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

Project Proposal
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 that contradicts 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]. 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 glitches from GW strain data has been a long-standing effort in order to improve the reliability of signal detection as well as 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 characterising GW sources [3]. One such pipeline is BILBY, a Python code which uses 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]:

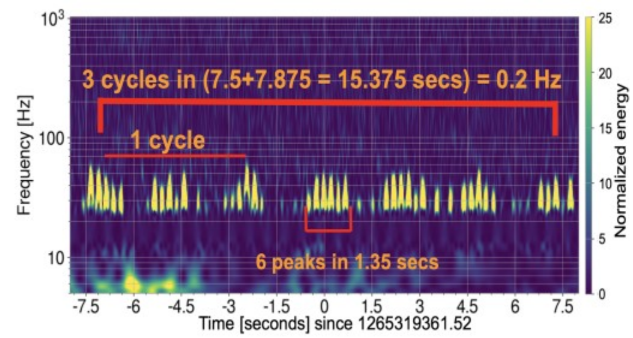

FIG. 1. A spectrogram of fast scattering triggers generated using the Q-transform. Fast scattering glitches occur as multiple sub-arches organised in a shape akin to a larger arch. Image reproduced from [1].

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

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 as 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]:

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

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. This assumption contradicts the non-Gaussianity of transient glitches, 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. The motivation for using BILBY is twofold: first, modelling scattered light glitches by way of Bayesian inference provides a more reliable method of subtracting glitches from detector data; second, modelled inference may also be able to test for the presence of unseen sub-arches which may otherwise be missed [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 subtract instances from 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 equation for the excess strain noise to model scattered light glitches over a time t [2, 3]:

$$h(t) = Asin\left[\frac{f_{gl}}{f_{mod}} sin(2\pi f_{mod} t) + \phi)\right], \tag{3}$$

where f_{gl} is the maximum glitch frequency, A is the amplitude of the noise produced by the glitch, and f_{mod} is the frequency of the oscillating surface.

The final model derived for this project, which will reproduce the waveform of fast scattering glitches, will serve as an extension of Equation 3. The aforementioned equation is specific to the waveform of slow scattering glitches. 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 modelled 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. TIMELINE

This project will persist over the duration of ten weeks. The first half of the project, tentatively, will consist of training and preparing data for analysis. The second half will involve execution and implementation. The following is a provisional project plan:

- Training & Data Preparation
  - Study BILBY syntax and coding procedure to obtain a posterior distribution when given a prior and likelihood
  - Learn how to test the validity of a model using BILBY parameter estimation
  - Collect test candidates that exhibit fast scattering glitches in GW data
- Execution
  - Mathematically derive models that emulate the waveform of fast scattering glitches
  - Compare the resulting waveforms of each model to true glitch data and eliminate those that do not agree until we achieve a final model
  - Test the validity and efficacy of the model by injecting it into BILBY and using it to subtract fast scattering glitches from GW data

Finally, the remaining weeks will be devoted to preparing the final presentation and report.

[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.