Bayesian Inference for Fast Scattering Glitches

Aislinn D. McCann
California State University, Northridge

Mentors: Rhiannon Udall and Derek Davis
LIGO, California Institute of Technology, Pasadena

Second Interim Report
LIGO Caltech SURF Program 2023

Data collected by gravitational wave (GW) interferometers such as the Laser Interferometer Gravitational-wave Observatory (LIGO) is permeated by noise as a result of environmental interference. Parameter estimation pipelines such as Bilby used to analyse LIGO data employs Bayesian inference, which assumes that the noise in GW data is Gaussian and stationary: an assumption contradicted by the nature of non-Gaussian transient noise "glitches" prevalent within the data. We have constructed a mathematical model that emulates the waveform of fast scattering glitches, which was tested via Bilby to determine the efficacy of glitch mitigation under the basis of the model. The implementation of this model will facilitate the efficient subtraction of real fast scattering glitch instances from GW strain data, allowing for improved analysis and signal detection for future observing runs.

I. INTRODUCTION

The Laser Interferometer Gravitational-wave Observatory (LIGO) is an observatory designed to detect gravitational waves (GWs) via the utilisation of two Fabry-Perot interferometer arms. Within the detector is a beam splitter which sends a laser through each arm, allowing the laser to cycle and rejoin to be analysed by a photodetector. When GWs pass through a detector, each arm experiences a slight displacement which creates instances of constructive or destructive interference from the recombined beam, thereby inducing a phase shift which is then converted into a measurable signal [1]. A high sensitivity is required for all GW detectors to receive data from distant sources such as compact binary coalescences (CBCs), which consequently hinders data analysis by also increasing the prevalence of persistent and short duration transient noise "glitches" produced by various sources of environmental interference or electronic malfunction [1–3].

One form of glitch known as scattered light glitches are the result of beam segments diverging from the main beam path and reflecting from objects of a conflicting relative velocity within the interferometer, which later rejoin the main beam and produce an additional phase shift [3]. Scattered light glitches present two main complication in analysis. One such impediment is that glitch instances may trigger false positives in GW search pipelines. The second and more common difficulty is that glitches may overlap on top of an existing signal, presenting the largest hindrance to analysis efforts. Our focus is to mitigate instances of fast scattering glitches, a form of scattering glitch which occurs as a result of increased ground activity in the anthropogenic band ($1-5$ Hz) and microseism band ($0.1-0.3$ Hz). Each of these sources affect the detector's sensitivity in the frequency band between 10 and 50 Hz [1]. Figure 1 provides an example of the short duration noise bursts characteristic of fast scattering glitches.

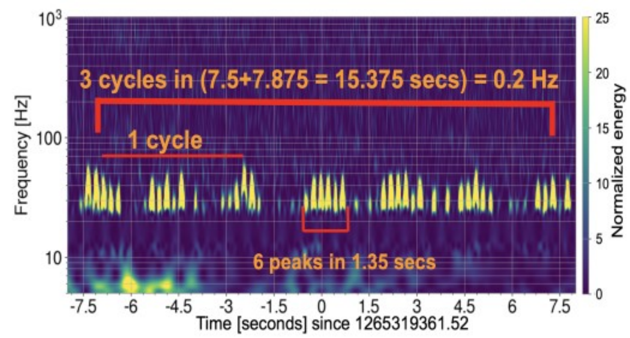

FIG. 1. A spectrogram of fast scattering triggers generated using the Q-transform. Fast scattering glitches occur as multiple sub-arches organised in a shape akin to a larger arch. Image reproduced from [1].

The process of removing noise and glitches from GW strain data has been a persistent effort in order to improve detector sensitivity. Furthermore, removing glitches from data improves the accuracy of CBC parameter estimation pipelines which analyse raw strain data collected by detectors to infer astrophysical properties that characterise GW sources [3]. One such pipeline is Bilby, a Python code which utilises Bayesian inference in order to perform accurate parameter estimations [4].

