Draft version September 19, 2016
Preprint typeset using LaTeX style emulateapj v. 12/16/11

IMPROVING THE ACCURACY OF ALIGO CALIBRATION: CALIBRATION ERROR MODELS FOR PRECISION TESTS OF GENERAL RELATIVITY

M. McIntosh[1]
Advised by C. Cahillane,[2] A. Weinstein,[2] K. Blackburn[2]

[1]Harvard University Department of Astronomy, 60 Garden Street, Cambridge, MA 02138, USA
[2]LIGO, California Institute of Technology, Pasadena, California 91125, USA
LIGO Laboratory Caltech SURF Program Progress Report #1
LIGO-T1600260
September 19, 2016
Draft version September 19, 2016

ABSTRACT

In 2015, the first direct detection of gravitational waves (GWs) was realized with the advanced Laser Interferometer Gravitational Wave Observatory (aLIGO). With the detection emerged the ability to test General Relativity (GR) in large velocity, highly dynamical and strong-field gravity regimes. If GWs were to reveal deviations from GR, the perturbations would be extremely small; characterizing and reducing uncertainties in aLIGO data would allow as much physical information to be recovered in the GW signal as possible. The intent of this project is twofold: to describe the calibration methods and uncertainties used for aLIGO and to also begin the discussion on the impact of calibration errors on precision tests of GR. We first describe the current calibration methods for aLIGO and the most important calibration error sources. Then, in a toy model, we generate GW signals with a non-GR perturbation in the merger-ringdown regime. We recover the perturbation through Bayesian parameter estimation to determine the precision with which such perturbations can be detected. We conclude with discussing how new calibration error models could increase that precision, and include a more sophisticated model setup in the Appendix.

I. INTRODUCTION

In 1915, Einstein published his General Theory of Relativity (GR). This theory and his following papers predicted the existence of gravitational waves (GWs), or oscillations in the gravitational field caused by the acceleration of massive bodies. In 1993, a Nobel Prize went to Hulse and Taylor [8] who discovered a pulsar system losing energy at the same rate as predicted by GW emission and thus implying the existence of GWs. Then in 2015, a direct detection of a GW occured. GW150914, identified as a result of a binary black hole merger, was observed with the Advanced Laser Interferometer Gravitational Wave Observatory (aLIGO) detector network [2]. With this detection, the field of direct GW observations emerged.

GWs allow us to observe strong-field dynamics of space-time and astrophysical phenomena inaccessible by electromagnetic radiation. With the direct detection of GWs, experiments to test GR in large velocity, highly dynamical, and strong-field gravity regimes can be conducted. Because GWs cause extremely small perturbations in spacetime (on the order of 10^{-20}m [2]) the aLIGO sensors and signal analyses need to be precise. Characterizing and reducing uncertainties in aLIGO data allows us to reclaim as much physical information from the GW signal as possible.

The goals of this work are to (1) describe both the calibration methods used for aLIGO and their uncertainties and (2) begin the discussion on the effects of this calibration uncertainty on precision tests of GR. It is unknown whether the calibration errors (CEs) in aLIGO are great enough to disguise any non-GR perturbations that may exist in GW signals which would otherwise be revealed by stacking many GW signals (a technique discussed in Section V). If they are great enough, it is unknown if implementing a calibration error model would reduce the CE impact enough to see these non-GR perturbations. So we begin the discussion of the effects of calibration uncertainty on precision tests of GR by developing a toy model in which we introduce a non-GR perturbation in the merger-ringdown regime and then attempting to recover the perturbation through Bayesian parameter estimation. We suggest next steps in this investigation, including implementing a frequency dependent calibration error model to see its impact.

The layout of this paper is as follows. First, we specify our focus on CEs in Section II. Next, we summarize how aLIGO detects GW and how these CEs impact the instrument in III. In IV, we describe the differential arm length (DARM) closed feedback loop transfer function, which contains the GW wave signals and CEs for aLIGO. We then describe our toy model in VI. We discuss next steps in VII and a more sophisticated modeling in X.

II. CALIBRATION ERRORS

Two sources of uncertainty in any instrument are statistical and systematic uncertainty. Statistical uncertainty— a zero-mean Gaussian distribution measurement variation— is unavoidable, but it can be reduced by taking additional observations. Systematic uncertainty is relatively avoidable and cannot be reduced by

additional measurements (unless GR is used to calibrate the data). Conversely, systematic uncertainty stems from an incorrect characterization of a detector and causes calibration errors. For aLIGO, careful calibration has to be maintained to accurately associate the frequency response of the detector with the motion of aLIGO's optics and consequently reduce systematic error. These systematic/calibration errors are the focus of this project.

For aLIGO, calibration errors (CEs) are errors that pertain to the conversion of instrumental readout to GW strain data and are contained in a differential arm length (DARM) control loop. CEs can affect detection rates and parameter estimations; here we are concerned with parameter estimation. Current CEs for aLIGO are estimated as an overall constant in wave amplitude and phase or as varying such that the errors can be fit with splines, but a new, functional frequency dependent estimation method also is discussed here. Previous works that have studied calibration error analysis for aLIGO include [15, 19].

This work uses a Bayesian approach to quantify potential CE effects on signal injections which mimic both GR and non-GR conforming GWs. It is possible that the CEs will blur our ability to distinguish between the two; we also investigate if our frequency dependent method of characterizing CEs will allow us to distinguish the GR and non-GR conforming signal injections. Though the few individual GW signals detected so far have had no statistically significant disagreement with GR and their statistical error exceeds their systematic error, the aLIGO detectors are not yet at their design sensitivities, louder GW sources may yet be detected, and the potential to combine or "stack" GW observations as in [16] to increase sensitivity all encourage this work [4].

III. HOW LIGO DETECTS GWS

aLIGO is a complex and cutting edge instrument; it consists of a modified Michelson interferometer with Fabry-Perot arms, power-recycling mirrors, and resonant sideband extraction which all allow it to measure minuscule phase propagation differences via the Pound-Drever-Hall technique [7, 10, 12]. However, we follow [19] in this analysis and reduce the entire instrument to a sensor with a single degree of freedom: differential arm length (DARM) perturbations. When a GW passes through aLIGO, the space in one direction stretches while the space in the other direction contracts; this arm length change/perturbation interrupts the constructive interference of laser light maintained in the arm cavities and some light escapes the arms to a photodetector (see Figure 1). This escaped light generates the signal we analyze to recover the GW parameters.

