Low-latency gravitational-wave data products in the multi-messenger era of the International Gravitational-Wave Network

A Thesis
SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA
BY

Andrew Toivonen

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy: Astrophysics

Advised by Michael Coughlin
UNIVERSITY OF MINNESOTA–TWIN CITIES
Minneapolis, Minnesota

2025
Defended September 4

Andrew Toivonen
ORCID: 0009-0008-9546-2035

ACKNOWLEDGEMENTS

My PhD was made possible by all of the amazing people around me. I give my unending appreciation and gratitude to Michael Coughlin, my advisor, mentor, and friend. Michael is the busiest and hardest working person I know, evidenced by his mornings spent simultaneously in two zoom calls at once–one on his phone, and one his computer. Yet somehow, he is still the fastest person I know at replying to emails. As my fellow students out there know, this can be a pretty good barometer for an advisor. Michael obviously passes with flying colors.

To each and every member of the expanded LIGO group here at the University of Minnesota, thank you for putting up with my rambling as I led group meetings, and I am sorry for the often late reminders or numerous times I left the slides on view-only. Many of you were great at adding to the slides and I appreciate you for guiding discussion. On the other hand, I think some of you wished I left them on view-only each week! Being serious, group meetings were formative for my public speaking skills and gaining confidence talking about science with my peers. Those are invaluable skills I will take with me for the rest of my career.

To Sushant, my "partner in crime" in the collaboration by his own words, and my replacement in Michael's group, thank you. You and I always hung out together at conferences–Banff being particularly memorable–and you were always the first person I would reach out to over Mattermost when I had a question. Sorry our overlap at UMN was short! If you are ever heading to a future conference, let me know, and I will try to join you.

To Vuk, thank you for your constant support and friendliness at group meetings. I appreciate your hospitality and generosity in welcoming the entire group to your home and backyard on multiple occasions.

Thank you to all of my friends and peers, and everyone I have interacted with at the University of Minnesota and beyond. Will for shared gravitational-wave pursuits and good times, Tyler and Brendan for collaboration and adventures in Germany, and Daniel for our office discussions.

To all of my students, both undergraduate and high school, thank you for your effort and contributions, many of which are reflected in this work. For those moving on to the next stage of your career, Megan and Seiya, I wish you the best of luck after your senior year, and Avi best of luck with your undergraduate studies.

Thank you to my collaborators in the LVK, all those I spent countless hours with on telecons and deliberated over merge requests with.

Thank you to Leo Singer, for hosting me at Goddard, advising me for a summer and beyond, and for taking me on awesome adventures in the DC area.

Thank you to Tim Dietrich, the Potsdam crew, and the larger NMMA group. You were amazing hosts, always ask motivating questions, and are a joy to meet at conferences.

To my PhD committee: Michael Couglin, Vuk Mandic, Ben Margalit, Pat Kelly, and Galin Jones–thank you. I appreciate the thorough review of my thesis, for attending my defense, and for all the guidance along the way.

And to my wife Megan, I am sorry you were woken up in the middle of the night countless times by my "kilonova alarm". At least that is what you called it. Unfortunately the name never befit the result. You would always ask me the next day, "did you find one (a kilonova)?" One day maybe I will be able to say yes.

Thank you to my entire family, both the Toivonen's, Farr's and significant others. Spurred on by Megan, they too often asked me if I had found a kilonova yet. They said: "Let me get this straight, you spent five years trying to find this kilonova, never found one, and now someone else is going to pay you to do the same?" Yeah. Pretty much. But they did promise to throw me a kilonova party if I ever find one, so I keep looking.

Now as I reach the end of my time at the University of Minnesota, I look back fondly on my experiences, and share a couple of mine with Michael. First, I remember chatting with Michael near the beginning of the current gravitational-wave observing run, O4. I think Michael was a bit stressed by his numerous responsibilities that day, and I remember him telling me: "I really hope we find a kilonova during O4, otherwise I am NOT going to get tenure. And at that point I think you would want to

find a new advisor." That did not worry me at all, however. It was no surprise to me when he was up for, and received, tenure a year early. And it's looking good for me so far in terms of graduating. Sure, we haven't found a kilonova yet, and everyone is feeling impatient, but doesn't that just mean our services are in high demand? But in seriousness, if it was easy, it wouldn't be worth searching for. And the universe certainly does not make it easy.

I will close on a second anecdote. One summer while we were hosting the ZTF Summer school with students from all around the world, Michael planned an excursion to the Mall of America, where we obviously spent the entire day at the theme park, showing off a highlight of Minneapolis. Michael, Tyler, and I got to ride the Log Flume together, which was pretty cool. You might wonder why I highlight this one random Log Flume ride? And not, for example, Michael and I dueling it out on the competitive Ghost Blasters ride? Well, mostly it is just a fond memory along the way. It is good to have some fun after all. But maybe a PhD is just like the Log Flume, some ups, some downs, but with a little luck you end up where you are meant to get to in the end.

ABSTRACT

The joint gravitational-wave (GW) and electromagnetic (EM) detection of the binary neutron star (BNS) merger GW170817 [9] and the associated kilonova AT2017gfo [7, 86, 217] ushered in the current era of Multi-Messenger Astronomy (MMA). These MMA observations are of great scientific interest due to the breadth of their impact in nuclear astrophysics as sites of r-process nucleosynthesis and heavy element production, in cosmology as means of measuring the universe's rate of expansion, and in tests of general relativity. Low-latency GW searches are the most promising means of detection for both BNS and neutron star-black hole (NSBH) mergers, which have proven to be exceedingly rare. By decreasing the latency at which we send alerts, improving current GW data products, and developing exciting new data products such as estimates of chirp mass and kilonova properties, I strive to maximize our chances of detecting the next multi-messenger neutron star merger and be prepared to access the wealth of scientific knowledge it can provide.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Megan Averill, Andrew Toivonen, et al. The Standard Siren approach using an end-to-end simulation of low-latency Gravitational-Wave Searches in the Mock Data Challenge. 2025.
A.T. advised M.A. as an undergraduate researcher for this work, it will soon be on arXiv and submitted to PASP.

[2] Sushant Sharma Chaudhary, Andrew Toivonen, et al. Low-latency gravitational wave alert products and their performance at the time of the fourth LIGO-Virgo-KAGRA observing run. *Proc. Nat. Acad. Sci.*, 121(18):e2316474121, 2024. doi: 10.1073/pnas.2316474121.
A.T. and S.S.C contributed equally to this work.

[3] Andrew Toivonen et al. What to Expect: Kilonova Light Curve Predictions via Equation of State Marginalization. *Publ. Astron. Soc. Pac.*, 137(3):034506, 2025. doi: 10.1088/1538-3873/adbcd7.
A.T. was lead author and led software development of the data product.

[4] Seiya Tsukamoto, Andrew Toivonen, Holton Griffin, Avyukt Raghuvanshi, Megan Averill, Frank Kerkow, Michael W. Coughlin, Man Leong Chan, and Leo Singer. Astrophysical or Terrestrial: Machine learning classification of gravitational-wave candidates using multiple-search information. *arXiv*, 8 2025. doi: 10.48550/arXiv.2508.14242.
A.T. advised S.T., H.G., A.R., M.A., and F.K. as undergraduate researchers for this work, and it has been submitted to PASP.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

AdVirgo. Advanced Virgo.

AGN. active galactic nucleus.

aLIGO. Advanced LIGO.

AUC. Area Under the Receiver Operating Characteristic Curve.

BAYESTAR. BAYESian TriAngulation and Rapid localization.

BBH. binary black hole (merger).

BCI. the coherent versus incoherent Bayes' Factor.

BH. black hole.

Bilby. Bayesian Inference Library for CBC GW signal.

BNS. binary neutron star (merger).

BSN. signal to noise coherent Bayes' Factor.

CBC. compact binary coalescence.

CMB. cosmic microwave background.

cWB. Coherent WaveBurst.

EM. electromagnetic.

EoS. equation of state.

FAR. false alarm rate.

FITS. Flexible Image Transport System.

FPR. False Positive Rate.

GCN. General Coordinates Network.

GPU. Graphical Processing Units.

GraceDB. GRAvitational-wave Candidate Event DataBase.

GRB. Gamma Ray Burst.

GstLAL. GStreamer LIGO Scientific Collaboration Algorithm Library.

GW. gravitational-wave.

GWTC. Gravitational-Wave Transient Catalog.

HEALPix. Hierarchical Equal Area isoLatitude Pixelization.

IGWN. International Gravitational-Wave Network.

KAGRA. KAmioka GRAvitational-wave observatory.

KDE. kernel density estimation.

KNN. K-Nearest Neighbors.

LIGO. Laser Interferometer GW Observatory.

LISA. Laser Interferometer Space Antenna.

LLAI. Low-Latency Alert Infrastructure.

LVK. LIGO-Virgo-KAGRA.

MBTA. Multi-Band Template Analysis.

MDC. Mock Data Challenge.

ML. Machine Learning.

MLP. Multilayer Perceptron.

MMA. Multi-Messenger Astronomy.

NMMA. The Nuclear Multimessenger Astronomy Framework.

NN. Neural Network.

NS. neutron star.

NSBH. neutron star-black hole (merger).

O1. aLIGO's first observing run.

O2. aLIGO's and AdVirgo's second observing run.

O3. aLIGO's and AdVirgo's third observing run.

O4. aLIGO's, AdVirgo's and KAGRA's fourth observing run.

O5. aLIGO's and AdVirgo's fifth observing run.

oLIB. Omicron+LIB.

PDF. probability density function.

PE. Parameter Estimation.

PyCBC. Python based Compact Binary Coalescence search software.

RAVEN. Rapid, on-source VOEvent Coincident Monitor.

RF. Random Forest.

ROC. Receiver Operating Characteristic.

SCiMMA. Scalable CyberInfrastructure for Multi-Messenger Astrophysics.

SNR. signal-to-noise ratio.

SPIIR. Summed Parallel Infinite Impulse Response.

ToO. target of opportunity.

TPR. True Positive Rate.

Chapter 1

INTRODUCTION

1.1 A Brief Historical Introduction to Gravitational Waves

GWs were first theorized by Albert Einstein in 1916 [99, 100], shortly following, and a consequence of, his 1915 theory of General Relativity that revolutionized modern physics. [98]. While the concept of GWs was intriguing, there were no instruments sensitive enough to directly detect them at the time, and that would be remain case for nearly a century after their formulation.

The first *indirect* detection of GWs was made in 1974 with observations of the Hulse–Taylor binary pulsar system [129]. A pulsar is a rapidly rotating NS that emits intense EM radiation from its poles, observable at a consistent period due to its rotation. It was observed that this pulsar had significant variations in its period, which was explained by a Doppler Shift due to an orbit with the other compact object, a second NS, in the binary. Due to the rapid acceleration of the massive compact objects, the orbit of such a binary should decay from the emission of gravitational radiation in the form of GWs, which was verified based on predictions from General Relativity. This discovery eventually led to the 1993 Nobel Prize.

The first *direct* detection of GWs was made in 2015 when the Laser Interferometer GW Observatory (LIGO) and Virgo collaborations detected GWs emitted from the BBH merger GW150914 [6]. The observations were made using the pair of LIGO laser interferometers, covered in Sec. 1.6. The 2017 Nobel Prize was awarded as a result of this detection and design of the interferometers.

1.2 Gravitational Waves

GWs are the result of Einstein's theory of relativity [98], and later described in [99, 100]. They are a consequence of gravity being described as the distortion of spacetime by massive objects. The acceleration of these masses produces waves which propagate through spacetime.

To understand GWs, we must start with the Einstein field equation:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.1}$$

where $R_{\mu\nu}$ is the Ricci tensor, $g_{\mu\nu}$ is the metric tensor, R is the Ricci scalar, G is the gravitational constant, and $T_{\mu\nu}$ is the stress-energy (or energy-momentum) tensor. Now, for the metric tensor, we start with Minkowski spacetime, composed of three spatial dimensions and time. The Minkowski metric, $\eta_{\mu\nu}$, is defined as:

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.2}$$

We then introduce a small perturbation $|h_{\mu\nu}| << 1$ to the otherwise flat Minkowski spacetime, giving us the following metric:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.3}$$

We then apply the weak-field limit, which is the limit of an observer far away from the mass, and discard higher order terms. This results in the linearized Einstein equations, the weak-field equations, defined in the harmonic gauge as:

$$\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu}. \tag{1.4}$$

Where $\Box$ is the d'Alenbertian operator:

$$\Box = \partial^\mu \partial_\mu = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2, \tag{1.5}$$

and $\bar{h}_{\mu\nu}$ is the trace-reversed metric perturbation:

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h. \tag{1.6}$$

In a vacuum, where $T_{\mu\nu} = 0$,

$$\Box \bar{h}_{\mu\nu} = 0, \tag{1.7}$$

which implies that GWs propagate at the speed of light. Solving the linearized Einstein equation, Eq. 1.4, the simplest solution is a monochromatic plane wave:

$$\bar{h}_{\mu\nu} = \Re\{A_{\mu\nu} e^{ik^\alpha x_\alpha}\} \tag{1.8}$$

,

where $\vec{k}$ is the wave vector and $A_{\mu\nu}$ is the polarization tensor. Now, applying the transverse-traceless gauge, we can determine the degrees of freedom for GWs. Let us assume a wave propagating through flat spacetime in the $\hat{x}$ direction, applying this gauge we get:

$$h_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & h_{yy} & h_{yz} \\ 0 & 0 & h_{yz} & -h_{yy} \end{pmatrix} \tag{1.9}$$

.

This shows GWs have two physical degrees of freedom, and that they deform space along the axes perpendicular to the direction of travel. GWs propagating in the $\hat{x}$ direction will deform space in the $\hat{y}$ and $\hat{z}$ direction, alternating stretching one dimension and compressing the other.

1.3 Neutron Stars and Black Holes

NSs are extremely dense remnants of collapsed stars, containing $\sim 1 - 3\ M_\odot$ within a radius on the order of 10 km. If the progenitor star is massive enough, the force of gravity overcomes the electron degeneracy pressure during collapse, resulting in the object collapsing further and forming a NS. There is still much to learn about NSs, such as the NS EoS, which defines both the maximum mass of a neutron star, above which it collapses into a BH, and the pressure-density, or equivalently

mass-radius, relation of the object. If the neutron star grows too massive, reaching the Tolman-Oppenheimer-Volkoff (TOV) limit somewhere between $\sim 2-3\ M_{\odot}$, the neutron degeneracy pressure is overcome and a BH is formed as there is nothing left to resist gravity and the matter collapses in on itself.

1.4 Sources of Gravitational Waves

GWs are produced by accelerating masses that cause perturbations in spacetime. Similarly to the force of gravity, the amplitude of these waves is quite small, however, unless produced in the most extreme astrophysical events resulting in drastic accelerations of large masses. One such class of astrophysical phenomena, are compact binary coalescences (CBCs). These CBCs are mergers of compact objects, such as NSs and BHs, and are the only source of GWs that has been directly observed so far. It is also expected that GWs could be observed from other compact objects, such as white dwarfs. These CBCs have been observed, thus far, in three types:

- BNS: mergers of two neutron stars
- BBH: mergers of two black holes
- NSBH: mergers of a neutron star and a black hole

CBCs occur when a pair of compact objects become gravitationally bound and in binary orbit around one another. As the objects orbit, they radiate energy in the form of GWs, slowly decaying their orbit. We call this the "inspiral," as the orbital radius shrinks and the objects accelerate to greater speeds. This inspiral can last thousands of years, but the GWs we observe in ground-based GW detectors are from the last violent seconds of the inspiral and merger. The amplitude of the GWs is impacted by the masses involved, as well as the inspiral itself. The innermost stable orbit, or r_{ISCO}, is the smallest stable, circular orbit, and is a defining aspect of the inspiral phase. The smaller the r_{ISCO}, the closer and faster the compact objects orbit, leading to more extreme accelerations and greater energy loss through the radiation of gravitational waves.

Of our confirmed GW detections, BBH mergers are the most numerous. Not only are there larger masses involved in these mergers (up to $\sim 100M_{\odot}$ and beyond), these events have a smaller r_{ISCO} than their NS counterparts. Larger masses and closer orbits means greater distortions of spacetime, and that the GWs emitted are of greater amplitude, making them easier to detect. BNS events, for comparison, are composed on NSs which, at most, can approach $\sim 3\ M_{\odot}$. These BNS events inspiral more slowly through the sensitive frequency range of our detectors, however, meaning the signals from BNS mergers may be detectable for tens of seconds or minutes, while the signal from BBH mergers only seconds or fractions of a second. We use modeled CBC searches to detected these GW signals, based on our understanding of the expected waveforms [18]. New machine learning-based searches capable of parameter estimation within seconds are also in development [167].

There are also other expected types of GWs that could be observed. Beyond CBCs, we expect core-collapse supernovae can produce an observable GW signal from oscillations caused by hydrodynamic instabilities during collapse [25]. There are "burst" GW searches, which target unmodeled signals of unknown or undetermined origin. These searches primarily search for coincident GW power in multiple detectors, indiscriminate of the signal morphology [17].

Finally, there are continuous and stochastic GW. Continuous GWs refer to those emitted continuously at the same frequency by a source, such as a spinning NS [20]. Stochastic GW refer to the GWs from the GW, or stochastic, background. This background signal is the cumulative "hum" of the universe in GWs, the combination of all the faint signals present in the universe, some a result of the Big Bang [72].

1.5 Detection of Gravitational Waves

GWs propagate at the speed of light, and lose amplitude and energy as they propagate through space. Unlike light, however, which falls off in brightness by the square of distance, the amplitude of GWs falls off proportional to distance, allowing them to be observable across great distances. The direct detection of GWs requires extraordinary precision, specifically the ability to measure changes in distance down

to approximately 1/10000th the width of proton[1].

Ground-based detectors like LIGO [2], Virgo [26], and Kagra [30], introduced in Sec. 1.6, are designed to be sensitive to a specific frequency range of GWs, from ~ 10 Hz to 10 kHz [14]. This range works well for detecting CBCs, as they emit GWs in this range at the very end of inspiral, when the signal is the strongest. While GWs are emitted during the entire inspiral that can take thousands of years, the distortions in spacetime become the strongest and most violent as the objects approach one another, increasing in orbital frequency, and eventually merge. It is these last few seconds, or even fractions of seconds, that fall within the sensitive band and that we hope to observe. We call this characteristic rise and spike in the frequency of the corresponding emitted GWs the “chirp".

We measure GWs in strain, h, which is the measure of how a length, L is compressed by GWs.

$$h = \frac{\Delta L}{L} \tag{1.10}$$

As GWs emitted from an astrophysical source pass through the detector, they cause a detectable change in the length of the detector and the path length of the laser used to measure the change. This is covered in detail in Sec. 1.2. The measured strain of a GWs will only be around 10^{-21} by the time the GWs reach Earth, a minuscule change in distance and one that requires impressive sensitivity to detect.

1.6 The International Gravitational-Wave Network

In order to detect, study, and localize GWs, we use a network of laser interferometers, currently composed of the LIGO detectors in Hanford, Washington and Livingston, Lousiana [2], the Virgo detector near Pisa, Italy [26], the Kagra detector near Hida, Japan [30], and the GEO600 detector near Hannover, Germany [95]. This network is seen in Fig. 1.1.

Laser interferometers, shown in Fig. 1.2, are L-shaped instruments with equal length arms and a beam splitter for laser, where the laser is split and then recombined to

[1] https://www.ligo.caltech.edu

observe the resulting interference pattern. When the arms are free of disturbances, the lasers from each arm are recombined out of phase and destructively interfere, canceling each other out. When the arms are disturbed, such as being expanded and contracted as GWs pass through, the split lasers will no longer be perfectly out of phase due to the change in path length, and there will be an observable signal when recombined.

The pair of LIGO detectors are currently the largest and most sensitive in the network, with 4 km arms and a BNS sensitivity of 150-160 Mpc, as seen in Fig. 1.3. The Virgo interferometer has 3 km arms with a sensitivity of 50-80 Mpc, while KAGRA has similar 3 km arms, with a lower sensitivity. Lastly, GEO600 is a 600 m detector primarily used for detector research and development. A new LIGO-like detector is also planned near Hingoli, India and will join the network, called LIGO India [207].

The network of detectors increases our chances of detection for a number of reasons. First, distinguishing GW signals from noise is difficult, so with detectors isolated from one another, we should not expect the same noise signals to be present in multiple detectors at the same time, unlike for GWs. Coherence between the signals in multiple detectors is a strong indication a signal is astrophysical. The network of detectors also allows for the signals to be localized in the sky via the differing detector responses, and allows for the detection of signals that are missed by one detector, whether that be due to a signal located in the blind spot of the detector antenna pattern, due to transient noise, or when a detector is down.

1.7 Gravitational-Wave Laser Interferometer

GW detectors are exceptionally precise instruments operating under a simple premise: when GWs pass through their arms, they distort spacetime, changing the path length of the laser. The LIGO detectors are Michelson Interferometers with 4 km Fabry Perot cavities, as seen in the simplified diagram in Fig. 1.2 [2]. The interferometer starts with a high power laser passes through a beam splitter, sending half of the laser's power in two perpendicular directions. The laser then passes into the Fabry Perot cavities.

These Fabry Perot cavities increase the effective length of the instrument by the

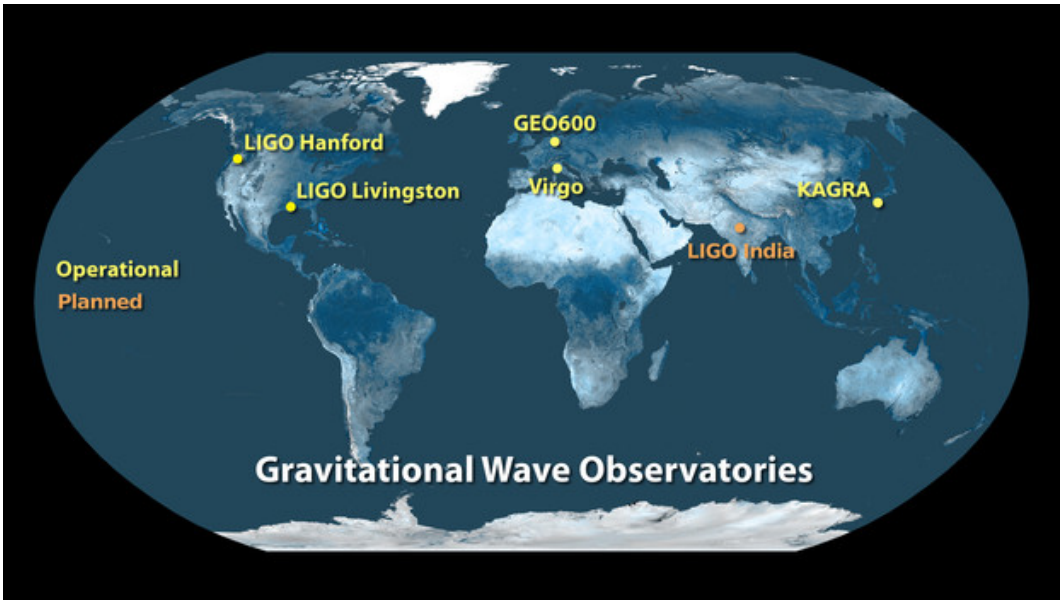

Figure 1.1: The International Gravitational-Wave Network of Observatories. Credit: `https://www.ligo.caltech.edu`

reflecting the laser back and forth an average of 300 times, making the effective length of the instrument 1200 km. Increasing the effective length in this manner increases our sensitivity to GWs, without further increasing the size of the interferometer. In addition, power recycling mirrors are used to increasing the power of the laser, again increasing the precision of the instrument. This is done by continuously reflecting light that has already passes through the instrument back in the Fabry Perot cavities.

Once the light from one arm of the interferometer has escaped the highly reflective Fabry Perot cavity, it recombined with the light from the other arm. When there is no disturbance, the two potions of the split laser are recombined out of phase with perfect destructive interference, and no signal is observed. When a GW signal passes through the detector, spacetime is deformed and the effective path length of the laser is impacted, which changes the outgoing phase of the laser and disrupts the destructive interference, resulting in an observable signal at the detector. Additional details can be found in [2].

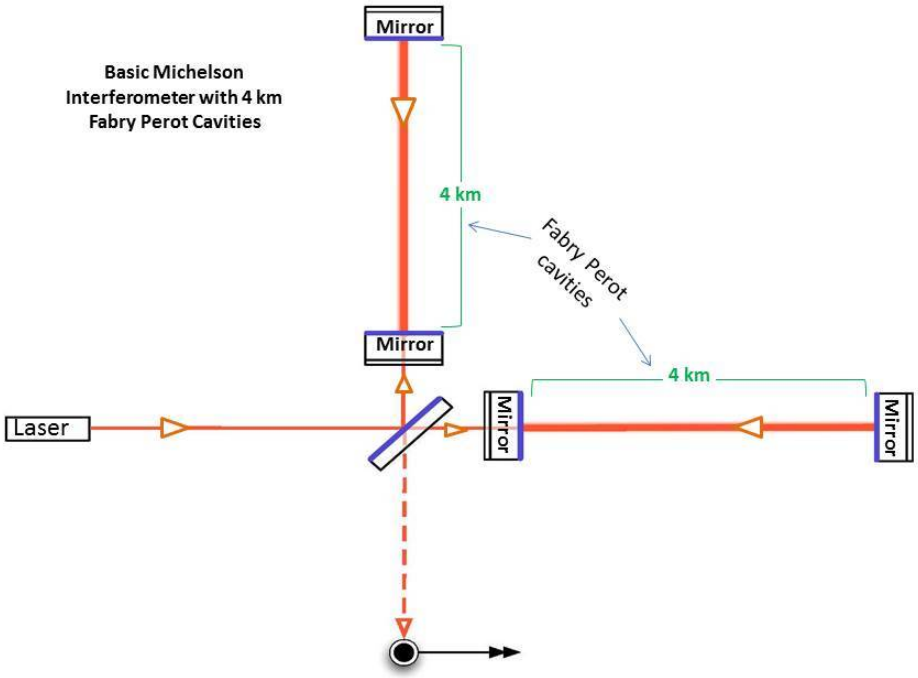

Figure 1.2: A basic diagram of a Michelson interferometer with 4 km Fabry Perot cavities, similar to LIGO interferometers. Credit: `https://www.ligo.caltech.edu`

1.8 Kilonovae

Kilonovae are astrophysical transient events that may result from either BNS or NSBH mergers. They are of particular scientific interest as sites of r-process nucleosynthesis, through which heavy elements can be produced. The radioactive decay and interactions of these r-process elements in the mass ejected from the NS(s) powers the EM emission we hope to observe [135, 151, 155, 171]. Observations of kilonovae, such as GW170817, allow for breakthroughs in nuclear astrophysics [135, 137, 166, 217, 234], cosmology [8, 83, 94], and tests of General Relativity [45, 88, 104].

