured Gaussian noise realisations. Note the $p(M_t, \chi_t)$ values are similar among all three cases, but a much larger difference is seen in ϵ values. For aLIGO at design sensitivity, the SNRs are low for sources at 1 Gpc and the BH parameters cannot be well constrained (error bars for ϵ are large). With Cosmic Explorer, an order of magnitude higher SNRs can be achieved compared to aLIGO. The uncertainty of the parameter estimates becomes much smaller in the high SNR regime (error bars for ϵ are small in Cosmic Explorer even with a source at 2 Gpc). The $p(M_t, \chi_t)$ value alone is no longer a good indicator of the recovery.

5.3 Events in extreme signal-to-noise regime

If a GW150914-like event at a luminosity distance of 410 Mpc is observed in Cosmic Explorer, we obtain a total ringdown SNR of ~ 600 . In such an extremely high SNR regime, the likelihood function peaks steeply in a very narrow parameter space, and a fine-resolution, zoomed-in analysis is required to obtain the most accurate constraints on the BH parameters. As shown in Fig. 13, the 90% credible region expands when starting the analysis at later times because the QNM decays over time, and thus, the SNR decreases. Vice versa, at an extremely high SNR, the 90% credible region shrinks to a small pixel and requires an extremely dense grid in the parameter space to resolve different credible intervals. With the same resolution in the (M, χ) plane as used in a lower SNR regime, the exact credible interval where (M_t, χ_t) lies cannot be resolved, resulting in an incorrectly quoted $p(M_t, \chi_t)$.

Fig. 26 shows $p(M_t, \chi_t)$ and ϵ for an injected GW150914-like NR waveform (without calibration errors) under three scenarios (as shown in the horizontal axis): source at 1 Gpc in aLIGO Hanford (design sensitivity), source at 1 Gpc in Cosmic Explorer, and source at 2 Gpc in Cosmic Explorer, respectively, with a total ringdown SNR of ~ 13 , ~ 250 , and ~ 125 . We would expect the injection at 1 Gpc in Cosmic Explorer to yield the best parameter estimates and thus the lowest $p(M_t, \chi_t)$. This is not the case, with the injection at 2 Gpc giving a $p(M_t, \chi_t)$ about 0.2 lower than the closer injection. However, the ϵ value does show a better parameter recovery (closer to the true value with a smaller uncertainty) for the closer injection at 1 Gpc in Cosmic Explorer. This indicates that $p(M_t, \chi_t)$ may not be a consistent indicator of performance across different scenarios (i.e., different SNRs or detectors), and may have issues at higher SNRs, although it serves as a good performance indicator in a lower SNR regime. If the resolution of the grid in the (M, χ) plane used for calculating likelihoods is not fine enough in the high SNR regime, an unresolved credible interval can lead to an incorrect $p(M_t, \chi_t)$. A much finer resolution in the (M, χ) plane or a much zoomed-in parameter space may be necessary.

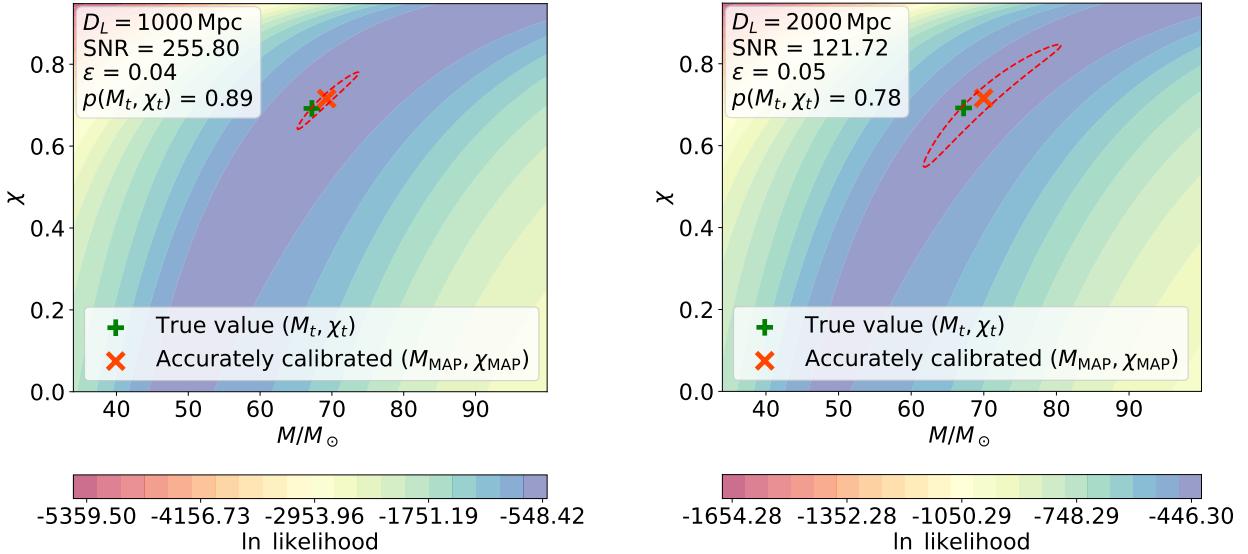


Figure 27: Joint posterior distribution of M and χ for a GW150914-like NR waveform injected into Cosmic Explorer (without calibration errors) at 1 Gpc (left), and 2 Gpc (right). The dashed contour indicates the 90% credible region found by integrating the joint posterior. The green plus and red cross indicate the true and maximum a posteriori (MAP) values, respectively.

In Fig. 27, we plot the joint posterior distribution of M and χ (coloured contours) and 90% credible region (red dashed contours) for a pair of injections in Cosmic Explorer from a source at 1 Gpc (left) and 2 Gpc (right). The MAP value is closer to the true value (a better recovery) for the injection at 1 Gpc, but a larger $p(M_t, \chi_t)$ value is obtained (a seemingly worse recovery) simply because the 90% credible region in the left panel is not well resolved and thus incorrectly calculated. In the right panel, the 90% credible region is broader compared to the left, but still narrow. The distance between the MAP and true values, ϵ , is generally a more reliable indicator in the high SNR regime.

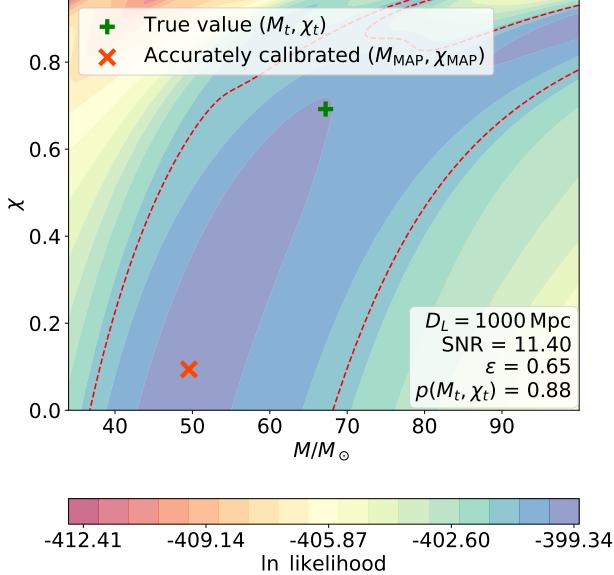


Figure 28: (Similar to Fig. 27) Joint posterior distribution of M and χ for a GW150914-like NR waveform injected into aLIGO Hanford (design sensitivity) at 1 Gpc.