Figure 2 illustrates a waveform that aLIGO might detect. From the amplitude and phase of this waveform, we can extract astrophysical parameters such as those listed in Table 1.

In this work, we focus on a single-parameter analysis in the merger-ringdown regime to test GR using multiple events, rather than characterizing a particular GW event. Subsequently, we marginalize over the parameters listed in Table 1.

IV. DARM FEEDBACK CONTROL LOOP

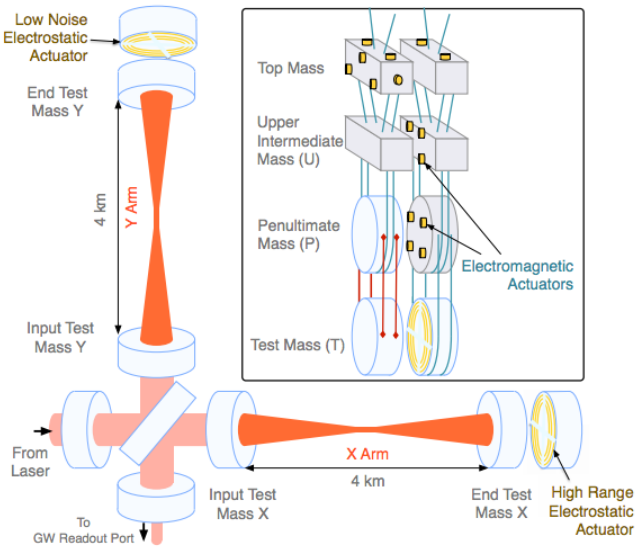

FIG. 1.— This diagram of aLIGO is from [15]. The arm cavities are contained between the reflective test masses. These test masses are suspended from a quadruple pendulum system and are adjusted spatially by an actuating system displayed in the upper right corner. These adjustments allow resonance to be maintained in the arm cavities. When resonance is disrupted by a passing GW stretching one arm cavity and shrinking the other, some light escapes to the GW Readout Port at the bottom of the figure.

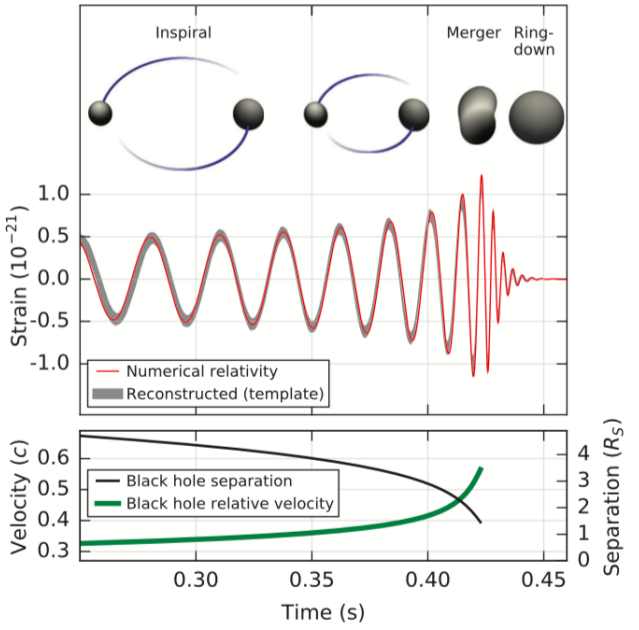

FIG. 2.— A figure illustrating an inspiral-merger-ringdown (IMR) of a compact binary system from [2]. According to GR, two objects in orbit will slowly spiral inwards due to a loss of energy and angular momentum via GWs. The frequency and amplitude of the emitted GWs increases as the orbital distance between the objects shrink. When the objects finally merge they can radiate GWs as a superposition of quasinormal ringdown modes. One mode will eventually dominate with a exponentially damped, constant frequency wave [4]. We observe this as a lower frequency inspiral phase, a post-merger peak at some fixed frequency, and then a higher frequency ringdown. In this work we investigate the differences due to non-GR conformity as described in a parameter pertaining to the ringdown of the GW.

TABLE 1
PARAMETERS USED TO CHARACTERIZE A GW DETECTION.

m_i	Masses
q	Mass ratio
$\mathcal{M}_c$	Chirp Mass
$\mathbf{s}_i$	Spins
χ_{eff}	Effective inspiral spin parameter
D_L/Mpc	Luminosity Distance
t_c	Coalescence time
ϕ_i	Phase of coalescence
α, δ	Sky location
$\cos\iota$	Orbital orientation
ψ	Polarization angle

The external differential arm length change, ΔL_{ext}, is related to the GW amplitude, called the "strain:"

$$h(f;t) = \frac{\Delta L_{ext}}{\langle L \rangle} \tag{1}$$

where $\langle L \rangle \approx 4000m$ is the arm cavity length gain of aLIGO. aLIGO's photodetector does not directly measure ΔL_{ext} but rather the current generated from the amount of light that has escaped from the arm cavities, d_{err} (the DARM error signal), in arbitrary units. ΔL_{ext} must be reconstructed from $d_{err}(f)$.

$d_{err}(f)$ is measured continuously in a closed feedback loop. The purpose of this feedback loop (see the reduced block schematic in Figure 3) is to recenter the mirrors used in aLIGO after the arms have been perturbed by a GW or noise so that the constructive interference/resonance of the laser is maintained. This allows the instrument to measure the next arm length differential as quickly as possible.