The name "kilonova" comes from the fact the peak luminosity can reach 1000 times that of a classical nova. At 10^{42} erg/s, these transients are approximately 100

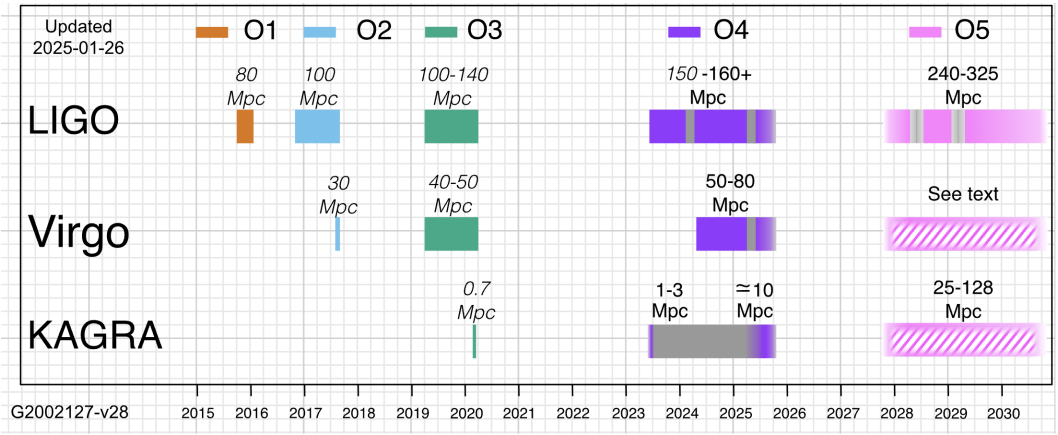

Figure 1.3: BNS Sensitivity of IGWN interferometers. Credit: `https://emfollow.docs.ligo.org/userguide/`

times fainter than a type Ia supernova at peak luminosity. Searches for kilonovae are challenging as they are short lived, can be relatively faint, and may not be well-localized. Specifically when following up GW events, sky localizations can span $\approx 100 - 10,000$ deg^2 [53, 77, 101, 105, 106, 123, 142, 202, 212, 214, 235]. GW178017, BNS merger, is the only NS merger that provided joint observations of a GW signal, kilonova, and sGRB. While GW follow-up is rightfully a focus, it is also possible to identify kilonovae associated with sGRBs [39, 133, 224], long GRBs [154, 193, 227], or even serendipitously in survey operations [31, 33, 34].

Kilonovae may result from the BNS or NSBH mergers as mass is ripped from the NS(s) during the extreme tidal forces of the coalescence, and the particle-driven winds that result. The radioactive decay and interactions of r-process elements in this mass ejected during merger is what powers the emission we can observe.

Mass ejecta (m_{ej}) is often described in two components: the dynamical ejecta and the wind ejecta. The dynamical ejecta ($m_{\rm ej}^{\rm dyn}$) is produced as the extreme gravitational forces of the inspiral tidally deform the NS and rip mass free [172]. The wind ejecta ($m_{\rm ej}^{\rm wind}$) is produced when matter is ejected from the accretion disk surrounding the merger by energetic outflows of particles, sometimes referred to as particle "winds" [172]. The total mass ejecta is expected to be strongly correlated with the peak luminosity of the kilonova emissions, so predicting the mass ejecta

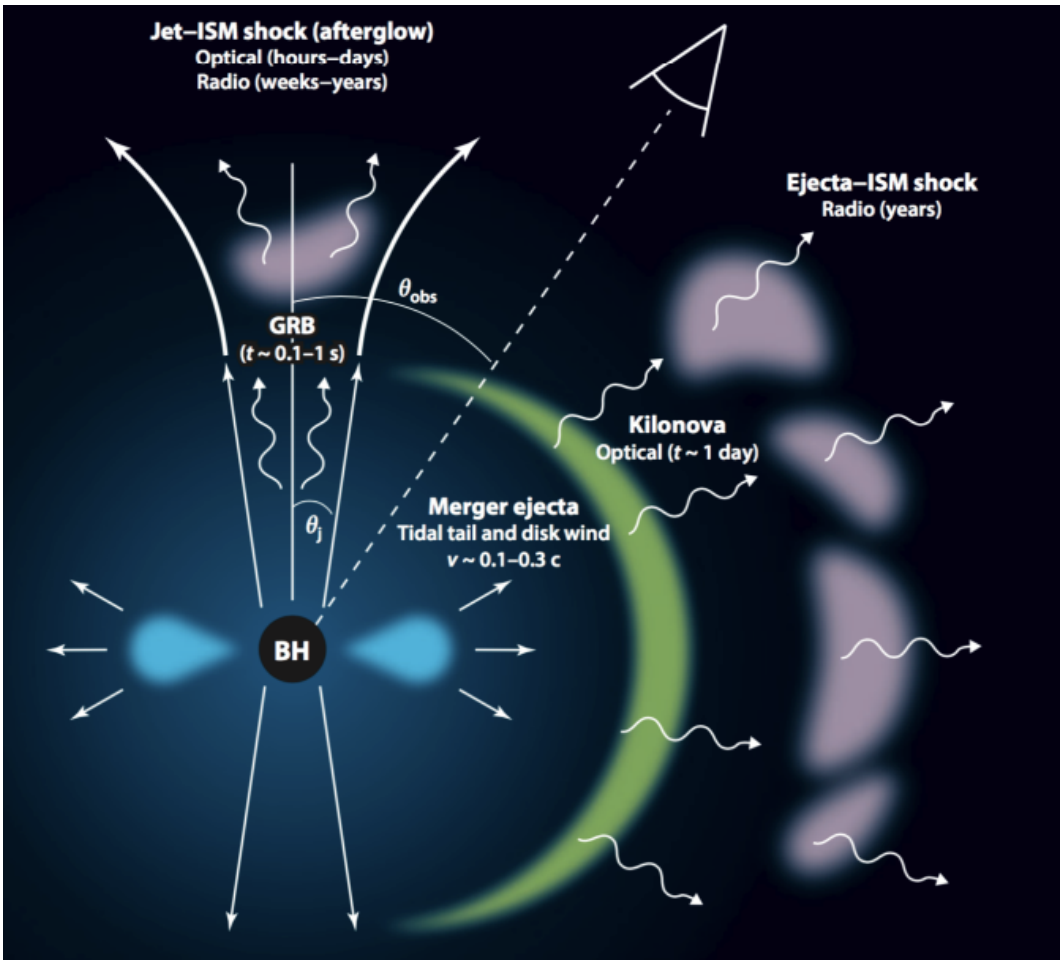

Figure 1.4: Summary of the electromagnetic counterparts of BNS and NSBH mergers and their dependence on the viewing angle with respect to the axis of the GRB jet. The kilonova, in contrast to the GRB and its afterglow, is relatively isotropic and thus represents the most promising counterpart for the majority of GW-detected mergers. *Figure and caption excerpt from:* Ref. [172]

for a given merger useful for understanding which events can produce an observable kilonova.

Models of kilonova emission often consider two components: a red, lanthanide-rich component, and a blue, lanthanide-free component. The red component is neutron rich and where heavy element nucleosynthesis in the r-process is expected to take place. This component is equatorial and high opacity, and peaks in the NIR [172].

The blue component is expected to be concentrated in the polar regions or isotropic outflows from the accretion disk. It has a lower neutron fraction and low opacity, and peaks in the visual bands.

Fig. 1.4 shows a summary of the different EM components that may follow a BNS or NSBH merger, including a kilonova, Gamma Ray Burst (GRB), and accompanying afterglow. The GRB will occur and peak within seconds of merger, concentrated in the polar regions, while the kilonova will reach the peak of emissions on the order of $\sim$ day after merger, and slowly fade. Chapter 4 covers proposed data products designed to maximize our chances at detecting a kilonova.

1.9 Gamma-Ray Bursts

GRBs are high energy astrophysical transients resulting from the supernovae or CBC merger events. As a general rule, if the burst lasts $< \sim$ 2s, it is classified as an short GRB (sGRB), and those with a greater burst duration as long GRBs. GRBs were first discovered by the Vela satellites by scientists at Los Alamos National Laboratory while watching for tests of nuclear weapons in space [140]. The bursts were found to be of extra-galactic origin, and hence extremely energetic, due to their isotropic distribution in the sky and lack of coincidence with the galactic plane. They also can have observable afterglows in longer wavelengths such as radio and X-ray, allowing for additional study.

As GRBs can be caused by CBC mergers, they are a captivating target for multi-messenger efforts. As covered in Sec. 1.11, multi-messenger observations of CBC have applications across nuclear astrophysics, general relativity, and cosmology. There are numerous efforts to find coincidences between GRBs, GWs, and neutrinos, and it is also possible to identify kilonovae associated with both sGRBs [39, 133, 224] and long GRBs [154, 193, 227].

1.10 Neutrinos

Neutrinos are weakly interacting particles, specifically only interacting through the weak force and gravity, making them difficult to detect. High energy neutrinos can be emitted from supernovae and CBCs, which make them targets of multi-messenger efforts. In order to detect neutrinos, large Cherenkov detectors are used to observe

the light emitted via Cherenkov radiation from secondary particles produced as neutrinos interact with other particles. Cherenkov radiation is produced when these high energy secondary particles travel faster than the speed of light in a medium, such as water or ice, and produce light. This light can then be detected and the particle interactions reconstructed to determine the energy and direction of the incoming neutrino.

The IceCube Neutrino Observatory is one such example of a Cherenkov neutrino detector. IceCube consists of digital optical modules with photomultiplier tubes covering a cubic kilometer of ice [1]. When a neutrino interacts with a molecule, such as the water in the ice, it creates a shower of charged particles that create Cherenkov light observable by the detector. As these high energy neutrinos can produced by transients such as GRB or kilonova resulting from CBCs, neutrinos coincident with GWs can indicate such a multi-messenger event.

1.11 GW150914 and GW170817: Heralds of the Multi-Messenger Era

On September 9th, 2015 the first direct detection of GWs was made via observations of a pair of black holes merging: a BBH merger [6]. The black holes involved in this merger were $\sim 36\ M_\odot$ and $\sim 29\ M_\odot$ respectively, and the observations matched the waveform predicted by General Relativity for such a merger into and ringdown of the resulting black hole. The detection of this event is shown in Fig. 1.5, with the expected waveform in the top panels and characteristic "chirp" in frequency seen in the boot panels. This Nobel Prize-winning discovery ushered in the era of GW astronomy, opening up a new means, or messenger, for studying the observable universe: GWs. These observations led to rapid growth in the field of GW astronomy, as a new wealth of information previously inaccessible became available. Alongside growth in countless areas of study, GW astronomy has lead to comprehensive studies of the population of merging binaries [21] and extensive tests of general relativity [19].

Then, with the combined detection of GWs from the BNS GW170817 [9] and EM counterparts in the kilonova AT2017gfo [7, 86, 217] and GRB170817A [10, 118, 208], the era of multi-messenger astrophysics was ushered in. Fig. 1.6 shows the

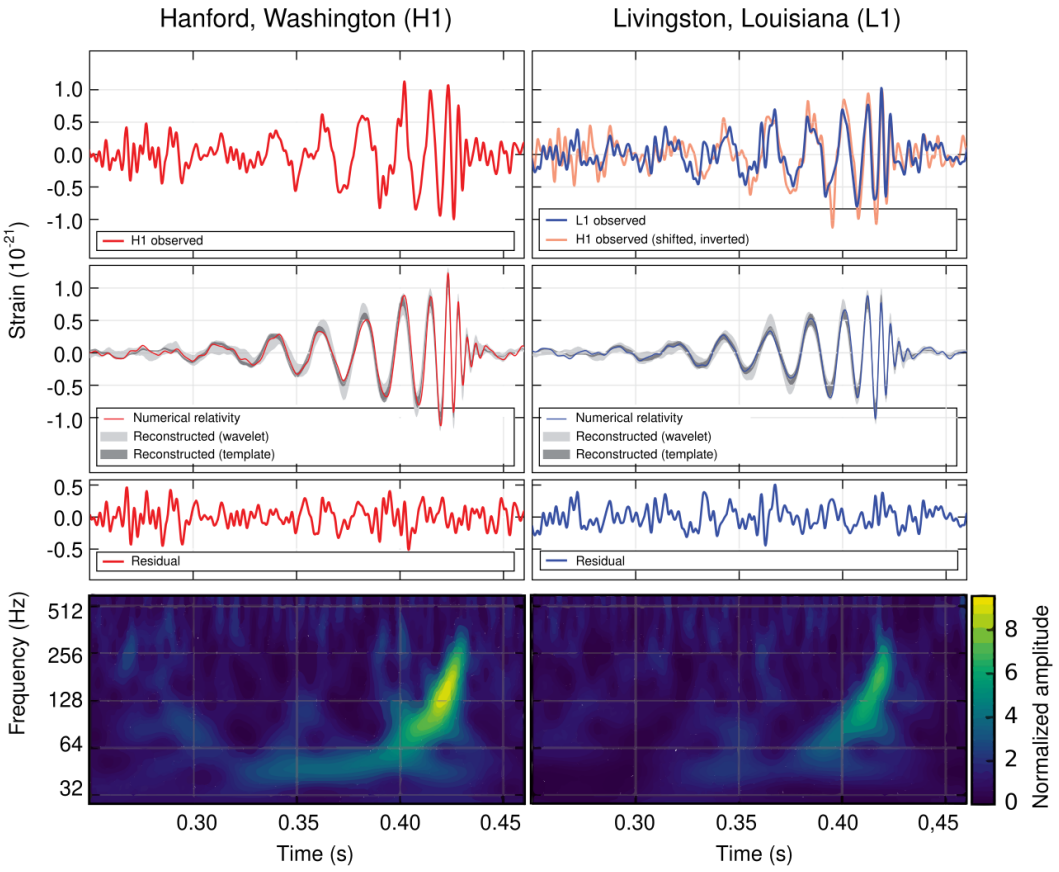

Figure 1.5: The gravitational-wave event GW150914 observed by the LIGO Hanford (H1, left column panels) and Livingston (L1, right column panels) detectors. Times are shown relative to September 14, 2015 at 09:50:45 UTC. *Figure and caption excerpt from:* Ref. [6]

coincident GW and GRB detection. This combined detection was a multi-messenger event, referring to the combination of information, or messengers, from light and GWs in this case, with another possible messenger in neutrinos. The joint GW and EM observations of the BNS and kilonova have led to insights across diverse fields of physics and astronomy. With this detection, kilonovae are shown to be sites of r-process nucleosynthesis, a means through which heavy elements are produced, and it is the radioactive decay and interactions of these r-process elements that powers the observable kilonova emission [135, 151, 155, 171]. Observations of GW170817

led to breakthroughs in nuclear astrophysics in our understanding of the merger dynamics and resulting emissions [135, 137, 166, 217, 234], as well as provided a measurement of the NS radius and EoS [11], and was the subject for numerous tests of General Relativity [45, 88, 104]. In addition, the event is of cosmological interest, as a means for standard siren measurement of the Hubble Constant [8, 83, 94]. There is still much to learn, and we hope additional future observations of these rare and elusive events will help further our understanding of the diversity of their intrinsic parameters, their EM and neutrino emissions, and the heavy elements produced by r-process nucleosynthesis.

1.12 GW170817: A Summary of Observations

On August 17, 2017, a short GRB was detected by Fermi-GBM, and an alert was immediately sent out to the community. The GRB was confirmed shortly after by INTEGRAL [10]. Around the same time, a GW signal was observed in the LIGO-Hanford detector. However, due to data quality issues in the LIGO-Livingston detector, the signal was not immediately picked up. After expedient offline analysis, including removing a transient noise artifact in the Livingston data and including Virgo, the result was a 32.4 SNR trigger observed over $\sim$ 100 s. This was the loudest and longest duration signal observed at the time. The combined GW network SNR of 32.4 corresponds to individual SNRs of 26.4, 18.8, and 2.0 in the LIGO-Hanford, LIGO-Livingston, and Virgo detectors, respectively.

The initial low-latency sky localization for this GW event, now know as GW170817, was 190 deg^2, but after additional offline analysis, this was reduced to 31 deg^2, and then 28 deg^2. The distance to the event was determined to be approximately 40 Mpc. Where the source was localized, as well as the comparison in sky localizations and distance posteriors can be see in Fig. 1.7. This drastic improvement in localization was possible by including all three detectors, LIGO-Hanford, LIGO-Livingston, and Virgo in the analysis. Even though Virgo contributed a low SNR, the inclusion of Virgo data was still of great benefit as it could be determined the source was near the blind spot of the detector, and regions of the sky could be ruled out.

Urged on by the prospects of a multi-messenger detection and armed with the sky

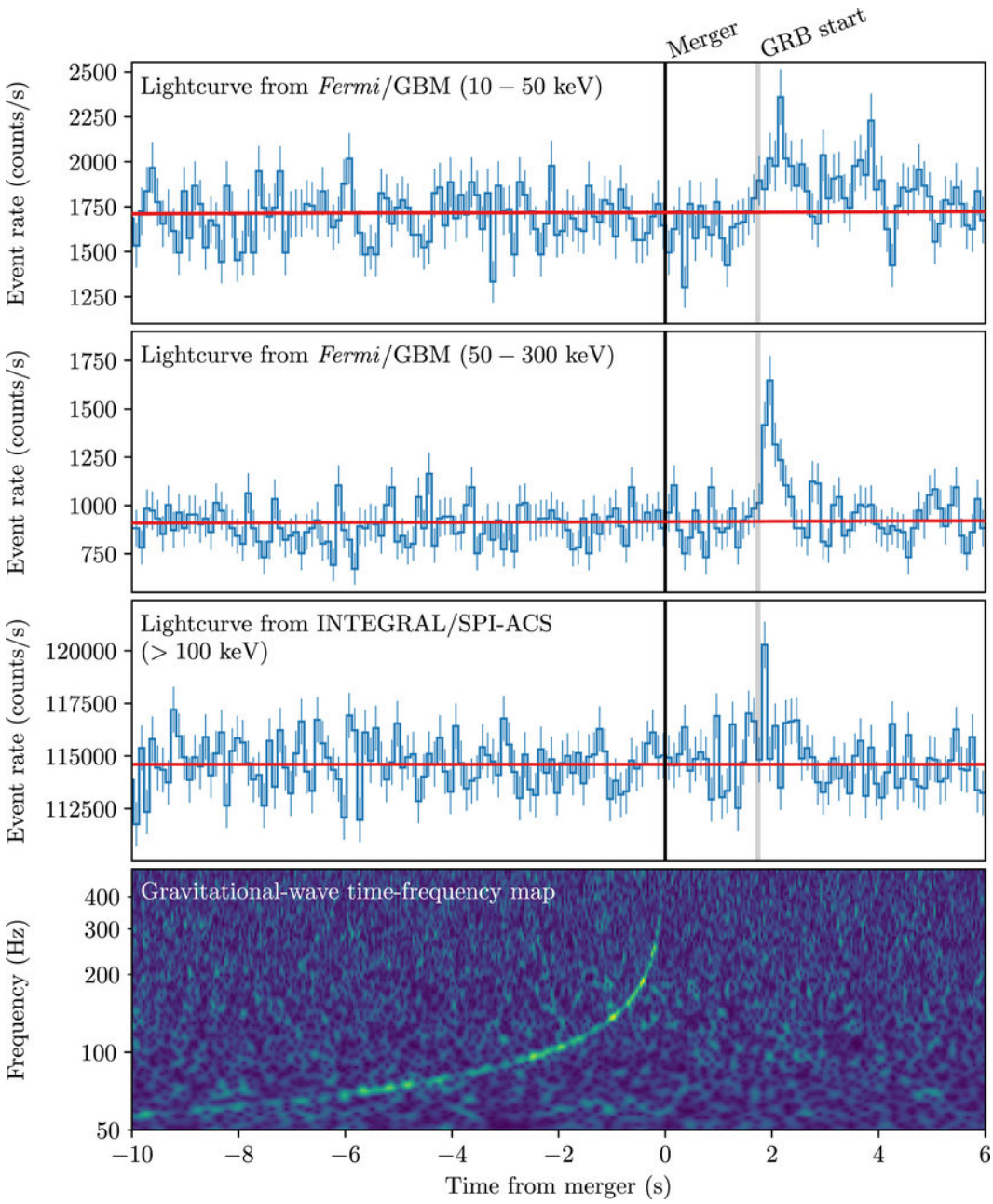

Figure 1.6: Joint, multi-messenger detection of GW170817 and GRB 170817A. *Figure and caption excerpt from:* Ref. [10]

localization, astronomers and surveys across the globe began searching for an EM counterpart. Approximately 11 hours after the GW detection, the source was found

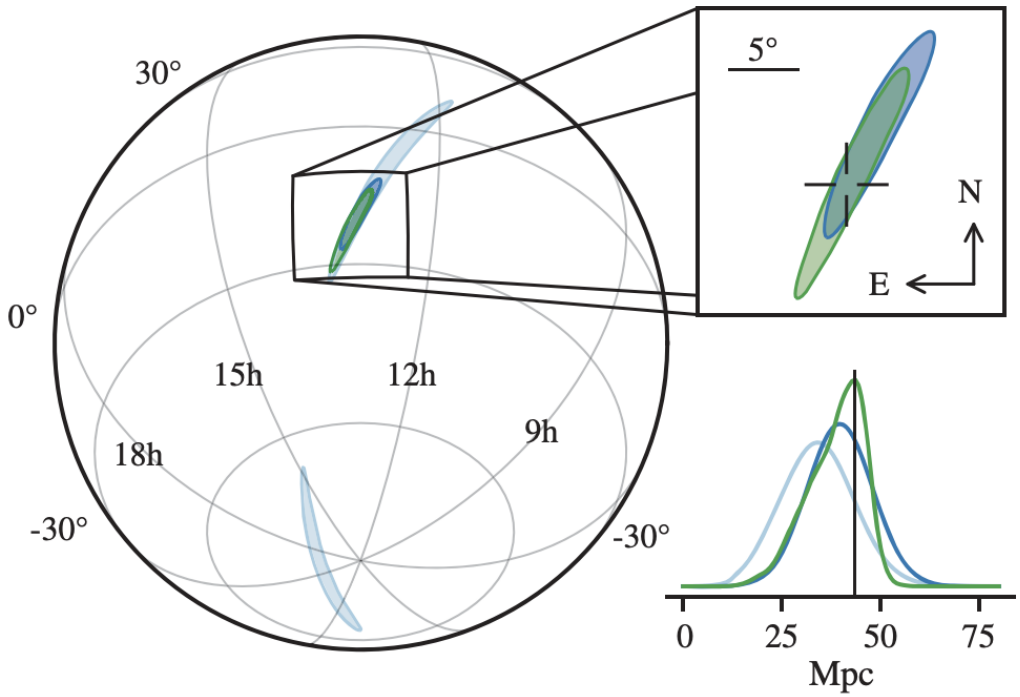

Figure 1.7: Sky location reconstructed for GW170817 by a rapid localization algorithm from a Hanford-Livingston (190 deg^2, light blue contours) and Hanford-Livingston-Virgo (31 deg^2, dark blue contours) analysis. A higher latency Hanford-Livingston-Virgo analysis improved the localization (28 deg^2, green contours). In the top-right inset panel, the reticle marks the position of the apparent host galaxy NGC 4993. The bottom-right panel shows the *a posteriori* luminosity distance distribution from the three gravitational-wave localization analyses. *Figure and caption excerpt from:* Ref. [9]

near the galaxy NGC 4993 by the Swope Telescope at Las Campanas Observatory, and soon confirmed by others [7, 10]. The source was determined to be at 40 Mpc, consistent with the distance estimates from the GW sky localization. With the source located, additional photometric and spectroscopic observations were taken using numerous telescopes, and would lead to scientific gains in nuclear astrophysics, general relativity, and cosmology [7].

1.13 Low-Latency Alert Infrastructure

The Low-Latency Alert Infrastructure (LLAI) is comprised of numerous software stacks working in unison to send out timely alerts to the greater scientific community and enable a wide breadth of research related to our GW detections. Upon detection and upload of a candidate event passing the FAR threshold by one of our GW search pipelines, public alerts are sent out to inform the greater astronomical community of GW candidate events, enabling multi-messenger follow-up of these events. It is imperative these alerts are sent with minimal latency, as astrophysical transients such as kilonovae and GRBs are short lived and reach their peak of emissions very quickly. These public alerts are distributed both via GCN and the Scalable Cyberinfrastructure to support Multi-Messenger Astrophysics[2] (SCiMMA).

A few major components of the LLAI are GRAvitational-wave Candidate Event DataBase (`GraceDB`)[3], `igwn-alert`[4], and `GWCelery`[5] [6]. `GraceDB` serves as both the database for candidate events, as well an internal and external web view for these same events. `igwn-alert` is an internal messaging system to communicate the state of events, and `GWCelery` is a task queue, to cluster, annotate, and orchestrate the events, as well as publish public alerts to the greater scientific community. The components and performance of the LLAI are covered in detail in Chapter 2, including the individual GWs search pipelines.