We also note that $p(M_t, \chi_t)$ alone may not be a consistent indicator of the performance for parameter estimation across different scenarios. For example, Fig. 28 shows an injection with a low SNR of ~ 11 . We find $p(M_t, \chi_t) = 0.39$, much lower than those in Fig. 27 where the SNRs are much higher. On the other hand, the $(M_{\text{MAP}}, \chi_{\text{MAP}})$ values are far away from the true values (ϵ is much larger than those in Fig. 27, indicating a much worse recovery). The $p(M_t, \chi_t)$ values are however informative, especially at a lower SNR regime, which is more relevant to the current-generation GW observatories, as it represents whether the true parameters are covered in the credible region when the uncertainty of the parameter estimation is high (e.g., the 90% credible regions are broad). The parameter distance, ϵ , is a more informative indicator of the accuracy of the parameter recovery at a high SNR regime when the uncertainty of the estimates gets low.

Further investigations are required to better understand the behaviours in the high (and extremely high) SNR regime to more accurately determine $p(M_t, \chi_t)$, and to define a better metric to quantify the accuracy and precision of the parameter recovery. We consider this as future work.

6 Conclusion

Gravitational waves (GW) emitted from compact binary mergers are encoded with information about the most extreme environments in our Universe. Accurately analysing these GW signals will allow us to explore the dynamics of dense matter, investigate populations of compact objects throughout cosmic time, and study fundamental physics and extreme

gravity [52]. Black hole (BH) spectroscopy methods aim to analyse the ringdown GWs from the remnant BH formed in a binary merger, and recover the properties of the remnant BH and the progenitor binary system. The accuracy of such analyses in part depends on the accuracy of the detector’s calibrated output data. Systematic errors in the detector calibration process may induce biases in BH spectroscopy and lead to incorrect astrophysical inferences. In this study, we aim to investigate the impact on BH ringdown analyses of the ‘worst-case’ calibration errors seen in current GW detectors, using a set of rational quasinormal filters (Sec. 4.1). The results inform calibration accuracy and precision required in GW observatories for robust BH spectroscopy.

In order to investigate the effect of certain calibration errors, we model calibration errors with four properties, namely the peak frequency, frequency-domain characteristic width, peak magnitude error, and peak phase error. We simulate artificial calibration errors based on real error examples commonly seen in the third observing run of aLIGO (Sec. 4.2). We then miscalibrate signal waveforms and use the rational quasinormal filter to analyse both the miscalibrated and accurately calibrated waveforms in different noise situations. We analyse injected signals in the detector frame without additive noise (Sec. 4.3), with white Gaussian noise (Sec. 4.4), and with coloured simulated detector noise (Sec. 5). The analysis conducted in simulated detector noise primarily informs the calibration requirements for accurate and robust BH ringdown analyses in current-generation (Sec. 5.1) and next-generation (Sec. 5.2) GW observatories. We use the currently achieved calibration accuracy as a point of reference, with magnitude errors $\lesssim 10\%$ and phase errors $\lesssim 10$ deg, and confirm such level of calibration errors does not bias BH ringdown studies in the current-generation GW observatories. However, calibration errors with peak magnitude error $\gtrsim 8\%$ or peak phase error $\gtrsim 8$ deg can lead to significantly biased results in BH ringdown analyses in the next-generation observatories. This is primarily due to the increased signal-to-noise ratio (SNR) that can be achieved by future observatories.

The results obtained in this study still have limitations: we have only investigated the dominant fundamental mode 220 and its first overtone, 221, for one GW150914-like system, with limited combinations of the calibration magnitude and phase errors, and in a limited SNR regime. Next-generation GW observatories will observe events at an extremely high SNR regime (with a ringdown SNR of ~ 600), which has not been scrutinised in this study. For such an extreme SNR, it is likely that a much more stringent calibration accuracy will be required for BH spectroscopy. Potential approaches for detailed investigations are outlined in Sec. 5.3. Further more, to determine a more comprehensive set of calibration requirements for future observatories, we aim to apply the analysis formulated in this study to a broader scope in the future. It would primarily involve applying more types of possible calibration errors and studying different binary merger systems. As discussed in Sec. 5.2, we only investigate certain limited combinations of peak magnitude and phase errors for next-generation GW observatories, but more combinations of the four calibration error properties need to be tested. Different binary merger systems excite different modes in the ringdown stage, not limited to the dominant fundamental 22 mode. An excitation of two fundamental modes at separate frequencies, e.g., 22 and 33 modes, with decent

SNRs, allows us to study the impact of calibration errors in a broader frequency range. A preliminary analysis of the multi-fundamental-mode scenario has been conducted, with examples provided in the Appendix. A comprehensive investigation on more complicated mode excitation mixed with calibration errors will be conducted as future work.

A Appendix

A.1 Estimates of (M, χ) with calibration errors at different frequencies

Fig. 29 shows the 90% credible regions (dashed contours) of the (M, χ) joint posterior for a manually constructed, controlled 220 and 221 signal from a GW150914-like system (not from an NR waveform), injected in aLIGO Hanford without additive noise. These signals are miscalibrated by a calibration error with $f_{\text{width}} = 50$ Hz, $\delta A_{\text{peak}} = 10\%$, and $\delta \phi_{\text{peak}} = -10$ deg, and a set of varying $\Delta f = f_{\text{peak}} - f_{220}$ as indicated in the top left corner of each subplot. Selected time-series data are shown on the top panel. This set of contour plots corresponds to the results shown in Fig. 15a. As evidenced in the contour plots, when $\Delta f = 0$, i.e., $f_{\text{peak}} = f_{220}$, the 90% credible region and the MAP values obtained from the miscalibrated data (blue) are most distinct from the accurately calibrated results (red).

A.2 Analysis of waveforms with two fundamental modes

Here, we present the preliminary results for investigating the scenario where two fundamental modes are excited, e.g., the 220 and 330 QNMs. Fig. 30 shows a similar set of results as Fig. 29, but with the injection being a manually constructed signal with two fundamental modes, 220 and 330, from a GW150914-like system. Overall, the biases caused by calibration errors are less significant compared to the analysis for 220 and 221 QNMs (the blue and red results are much more consistent). The biggest impact still happens when f_{peak} is near any of the QNM frequencies.