We can reconstruct $h(f;t)$ from the DARM control loop (see Figure 3):

$$\langle L \rangle h(f;t) - \Delta L_{ctrl} = \Delta L_{res} \tag{2}$$

$$h(f;t) = \frac{1}{\langle L \rangle}\left(\frac{1}{C(f;t)} D(f) d_{err} + A(f;t) d_{ctrl}\right) \tag{3}$$

$$h(f;t) = \frac{1}{\langle L \rangle}\left(\frac{1+G(f;t)}{C(f;t)} d_{err}\right) \tag{4}$$

$$h(f;t) = \frac{1}{\langle L \rangle} R_e(f;t) d_{err} \tag{5}$$

$$\text{where} \quad R_e(f;t) = \frac{1 + A(f;t)D(f)C(f;t)}{C(f;t)} \tag{6}$$

$$= \frac{1+G(f;t)}{C(f;t)} \tag{7}$$

Here, $C(f;t)$ is the transfer function of the arm cavity or the sensing function, $D(f)$ is a digital filter, and $A(f;t)$ is the actuation function that corrects mirror position. d_{ctrl} is the signal sent to the actuators describing how to move the test masses to recover resonance, ΔL_{ctrl} is the length perturbation the actuation function applies to the test masses, and ΔL_{res} is any residual length change the actuation function happens to not correct. We take the various transfer functions into $G(f;t) = C(f;t)D(f)A(f;t)$, the DARM open loop gain. We further rearrange this equation into a response function, $R_e(f;t)$, which lets us estimate uncertainty more easily. $D(f)$ is known precisely, so the uncertainty in our GW strain, $\sigma_h(f;t)$, is dominated by the uncertainty in $R_e(f;t)$: $\sigma_R(f;t)$. Equation 7 is derived and its components described more thoroughly in [19, 15]. More complex calibration loop treatments are given in [1, 13].

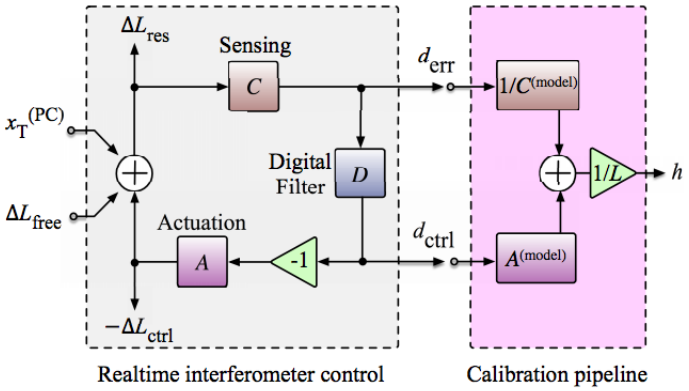

FIG. 3.— A block diagram of the DARM feedback control servo from [15]. This schematic shows that aLIGO's output is dependent on the performance of the feedback loop. Each component of the feedback loop is described by a transfer function and the uncertainty on the overall loop transfer function yields the CEs on the GW strain detection. The transfer functions of all of the components in the feedback loop are necessary to reconstruct the GW signal. The subsystems are described more thoroughly in [15, 19].

In equation 5, $R_e(f;t)$ gives the theoretical or exact response function, but the measured length function $R_m(f;t)$ includes CEs from the sensing $C(f;t)$ and actuation $A(f;t)$ functions as well as the slow, time-dependent drift in these functions. Because detectors are noisy, drift with time, and can glitch, $R_e(f;t)$ and $R_m(f;t)$ can differ. This leads to systematic errors in GW strain reconstruction.

The frequency dependent and time dependent parameters of $R(f;t)$ are what impact the response function uncertainty, $\sigma_R(f;t)$. We examine these parameters for $C(f;t)$ and $A(f;t)$.

IV.I. *Sensing Function, $C(f;t)$*

The sensing function (see Figure 4) "senses" GW strain. It represents the interferometer optical plant and is approximated by a coupled-cavity single pole function [9]:

$$C(f;t) = \frac{\kappa_C(t) C_R(f)}{1 + if/f_{CC}(t)} e^{-2\pi i f \tau_C} \tag{8}$$

$\kappa_c(t)$: optical gain
$fcc(t)$: coupled cavity pole
$C_R(f)$: "sensing residual" after cavity pole is divided out
τ_C : sensing function time delay

The optical gain $\kappa_c(t)$ depends on the laser power in the optical arm cavities and has a time dependent scalar gain factor. Equation 7 indicates $R_e(f;t) \approx 1/C(f;t)$ when $|G(f;t)| << 1$. Changes in the optical gain produce the largest systematic errors at frequencies above the unity gain frequency (40 Hz and 56 Hz for H1 and

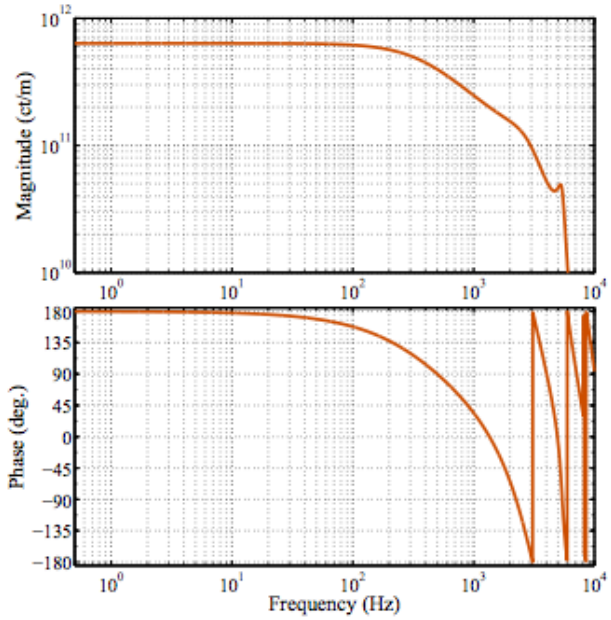

FIG. 4.— Figure from [15]. The sensing function model is shown in magnitude (top) and phase (bottom). The frequency dependence of $C(f;t)$ is determined by f_{cc} up until 1 kHz; afterwards, analog-to-digital conversion factors (in $\kappa_c(t)$) dominate. The features at higher frequencies are due to photodiode electronics and processing filters [15].

