Calibration uncertainty for Advanced LIGO's first and second observing runs

Craig Cahillane,[1,*] Joe Betzwieser,[2] Duncan A. Brown,[3] Evan Goetz,[4] Evan D. Hall,[5] Kiwamu Izumi,[4]
Shivaraj Kandhasamy,[2] Sudarshan Karki,[6] Jeff S. Kissel,[4] Greg Mendell,[4] Richard L. Savage,[4]
Darkhan Tuyenbayev,[7] Alex Urban,[1] Aaron Viets,[8] Madeline Wade,[9] and Alan J. Weinstein[1]

[1]*LIGO Laboratory, California Institute of Technology, Pasadena, California 91125, USA*
[2]*LIGO Livingston Observatory, Livingston, Louisiana 70803, USA*
[3]*Syracuse University, Syracuse, New York 13244, USA*
[4]*LIGO Hanford Observatory, Richland, Washington 99352, USA*
[5]*LIGO Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[6]*University of Oregon, Eugene, Oregon 97403, USA*
[7]*University of Texas Rio Grande Valley, Brownsville, Texas 78520, USA*
[8]*University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201, USA*
[9]*Kenyon College, Gambier, Ohio 43022, USA*
(Received 11 August 2017; published 8 November 2017)

Calibration of the Advanced LIGO detectors is the quantification of the detectors' response to gravitational waves. Gravitational waves incident on the detectors cause phase shifts in the interferometer laser light which are read out as intensity fluctuations at the detector output. Understanding this detector response to gravitational waves is crucial to producing accurate and precise gravitational wave strain data. Estimates of binary black hole and neutron star parameters and tests of general relativity require well-calibrated data, as miscalibrations will lead to biased results. We describe the method of producing calibration uncertainty estimates for both LIGO detectors in the first and second observing runs.

I. INTRODUCTION

The Laser Interferometer Gravitational-wave Observatory (LIGO), with its twin detectors in Hanford, Washington (H1) and Livingston, Louisiana (L1) has directly observed transient gravitational wave (GW) signals [1–3]. These events are consistent with binary black hole coalescences [4], whose detections have ushered in a new era of gravitational wave astronomy. Observing Run 1 (O1) saw the first Advanced LIGO GW strain data taken between September 18th, 2015 through January 12th, 2016. Observing Run 2 (O2) started on November 30th, 2016, and ended August 25th, 2017.

GW signals are extremely rich sources of information from previously unexplored astrophysical phenomena. The uncertainty in the estimated amplitude and phase of the GW directly impacts the astrophysics we can learn from both transient and long-duration signals. For compact binary coalescence GW signals, estimates of the progenitor masses, spins, luminosity distance, orbital plane inclination, final mass, and sky location are derived from the detected waveforms, and each are potentially limited by calibration accuracy [5]. The rate at which such systems coalesce in the universe can be drawn from detected events, but as the number of observations increases, rate estimates

will become limited by strain amplitude uncertainty [3,6]. Testing general relativity has begun with the first detections [3,7], but as the detectors' sensitivity improves and there are more high signal-to-noise ratio events, calibration uncertainty will limit our test results, and calibration error will bias our test results [8,9]. Upper limits and observations of sources of continuous gravitational waves, such as rapidly rotating neutron stars, depend on calibration uncertainty [10,11]. Upper limits and observations of the GW stochastic background of unresolvable sources depend on the amplitude calibration uncertainty [12,13]. Using many GW detections to refine estimates of the Hubble constant will be fundamentally limited by calibration uncertainty [14,15].

The total calibration uncertainty budget consists of statistical uncertainty and systematic error. Statistical uncertainty is the intrinsic uncertainty associated with measurements. Systematic error is the bias quantifying the difference between model and measurement. These quantities will be further defined in Sec. II.

This paper presents a refined method of producing calibrated GW strain data error and uncertainty budgets, discusses the error and uncertainty budget's evolution over time throughout the first two observing runs, and highlights the results relevant to the subsequent transient detections. The work here builds on previous work presented for the detection of GW150914 [16].

*ccahilla@caltech.edu

Section II reviews the detector fundamentals, how the strain time-series $h(t)$ is constructed from the detector output, the model of the detector response used to construct that estimate, and measurements supporting that model. Section III explains how the statistical uncertainty and systematic error budget on the strain time-series is constructed. Section IV shows error and uncertainty budgets for observational data sets to date, with focus on the times of the GW detections. Section V presents ideas for future work to further improve the calibration uncertainty. Section VI discusses the implications of the calibration uncertainty results.

II. DETECTORS

The Advanced LIGO detectors are Michelson interferometers whose arms are enhanced with 4 km long Fabry-Pérot resonant cavities. The cavities are filled with continuous carrier laser light from an Nd:YAG 1064 nm laser. Additional recycling cavities at the Michelson's input and output ports further improve the detector sensitivity to GWs [17,18].

A gravitational wave incident on the detector modifies the distance between the input and end mirrors of the arm cavities. This causes an apparent differential change in length of two arms, ΔL_{free}, relative to the average length of the arms, L. Differential arm (DARM) displacement is defined as

$$\Delta L_{\text{free}} = \Delta L_{\text{x}} - \Delta L_{\text{y}} \equiv hL, \qquad (1)$$

where ΔL_{x} and ΔL_{y} are changes in the X and Y arm lengths, and h is the detector's reconstructed GW strain signal and contains both the desired astrophysical information and unwanted noise. We report precision calibration for gravitational waves in the frequency band between 10 and 5000 Hz.

DARM displacement generates laser power fluctuations on the antisymmetric port photodetector, shown by the GW readout port in Figure 1 [19]. When the interferometer DARM degree of freedom is held on resonance, light in the antisymmetric port destructively interferes, and laser power fluctuations on the antisymmetric photodetector are quadratic in strain. "On resonance" means the round-trip DARM length is an integer number of laser wavelengths. A small length offset is introduced such that the interference of the laser beams from two cavities is not completely destructive in the direction of the antisymmetric photodetector. With the offset, laser power fluctuations are approximately proportional to strain, allowing us to directly read out incident GW strain from the laser fluctuations.

The arm cavity mirrors are suspended from multistage cascading pendula [20,21] and active seismic isolation systems [22] to suppress DARM displacement from ground
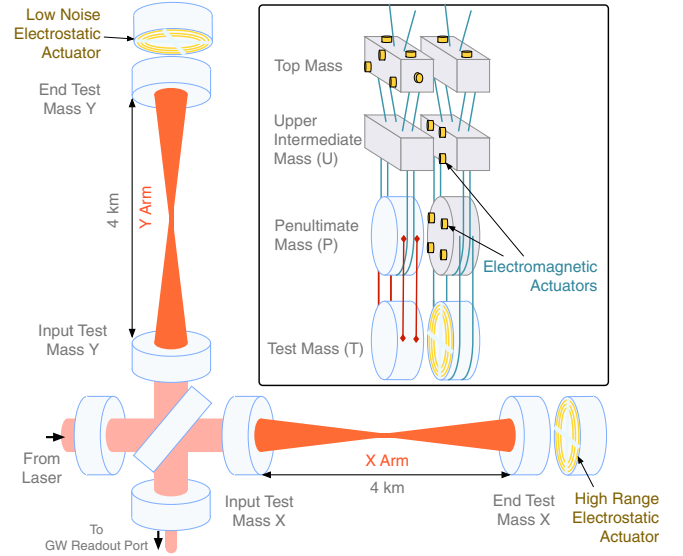

FIG. 1. Simplified Advanced LIGO interferometer layout. Laser light enters the interferometer at the lower left through a power-recycling mirror, and is split by a 50/50 beamsplitter. Two 4 km long Fabry-Pérot resonating arm cavities are formed from four highly reflective test masses. Laser light builds up in the cavities, reaching about 100 kW of laser power in O1 and O2. A signal-recycling mirror between the beamsplitter and GW readout photodetector modifies the detector response to increase the detector bandwidth. Inset: one of the quadruple pendulum suspension systems which holds each of the four test masses is shown.

motion and other force noise. Still, DARM displacement must be further controlled to hold the interferometer on resonance. This requires a global arm length control system, with open loop transfer function G, to suppress the free displacement ΔL_{free} to a smaller residual displacement via actuators present on the cascaded pendula [23] to a residual differential arm length ΔL_{res},

$$\Delta L_{\text{res}} = \Delta L_{\text{free}}/[1 + G]. \qquad (2)$$

We define three independently quantifiable transfer functions of the DARM control loop, shown schematically in Fig. 2. The sensing function $C = d_{\text{err}}/\Delta L_{\text{res}}$ defines the measured interferometric laser power response to DARM displacement and the digitization process of the power fluctuations to form the digital error signal d_{err}. Digital filters $D = d_{\text{ctrl}}/d_{\text{err}}$ convert the loop error signal to the loop control signal. The actuation function $A = \Delta L_{\text{ctrl}}/d_{\text{ctrl}}$ generates force on the optical cavity pendula to largely cancel any detected DARM displacement within the DARM loop bandwidth. All transfer functions are complex-valued functions of frequency, with quantifiable magnitude and phase. The digital filters D shape the DARM loop frequency response and are known to negligible uncertainty. The DARM loop transfer functions
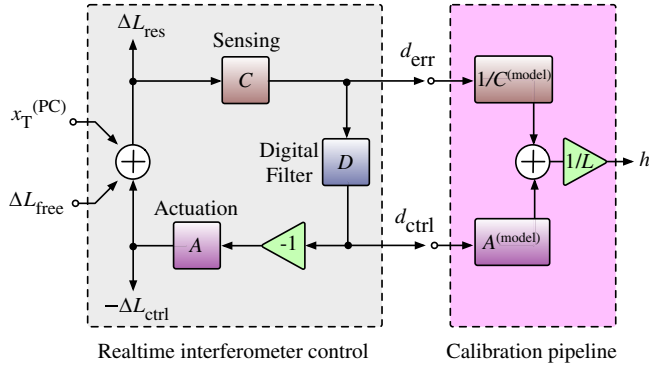
FIG. 2. DARM control loop and calibration procedure. The DARM control loop is shown in the grey box on the left. The sensing plant C produces the detector output d_{err} in linear response to residual differential arm motion ΔL_{res}. The digital filters D are known filters conditioning the detector output d_{err} into a control signal d_{ctrl}. The actuation plant A takes the control signal d_{ctrl} and actuates on the optics by ΔL_{ctrl} to maintain cavity resonance. The pink box on the right shows the calibration procedure, consisting of an inverse sensing model $1/C^{(\mathrm{model})}$ and actuation model $A^{(\mathrm{model})}$. The output of the calibration pipeline is GW strain data $h(t)$.