Fig. 31 shows the joint posterior quantile values that the true BH parameters correspond to, $p(M_t, \chi_t)$, for a manually constructed, controlled injection of the 220 and 330 QNMs from a GW150914-like system without additive noise (similar to Fig. 15 in Sec. 4.3.3). Note that the lower the value of $p(M_t, \chi_t)$, the better the BH parameters are recovered. In panel (a), the bimodal feature around the two QNM peak frequencies demonstrates that when the calibration error peak frequency f_{peak} is close to any of the QNM frequencies, the impact from the miscalibration is maximised. Panel (b) shows that a calibration error with f_{width} up to ~ 500 Hz can still lead to biased parameter estimation. Unlike the analysis with 220 and 221 modes, which is restricted to a narrower band, signals with 220 and 330 QNMs span a wider frequency band, and thus a wider calibra-

tion error may still matter. Panels (c) for $\delta\mathcal{A}_{\text{peak}}$ and (d) for $\delta\phi_{\text{peak}}$ also show more of an asymmetric tread compared to Fig. 15. Notably, the parameter recovery in Fig. 31 is generally better compared to Fig. 15, as evidenced by the ~ 0.2 lower $p(M_t, \chi_t)$ values. Two QNMs at distinct frequencies present in the signal may make the ringdown analysis less susceptible to calibration errors because the calibration errors are less likely to mimic the behavior of two distinct modes at the same time. Note that in this example, we manually construct a controlled signal with an enlarged 330 component, comparable to the dominant 220 mode. However, the 330 mode or other angular modes are usually much weaker and thus may not contribute enough power to reduce the impact of calibration errors. A more detailed investigation will be continued as future work in order to obtain a more comprehensive understanding of calibration requirements for BH spectroscopy.

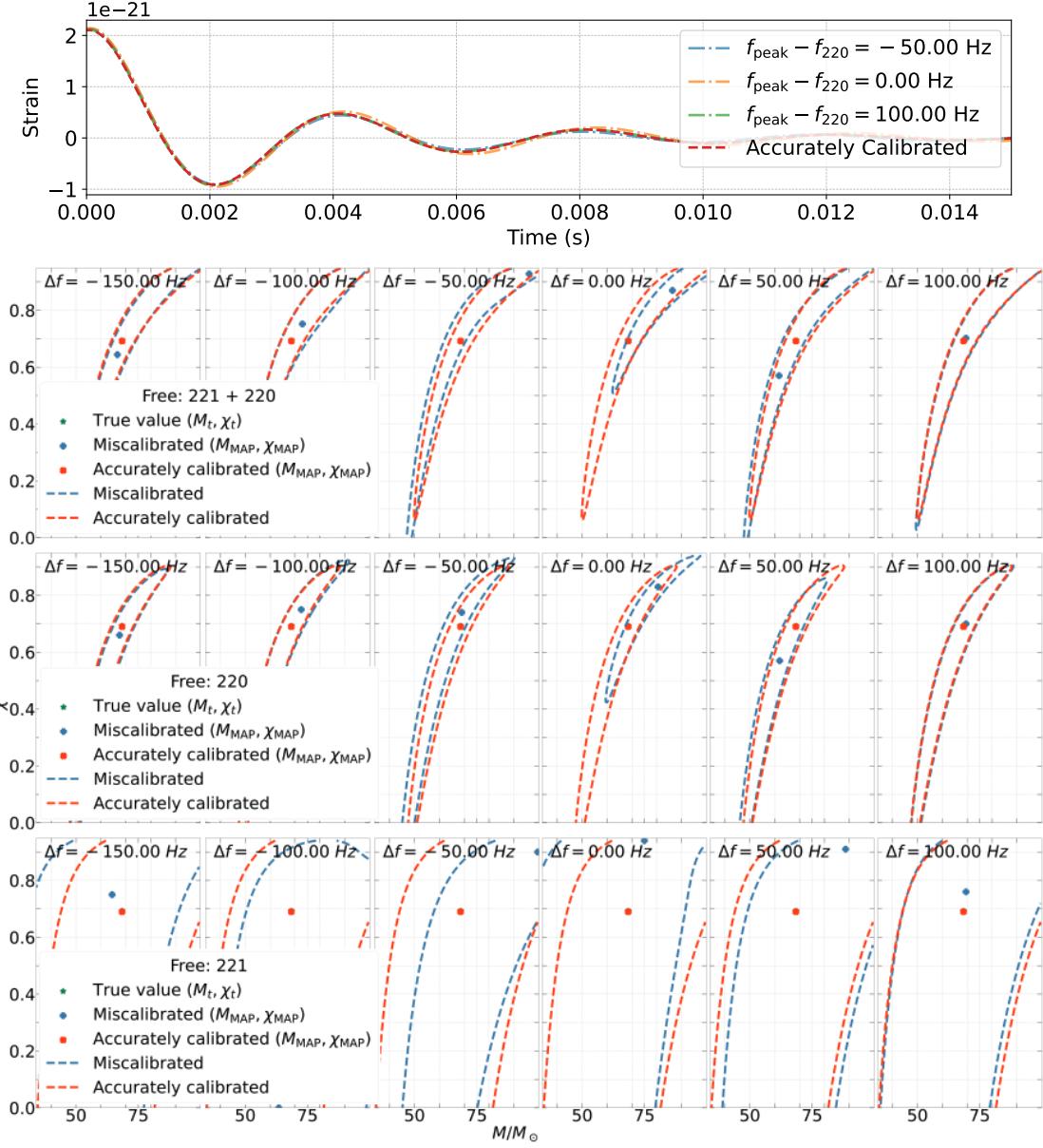


Figure 29: The 90% credible region (dashed contours) of the (M, χ) joint posterior for a manually constructed 220 and 221 signal from a GW150914-like system (not from an NR waveform) in aLIGO Hanford without additive noise (SNR = 42). The red and blue contours stand for the results from the accurately calibrated and miscalibrated data, respectively. The red cross and blue plus markers are the maximum a posteriori (MAP) values from the data without and with calibration errors, respectively. The green star denotes the true BH parameter values. Each vertical column corresponds to a different f_{peak} , as labelled in the top left corner of each subplot. The other calibration error parameters are $\delta\mathcal{A}_{\text{peak}} = 10\%$, $\delta\phi_{\text{peak}} = -10 \text{ deg}$, and $f_{\text{width}} = 50 \text{ Hz}$. Select time-series data with and without calibration errors are shown in the top panel (see $\Delta f = f_{\text{peak}} - f_{220}$ in legend). Each row corresponds to the parameter estimation using a different set of QNMs. In the first row, the 220 and 221 modes are fitted together. In the second (third) row, the 221 (220) mode is removed before fitting for the 220 (221) mode only. The mode(s) used in the fitting are denoted by the ‘‘free’’ mode(s) in the legend.