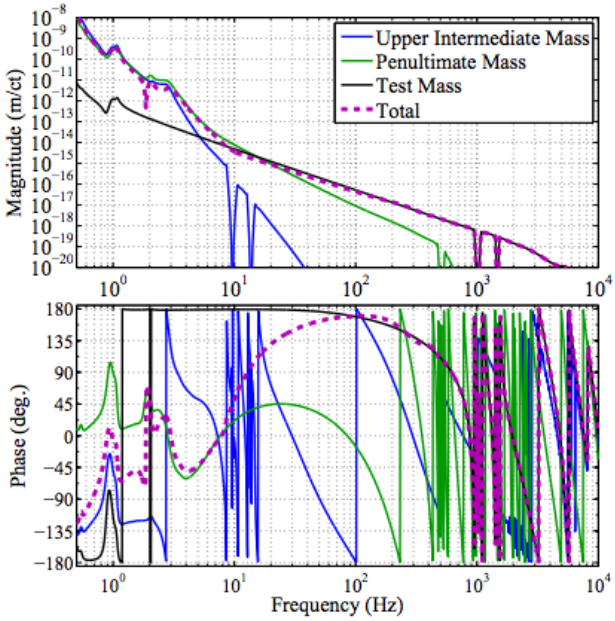

FIG. 5.— Figure from [15]. The actuation function $A(f;t)$ shown in magnitude (top) and phase (bottom), includes the actuation functions for each pendulum stage. As frequency increases, the dominating pendulum stage progresses to the lower pendulum stages. Digital notch filters are apparent at select frequencies with the purpose of avoiding mechanical instabilities [15].

L1, respectively [15]). The coupled cavity pole frequency $f_{cc}(t)$ depends on the reflectivities of the interferometer optics and has a time dependency due to cavity length and alignment changes [14].This changes the shape of the sensing function at frequencies close to the nominal coupled cavity pole frequencies. At high frequencies, $\kappa_c(t)$ and $fcc(t)$ contribute systematic errors in the magnitude of ΔL_{ext} [17].

IV.II. *Actuation Function $A(f;t)$*

The actuation function (see Figure 5) describes the physical actuators that spatially adjust the test masses hung in quadrature (see Figure 1). The test mass stage are labeled as follows: Top, Upper Intermediate, Penultimate, and Test. All except for the top stage are actuated upon and so contribute a term to the actuation function. The Upper-Intermediate and Penultimate masses have Optical Sensor and Electromagnetic (OSEM) actuators while the test mass, the mass that the laser light hits, has a ElectroStatic Drive (ESD) actuator. The ESD allows for finer spatial adjustments but can build up charge from residual gas via ion vacuum pumps in aLIGO. Consequently, its strength changes with time.

$$A(f;t) = \left[\kappa_T(t)A_T(f) + \kappa_P(t)A_P(f) + \kappa_P(t)A_U(f)\right] e^{-2\pi i f \tau_A} \tag{9}$$

$\kappa_T(t)$: ESD actuation strength
$\kappa_P(t)$: OSEM actuation strength
$A_T(f)$: Test mass actuation function
$A_P(f)$: Penultimate mass actuation function
$A_U(f)$: Upper-Intermediate mass actuation function
τ_A : computational time delay in digital-to-analog conversion

Equation 7 indicates $R_e(f;t) \approx A(f;t)D(f)$ when $|G(f;t)| >> 1$. $A_T(f)$ is the dominant term in $A(f;t)$ for frequencies greater than 20 Hz; the largest systematic errors contained in $A(f;t)$ ($\pm 15\%$) are due to variations in the actuation strength of the ESD from $20 - 50$ Hz and are contained in $A_T(f)$ [17].

IV.III. *Full Parametrization of the Response Function and Real-Time Calibration Measures*

With our parameters from $C(f;t)$ and $A(f;t)$, we can rewrite our actuation function beginning with equation 7 and dropping the time delays as:

$$R_e(f;t) = \frac{1 + A(f;t)D(f)C(f;t)}{C(f;t)} \tag{10}$$

$$R_e(f;t) = \frac{1}{C(f;t)} + D(f)A(f;t) \tag{11}$$

$$R_e(f;t) = \left(\frac{1 + if/f_{CC}(t)}{\kappa_C(t)C_R(f)}\right) + D(f)\left(\kappa_T(t)A_T(f) + \kappa_P(t)A_P(f)\kappa_P(t)A_U(f)\right) \tag{12}$$

The time dependent parameters ($\kappa_T(t)$,$\kappa_P(t)$,$\kappa_C(t)$,$f_{CC}(t)$) are monitored for variation using intentional injections at four frequencies (called calibration lines) into the DARM spectrum throughout an observation run (see Table 2). The calibration lines reveal any short-term gain fluctuations in optical plant or actuation strengths.

TABLE 2
Calibration line table recreated from [17]. Lines 1-3 are used to estimate κ_T and κ_P and line 4 for κ_C and f_c for the LIGO Hanford (H1) and LIGO Livingston (L1) detectors.

#	Signal	Freq. H1.	(Hz) L1	Line Purpose
1	x_T	35.9	35.3	ESD actuation strength
2	ΔL_{pcal}	36.7	34.7	DARM actuation
3	x_{ctrl}	37.3	33.7	Penultimate & Upper-Intermediate actuation strengths
4	ΔL_{pcal}	331.9	331.3	Optical gain and coupled cavity pole frequency

The frequency dependent parameters in our response function ($A_U(f)$, $A_P(f)$, $A_T(f)$, $C_R(f)$) are measured between observation runs using swept sine calibration; a sine wave displacement signal is applied to a test mass while the interferometer is locked. Then, the frequency is slowly swept over the GW detection band and the error signal is measured as a function of the displacement. This yields the closed loop transfer function equation. During observation runs, the frequency dependent parameters are measured a few times per run using wide-band psueudo-random signals with amplitudes below the noise level. The disturbances created from this are small and distributed over the whole gravitational wave band. While having the advantage of not impacting any GW signal observations, below noise-level calibration signals have the disadvantage of needing a longer integration time to extract. This results in obtaining only a few frequency calibration signals per observation run. It is assumed that the frequency dependence of the control loop will not vary in between these calibration signal extractions [13].