C and A must be measured and modeled in the frequency domain between 5 and 5000 Hz. Both C and A contribute to the total calibration uncertainty budget.

The digital signals d_{err} and d_{ctrl} are digitally filtered to form a time-series estimate of the GW strain $h(t)$ used for astrophysical searches. The digital filters applied to d_{err} and d_{ctrl} are constructed from models of the sensing function $C^{(\mathrm{model})}$ and actuation function $A^{(\mathrm{model})}$:

$$h = \frac{1}{L}\left[\frac{1}{C^{(\mathrm{model})}} * d_{\mathrm{err}} + A^{(\mathrm{model})} * d_{\mathrm{ctrl}}\right], \qquad (3)$$

where $*$ indicates convolution in the time domain, or multiplication in the frequency domain, and L is the length of the interferometer arms, known to negligible uncertainty. The accuracy and precision of the models $C^{(\mathrm{model})}$ and $A^{(\mathrm{model})}$ define the systematic error and statistical uncertainty in the estimated time series $h(t)$.

We define a transfer function called the response function R,

$$h = R * d_{\mathrm{err}} = \frac{1}{L}\left(\frac{1+G}{C}\right)d_{\mathrm{err}} \qquad (4)$$

where the DARM open loop gain $G = C * D * A$. Equation (4) illustrates that in the frequency domain, response function error δR is equivalent to the GW strain data error δh and response function uncertainty σ_R is equivalent to the GW strain data uncertainty σ_h.

Throughout this paper, the response error and uncertainty relative to the calibration pipeline model $R^{(\mathrm{model})}$ are quantified as a function of frequency f with time dependence t:

$$\frac{\delta R(f,t)}{R^{(\mathrm{model})}} = \frac{\delta h(f,t)}{h}, \qquad \frac{\sigma_R(f,t)}{R^{(\mathrm{model})}} = \frac{\sigma_h(f,t)}{h}. \qquad (5)$$

The DARM loop transfer functions C and A are measured and modeled in the frequency domain. Additionally, the values of C and A can drift slowly over time, giving functions of frequency that vary in time $C(f,t)$ and $A(f,t)$. However, our online calibration pipeline digital filters $1/C^{(\mathrm{model})}$ and $A^{(\mathrm{model})}$ are not perfect representations of our understanding of the interferometer. This leads to known systematic errors in our $h(t)$ reconstruction, governed by the sensing and actuation systematic errors $\delta C(f,t)$ and $\delta A(f,t)$. The systematic errors relative to $C^{(\mathrm{model})}$ and $A^{(\mathrm{model})}$ are quantified as

$$\frac{\delta C(f,t)}{C^{(\mathrm{model})}} = \frac{C(f,t)}{C^{(\mathrm{model})}} - 1, \qquad \frac{\delta A(f,t)}{A^{(\mathrm{model})}} = \frac{A(f,t)}{A^{(\mathrm{model})}} - 1, \qquad (6)$$

where $C(f,t)$ and $A(f,t)$ represent the measured sensing and actuation transfer functions.

Systematic errors δC and δA propagate forward to the relative response function systematic error $\delta R/R^{(\mathrm{model})}$:

$$\begin{aligned}
\frac{\delta R(f,t)}{R^{(\mathrm{model})}} &= \frac{R(f,t)}{R^{(\mathrm{model})}} - 1 \\
&= \left(\frac{1+G(f,t)}{C(f,t)}\right)\Big/\left(\frac{1+G^{(\mathrm{model})}}{C^{(\mathrm{model})}}\right) - 1 \\
&= \frac{\left(G^{(\mathrm{model})}\frac{\delta A(f,t)}{A^{(\mathrm{model})}} - \frac{\delta C(f,t)/C^{(\mathrm{model})}}{1+\delta C(f,t)/C^{(\mathrm{model})}}\right)}{1 + G^{(\mathrm{model})}}.
\end{aligned} \qquad (7)$$

The response function uncertainty $\sigma_R(f,t)$ is in general a 2×2 matrix to represent uncertainty in the complex plane. The off-diagonal terms are capable of capturing covariance between the two basis vectors. In this paper, we will be using Bendat and Piersol relative magnitude uncertainty and absolute phase uncertainty, seen in Eq. (18) [24]. All uncertainties will be propagated in the relative magnitude and absolute phase basis with no covariance, as given from Bendat and Piersol.

In previous calibration uncertainty work [16], covariance between actuation stages was found to be non-negligible and was included in the uncertainty budget. In this work, improved measurement techniques broke this covariance, rendering its effect negligible. There is also no covariance between the sensing function and any actuation stage.

The 2×2 uncertainty matrices $\sigma_C(f,t)$ and $\sigma_A(f,t)$ propagate to the relative response function uncertainty $\sigma_R/R^{(\text{model})}$:

$$\frac{\sigma_R(f,t)}{R^{(\text{model})}} = \frac{1}{R^{(\text{model})}} \sqrt{\left(\frac{\partial R}{\partial C}\right)^2 \sigma_C^2 + \left(\frac{\partial R}{\partial A}\right)^2 \sigma_A^2} = \frac{C^{(\text{model})}}{1 + G^{(\text{model})}} \sqrt{\frac{1}{C(f,t)^4} \sigma_C(f,t)^2 + D(f)^2 \sigma_A(f,t)^2}. \tag{8}$$

Together $\delta R/R^{(\text{model})}$ and $\sigma_R/R^{(\text{model})}$ make up the entire calibration error and uncertainty budget.

A. Sensing model

The core of sensing function model $C^{(\text{model})}$ is the interferometric transfer function from DARM displacement to laser power on the antisymmetric port photodetectors. The photodetectors generate photocurrent in response, which is then run through a transimpedence amplifier, whitening filters, anti-aliasing filters, and an analog-to-digital converter to produce DARM error counts d_{err}. The transfer function from DARM displacement in picometers to milliamps of photodetector output current is shown in Fig. 3.

Advanced LIGO's dual-recycled, Fabry Perot Michelson detectors operate in the "resonant signal extraction" configuration: the signal recycling mirror is purposefully detuned from resonance to increase the bandwidth of the detector. When detuned exactly to 90 degrees, as designed, the sensing model may be approximated by as a single pole system. This model was used for estimating the uncertainty and error of GW150914 [16]. However, measurements

have revealed both detectors' signal recycling cavities are slightly offset from 90 degrees detuning, inducing an optical antispring which reduces the displacement response at low frequencies.

We employ another approximation to the interferometric response in the sensing model, $C^{(\text{model})}(f,t,\vec{\lambda}_C)$ [25–27]. The model contains all terms from [16] but additionally includes an optical antispring term defined by f_S and Q_S, the optical antispring pole frequency and quality factor,

$$C^{(\text{model})}(f,t,\vec{\lambda}_C) = \frac{\kappa_C(t) H_C}{1 + if/f_{CC}} C_R(f) e^{-2\pi i f \tau_C}$$
$$\times \frac{f^2}{f^2 + f_S^2 - if f_S Q_S^{-1}} \tag{9}$$

The optical gain H_C defines the scale of the sensing function in units of error signal counts d_{err} per DARM displacement in meters. It collects all individual scale factors from the interferometric response in watts/meter, the optical efficiency of the photodiodes in amps/watt,
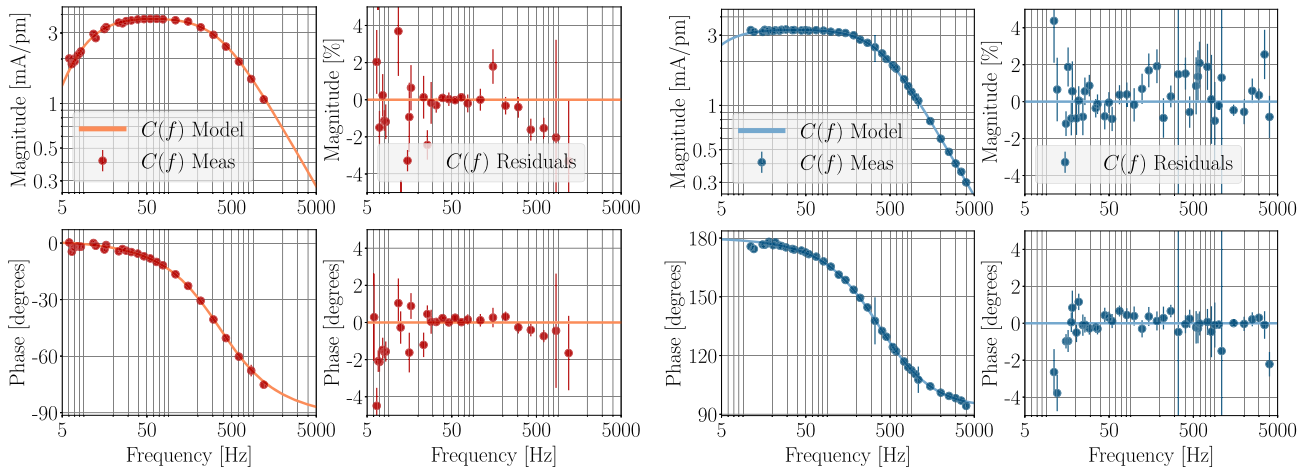