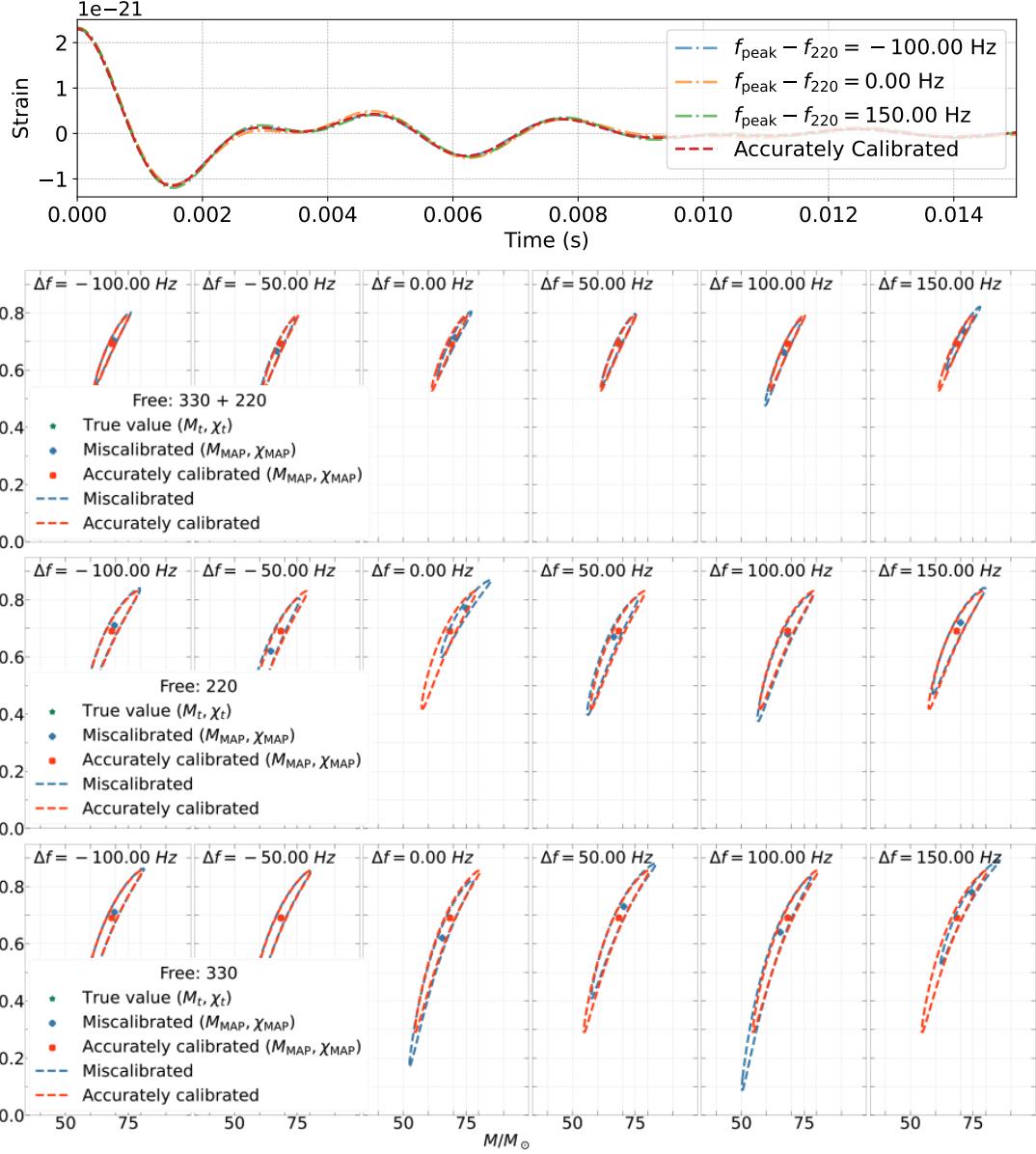


Figure 30: (Similar to Fig. 29) The 90% credible region (dashed contours) of the (M, χ) joint posterior for a manually constructed 220 and 330 signal from a GW150914-like system (not from an NR waveform) in aLIGO Hanford without additive noise (SNR = 42). (See Fig. 29 caption for details.)