Because this does not always hold true, the interpreted GW strain will be different from the true GW strain in both phase and amplitude. As discussed in Section III, this affects not only the precision measurement of astrophysical parameters like masses, sky location, distance, and etc, but also the precision measurement of universal parameters like those that describe variations from GR. To decrease the difference between the measured length function and the exact length function, we seek to better characterize aLIGO's CEs through Bayesian parameter estimation.

V. Bayesian Parameter Estimation

We can compare a parametrized GW waveform model to the detected strain signal, using a matched filtering technique with template banks, further described in [6]. It is an accurate and time sensitive method to identify potential GWs from compact binary coalescence (CBC). Figure 2 shows the GW waveform model used for GW150914.

With the matched filtering results, we can construct probability density functions (PDFs) for each of the parameters in the GW detection. To be explicit, we begin with Bayes' theorem; the probability that a parameter, θ, is the correct value given some data, x, is equal to the probability of getting the data given the parameter times the probability that the parameter is the correct value, and divided by the evidence of the data:

$$P(\theta|x) = \frac{P(x|\theta) \times P(\theta)}{P(x)} \tag{13}$$

Here, $P(\theta|x)$ is the posterior probability, $P(x|\theta)$ is the likelihood, $P(\theta)$ the prior, and $P(x)$ the evidence.

We can then express the probability that the strain data $x(f)$ came from an astrophysical system with parameter θ with the log-likelihood, $\ln P(x|\theta)$:

$$\ln P(x|\theta) = -1/2 \int_0^\infty \frac{|h(f,\theta) - x(f)|^2}{S_{nn}(f)} \tag{14}$$

where $x(f)$ is the strain data from the detector, $h(f,\theta)$ is the GR prediction for the strain with parameter θ, and $S_{nn}(f)$ is the power spectral density of the detector's strain noise.

We can update this probability as more data becomes available. We can "stack" GW observations like so:

$$\begin{aligned} P(\theta|x,y) &= \frac{P(x,y|\theta) \times P(\theta)}{P(x,y)} \\ &= \frac{P(y|\theta,x) \times P(\theta|x)}{P(x,y)} \quad \text{Substitute in Eqn. 13} \\ &= \frac{P(y|\theta) \times P(x|\theta) \times P(\theta)}{P(x)P(y)} \\ &\quad \text{Assuming GWs are uncorrelated the probability of } y \text{ does not depend on } x \end{aligned} \tag{15}$$

where $P(\theta|x,y)$ is the posterior probability that θ is the correct value given that x, our data or a GW detection, and y, new data or another GW detection, exist. The normalization constants of these models are typically ignored in favor of simply comparing two competing models by taking the ratio (called the odds ratio) of posterior probabilities to the evidence/potential GW signal:

$$O_{i,j} = \frac{P(\mathcal{H}_i|x)}{P(\mathcal{H}_j|x)} B_{ij} \tag{16}$$

where $\mathcal{H}_i$ is some hypothesis and is compared to another hypothesis, $\mathcal{H}_j$. The Bayes factor or evidence ratio, B_{ij}, is the ratio of likelihoods between the models. It is often used as a statistic to describe confidence in the correctness of a model or to rank competing hypotheses given the observed data. From equation 15 we see that as we amass new GW detections we update the probability by multiplying the posterior probability distributions for θ for each event together. For an arbitrary number of detections, the probability of the parameter is then given by:

$$P(\theta|x_i) \propto \left[\prod_i P(x_i|\theta)\right] \times P(\theta) \tag{17}$$

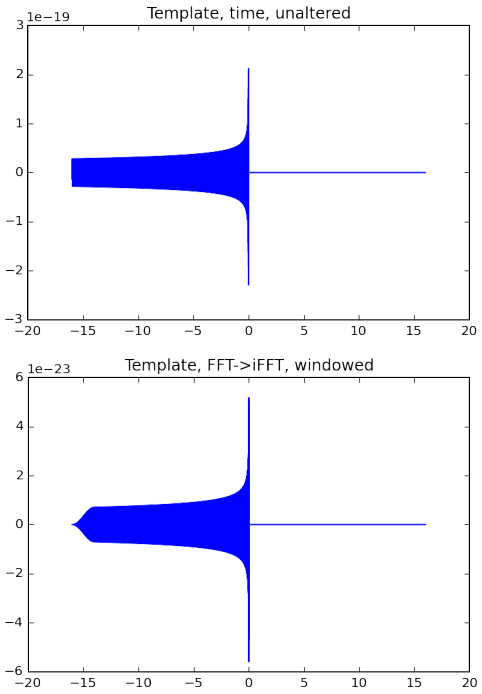

FIG. 6.— The unaltered GW151226 template (above), available from `https://losc.ligo.org/s/events/GW151226/LOSC_Event_tutorial_GW151226.html`, and the windowed template (below).

Stacking GW strain signals in this way can constrain parameter estimation better than a single detection can. For computational simplicity, we limit our toy model to one parameter.

VI. THE TOY MODEL

Though up to 100's of observations per year are expected by the time LIGO operates at design sensitivity [3], only three GW detections have been published as of the time of this paper. For now, we use simulated GW signals instead of real GW signals. GW waveform models are based on an analytical inspiral-merger-ringdown (IMR) model (see Figure 2)

Our toy model begins with the GW151226 template windowed with a Tukey/tapered cosine window to remove effects at the beginning and end of the template (See Figure 6). We modified a copy of the template to create the base for our mock data by smoothly decreasing the amplitude of the ringdown (see Figure 7) by 50% with a logistic function. This represents a possible non-GR perturbation in our GW template. We then created 550 templates of these smoothly damped in ringdown templates in order to populate a template bank for match filtering. We additionally scaled these templates by mass, following a m^{-1} power law from 5 to $45 M_\odot$ (See Figure 8).

Returning to our original ringdown damped template, we added colored (frequency-dependent) noise to achieve a mock signal (see Figure 9). The SNR of the noise was drawn from a SNR^{-4} distribution (see Figure 10).

