Understanding and improving the accuracy of Advanced LIGO Calibration

M. McIntosh[1]

Advised by C. Cahillane[2], A. Weinstein[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

June 28, 2016

1. First Progress Report Questions

Write in some detail the motivation for your project. It should include background and an overview of the ongoing work in the laboratory. You should include references.

See Section 2 for motivation and background (it is a draft of the introduction for my final paper). Ongoing work includes reading TIGER and testing-GR papers as recommended by Alan.

Discuss the problem you are working on and explain how it fits into the ongoing work. Explain your approach and outline the methods you expect to use.

I am working on running parameter estimation through the LALInferencemcmc pipeline for various calibration parameters. This is the heart of the project as it is how I will recover my parameter estimations after forming software signal injections. My approach includes using online documentation and my mentors' recommendations on running instructions and parameter values.

Discuss the progress you have made on your project, your goals for the next month, and the methods or approach you will use to reach your research goals.

In the last two weeks, I have worked on/attended/been introduced to:

- introductory lecture on gravitational wave astronomy
- parameters for compact binary coalescence/Kerr parameters, black hole perturbation theory
- intro lecture on linear time invariant systems control, convolution theory
- lectures on LIGO instrumentation

- lecture on ipython notebook, binder, signal processing tutorials; masses in units of seconds
- read relevant LIGO instrumentation papers
- introduction to the parameter estimation pipeline LALInference with Alan & Kent
- ran lalinfmcmc, varying only neff; neff5000 through 500 (discovered a hard cut of at neff2000)
- created condor submit files and a dag file by hand
- python plots of the results

My ipython notebook is uploaded on my GitHub: `https://github.com/mmcintosh27/LIGO` and Figure 1 shows the one of plots I made of lalinfmcmc at various neff to determine if the parameter estimates varied at a function of neff, but it doesn't look like they do.

Fig. 1.— More information about the runs are on `https://ldas-jobs.ligo.caltech.edu/~melissa.mcintosh/neff1000/posplots.html` and urls of the like.

My goals for next month include finishing my literature review, selecting my calibration focus, understanding the calibration uncertainty to a greater extent, learning how to modify current calibration models, developing methods to propagate calibration uncertainty for astrophysical and precision-GR parameters, quantitatively evaluating systematic and statistical errors on these parameters, and estimating the contributions from the calibration uncertainties in impacting these errors.

What are the challenges and problems you have met so far and what challenges and problems do you anticipate?

Running lalinfmcmc seems to be done with an “ini” file, but system updates has caused the file to not be functional for the past two weeks. In its place I have been using the command line to run lalinfmcmc and creating sub files for condor and dag files by hand. This is not as straightforward but also has the benefit of a lower-level understanding of the function of the ini file. In the upcoming weeks I anticipate the difficulty of “stacking” injected signals and understanding ways of testing GR.

What resources will you require?

A patient mentor and perhaps some tech support.

Questions I have

Am I injecting signals into simulated or real data or both? When I'm looking at the ringdown parameter, am I also looking at higher order mode oscillations or is that entirely different? Am I creating injections or using them from a library? I've seen spinTaylor, TaylorF2, TaylorT4 and effective-one-body mentioned.

In [5], “ Similarly, model selection is not much affected by calibration errors. Fig. 4 shows the effect of calibration errors, modeled exactly as in [65], on the log odds ratio background distribution. As expected, the effect is minor ... and calibration errors will not affect the performance of TIGER.”

So why am I focusing on calibration errors and using TIGER?

The rest of this document is a start on my final paper.

2. Introduction

In 1915, Einstein published his General Theory of Relativity (GR). This theory and his following papers predicted the existence of gravitational waves (GWs). In 1993, a Nobel Prize went to Hulse and Taylor [9] who discovered pulsar system losing energy at the same rate predicted by GW emission, and thus implying the existence of GWs. Then in 2015, direct detection of GW150914, which was identified as a result of a binary black hole merger, occurred at the Advanced Laser Interferometer Gravitational Wave Observatory (aLIGO) detector network [2]. With this detection, studying GW observations became feasible.

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 aLIGO (on the order of 4×10^{-20}m [2]) the sensors and signal analysis needs to be precise. Characterizing and reducing uncertainties in aLIGO data allows us to reclaim as much physical information from the GW signal (measured in 'strain') as possible, increasing the detail of any GW analysis.