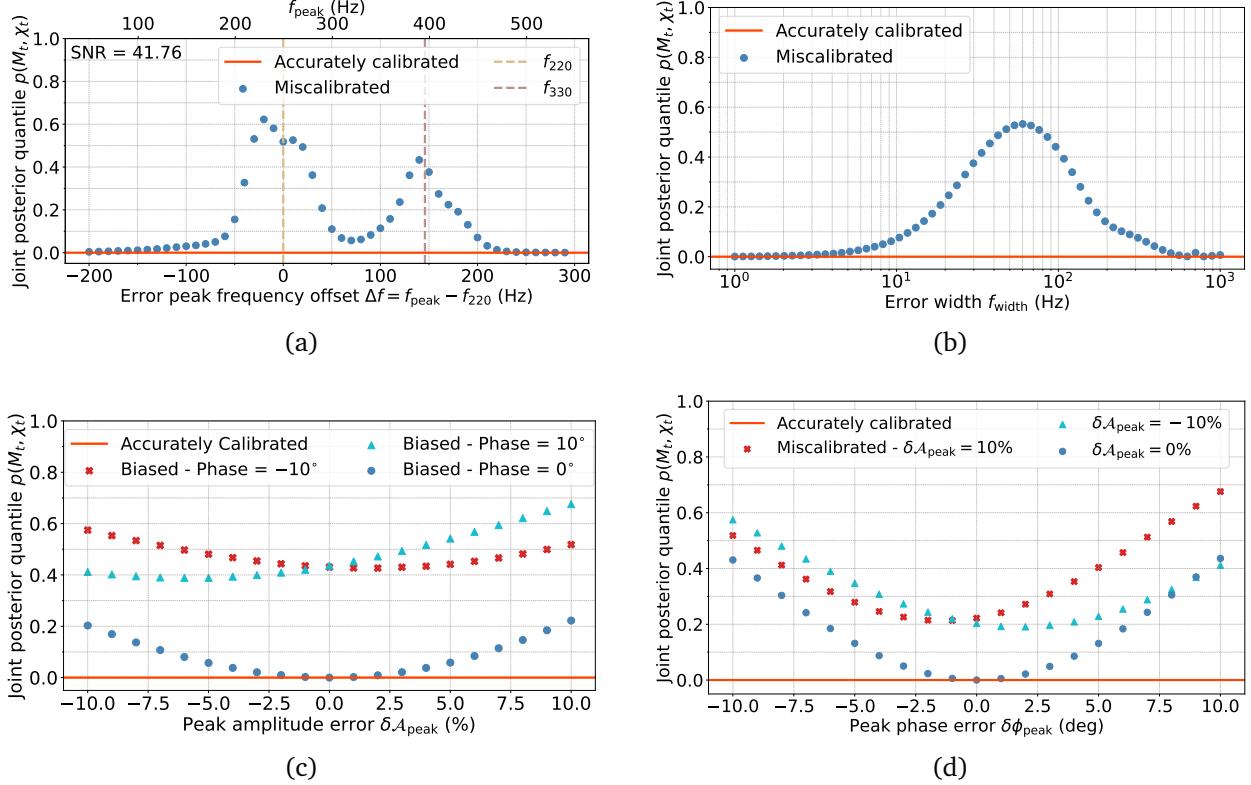


Figure 31: Joint posterior quantile values, $p(M_t, \chi_t)$, for a manually constructed 220 and 330 signal from a GW150914-like system (not from an NR waveform) in aLIGO Hanford without additive noise as a function of the calibration error properties: (a) the offset between the peak frequency of the error and the 220 QNM frequency $\Delta f = f_{\text{peak}} - f_{220}$ ($f_{\text{width}} = 50$ Hz, $\delta A_{\text{peak}} = 10\%$, and $\delta \phi_{\text{peak}} = -10$ deg), (b) error width f_{width} ($\Delta f = 0$, $\delta A_{\text{peak}} = 10\%$, and $\delta \phi_{\text{peak}} = -10$ deg), (c) peak magnitude error δA_{peak} ($\Delta f = 0$, $f_{\text{width}} = 50$ Hz), and (d) peak phase error $\delta \phi_{\text{peak}}$ ($\Delta f = 0$, $f_{\text{width}} = 50$ Hz). In the second row, the colored markers indicate different fixed values of $\delta \phi_{\text{peak}}$ and δA_{peak} on the left and right, respectively. The red horizontal lines indicate the perfect recovery of (M, χ) from correctly calibrated data, with all $p(M_t, \chi_t)$ aligned with 0.

References

- [1] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016. doi: 10.1103/PhysRevLett.116.061102. URL <https://link.aps.org/doi/10.1103/PhysRevLett.116.061102>.
- [2] B. P. Abbott, R. Abbott, T. D. Abbott, et al. GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs. *Physical Review X*, 9(3), sep 2019. doi: 10.1103/physrevx.9.031040. URL <https://doi.org/10.1103%2Fphysrevx.9.031040>.
- [3] R. Abbott, T. D. Abbott, S. Abraham, et al. GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. *Physical Review X*, 11(2), jun 2021. doi: 10.1103/physrevx.11.021053. URL <https://doi.org/10.1103%2Fphysrevx.11.021053>.
- [4] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, T. D. Abbott, et al. Gwtc-2.1: Deep extended catalog of compact binary coalescences observed by ligo and virgo during the first half of the third observing run, 2022.
- [5] The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al. Gwtc-3: Compact binary coalescences observed by ligo and virgo during the second part of the third observing run, 2021.
- [6] Sizheng Ma, Ling Sun, and Yanbei Chen. Black hole spectroscopy by mode cleaning. *Physical Review Letters*, 130(14), apr 2023. doi: 10.1103/physrevlett.130.141401. URL <https://doi.org/10.1103%2Fphysrevlett.130.141401>.
- [7] Sizheng Ma, Keefe Mitman, Ling Sun, Nils Deppe, François Hébert, Lawrence E. Kidder, Jordan Moxon, William Throwe, Nils L. Vu, and Yanbei Chen. Quasinormal-mode filters: A new approach to analyze the gravitational-wave ringdown of binary black-hole mergers. *Physical Review D*, 106(8), oct 2022. doi: 10.1103/physrevd.106.084036. URL <https://doi.org/10.1103%2Fphysrevd.106.084036>.
- [8] Sizheng Ma, Ling Sun, and Yanbei Chen. Using rational filters to uncover the first ringdown overtone in gw150914. *Phys. Rev. D*, 107:084010, Apr 2023. doi: 10.1103/PhysRevD.107.084010. URL <https://link.aps.org/doi/10.1103/PhysRevD.107.084010>.
- [9] Michael Boyle, Daniel Hemberger, Dante A B Iozzo, Geoffrey Lovelace, Serguei Ossokine, Harald P Pfeiffer, Mark A Scheel, Leo C Stein, Charles J Woodford, Aaron B Zimmerman, Nousha Afshari, Kevin Barkett, Jonathan Blackman, Katerina Chatzioannou, Tony Chu, Nicholas Demos, Nils Deppe, Scott E Field, Nils L Fischer, Evan Foley, Heather Fong, Alyssa Garcia, Matthew Giesler, Francois Hebert, Ian Hinder, Reza Katebi, Haroon Khan, Lawrence E Kidder, Prayush Kumar, Kevin Kuper, Halston Lim, Maria Okounkova, Teresita Ramirez, Samuel Rodriguez, Hannes R Rüter,

Patricia Schmidt, Bela Szilagyi, Saul A Teukolsky, Vijay Varma, and Marissa Walker. The SXS collaboration catalog of binary black hole simulations. *Classical and Quantum Gravity*, 36(19):195006, sep 2019. doi: 10.1088/1361-6382/ab34e2. URL <https://doi.org/10.1088%2F1361-6382%2Fab34e2>.

- [10] SXS Gravitational Waveform Database. URL <https://data.black-holes.org/waveforms/index.html>.
- [11] Alexandre Le Tiec and Jérôme Novak. Theory of gravitational waves. In *An Overview of Gravitational Waves*, pages 1–41. WORLD SCIENTIFIC, feb 2017. doi: 10.1142/9789813141766_0001. URL https://doi.org/10.1142%2F9789813141766_0001.
- [12] Éanna É Flanagan and Scott A Hughes. The basics of gravitational wave theory. *New Journal of Physics*, 7:204–204, sep 2005. doi: 10.1088/1367-2630/7/1/204. URL <https://doi.org/10.1088%2F1367-2630%2F7%2F1%2F204>.
- [13] A.G. Abac et al. Observation of gravitational waves from the coalescence of a 2.5 – 4.5 m_{\odot} compact object and a neutron star, 2024.
- [14] M. Mezcua. Black holes. *arXiv e-prints*, art. arXiv:2110.08629, October 2021. doi: 10.48550/arXiv.2110.08629.
- [15] Vitor Cardoso and Paolo Pani. Testing the nature of dark compact objects: a status report. *Living Reviews in Relativity*, 22(1), jul 2019. doi: 10.1007/s41114-019-0020-4. URL <https://doi.org/10.1007%2Fs41114-019-0020-4>.
- [16] Vitor Cardoso and Leonardo Gualtieri. Testing the black hole ‘no-hair’ hypothesis. *Classical and Quantum Gravity*, 33(17):174001, August 2016. ISSN 1361-6382. doi: 10.1088/0264-9381/33/17/174001. URL <http://dx.doi.org/10.1088/0264-9381/33/17/174001>.
- [17] Emanuele Berti, Vitor Cardoso, and Andrei O Starinets. Quasinormal modes of black holes and black branes. *Classical and Quantum Gravity*, 26(16):163001, jul 2009. doi: 10.1088/0264-9381/26/16/163001. URL <https://doi.org/10.1088%2F0264-9381%2F26%2F16%2F163001>.
- [18] E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence. Metric of a Rotating, Charged Mass. *Journal of Mathematical Physics*, 6(6):918–919, June 1965. doi: 10.1063/1.1704351.
- [19] Kostas D. Kokkotas and Bernd G. Schmidt. Quasi-normal modes of stars and black holes. *Living Reviews in Relativity*, 2(1), September 1999. ISSN 1433-8351. doi: 10.12942/lrr-1999-2. URL <http://dx.doi.org/10.12942/lrr-1999-2>.
- [20] Saul A Teukolsky. The kerr metric. *Classical and Quantum Gravity*, 32(12):124006, jun 2015. doi: 10.1088/0264-9381/32/12/124006. URL <https://doi.org/10.1088%2F0264-9381%2F32%2F12%2F124006>.

