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

First 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 intend to construct a model that emulates the waveform of fast scattering glitches and implement a refined iteration of the model into Bilby to determine the efficacy of glitch subtractions under the basis of the model. The implementation of this model will facilitate the subtraction of fast scattering glitch data 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) by converting phase shifts produced by GW sources into a measurable signal [1]. A high sensitivity is required for all GW detectors to receive data from GW sources, but this simultaneously hinders the collection of raw strain data accumulation by also increasing rates of persistent and short duration transient noise "glitches" produced by various sources of environmental interference or electronic malfunction [1–3].

Glitches are the result of scattered light diverging from the main beam path and reflecting from moving objects within the interferometer, which later rejoins the main beam and produces an additional phase shift [3]. Scattering glitches may either present as false instances of a GW signal or overlap on top of an existing signal, impeding further analysis. Of interest to us are fast scattering glitches, which occur 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.

The process of removing noise and glitches from GW strain data has been a persistent effort in order to improve the reliability of signal detection and detector sensitivity. Particularly, removing glitches from data is a requirement for the functionality of parameter estimation pipelines which analyse raw strain data collected by the 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 utilises 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]:

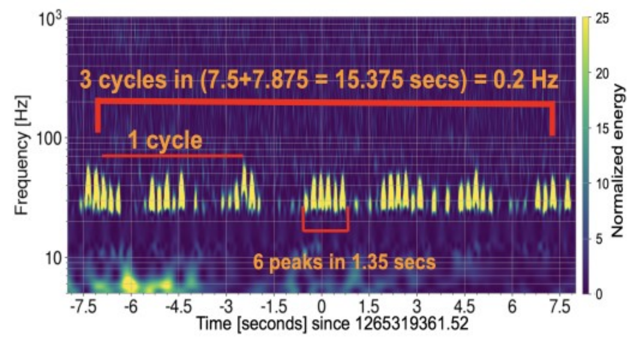

FIG. 1. A spectrogram of fast scattering triggers generated using the Q-transform. Fast scattering glitches occur as multiple sub-arches organised in a shape akin to a larger arch. Image reproduced from [1].

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

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 intend to construct a model which will provide a baseline to identify fast scattering glitches from GW data and test the model by implementing Bilby. Bilby provides a more reliable method of both the subtraction and marginalisation of long duration glitches–including scattering glitches–as opposed to other Bayesian inference algorithms such as BayesWave, which are more proficient in the subtraction of 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 will allow us to emulate the behaviour and conditions of data 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.

A successfully derived model for fast scattering glitches will allow us to better distinguish true GW signals from fast scattering glitches and accurately mitigate glitch instances in data, thereby providing a simpler means of improving data analysis and signal detection for future GW observation runs. The validity of the model will be tested on instances of fast scattering glitches present in preexisting GW data.

III. APPROACH

We will begin by mathematically deriving a series of models which produce a waveform similar to that seen for fast scattering glitches. After acquiring sufficient test candidates of these glitches, we will then compare these with the waveforms fabricated by the models and determine the accuracy at which they align with the waveform of these candidates. Upon doing so, we can isolate the model which most closely agrees with the candidate spectrograms. If we assume that the motion of the surface reflecting this scattered light is a simple harmonic oscillator, we may follow the form of the undermentioned generic scattering model related to the motion of the surface $x(t)$ which produces scattered light of wavelength λ over time t [2]:

$$h(t) = Asin\left[\frac{4\pi}{\lambda}x(t) + \phi\right], \quad (3)$$

where A is the amplitude of the noise produced by the glitch with a phase shift ϕ.

We intend to perform a derivation of Equation 3 applied for the case of fast scattering glitches by employing the same process used to determine the following waveform model for slow scattering glitches [2, 3]:

$$h(t) = Asin\left[\frac{f_{gl}}{f_{mod}}sin(2\pi f_{mod}t) + \phi)\right], \quad (4)$$

where f_{gl} is the maximum glitch frequency and f_{mod} is the frequency of the oscillating surface.

Unlike slow scattering glitches, whose waveform is governed by a singular frequency and presents itself as many harmonics that appear on top of each other, a fast scattering glitch persists with multiple driving frequencies. Each driving frequency must subsequently be measured in order to accurately identify and clean instances of fast scattering glitches from GW data.

After a final model is mathematically obtained and is verified to produce an accurate fast scattering waveform using various fast scattering glitch examples, we will apply it to datasets that are saturated with such glitches, the parameters of which are known. Studying the resulting posterior distributions will test the model's ability to clean the glitches from GW source data by injecting our custom model in place of the standard GW waveform template provided by Bilby [2].

IV. PROGRESS

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.

A. 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 greater 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 either constructively or destructively with one another. As our objective for this project is to mathematically model a reliable waveform of fast scattering glitches, 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 I completed were necessary to understand how I will proceed for the duration of this project. Because I intend to mathematically construct the basis of a fast scattering waveform and implement it into Bilby to determine the efficacy at which it emulates fast scattering glitch data, it was 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.

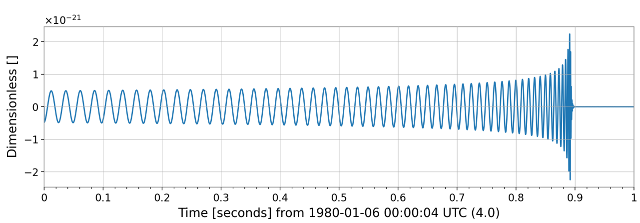

FIG. 2. The waveform of an example compact binary coalescence (CBC) signal produced during the GW Open Data Workshop series.

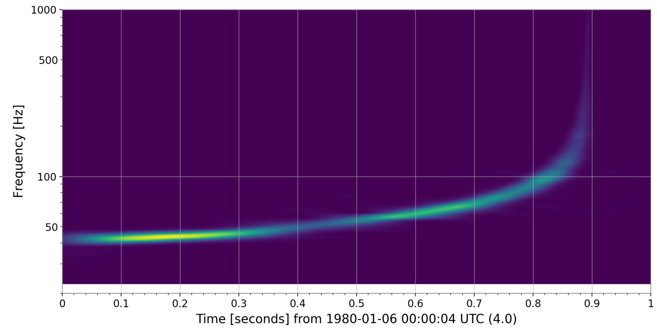

FIG. 3. The spectrogram of an example CBC signal generated during the GW Open Data Workshop series corresponding to Figure 2.

B. Challenges

While the work I have done on this project thus far has been strictly introductory, I have still faced some challenges while completing the needed preliminaries. One such difficulty was re-familiarising myself with Python coding procedures and syntax, a somewhat time consuming task. Furthermore, I found that I had undergone much trial and error in expanding upon my current Python knowledge with new information on performing computations related to defining a waveform and visualising GW signals. With much effort, I am gradually cultivating my skills in coding and improving my understanding of performing computations involved in GW research.

In the future, I expect to encounter more complications as I overcome the learning curve that may impede my progress in the work involved for this project. I expect that it will take several attempts in order to produce a waveform that provides a reliable model of a fast scattering glitch. Additionally, because the various iterations I will construct will each be injected into Bilby as a custom model in order to test their accuracy in relation to fast scattering examples, I believe that it will take some time to become familiar with the coding methodology associated with performing the injection and determining the validity of each model. I intend to better acquaint myself with Bilby syntax prior to performing the injections so as to avoid spending unnecessary time on solving this problem and thereby hindering future progress. Despite these challenges I anticipate, the experience I will gain in overcoming them will be very rewarding for my future as a scientist.

[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.