The two sources of uncertainty in aLIGO are statistical and systematic uncertainty. Though statistical uncertainty, a zero-mean Gaussian distribution strain measurement variation, is unavoidable, it can be reduced by taking additional strain measurements. Systematic uncertainty is relatively avoidable and cannot be reduced by additional measurements. Systematic uncertainty stems from incorrect characterization of the detector and causes calibration errors. 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 calibration errors are the focus of this project.

Calibration errors are errors that pertain to the conversion of instrumental readout to GW strain and are subsumed in the differential arm length transfer function (DARM) of aLIGO. Previous works that have forayed into calibration error analysis include [13, 15]. Current calibration errors for aLIGO are estimated as an overall constant in wave amplitude and phase, but a new, frequency dependent estimation method also is discussed here.

This work has three goals:

1. Describe the calibration methods used for aLIGO and their uncertainties
2. Estimate the effects of this calibration uncertainty on astrophysical parameters such as the source distance, sky location, and the progenitor masses and spins
3. Estimate the effects of this calibration uncertainty on precision tests of GR

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

The layout of this paper is as follows. First, we describe the differential arm length (DARM) closed feedback loop transfer function, which contains the GW wave signals and calibration errors for aLIGO. Next we describe how we estimate astrophysical and calibration parameters using the parameter estimation pipeline `LALInference` [14]. Then we relate previous research on the impact of calibration errors on this parameter estimation method. Afterwards, we describe the software signal injections used to mimic GR and non-GR conforming GW signals and their use in `TIGER` [5], a data analysis pipeline for testing the strong-field dynamics of GR with GW signals. We then attempt to recover the parameters used to generate the software signal injections using `LALInference` with both a constant calibration error model and also a frequency dependent calibration error model. Finally, we discuss the effectiveness the two calibration error models have on recovering astrophysical parameters and precision GR test parameters (in particular, we focus on a single GW ringdown parameter).

3. DARM Closed Feedback Loop Transfer Function

aLIGO is a complex and cutting edge instrument; it consists of a modified Michelson interferometer with Fabry-Perot arms that uses power-recycling mirrors and resonant sideband extraction that allow it to measure minuscule phase propagation differences via the Pound-Drever-Hall technique [8, 10, 11]. However, we follow [15] in this analysis and reduce the entire instrument to a sensor with a single degree of freedom: differential arm length (DARM) perturbations. These differentials contain GW, error, and other signals and are measured continuously in a closed feedback loop. The purpose of this feedback loop (see the reduced block schematic in Figure 2) is to recenter the mirrors used in aLIGO after the arms have been perturbed so that constructive interference of the laser is maintained. This allows the instrument to measure the next arm length differential as quickly as possible.

Each component of the feedback loop is described by a transfer function and the uncertainty on the overall loop transfer function yields the calibration error on the GW strain detection. The length response function/overall loop transfer function of the closed

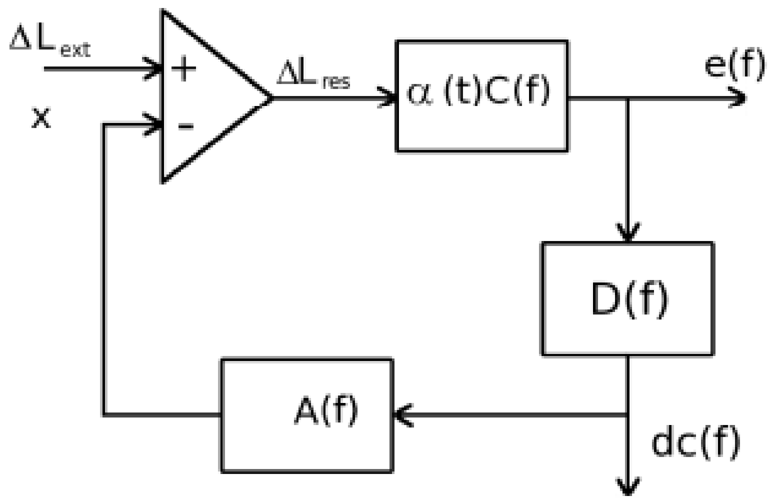

Fig. 2.— Figure from [15]. This reduced box schematic shows that the interferometer output is dependent on the performance of the feedback loop such that the transfer functions of all of the components in the feedback loop are necessary to reconstruct the GW signal. The subsystems are described more thoroughly in [15, 13].

loop feedback system (ignoring slow time dependency) is given by

$$R(f) = \frac{1 + G(f)}{C(f)} \tag{1}$$

where $G(f) = A(f)C(f)D(f)$, or the open loop gain of the system. Here, $C(f)$ is the transfer function of the arm cavity or the sensing function, $D(f)$ is a digital filter, and $A(f)$ is the actuation function that corrects mirror position. This equation is derived and its components described more thoroughly in [15, 13]. More complex calibration loop treatments are given in [1, 12]

Knowing the length response function, the GW strain can be calculated by

$$d(f) = \frac{\Delta L_{ext}}{L} = \frac{R(f)e(f)}{L} \tag{2}$$

where $e(f)$ is the error signal/detector output given by the DARM feedback loop, ΔL_{ext} the external length perturbation caused by the GW, and L the length of the arm of the interferometer in the absence of perturbations.