- [21] Maximiliano Isi, Matthew Giesler, Will M. Farr, Mark A. Scheel, and Saul A. Teukolsky. Testing the no-hair theorem with GW150914. *Physical Review Letters*, 123(11), sep 2019. doi: 10.1103/physrevlett.123.111102. URL <https://doi.org/10.1103/2Fphysrevlett.123.111102>.
- [22] B. F. Schutz and C. M. Will. Black hole normal modes - A semianalytic approach. , 291:L33–L36, April 1985. doi: 10.1086/184453.
- [23] Gary T. Horowitz and Veronika E. Hubeny. Quasinormal modes of ads black holes and the approach to thermal equilibrium. *Physical Review D*, 62(2), June 2000. ISSN 1089-4918. doi: 10.1103/physrevd.62.024027. URL <http://dx.doi.org/10.1103/PhysRevD.62.024027>.
- [24] Hans-Peter Nollert and Bernd G. Schmidt. Quasinormal modes of schwarzschild black holes: Defined and calculated via laplace transformation. *Phys. Rev. D*, 45:2617–2627, Apr 1992. doi: 10.1103/PhysRevD.45.2617. URL <https://link.aps.org/doi/10.1103/PhysRevD.45.2617>.
- [25] Edward W. Leaver. Spectral decomposition of the perturbation response of the schwarzschild geometry. *Phys. Rev. D*, 34:384–408, Jul 1986. doi: 10.1103/PhysRevD.34.384. URL <https://link.aps.org/doi/10.1103/PhysRevD.34.384>.
- [26] Xiang Li, Ling Sun, Rico Ka Lok Lo, Ethan Payne, and Yanbei Chen. Angular emission patterns of remnant black holes. *Phys. Rev. D*, 105:024016, Jan 2022. doi: 10.1103/PhysRevD.105.024016. URL <https://link.aps.org/doi/10.1103/PhysRevD.105.024016>.
- [27] Maximiliano Isi and Will M. Farr. Analyzing black-hole ringdowns, 2021.
- [28] S. Gossan, J. Veitch, and B. S. Sathyaprakash. Bayesian model selection for testing the no-hair theorem with black hole ringdowns. *Physical Review D*, 85(12), June 2012. ISSN 1550-2368. doi: 10.1103/physrevd.85.124056. URL <http://dx.doi.org/10.1103/PhysRevD.85.124056>.
- [29] Olaf Dreyer, Bernard Kelly, Badri Krishnan, Lee Samuel Finn, David Garrison, and Ramon Lopez-Aleman. Black-hole spectroscopy: testing general relativity through gravitational-wave observations. *Classical and Quantum Gravity*, 21(4):787–803, January 2004. ISSN 1361-6382. doi: 10.1088/0264-9381/21/4/003. URL <http://dx.doi.org/10.1088/0264-9381/21/4/003>.
- [30] Emanuele Berti, Vitor Cardoso, and Clifford M. Will. Gravitational-wave spectroscopy of massive black holes with the space interferometer lisa. *Physical Review D*, 73(6), March 2006. ISSN 1550-2368. doi: 10.1103/physrevd.73.064030. URL <http://dx.doi.org/10.1103/PhysRevD.73.064030>.
- [31] Maximiliano Isi and Will M. Farr. Revisiting the ringdown of gw150914, 2022.

- [32] Roberto Cotesta, Gregorio Carullo, Emanuele Berti, and Vitor Cardoso. Analysis of ringdown overtones in gw150914. *Physical Review Letters*, 129(11), September 2022. ISSN 1079-7114. doi: 10.1103/physrevlett.129.111102. URL <http://dx.doi.org/10.1103/PhysRevLett.129.111102>.
- [33] Eliot Finch and Christopher J. Moore. Searching for a ringdown overtone in gw150914. *Physical Review D*, 106(4), August 2022. ISSN 2470-0029. doi: 10.1103/physrevd.106.043005. URL <http://dx.doi.org/10.1103/PhysRevD.106.043005>.
- [34] J. H. Taylor and J. M. Weisberg. A new test of general relativity - Gravitational radiation and the binary pulsar PSR 1913+16. , 253:908–920, February 1982. doi: 10.1086/159690.
- [35] M. E. Gertsensht\v{e}in and V. I. Pustovoit. On the Detection of Low-Frequency Gravitational Waves. *Soviet Journal of Experimental and Theoretical Physics*, 16:433, January 1963.
- [36] Rainer Weiss. Nobel lecture: Ligo and the discovery of gravitational waves i. *Rev. Mod. Phys.*, 90:040501, Dec 2018. doi: 10.1103/RevModPhys.90.040501. URL <https://link.aps.org/doi/10.1103/RevModPhys.90.040501>.
- [37] J Aasi et al. Advanced ligo. *Classical and Quantum Gravity*, 32(7):074001, mar 2015. doi: 10.1088/0264-9381/32/7/074001. URL <https://dx.doi.org/10.1088/0264-9381/32/7/074001>.
- [38] John Whelan. LIGO-T1300666-v3: Lecture Notes on Gravitational Wave Data Analysis, July 2013. URL <https://dcc.ligo.org/LIGO-T1300666/public>.
- [39] Craig Cahillane and Georgia Mansell. Review of the advanced ligo gravitational wave observatories leading to observing run four. *Galaxies*, 10(1):36, February 2022. ISSN 2075-4434. doi: 10.3390/galaxies10010036. URL <http://dx.doi.org/10.3390/galaxies10010036>.
- [40] M. Bailes, B. K. Berger, P. R. Brady, M. Branchesi, K. Danzmann, M. Evans, K. Holley-Bockelmann, B. R. Iyer, T. Kajita, S. Katsanevas, M. Kramer, A. Lazzarini, L. Lehner, G. Losurdo, H. Lück, D. E. McClelland, M. A. McLaughlin, M. Punturo, S. Ransom, S. Raychaudhury, D. H. Reitze, F. Ricci, S. Rowan, Y. Saito, G. H. Sanders, B. S. Sathyaprakash, B. F. Schutz, A. Sesana, H. Shinkai, X. Siemens, D. H. Shoemaker, J. Thorpe, J. F. J. van den Brand, and S. Vitale. Gravitational-wave physics and astronomy in the 2020s and 2030s. *Nature Reviews Physics*, 3(5):344–366, April 2021. doi: 10.1038/s42254-021-00303-8.
- [41] B. P. Abbott et al. Calibration of the advanced ligo detectors for the discovery of the binary black-hole merger gw150914. *Physical Review D*, 95(6), March 2017. ISSN 2470-0029. doi: 10.1103/physrevd.95.062003. URL <http://dx.doi.org/10.1103/PhysRevD.95.062003>.