[2] `https://scimma.org/`
[3] `https://gracedb.ligo.org/`
[4] `https://igwn-alert.readthedocs.io`
[5] `https://git.ligo.org/emfollow/gwcelery`
[6] `https://rtd.igwn.org/projects/gwcelery/en/latest/`

Chapter 2

LOW-LATENCY GRAVITATIONAL WAVE ALERT PRODUCTS AND THEIR PERFORMANCE AT THE TIME OF THE FOURTH LIGO-VIRGO-KAGRA OBSERVING RUN

Work reproduced from:

Sushant Sharma Chaudhary, Andrew Toivonen, et al. Low-latency gravitational wave alert products and their performance at the time of the fourth LIGO-Virgo-KAGRA observing run. *Proc. Nat. Acad. Sci.*, 121(18):e2316474121, 2024. doi: 10.1073/pnas.2316474121.
A.T. and S.S.C contributed equally to this work.

2.1 Motivation

Multi-messenger searches for BNS and NSBH mergers are currently one of the most exciting areas of astronomy. The search for joint electromagnetic and neutrino counterparts to GWs has resumed with O4. To support this effort, public semi-automated data products are sent in near real-time and include localization and source properties to guide complementary observations. In preparation for O4, we have conducted a study using a simulated population of compact binaries and a MDC in the form of a real-time replay to optimize and profile the software infrastructure and scientific deliverables. End-to-end performance was tested, including data ingestion, running online search pipelines, performing annotations, and issuing alerts to the astrophysics community. We present an overview of the low-latency infrastructure and the performance of the data products that are now being released during O4 based on the MDC. We report the expected median latency for the preliminary alert of full bandwidth searches (29.5 s) and show consistency and accuracy of released data products using the MDC. For the first time, we report the expected median latency for triggers from early warning searches (-3.1 s), which are new in O4 and target neutron star mergers during inspiral phase. This paper provides a performance overview for LVK low-latency alert infrastructure and data

products using the MDC and serves as a useful reference for the interpretation of O4 detections.

2.2 Low-Latency Gravitational Wave Searches

As of May 24 2023[1], aLIGO's, AdVirgo's and KAGRA's fourth observing run (O4) is underway, following a series of observing runs, which have reported the detection of the first BBH in aLIGO's first observing run (O1) [24], the detection of the first BNS merger [9] and associated electromagnetic counterparts AT2017gfo [7, 86, 217] and GRB170817A [10, 118, 208] in the aLIGO's and AdVirgo's second observing run (O2),and NSBH [15] in the O3. Focusing on NS mergers, there are a variety of science cases for their multi-messenger counterpart searches and detections, including measurements of the NS EoS [37, 48, 79–81, 93, 131, 150, 164, 177, 192], the Hubble constant [8, 82, 83, 93, 128], and r-process nucleosynthesis [71, 86, 87, 136, 187, 201, 217, 234].

The LIGO-Virgo-KAGRA (LVK)'s real time alert infrastructure depends on several components. Broadly this includes low-latency data calibration and transfer, running modeled and unmodeled online searches, and maintaining the state of events in `GraceDB` following the discovery. In addition to `GraceDB`,[2] which serves as both the database and as an internal and external web view, the alert infrastructure includes `igwn-alert`,[3] an internal messaging system to communicate the state of events, and `GWCelery`,[4] [5] a task queue, to cluster, annotate, and orchestrate the events, as well as publish public alerts[6] for the community to subscribe to. Figure 2.1 shows the task flow of the LLAI for candidate events. The current LLAI is a significantly upgraded version of the infrastructure used earlier, described in [157], and used in the more recently reported early-warning system, reported in [162]. The primary changes compared to previous observing run are the transition from a XMPP based pubsub system internally to Kafka-based messaging provided by the SCiMMA bro-

[1] `https://observing.docs.ligo.org/plan`
[2] `https://gracedb.ligo.org/`
[3] `https://igwn-alert.readthedocs.io`
[4] `https://git.ligo.org/emfollow/gwcelery`
[5] `https://rtd.igwn.org/projects/gwcelery/en/latest/`
[6] For example, those hosted by SCiMMA and NASA

ker, removal of timeouts at various places of synchronization and instead relying on labels on `GraceDB` to keep track of the state of the superevent, and reconfiguring the snapshotting configuration for `Redis` database to strike a balance between fault tolerance and avoid filling disk quota. Aside from this, upgrades to software versions of the dependencies and running computationally expensive resources on specific pool of modern hardware contributed to a improvement compared to previous observing run. The LVK Alert User Guide [7] constitutes a living document where information and updates about this system are regularly communicated to the broader community. Further discussion of `GraceDB`, `igwn-alert`, and `GWCelery` is provided in the supplementary material.

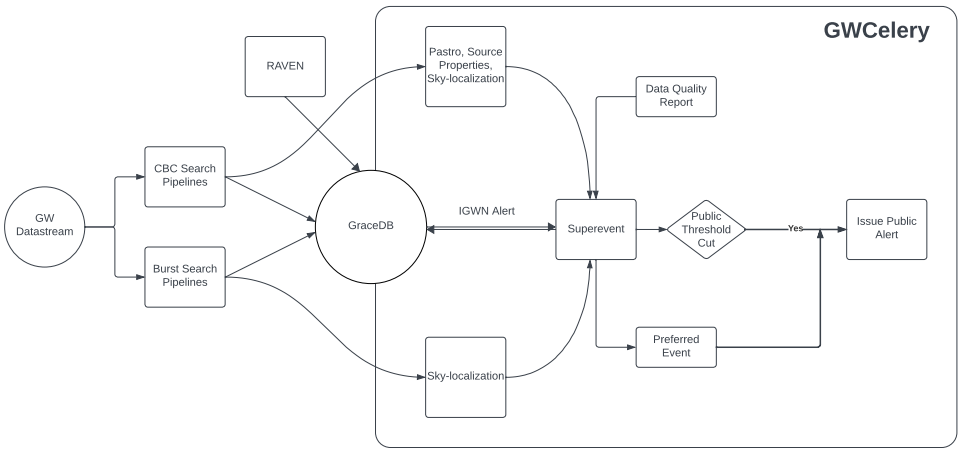

Figure 2.1: Task flow of low-latency alert infrastructure. The process begins with search pipeline trigger(s) on a candidate in the GW datastream, which are passed through data quality checks and compiled into a superevent. If the preferred event from the superevent passed the significant FAR cut, a preliminary alert is sent out to the public.

To prepare for O4 and demonstrate performance across a variety of software and alert system improvements, we carried out a Mock Data Challenge (MDC). This MDC constitutes a testing environment for the LLAI to prepare for O4, producing repeated sets of 40 days of data from O3, with associated simulations of CBCs to

[7] `https://emfollow.docs.ligo.org/userguide/`

stress test the system. While the rates of simulated events (see Section 2.2 for a description of the data set) were much higher than that expected for O4, this high rate was designed to test the various components of searches, the alert system and the scientific deliverables before heading into O4; these include, for example, tests of the detection efficiency of online real time low-latency searches, the rapid estimation of the binary system properties, and their associated sky localizations.

In this paper, we describe the details of the alert system for O4 and its performance based on this MDC. In addition, we provide an overview of the detection performance of real time searches, along with the consistency and accuracy of alert data products. Sections 2.2 - 2.2 provide an overview of the LLAI and the scientific data products reported, while Section 2.2 describes the properties of the MDC, including the motivations for the choices made. Section 2.3 reports the properties of the LLAI as of the beginning of O4, as measured by the MDC, and Section 2.4 describes the conclusions and prospects for future development.

Searches

Low-latency GW searches consist of two categories: "modeled" CBC [160] and "unmodeled" Burst [159] searches. Modeled CBC searches target BNS, NSBH, or BBH; unmodeled searches look for signals with generic morphologies from a wide variety of astrophysical sources like core-collapse of massive stars, magnetar star-quakes, and other sources, in addition to compact binary mergers [16, 159]. For the purpose of this MDC analysis, we focus on CBC searches, but also report latencies of injections found by Burst pipelines. CBC searches can be categorized as early warning, referring to pre-merger searches [146, 205], or full bandwidth, referring to post-merger, based on how the search truncates their templates. Each search produces candidate GW triggers and assigns them ranking statistic values and FARs; the FAR for a trigger in a given search pipeline is defined as the expected rate of triggers due to detector noise, in that pipeline, with equal or higher ranking. Each search pipeline has different and independent methods of generating and ranking triggers and estimating the noise background and thus the FAR; for details see [43, 89, 127, 156, 170, 188, 231]. In addition, the probability of astrophysical origin, p_{astro}, for a trigger is calculated for CBC searches, which is briefly introduced in

Section 2.2.

The MDC data was analyzed by four CBC searches (GstLAL [103, 170, 228], Multi-Band Template Analysis (MBTA) [43], Python based Compact Binary Coalescence search software (PyCBC) [89, 181], Summed Parallel Infinite Impulse Response (SPIIR) [74, 146]), two burst searches (Coherent WaveBurst (cWB) [96, 143, 144], and Omicron LALInferenceBurst (oLIB) [161]), and an external coincidences search (Rapid, on-source VOEvent Coincident Monitor (RAVEN) [70, 188, 230]). A brief description of these pipelines is given in the supplement section.

Selection of the public GW event candidate

When a potential GW signal appears in the detector, the low-latency search pipelines analyze the signal and produce event candidates. Each pipeline can report multiple candidates for a single GW signal. The event candidates, reported within a specific time window (1 s around coalescence time for CBC searches, and 1 s around trigger time for burst searches), are collected and grouped as a *superevent*. In the collection, one GW event candidate is identified as the preferred event (defined as the event with the highest network SNR for CBC pipelines whereas lowest FAR for burst pipelines), and its properties and data products are prepared for release to the public, which include the merger time, FAR, sky localization, and classification (see Section 2.2).

Alerts are released publicly when a FAR passes the public alert threshold, currently FAR $\leq 1.6 \times 10^{-4}$ Hz (fourteen per day). An alert is labeled as *significant* when a CBC alert passes a FAR threshold of FAR $\leq 3.9 \times 10^{-7}$ Hz (one per month) or when an unmodeled burst alert passes a FAR threshold of FAR $\leq 3.2 \times 10^{-8}$ Hz (one per year). Since multiple CBC and burst searches run in low-latency, to account for the trials factor from these different searches with statistically independent false alarms, events from CBC searches are labeled significant when a FAR passes a threshold of FAR $\leq 7.7 \times 10^{-8}$ Hz (one per 5 months), whereas events from burst target searches require a FAR $\leq 7.9 \times 10^{-9}$ Hz (one per 4 year). In the MDC, however, CBC trials factor of 6 was used due to which CBC events were labeled significant when a FAR passes a threshold of FAR $\leq 6.4 \times 10^{-8}$ Hz (one per 6 months). Considering the trials factor accounting for all searches, the public alert

threshold is FAR $\leq 2.3 \times 10^{-5}$ Hz (two per day). Alerts that meet the public alert threshold but not the significant threshold are labeled *low significance*. RAVEN only uses the significant FAR thresholds when assessing its joint FAR for publication, with additional trials factors to compensate for listening to multiple GW pipelines. The LVK Alert User Guide[8] should be referenced for up-to-date information on trials factors during observing runs.

Alert Contents

Public alerts are sent in order to inform the greater astronomical community of GW events, enabling multi-messenger follow-up of these events. These alerts are distributed both via GCN and the Scalable Cyberinfrastructure to support Multi-Messenger Astrophysics[9] (SCiMMA) project in order to reach maximum consumers through two broad bases of subscribers. The alerts come in two types: notices, which are machine-readable and come in a variety of formats, and GCN circulars, which are human-readable.

There are five types of notices that may be sent out for a candidate event: Early Warning, Preliminary, Initial, Update, and Retraction. Early Warning Notices arise from dedicated pre-merger search pipelines, potentially enabling the release of alerts seconds before merger [162]. A first Preliminary Notice is sent out when an event candidate of a superevent exceeds the public FAR threshold. Following a timeout, the preferred event is determined and a second Preliminary Notice is then issued (even if the preferred event candidate remains unchanged). Both Early Warning and Preliminary Notices are sent out if the candidate passes automatic data quality checks [38]. These data quality checks are carried out by the Data Quality Report framework[10], and include checks for terrestrial noise and stationarity of the data, among others. In certain cases, such as when manual data quality checks yield suspicions on the astrophysical nature of the candidate, a Retraction Notice may be sent. If, however, the Early Warning or Preliminary Notice passes human vetting, then an Initial Notice is sent out accompanied by a GCN Circular to announce

[8] `https://emfollow.docs.ligo.org/userguide/`
[9] `https://scimma.org/`
[10] `https://docs.ligo.org/detchar/data-quality-report/`

Compact Object Properties					
Binary Type	Object	$m/M_\odot$ (min/max)	m distribution	Max a	a distribution
BNS	Primary	1.0 - 2.05	uniform	0.4	uniform & isotropic
	Secondary	1.0 - 2.05	uniform	0.4	uniform & isotropic
NSBH	Primary	1.0 - 60.0	m^{-1}	0.998	uniform & isotropic
	Secondary	1.0 - 2.05	uniform	0.4	uniform & isotropic
BBH	Primary	2.05 - 100	$m^{-2.35}$	0.998	uniform & isotropic
	Secondary	2.05 - 100	m^{1}	0.998	uniform & isotropic

Table 2.1: Distribution of intrinsic properties (component masses m and spins a) of binary systems in the injection sample. The spin distributions are uniform in magnitude and isotropic in orientation, as seen in the last column.

the detection. The final type of notice, an Update Notice, is used to send out improved estimates of alert contents based on parameter estimation when they become available. Included in each alert is an estimate of the event's probability of astrophysical origin, or p_{astro}. Both Initial and Update Notices are accompanied by human-readable GCN Circulars. This is broken up into four categories that sum to 1 by definition: P(BNS), P(NSBH), P(BBH), and P(Terrestrial), where the mass boundary between NS and BH is set at $3\,M_\odot$. The specifics of individual pipeline p_{astro} calculations can be found in the supplemental text. If the superevent is coincident with a GCN candidate, the various data products concerning the joint candidate are included, such as the time delay, joint FAR [11], and combined sky map if applicable.

Sky localization One of the key data products to enable multi-messenger follow-up is the rapid inference of the sky localization from GW observations. This sky localization consists of the posterior probability distribution of the source location in the sky. The sky localization, mapped either over a 2D map of right ascension and declination, or a 3D volume which also includes a distance estimate, is known as a "sky map." Sky localization (and parameter estimation more generally) is conducted in multiple stages once a candidate is identified.

For CBC sources, BAYESian TriAngulation and Rapid localization (`BAYESTAR`), a rapid sky localization algorithm [216], is used to generate sky maps, and may be

[11] `https://ligo-raven.readthedocs.io`

updated by `Bilby` (Section 2.2), a python-based parameter estimation pipeline that uses stochastic sampling methods [41, 200]. Sky maps from `BAYESTAR` are released with Preliminary Notices and sky maps from `Bilby` are released in Update Notices. Additionally, cWB also generates localizations for burst event candidates[141].

The sky map is stored as FITS file using the Hierarchical Equal Area isoLatitude Pixelization (HEALPix) [119] framework in the Multi-Order Coverage (MOC) representation [111]; flattened versions at a fixed HEALPix grid size are also available for superevents. MOC sky maps use adaptive division of the HEALPix grid, focusing areas of highest resolution on regions of highest probability with minimal information loss. Sky maps are made available both through the distributed alert as well as uploaded on `GraceDB`, where they are available for direct download. If there is a coincidence with a GRB candidate that has a sky localization, we compute the overlap integral with the GW sky map information. This weighted sky map is then included in the alert. The technique of combining sky maps from two independent datasets is laid out for the first time in [230], under the signal hypothesis of the Bayesian framework, and is presented in an accessible manner in [40].

EM-Bright EM-Bright[12] is a pipeline designed to assess whether a GW candidate is capable of producing an electromagnetic counterpart [67]. A rapid assessment of EM-Bright properties, `HasNS` and `HasRemnant`, is essential to trigger target of opportunity (ToO) follow-up by ground and space-based observatories. In this regard, `HasNS` and `HasRemnant` quantities are reported as a part of the automated and update discovery notices. The `HasNS` is the probability of the binary having a NS component, while `HasRemnant` is the probability of the merger leaving remnant matter post-merger in the form of dynamical or tidal ejecta.

The exact nature of EM emission from the merger is complex and depends on several factors like the properties of the ejecta, the NS EoS, and the BH mass and spin. Detailed analyses are required to assess the accurate properties of EM counterparts (see, for example, [210] for a review). These are, however, impractical in a real-time setting. Aside from theoretical uncertainties, measurement uncertainties predom-

[12] `https://git.ligo.org/emfollow/em-properties/em-bright`

inantly affect the assessment of EM-brightness in real-time. Note that the only real-time data products available from match filter CBC searches are the template parameters that maximize the likelihood of detection, and these values are internal. Bayesian parameter estimation from computationally cheap waveform models may be available in $\sim$ hours, as discussed later, but it is not available in the seconds after a trigger is registered. Hence, inference from the template parameters and real-time detection statistics is the feasible solution.

To this end, [67] showed the application of supervised machine-learning trained on a feature space involving the template parameters and detection statistics to make this inference. Training is done using large-scale simulation campaigns where the ground truth and the recovery of search pipelines are registered. The ground truth is labeled based on its intrinsic source-frame mass as having a NS component, or both mass and spin components as leaving remnant matter behind post merger, based on a phenomenological fit to numerical relativity simulations by [114]. The NS EoS plays a crucial role in the labeling as stiffer EoS favor tidal disruption, and therefore prefer larger ejecta masses. While in [67] a single, stiff NS EoS was used on conservative grounds, here we extend the analysis to multiple EoSs, and reweight the score based on Bayes factors computed against GW170817 [158] tidal deformability measurements for several NS EoSs presented in [116]. The score presented is therefore marginalized over several EoSs.

In addition to `HasNS` and `HasRemnant`, a new quantity `HasMassGap`, the probability that at least one component of the binary merger is in the lower mass-gap region i.e. source-frame mass between 3 $M_\odot$ to 5 $M_\odot$ is computed. The technique used in computing `HasMassGap` is similar to the original EM-Bright quantities, except the labeling is different and does not involve the knowledge of the NS EoS. The values reported for `HasNS` and `HasRemnant` use a nearest-neighbor classifier algorithm, while that used for `HasMassGap` use a random-forest classifier algorithm.

Similar to the sky maps, these quantities are updated from online parameter estimation samples, which are made publicly available $\sim$ hours after discovery. The parameter estimation samples allow for these quantities to be computed directly.

Low-latency Parameter Estimation

CBC signal candidates labeled as significant (see 2.2) are further investigated via automated Bayesian parameter estimation analysis with the `Bilby` library. It employs the nested sampling technique implemented in the `Dynesty` library [220] to explore the full parameter space of masses and spins, producing accurate inference results immune to biases included in the point estimates of masses and spins from search pipelines. It also takes into account uncertainties in detector calibration and marginalizes the posterior probability distribution over them.

To accelerate the analysis, we employ the reduced order quadrature (ROQ) technique [61, 175, 218], which approximates gravitational waveform with ROQ basis elements to reduce the computational cost of likelihood evaluations. The ROQ basis elements employed in the automated parameter estimation of O4 are presented in [176].

For BNS candidates, the analysis assumes that dimensionless spins have norms less than 0.05 and are aligned with the orbital angular momentum, and employs the `IMRPhenomD` waveform approximant [130, 138] to recover the observed signals. With the acceleration technique, the sampling completes typically in less than 10 minutes. The actual time from detection to upload of results from `Bilby` is a few tens of minutes since this analysis starts around 5 minutes after signal detection, and preparing input data and post-processing results take several minutes. Currently, the output of this analysis is not automatically made public but manually sent after it passes human vetting. Hence it is sent at the earliest when an Initial Notice is sent, and the actual latency of the update is higher than the latency of upload of the results.

In addition to this automated analysis, more costly manual analyses incorporating general spin configurations and tidal deformation of colliding objects may follow, depending on the significance of the signal. For candidates with higher masses, the automated analysis takes into account general spin configurations, and employs `IMRPhenomXPHM` [190] if its ROQ basis elements are available in the target mass range, and `IMRPhenomPv2` [124] otherwise. This analysis typically takes hours to complete. If this analysis completes after an Initial Notice is sent, its output is released as an Update Notice. The UV-optical radiation from a kilonova is expected

Latency Measure	Description	50% (s)	90% (s)
Superevents	$t_{\text{superevent}}$- t_0	9.4	18.1
CBC Events	t_{event}- t_0	12.3	41.4
Burst Events	t_{event}- t_0	72.3	671.3
Early Warning Events	t_{event}- t_0	-3.1	2.9
GW Advocate Request	$t_{\text{ADV\_REQ}}$- t_0	12.7	40.1
GCN Preliminary Sent	$t_{\text{GCN\_PRELIM}}$- t_0	29.5	171.8
Coincidence with GRB Found	$t_{\text{EM\_COINC}}$- t_0	32.9	44.4
RAVEN Alert Triggered	$t_{\text{RAVEN\_ALERT}}$- t_0	35.3	48.4

Table 2.2: Measured latencies for a number of steps in the pipeline. t_0 corresponds to the event merger time reported by the pipeline, while $t_{superevent}$ and t_{event} correspond to the time of superevent or event creation. For the case of superevent latencies, t_0 is determined by the preferred event.

to fade away within $\sim$ 48 hours [7], so parameter estimation updates within $\sim$ hour are sufficient for follow-up purposes.

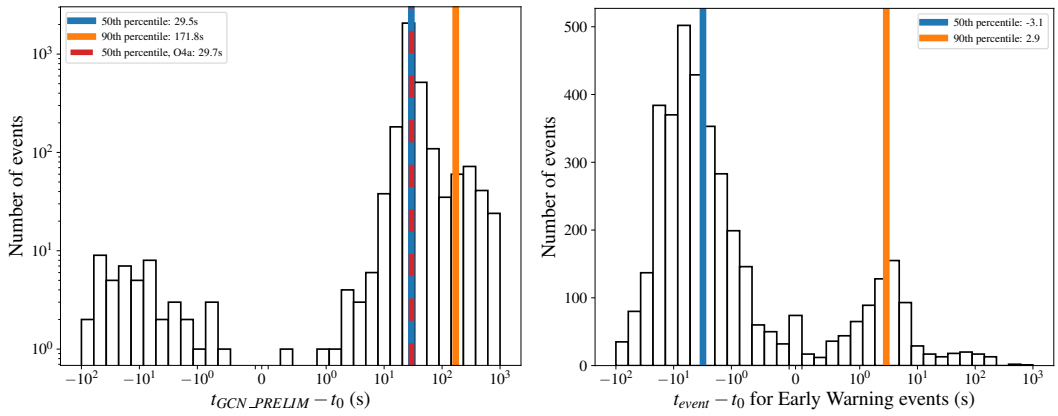

Figure 2.2: *Left*: Histogram of latencies for the sending of the GCN preliminary alert. We compare the MDC latencies to the median latency measured during O4a. The O4a measurement includes data calibration, construction, and transfer time, while the MDC latencies do not. *Right*: Histogram of latencies for Early Warning alerts. t_{GCN_PRELIM} corresponds to the time the GCN preliminary alert is sent, t_0 corresponds to the preferred event merger time, and t_{event} corresponds to the time of event creation.

Mock Data Challenge

To create a background for the MDC, we consider the stretch of data taken between Jan 05, 2020 – Feb 14, 2020 by the LVK instruments during O3. A total of 5×10^4 simulated CBC waveforms with mass and spin distributions mentioned in Table 2.1 are injected into the O3 data with an interval of $\sim$ 1 minute between injections. The optimal network SNR is greater than 4 for all the injections. This is done to prevent "hopeless" injections, which are improbable to be detected in reality. The `IMRPhenomPv2_NRTidalv2` waveform approximant is used to consider matter effects in case of NS components of the injections. For BHs, the same waveform is used with the tidal parameters set to zero. In order to label a component as a NS, the SLy [65] NS EoS is used, which allows for a maximum mass of $\sim$ 2.05 $M_\odot$. Hence, in this scheme, the tidal deformability of component masses above this limit are set to zero consistent with being a BH. In particular, all injections above the SLy maximum mass are assumed to be BHs, and the appropriate relative rate is used for the same. These injected signals are primarily recovered by CBC pipelines, and occasionally by Burst pipelines. In the MDC exercise, we have focused most of our analysis on the output and data products of the CBC pipelines. We also note that the injection rate density used in the study is artificially high and not representative of the true discovery rate in O4. We expect $O(10^2)$ CBC detections during the full duration of O4 [139], compared to $O(10^3)$ of detections across the 40 day MDC cycle. Therefore, quantities like p_{astro}which rely on the background distribution, may not be the true representation as compared to a realistic signal density. The CBC injection set consists of 40.9% BNS, 35.8% NSBH, and 23.3% BBH injections.

The events are distributed uniformly in co-moving volume assuming flat ΛCDM cosmology with $H_0 = 67.3$ km s^{-1} Mpc^{-1} and $\Omega_m = 0.3$ based on Planck 2018 results mentioned in Table 1 of [28]. The BNS systems are distributed up to a maximum redshift of $z = 0.15$, the neutron-star black-hole systems up to $z = 0.25$, and the BBHs up to $z = 1.9$. The simulated strain is projected on to the detector geometries, shifted in time to the time of experiment and streamed as 1 s segments for the search and annotation pipelines to analyze in real-time (see Section 2.2). The triggers and their annotations were reported in `GraceDB` for post processing studies.

This exercise is repeated in several cycles for benchmarking analysis and will continue internally during the observing run to continuously track improvements in the alert infrastructure and provide avenues for pipelines to test their changes. The numbers reported here are those from a single cycle of 5×10^4 injections where the status of most analyses were close to their O4 configurations.