Equation 1 gives the theoretical or exact length function, but the measured length function includes calibration errors. Because the measured length function differs from the exact one, the interpreted strain will be different from the true strain in both phase and amplitude. These calibration errors affect the signal-to-noise ratio (SNR) of potential GW signals found in detection pipelines, though only to second order terms [7]. This affects both the precision measurement of astrophysical parameters like masses, sky location, distance, inclination, and orientation and also the measurement of universal parameters like those that describe variations form GR. To decrease the difference between the measured length function and the exact length function, we seek to better characterize aLIGO's calibration errors through Bayesian parameter estimation.

4. Bayesian Parameter Estimation with LALInference_mcmc

Using the parameter estimation pipeline `LALInference` [14] in a similar way to [DCC: P1500105] we compare a parametrized GW waveform model to the detected strain signal. Figure 3 shows the model used for GW150914. This matched filtering technique using template banks, further described in [6], is an accurate and time sensitive method to identify potential gravitational waves.

Using `LALInference`'s results, we can construct probability density functions (PDFs) for each of the parameters in the GW detection. To be explicit, we reiterate part of Bayes' theorem; the probability that a parameter, θ, is the correct value given some data, x, is proportional to the probability of getting the data given the parameter times the probability that the parameter is the correct value:

$$P(\theta|x) \propto P(x|\theta) \times P(\theta) \tag{3}$$

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

$$\begin{aligned} P(\theta|x,y) &\propto P(x,y|\theta) \times P(\theta) \\ &\propto P(y|\theta,x) \times P(\theta|x) \quad \text{We can now substitute in Eqn. 3} \\ &\propto P(y|\theta) \times P(x|\theta) \times P(\theta) \quad \text{Because GWs are uncorrelated} \\ &\quad\quad \text{the probability of } y \text{ does not depend on } x \text{ so we drop } x \end{aligned} \tag{4}$$

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 constant of these models are typically ignored in favor of simply comparing two competing models by taking the ratio (called the odd's ratio) of posterior probabilities to the evidence/potential GW signal:

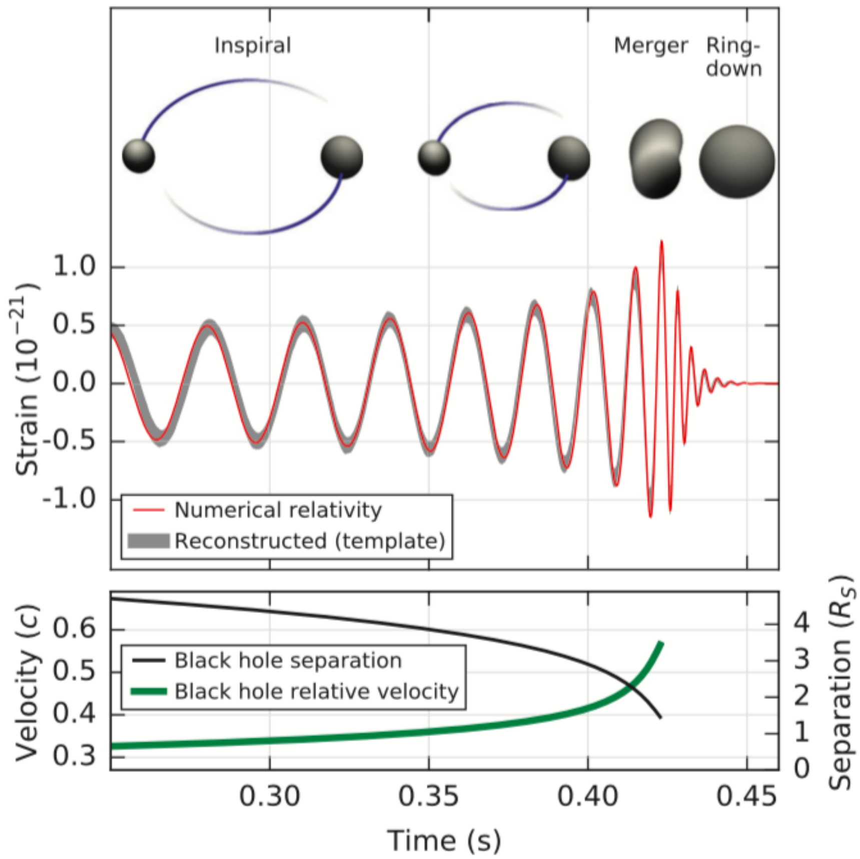

Fig. 3.— Figure from [2]. According to GR, two objects in orbit will slowly spiral inwards due to a loss of energy and angular momentum via GWs. The frequency and amplitude of the emitted GWs increases as the orbital distance between the objects shrink. When the objects finally merge, if they are BHs, they radiate GWs as a superposition of quasinormal ringdown modes, with one mode eventually dominating 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.

$$O_{i,j} = \frac{P(\theta_i|x)}{P(\theta_j|x)} B_{ij} \tag{5}$$

where θ_i is some parameter model and is compared to another parameter model, θ_j. The Bayes factor or evidence ratio, B_{ij}, is the ratio of likelihoods between the models. It is often used as a statistic describing how confident we are that the model is correct or to rank competing hypotheses given the observed data. From equation 4 we see that as we amass new GW detections we update the probability by multiplying the detections together. For an arbitrary number of detections, the probability of the parameter is then given by:

$$P(\theta|x_i) \propto \left[\prod_i P(x_i|\theta)\right] \times P(\theta) \tag{6}$$

Stacking GW strain signals in this way can constrain parameter estimation better than a single detection can. However, Bayesian inference tends to be computationally expensive, due to a large number of parameters (15 for the most simple compact binary merger models, excluding instrumental and calibration parameters), complex multi-modal likelihood functions, and the computationally costly process of generating the model waveforms [14]. 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 specifically uses the MCMC routine `LALInference_mcmc`.

In short, `LALInference_mcmc` subtracts the input GW model from the potential GW signal strain and compares the result with Gaussian white noise [14]. It uses a MCMC to step through the 15+ parameter space, using Bayesian inference to calculate the likelihood of the subtracted signal to be Gaussian white noise as a function of the step's parameter values. `LALInference_mcmc` then takes the most probable template and outputs the PDFs of those parameters. With the PDFs we can compute parameter expectations and confidence intervals.

5. Potential affect of Calibration Errors on Bayesian Parameter Estimation

6. Software Injections

Though up to 100's of observations per year are expected by the time LIGO operates at design sensitivity [3], only two GW detections have been published as of the time of this paper. For now, we use simulated signals injected by software instead of real GW signals. We construct these injections to share a common non-GR parameter for us to constrain in hopes that the technique can be used on real data when more GW detections are available. Additionally, we can construct confidence intervals from these PDFs as a function of the number of injections/detections. This will allow us to estimate how many injections/detections are required to constrain the non-GR parameter in future studies.

GW waveform models are based off of an analytical inspiral-merger-ringdown (IMR) model and usually calibrated against waveforms from direct numermical integration of the Einstein equations [4]

[how injections/waveforms are made/what libraries we take them from]

TIGER (Test Infrastructure for GEneral Relativity) [5] is a data analysis pipeline for model-independent testing the strong-field dynamics of general relativity with GW signals. It relies on Bayesian model selection to combine information from multiple observations. It then compares the stacked data to both a GW waveform model consistent with GR and a model with parametrized deformations of the GW waveform model, as given by additional parameters. 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.

In this work, we focus on a single-parameter analysis in the merger-ringdown regime. We fix all other parameters to be consistent with GR.

7. Parameter Recovery and Error Propagation

8. Comparison of Calibration Error Models and effectiveness of Parameter Recovery

9. Conclusion

Next steps could include expanding our single-parameter analysis to a multi-parameter analysis

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 and it can happen in the excitement of new and big discoveries. 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.

REFERENCES

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, Dec. 2010.

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, Feb. 2016.

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, Feb. 2016.

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.

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, Apr. 2014.

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, Jan. 2013.

D. A. Brown and LIGO Scientific Collaboration. Testing the LIGO inspiral analysis with hardware injections. *Classical and Quantum Gravity*, 21:S797–S800, Mar. 2004.

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.

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

J. Mizuno, K. Strain, P. Nelson, J. 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.

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.

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

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*, Feb. 2016.

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, Feb. 2015.

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, Mar. 2012.

This manuscript was prepared with the AAS LaTeX macros v5.2.