Bayesian inference incorporates Bayes' theorem to produce the posterior probability distribution of GW source parameters by incorporating the prior distribution of these source parameters with a model hypothesis. The posterior probability may be computed using Bayes' theorem with data d and source parameters θ [2, 5]:

$$p(\theta|d,\mathscr{M}) = \frac{\mathscr{L}(d|\theta,\mathscr{M})\pi(\theta|\mathscr{M})}{\mathcal{Z}(d|\mathscr{M})}, \tag{1}$$

where $\mathscr{L}(d|\theta, \mathscr{M})$ is the likelihood, $\pi(\theta|\mathscr{M})$ is the prior probability, and $\mathcal{Z}(d|\mathscr{M})$ is the model evidence, each given a model $\mathscr{M}$.

Parameter estimation pipelines such as Bilby assume GW noise data to be stationary and Gaussian [3]. The likelihood for transient behaviours present in GW strain data is thus expressed using the following Gaussian noise likelihood $\mathscr{L}$, with a data value d_k at a frequency bin index k [2, 4]:

$$\ln \mathscr{L}(d|\theta) = -\frac{1}{2}\sum_k \left\{ \frac{[d_k - \mu_k(\theta)]^2}{\sigma_k^2} + \ln(2\pi\sigma_k^2) \right\}, \quad (2)$$

where σ_k is the amplitude spectral density for the noise at a given frequency bin and $\mu_k(\theta)$ is the waveform in that frequency bin. The non-Gaussianity of transient glitches contradicts this assumption, further demonstrating the importance of producing a means to remove these triggers from GW data.

II. OBJECTIVE

We have constructed a model which provides a baseline to identify fast scattering glitches from GW data and mitigate these instances for improved analysis. Because the model characterises long duration scattering glitches, we performed model testing through Bilby. Bilby provides a more reliable method of both the subtraction and marginalisation of long duration glitches as opposed to other Bayesian inference algorithms such as BayesWave, which are more proficient in subtracting short duration glitches. Modelled inference performed by such algorithms provides a more robust probe of glitch morphology, incorporating information on the nature of these glitches to assess the presence of unseen sub-arches for the case of slow scattering glitches and additional arches within fast scattering glitch clusters [2]. The construction of this model allows us to emulate the behaviour and conditions of triggers produced by fast scattering glitches by inferring their parameters and evaluating the likelihood that a particular set of configurations may approximate a fast scattering glitch as seen in GW strain data.

III. PROGRESS

A. Preliminary Work

The first few weeks of the program were devoted solely to training and attaining a better understanding of GWs, the LIGO detector, and the various forms of noise that permeate GW source data. My training began by undergoing a tutorial to familiarise myself with the Python coding procedures involved in performing computations associated with an inspiral binary system of given masses, including the orbital separation of the two objects, the orbital period and velocity associated with this distance, as well as the orbital frequency and the corresponding GW frequency. From these values, I calculated and plotted the rate of energy loss and the GW strain consistent with this theoretical binary system. Additionally, I attended a two-day GW Open Data Workshop in which I learned how to create a spectrogram to display the time and frequency information for a GW signal produced via the Q-transform, as well as how to plot a LIGO noise curve and the various ways noise may obscure a true GW signal. In particular, I also learned about how parameter estimation is performed given a GW waveform, a lecture which I found to be particularly enlightening for the purposes of my project. The skills and knowledge I have attained thus far will be essential when working towards our objective of constructing a fast scattering waveform model which may be used as a means of glitch subtraction from true GW data.

1. Motivation

In order to move forward in this project, it is necessary to possess a thorough understanding of the construction of the LIGO detector and the sources of the noise associated with GW data, as well as how noise may obscure a signal. Because scattered light from test mass mirrors within the LIGO detectors reflects from surfaces of conflicting relative motion and rejoin the main beam path, fast scattering glitches are a persistent issue within GW strain data and often conceal true signals. Additionally, fast scattering glitches persist as multiple driving frequencies interacting with one another. As our objective for this project is to use a mathematical model to reliably mitigate fast scattering glitch instances, it is thus necessary to understand the sources associated with the driving frequencies that compose fast scattering glitches in order to properly consider each within the model.