2.3 Results

Latency Measures

Due to the desire for timely follow-up by the multi-messenger community, a key feature of the LLAI is dissemination of results as quickly as possible. The goal for the LLAI system is to send alerts for events within 30 s of merger time; this number sets the timescale for comparison below. Here, we perform a systematic study of the alert latency for three of the key pieces of the pipeline (a fourth, the data calibration, construction, and transfer between sites, which takes $\sim$ 5-10 s, is not captured here, as well as latency from the ingestion and redistribution by GCN or SCiMMA). Latency comes primarily from these three components: (i) the search pipelines, (ii) the event orchestrator `GWCelery`, and (iii) `GraceDB`. We note that technical issues during the MDC may also cause some high-latency outliers, so the results presented are conservative when excluding the time needed for transfer and construction of the strain data.

To calculate event latencies, we compare the time an event is created and appears on `GraceDB` to that of the known merger time. For GW event candidates during an observing run, the merger time is defined as the time the signal peak reaches Earth's center. For CBC pipelines, we find a median (90%) latency of 12.3 s (41.4 s); for Burst pipelines, we find a median (90%) latency of 72.3 s (671.3 s) We then compare this number to the creation of a superevent; we find a median (90%) latency of 9.4 s (18.1 s). The median superevent latency is lower than the event latency simply due to the fact that the superevent may be created upon the first trigger, and that a superevent often consists of multiple events. We also make the same measurement for Early Warning alerts, shown on the right of Figure 2.2; we find a median (90%) latency of -3.1 s (2.9 s). Considering the joint candidates, we find that it takes a median (90%) latency of 32.9 s (44.4 s) to find a coincidence with a GRB injection

and a median (90%) latency of 35.3 s (48.4 s) to trigger a RAVEN alert.

Once the event(s) have been created, there is a request for human vetting of the alert, called the Advocate Request; we find a median (90%) latency of 12.7 s (40.1 s) to notify the advocate. To measure the latency of event communication to the community, we also measure the latency for sending of the GCN preliminary alert, which occurs for superevents that pass automated data quality checks; we find a median (90%) latency of 29.5 s (171.8 s). We show this statistic for GCN preliminary alerts on the left of Figure 2.2. This latency reported specifically measures that time until the GCN preliminary label is applied. We also compare this to the measured median GCN latency during O4a, the first half of O4, which, including data calibration, construction, and transfer time, is 29.7 s. The agreement between the GCN latency during the MDC and O4a shows our latency has not increased, and has likely slightly decreased as the O4a measure includes data calibration, construction, and transfer time, while the MDC measurement does not. Table 2.2 shows a compilation of these latency statistics for comparison. The number of candidate events within a superevent was not shown to have a noticeable effect on the latency of that event and its corresponding preliminary alert.

Probability of astrophysical origin

Each CBC pipeline uploads its own estimate of p_{astro}, as described in Section 2.2, and is inherited by the superevent if the event is the preferred event. By matching the MDC's injected parameters to the recovered p_{astro}values from pipelines, we can test the accuracy of p_{astro}. Figure 2.3 shows the recovered P(source) for true sources (e.g., P(BNS) for true injected BNS events), for superevents which pass the public alert FAR threshold. In matching these injections, we place the cut between NS and BH at 3 $M_{\odot}$. "True" terrestrial events correspond to superevents which were not temporally matched to injections, indicating that they arose from detector noise. Injected BBHs are typically recovered confidently, with the vast majority resulting in P(BBH) > 0.5. Over 90% of BBHs are recovered with P(BBH) > 0.9. The P(BNS) and P(NSBH) distributions are less confident than the P(BBH) distribution; about 10% of true BNSs and NSBHs are recovered with P(BNS) or P(NSBH) < 0.1. As a check against contamination across P(BNS) and P(NSBH) due to errors in recovered

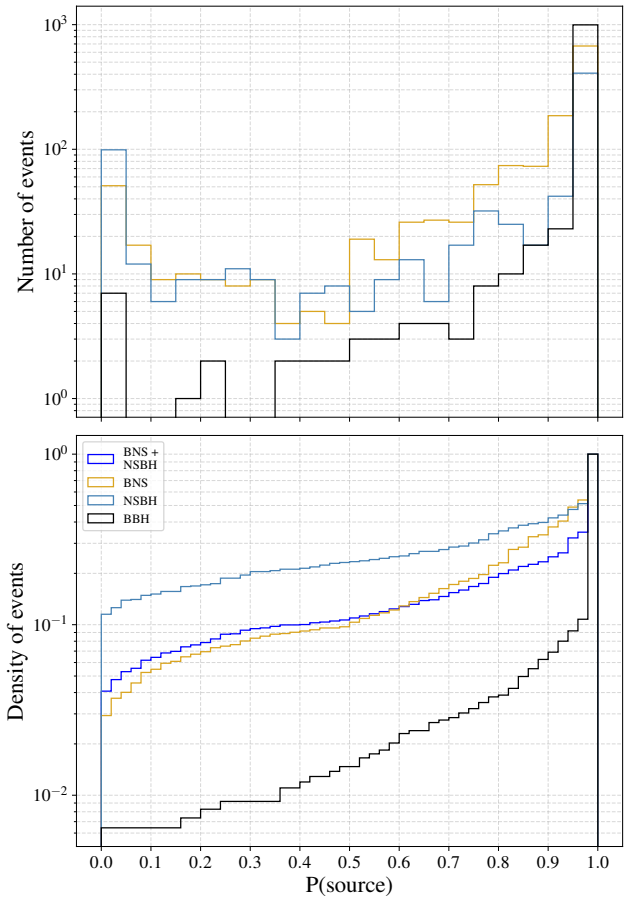

Figure 2.3: *Top:* Histogram of "preferred" recovered P(source) for true sources, for superevents which pass the significant public alert threshold. The possible source classes are BNS, NSBH, and BBH. This excludes early warning events, which were not fully functional during the time of this analysis. *Bottom:* Cumulative density of the same data. We also include the distribution of correctly recovered BNS or NSBH, which checks for contamination between the two classes due to misrecovered secondary masses or effects from varying the NS/BH mass boundary. As this is a cumulative histogram, the fraction of events above a certain P(source) corresponds to the TPR. We see that the majority of events with a P(source) greater than 0.5 correctly recover the injection source type.

masses or variation in the definitions of the mass border between NS and BH (i.e., the Tolman–Oppenheimer–Volkoff mass), we can also look at the distribution of the sum of P(BNS) and P(NSBH) for injected events where $m_2 \leq 3M_\odot$. This is

shown in the bottom panel of Figure 2.3. Compared to P(BNS) or P(NSBH) alone, P(BNS) + P(NSBH) performs slightly better, with $\sim$ 75% of true BNS or NSBHs receiving P(BNS) + P(NSBH) $>$ 0.9, compared to 60% and 65% for NSBH and BNS respectively. If instead we use a threshold P(source) of .5, we find a TPR of $\sim$ 98 % for BBH, $\sim$ 90 % for BNS, and $\sim$ 76 % for NSBH injections.

Sky localization

In order to evaluate localization performance, in the following, we focus on three metrics: (i) localization area, (ii) retrieved median distance of the source, and (iii) searched area. Localization area is the area (measured in deg^2) which encloses a given probability contour in the sky map; in this paper, we use 90% as the total cumulative probability threshold. Retrieved distance refers to the median of the distance distribution along the line of sight of the injected sky position. Searched area is the smallest 2D area, starting with the regions of highest probability, that contains the true location of the source; it represents a measurement of the sky area that a telescope with a small FOV relative to the sky map size would need to cover before imaging the true location. We refer the reader to [216] for more details on these parameters, and use the `ligo.skymap`[13] package to compute all metrics.

When evaluating sky map performance, we consider the preferred events for superevents that fall under the significant FAR threshold before trials factor and were detected by more than one interferometer. We exclude single detector triggers as most resulted from injections that occurred during a portion of O3 replay data where one or more detectors was not in science mode, and they may have sky localizations on the order of the entire sky. For the left half of Figure 2.4, we also exclude a slice of parameter space for injected NSs of mass $\leq 2M_{\odot}$ with spins $\geq .05$, as these events may not have a match within the pipeline template banks.

Further, in the top left panel of Fig 2.4, we show the accuracy of `BAYESTAR` sky maps through a *P-P* plot. *P-P* plots of this format show the fraction of injections found within a given credible interval across all levels of credible intervals. The three gray lozenges around the diagonal shows the three different levels of confidence

[13] `https://git.ligo.org/leo-singer/ligo.skymap`

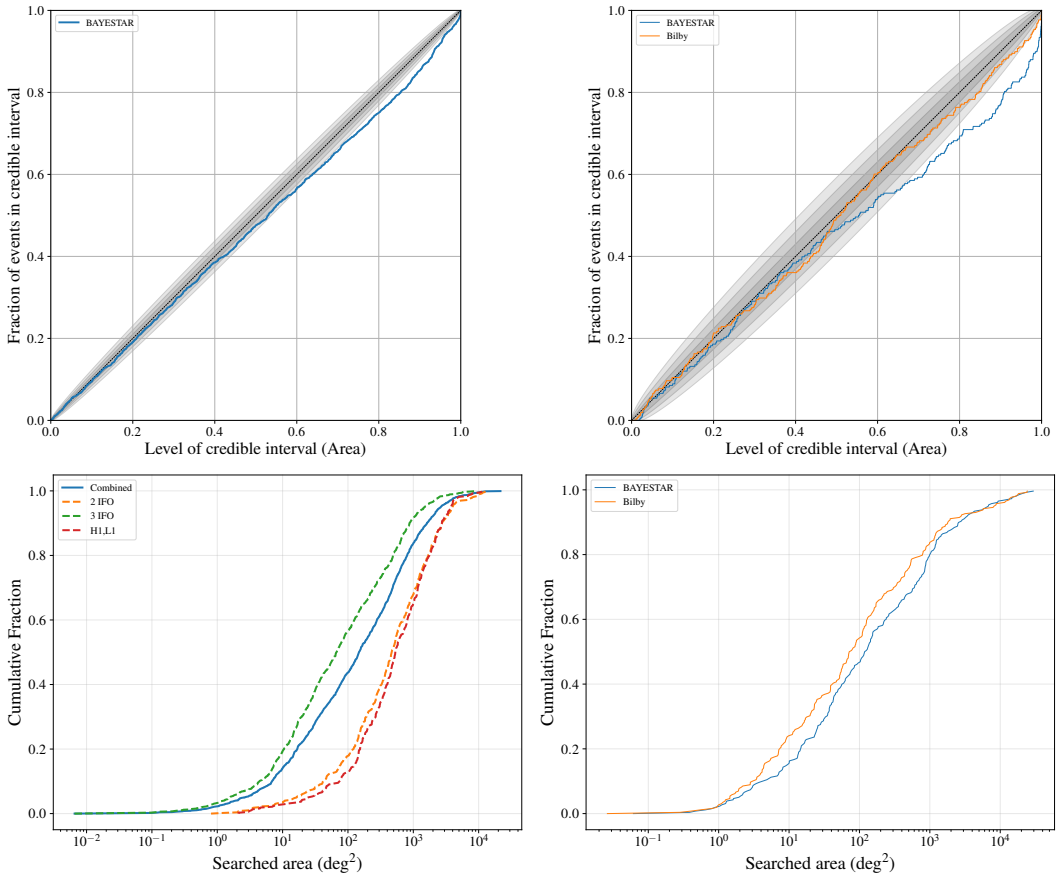

Figure 2.4: *Top left:* A *P-P* plot showing the `BAYESTAR` sky map statistics for the preferred event. The credible intervals shown in gray are based on the total number of events. *Bottom left:* Cumulative histograms of `BAYESTAR` searched area for all events (blue), compared to two-interferometer (Orange), three-interferometer (Green), and Hanford (H1) and Livingston (L1) (Red) events. We see that the three-interferometer events produced smaller searched areas than the two-interferometer events by almost an order of magnitude as discussed in Section 2.3. *Top right:* A *P-P* plot showing the performance of `BAYESTAR` (blue) and `Bilby` (orange) generated sky maps for BNS events. The credible intervals shown in gray based on the total number of such preferred events where both `BAYESTAR` and `Bilby` sky maps are available. *Bottom right:* Cumulative histograms showing searched area statistics for `BAYESTAR` and `Bilby` sky maps. We observe that `Bilby` sky maps give a lower searched area and tend to be more precise than their `BAYESTAR` counterparts as discussed in 2.3.

(1-3 σ) for the combined BAYESTAR map sample. We find that the BAYESTAR sky maps fall just outside the credible intervals for higher credible intervals. This tells us BAYESTAR slightly overstates the precision of its sky localizations. We also show the performance of different pipelines that detected the preferred event in the given sample. See the supplement for a discussion of sky map performance for each individual pipeline. In the bottom left panel of Figure 2.4, we show the cumulative trend of the searched area statistics from the combined sample. We find a median searched area of $100-200\,\mathrm{deg}^2$. We then compare the two interferometer events with the three interferometer events, to show that the latter produced smaller searched areas. We also include a line for the two-interferometer case where Hanford (H1) and Livingston (L1) specifically observed the event, as those are the interferometers currently being used during O4. The LVK Alert User Guide[14] provides up-to-date information about the interferometers currently in use. In this case, we see that there is marginal difference between the two-interferometer line, and the H1, L1 line.

Further, we probed the accuracy of the BAYESTAR sky maps compared with Bilby sky maps for BNS events. This comparison *P-P* plot and searched area histogram can be seen in the right half of Figure 2.4 for the preferred event of all BNS superevents for which both sky maps were produced. In this figure, we include the events excluded in the left half of Figure 2.4 to demonstrate Bilby's performance even without the cuts. A plot with those cuts applied can be found in the appendix. From the top right panel of Figure 2.4, we see that the BAYESTAR sky maps tend to sag below Bilby's implying that their precision was overstated as compared to Bilby. There is a trade-off between the latency of BAYESTAR sky maps available with the Preliminary GCN alert, and the improved accuracy of the Bilby sky maps that are available with the completion of parameter estimation. The cumulative searched area plot in the bottom right panel of Figure 2.4 shows that typically Bilby sky maps have lower searched area by factor of 2 or more.

EM-Bright

We show the performance of the EM-Bright classifiers [67] across all four CBC pipelines via their Receiver Operating Characteristic (ROC) curves. For this, the

[14] https://emfollow.docs.ligo.org/userguide/

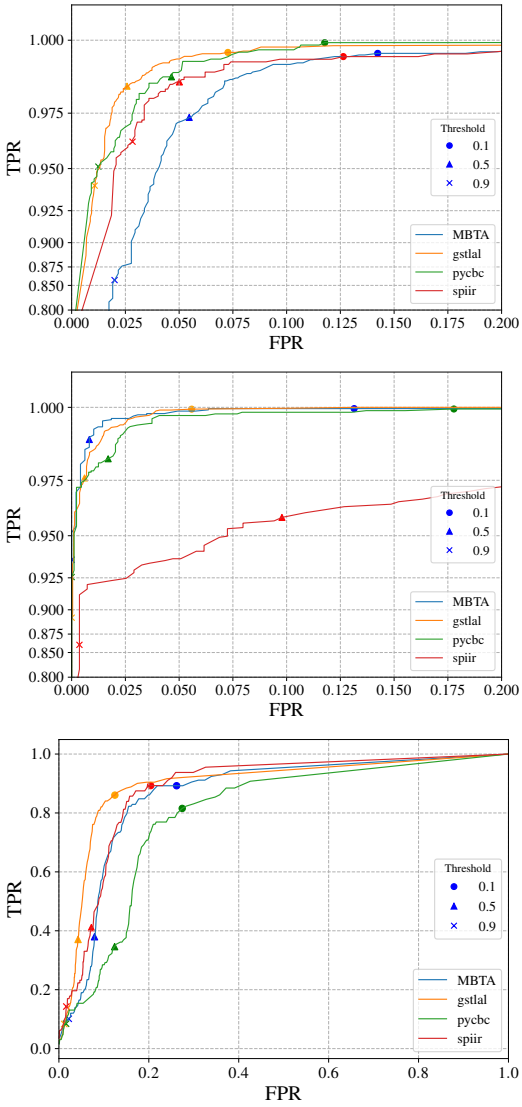

Figure 2.5: The ROC curves for the different EM-Bright classifiers are shown here for MDC events. The top, middle, and bottom panels refer to `HasRemnant`, `HasNS`, and `HasMassGap`quantities respectively. The markers denote different representative thresholds along the curve.

NSBH boundary is chosen according to SLy EoS but the probabilities are EoS marginalized, meaning we include some uncertainty in the NS EoS in our classifications. The markers indicates three different representative thresholds (a score

above which events are considered to be positively classified as the source type in question) along the ROC curves for each pipeline. Since EM-Bright classifiers are trained only using GstLAL injection-recovery data [67], evaluating performance across various CBC pipelines is crucial.

In Figure 2.5, we see that the `HasRemnant` quantity for all four pipelines has greater than a 95% TPR for a 5% False Positive Rate (FPR). In the middle panel, we see (GstLAL, PyCBC, MBTA) perform consistently at $\sim$ 97% TPR at $\sim$ 3% FPR for `HasNS` classifier. The SPIIR pipeline purity is slightly lower compared to the other pipelines, $\sim$ 94% TPR at the same misclassification fraction. One possible mitigation technique is to use more training data from the pipeline in the training process. In the last panel, for `HasMassGap`, we see (GstLAL, SPIIR, MBTA) perform with $\sim$ 80% TPR at $\sim$ 20% FPR while PyCBClags slightly below. We expect to enhance the performance of these classifiers further in the near future by retraining the classifiers using the O3 MDC data and considering all pipelines.

2.4 Conclusion

In this paper, we present the performance of the low-latency alert infrastructure and associated data products based on the O3 MDC. A large simulation campaign of compact binaries i.e. BNS, NSBH, and BBH are injected into a stretch of real data from O3. The data is taken through the entire end-to-end alert infrastructure starting from the search pipeline, data products computation, and alert generation. We demonstrate that for full bandwidth searches automated preliminary alerts, excluding time for data transfer and construction, are delivered with a median latency of $\lesssim$ 30 s, which is an improvement since O3 [18]. We show that low-mass BNS injections are successfully detected by early warning searches. Annotations and alert delivery is achieved for a significant fraction of such signals before merger time with a median alert latency of $\sim$ -3 s. It is to be noted however, that alert delivery before merger time does not happen for all early warning events. In addition, through the use of this MDC dataset, we demonstrate that the data products, produced in the same workflow as planned for O4, are statistically consistent with simulated values.

The p_{astro}values giving probability of an astrophysical BBH, BNS, or NSBH were

found to correctly classify the source for the majority of events. For a threshold P(source) of 0.5, we find a TPR of ~ 98 % for BBH, ~ 90 % for BNS, and ~ 76 % for NSBH injections.

The distribution of injected sky positions is found to be well recovered by the sky maps produced by both `BAYESTAR` and `Bilby`, as evidenced by Figure 2.4. `BAYESTAR` provides low-latency sky maps that slightly overstate the precision, while `Bilby` provides improved accuracy upon completion of parameter estimation. The median searched area is found to be $100 - 200$ deg^2, with slight variations between method and pipelines. We observed that `Bilby` sky maps have better precision, and typically gave a smaller searched area.

The EM-Bright values corresponding to the probabilities of `HasNS` and `HasRemnant` have a TPR of above ~ 95% at ~ 5% FPR across GstLAL, PyCBC, MBTA pipelines. The SPIIR pipeline is performing similarly for `HasRemnant` but its performance is slightly lower for `HasNS`. `HasMassGap`, on the other hand, has a TPR of above ~ 80% at ~ 20% FPR.

This paper presents the low-latency data products for O4 and their expected performance. Additional data products that expand on the current EM-Bright products for determining the likelihood of a kilonova are being developed that include predictions of mass ejecta for BNS and NSBH events, as well as peak magnitudes for corresponding kilonovae. We hope to make these data products public in the future.

2.5 Appendix: Individual Searches

This MDC study covers seven searches. There are six direct searches for gravitational waves, four CBC searches and two burst, and then one search that looks for coincidences between gravitational wave candidates found by the other searches and EM signals. Below we cover each those searches in detail.

GstLAL

GstLAL is a stream-based matched filtering algorithm capable of detecting GW signals within seconds of their arrival on Earth [103, 170, 228]. GstLAL uses a template bank of ~ 10^6 CBC waveforms in order to filter the full BNS, NSBH, and

BBH regions of the parameter space [206]. The template bank is divided into $\sim 10^3$ bins of time-sliced singular value decomposition (SVD) waveforms according to the Low Latency Online Inspiral Detection (LLOID) method [63]. These waveforms are used to filter the strain data producing an output SNR timeseries. Peaks in the SNR time-series which pass a threshold of 4.0 are stored as "triggers". These form candidates which may be coincident among two or more detectors or observed in only a single detector. Significance is assigned to each candidate using the likelihood ratio ranking statistic which is then mapped to a false alarm probability and corresponding FAR [62]. Candidates are finally uploaded to `GraceDB` after aggregating them across SVD bins by maximum SNR. GstLAL carries out both an early warning search, and a full bandwidth search. The early warning search targets low redshift BNS events that can be detected $\sim$ 10 - 60 s before merger, using templates with non-spinning component masses between 0.95 $M_\odot$ and 2.4 $M_\odot$ [205]. The full bandwidth search covers the entire CBC template bank parameter space.

GstLAL uses the multi-component FGMC method for assigning a probability of astrophysical origin to candidates [107, 134]. The probability that the signal originates from each CBC source category is also assigned. Triggers from each category (BNS, NSBH, BBH and terrestrial) are treated as realizations of independent Poisson processes. The rate of detectable triggers characterizing the Poisson process corresponding to each foreground category is approximated from the astrophysical rate estimates yielded by offline FGMC analyses of past observing runs while accounting for the change in sensitive spacetime volume between the past and ongoing runs. Misclassification among astrophysical source categories is accounted for by computing the probability of migration between injected and recovered templates across the entire bank semi-analytically under the Gaussian noise approximation [112]. With the rates and migration probabilities precomputed, $p_{astro}$is estimated in low-latency from trigger data comprising the likelihood ratio ranking statistic assigned to said trigger and the matched template.

MBTA

The MBTA pipeline carries out an early warning and main, full bandwidth search and performs matched filtering per frequency band to reduce computational costs

[43]. The main instance of the pipeline is searching for binaries with total masses ranging from 2 to 500 $M_\odot$ and mass ratio smaller than 50.

The pipeline includes signal-consistency checks to help distinguish astrophysical signals from background. As for the O3 offline analysis [36] the probability of astrophysical origin of GW candidate events is derived from the expected rate of astrophysical events and background candidates at the recovered chirp mass, mass ratio and ranking statistic. The foreground distribution has been estimated by performing injections of simulated BNS, BBH and NSBH signals into LIGO-Virgo O3a data which are then analyzed by the MBTA pipeline. This method computes $p_{astro}$and source classification. It has been extended to also provide EM-Bright information to determine the likelihood of an electromagnetic counterpart, as covered in Section C of the introduction in the main text.

PyCBC Live

PyCBC Live is a matched filtering pipeline designed to detect CBC events by comparing the incoming GW signal to a template bank of waveforms [89, 181]. Two PyCBC Live searches were employed; one is a full bandwidth search for a wide range of signals, the other is an early-warning configuration for which the templates are truncated at certain frequencies before merger [182]. The full bandwidth template bank contains 412,575 templates, covering total masses from 2 to 500 $M_\odot$, and mass ratios from 1 to 100 [203, 204]. The early-warning template bank contains ~4700 templates with component masses in the range 1 to 3 $M_\odot$, truncated at a set of frequencies designed to give early warnings at regular intervals before the merger.

The matched filtering algorithm produces a time series of SNR values, and only triggers with SNR $\geq$ 4.5 are considered for further analysis. The SNR is then re-weighted according to signal-consistency tests in each detector, and using multi-detector properties determined by the distribution of source extrinsic parameters (time difference, phase difference and amplitude ratios) for the signal population [180].

In order to assess the frequency of a coincident noise signal which would be ranked greater or equal to a given detection, PyCBC Live assigns a FAR value by comparing

the candidate to time-shifted background from the last several hours [181]. The FAR values for injections recovered during the MDC are subject to a substantial upward bias due to the high rate of high-SNR injected events, which significantly influences the background estimation.

For single-detector candidates, if strict criteria on signal consistency tests are passed, a FAR is assigned by comparing the candidate's re-weighted SNR to the noise trigger distribution via a template-dependent exponential fit [89]. The exponential fit is performed using the original data without injections, thus the FAR calculation for single-detector events is not subject to contamination from injections. Single-detector candidate signals are only considered for potentially electromagnetically bright signals, as BBH signals in low-latency are unlikely to yield multimessenger counterparts, and higher-mass templates are more susceptible to glitch contamination due to their shorter duration. Single-detector early-warning candidate signals are also not considered, as the poor localization of single-detector events is not of interest for pre-merger alerts.

For full bandwidth events, PyCBC Live also calculates the probability of astrophysical origin p_{astro}, based on the FAR value, the trigger SNR, the approximate distributions of signal and noise events over template chirp mass, and the sensitivities of observing detectors [91]. This $p_{astro}$is then combined with estimates of the relative probabilities of different source classes, also based on template chirp mass; for details see [233].

In the later part of O3, an additional step was performed to optimize event SNR over template masses and spins [89]. Our MDC results include additional events produced via this optimization; however, it was removed from the search configuration deployed at the start of O4 in order to reduce complexity and computational load. We do not expect major differences in the search outputs detailed here due to the change.

SPIIR

The SPIIR pipeline is designed to achieve lower delays in signal detection and differs from other pipelines in multiple aspects. As the name suggests, SPIIR uses

a time domain counterpart of matched filtering [127, 156] as its primary filter. This method breaks down millions of CBC templates into a few hundred thousand IIR filters to perform match filtering in the time domain, further accelerated by the use of GPUs. The SPIIR pipeline implements a computational coherent network search approach [55, 125] to select the GW candidate events with low-latency, achieved with the help of SVD [236]. The pipeline performs a full bandwidth search for BBH, BNS, and NSBH sources, and an early warning search for BNS and NSBH sources. SPIIR has demonstrated its performance in past LIGO-Virgo runs [74, 146].