- [42] A. Buikema et al. Sensitivity and performance of the advanced ligo detectors in the third observing run. *Physical Review D*, 102(6), September 2020. ISSN 2470-0029. doi: 10.1103/physrevd.102.062003. URL <http://dx.doi.org/10.1103/PhysRevD.102.062003>.
- [43] LIGO. Ligo scientific collaboration: Our evolving detectors, 2024. URL <https://www.ligo.caltech.edu/page/ligo-evol>. Accessed: 2024-05-21.
- [44] B. P. Abbott et al. Gw170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017. doi: 10.1103/PhysRevLett.119.161101. URL <https://link.aps.org/doi/10.1103/PhysRevLett.119.161101>.
- [45] Virgo. Virgo: Virgo history, 2024. URL <https://www.virgo-gw.eu/about/virgo-history/>. Accessed: 2024-05-21.
- [46] B. P. Abbott et al. Properties of the binary black hole merger gw150914. *Physical Review Letters*, 116(24), June 2016. ISSN 1079-7114. doi: 10.1103/physrevlett.116.241102. URL <http://dx.doi.org/10.1103/PhysRevLett.116.241102>.
- [47] T. Akutsu et al. Overview of kagra: Detector design and construction history, 2020.
- [48] David Reitze et al. The us program in ground-based gravitational wave science: Contribution from the ligo laboratory, 2019.
- [49] Matthew Evans, Riccardo Sturani, Salvatore Vitale, and Evan Hall. Unofficial sensitivity curves (asd) for aligo, kagra, virgo, voyager, cosmic explorer, and einstein telescope. Technical report, LIGO, September 2023. URL <https://dcc.ligo.org/LIGO-T2300041/public>. T1500293-v13.
- [50] Peter Fritschel, Kevin Kuns, Jenne Driggers, Anamaria Effler, Brian Lantz, David Ottaway, Stefan Ballmer, Katherine Dooley, Rana Adhikari, Matthew Evans, Benjamin Farr, Gabriela Gonzalez, Patricia Schmidt, and Sendhil Raja. Report of the lsc post-o5 study group. Technical Note LIGO-T2200287-v3, LIGO Laboratory, September 2023. Available online at <https://dcc.ligo.org/LIGO-T2200287/public>.
- [51] Kevin Kuns and Peter Fritschel. A# Strain Sensitivity. Technical report, LIGO, April 2023. URL <https://dcc.ligo.org/LIGO-T2300041/public>. T2300041-v1.
- [52] Matthew Evans, Rana X Adhikari, Chaitanya Afle, Stefan W. Ballmer, Sylvia Biscoveanu, Ssohrab Borhanian, Duncan A. Brown, Yanbei Chen, Robert Eisenstein, Alexandra Gruson, Anuradha Gupta, Evan D. Hall, Rachael Huxford, Brittany Kamai, Rahul Kashyap, Jeff S. Kissel, Kevin Kuns, Philippe Landry, Amber Lenon, Geoffrey Lovelace, Lee McCuller, Ken K. Y. Ng, Alexander H. Nitz, Jocelyn Read, B. S. Sathyaprakash, David H. Shoemaker, Bram J. J. Slagmolen, Joshua R. Smith, Varun Srivastava, Ling Sun, Salvatore Vitale, and Rainer Weiss. A horizon study for cosmic explorer: Science, observatories, and community, 2021.

- [53] Matthew Evans, Alessandra Corsi, Chaitanya Afle, Alena Ananyeva, K. G. Arun, Stefan Ballmer, Ananya Bandopadhyay, Lisa Barsotti, Masha Baryakhtar, Edo Berger, Emanuele Berti, Sylvia Biscoveanu, Ssohrab Borhanian, Floor Broekgaarden, Duncan A. Brown, Craig Cahillane, Lorna Campbell, Hsin-Yu Chen, Kathryne J. Daniel, Arnab Dhani, Jennifer C. Driggers, Anamaria Effler, Robert Eisenstein, Stephen Fairhurst, Jon Feicht, Peter Fritschel, Paul Fulda, Ish Gupta, Evan D. Hall, Giles Hammond, Otto A. Hannuksela, Hannah Hansen, Carl-Johan Haster, Keisi Kacanja, Brittany Kamai, Rahul Kashyap, Joey Shapiro Key, Sanika Khadkikar, Antonios Kontos, Kevin Kuns, Michael Landry, Philippe Landry, Brian Lantz, Tjonne G. F. Li, Geoffrey Lovelace, Vuk Mandic, Georgia L. Mansell, Denys Martynov, Lee McCuller, Andrew L. Miller, Alexander Harvey Nitz, Benjamin J. Owen, Cristiano Palomba, Jocelyn Read, Hemantakumar Phurailatpam, Sanjay Reddy, Jonathan Richardson, Jameson Rollins, Joseph D. Romano, Bangalore S. Sathyaprakash, Robert Schofield, David H. Shoemaker, Daniel Sigg, Divya Singh, Bram Slagmolen, Piper Sledge, Joshua Smith, Marcelle Soares-Santos, Amber Strunk, Ling Sun, David Tanner, Lieke A. C. van Son, Salvatore Vitale, Benno Willke, Hiro Yamamoto, and Michael Zucker. Cosmic explorer: A submission to the nsf mpsac nggw subcommittee, 2023.
- [54] Kevin Kuns, Evan Hall, Varun Srivastava, Joshua Smith, Matthew Evans, Peter Fritschel, Lee McCuller, Christopher Wipf, and Stefan Ballmer. Cosmic Explorer strain sensitivity. Technical Report CE-T2000017, Cosmic Explorer, 2022. URL <https://dcc.cosmicexplorer.org/CE-T2000017-v5/public>.
- [55] S. Hild et al. Sensitivity studies for third-generation gravitational wave observatories. *Classical and Quantum Gravity*, 28:094013, 2011. doi: 10.1088/0264-9381/28/9/094013.
- [56] Michele Maggiore, Chris Van Den Broeck, Nicola Bartolo, Enis Belgacem, Daniele Bertacca, Marie Anne Bizouard, Marica Branchesi, Sebastien Clesse, Stefano Foffa, Juan García-Bellido, Stefan Grimm, Jan Harms, Tanja Hinderer, Sabino Matarrese, Cristiano Palomba, Marco Peloso, Angelo Ricciardone, and Mairi Sakellariadou. Science case for the Einstein telescope. *Journal of Cosmology and Astroparticle Physics*, 2020(03):050, mar 2020. doi: 10.1088/1475-7516/2020/03/050. URL <https://dx.doi.org/10.1088/1475-7516/2020/03/050>.
- [57] M. Punturo et al. The Einstein Telescope: A third-generation gravitational wave observatory. *Classical and Quantum Gravity*, 27:194002, 2010. doi: 10.1088/0264-9381/27/19/194002.
- [58] Einstein Telescope Science Team. Einstein gravitational wave Telescope conceptual design study. Technical Report ET-0106C-10, European Gravitational Observatory, 2011. URL <http://www.et-gw.eu/>.
- [59] F Acernese, M Agathos, Ain Ain, S Albanesi, C Alléné, Annalisa Alcocka, A. Amato, M Andia, Tatiana Andrade, Nilson Andres, M Andrés-Carcasona, Tomislav Andric, S. Ansoldi, S Antier, T Apostolatos, E Appavuravther, M Arène, N Arnaud, M Assiduo,