FIG. 3.   Sensing measurements $C^{(\text{meas})}(f)$, sensing model $C^{(\text{model})}(f,t,\vec{\lambda}_C)$, and their residuals $\delta C(f,t)/C^{(\text{model})}(f,t,\vec{\lambda}_C)$. The H1 sensing reference measurement from January 4th, 2017 is shown in the four panels on the left in red. The L1 sensing reference measurement from November 26th, 2016 is in the four right panels in blue. The first and third columns are the Bode plots of C, while the second and fourth columns are the residuals. The model parameters $\vec{\lambda}_C$ were found via an MCMC. Physically, the magnitude Bode plots represent how many milliamps of current are generated at our transimpedence photodetector per picometer of differential arm motion from 5 to 5000 Hz. The drop in sensitivity at low frequencies shows the effect of detuning at both detectors. The 180 degree phase difference between H1 and L1 is a sign convention difference between the detectors.

through the transimpedance analog electronics in volts/amp, and recorded in analog-to-digital converter counts/volt. The time dependent scale factor $\kappa_C(t)$, initially set to 1 at the reference time, accounts for slow changes in H_C as the detector's interferometric response evolves due to mirror alignment drift and thermal loading [28]. The coupled cavity pole f_{CC} defines the detector bandwidth. The sensing time delay τ_C includes the light travel time over the length of the arms, computational delay in the digital acquisition system, and a compensation for the exclusion of additional, high-frequency response of the Fabry-Pérot arm cavities beyond the single coupled cavity pole model [29]. The model time delay τ_C is 77.6 μs for both detectors. The frequency dependent function $C_R(f)$ is the response of the digital acquisition system, including transimpedance electronics and anti-aliasing filters, all known to negligible uncertainty. The parameter vector $\vec{\lambda}_C$ defines a set of the time-independent, reference sensing parameters whose values are fit to non-negligible precision: $\vec{\lambda}_C = (H_C, f_{CC}, \delta\tau_C, f_S, Q_S^{-1})^T$, where $\delta\tau_C$ is a correction time delay factor on the model time delay τ_C. The nominal values of the reference sensing parameters $\vec{\lambda}_C$ for each detector are found in Table I.

Our model of the sensing function $C^{(\text{model})}(f, t, \vec{\lambda}_C)$ is an approximation. The true detector sensing function changes over time and deviates from the sensing model at high frequencies. The sensing model dynamically corrects for $\kappa_C(t)$ with real-time measurement. However, f_{CC},

f_S, and Q_S^{-1} are also changing in time, but are not corrected for in the model. At present, the time dependence in f_{CC} is included in the calibration uncertainty budget as a known systematic error, since it is tracked via real-time measurement but cannot yet be dynamically corrected for in the model. The time dependence in f_S and Q_S^{-1} results in expanded uncertainty at low frequency. The total systematic error in the sensing function, $\delta C(f, t)$, is

$$\frac{\delta C(f, t)}{C^{(\text{model})}} = \left(\frac{1 + if/f_{CC}}{1 + if/f_{CC}(t)}\right)\frac{\delta C^{GP}(f)}{C^{(\text{model})}}e^{-2\pi i f \delta\tau_C}. \quad (10)$$

The first term is the explicit correction for time dependence of the coupled cavity pole, $f_{CC}(t)$. A correction time delay factor $\delta\tau_C$ modifies the original time delay τ_C included in the model. Further systematic errors may originate from the uncorrected time dependence of f_S and Q_S^{-1} or additional unknown systematic errors. Any remaining frequency dependent systematic errors are covered by a Gaussian Process regression $\delta C^{GP}(f)$. Quantifying errors $\delta C^{GP}(f)$ is explained further in Sec. III.

B. Actuation model

The Advanced LIGO test masses are suspended via quadruple cascaded pendula. Each suspension stage has independent actuators, as shown in Fig. 1. The control signal, d_{ctrl}, is digitally distributed as a function of frequency to each stage's actuators via a digital-to-analog converter and signal processing electronics to create the control displacement, ΔL_{ctrl}. The distribution filters are designed taking into account all actuators' authority to displace the test mass. On the upper intermediate and penultimate stage, the digital-to-analog converter drives electromagnets on the reaction stage creating a force on magnets attached to the suspended stage. On the test mass stage, the digital-to-analog converter drives an electrostatic system which creates a force, quadratic in the applied potential, via dipole-dipole interactions between the test mass and a pattern of electrodes on the reaction mass (see Fig. 1). With a large bias voltage and low control voltage, the requested actuation forces on the electrostatic system are in the linear regime. Any time-dependent change to the slope of the linear response due to quadratic terms is measured continuously, as described below.

The sum of the paths the digital control signal, d_{ctrl}, takes through each stage to displace the test mass, ΔL_{ctrl}, makes up our total actuation model:

$$\begin{aligned}A^{(\text{model})}(f, t, \vec{\lambda}_A) = &[\kappa_T(t)F_T(f)H_TA_T(f)\\&+ \kappa_{PU}(t)(F_P(f)H_PA_P(f)\\&+ F_U(f)H_UA_U(f))]e^{-2\pi i f \tau_A}\end{aligned} \quad (11)$$

TABLE I. Sensing function model parameters $\vec{\lambda}_C$ MCMC fit values and uncertainties for H1 (top) and L1 (bottom). The fits were performed on H1's January 4th, 2017 reference measurement and L1's November 26th, 2016 reference measurement. The model corresponding to these parameters can be seen in Fig. 3. The corner plot showing the MCMC results from the H1 reference measurement is shown in Fig. 6.

H1 Parameters	Variable	Value$^{+1\sigma}_{-1\sigma}$	Units
Optical Gain	H_C	$3.834^{+0.003}_{-0.003}$	mA/pm
Coupled Cavity Pole	f_{CC}	360^{+2}_{-2}	Hz
Time Delay	$\delta\tau_C$	$0.6^{+1.3}_{-1.3}$	μs
Optical Spring Frequency	f_S	$6.87^{+0.03}_{-0.03}$	Hz
Optical Spring Inverse Q	Q_S^{-1}	$0.034^{+0.004}_{-0.004}$	none

L1 Parameters	Variable	Value$^{+1\sigma}_{-1\sigma}$	Units
Optical Gain	H_C	$3.288^{+0.007}_{-0.007}$	mA/pm
Coupled Cavity Pole	f_{CC}	$369.5^{+1.0}_{-0.9}$	Hz
Time Delay	$\delta\tau_C$	$-0.84^{+0.13}_{-0.13}$	μs
Optical Spring Frequency	f_S	$2.6^{+0.2}_{-0.2}$	Hz
Optical Spring Inverse Q	Q_S^{-1}	$0.005^{+0.009}_{-0.004}$	none

where U, P, and T represent the three stages used for control; the upper-intermediate, penultimate, and test mass stages, respectively. Each stage is composed of the normalized electromechanical frequency response of the pendulum and its actuators, $A_i(f)$, the digital distribution filter, $F_i(f)$, a dimensionful scale factor, H_i, and an overall digital delay, τ_A, defined by the common computational delay from each stage. The model time delay τ_A is 45 μs for L1 and 61 μs for H1. $\kappa_{PU}(t)$ is the time dependence of the penultimate and upper intermediate scale factor, and $\kappa_T(t)$ is the time dependence of the test mass scale factor [28].

The penultimate and upper intermediate scale factor $\kappa_{PU}(t)$ is not expected to vary much over time, as it represents the change in the electromagnetic coil actuators' strength. The test mass scale factor $\kappa_T(t)$ does vary significantly over time as the electric charge on the test mass builds up, changing the actuation strength of the electrostatic drive.

The reference scale factor for each stage, H_i, collects scale factors from that of the digital-to-analog converter in volts/count, each stage's drive electronics in amps/volt or volts/volt, the actuator itself in newtons/amp or newtons/volt depending on the stage, and the stiffness of the suspension in meters / newton. Time delay correction factors for each stage $\delta\tau_i$ are extracted from measurements as stage-specific corrections to the overall actuation delay τ_A. The electromechanical transfer functions, A_i, for each stage are independently measured and included in the model with negligible uncertainty. Remaining scale factor and delay parameters dominate the actuation function uncertainty, and are thus collected in the set of reference parameters $\vec{\lambda}_A = (H_U, \delta\tau_U, H_P, \delta\tau_P, H_T, \delta\tau_T)^T$. The values of these reference parameters λ_A are found in Table II.

The digital filters, F_i, are known *a priori*, and time-dependent corrections κ_{PU} and κ_T are dynamically corrected for when estimating $h(t)$. The remaining components of the actuation stage model, $[H_iA_i]^{(\text{model})}(f,\vec{\lambda}_i)$, may contain systematic errors. We allow for and quantify systematic errors in each actuation stage as

$$\frac{\delta A_i(f)}{A_i^{(\text{model})}} = \frac{\delta A_i^{GP}(f)}{A_i^{(\text{model})}} e^{-2\pi i f \delta\tau_i} \tag{12}$$

where $\delta\tau_i$ is a time delay phase error on each stage, and $\delta A_i^{GP}(f)$ is the systematic error in scale or frequency dependence from the Gaussian process regression done on each stage's measurement residuals. Systematic error calculations are explained fully in Sec. III, subsection III B.

C. Measurements

In this section, we first describe our fundamental displacement reference, the radiation pressure actuator.

TABLE II. Actuation function model parameters $\vec{\lambda}_A$ MCMC fit values and uncertainties for H1 (top) and L1 (bottom). The fits were performed on H1's January 4th, 2017 reference measurements and L1's November 26th, 2016 reference measurement. The models corresponding to these parameters can be see in Fig. 4. To get from Newtons/count units in this table to meters/count in Fig. 4, we multiply by the suspension models which have units of meters/Newton and are known to negligible uncertainty.