SPIIR is introducing a new two-step p_{astro}calculation for O4. In the first step, the pipeline calculates the two-component $p_{astro}$of the trigger based on the FGMC two-component method by [107] and [134]. This assigns the probability of the trigger's astrophysical or terrestrial origin. In the second step, it further classifies the probability of astrophysical origin into NSBH, BBH, and BNS, based on the chirp mass method [233].

Coherent WaveBurst

cWB [96, 143, 144] is an excess power algorithm using minimal assumptions on the GW signature. cWB decomposes the GW strain data using a wavelet transform ([178]). It then selects coherent signal power in multiple detectors and applies a maximum likelihood approach to select GW events. The calculation of the likelihood over the sky allows for building a sky map that characterizes the probability of the GW source sky location. A new feature with respect to O3 cWB analyses has been implemented for the significance assessment - a machine learning algorithm based on XGBoost [173, 222]. In low-latency, cWB analyzes 180 s data segments overlapping every 30 s. The alerts are created up to a latency of around 1 minute.

oLIB

The omicron-LALInferenceBurst (oLIB) pipeline is a short duration ($\lesssim$ 1 second) unmodeled detection pipeline that is sensitive to a wide variety of sources that includes, but is not limited to, CBCs [161]. As such, oLIB makes very minimal assumptions about the astrophysical source type of the emission. The search is performed hierarchically. First, data from individual interferometers is analyzed

with the Omicron trigger generator algorithm [199]. Omicron identifies regions in the time-frequency plane of excess power. Triggers that are coincident in time and frequency between interferometers are then followed up with a coherent Bayesian analysis using LIB. LIB models the data with a single sine-Gaussian wavelet, calculating two Bayes factors. Each of these Bayes factors is expressed as the natural logarithm of the evidence ratio of two hypotheses: (1) a GW signal versus Gaussian noise (BSN) and (2) a coherent GW signal versus incoherent noise transients (BCI). Ultimately, these two Bayes factors are used to construct a likelihood ratio Λ that is used as the final search statistic.

RAVEN

RAVEN is a multi-messenger pipeline that searches for coincidence between GW candidates and other astronomical detections, such as gamma-ray bursts (GRBs) and neutrino bursts [70, 188, 230]. RAVEN ingests events submitted to the General Coordinates Network (GCN)[15] ([215]) into `GraceDB`, queries `GraceDB` to look for a corresponding GW candidate, and calculates the joint false alarm rate to determine whether to send a public alert.

2.6 Appendix: Low-Latency Infrastructure

GraceDB

`GraceDB`[16] is the central location that houses GW event candidates and analyses for transient searches. GW event candidate data in `GraceDB` can be viewed and manipulated on the web or through the use of a RESTful API. A permission structure exists to show only proprietary data to LVK users versus data that is available to the general public. State changes in `GraceDB` (which may take the form of new event uploads/annotations, new superevent uploads/annotations, log/file updates, etc.) are communicated to LVK and external users and processes via the `igwn-alert` system.

At its core, `GraceDB` is, architecturally, a standard Web/API application. `GraceDB` is hosted in a high-availability configuration in Amazon AWS. A PostgreSQL backend is powered by a Django web framework. External requests are served by Apache

[15] `https://gcn.nasa.gov/`
[16] `https://gracedb.ligo.org/`

acting as a reverse-proxy for a Gunicorn-based WSGI HTTP server. Files are stored on an NFS (Amazon EFS) filesystem, and low-latency analyses stream data from the detectors and upload candidate events to `GraceDB` via a representational state transfer (REST) API.

igwn-alert

`igwn-alert`[17] is an alert data stream based on kafka and leverages SCiMMA (Scalable Cyberinfrastructure for Multimessenger Astronomy) infrastructure for data delivery. Client-side tools maintained by IGWN Computing and Software (CompSoft) allow users to listen and respond to `igwn-alert` messages. `igwn-alert` messages are machine-readable (JSON), so as to be read by automated followup processes. `igwn-alert` listeners act on notifications from `GraceDB` and are used to launch follow-up analyses (e.g., Superevent creation, parameter estimation, sky localization, etc.). Results from follow-up analyses are then uploaded and stored in `GraceDB`. `GraceDB` and `igwn-alert` are the orchestrator and source-of-truth for external observers and follow-up processes.

GWcelery

`GWCelery`[18] is a distributed task queue application for orchestrating and annotating GW alerts. At its core, it is a `Celery` [64] application. Some of the advantages of a distributed task queue, like `Celery`, include handling of asynchronous tasks, easy scalability based on requirement, designing *canvas* workflows, setting conditions to retry individual parts of a canvas, easy error handling, and running periodic tasks. `Celery` is fault tolerant and preserves the state of tasks in a result backend, and communicates with it using a messaging broker. `GWCelery` uses `Redis` as both a broker and a backend for the application. `GWCelery` also contains a `Flask` application to provide a web interface to run routine tasks which require human interaction with the application, or to handle situations where part of a canvas are to be executed manually, overriding the automated processing. We run all `GWCelery` processes using HTCondor, used for job scheduling in the LIGO Data Grid. The

[17] `https://igwn-alert.readthedocs.io`
[18] `https://git.ligo.org/emfollow/gwcelery`

major subsystems of `GWCelery` are:

- The listener for IGWN alerts, which is pubsub system that `GraceDB` uses to push machine-readable notifications about its state.
- The Superevent Manager, which clusters individual GW candidates into *superevents*[19].
- The External Trigger Manager which listens for and correlates candidates from external facilities to spot coincidences with GW events.
- The GCN and SCiMMA alert producer that disseminates GW candidate information for external consumption.
- The Orchestrator, which executes the per-event annotation workflow. This involves having the data products ready for sending alerts for the superevent, and broadly includes computing rapid sky-localization and source properties for individual events and updating the state of the superevent, and launching parameter estimation runs.

2.7 Appendix: Mock Data Challenge

Pipeline Performance

In order to demonstrate LLAI readiness for O4, we compare CBC triggers and their corresponding data products uploaded to `GraceDB` with the injection set. Triggers and injections are matched using the merger time; all triggers within 1 s of an injection are matched to that injection and included in the analysis. This process removes noise triggers between injections and ensures trigger times correspond to an injection time window. It is possible for a few noise triggers to be coincident with injection time window. Matching triggers to injections allows us to evaluate data products by comparing the results with the injected quantities. These data products include basic parameter estimates, sky maps, and p_{astro}values. However, not all injections are recovered in the form of trigger. There are four main reasons why an

[19] `https://gwcelery.readthedocs.io/en/latest/gwcelery.tasks.superevents.html\#gwcelery-tasks-superevents-module`

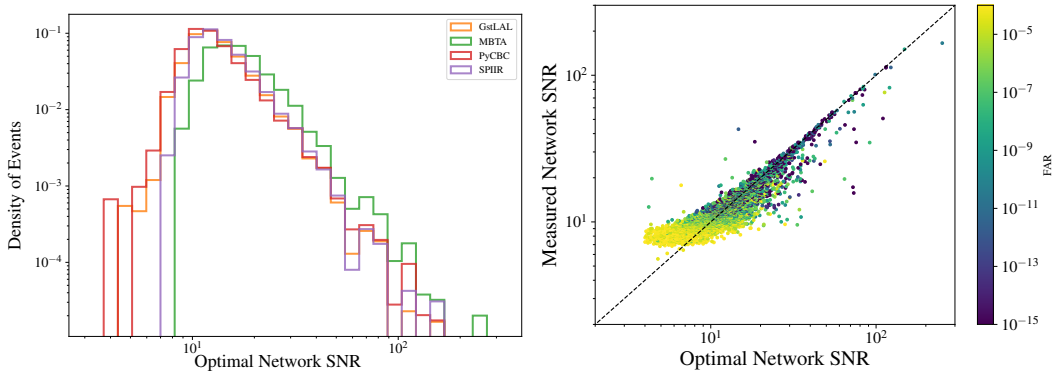

Figure 2.6: *Left*: Histograms of the optimal, or injected, network SNR, normalized for each CBC pipeline, for triggers below the significant FAR threshold. All pipelines were found to recover injections across the range of injected SNR values. *Right*: The measured network SNR recovered during the MDC compared to the optimal, or injected, network SNR with the points colored by FAR. We find SNR is recovered more accurately for higher values.

injection may not be found by the search pipelines: (i) some injections are distant or have a low SNR, due to the cosmological distribution preferring larger distances, and can not be distinguished from background noise, (ii) there are stretches of the O3 replay where one or more detectors were not operational and in science mode, (iii) there were some temporary technical issues on computing resources used during this MDC cycle, and (iv) there may be data quality issues that overlap with an injection, such as loud or long glitches.

As mentioned in the previous section, for this analysis we focus on the MDC cycle used for the review of pipeline performance, which ran from February 16 through March 28, 2023 consisting of 5×10^4 injections. During this MDC, 1489 BNS, 1105 NSBH, and 1920 BBH injections were recovered. As seen in the left panel of Figure 2.6, each of the CBC search pipelines, PyCBC, GstLAL, MBTA, and SPIIR, successfully uploaded events below the significant FAR threshold across the range of injected SNR values. Burst searches, cWB and oLIB, make little assumptions of source type, and so are only considered for measures of latency in this paper. We plot the simulated vs. recovered network SNR in the right panel of Figure 2.6. In general, signals with moderate to high SNR are recovered well, with some bias

at low SNRs due to the FAR threshold imposed for upload. Additional scatter in SNR recovery is expected since the simulated optimal SNRs were calculated using fixed detector sensitivities, whereas actual detector data has significant fluctuations in sensitivity over time. The optimal SNRs for injections are calculated using a global average PSD, instead of using a local estimate of the PSD, which may cause some of the off-diagonal outliers. For the 4514 GW injections found, we created 469 multimessenger coincidences by injecting simulated GRB candidates at times surrounding the GW injections. We found 356 of these joint candidates triggered a RAVEN alert as a result of passing the *significant* FAR threshold.

HasMassGap Training

The data set used for training contains additional mass-gap injections done separately on O2 dataset whereas the feature space used to train the algorithm is the same as [67] – a five-dimensional space involving the triggered template masses $m_{1,2}$, the aligned dimensionless spins, $\chi_{1,2}$, and the network SNR. The duration from which the mass-gap injections were taken from is shown in Table 2.3.

GstLAL Chunks used for Training `HasMassGap` Classifier

Start date	End date
Sun 2017-01-22 08:00:00 UTC	Fri 2017-02-03 16:20:00 UTC
Tue 2017-02-28 16:30:00 UTC	Fri 2017-03-10 13:35:00 UTC
Fri 2017-06-30 02:30:00 UTC	Sat 2017-07-15 00:00:00 UTC
Sat 2017-08-05 03:00:00 UTC	Sun 2017-08-13 02:00:00 UTC
Sun 2017-08-13 02:00:00 UTC	Mon 2017-08-21 01:05:00 UTC

Table 2.3: Calendar times for the detector chunks of LIGO O2 data. We consider the mass-gap injections performed by the GstLAL search in these duration along with previously existing set in [67] for the study.

2.8 Sky map Performance

Continuing the discussion from the main text, we provide additional plots to characterize sky map performance. In Figure 2.7, we compare the 90% localization area with the recovered distance, with the points colored by searched area. We find a positive correlation between the retrieved distances of the injection and the localization areas, as in general greater distances lead to larger localizations. This

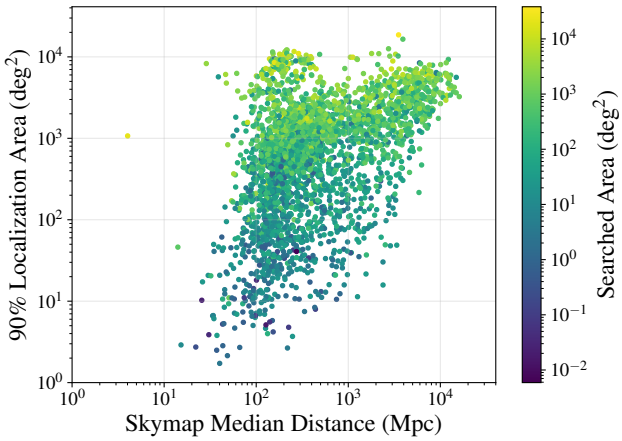

Figure 2.7: Sky localization distribution as a function of recovered median distance from the sky map, with the trend of Searched Area for the preferred event. We find the more distant the event, the larger is the localization area. The color bar shows that the searched area associated with the event also increases with the localization area and sky map median distance.

same trend was seen for the searched areas for these superevents, where at greater distances one would typically encounter larger searched areas, which aligns with the expected behavior.

The remaining plots shown in this section are provided to demonstrate `Bilby` and `BAYESTAR` produce accurate sky localizations from injections found by each individual CBC pipeline. The left panel of Figure 2.8 shows the `BAYESTAR` performance of preferred events from each pipeline falls within the credible intervals besides minor deviations, and uses the same set of events and cuts as found in Figure 4 in the main text. To demonstrate performance for the events most likely to be subject to extensive follow-up, the right panel of Figure 2.8 specifically presents only BNS injections that pass the cuts applied in Figure 4 in the main text. With these cuts we find the combined performance for both `Bilby` and `BAYESTAR` falls within for within credible intervals.

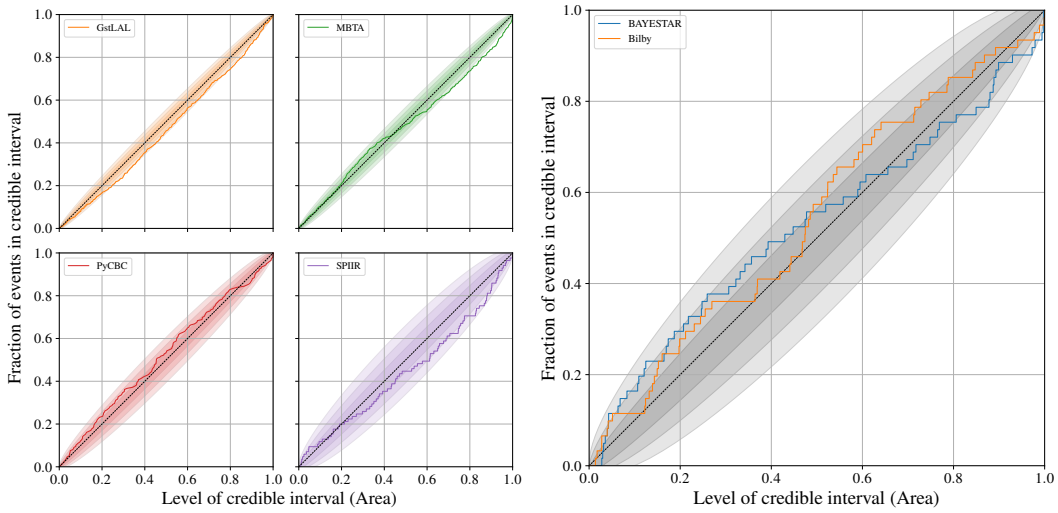

Figure 2.8: *Left*: *P-P* plots showing `BAYESTAR` performance for pipelines that detected the preferred event. All the sky maps generated show that the performance of the sky maps are within the confidence bands. *Right*: *P-P* plot comparing the performance of `Bilby` and `BAYESTAR` for BNS injections likely to be the subject of follow-up, including the cuts on mass, spin, FAR, and number of interferometers as covered in the main text. We see both `Bilby`'s and `BAYESTAR`'s performance is improved and within the credible intervals when including these cuts.

Chapter 3

ASTROPHYSICAL OR TERRESTRIAL: MACHINE LEARNING CLASSIFICATION OF GRAVITATIONAL-WAVE CANDIDATES USING MULTIPLE-SEARCH INFORMATION

Work reproduced from:

Seiya Tsukamoto, Andrew Toivonen, Holton Griffin, Avyukt Raghuvanshi, Megan Averill, Frank Kerkow, Michael W. Coughlin, Man Leong Chan, and Leo Singer. Astrophysical or Terrestrial: Machine learning classification of gravitational-wave candidates using multiple-search information. *arXiv*, 8 2025. doi: 10.48550/arXiv.2508.14242.
A.T. advised S.T., H.G., A.R., M.A., and F.K. as undergraduate researchers for this work, and it has been submitted to PASP.

3.1 Motivation

Low-latency gravitational-wave alerts provide the greater multi-messenger community with information about the candidate events detected by the IGWN. Prompt release of data products such as the sky localization, FAR, and p_{astro}values allow astronomers to make informed decisions on which candidate gravitational-wave events merit ToO follow-up. However, false alarms, often referred to as “glitches", where a gravitational-wave candidate, or trigger, is the result of terrestrial noise, are an inherent part of gravitational-wave searches. In addition, with the presence of multiple gravitational-wave searches, different searches may have varying assessments of the significance of a given trigger. As a complement to quantities such as p_{astro}, we provide a Machine Learning (ML) based approach to determining whether candidate events are astrophysical or terrestrial in nature, specifically a classifier that utilizes information provided by multiple low-latency search pipelines its feature space. This classifier has a performance an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.96 and accuracy of 0.90 on the Mock Data Challenge training set and an AUC of 0.93 and accuracy of 0.86 on events from the O3 observing run.

3.2 Determining Significant of Gravitational-Wave Triggers

The first direct detection of gravitational waves, which originated from the BBH merger event GW150914 [6], provided a new means for studying the universe. The subsequent joint detection of gravitational waves from the BNS merger GW170817 [9] and the optical counterparts in kilonova AT2017gfo [7, 86, 217] and GRB170817A [10, 118, 208] led to enormous growth of interest in multi-messenger astronomy and GW follow-up searches. Observations of kilonovae, astrophysical transients that can be produced by BNS or NSBH mergers, are of great interest across nuclear astrophysics [135, 137, 166, 217, 234], cosmology [8, 83, 94], and for tests of General Relativity [45, 88, 104]. Observations from GW170817 and AT2017gfo specifically revealed r-process nucleosynthesis took place and powered kilonova emissions, as evidenced by the presence of r-process elements in the ejecta post-merger [135, 151, 155, 171].

In this era of multi-messenger astronomy, where we hope to combine GW detections with EM or neutrino observations, the timely release of GW alerts and derived data products is essential. The search for GW events and their associated counterparts [5] continues with O4, which began on May 23, 2023[1] and is in progress as of the time of writing. Kilonovae are faint and short lived, and can have large sky localizations [53, 77, 101, 105, 106, 123, 142, 202, 212, 214, 235], making them difficult targets for follow-up. With these challenges in mind, sending out accurate, low-latency alerts is important for maximizing our chances of locating a transient and observing the peak of emissions.

Prompt release of GW candidate event information relies on GW search pipelines and the LVK's LLAI [12, 69]. Alerts are distributed both via GCN[2] and the Scalable Cyberinfrastructure to support Multi-Messenger Astrophysics[3] (SCiMMA) and come in two types: notices, which are machine-readable and come in a variety of formats, and GCN circulars, which are human-readable.

There are two categories of low-latency GW searches: "modeled" CBC [22] searches

[1] `https://observing.docs.ligo.org/plan`
[2] `https://gcn.nasa.gov/`
[3] `https://scimma.org/`

and “unmodeled” Burst [17] searches. Modeled CBC searches target waveforms from BNS, NSBH, or BBH mergers, while unmodeled searches look for GW signals of generic morphologies from a wide variety of astrophysical sources, which may include core-collapse of massive stars, magnetar star-quakes, and other sources, in addition to CBCs [16, 17]. Candidate GW events that meet a certain FAR threshold are released publicly, in the form of *low-significance* and *significant* alerts [69]. Up-to-date information on the current FAR threshold and the searches running in low-latency can be found in the LVK userguide[4].

Significant alerts, those which pass the most stringent FAR threshold, are often the subject of downstream analyses and targets of ToO searches. Candidate events may be uploaded by multiple search pipelines. When these events are temporally coincident around the same signal, they are grouped together into what is termed a *superevent*. This superevent corresponds to a single candidate GW detection, and the individual candidate event with the highest SNR that passes the FAR threshold becomes the *preferred event*. When passing the public FAR threshold, certain information about preferred event becomes public, and is the subject of our alerts. It is possible for the preferred event to change as new events are uploaded, and given the different search pipelines, it is also possible that different events within a superevent are not in perfect agreement on certain aspects of the event.

If after additional analyses a significant candidate event is deemed to not be of astrophysical origin, a retraction is issued for that candidate event. As opposed to astrophysical events, these events are referred to as “terrestrial", meaning they have resulted from sort of environmental or detector noise source. Terrestrial events, often referred to as noise triggers, and their eventual retraction are an inherent part of GW searches as setting a FAR threshold defines your tolerance for false alarms and the rate at which you expect them. However, minimizing the number of retractions, promptly retracting noise triggers, and providing information on the likelihood a candidate event is astrophysical are all important parts of improving the reliability of GW candidates.

There have been over 200 significant GW candidate events detected during O4,

[4] `https://emfollow.docs.ligo.org/userguide/`

excluding retractions. While only a small fraction of significant events are retracted–alongside the first 200 significant events were 25 retractions–there is demand for data products for assessing confidence in GW candidate events. Certain events, such as those likely to include at least one NS, are often targets of extensive ToO follow-up, using notable person-power and telescope time. Avoiding using such resources on noise triggers is important for efficient follow-up of GW events, especially looking forward to next generation GW detectors such as LISA [44], Cosmic Explorer [102, 196], and Einstein Telescope [3] where the expected rate of events is much higher.

For training our classifiers, we focus on four CBC searches which participated in the pre-O4 MDC [69]: (GstLAL [103, 170, 228], MBTA [43], PyCBC [89, 181], and SPIIR [75, 146]). Each search pipeline has independent methods of estimating the noise background and calculating quantities such as FAR and p_{astro}, the probability of astrophysical origin; for details see [43, 89, 127, 156, 170, 231]. There have also been efforts to combine p_{astro}information from multiple search pipelines for a "unified" p_{astro}[46], which would provide a similar use as our classifier.

Alongside the search pipelines, the LLAI encompasses parameter estimation and source classification efforts which are included in alerts and crucial for multi-messenger searches [42, 52, 67, 216]. Additional machine learning-based searches [168] and rapid parameter estimation techniques are also in development [68, 90]. Currently, pipeline p_{astro}values and `GWSkyNet` [23, 59, 66, 194, 195] provide statements of confidence for the astrophysical nature of candidate events; we envision our classifier(s) as a complementary data product to those efforts, in addition to possible future data products aimed at multi-messenger sources [52, 225].

3.3 Data

Mock Data Challenge

The MDC [69] is a real-time simulation campaign which was carried out in anticipation of O4 to stress-test the LLAI and evaluate the performance of GW searches and data products. The MDC takes a 40 day stretch of background data taken by LVK interferometers during O3, and injects CBC waveforms across the BNS, NSBH, and

BBH parameter space. A total of 5×10^4 simulated CBC waveforms of varying mass and spins were injected into the aforementioned stretch O3 of data, with the GW search pipelines running on the data stream. This results in thousands of recovered injections with significant FARs, and a realistic dataset for downstream analyses.

Due to the numerous recovered injections and comprehensive nature of the dataset, we use this MDC as our training data. However, we must note the that the rate of injections is artificially high and not representative of the true discovery rate in O4. This impacts FAR values, and consequentially p_{astro}calculations reliant on those FAR values. As certain searches rely on a local background estimation for FAR calculations, their FAR values are subject to a upward bias due to large density of high SNR injections in the data [69]. As discussed later in Sec. 3.6, this will impact the performance of our classifier when evaluating accuracy on real candidate events from an observing run, such as those from O3. Resolving this upward bias in FAR is a high priority issue ahead of the MDC that will be carried out in anticipation of aLIGO's and AdVirgo's fifth observing run (O5), using a stretch of O4 data. One proposed solution would be to have pipelines evaluate the background based on the data stream alone, without any injections present.

O3 GW Detections

We also evaluate the performance of the classifier on real GW events from a past observing run, to ensure the model can still make reliable predictions for these real events. In order to assess our accuracy, however, we must assume true labels for our dataset, and provide events with true labels of both astrophysical and terrestrial. We assign all pipeline events corresponding to a confirmed significant O3 event in the Gravitational-Wave Transient Catalog (GWTC) catalog [22] a true label of astrophysical, and take the retracted significant events as our terrestrial class. This is a reasonable assumption, as all events in the GWTC catalog have undergone significant human-vetting, while retracted events can be assumed to be terrestrial due to their data quality issues, glitches present in the strain, and lack of significant GW signal. As a note for O3 events, since a superevent is a grouping of individual pipeline events, there are some individual pipeline events which correspond to a significant superevent that do not pass the significant FAR threshold on their own.

We still assign these events a true label of astrophysical and evaluate our model on them. In the future, once candidate events from O4 have been added to the GWTC catalog, we can also release the performance on those events.

3.4 The Bayes Factors

We take a Bayesian approach. Bayes theorem is stated as:

$$P(H_i|D) = \frac{P(D|H_i) \cdot P(H_i)}{P(D)}, \tag{3.1}$$

where D is the GW event and H_i is the hypothesis. $P(H_i|D)$ and $P(D|H_i)$ are conditional probabilities defined as the posterior and likelihood under hypothesis i. $P(H_i)$ and $P(D)$ are defined as the prior and marginal probabilities. As the marginal is constant throughout different hypotheses, the relation between the ratio of priors and posteriors between hypothesis' can be described through the Bayes Factor:

$$BF_{i,j} = \frac{P(D|H_i)}{P(D|H_j)}. \tag{3.2}$$

`BAYESTAR` provides the signal to noise coherent Bayes' Factor (BSN) [216, 232] which is defined as:

$$\log(BSN) = \log\left(\frac{P(D_{HL}|H_{signal+noise})}{P(D_{HL}|H_{signal})}\right). \tag{3.3}$$

`BAYESTAR` also provides the coherent versus incoherent Bayes' Factor (BCI) which is defined as:

$$\log(BCI) = \frac{\log(BSN)}{\sum \log(BSN_{incoherent})}, \tag{3.4}$$

where $BSN_{incoherent}$ is the BSN of each interferometer independently. These are the two Bayes Factors that will be used to aid in classifying the GW events from the MDC data [69].