and J-P Zendri. Advanced virgo plus: Future perspectives. *Journal of Physics: Conference Series*, 2429:012040, 01 2023. doi: 10.1088/1742-6596/2429/1/012040.

- [60] David Reitze. A roadmap for the ligo observatories in the era of cosmic explorer. Presentation, June 2023. URL https://www.nsf.gov/mps/phy/nggw/present_ligo.pdf. LIGO-G2301130-v1 Form F0900043-v1.
- [61] Brian D. Metzger. Kilonovae. *Living Reviews in Relativity*, 23(1), December 2019. ISSN 1433-8351. doi: 10.1007/s41114-019-0024-0. URL <http://dx.doi.org/10.1007/s41114-019-0024-0>.
- [62] Vicky Kalogera et al. The next generation global gravitational wave observatory: The science book, 2021.
- [63] David McClelland et al. 3g rd: Rd for the next generation of ground-based gravitational-wave detectors, 2021.
- [64] R X Adhikari, K Arai, A F Brooks, C Wipf, O Aguiar, P Altin, B Barr, L Barsotti, R Bassiri, A Bell, et al. A cryogenic silicon interferometer for gravitational-wave detection. *Class. Quantum Grav.*, 37(16):165003, Jul 2020. ISSN 1361-6382. doi: 10.1088/1361-6382/ab9143. URL <http://dx.doi.org/10.1088/1361-6382/ab9143>.
- [65] Ling Sun, Evan Goetz, Jeffrey S Kissel, Joseph Betzwieser, Sudarshan Karki, Aaron Viets, Madeline Wade, Dripta Bhattacharjee, Vladimir Bossilkov, Pep B Covas, Laurence E H Datrier, Rachel Gray, Shivaraj Kandhasamy, Yannick K Lecoeuche, Gregory Mendell, Timesh Mistry, Ethan Payne, Richard L Savage, Alan J Weinstein, Stuart Aston, Aaron Buikema, Craig Cahillane, Jenne C Driggers, Sheila E Dwyer, Rahul Kumar, and Alexander Urban. Characterization of systematic error in advanced ligo calibration. *Classical and Quantum Gravity*, 37(22):225008, oct 2020. doi: 10.1088/1361-6382/abb14e. URL <https://dx.doi.org/10.1088/1361-6382/abb14e>.
- [66] M Wade, A D Viets, T Chmiel, M Stover, and L Wade. Improving ligo calibration accuracy by using time-dependent filters to compensate for temporal variations. *Classical and Quantum Gravity*, 40(3):035001, January 2023. ISSN 1361-6382. doi: 10.1088/1361-6382/acabf6. URL <http://dx.doi.org/10.1088/1361-6382/acabf6>.
- [67] Yiwen Huang, Hsin-Yu Chen, Carl-Johan Haster, Ling Sun, Salvatore Vitale, and Jeff Kissel. Impact of calibration uncertainties on hubble constant measurements from gravitational-wave sources, 2022.
- [68] Ethan Payne, Colm Talbot, Paul D. Lasky, Eric Thrane, and Jeffrey S. Kissel. Gravitational-wave astronomy with a physical calibration model. *Phys. Rev. D*, 102: 122004, Dec 2020. doi: 10.1103/PhysRevD.102.122004. URL <https://link.aps.org/doi/10.1103/PhysRevD.102.122004>.

- [69] Salvatore Vitale, Carl-Johan Haster, Ling Sun, Ben Farr, Evan Goetz, Jeff Kissel, and Craig Cahillane. Physical approach to the marginalization of ligo calibration uncertainties. *Phys. Rev. D*, 103:063016, Mar 2021. doi: 10.1103/PhysRevD.103.063016. URL <https://link.aps.org/doi/10.1103/PhysRevD.103.063016>.
- [70] Craig Cahillane, Joe Betzwieser, Duncan A. Brown, Evan Goetz, Evan D. Hall, Ki-wamu Izumi, Shivaraj Kandhasamy, Sudarshan Karki, Jeff S. Kissel, Greg Mendell, Richard L. Savage, Darkhan Tuyenbayev, Alex Urban, Aaron Viets, Madeline Wade, and Alan J. Weinstein. Calibration uncertainty for advanced ligo's first and second observing runs. *Phys. Rev. D*, 96:102001, Nov 2017. doi: 10.1103/PhysRevD.96.102001. URL <https://link.aps.org/doi/10.1103/PhysRevD.96.102001>.
- [71] Ling Sun, Evan Goetz, Jeffrey S. Kissel, Joseph Betzwieser, Sudarshan Karki, Dripta Bhattacharjee, Pep B. Covas, Laurence E. H. Datrier, Shivaraj Kandhasamy, Yannick K. Lecoeuche, Gregory Mendell, Timesh Mistry, Ethan Payne, Richard L. Savage, Aaron Viets, Madeline Wade, Alan J. Weinstein, Stuart Aston, Craig Cahillane, Jennifer C. Driggers, Sheila E. Dwyer, and Alexander Urban. Characterization of systematic error in advanced ligo calibration in the second half of o3, 2021.
- [72] F Acernese et al. Calibration of advanced virgo and reconstruction of the detector strain $h(t)$ during the observing run o3. *Classical and Quantum Gravity*, 39(4):045006, January 2022. ISSN 1361-6382. doi: 10.1088/1361-6382/ac3c8e. URL <http://dx.doi.org/10.1088/1361-6382/ac3c8e>.
- [73] Matthew Giesler, Maximiliano Isi, Mark A. Scheel, and Saul A. Teukolsky. Black hole ringdown: The importance of overtones. *Physical Review X*, 9(4), dec 2019. doi: 10.1103/physrevx.9.041060. URL <https://doi.org/10.1103%2Fphysrevx.9.041060>.