We then run our mock data through match filtering against our template bank to generate PDFs for different SNRs, noting which templates were returned as the best fits (see Figure 11). Next, we stacked our observations by multiplying the posteriors (assuming a flat prior). After renormalizing the PDFs, we measured their variance to represent our uncertainty of the parameter measurement and plotted that uncertainty against the number of events stacked (see Figure 12)

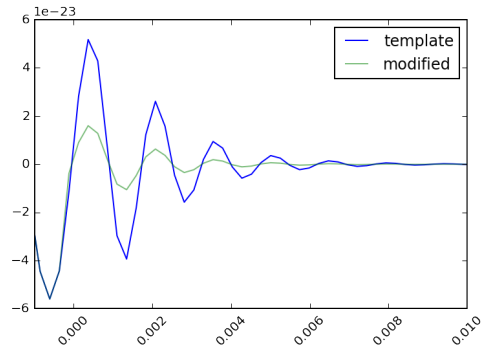

FIG. 7.— A smoothly damped ringdown

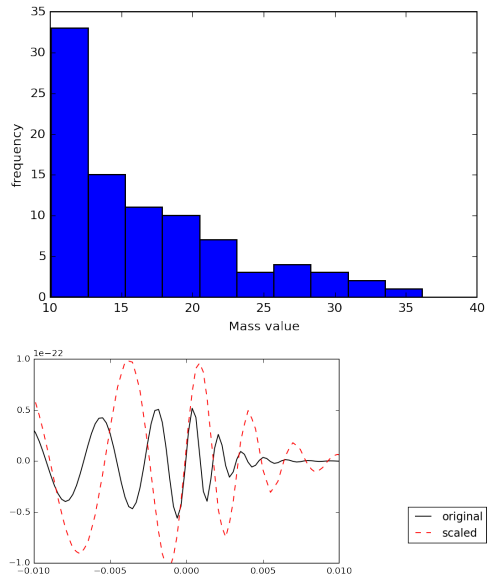

FIG. 8.— The mass distribution sampled from to scale the templates and generate mock data (above) and a resulting scaled ringdown (below)

VII. NEXT STEPS

Next steps for this toy model in particular include adding CE to the mock data representing optical gain uncertainty error along with CE models to see if we would be able to observe the non-GR perturbation and repeating this for very low SNR ($\sim$ 10). Once this is implemented, we can test ides such as: What calibration accuracy do we need before statistical error dominates? For large SNR, how well can we constrain a parameter? With an additional distorted calibration, how well can we constrain that parameter? Which parameters in CE models are impacted by calibration errors the most? For which BH masses does the CE matter the most? Which parameters should we do our best to measure? How well should we measure them for N events?

The answers to such questions will inform us in our estimation of the impact of aLIGO calibration uncertainty on precision tests of GR using observations of GWs from BBH mergers. Additionally, they will direct us to the most important questions to ask of more sophisticated

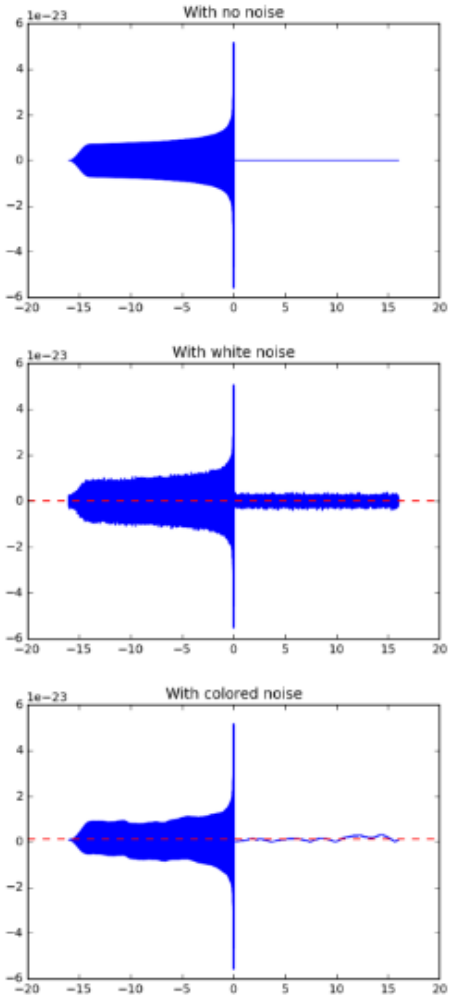

FIG. 9.— The original modified template (top) followed by the template with white/Gaussian noise (middle) and finally colored/frequency dependent noise (bottom).

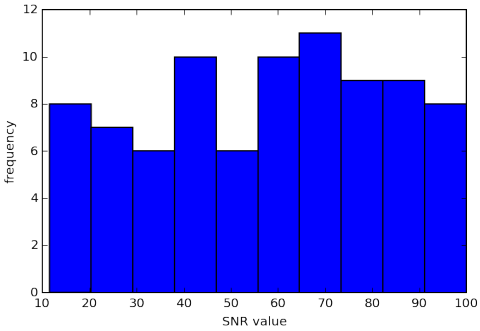

FIG. 10.— The SNR distribution from which the SNR for our mock data was generated.

models on this subject, such as that described in Section X.

VIII. CONCLUSION

We provided an overview of the calibration methods used for aLIGO and their uncertainties, describe how aLIGO detects GWs, and how CEs in the DARM closed feedback loop transfer function impact our detections. We note that the optical gain uncertainty was the most severe. We presented a toy model that generates mock data, match filters, stacks observations, and uses Bayesian parameter estimation to begin to explore the effects of calibration uncertainty on precision tests of GR. We suggested next steps in the investigation, including the implementation of a frequency-dependent calibration error model and a more sophisticated model set up in Section X.

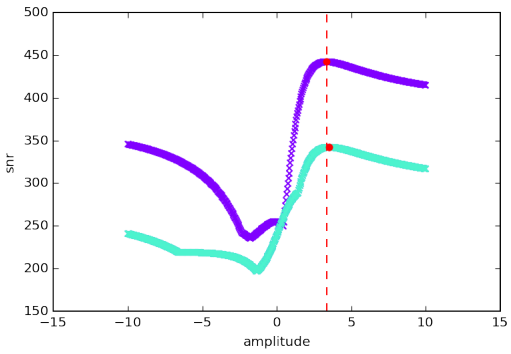

FIG. 11.— The SNR returned by the matched filtering function versus the selected template's amplitude. The true amplitude, 3.35, is marked by the red dashed line, and the maximum SNR is marked by a red circle.