FAR and SNR

FAR is a measure of significance produced by each of the pipeline for each potential candidate [4], this metric helps determine whether or not we send out an alert for a

candidate event, and lower FAR values are correlated with events of astrophysical nature. This is defined as:

$$FAR(\rho) = \int_{\rho}^{\rho_{\max}} \Lambda_n p_n(\rho')(d\rho'), \tag{3.5}$$

where Λ_n is the mean Poisson rate of signal and noise triggers, ρ is the SNR, and p_n is the PDF [60]. The SNR, signal-to-noise ratio, provides an estimate of the signal strength is by comparing it to the noise produced in a given detection. High SNR events are correlated with events of astrophysical nature. However, glitches with high power and non-Gaussian distribution can cause large SNR to be present in non-astrophysical triggers [9]. SNR is correlated with the BSN as seen in Fig. 3.1. The FAR and SNR will be implemented as additional features for the classification algorithms used.

p_{astro}

The probability of astrophysical origin, termed p_{astro}, is provided for GW event candidates and used to help determine whether an event is astrophysical or terrestrial in nature. $p_{astro}$is dependent on FAR and SNR, and each search pipeline calculates p_{astro}differently [69]. Our data product is meant to be complementary to p_{astro}, and we later evaluate our model performance with p_{astro}as a baseline for comparison in Sec. 3.6.

3.5 Model Architectures

Noise triggers are often single-pipeline events, so there is little overlap between CBC pipelines on terrestrial events. Because of this, the relationship between all data products from different CBC pipelines was not able to be modeled in a single architecture. In contrast, `BAYESTAR` is run on every event, so the relationship between its data products, specifically the BSN and BCI, with individual CBC search pipelines was able be modeled. In order to resolve this limitation, classification was split into two stages. In the first stage, we classify using each CBC search pipeline's preferred event and `BAYESTAR`'s data products in order to assign scores between 0 and 1 for each pipeline. If a pipeline does not recover an event, a value of -1 is assigned to indicate the lack of detection. In the second stage, the individual were

used to assign a final score for the superevent using the same model architecture for both stages of classification. The hyperparameters are shown in appendix Sec. 3.8, and were each tuned through cross-validation grid search.

K-Nearest Neighbors

KNN is an algorithm that makes predictions based on the proximity of a test data point in the parameter space to the data points in the training set. The algorithm assigns labels and probabilities based on a set of nearest neighbors of size K defined by a distance metric. We use the open-source Python implementation of KNN from `scikit-learn` for our classifier. More detail is provided in the appendix, Sec. 3.8.

Random Forest

RF is an ensemble learning method used for classification by constructing many decision trees during training. The hierarchical structure of the tree starts with a top node that contains the entire data set, with each tree trained independently of each other. At each branch the data is subdivided into two child nodes from a decision that is based on the features given. The splitting repeats, recursively, until it reaches a predefined stopping criteria. The end result is the best prediction given the dataset and the labels. We use the open-source Python implementation of RF from `scikit-learn` for our classifier. Additional details are provided in the appendix, Sec. 3.8.

Neural Network

Multilayer Perceptron (MLP) is a simple NN that contains a set of hidden layers and activation functions that perform linear and non-linear operations in the data. Using the gradient decent optimizer `ADAM`, neurons in each hidden layer can be trained to reduce binary cross entropy loss. We used the open-source Python implementation of NN's from `PyTorch` for our classifier. The hyperparameters were tuned through a cross-validated grid search resulting in 6 hidden layers with 10 neurons each. More detail is provided in the appendix, Sec. 3.8.

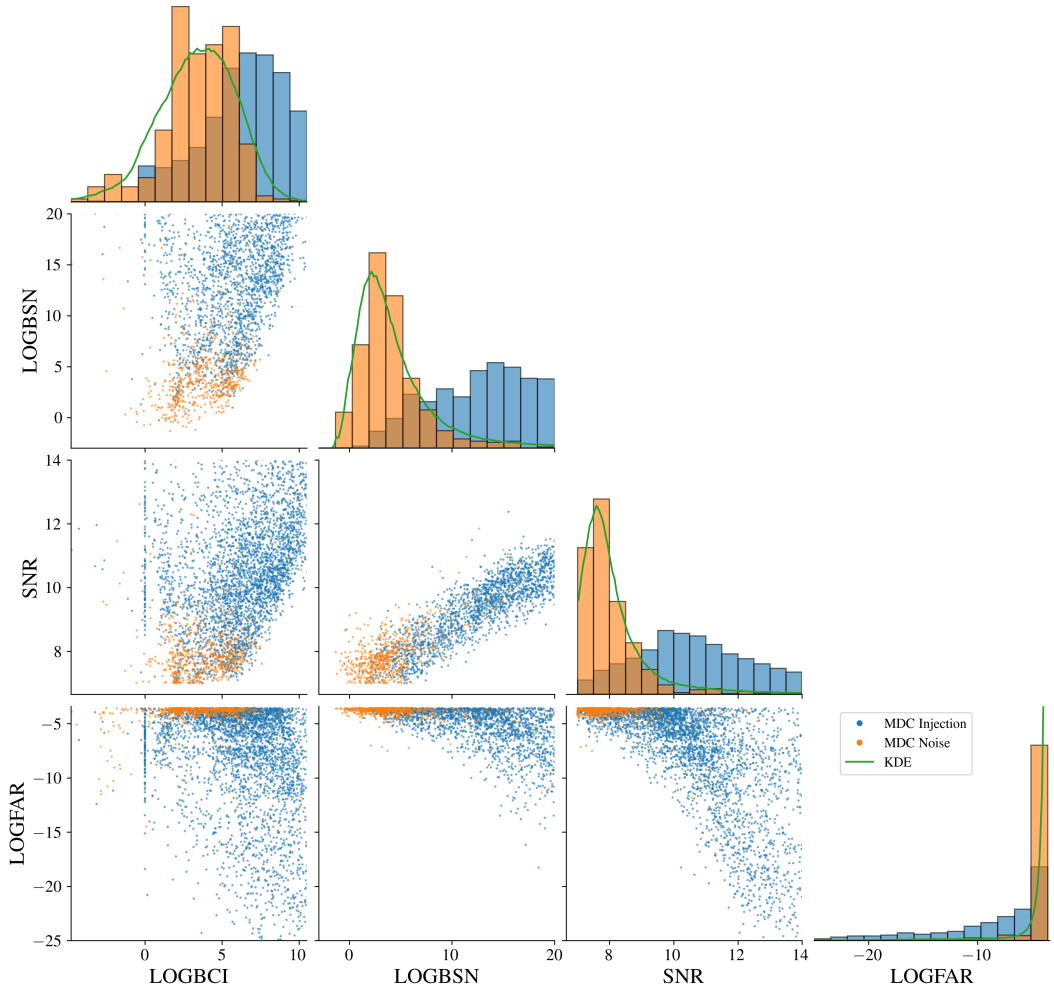

Figure 3.1: Corner plot of the GstLAL and BAYESTAR pipeline data products from MDC events windowed around noise events. The histograms are normalized and the green line shows the PDF of the KDE.

Feature Selection

We utilize event information from four CBC pipelines, GstLAL, MBTA, PyCBC, and SPIIR, in our feature space. This multiple pipeline approach allows us to capture the information provided by each of these searches. In addition, events detected by a only single pipeline are much more likely to be terrestrial than those found by multiple pipelines, while events detected by multiple-pipelines are much more likely to be astrophysical. Specifically, we used the highest SNR event upload that passes the FAR threshold for each pipeline in a superevent.

We use the following features:

The Bayes factors, BCI and BSN, are derived from the sky map FITS file for each event generated by `BAYESTAR` [216]:

1. Logarithm of the Bayes factor for the signal versus noise hypothesis (LOGBSN)
2. Logarithm of the Bayes factor for the coherence versus incoherence hypothesis (LOGBCI)

And inputs derived from low-latency pipelines:

3. Preferred SNR
4. Logarithm of preferred FAR (LOGFAR)

Dataset

Our dataset is composed of simulated MDC events run through CBC search pipelines and `BAYESTAR` [69]. The events are then labeled astrophysical or terrestrial by cross-matching the detected events with the set of injections. FAR and SNR were taken from the highest SNR event that passes the FAR threshold for each pipeline, also referred to as the preferred event. The total number of events labeled as astrophysical is 5726 and the total terrestrial is 1834, resulting in a class imbalance. In order to pre-process LOGBSN, LOGBCI, SNR, and LOGFAR into a Gaussian distribution,

a logit transform was used on LOGFAR, as it non-Gaussian and instead asymptotic towards high values, and a Yeo-Johnson transform was performed on LOGBSN and SNR, as they were a skewed Gaussian distribution. Using the transformed data products, a multivariate Gaussian kernel density estimation (KDE) was used to sample an adequate number of terrestrial events.

A KDE is a way to estimate the PDF of a variable based on applying kernels at data points. The KDE is described by

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - x_i}{h}) \tag{3.6}$$

where h is the kernel bandwidth, and n is the number of data points. We chose a multivariate gaussian kernel, and calibrated the KDE for each pipeline separately. The kernel bandwidth was chosen using a leave one out maximum likelihood estimate.

A multivariate KDE was created for each pipeline, and an example of one can be seen in Fig. 3.1 for the pipeline GstLAL. With a sufficiently large dataset, we could instead sample events for each combination of the four pipelines, but since 1778 of 1834 terrestrial events are only recovered by a single pipeline, there are too few events for this approach. This combination of many more astrophysical events than terrestrial, and the fact that these are not evenly distributed across the different pipeline configurations means there is a large class imbalance. In order to resolve the class imbalance and ensure sufficient astrophysical and terrestrial events for each pipeline configuration, class weights equivalent to the inverse of class size were used in KNN and RF. Downsampling was used for NN while resampling the training set when the model performance plateaus.

3.6 Results

Performance of Classifiers on simulated MDC Events

Fig. 3.2 shows the ROC curves for the all classifiers using ten fold cross validation. NN and RF had an AUC of 0.95 and 0.96, respectively, outperforming KNN which had an AUC of 0.94. Each of our classifiers outperform p_{astro}. The KNN ROC exhibits a straight line for higher FPR values as it operates in the regime with predominantly multiple-pipeline and true-label astrophysical events. The correspond-

ing confusion matrices in Fig. 3.3 show each classifiers' performance compared to p_{astro}using a threshold of 0.10, 0.52, 0.37, and 0.50 for KNN, RF, NN, and p_{astro}respectively. The thresholds of the three classifiers trained here were decided by finding the threshold in the ROC that minimized the distance to the top left corner. Since p_{astro}had large overclassification in opposite classes between MDC and O3, so the threshold was kept at 0.50. The RF and NN slightly outperform the KNN for astrophysical events, with RF outperforming the others for terrestrial events.

Fig. 3.4 shows permutation feature importance of the RF model as an example demonstrating that each pipeline has differing feature importance. This validates our approach to have separate models for each pipeline. In general, LOGBSN was found to have the highest feature importance across pipelines. Although permutation tests show that LOGFAR contributes the least to per-pipeline importance, it still provides a marginal gain, particularly for single-pipeline events, so we retain it for completeness. As explained in [69], the high rate of injections in the MDC leads to biased FAR values in our training set, something we hope would be resolved for a future pre-O5 MDC and should increase the LOGFAR feature importance.

In Fig. 3.5 we compare our classifier's performance to that of `GWSkyNet` [66]. This plot displays ROC curves for the subset of MDC events `GWSkyNet` was run on, specifically those with FAR $\leq$ 2 per day, network SNR $\geq$ 7, and two or more individual detector SNR $\geq$ 4.5. We see that our classifiers and `GWSkyNet` significantly outperform p_{astro}, and have similar performance to one another on this subset of events. We intend for our classifier to be complementary to `GWSkyNet`, with the major difference being our classifiers take advantage of multiple-pipeline information.

Performance on O3

Fig. 3.6 shows the ROC curves for the all classifiers on the O3 data. The NN and RF have matching performance with an AUC of 0.93 outperforming KNN with a performance of 0.92. The corresponding confusion matrix in Fig. 3.7 shows all classifiers' performance in comparison to p_{astro}using the same thresholds described in Sec. 3.6. KNN and RF has disagreement in three astrophysical events and four

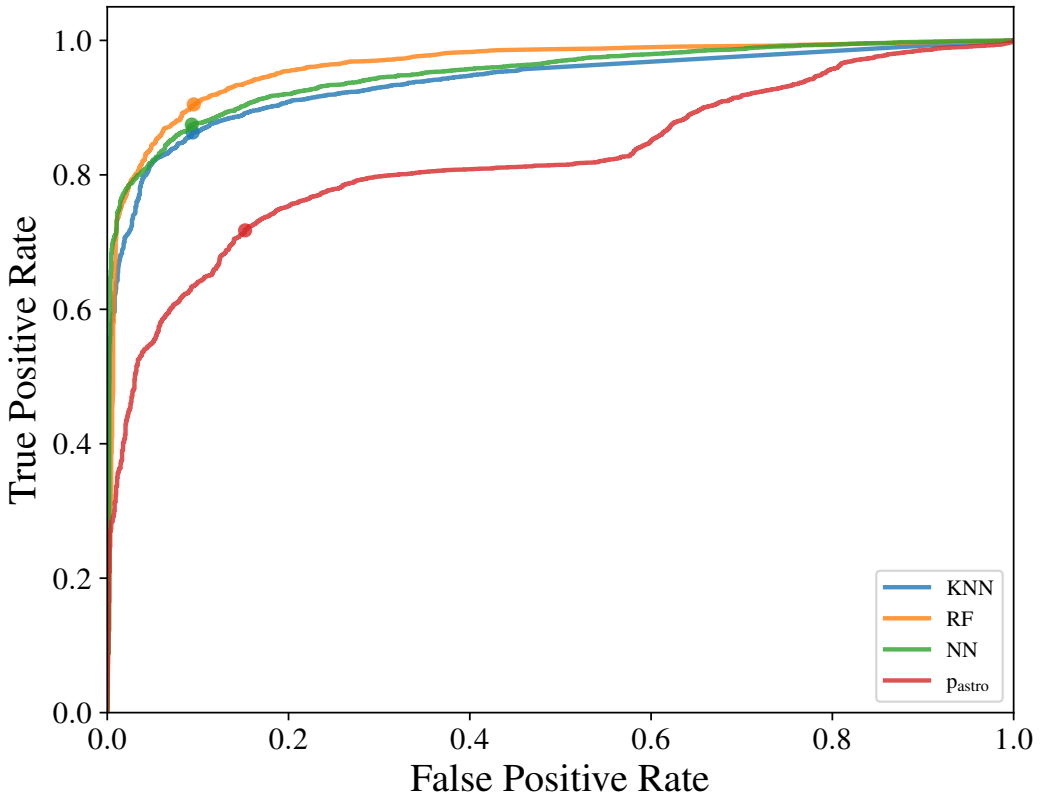

Figure 3.2: ROC curves obtained from the MDC. We use the the $p_{astro}$from the preferred event of each superevent for comparison. The straight line in KNN model is a result of neighborhoods that only include astrophysical events most prominently in multiple pipeline detections. The points on the curve denote the threshold that minimizes the distance to the upper left corner of the plot.

retractions. RF and NN has disagreement in four astrophysical events and four retractions. NN and KNN has disagreement in three astrophysical events and six retractions.

The true labels for the O3 dataset are defined differently than for the MDC dataset. Confirmed significant events that make the GWTC catalog [22] are given true astrophysical labels, while retracted events are given terrestrial labels. Seeing as each of these events have undergone extensive human-vetting to be a part of the catalog, or to be retracted, this is a fair assumption. However, it is possible there are

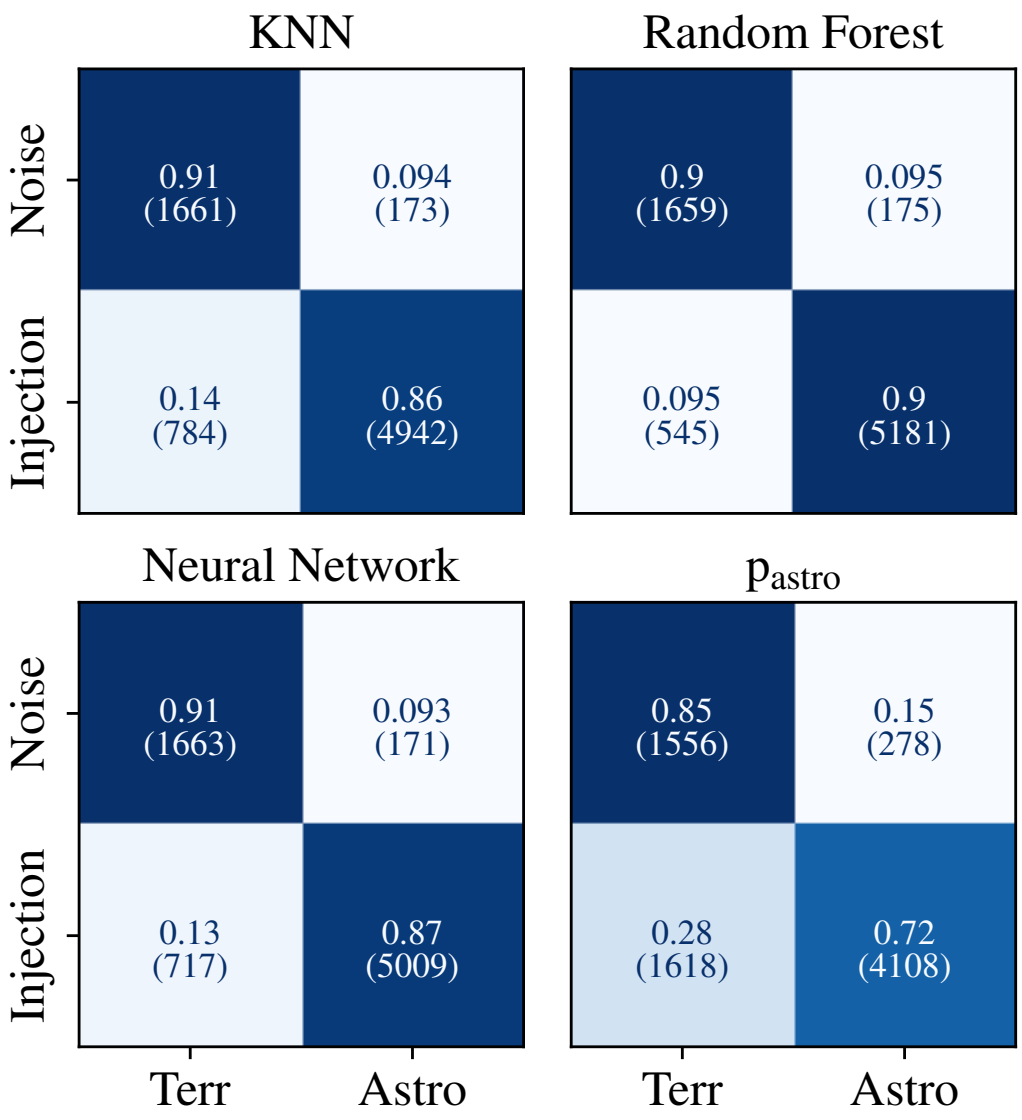

Figure 3.3: Confusion matrix of model performance on the MDC dataset. We used the the p_{astro} from the preferred event of each superevent for comparison.

terrestrial events that have made it into the catalog, which may slightly impact our results. Luckily, training is not impacted as that is done with MDC injections. With this in mind, all O3 superevents must have at least one event with FAR greater than the significance threshold, and all individual pipeline's preferred event within that

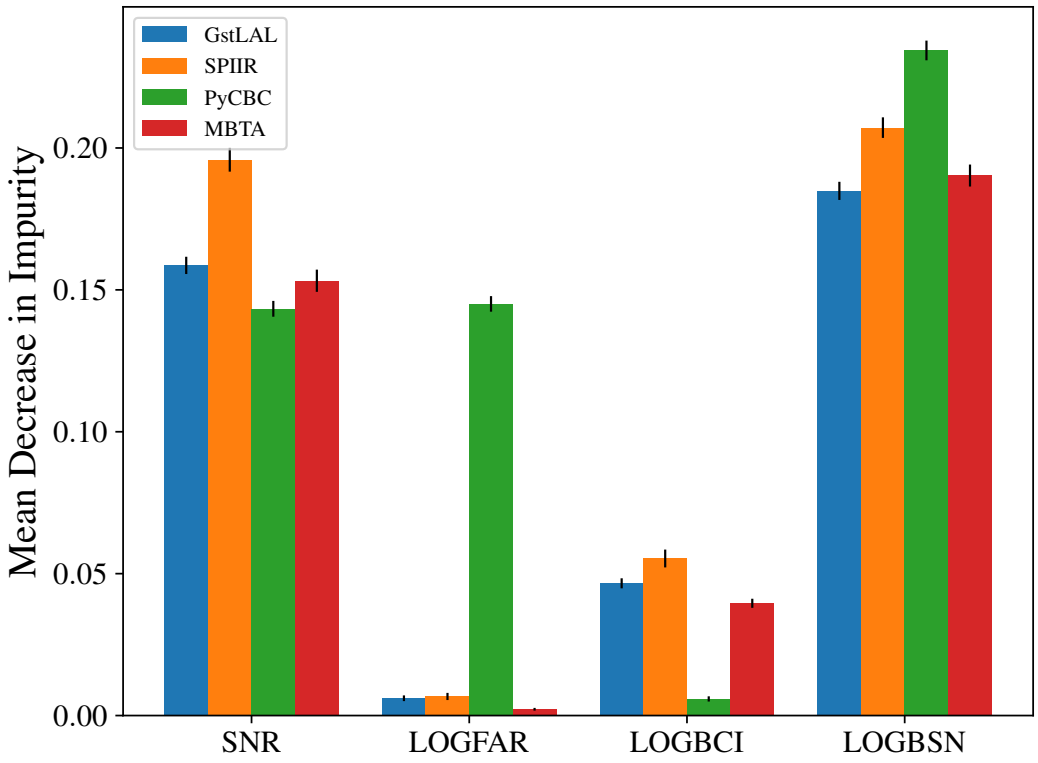

Figure 3.4: Permutation importance for the random forest classifier for each pipeline.

superevent are then considered. There is one exception in O3, S200105ae, which was originally subthreshold but published after manual review. All O3 retractions were single pipeline events meaning there is no subthreshold event labeled as a retraction. MDC noise triggers above the significance threshold, corresponding to a LOGFAR of -7.1, are rare, as can be seen in Fig. 3.1. The disparity in the proportion of labels between MDC and O3 is also accentuated by the upwards bias of FAR from the high rate of injections in the MDC [69]. Overall, this bias may cause the classifier for at least one pipeline to have a higher rate of false positives in the O3 dataset compared to the MDC dataset.