The tutorial and workshop series were an imperative step to become familiar with the relevant Python syntax in order to define and use a waveform. Figures 2 and 3 are an example waveform and corresponding spectrogram, respectively, generated for practice during the workshop series.

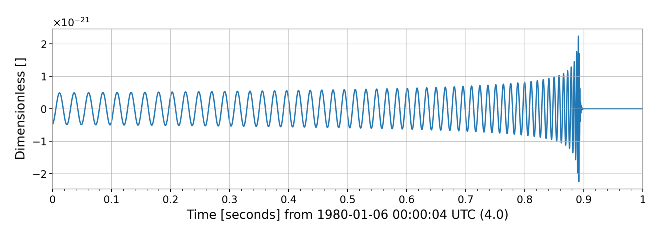

FIG. 2. The waveform of an example CBC signal produced during the GW Open Data Workshop series.

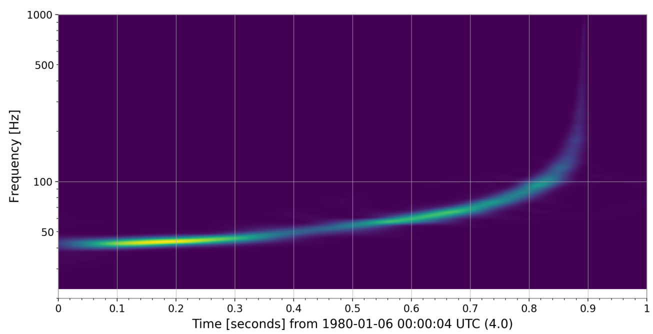

FIG. 3. The spectrogram of an example CBC signal generated during the GW Open Data Workshop series corresponding to Figure 2.

B. Methods

1. Model Construction

In constructing a mathematical model for a generic fast scattering glitch waveform, we began by following the form of the undermentioned equation that describes the excess strain noise $h(t)$ related to the motion of the surface $x(t)$ produced by scattered light of wavelength λ over time t [2]:

$$h(t) = \bar{A} sin\left[\frac{4\pi}{\lambda}x(t) + \phi\right], \tag{3}$$

where $\bar{A}$ is the amplitude of the noise produced by the glitch with a phase shift ϕ.

Assuming the movement of the relevant surface as a harmonic oscillator, its motion may be presented as such with the incorporation of the two driving frequencies f_1 and f_2, the two of which interact constructively and destructively to generate fast scattering glitches within data:

$$x(t) = A_1 sin(2\pi f_1 t) + A_2 sin(2\pi f_2 t), \tag{4}$$

where A_1 and A_2 are the respective amplitudes associated with the two driving frequencies.

Incorporating the above equation into Equation 3, we thus arrive at our model for a fast scattering glitch:

$$h(t) = \bar{A} sin\left[A_1 sin(2\pi f_1 t) + A_2 sin(2\pi f_2 t) + \phi\right]. \tag{5}$$

2. Model Testing: Spectrogram Tests

Testing the validity of the model began by determining predictions for each driving frequency and their associated amplitudes by incorporating the following relation between the wavelength of the scattered light beam λ, its frequency f, and velocity of propagation $v(t)$:

$$f = \left|\frac{v(t)}{\lambda}\right|. \tag{6}$$

We thus determine the velocity associated with the motion from Equation 4, and by simplifying for the case of the maximum frequency f_{max} where we may assume $cos(2\pi ft) = 1$:

$$A_1 f_1 + A_2 f_2 = f_{max}. \tag{7}$$

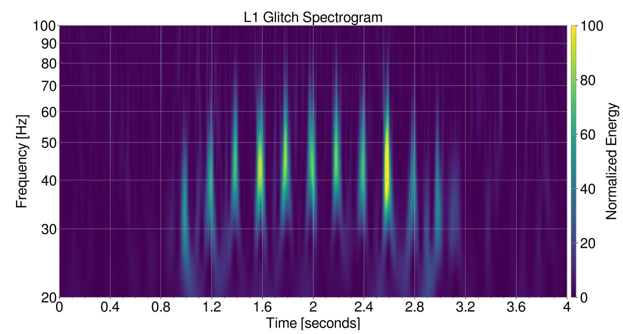

FIG. 4. A spectrogram produced by the mathematically derived fast scattering model with parameters $\bar{A} = 3 \times 10^{-22}$, $f_1 = 0.2$, $f_2 = 5$, $d_1 = 15$, $d_2 = 35$, and $\phi = 0$, where $d_1 = \frac{A_1}{f_1}$ and $d_2 = \frac{A_2}{f_2}$.

Because fast scattering glitches affect the detector sensitivity from $10 - 50$ Hz, we assume that $f_{max} = 50$ when determining the possible values of each parameter associated with Equation 7. Figure 4 provides the resulting spectrogram for one of the possible parameter combinations of our fast scattering model, reproducing the long duration, short-burst arches characteristic of traditional fast scattering glitch cases.

3. Model Testing: Bilby

The next step in testing the efficacy of our model in fast scattering glitch emulation and mitigation involved determining the accuracy at which various given parameter injections align with the posterior results of each case. This was performed by running the model with our chosen values through Bilby given the logical priors for each parameter. Figure 5 displays a corner plot associated with a given set of injection parameters. The posterior probability distributions for each possible value between each parameter are shown with the specified injection values on each contour as a single point.

Further testing of the validity of the glitch model involved generating a spectrogram from the injections on the basis of the model produced by Bilby. The resulting

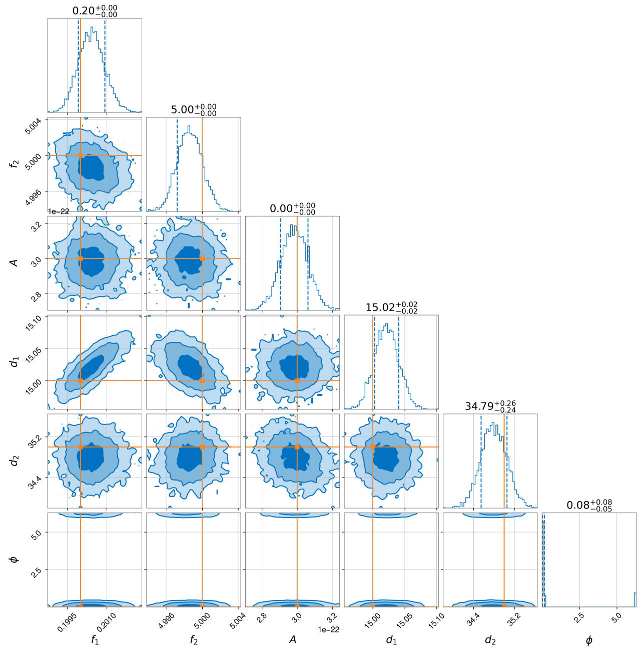

FIG. 5. A corner plot produced from Bilby given a set of injection parameters in association with those specified for Figure 4. The parameters d_1 and d_2 describe the amplitudes associated with the driving frequencies f_1 and f_2.

spectrogram was found to fall in close agreement with the spectrogram produced directly from the mathematical model, as shown in Figure 6.