H1 Parameters	Variable	Value$^{+1\sigma}_{-1\sigma}$	Units
Upper Intermediate Gain	H_U	$8.205^{+0.004}_{-0.004} \times 10^{-8}$	N/cts
Upper Intermediate Delay	$\delta\tau_U$	57^{+45}_{-46}	μs
Penultimate Gain	H_P	$6.768^{+0.002}_{-0.002} \times 10^{-10}$	N/cts
Penultimate Delay	$\delta\tau_P$	$0.4^{+0.6}_{-0.6}$	μs
Test Mass Gain	H_T	$4.3573^{+0.0008}_{-0.0008} \times 10^{-12}$	N/cts
Test Mass Delay	$\delta\tau_T$	$2.8^{+0.4}_{-0.4}$	μs

L1 Parameters	Variable	Value$^{+1\sigma}_{-1\sigma}$	Units
Upper Intermediate Gain	H_U	$7.24^{+0.03}_{-0.03} \times 10^{-8}$	N/cts
Upper Intermediate Delay	$\delta\tau_U$	102^{+56}_{-56}	μs
Penultimate Gain	H_P	$6.41^{+0.02}_{-0.02} \times 10^{-10}$	N/cts
Penultimate Delay	$\delta\tau_P$	$-8.7^{+6.2}_{-6.1}$	μs
Test Mass Gain	H_T	$2.513^{+0.004}_{-0.004} \times 10^{-12}$	N/cts
Test Mass Delay	$\delta\tau_T$	$-4.5^{+1.4}_{-1.4}$	μs

Then we explore how that reference is used to measure the detector's response to DARM motion, or in other words, calibrate the detector.

The DARM model functions $C(f,t)$ and $A(f,t)$ are measured from swept sine transfer functions of the DARM control loop. A swept sine transfer function is a collection of single frequency excitations applied in successive steps across the relevant frequency band of the detector. The cross-correlation of actuator excitation against detector response during the excitation forms the transfer function.

The swept sine transfer functions are then manipulated to give transfer function measurements of each of the actuation stages and the sensing function. Measurements of the detectors' DARM control loops require the detectors to be running at low-noise observation sensitivity. Once a full suite of reference measurements is taken, the complete response of the detector to GWs can be estimated.

1. Radiation pressure actuator

Two 1047 nm auxiliary laser systems known as photon calibrators (PCAL) displace each end test mass via radiation pressure [30]. The PCAL serves as a reference actuator on the test mass controlling

the DARM loop. The displacement of the test mass caused by the photon calibrator $x_T^{(PC)}$ is several orders of magnitude larger than the nominal displacement noise of the detector ΔL_{free} with integration time of 1 minute, which allows for high-precision characterization of the global control system and detector readout (See Fig. 2). Acousto-optic modulators (AOMs) are used to modulate the laser power incident on the test masses with arbitrary waveforms. The power incident on the test masses is recorded via two photodiodes coupled to integrating spheres, one after the AOM before transmission onto the test mass, the other upon reflection off of the test mass. Each photodiode's readout is then digitally recast as a displacement, $x_T^{(PC)}$, which is the amount of PCAL-induced displacement contributing to ΔL_{free}. The reflection photodiode was used for reference in all measurements described below.

The PCAL laser can introduce elastic deformations on the test mass, which can affect the calibration accuracy above 1 kHz. Elastic deformation can be largely mitigated through the use of two beams symmetrically displaced from the center of the test mass [30]. The uncertainty budget does not include error from elastic deformation, assuming this effect is negligible up to 5 kHz. The full suspension dynamics are incorporated into the transfer function from the PCAL power modulation to the test mass length modulation, giving an accurate frequency response at and below the suspension resonant frequency.

We are sometimes susceptible to systematic errors from PCAL "clipping" where the photon calibrator laser slightly misses the receiving photodiode, causing miscalibrations. Fortunately, PCAL clipping systematic errors are quantifiable and included in the error budget. The relative PCAL actuation strength correction factor, $H_{\text{PCAL}}(t)$, tracks the actuation strength of the PCAL over time. $H_{\text{PCAL}}(t)$ has a value of 1 during times of no clipping, and a value less than 1 during times of clipping. $H_{\text{PCAL}}(t)$ has a relative uncertainty of 0.79% over all time. This will affect our total calibration uncertainty budget directly in Section III C. More on PCAL clipping is discussed in Sec. IV. Further details of the PCAL and the composition of their uncertainty can be found in [30].

Checks of gross systematic errors in the photon calibrator system have been performed using other, less precise displacement (or equivalent there-of) references and found agreement with the PCAL to within 10% [16].

2. Measurement techniques

To measure the PCAL to DARM transfer function, a known photon calibrator sine wave excitation $x_T^{(PC)}$ is

applied to the detector while the DARM error signal d_{err} is recorded. This excitation is suppressed by the DARM control loop, forming the transfer function

$$\frac{d_{\text{err}}(f)}{x_T^{(PC)}(f)} = \frac{C(f)}{1 + G(f)}. \qquad (13)$$

The measurement suite is a collection of discrete sine waves swept over the frequency range 5 Hz $< f <$ 1 kHz. The closed loop suppression, $1/[1 + G(f)]$, is then measured independently with the standard in-loop suspension actuators at the same frequencies as Equation (13). During times of clipping, we underestimate the excitation $x_T^{(PC)}$ by the relative actuation strength $H_{\text{PCAL}}(t)$, and must divide $x_T^{(PC)}$ by $H_{\text{PCAL}}(t)$ to correct for this. The measured sensing function is then constructed as a function of frequency:

$$C^{(\text{meas})}(f) = H_{\text{PCAL}}(t)[1 + G(f)]\frac{d_{\text{err}}(f)}{x_T^{(PC)}(f)}. \qquad (14)$$

Above 1 kHz, the photon calibrator's signal-to-noise ratio and actuation strength are low. In this region, the open loop gain $G(f)$ is negligible, so

$$\frac{d_{\text{err}}(f)}{x_T^{(PC)}(f)} \approx \frac{C^{(\text{meas})}(f)}{H_{\text{PCAL}}(t)}, \qquad f > 1 \text{ kHz} \qquad (15)$$

We obtain the sensing function at high frequency by performing a long-duration swept sine transfer function measurement. Each single frequency is driven for many hours, and the response is compensated for time dependence using only $\kappa_C(t)$.

To measure the three actuation stages, similar swept sine excitations, $x_i(f)$, are applied to each stage at points upstream of the known distribution filters, $F_i(f)$, such that the detector readout measures

$$\frac{d_{\text{err}}(f)}{x_i(f)} = \frac{H_i A_i(f)C(f)}{1 + G(f)} \qquad (16)$$

where the index i indicates either the upper intermediate U, penultimate P, or test mass T stages. These excitations are then compared to an excitation from the photon calibrator to isolate each actuation plant, as in Eq. (13), to form

$$[H_i A_i(f)]^{(\text{meas})} = \frac{1}{H_{\text{PCAL}}(t)}\frac{x_T^{(PC)}(f)}{d_{\text{err}}(f)}\frac{d_{\text{err}}(f)}{x_i(f)}. \qquad (17)$$

The relative magnitude uncertainty and absolute phase uncertainty in a transfer function swept sine measurement point is calculated by Bendat and Piersol [24]:

$$\sigma^{(\mathrm{meas})}(f) = \sqrt{\frac{1 - \gamma^2(f)}{2N_{\mathrm{avg}}\gamma^2(f)}} \qquad (18)$$

where $\gamma^2(f)$ is the coherence between excitation and readout, and N_{avg} is the number of averages at each excitation frequency point.

To capture the time dependence of the calibration during a run, "calibration lines" are applied to the detectors during all observation times. A calibration line is a single-frequency excitation applied to the detector via the photon calibrator and suspension actuators. Using four calibration lines, we are able to capture changes in the detector calibration and partially correct for them in real time.

The calibration lines' response to the applied excitation is recorded in the detector readout d_{err}. These transfer functions are recast into each time dependent parameter, κ_T, κ_{PU}, κ_C, and f_{CC}. The calibration lines are driven with high signal-to-noise ratio such that the time-dependent parameter uncertainties are small relative to the parameter values. The calculation of the time-dependent parameters from calibration lines is derived in [28].

The statistical uncertainty in a time-dependent parameter $\sigma_{\kappa_i}(t)$, at any given time, t, is derived from the measured coherence of the calibration lines used to form them (see

Eq. (18), propagated as in [28]). That uncertainty is then used to form a distribution, with mean and standard deviation of $\kappa_i(t)$ and $\sigma_{\kappa_i}(t)$. The posteriors of this distribution are sampled and propagated through to the total uncertainty for that given time, as described below in Sec. III C. See Fig. 5 for an example result from this process for L1's $\kappa_C(t)$ at the time of GW170104.

III. UNCERTAINTY AND ERROR ESTIMATION

Our uncertainty budget is numerically evaluated by producing a large number of realizations of the response function. To do this, we first estimate the DARM model parameters using a Markov Chain Monte Carlo (MCMC) method. Next, we stack all measurement residuals and estimate any deviations from the model using a Gaussian process regression. Then, we sample our MCMC and regression results to form ten thousand resultant response functions. These response functions stacked form the calibration error and uncertainty budget.

A. DARM model parameter estimation

First, a measurement $\vec{d} = C^{(\mathrm{meas})}(f)$ or $A^{(\mathrm{meas})}(f)$ is obtained as described in Sec. II C. Next, the models $\vec{M} = C^{(\mathrm{model})}(f, t, \vec{\lambda}_C)$ or $A^{(\mathrm{model})}(f, t, \vec{\lambda}_A)$ are fit to the measurement by varying the model parameters $\vec{\lambda} = \vec{\lambda}_C$ or $\vec{\lambda}_A$ via a Markov Chain Monte Carlo (MCMC).