Data quality issues in O3 events can also impact the performance of our model as the model only takes in the data products produced in low latency. High amplitude glitches can cause terrestrial events to have parameters similar to astrophysical

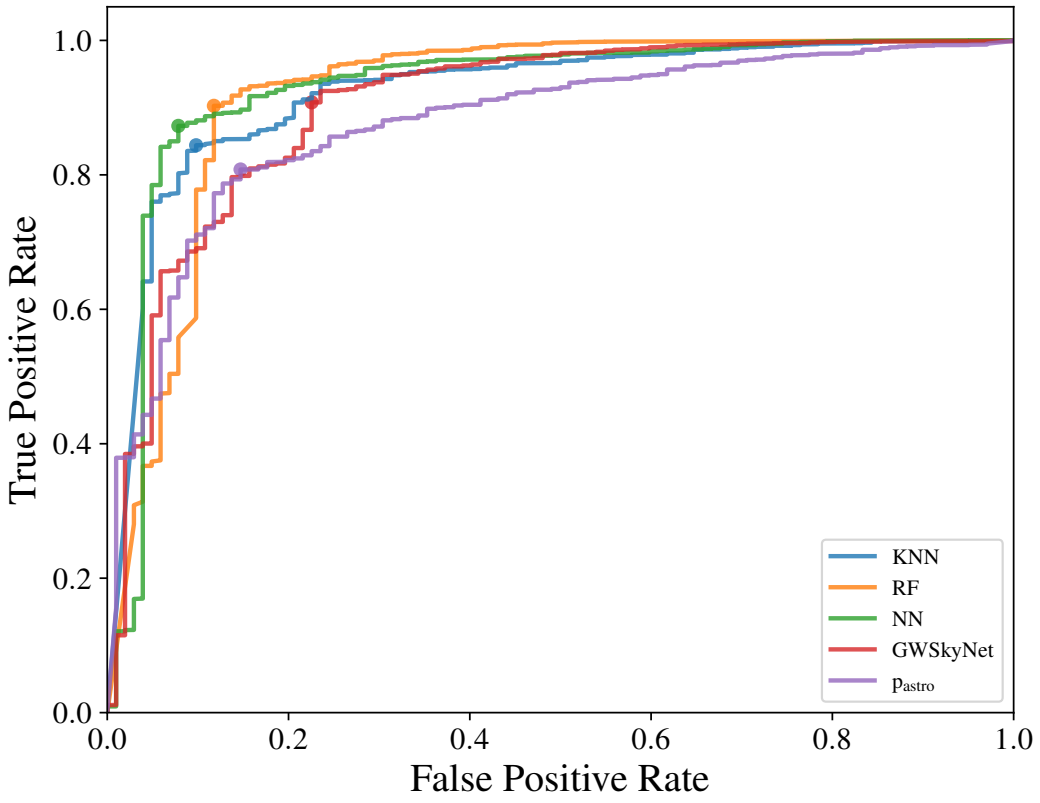

Figure 3.5: ROC curves from the subset of MDC events `GWSkyNet` was run on, specifically those with FAR $\leq$ 2 per day, network SNR $\geq$ 7, and two or more individual detector SNR $\geq$ 4.5. We used the the $p_{astro}$from the preferred event of each superevent for comparison. The points on the curve denote the threshold that minimizes the distance to the upper left corner of the plot for this subset of events.

events causing false positives. For example, S191220af had a glitch in L1 with a LOGBSN of 15.56, S191120aj had glitches in H1 and L1 with a LOGBSN of 27.58, and S191225aq had glitches in L1 and V1 with a LOGBSN of 110.47. To our model, these strong feature values mimic astrophysical events, even when there are clear data quality issues upon human inspection. Other events, such as S190829u with LOGBSN of 43.80 and S190822c with LOGBSN of 276.94, had no high strain glitches yet can still produce inaccurate data products due to detector background. Astrophysical events with low signal strength account for false negatives as associated values are similar to events due to noise, such as

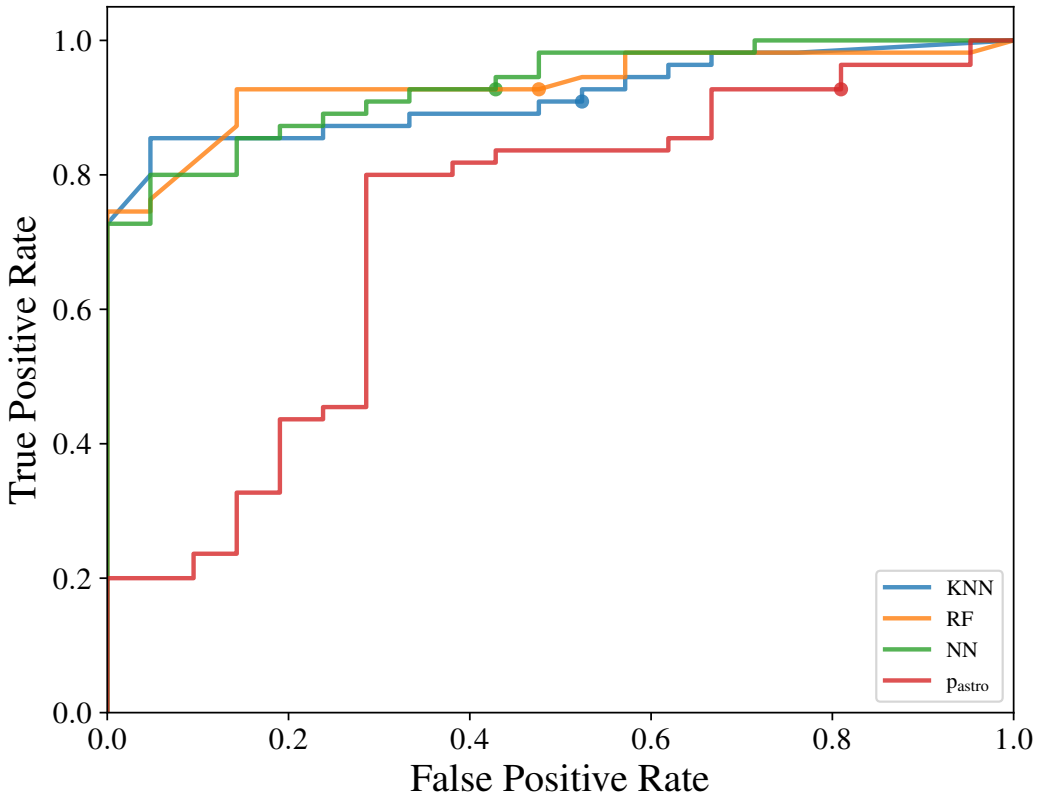

Figure 3.6: ROC curves obtained from the testing the on O3 dataset.

S191205ah, S190910h, and S190718y for the NN.

Fig. 3.8 shows FAR distributions of MDC and O3 events for each pipeline, as well as their true labels. We note that there are some O3 events that do not pass the significant FAR threshold, which are those corresponding to confirmed significant superevent in the GWTC catalog. These events are correctly labeled regardless of their high FAR, due to the multiple-pipeline nature of our model.

3.7 Conclusion

In this paper, we present the performance of our classifiers for assessing significance of low-latency GW candidate events trained data from the MDC [69]. We cover three distinct model architectures, KNN, RF, and NN, to assess event data products and classify those events as either astrophysical or terrestrial in origin. The NN

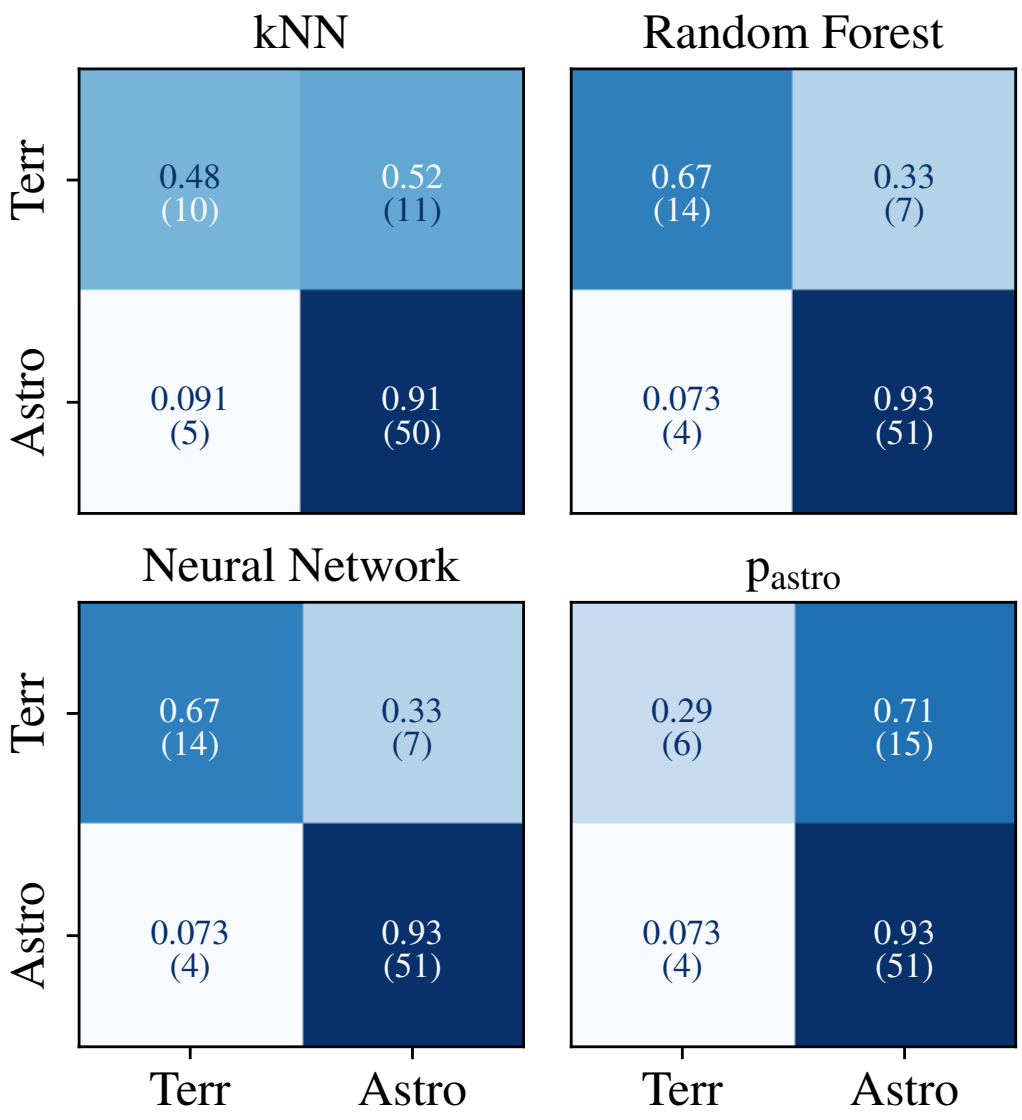

Figure 3.7: Confusion matrix of model performance on the O3 dataset.

and RF classifiers have matching performance with an AUC of 0.93 on O3 events slightly outperforming the KNN algorithm with a performance of 0.92. For MDC events, the RF and NN have an AUC of 0.96 and 0.95, while the KNN is 0.94. In both cases each of our classifiers significantly outperforms p_{astro}, and could be a useful complementary data product alongside it.

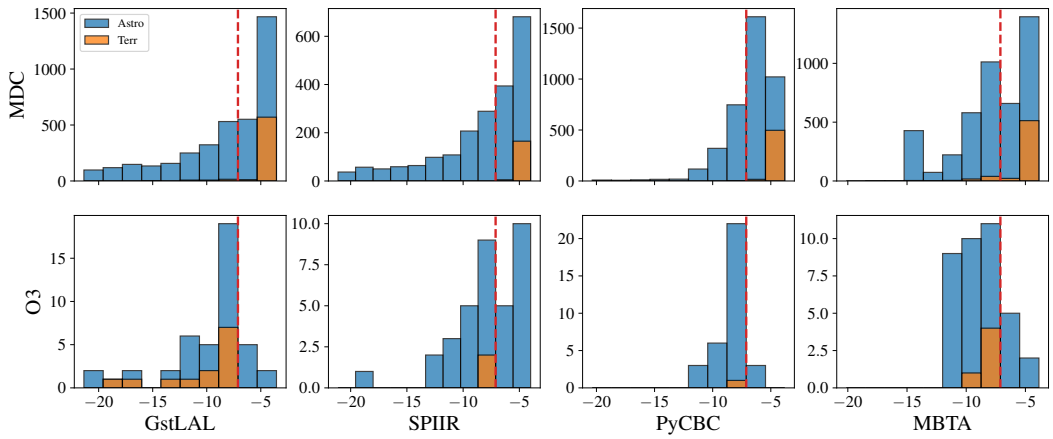

Figure 3.8: Distributions of MDC and O3 events against LOGFAR and their true labels. The red dotted line shows the five month significant threshold used on O3.

With the limited information available in low-latency, we can reliably determine whether an event is astrophysical or terrestrial, which can be extremely valuable for EM follow-up efforts. The most prominent limitation of the classifier is the disconnect between the data products and the detection data quality. The actual GW strain data is not included in the decision process for each event classification, so data quality issues such as glitches may inflate the significance of our features and lead to some false positives. For example, loud glitches caused false positives in a handful of O3 events: S191225aq, S191220af, S191120aj, and S190808ae. To improve on this aspect, including a metric for the model to consider a quantitative measure of the data quality of GW strain. Another improvement would be a larger and more comprehensive set of terrestrial events to balance the true classes or allow us to more accurately sample the data with a KDE. In particular, multiple-pipeline terrestrial events would be useful, but these are rare.

We should also note, that it is possible for a loud terrestrial glitch to be coincident with an astrophysical event, just as was the case with GW170817 [9]. Our classifier has not been explicitly trained or tested on this class of events, and we do not make assumptions about how our model will perform with such events, as quantities such as FAR and SNR will certainly be impacted. We can also assume extensive human intervention and vetting will be necessary in this case.

A similar score is produced by GWSkyNet, which is a pipeline using a series of convolutional neural networks trained on BAYESTAR localization and volume sky maps along with numerical inputs including the Bayes factors used here [59, 195]. In contrast, our pipeline uses internal information from *multiple* CBC search pipelines directly. This means a different parameter space is used compared to GWSkyNet, and the two can make complementary data products alongside p_{astro}.

The events in the MDC are created by injecting waveforms onto existing O3 data at a high rate. This causes an upwards bias of FAR values if the pipeline calculates FAR using an estimation of the local noise background [69]. This means that when using real GW events from O3, our performance may be negatively impacted. Luckily, a pre-O5 MDC using O4 data is planned and will hopefully resolve this issue. Retraining and tuning on this dataset will need to be done, as well as evaluating on O4 events after their public data release. This pre-O5 MDC may also be comprised of a different set of CBC search pipelines, which is something we can easily accommodate. Any additions can be implemented by adding classifiers in the first stage for additional CBC search pipelines and adding parameters for additional post CBC search pipelines. We envision this classifier could be used to determine significance of O5 candidate events alongside other data products such as $p_{astro}$and GWSkyNet. This addresses the issue where search pipelines may have differing assessments of significance for a given candidate event and provides a means of combining information from multiple searches into one cohesive statement of significance. We will advocate for the implementation and public release of our classifier's outputs in the future.

3.8 Model Architecture and Hyperparameters

Fig. 3.9 outlines the multiple-pipeline model architecture and corresponding features, as covered in Sec. 3.5.

K-Nearest Neighbors (KNN)

Stage 1 KNN: 6 neighbors, distance weight function, and Minkowski metric.

Stage 2 KNN: Same as stage 1 except with 693 neighbors.

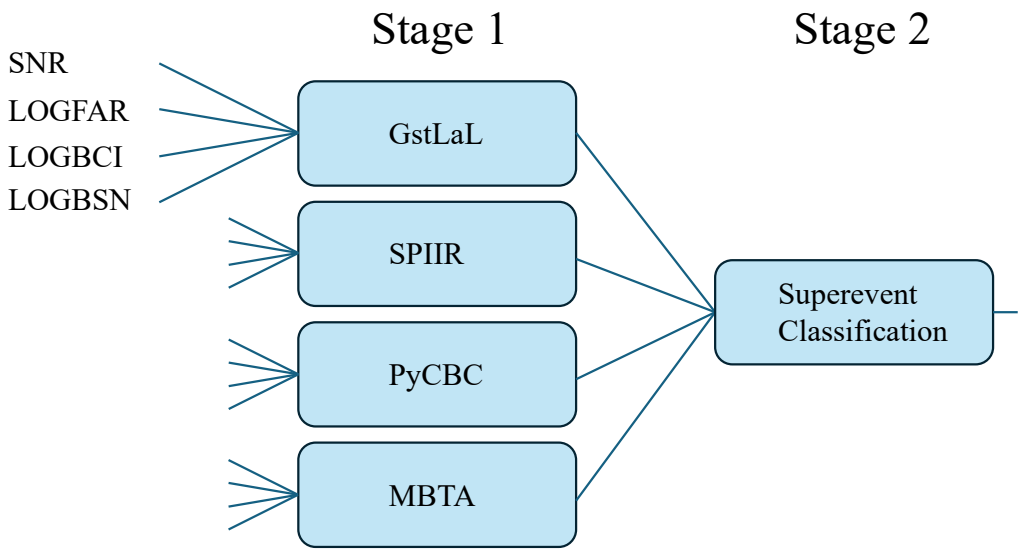

Figure 3.9: Diagram of model architecture, where each block is either RF, KNN, or NN. The NN, consists of 6 hidden layers with ReLU activation function. If a pipeline is not used in a superevent, a value of -1 is sent to superevent classification.

Random Forest (RF)

Stage 1 RF: 10 estimators, max depth of 4, Gini criterion, minimum samples to split of 2, minimum samples of each leaf of 1, and maximum features of 4.

Stage 2 RF: Same as stage 1 except with 310 estimators and max depth of 7.

Neural Network (NN)

Stage 1 NN: 10 hidden layers with 59 neurons each, L2 regularization penalty of 0.0035, learning rate of 0.0063, batch size of 122, ReLU activation function is used for all hidden layers, binary cross entropy loss. Training is done using a 90/10 train/validation split and training stops when validation loss plateaus.

Stage 2 NN: 2 hidden layers with 52 neurons each, L2 regularization penalty of 1.72e-06, learning rate of 0.008, batch size of 675, ReLU activation function is used for all hidden layers, binary cross entropy loss. Since train set is redefined at plateau due to downsampling, training stops after the third plateau.

3.9 Supporting Figures

Fig. 3.10 shows the individual confusion matrices for each of the pipelines trained on for both the MDC and O3. The individual single-pipeline models have lesser performance on their own, but when combined together as shown in Fig. 3.9 their performance improves significantly, as seen in Figs. 3.3 and 3.7.

Shown here in Figs. 3.11, 3.12, and 3.13 are the corner plots corresponding to the MBTA, PyCBC, and SPIIR MDC data and KDEs, complementary to the one shown in Fig. 3.1 for GstLAL.

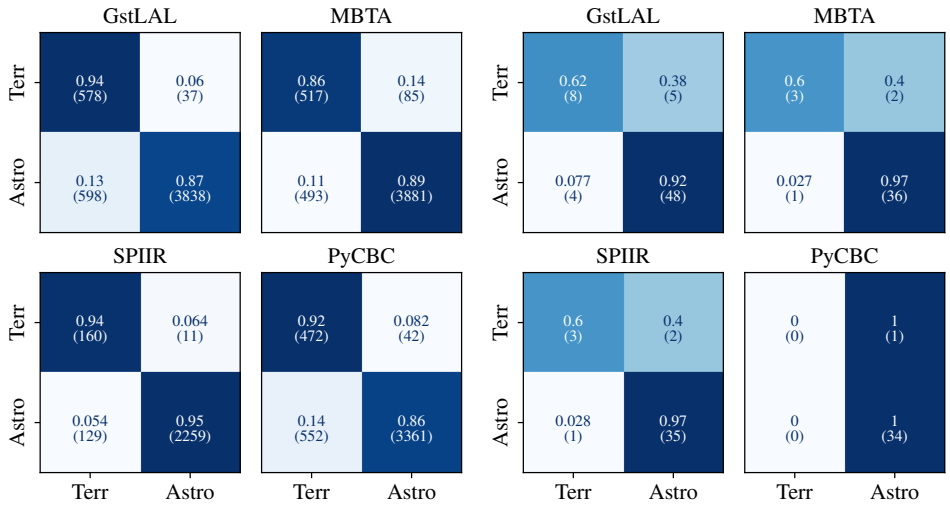

Figure 3.10: Confusion matrices for MDC (left) and O3 (right) on individual pipelines for NN.

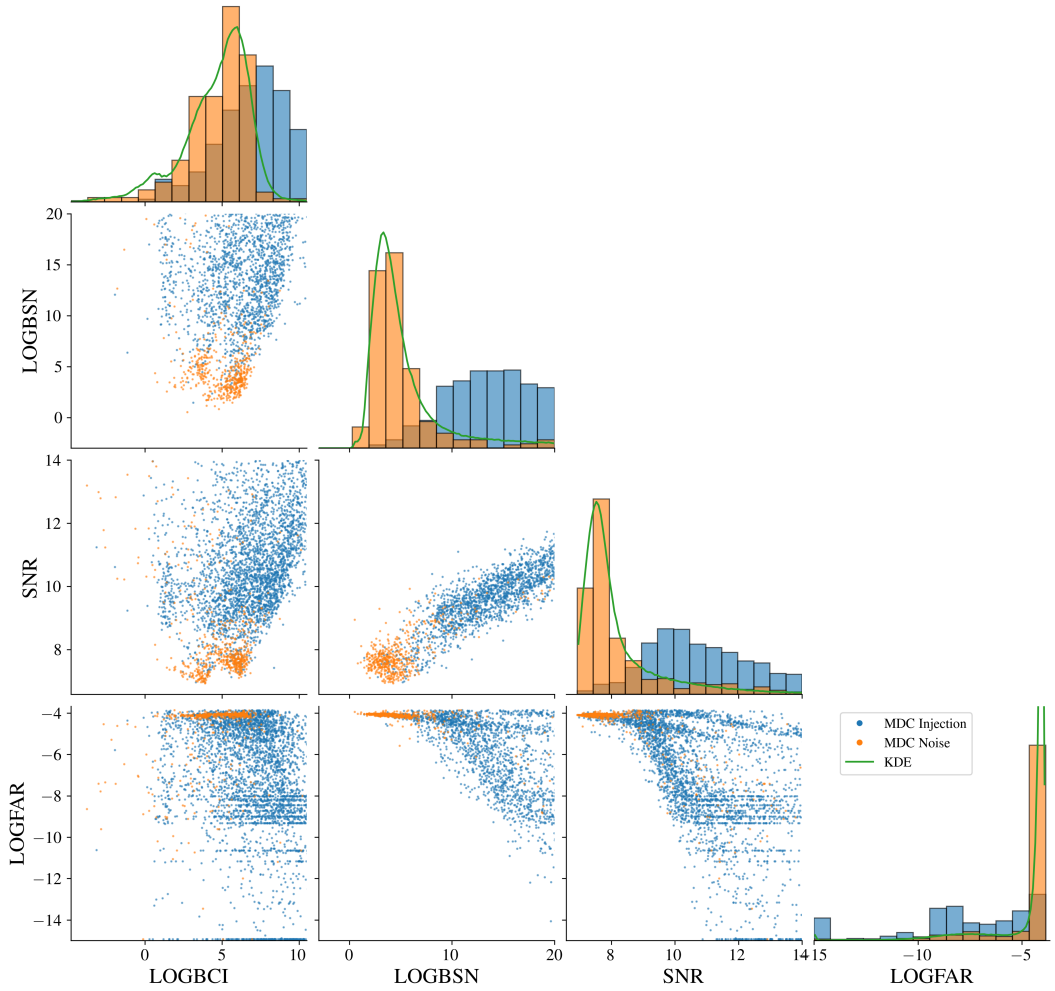

Figure 3.11: Corner plot of the MBTA and BAYESTAR pipeline data products from MDC events windowed around noise events. The histograms are normalized and the green line shows the PDF of the KDE.

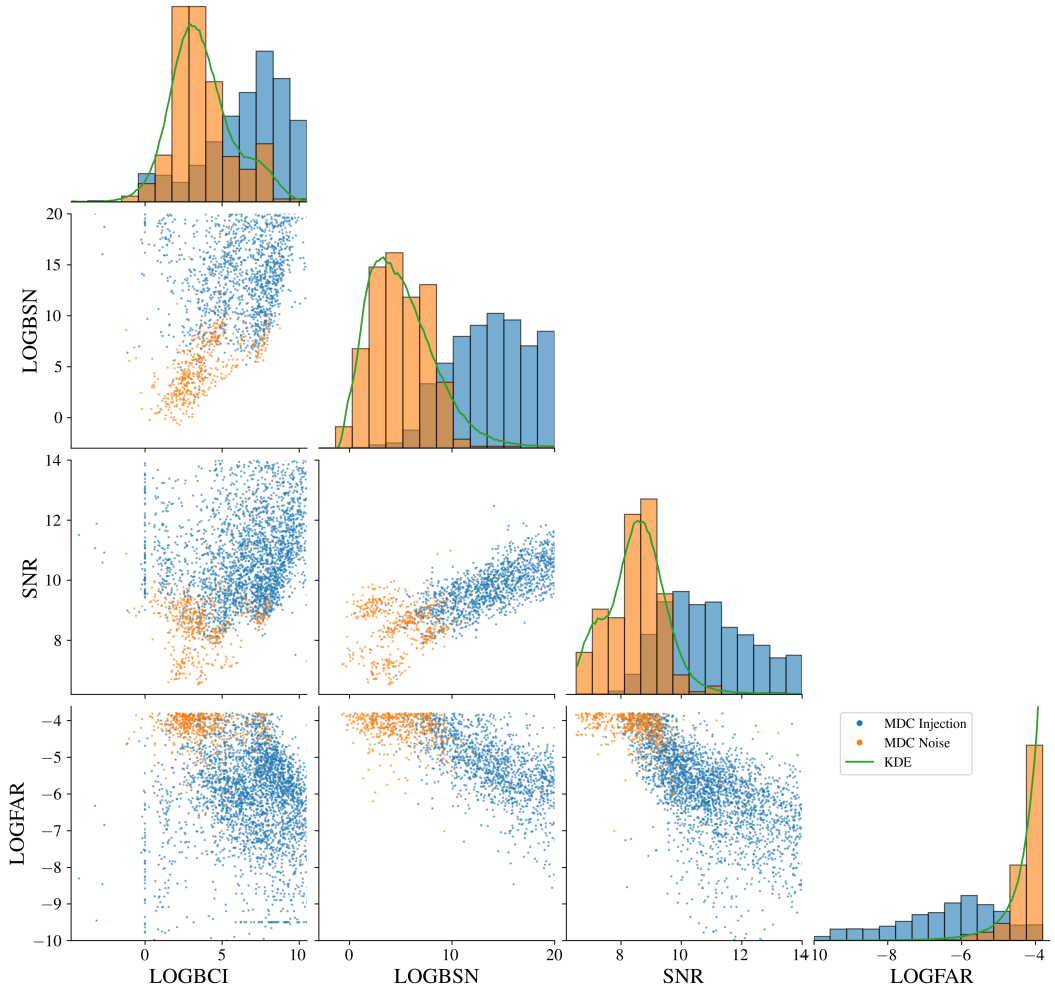

Figure 3.12: Corner plot of the PyCBC and BAYESTAR pipeline data products from MDC events windowed around noise events. The histograms are normalized and the green line shows the PDF of the KDE.