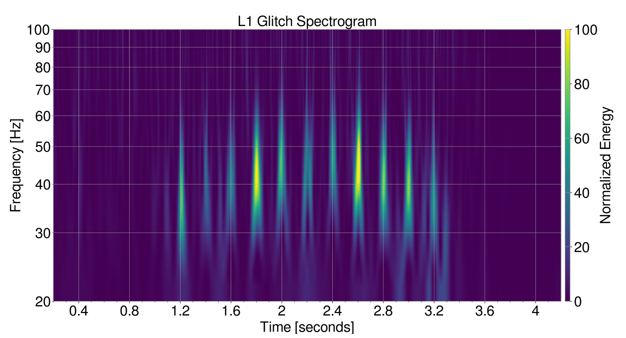

FIG. 6. The resulting spectrogram generated from Bilby given the previously specified set of injection parameters.

4. *Model Testing: P-P Tests*

The final validity test which we have performed thus far consisted of P-P testing, or parameter-parameter testing, to ensure unbiased posterior results given a set of samples taken from the prior. Figure 7 displays the associated P-P plot for a sample size of $N = 100$. The events lying within each confidence interval tend to closely follow a linear trend, suggesting that the results are indeed unbiased.

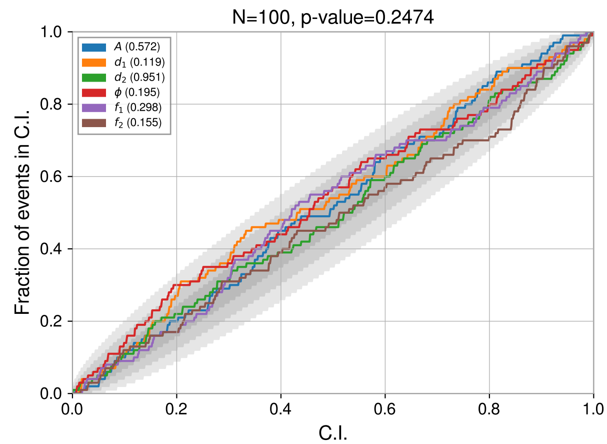

FIG. 7. The P-P plot produced for the model with a sample size of $N = 100$.

C. Next Steps

For the remainder of the summer, we intend to test our waveform model on real fast scattering glitch instances saturating GW strain data in order to determine its capability in glitch emulation and mitigation. If successful, such efforts will provide a reliable method to reduce glitch prevalence in data, thereby improving efforts to analyse GW source data and allow for greater detector sensitivity for future observing runs. Furthermore, if time allows, we intend to perform joint "CBC+glitch" injection testing to determine the model's success in uncovering obscured CBC strain data. The remainder of the summer will be spent preparing the final presentation.

[1] S Soni, C Austin, A Effler, RMS Schofield, G González, VV Frolov, JC Driggers, A Pele, AL Urban, G Valdes, et al. Reducing scattered light in ligo's third observing run. *Classical and Quantum Gravity*, 38(2):025016, 2020.

[2] Rhiannon Udall and Derek Davis. Bayesian modelling of scattered light in the ligo interferometers. *arXiv preprint arXiv:2211.15867*, 2022.

[3] Arthur E Tolley, Gareth S Cabourn Davies, Ian W Harry, and Andrew P Lundgren. Archenemy: Removing scattered-light glitches from gravitational wave data. *arXiv preprint arXiv:2301.10491*, 2023.

[4] Gregory Ashton, Moritz Hübner, Paul D Lasky, Colm Talbot, Kendall Ackley, Sylvia Biscoveanu, Qi Chu, Atul Divakarla, Paul J Easter, Boris Goncharov, et al. Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *The Astrophysical Journal Supplement Series*, 241(2):27, 2019.

[5] Isobel M Romero-Shaw, C Talbot, S Biscoveanu, V D'emilio, G Ashton, CPL Berry, S Coughlin, S Galaudage, C Hoy, M Hübner, et al. Bayesian inference for compact binary coalescences with bilby: validation and application to the first ligo–virgo gravitational-wave transient catalogue. *Monthly Notices of the Royal Astronomical Society*, 499(3):3295–3319, 2020.