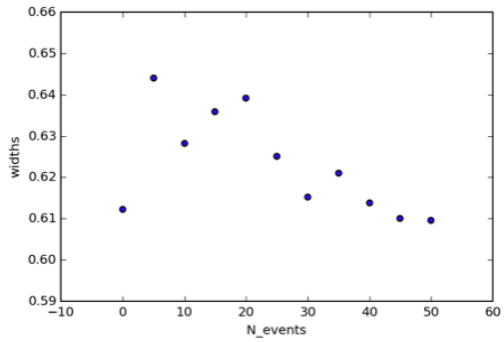

FIG. 12.— The number of stacked mock events versus the uncertainty of the parameter measurement. It looks as if the uncertainty reaches a constant value around 50 stacked events, but more runs would clarify the result.

The first GW has only recently been detected; the universe as illustrated by GWs is an emerging perspective in astronomy. Observing things in the universe for the first time is exciting and impactful, but verification that the detections are accurate is vital. Publishing uncertain/inaccurate detections can confuse and delay our understanding of the universe. Consequently, good calibrations of instrumentation is essential if we are to be confident about collected data and use it to understand how our universe works. Though only three GW detections have been observed, many more expect to be seen in the future. Analyzing our calibration uncertainties and their impact on important questions like the existence non-GR perturbation now will help us prepare our instrumentation for the next leg of direct GW detection astronomy.

IX. ACKNOWLEDGMENTS

The authors would like to acknowledge the support of the United States National Science Foundation (NSF) for the construction and operation of aLIGO and its and

the Caltech Institute of Technology's REU LIGO-SURF program.

X. APPENDIX

A more sophisticated investigation into the impact of CEs on precision tests of GR could use the the parameter estimation pipeline `LALInference` [18] and `TIGER` [5], a data analysis pipeline for testing the strong-field dynamics of GR with GW signals.

In this section, we outline astrophysical and calibration parameter estimation using `LALInference` and we describe the software signal injections used to mimic GR and non-GR conforming GW signals in `TIGER`.

`LALInference` compares a parametrized GW waveform model to the detected strain signal, using a more sophisticated matched filtering technique than that described in V. Using `LALInference`'s results, we can construct PDFs for each of the parameters in the GW detection, and stack them to increase our SNR, as also described in described in V, but for all parameters instead of one.

However, Bayesian inference tends to be computationally expensive, especially with a large number of parameters (15 for the most simple compact binary merger models, excluding instrumental and calibration parameters, see Table 1). It additionally has complex multi-modal likelihood functions, and the computationally costly process of generating the model waveforms [18]. As a result, stochastic sampling techniques, like Markov Chain Monte Carlo (MCMC), Nested Sampling, and MultiNest/BAMBI, have been explored and developed for Bayesian inference for GW data and have been packaged into `LALInference` to speed up the process. This work could use the Nested Sampling routine `LALInference_nest`, for example.

CEs, as discussed previously, will introduce bias in the posterior distributions of the measured parameters. `LALInference` currently uses a spline-fitted CE model, but a new frequency-dependent CE model is being developed a comparison between the effectiveness in parameter retrieval of two CE models is one possible next step. To compare CE models, we can use an effect size statistic to measure any shift in the mean of a parameter from its true injection value and weight it by the standard deviation:

$$\sigma_\theta \equiv \frac{\hat{\theta}_{true} - \hat{\theta}_{approx}}{\sigma_{\theta,approx}} \tag{18}$$

`TIGER` could be used for generating the common non-GR parameter. `TIGER` (Test Infrastructure for GEneral Relativity) is a data analysis pipeline for model-independent testing the strong-field dynamics of general relativity with GW signals [5]. It relies on Bayesian model selection to combine information from multiple observations. It then compares the stacked data evidence between two hypotheses: a GW waveform model consistent with GR, H_{GR}, and a model with parametrized deformations of the GW waveform model, H_{modGR}, as given by additional parameters. TIGER uses an odds ratio to compare these models:

$$O_{GR}^{modGR} \equiv \frac{P(H_{modGR}|d,I)}{P(H_{GR}|d,I)} \tag{19}$$

where d is the data and I is our prior.

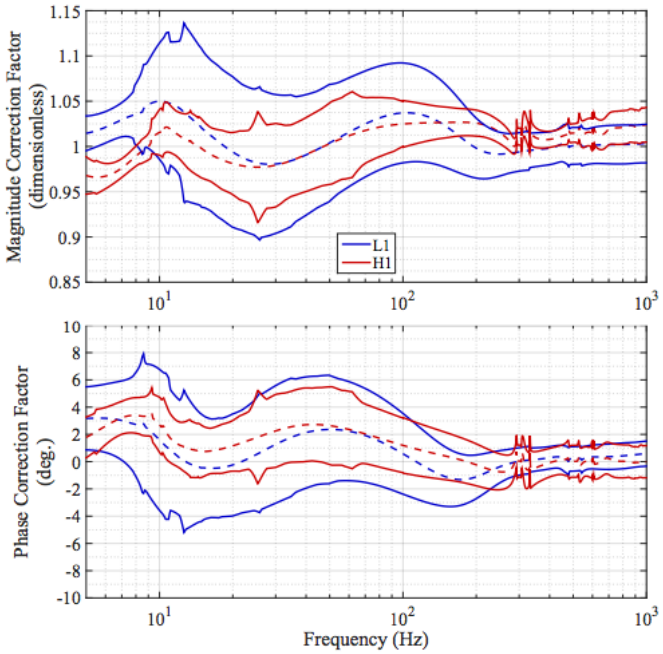

FIG. 13.— Figure from [15]. Differences in model and measured $R(f;t)$ result in correction factors, $1 + \delta R(f;t)/R(f;t)$, shown in magnitude (upper) and phase (lower). The dashed lines show the systematic errors, known from model parameters, and the solid lines the total statistical uncertainty [15]. This $R(f;t)$ incorporates the frequency dependent calibration model that is being developed for `LALInference`.