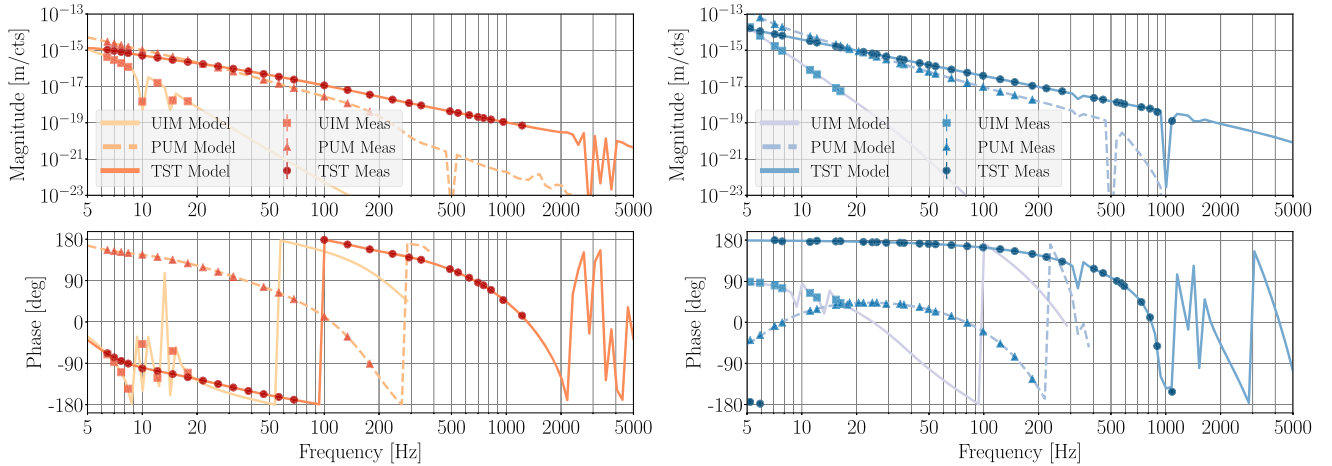An MCMC algorithm can quickly approximate the posterior probability distributions on the values of the model

FIG. 4. Actuation stage measurements $[H_i A_i(f)]^{(\mathrm{meas})}$ and models $[H_i A_i(f, \vec{\lambda}_i)]^{(\mathrm{model})}$. Each index i is one of the actuation stages U, P, or T. The H1 actuation reference measurements from January 4th, 2017 are shown in the two left plots in red. The L1 actuation reference measurements from November 26th, 2016 are in the two right panels in blue. The model parameters $\vec{\lambda}_A$ for $A_i(f, \vec{\lambda}_i)$ have been found via the MCMC method. The actuation strength magnitude is in units of meters per d_{ctrl} count. Notches seen in the magnitude plot are purposefully placed to avoid ringing up suspension violin modes at specific frequencies. Each stage's phase is sensible for frequencies at which each actuation stage dominates, but then rolls rapidly as it loses authority at high frequencies. For this reason, the UIM and PUM stage phase plots are cut off at 300 Hz and 400 Hz respectively.
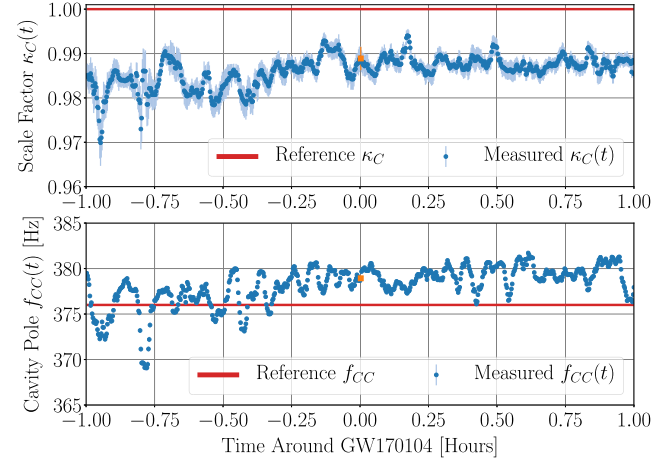
FIG. 5. Time dependent parameters in L1 around GW170104. The time series of the optical gain scale factor $\kappa_C(t)$ and coupled cavity pole $f_{CC}(t)$ in L1 for the two hours surrounding GW170104. The reference values are plotted as a red line, while the data is blue dots with uncertainty bars. The values at the time of GW170104 are plotted in orange. Changes in $\kappa_C(t)$ are updated in the reference sensing model $C^{(\text{model})}(f,t)$. Changes in $f_{CC}(t)$ are not reflected in the reference sensing model. This represents a systematic error in the calibration pipeline strain data $h(t)$.

FIG. 6. Posterior distribution on the H1 sensing parameters $\vec{\lambda}_C$. Each column represents one of the five sensing parameters: optical gain H_C, coupled cavity pole f_{CC}, time delay correction $\delta\tau_C$, optical spring f_S, and optical spring inverse quality factor Q_S^{-1}. Each point represents a sample in five dimensional parameter space. The diagonal plots represent the variance on each parameter, while the off-diagonal plots show the covariance of each parameter with another. The dashed vertical lines on the diagonal plots represent the median and 1σ values for each parameter.

parameters given a log likelihood function and assumed prior distribution. The log likelihood, $\log\mathcal{L}(\vec{M}|\vec{\lambda},\vec{d})$, is a simple least squares comparison between the model values $\vec{M}(\vec{\lambda})$ (where $\vec{M}$ represents $C^{(\text{model})}$ or $A^{(\text{model})}$) given model parameters $\vec{\lambda}$ (namely $\vec{\lambda}_C$ or $\vec{\lambda}_A$) and measurement data $\vec{d}$ (as described in Sec. II C). All initial parameter estimates in $\vec{\lambda}_C$ and $\vec{\lambda}_A$ were assumed to have flat prior distributions. The maximum *a posteriori* (MAP) values of the posterior distributions are taken as the best fit values. The ensemble of MCMC distributions are saved to be sampled for the total uncertainty budget in subsection III C.

The MCMC posteriors are found for both detector's frequency dependent models: $C^{(\text{model})}(f,t,\vec{\lambda}_C)$ and $A_i^{(\text{model})}(f,t,\vec{\lambda}_i)$. The best fit values are reported in Tables I and II. The plots of the model fits can be seen in Figs. 3 and 4. The one- and two-dimensional posterior distributions for the H1 sensing model parameters $\vec{\lambda}_C$ are shown in Fig. 6. The MCMCs were performed using the python EMCEE toolbox [31,32]. The plot was produced with the CORNER python plotting package [33].

B. Quantifying frequency dependent error and uncertainty

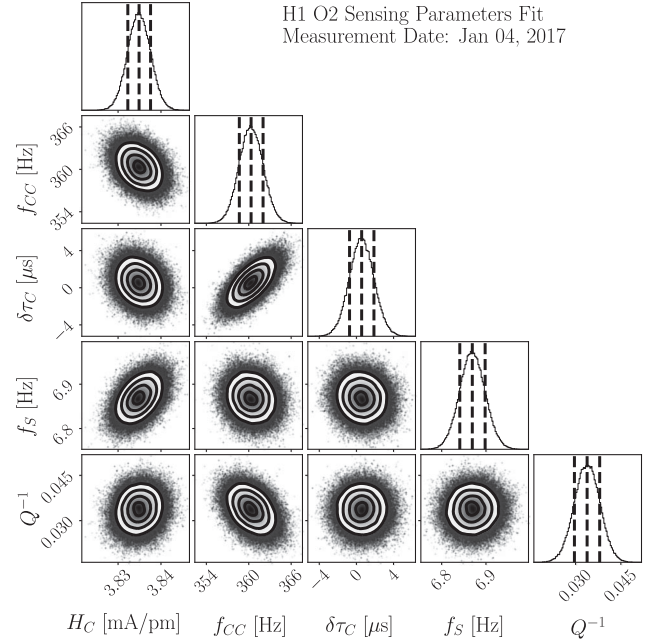Throughout observing runs, collections of detector measurements are taken regularly. Every measurement taken is run through the MCMC method as detailed in subsection III A. The measurement is then divided by its best fit DARM model to produce a residual, as seen in Eq. (6).

All of the residuals are gathered together into a collection of all measurements taken over the observing run. These residuals have all known systematic errors removed, but still contain information about unknown systematic errors. We create a distribution of functions that could describe this residual systematic error, then we incorporate this distribution into the calibration uncertainty budget. To accomplish this, we use a Gaussian process regression [34,35].

A Gaussian process is a method of producing distributions over random functions. The Gaussian process regression takes in data and a user-defined covariance matrix, tunes the covariance matrix hyperparameters to fit the given data, and outputs a posterior of potential function fits to the data. This allows an uncertainty budget to be produced for arbitrary frequencies, creating a continuous posterior distribution from discrete data.

From the resulting posterior distribution, we can extract a most probable fit function, known as the mean function.
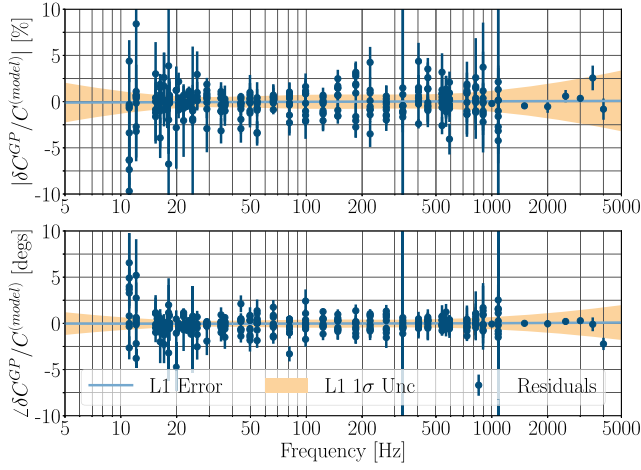
FIG. 7. Gaussian process regression of L1's sensing systematic error $\delta C^{GP}(f)$. The dark blue points are all the sensing measurement residuals, $\delta C/C^{(model)}(f, t, \vec{\lambda}_C)$, taken over the entire observation run. This includes the residuals from the L1 reference measurement in the far right plots of Fig. 3. The light blue line is the mean function representing systematic error. The light orange envelope is the 1σ uncertainty on the systematic error.

