Crowdsourcing GWs: Emergent Properties of Gravitational Wave Searches

Oleksandra Lukina
University of South Dakota

Mentor: Derek Davis
LIGO, California Institute of Technology
(Dated: May 12, 2023)

Project Plan
LIGO Caltech SURF Program 2023

Determining whether a gravitational wave (GW) signal is of astrophysical origin or is caused by terrestrial noise still presents a challenge to the GW community. Current searches estimate the significance of events by calculating the false alarm rate (FAR) and p_{astro}, but these results are limited to a single search pipeline. In this work, we are going to investigate different ways of combining GW information to learn more about observed compact mergers. This will include analysing the correlation between different pipelines, running simplified searches, designing new methods of combining results from multiple pipelines, and testing whether they result in a meaningful estimate of astrophysical significance of events in realistic datasets.

I. INTRODUCTION

Since the first detection of gravitational waves (GWs) [1], the total number of GW candidates reported by LIGO-Virgo-KAGRA (LVK) Collaboration reached 90 [2] and will continue to grow [3]. At the same time, determining whether a certain signal has astrophysical origin or is caused by terrestrial noise still remains a challenge, which leads to the uncertainty in the number of detected compact mergers [4].

Noisy local environments, observations with multiple detectors, and inability to shield the instruments from GW signals results in the dependence of the estimated significance on a search analysis [5]. Two main approaches that are used to search for events include matching signals to the compact binary coalescence waveform templates and searching for transient signals across the network of detectors with minimal modeling. Variations of the first method are used in PyCBC [6], GstLAL [7], and IAS Search (IASS) [8] pipelines, and the second method is used in the cWB pipeline [2]. There are several technical differences between search pipelines, so they result in different estimations of the probability of an astrophysical origin. However, all pipelines are designed to search for the same signals, so their results are not fully independent and should correlate.

Fig. 1 illustrates the correlation between two search pipelines in a toy model described in [9]. Example pipeline 3 only uses the first half of the points in a segment of data, and pipeline 4 uses the second half of the points. Although these two pipelines are statistically independent, their combination shows a correlation. The figure shows that both pipelines can see about a half of the signals and their results differ, but the joint fit accounts for this bias and shows the total signal count.

In order to apply this reasoning to the calculation of the significance of events, we introduce two quantities, false alarm probability (FAP) and p_{astro}. FAP is a probability of observing a coincidence or a "false alarm" with a signal-to-noise ratio (SNR) higher than a certain value. As a result, to confirm the presence of a signal, one must show that the probability to obtain the observed event in a dataset that only contains noise is smaller than a given threshold [5]. It is also possible to convert FAP into a related quantity called the false alarm rate (FAR), which is measured in yr^{-1} and is usually included in GW catalogs, e.g., [2].

On the other hand, p_{astro} is defined as a probability that a GW candidate has astrophysical origin and is not caused by terrestrial noise. It is calculated by combining the rates at which triggers – outputs of a search pipeline – are generated by both astrophysical and noise sources, i.e., both false and true alarm rates [10].

p_{astro} can be described in terms of Bayesian statistics as suggested in [9]:

$$p_{astro}(x) = \frac{p(S|x, \Phi_s, \Phi_n)}{p(S|x, \Phi_s, \Phi_n) + p(\varnothing|x, \Phi_n)}, \tag{1}$$

where x is a trigger statistic, S is a signal hypothesis, $\varnothing$ is a noise hypothesis, and Φ_s and Φ_n are some signal and noise parameters. Thus, $p(S|x, \Phi_s, \Phi_n)$ and $p(\varnothing|x, \Phi_n)$ are posterior probabilities of signal and noise hypotheses, respectively.

Applying Bayes' theorem, one can rewrite equation (1) as follows:

$$p_{astro}(x) = \frac{\pi_s p(x|S)}{\pi_s p(x|S) + \pi_n p(x|\varnothing)}, \tag{2}$$

where π_s and π_n are prior probabilities of having a signal or noise and $p(x|S)$ and $p(x|\varnothing)$ are the corresponding likelihoods of getting a trigger x in a dataset containing signal or only noise.

In order to write a similar expression for a unified p_{astro}, we define $\vec{x}$ as a vector of triggers from multiple pipelines and obtain

$$p_{astro}(x) = \frac{\pi_s p(\vec{x}|S)}{\pi_s p(\vec{x}|S) + \pi_n p(\vec{x}|\varnothing)}. \tag{3}$$

It is important to note that, since pipelines are correlated, $p(\vec{x}|S)$ and $p(\vec{x}|\varnothing)$ are not independent and the relationship between them has to be determined separately.

II. OBJECTIVES

The goal of this project is to investigate different ways of combining gravitational wave information to learn more about the observed compact mergers. It will include the analysis of differences and similarities between search pipelines in order to compute the covariance of their search results. This result will become the basis for combining probabilities of event significance and will help us to contribute to the solution of the problem of the joined p_{astro}.

Expanding the illustrative work of [9], we plan to come up with an updated list of GW candidates that will include an estimation of their astrophysical probabilities and other properties from different sources. This work will help to better understand the emergent properties of already detected candidates and provide insights for new GW detections that will come in the O4 observing run that will have begun shortly before the start of this project.

III. APPROACH

To better understand the properties of the search pipelines, we are going to run a simplified version of one of the searches and begin with a binary analysis of events that classifies them as either astrophysical or terrestrial. Then, we will use this information to estimate the covariance of search results and calculate their FAR, joint likelihoods, and unified p_{astro}.