This method is considered model independent because any/all of the additional parameters are allowed to vary from zero (where they agree with GR) such that many different waveforms could be well fit. Each possible waveform is considered a sub-hypothesis and the Bayes factors for all of the sub-hypotheses can be merged into a single odds ratio with which to compare the GR consistent model.

Detector noise can sometimes mimic GR violations. To allay this, the odds ratio should be compared with a noisy background distribution; injecting many simulated GW signals with different astrophysical parameters into data surrounding the GW signal of interest can accomplish this [5]. The odds ratio can then be calculated for many GR consistent injections/noisy background sets. Then, a distribution of the odds ratio for GR consistent GWs can be calculated with an accompanying p-value. From this a threshold can be set for non-GR conforming GW model odds ratios to overcome.

For signal injections, `LALInference_pipe`, part of the `LaLApps` Suite [11], could be used. On such method could be to select a section of aLIGO data free of signals, undue noise, or other injections, and then add ~ 50 GW injections from systems with total masses of $10 - 50 M_\odot$ and spins $0 - 0.99$. A second set of these injections could be constructed to share a common non-GR parameter. This would allow us to estimate how many injections/detections are required to constrain a non-GR parameter if one exists.

This more sophisticated analysis would be a great follow-up to our introductory toy model, and would provide better insight to how limited aLIGO is in detecting small, non-GR perturbations. Such a detection would

indeed be a very important development in physics.

REFERENCES

[1]J. Abadie, B. P. Abbott, R. Abbott, M. Abernathy, C. Adams, R. Adhikari, P. Ajith, B. Allen, G. Allen, E. Amador Ceron, and et al. Calibration of the LIGO gravitational wave detectors in the fifth science run. *Nuclear Instruments and Methods in Physics Research A*, 624:223–240, December 2010.

[2]B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, 116(6):061102, February 2016.

[3]B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo. *Living Reviews in Relativity*, 19, February 2016.

[4]B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Tests of General Relativity with GW150914. *Physical Review Letters*, 116(22):221101, June 2016.

[5]M. Agathos, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, J. Veitch, and S. Vitale. TIGER: A data analysis pipeline for testing the strong-field dynamics of general relativity with gravitational wave signals from coalescing compact binaries. *Phys. Rev. D*, 89(8):082001, April 2014.

[6]S. Babak, R. Biswas, P. R. Brady, D. A. Brown, K. Cannon, C. D. Capano, J. H. Clayton, T. Cokelaer, J. D. E. Creighton, T. Dent, A. Dietz, S. Fairhurst, N. Fotopoulos, G. González, C. Hanna, I. W. Harry, G. Jones, D. Keppel, D. J. A. McKechan, L. Pekowsky, S. Privitera, C. Robinson, A. C. Rodriguez, B. S. Sathyaprakash, A. S. Sengupta, M. Vallisneri, R. Vaulin, and A. J. Weinstein. Searching for gravitational waves from binary coalescence. *Phys. Rev. D*, 87(2):024033, January 2013.

[7]R. W. P. Drever, J. L. Hall, F. V. Kowalski, J. Hough, G. M. Ford, A. J. Munley, and H. Ward. Laser phase and frequency stabilization using an optical resonator. *Applied Physics B*, 31(2):97–105, 1983.

[8]R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *ApJ*, 195:L51–L53, January 1975.

[9]K. Izumi and D. Sigg. Advanced LIGO:Length Sensing and Control in a Dual Recycled Interferometric Gravitational Wave Antenna. *LIGO Public Technical Document LIGO-P1500277-v3*, 2016.

[10]J. Mizuno, K.A. Strain, P.G. Nelson, J.M. Chen, R. Schilling, A. Rdiger, W. Winkler, and K. Danzmann. Resonant sideband extraction: a new configuration for interferometric gravitational wave detectors. *Physics Letters A*, 175(5):273 – 276, 1993.

[11]LIGO Observatory. Inspiral search programs, 2016.

[12]M. W. Regehr, F. J. Raab, and S. E. Whitcomb. Demonstration of a power-recycled Michelson interferometer with Fabry-Perot arms by frontal modulation. *Optics Letters*, 20:1507–1509, July 1995.

[13]D. Sigg. Strain Calibration in LIGO. *LIGO Public Technical Document T970101-x0*, 2003.

[14]A Staley, D Martynov, R Abbott, R X Adhikari, K Arai, S Ballmer, L Barsotti, A F Brooks, R T DeRosa, S Dwyer, A Effler, M Evans, P Fritschel, V V Frolov, C Gray, C J Guido, R Gustafson, M Heintze, D Hoak, K Izumi, K Kawabe, E J King, J S Kissel, K Kokeyama, M Landry, D E McClelland, J Miller, A Mullavey, B OReilly, J G Rollins, J R Sanders, R M S Schofield, D Sigg, B J J Slagmolen, N D Smith-Lefebvre, G Vajente, R L Ward, and C Wipf. Achieving resonance in the advanced ligo gravitational-wave interferometer. *Classical and Quantum Gravity*, 31(24):245010, 2014.

[15]The LIGO Scientific Collaboration and B. P. Abbott. Calibration of the Advanced LIGO detectors for the discovery of the binary black-hole merger GW150914. *ArXiv e-prints*, February 2016.

[16]The LIGO Scientific Collaboration, the Virgo Collaboration, B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, and et al. Binary Black Hole Mergers in the first Advanced LIGO Observing Run. *ArXiv e-prints*, June 2016.

[17]D. Tuyenbayev, S. Karki, J. Betzwieser, C. Cahillane, E. Goetz, K. Izumi, S. Kandhasamy, J. Kissel, G. Mendell, M. Wade, A. J. Weinstein, and R. L Savage. Improving LIGO calibration accuracy by tracking and compensating for slow temporal variations. *LIGO Document LIGO-190 P1600063*, 2016.

[18]J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D*, 91(4):042003, February 2015.

[19]S. Vitale, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, I. Mandel, B. Aylott, and J. Veitch. Effect of calibration errors on Bayesian parameter estimation for gravitational wave signals from inspiral binary systems in the advanced detectors era. *Phys. Rev. D*, 85(6):064034, March 2012.