The mean function becomes the systematic error $\delta C^{GP}(f)$ and $\delta A_i^{GP}(f)$ in Eqs. (10) and (12). We can also draw frequency dependent uncertainties $\sigma_{\delta C}^{GP}$ and $\sigma_{\delta A_i}^{GP}$ on the systematic error. Posteriors representing $\sigma_{\delta C}^{GP}$ and $\sigma_{\delta A_i}^{GP}$ will be sampled for the total uncertainty budget in subsection III C.

Our Gaussian process regression trains on the residual data with the following covariance kernel

$$k(\log(f), \log(f')) = \gamma_1^2 + \log(f) \cdot \log(f')$$
$$+ (\gamma_2^2 + \log(f) \cdot \log(f'))^2$$
$$+ \gamma_3^2 \exp\left(-\frac{(\log(f) - \log(f'))^2}{2\ell^2}\right)$$

(19)

where $\{\gamma_1, \gamma_2, \gamma_3, \ell\}$ are the hyperparameters of the covariance kernel. The hyperparameters are tuned by the Gaussian process via gradient descent to best match the training data. This kernel assumes the detector plants' systematic error should be characterized in the log frequency domain, and that the error is relatively smooth and can be captured by a squared exponential and quadratic kernel.

An example collection of measurement residuals for the L1 detector's sensing function and the resulting Gaussian process regression is shown in Fig. 7. Here we show the same data from Fig. 3, but with additional measurements from the entire observation run.

C. Total calibration uncertainty budget

The total calibration uncertainty budget for any given time is constructed from many sampled response functions $R(f, t)$. Each sample response function is constructed by sampling from the posteriors of the response function components. The response function components are

(1) The sensing DARM model parameters: $\vec{\lambda}_C = (H_C, f_{CC}, \delta\tau_C, f_S, Q_S^{-1})^T$

(2) The actuation DARM model parameters: $\vec{\lambda}_A = (H_U, \delta\tau_U, H_P, \delta\tau_P, H_T, \delta\tau_T)^T$
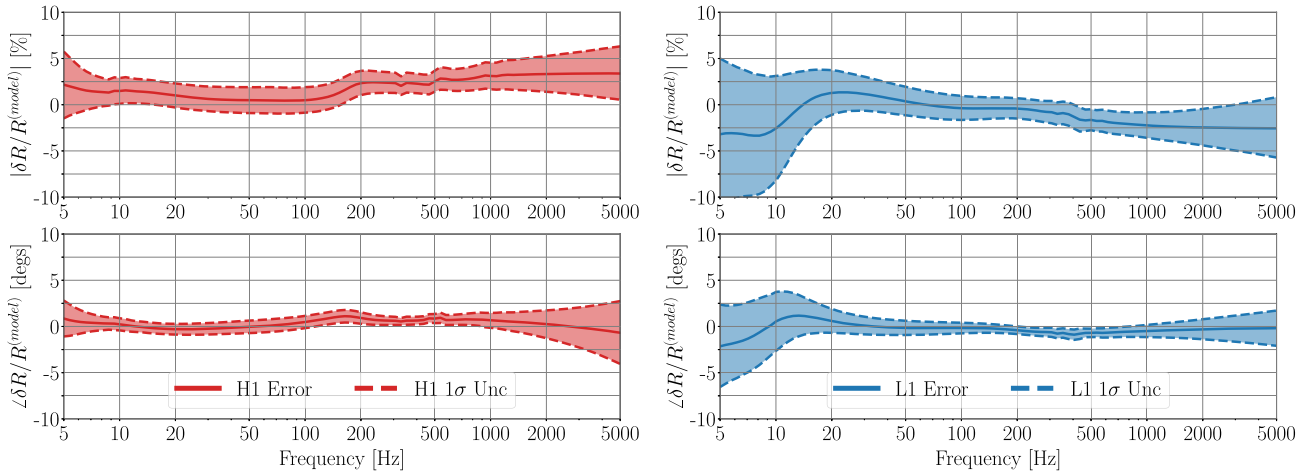

FIG. 8. Total calibration error and uncertainty budget at the time of GW170104. The uncertainty in the calibrated response function for the H1 detector is on the left, and for L1 is on the right. The y axis is relative response error $\delta R/R^{(model)}$ and uncertainty $\sigma_R/R^{(model)}$, with magnitude on top and phase on the bottom. The solid line is the median relative response, interpreted as the frequency dependent systematic error on the model response $R^{(model)}$. The dashed lines represent the 1σ uncertainty on this error. Stacking ten thousand drawn response function samples produces the numerical uncertainty budget shown here. The extreme 1σ uncertainties are presented in Table III.

TABLE III.   GW170104 extreme calibration uncertainty values for H1 and L1 in the 20-1024 Hz frequency range. "Extreme uncertainty" refers to the maximum and minimum of error $\pm 1\sigma$ uncertainty. The plots informing this table can be seen at Fig. 8.

GW170104 Uncertainty	H1	L1
$+1\sigma$ Magnitude [%]	4.6%	3.7%
-1σ Magnitude [%]	−1.0%	−3.7%
$+1\sigma$ Phase [degrees]	1.8°	1.9°
-1σ Phase [degrees]	−0.9°	−1.4°

(3) The sensing Gaussian process systematic error: $\delta C^{GP}(f)$

(4) The actuation Gaussian process systematic errors: $\delta A_U^{GP}(f)$, $\delta A_P^{GP}(f)$, $\delta A_T^{GP}(f)$

(5) The time dependent parameters: $\kappa_T(t)$, $\kappa_{PU}(t)$, $\kappa_C(t)$, $f_{CC}(t)$

(6) The photon calibrator radiation pressure strength: $H_{\mathrm{PCAL}}(t)$

Each of these components to the response have had posterior distributions constructed previously: (1) and (2) from the MCMC ensemble results on the reference measurements, (3) and (4) from the Gaussian process regressions on the residuals to incorporate unknown systematic errors, (5) from the calibration line measurements and coherence, and (6) from the 0.79% uncertainty in $H_{\mathrm{PCAL}}(t)$ from the photon calibrator paper [30].

Ten thousand samples are drawn from each of these posterior distributions and combined into ten thousand response function samples according to Eq. (4). Each of these response functions is then divided by the nominal response function, $R^{(\mathrm{model})}(f, t)$, which is constructed from the sensing model $C^{(\mathrm{model})}(f, t, \vec{\lambda}_C)$ and actuation model $A^{(\mathrm{model})}(f, t, \vec{\lambda}_A)$. This gives ten thousand relative response functions, each of which is plotted in Fig. 8. The median of this relative response function distribution constitutes the overall systematic error, and the 68th percentile upper and lower contours are the statistical uncertainty, both a function of frequency.

Figure 8 shows the calibration uncertainty at the time of the most recent detection, GW170104. Table III reports the "extreme uncertainty" for calibration between 20–1024 Hz during GW170104. Extreme uncertainty refers to the maximum and minimum of the systematic error $\pm 1\sigma$ uncertainty within a certain frequency band. This quantity is useful for searches requiring single number calibration uncertainty values, and ignore calibration systematic errors or frequency-dependent calibration uncertainty.

D. Calibration uncertainty for entire observing runs

Calibration error and uncertainty evolves over observing runs, affecting the results of continuous and stochastic

gravitational wave searches [10–13]. To assess the uncertainty of the detectors throughout an observing run, a total calibration uncertainty budget is made for every hour of observing data.

Collapsing the uncertainty budgets along the time axis, the 68th, 95th, and 99th percentile (1σ, 2σ and 3σ) limits are reported. The entire run's calibration error and uncertainty is often reduced to a single statement such as "over the course of an observing run, the 1σ uncertainty is no larger than XX % in magnitude and YY degrees in phase." To do so, the extreme uncertainty is taken in magnitude (XX%) and phase (YY degrees) using the 68th percentile contour over the relevant frequency band.

IV. RESULTS

The final calibration uncertainty budget for GW170104 is shown in Fig. 8. The "extreme uncertainties," or the maximum and minimum of error $\pm 1\sigma$ uncertainty, are reported in Table III.

The previous uncertainty quantification method from [16] conservatively reported 10% and 10 degrees uncertainties for GW150914 and the calibration uncertainties for all three O1 events in Table III in [4]. The uncertainty quantification method used for GW170104 was repeated on the O1 events. These results are reported in Appendix A, with plots of the uncertainty budgets for GW150914, LVT151012, and GW151226 in Fig. 10 and extreme uncertainties reported in Table IV.

Systematic errors are known discrepancies between the detector model and measurement. At low frequency, the systematic error is dominated by the Gaussian process regression on the actuation function residuals. At high frequency, fluctuations in the coupled cavity pole $f_{CC}(t)$, which are not corrected for in the calibration procedure, dominate the error budget.

Uncertainty everywhere is dominated by the Gaussian process regression on both functions. The uncertainty from the MCMC parameter fits on $\vec{\lambda}_C$ and $\vec{\lambda}_A$, and the uncertainty in the time dependent parameters $\kappa_T(t)$, $\kappa_{PU}(t)$, $\kappa_C(t)$, and $f_{CC}(t)$ tend to be about an order of magnitude smaller than the Gaussian process regression results. The 0.79% uncertainty in the photon calibration strength $H_{\mathrm{PCAL}}(t)$ contributes only to magnitude uncertainty.