Later, we are going to use the conclusions of the simplified search to come up with new methods of combining search results. Using the current techniques for calculating FAR and p_{astro} and the framework suggested in [9], we would like to suggest a method that would still result in a single number related to the astrophysical probability, but is simple enough computationally to be used with multiple pipelines and realistic data.

In order to achieve these goals, I intend to familiarize myself with the principles of work of three pipelines, PyCBC, GstLAL, and IASS. In addition, I plan to develop skills in working with computer clusters and HTCondor software in particular, developing Python packages, and using statistical methods for data analysis.

IV. TIMELINE

The 10-week duration of the program can be split in two main parts, training and implementation, for which we set the following goals:

1. Training
   - Working with literature;
   - Building knowledge and skills on computer clusters, package development, and statistics;
   - Understanding the pipelines;
   - Developing a simplified search model;
   - Analysing the results of the simplified search;
2. Implementation
   - Designing new methods for combining GW data from multiple sources;
   - Applying suggested methods to realistic data;
   - Analysing whether suggested methods provide meaningful estimates of event significance and have the desired computational complexity;
   - Summarizing the results in the final report.

---

[1] LIGO Scientific Collaboration and Virgo Collaboration. GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. *Phys. Rev. D*, 93:122003, Jun 2016.

[2] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. *arXiv:2111.03606*, Nov 2021.

[3] https://observing.docs.ligo.org/plan/.

[4] Floor S. Broekgaarden. ChatGPT scores a bad birdie in counting gravitational-wave chirps. *arXiv:2303.17628*, Apr 2023.

[5] C. Capano, T. Dent, C. Hanna, et al. Systematic errors in estimation of gravitational-wave candidate significance. *Phys. Rev. D*, 96:082002, Oct 2017.

[6] S. A. Usman, A. H. Nitz, I. W. Harry, et al. The PyCBC search for gravitational waves from compact binary coalescence. *Classical and Quantum Gravity*, 33(21):215004, Oct 2016.

[7] S. Sachdev, S. Caudill, H. Fong, et al. The GstLAL Search Analysis Methods for Compact Binary Mergers

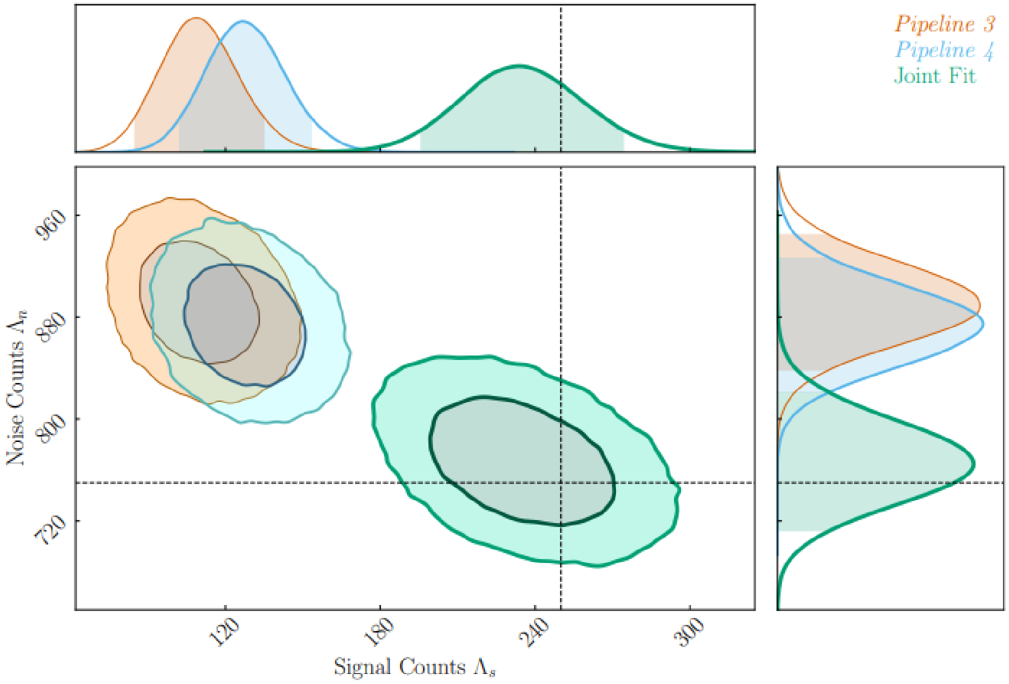

FIG. 1. Illustration of the correlation between two pipelines in a toy model. Image reproduced from [9].

in Advanced LIGO's Second and Advanced Virgo's First Observing Runs. *arXiv:1901.08580*, Jan 2019.

[8] T. Venumadhav, B. Zackay, J. Roulet, et al. New search pipeline for compact binary mergers: Results for binary black holes in the first observing run of Advanced LIGO. *Phys. Rev. D*, 100:023011, Jul 2019.

[9] S. Banagiri, C. P. L. Berry, G. S. Cabourn Davies, et al. A Unified $p_{\rm astro}$ for Gravitational Waves: Consistently Combining Information from Multiple Search Pipelines. *arXiv:2305.00071*, May 2023.

[10] LIGO Scientific Collaboration and Virgo Collaboration. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Phys. Rev. X*, 9:031040, Sep 2019.