The uncertainty and error for O2 strain data from November 19 through June 19 is shown in Fig. 9. This percentile plot was created by taking all observing time, producing an uncertainty budget for each hour, then compiling each budget into the percentiles shown. Overall, the detector calibration is stable over time. This consistency is largely due to the correction of the scale factors $\kappa_T(t)$, $\kappa_{PU}(t)$, and $\kappa_C(t)$ in the calibration pipeline models. Uncorrected systematic errors in the cavity pole $f_{CC}(t)$ are particularly visible at L1 at high frequency.
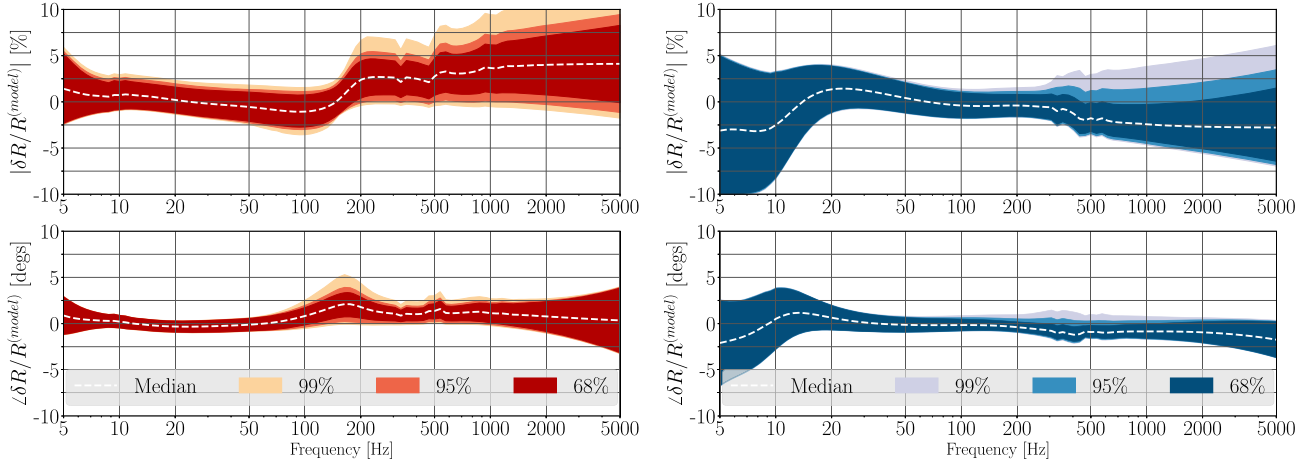
FIG. 9.   Total calibration uncertainty percentiles for observing run two. The percentiles are created for all of O2 data from November 30, 2016 to August 25th, 2017. H1's uncertainty is on the left, and L1's is on the right. The *y* axis is relative response $\delta R/R^{(\mathrm{model})}$ magnitude (top) or phase (bottom), stacked for all times in the observing run. The dashed white line is the median relative response, while the colors represent the 1σ calibration uncertainty for 68%, 95%, and 99% of the run's time. The largest changes in the calibration at H1 were due to clipping of the photon calibrator laser misreporting the strength of our response. The largest calibration changes at L1 were due to fluctuations in the coupled cavity pole, which changes in time but is not yet corrected for in our calibrated data.

During some parts of the second observing run, we have found that the reflection photodetector of the PCAL system at the H1 detector had suffered from clipping. Clipping means that the PCAL laser light incident on the photodetector was slightly off, giving a false low reading of how much power the PCAL was outputing. This means any measurement taken using the reflection photodiode as reference had a systematic error in scale. This includes the scale of any continuously measured time-dependent model parameters which are applied as correction factors for the estimated detector output, $h(t)$. We have quantified this systematic error using the same system's transmission photodiode, and included it as systematic error in the overall response. The systematic error was on the order of a few percent, and can be seen reflected in the upper percentiles of the H1 uncertainty in Fig. 9.

V. FUTURE WORK

There is much to be done to build upon this work. First, we will make use of calibration lines to track the detuning spring frequency f_s and Q values in real time. This will ensure the sensing plant is not severely detuned, or changing rapidly during detector operation. Second, we will employ time domain filters capable of correction for frequency-dependent changes in the plant. This will allow us to correct for changes in the coupled cavity pole f_{CC}, the antispring frequency f_s and quality factor Q, once these are successfully tracked. Third, the frequency-dependent systematic errors found from the Gaussian process regressions will be applied directly to the calibrated GW strain data $h(t)$ as it is produced, again through

time-domain filters. The above work would completely eliminate all known systematic errors from our calibrated data.

As we reduce the calibration uncertainty, properly characterizing systematic errors becomes much more important for precision astrophysics. Any systematic errors left unaccounted for in the calibrated data can result in systematic errors in binary black hole source parameters, compact binary merger rates, or tests of general relativity. Our direct measurements of our detector control loop plants combined with the physics-motivated response function model provide a sanity check that our understanding of the interferometer is close to correct.

There are a few considerations requiring quantification at the new low levels of uncertainty. One is understanding the difference between the quadruple pendulum response to an actual gravitational wave versus its response to the photon calibrator. In general, we care about the response of the test mass to external displacement, which causes light to be phase shifted out of the inteferometer's antisymmetric port. We simulate a gravitational wave by pushing on only a single end test mass with the photon calibrator laser. However, a real gravitational wave stretches space in the entire detector, in particular, the upper stages of the pendulum and the input test masses. The effect of this difference on calibrated GW data is on the order of about 1% at 10 Hz, and increases at lower frequencies. This now must be considered quantitatively as uncertainties approach this level.

Another consideration is the photon calibration actuation strength $H_{\mathrm{PCAL}}(t)$. Currently, the relative uncertainty in

$H_{\mathrm{PCAL}}(t)$ is 0.79% [30]. This is the fundamental limit on our uncertainty in the response R and therefore the GW strain data h. The uncertainty in $H_{\mathrm{PCAL}}(t)$ is dominated by uncertainty in the laser power and test mass rotation [30]. To push this fundamental limit lower, better measurements of the photon calibrator laser power and test mass rotation must be made, or more precise methods of calibration outside of the photon calibrator may need to be considered.

The uncertainty budget does not include error from test mass elastic deformation due to the PCAL laser exciting test mass vibrational modes. Preliminary evidence suggests that above around 3 kHz, elastic deformation has a significant effect on the calibration accuracy. Elastic deformation due to the PCAL must be further understood, monitored, and included in the uncertainty budget directly.

VI. CONCLUSION

The uncertainty and systematic error estimates reported in this paper represent a comprehensive characterization of our H1 and L1 detector calibrations for observing run two. In Advanced LIGO's lowest noise region, from about 20 Hz to 1 kHz, the uncertainty in the calibrated data has been reduced from what was previously reported in [16]. The uncertainty estimates for O2 give more refined results, with uncertainty growing at extreme frequency regions below 20 Hz and above 1 kHz, and reduced uncertainty in the low noise frequency region.

GW170104's detection and parameter estimation are primarily limited by noise, and not by calibration uncertainty. As Advanced LIGO becomes more and more sensitive, the signal-to-noise ratio of some detections will become quite large (as high as 100 or more), and calibration uncertainty will begin contributing significantly to source parameter estimation uncertainty. With more observing time comes more detections, enabling new tests of general relativity which will be limited by the precision of our detector data. Precision astrophysics demands the best understanding of our calibrated data possible. The methods described in this paper were developed primarily to enable the best science possible from LIGO's gravitational wave detections.

APPENDIX: OBSERVING RUN ONE RESULTS

During O1, there were two loud gravitational wave detections, GW150914 and GW151226 [1,2]. There was also a relatively quiet transient, LVT151012, that was likely a gravitational wave [4]. The previous report of the calibration uncertainty on these events did not include the new methods developed for O2 [16]. We have returned to the O1 data to recalculate the calibration uncertainties at the time of the three O1 events. The old calibration uncertainty method results are reported in Table III of [4]. The results shown are for version 2 (C02) of the calibrated GW strain data, available shortly after the run ended in January 2016.

Figure 10 shows plots of the calibration uncertainties at the times of the O1 events. Table IV reports the extreme uncertainties at the times of the O1 events. Figure 11 shows the calibration uncertainty for all of O1, meaning the GW strain data from September 14, 2015 through January 19, 2016.
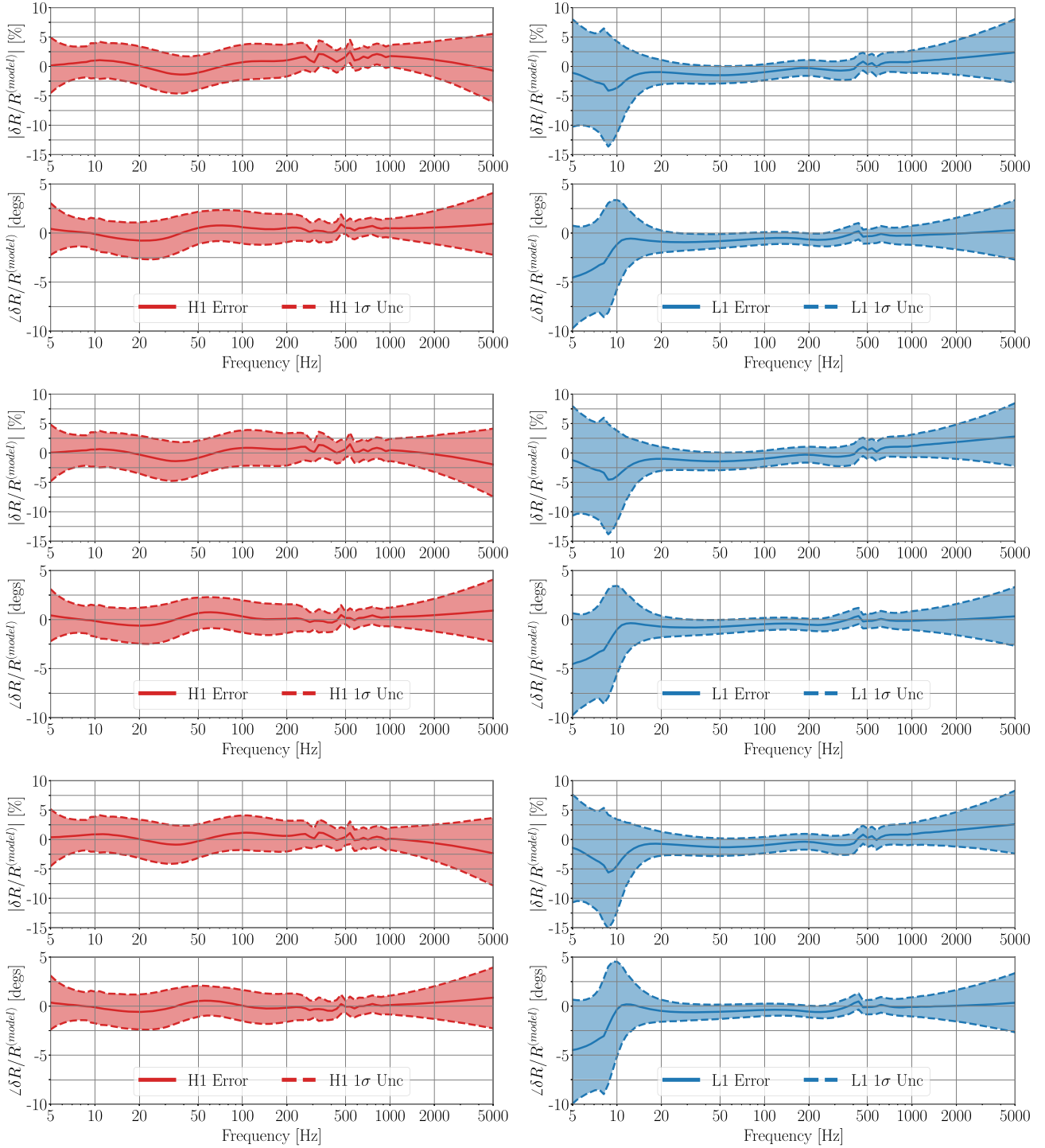
FIG. 10. Total calibration uncertainty budgets for GW150914, LVT151012, and GW151226. Uncertainties for GW150914 are on top, LVT151012 in the middle, and GW151226 on the bottom. The uncertainty in the calibrated response function for the H1 detector is on the left, and for L1 is on the right. The y axis is relative response error $\delta R/R^{(\mathrm{model})}$ and uncertainty $\sigma_R/R^{(\mathrm{model})}$, with magnitude on top and phase on the bottom. The uncertainties for these O1 events have been calculated using the refined O2 methods described in this paper. All the budgets are quite similar with slight differences coming from the uncorrected time dependent cavity pole. The jagged lines around 300–500 Hz come from the actuation function notches, like those seen in Fig. 4. These budgets report a smaller uncertainty in the region around 100 Hz than reported in past calibration uncertainty publications, with significantly smaller systematic error fluctuations [16]. The refined calibration uncertainty budget methods give a more sensible uncertainty budget for O1 events, with uncertainty expanding at extreme frequencies and reduced in the lowest noise regions.

TABLE IV. GW150914, LVT151012, and GW151226 extreme calibration uncertainty values for H1 and L1 in the 20–1024 Hz frequency range. "Extreme uncertainty" refers to the maximum and mininum of error $\pm 1\sigma$ uncertainty. The plots informing these tables can be seen in Fig. 10.

GW150914 Uncertainty	H1	L1
$+1\sigma$ Magnitude [%]	3.7%	1.1%
-1σ Magnitude [%]	-4.6%	-3.0%
$+1\sigma$ Phase [degrees]	$2.4°$	$0.3°$
-1σ Phase [degrees]	$-2.7°$	$-2.0°$

LVT151012 Uncertainty	H1	L1
$+1\sigma$ Magnitude [%]	3.8%	1.0%
-1σ Magnitude [%]	-4.8%	-3.0%
$+1\sigma$ Phase [degrees]	$2.3°$	$0.4°$
-1σ Phase [degrees]	$-2.5°$	$-1.8°$

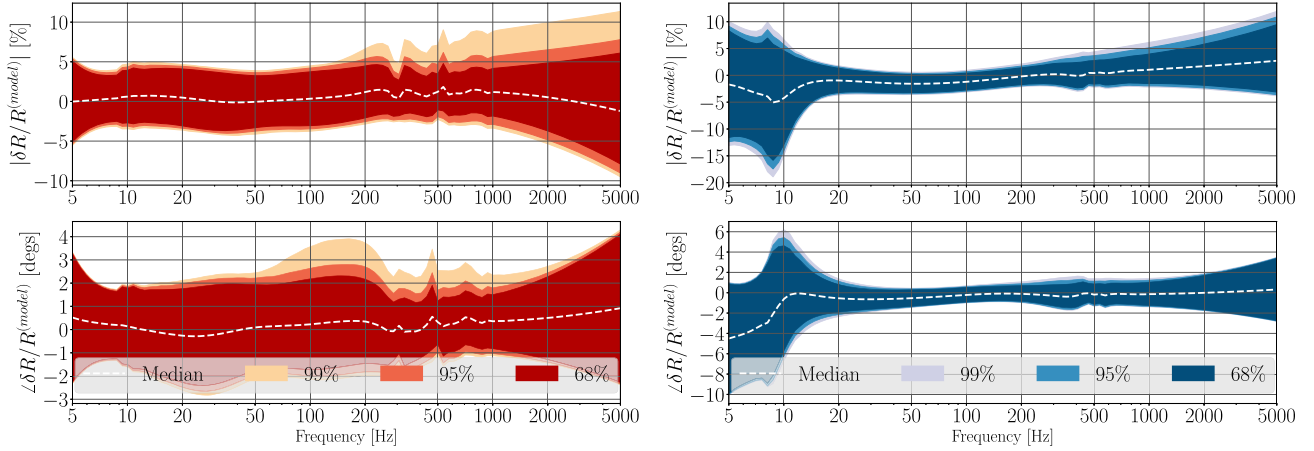| GW151226 Uncertainty	H1	L1
$+1\sigma$ Magnitude [%]	4.1%	1.3%
-1σ Magnitude [%]	-4.1%	-2.8%
$+1\sigma$ Phase [degrees]	$2.1°$	$0.6°$
-1σ Phase [degrees]	$-2.4°$	$-1.6°$

FIG. 11. Total calibration uncertainty percentiles for all of observing run one. The uncertainty in the calibrated response function for the H1 detector is on the left, and for L1 is on the right. The y axis is relative response error $\delta R/R^{(\mathrm{model})}$ and uncertainty $\sigma_R/R^{(\mathrm{model})}$, with magnitude on top and phase on the bottom. The dashed white line is the median relative response, while the colors represent the 1σ calibration uncertainty for 68%, 95%, and 99% of the run's time. Both interferometer's uncertainty and systematic error are fairly consistent over the course of the run, with the larger fluctuations coming from changes in the coupled cavity pole, which is not corrected for in O1 calibrated data.

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 061102 (2016).

[2] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 241103 (2016).

[3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **118**, 221101 (2017).

[4] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **6**, 041015 (2016).

[5] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 241102 (2016).

[6] B. P. Abbott *et al.*, Astrophys. J. Lett. **833**, L1 (2016).

[7] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 221101 (2016).

[8] N. Yunes, K. Yagi, and F. Pretorius, Phys. Rev. D **94**, 084002 (2016).

[9] K. Chamberlain and N. Yunes, Phys. Rev. D **96**, 084039 (2017).

[10] B. P. Abbott *et al.*, Astrophys. J. **839**, 12 (2017).

[11] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration) arxiv:1707.02669

[12] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **118**, 121101 (2017).

[13] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **118**, 121102 (2017).

[14] W. Del Pozzo, T. G. F. Li, and C. Messenger, Phys. Rev. D **95**, 043502 (2017).

[15] D. F. Chernoff and L. S. Finn, Astrophys. J. **411**, L5 (1993).

[16] B. P. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **95**, 062003 (2017).

[17] J. Aasi *et al.*, Classical Quantum Gravity **32**, 074001 (2015).

[18] B. P. Abbott *et al.* (LIGO Scientific C and Virgo Collaborations), Phys. Rev. Lett. **116**, 131103 (2016).

[19] K. Izumi and D. Sigg, Classical Quantum Gravity **34**, 015001 (2017).

[20] N. A. Robertson *et al.*, Classical Quantum Gravity **19**, 4043 (2002).

[21] S. M. Aston *et al.*, Classical Quantum Gravity **29**, 235004 (2012).

[22] F. Matichard *et al.*, Precis. Eng. **40**, 273 (2015).

[23] L. Carbone *et al.*, Classical Quantum Gravity **29**, 115005 (2012).

[24] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures* (John Wiley & Sons, New York, 2011), Vol. 729.

[25] A. Buonanno and Y. Chen, Phys. Rev. D **65**, 042001 (2002).

[26] R. L. Ward, Ph.D. thesis, California Institute of Technology, 2010, https://thesis.library.caltech.edu/5836/.

[27] E. D. Hall, Ph.D. thesis, California Institute of Technology, 2017, https://thesis.library.caltech.edu/10031/.

[28] D. Tuyenbayev *et al.*, Classical Quantum Gravity **34**, 015002 (2017).

[29] M. Rakhmanov, R. Savage, D. Reitze, and D. Tanner, Phys. Lett. A **305**, 239 (2002).

[30] S. Karki *et al.*, Rev. Sci. Instrum. **87**, 114503 (2016).

[31] J. D. Hunter, Comput. Sci. Eng. **9**, 90 (2007).

[32] D. Foreman-Mackey *et al.*, Publ. Astron. Soc. Pac. **125**, 306 (2013).

[33] D. Foreman-Mackey, J. Open Source Software **1**, 24 (2016).

[34] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006), Vol. 1.

[35] F. Pedregosa *et al.*, J. Mach. Learn. Res. **12**, 2825 (2011).