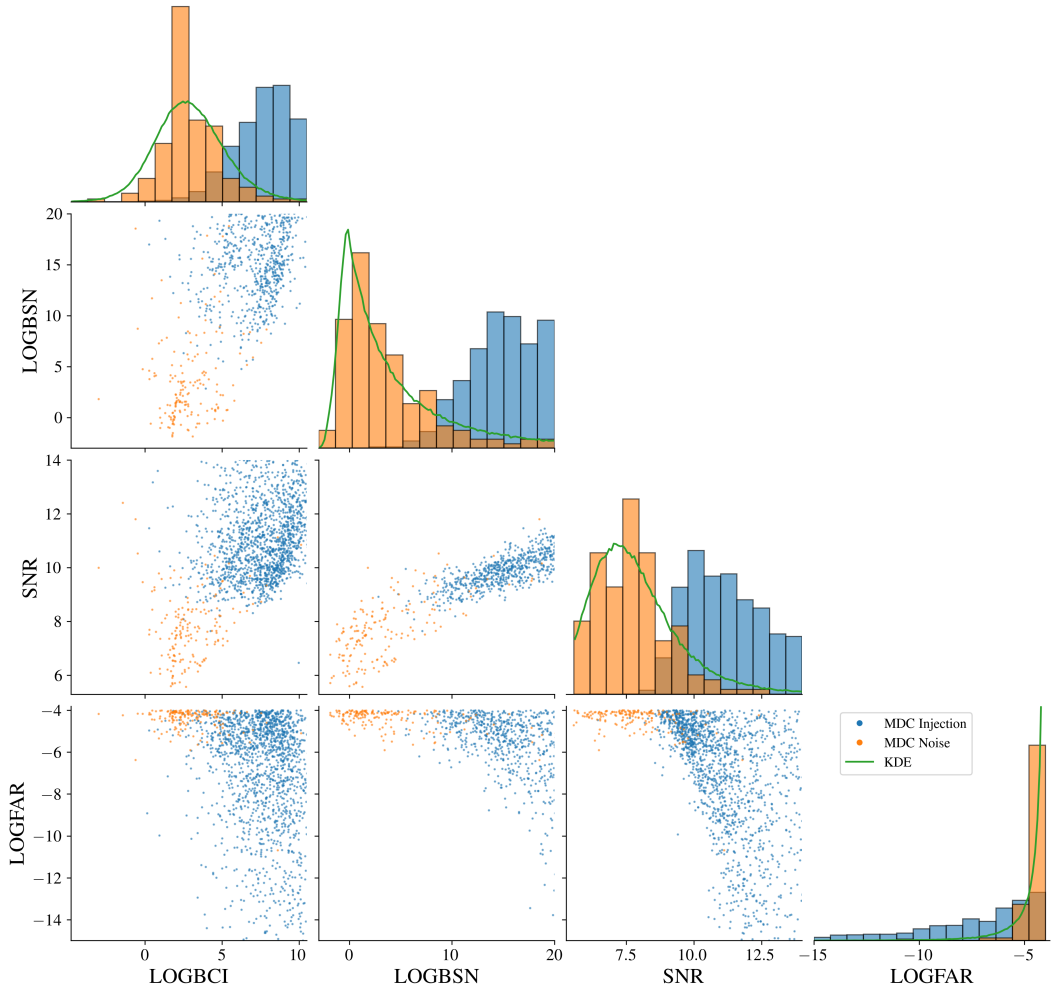

Figure 3.13: Corner plot of the SPIIR and `BAYESTAR` pipeline data products from MDC events windowed around noise events. The histograms are normalized and the green line shows the PDF of the KDE.

Chapter 4

WHAT TO EXPECT: KILONOVA LIGHT CURVE PREDICTIONS VIA EQUATION OF STATE MARGINALIZATION

Work reproduced from:

Andrew Toivonen et al. What to Expect: Kilonova Light Curve Predictions via Equation of State Marginalization. *Publ. Astron. Soc. Pac.*, 137(3):034506, 2025. doi: 10.1088/1538-3873/adbcd7.
A.T. was lead author and led software development of the data product.

4.1 Motivation

Efficient multi-messenger observations of gravitational waves from compact object mergers rely on data products reported in low-latency by the International Gravitational-wave Network (IGWN). While data products such as `HasNS`, the probability of at least one neutron star, and `HasRemnant`, the probability of remnant matter forming after merger, exist, these are not direct observables for a potential kilonova. Here, we present new kilonova light curve and ejecta mass data products derived from merger quantities measured in low latency, by marginalizing over our uncertainty in our understanding of the neutron star EoS and using measurements of the source properties of the merger, including masses and spins. Two additional types of data products are proposed. The first is the probability of a candidate event having mass ejecta ($m_{\rm ej}$) greater than $10^{-3} M_{\odot}$, which we denote as `HasEjecta`. The second are $m_{\rm ej}$ estimates and accompanying *ugrizy* and *HJK* kilonova light curves predictions produced from a surrogate model trained on a grid of kilonova light curves from `POSSIS`, a time-dependent, three-dimensional Monte Carlo radiative transfer code. We are developing these data products in the context of the IGWN low-latency alert infrastructure, and will be advocating for their use and release for future detections.

4.2 Searching for Kilonovae: A multi-disciplinary multi-messenger effort

The combined detection of the kilonova AT2017gfo [7, 86, 217] and gravitational-wave (GW) observations resulting from the binary neutron star merger GW170817 [9] has led to immense interest in the field of multi-messenger astrophysics. Kilonovae are short lived astrophysical transients that may result from either BNS or NSBH mergers and are of particular interest as they are sites of r-process nucleosynthesis, through which heavy elements can be produced. The radioactive decay and interactions of these r-process elements are what powers the kilonova emission we hope to observe [135, 151, 155, 171]. While GW170817 led to breakthroughs in nuclear astrophysics [135, 137, 166, 217, 234], cosmology [8, 83, 94], and tests of General Relativity [45, 88, 104], much remains to be learned about these rare events, including the diversity of their intrinsic parameters, their emission, and the heavy elements produced by r-process nucleosynthesis in these mergers.

O4 is underway as of May 23, 2023[1], and the search for gravitational wave events and their counterparts [5] has resumed. Searches for kilonovae are challenging due to the fact that they are short lived events, can be relatively faint, and may not be well localized. Sky localizations for gravitational wave events can span $\approx 100-10,000\,\mathrm{deg}^2$ [53, 77, 101, 105, 106, 123, 142, 202, 212, 214, 235]. However, in spite of these challenges, it is imperative to locate the transient as quickly as possible in the hope of observing the peak of emissions. A number of wide-field survey telescopes are used to try and cover these large sky localizations, such as: the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) [174], Asteroid Terrestrial-impact Last Alert System (ATLAS) [226], the Zwicky Transient Facility (ZTF) [29, 51, 120], and in the near future BlackGEM [122], the Vera Rubin Observatory, formerly known as the Large Synoptic Survey Telescope (LSST) [132], the Nancy Grace Roman Space Telescope[2] [35], and Ultraviolet Explorer (UVEX) [148]. While GW178017 is the only NS merger event that has provided us with joint observations of a GW signal, kilonova, and short gamma-ray burst (sGRB), it is also possible to identify kilonovae associated with sGRBs [39, 133, 224], long GRBs [154, 193, 227], or even serendipitously in survey operations [31, 33, 34].

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

[2] https://roman.gsfc.nasa.gov/

These searches are aided by source classification efforts, which we can use to determine the origin of GW events [52, 67]. In addition, there have been multiple efforts to simulate GW detections and constrain rates for O4 and beyond [76, 139, 186]. With O4 underway, we hope for additional BNS detections and follow-up opportunities. These observing scenarios simulations tell us what we can expect to observe, and can even help us constrain poorly measured parameters, like the inclination angle, as discussed in Sec. 4.3. The parameters of the binary before merger, such as the mass ratio and the masses of the objects involved, along with the EoS, can help predict the mass ejected from the merger and the light curves associated with a possible kilonova [9, 47, 48, 92, 189, 192]. The relationship between light curves and binary parameters can also be used to place constraints on the character of the progenitor systems and the mass ejected [57, 78, 85, 217].

Predictions of whether we can expect to see a kilonova, how bright that kilonova may be, and where it will be localized in the sky are all crucial pieces of information for astronomers searching for counterparts to GW events. Currently, the sky localization is provided through sky maps, produced by `BAYESTAR` [216] and `Bilby` [42], These efforts are intended to allow for more informed and efficient follow-up searches. There are also existing properties, `HasNs` and `HasRemnant` [67], which are provided by the `EM-Bright` pipeline and encode the probability that there will be at least one NS and non-zero $m_{\rm ej}$, respectively, by providing estimates of the amount of $m_{\rm ej}$ produced and the absolute magnitude of the kilonova.

In this paper, we propose two new data products, focused on observable properties directly tied to kilonovae, partially building on the work of [221]. We use similar ejecta fits and also define a quantity `HasEjecta` to describe the likelihood of having a significant amount of $m_{\rm ej}$. One noticeable difference is our focus on using parameter estimation, instead of template based estimates. The first data product is the probability of a candidate event having $m_{\rm ej}$ greater than $10^{-3}M_{\odot}$; we denote this quantity as `HasEjecta`. This quantity is useful for establishing the likelihood of a kilonova counterpart, regardless of brightness. The $10^{-3}M_{\odot}$ cutoff was chosen for three reasons, (i) our kilonova surrogate models were not trained on grid points below this value, (ii) ejecta fits can have large uncertainties that dominate at low

values, and (iii) this provides a reasonable floor estimate for the total $m_{\rm ej}$ capable of producing an observable kilonova. To directly support electromagnetic counterpart searches, we also propose to provide predicted *ugrizy* and *HJK* light curves. To do so, we use a surrogate model based on the time-dependent, three-dimensional Monte Carlo radiative transfer code `POSSIS` [56]. Using `POSSIS`, we produce kilonova light curves by simulating packets of photons released by the radioactive decay of r-process nuclei in a kilonova. `POSSIS` takes input in the form of a model that defines densities, compositions, and geometry of the ejecta, and outputs light curves and spectra as a function of inclination angle. `POSSIS` simulations are computationally expensive, so a grid of kilonova light curves are computed and then used to train a surrogate light curve model with The Nuclear Multimessenger Astronomy Framework (`NMMA`)[3] [184]. `NMMA` is a nuclear physics and cosmology library used for analysis of BNS and NSBH systems, as well as potential counterparts. To create the light curve model, `NMMA` starts with a grid of `POSSIS` simulations across input parameters, and uses a neural network to interpolate between that grid and generalize to arbitrary parameters. For this work, we use the light curve models referred to as `Bu2019lm` for BNS samples and `Bu2019nsbh` for NSBH samples [56], which depend on the dynamical mass ejecta ($m_{\rm ej}^{\rm dyn}$), wind ejecta ($m_{\rm ej}^{\rm wind}$), inclination angle (θ), and the opening angle (ϕ), as will be discussed more later on. `Bu2019lm` is trained on a grid of simulated BNS mergers, and `Bu2019nsbh` on a NSBH grid.

Sec. 4.3 provides the workflow for producing the proposed data products from what is available in low latency from IGWN searches. We present performance for the proposed data products in Sec. 4.4, while Sec. 4.5 provides our conclusions and plans for future research.

4.3 Modeling

Workflow overview

Fig.4.1 provides an overview of the workflow of our kilonova ejecta and light curve predictions. We start with a GW detection by one of the CBC searches, and then use initial mass and spin estimates from either the point estimate, or parameter estimation, as discussed in Sec. 4.3. The point estimate has high uncertainties, but

[3] `https://github.com/nuclear-multimessenger-astronomy/nmma`

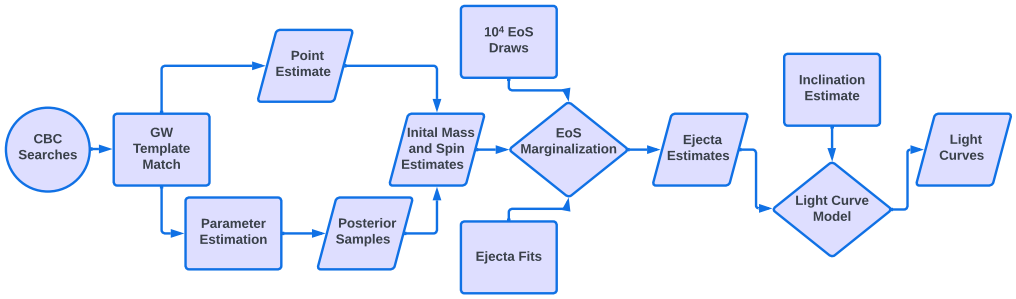

Figure 4.1: Workflow for the light curve and ejecta mass data products. We start with initial mass and spin estimates from a GW candidate as input, then marginalize over a number of EoS realizations per sample and apply mass ejecta fits. Finally, we use our light curve model to calculate light curves after drawing inclination angles if needed.

is also available in low-latency at the time the alert is sent out. Parameter estimation is more accurate, and covers a range of mass and spin values, but is not available until $\sim$ hours after merger. From there, we marginalize over EoS realizations and use the ejecta fits covered in Sec. 4.3 to calculate the $m_{\rm ej}$. Each sample is run with a set number of random EoS realizations to effectively cover the parameter space. Finally, adding in an inclination distribution and opening angle, covered in Sec. 4.3 and using our kilonova model covered in Sec. 4.3, we generate kilonova light curves. There are separate ejecta fits and light curve models used for BNS and NSBH samples, as is discussed later in Sec. 4.3 and 4.3. The maximum mass of a NS is defined by the EoS used.

Chirp mass and mass ratio

The chirp mass ($\mathcal{M}$) and mass ratio (q) of a compact object merger are important quantities for characterization of a merger event. $\mathcal{M}$ is the simplest merger quantity to measure from a GW signal, due to its relationship with the frequency evolution of the signal. $\mathcal{M}$ is defined as:

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \tag{4.1}$$

where m_1 is the primary, or larger, mass, while m_2, the secondary, or smaller, mass. The mass ratio, q, is much more difficult to measure accurately from GW data, and is the ratio between the component masses.

$$q = \frac{m_2}{m_1} \tag{4.2}$$

Estimates of q can be made using the phase of the GW signal, but they include significant error. $\mathcal{M}$ is recovered quite accurately by parameter estimation in most cases, however, there may still be a wide range of possible component mass values due to the poorly constrained q values, meaning the source classification cannot always be clearly ascertained.

We have two possible options for our initial mass and spin estimates: (i) the mass estimates produced by `Bilby` [42], an automated Bayesian parameter estimation analysis library, and (ii) the point estimate from the CBC template match, e.g. `GstLAL` [103, 170, 228], `MBTA` [36], `PyCBC` [89, 181], `SPIIR` [127, 156]. While parameter estimation is the starting point with the most precise measurements, given it provides (i) the most accurate estimate of the true source parameters and (ii) uncertainties, it takes $\sim$ hours to be completed post-merger. If a low-latency prediction is desired, as may benefit searches for counterparts in the hours after merger, we must use the point estimate as a starting point. There are also promising prospects of real-time machine learning based parameter estimation, which may be available in the future [90], but for now we must rely on the point estimate. In this case, we can use the point estimate, or take $\mathcal{M}$ from the point estimate and draw a q value consistent with that $\mathcal{M}$ value from population level distributions [84, 139] informed by GWTC-3 catalog fits [22]. From $\mathcal{M}$ and q, we compute the component masses that will be inputs for the simulated ejecta quantities and light curves (see below). For the analysis of MDC [69] events to follow, we use parameter estimation as the starting point for the reasons outlined above.

Equation of state and spin distributions

The NS EoS impacts the amount of mass expected to be ejected from a merger event as it influences how the NS is tidally disrupted. While the NS EoS remains unknown, there are certain popular EoS in the literature useful for comparisons. In

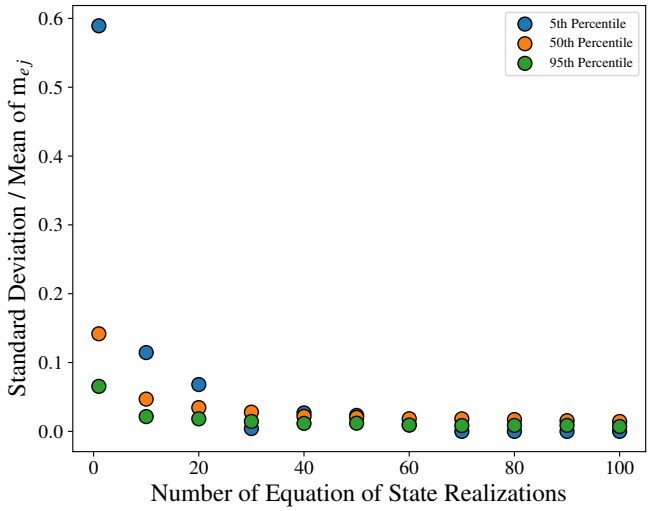

Figure 4.2: Comparison of relative error between 100 realizations of predictions based on parameter estimation for a single MDC event as a function of the number of EoS realizations per sample. We find that the error falls off asymptotically as the number of EoS realizations increasing, and that there are diminishing returns past 50 EoS realizations, where the error is less than a few percent.

the following, when we need a single EoS to compare to, we will use the SLy [65] EoS for comparisons to our marginalized range of predictions. This is to provide a singular point of reference for comparison, and to ensure that our predictions are reasonable. The SLy EoS was chosen as it has support from mass-radius posteriors from GW170817 [11].

Within our pipeline, due to the uncertainty in our understanding of the NS EoS, we marginalize over a number of EoS realizations per sample in order to produce our ejecta estimates. For each component mass pair, we draw a specified number of EoS realizations from a set of 10^4 equally weighted EoS realizations in order to cover the parameter space. The EoS realizations are drawn from a Gaussian-process posterior [152, 153] conditioned on a radio pulsar mass measurement [240] and gravitational-wave mass and tidal deformability measurements [9, 13]. For the purpose of our workflow, each sample is run with a different subset but same number of EoS

realizations. This number of realizations is a variable within the workflow and can be set as desired. To establish the number sufficient for consistent results, we ran a series of tests; Fig. 4.2 shows how the relative error, defined here as the standard deviation over the median $m_{\rm ej}$, varies for different numbers of EoS realizations. The standard deviation and mean was found across 100 realizations of the predictions for each number of EoS realizations. We find that 50 EoS realizations reduces the error to less than a few percent, giving us more than sufficient consistency between runs while still being computationally feasible. Beyond this point improvements are negligible. 50 EoS realizations also ensures that even in the case where only a small fraction of the posterior samples produce significant amounts of $m_{\rm ej}$, we will still adequately cover the parameter space. If running on a point estimate or similar small sample size, all 10^4 samples can be employed.

The NS and BH spins may also impact the amount of $m_{\rm ej}$. As will be covered in Sec. 4.3, the BNS ejecta fits used do not depend on spin, while the NSBH ones depend on the total effective spin $\chi_{\rm eff}$:

$$\chi_{\rm eff} = \frac{\chi_{1,\rm z} m_1 + \chi_{2,\rm z} m_2}{m_1 + m_2} \tag{4.3}$$

where $\chi_{1,\rm z}$ and $\chi_{2,\rm z}$ are the individual spin components of the dimensionless spin vectors χ_1 and χ_2 aligned with the orbital angular momentum ($\hat{\mathbf{L}}$) of the merger.

$$\chi_{\rm n,z} = \chi_n \cdot \hat{\mathbf{L}} \tag{4.4}$$

This means that our data products will be unchanged by varying spin values for BNS events but impacted for NSBH events. For simulated events from the MDC that we will compare to in Sec. 4.4, the spin distributions are injected uniform in magnitude and isotropic in orientation, with magnitudes up to 0.4 and 1 for NS and BH respectively [69]. Just as for candidate GW events, parameter estimation provides spin estimates alongside the component masses, which can then be used for ejecta predictions.

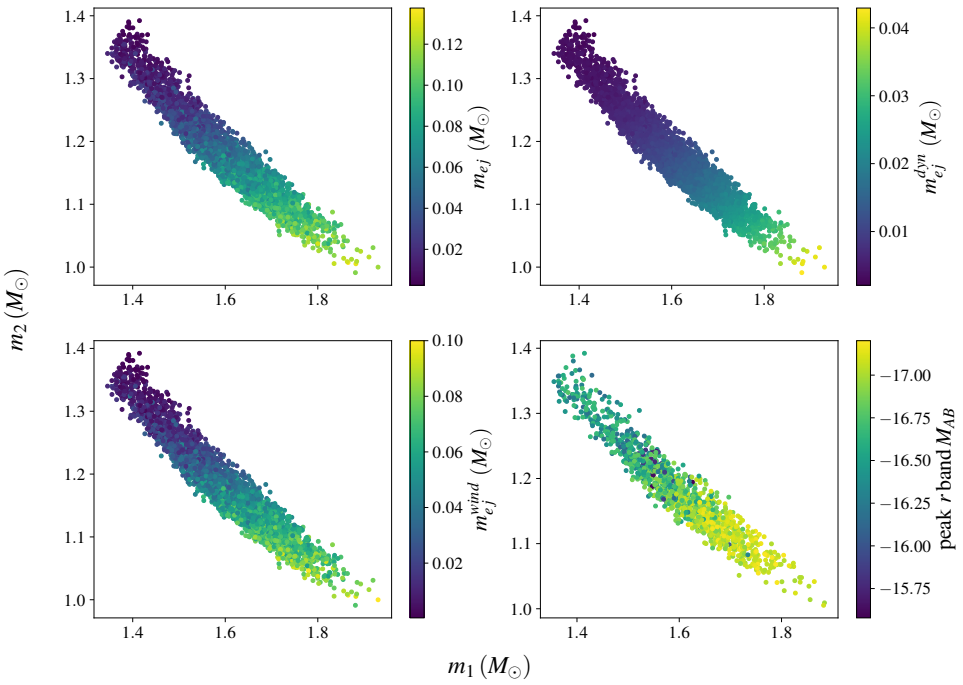

Figure 4.3: Scatter plots of $m_{\rm ej}$ and light curve predictions using EoS marginalization and parameter estimation from a $m_1 = 1.40M_\odot$, $m_2 = 1.34M_\odot$ injection. On the upper left we have $m_{\rm ej}$, upper right $m_{\rm ej}^{\rm dyn}$, lower left $m_{\rm ej}^{\rm wind}$, and on the lower right the peak r-band M_{AB}. We see a strong correlation between all four quantities with some slight variations.

Dynamical and disk wind ejecta from fits: BNS and NSBH

The total $m_{\rm ej}$ is calculated in two components: the dynamical ejecta, $m_{\rm ej}^{\rm dyn}$, and the disk wind ejecta, $m_{\rm ej}^{\rm wind}$. The first component, $m_{\rm ej}^{\rm dyn}$ is produced as the extreme tidal forces of the inspiral tidally deforms and rips mass from a neutron star [172]. The other, $m_{\rm ej}^{\rm wind}$, is produced by matter that is ejected from the accretion disk of the merger by energetic outflows of particles, sometimes referred to as particle "winds" [172]. Estimates of $m_{\rm ej}$ are found from fit by combining the source properties with the EoS used. The EoS provides tables of mass, radius, and tidal deformability information [153] which can be used to determine compactness needed for the fits.

When making predictions, it is possible to (i) marginalize over EoS in order to effectively cover the uncertainty in this parameter space, or (ii) pick a single named EoS as input. Regardless of the EoS, we use ejecta fits in order to to calculate both the dynamical and disk wind ejecta, which are found from separate fits for BNS mergers and NSBH mergers. These are found in Eqs. 4.5-4.12 below. We use the fits as implemented in `NMMA` [184].

The total $m_{\rm ej}$ is defined by [94] in Eq. 4.5, where $M_{\rm disk}$ is the mass of the disk and ζ is the fraction of mass that becomes unbound from the system. Therefore $\zeta M_{\rm disk}$ is the disk wind ejecta: $m_{\rm ej}^{\rm wind}$. We use $\zeta = 0.30$ for both BNS and NSBH mergers [73, 109, 110, 213]). The total $m_{\rm ej}$ is then simply the sum of the contributions from the $m_{\rm ej}^{\rm dyn}$ and $m_{\rm ej}^{\rm wind}$ components.

$$m_{\rm ej} = M_{\rm ej}^{\rm dyn} + \zeta M_{\rm disk} \tag{4.5}$$

For BNS mergers, we use the following fit in Eq. 4.6 from [147] for $m_{\rm ej}^{\rm dyn}$:

$$\begin{aligned} &\frac{M_{\rm ej}^{dyn}}{10^{-3} M_\odot} = \\ &\max\left(0, \left[\frac{a}{C_1} + b + \left(\frac{m_1}{m_2}\right)^n + cC_1\right] m_1 + [1 \leftrightarrow 2]\right) \end{aligned} \tag{4.6}$$

In these expressions, m_1 is the mass of the primary, or largest component mass, while m_2 is the secondary, or smaller mass. For BNS both are neutron stars, while for NSBH m_1 is always the black hole, while m_2 is the neutron star. In the same way, C_1 is the compactness of the primary mass, while C_2 is the compactness of the secondary. From the fit, the coefficients are: $a = -9.335$ $b = 114.17$, $c = -337.56$, and $n = 1.5465$. The term $[1 \leftrightarrow 2]$ simply refers to an addition of the same terms with indices swapped.

Then, for $m_{\rm ej}^{\rm wind}$, again for BNS mergers, we use the following in Eq. 4.7 from [94]:

$$\log_{10}\frac{M_{\rm disk}}{M_\odot} = \max\left(-3, a\left(1 + b\tanh\left(\frac{c - m_{\rm tot}/M_{\rm threshold}}{d}\right)\right)\right) \tag{4.7}$$

where a and b are given by

$$a = a_o + \delta a \cdot \xi\,, \qquad b = b_o + \delta b \cdot \xi\,, \tag{4.8}$$

and a_o, b_o, δa, δb, c, and d are all free parameters. The parameter ξ is given by

$$\xi = \frac{1}{2}\tanh\left(\beta\left(q - q_{\rm trans}\right)\right)\,, \tag{4.9}$$

where β and $q_{\rm trans}$ are free parameters and $q \equiv m_2/m_1 \leq 1$ is defined as the mass ratio. The best-fit model parameters are $a_o = -1.581$, $\delta a = -2.439$, $b_o = -0.538$, $\delta b = -0.406$, $c = 0.953$, $d = 0.0417$, $\beta = 3.910$, $q_{\rm trans} = 0.900$. The threshold mass $M_{\rm threshold}$ for a given EoS is estimated by the following [27]:

$$M_{\rm threshold} = \left(2.38 - 3.606\frac{M_{\rm TOV}}{R_{1.6}}\right) M_{\rm TOV}, \tag{4.10}$$

where $M_{\rm TOV}$ is the maximum mass of a non-spinning NS and and $R_{1.6}$ is the radius of a $1.6M_\odot$ NS.

Now moving to NSBH mergers, we use the $m_{\rm ej}^{\rm dyn}$ fit from [147], shown in Eq. 4.11.

$$M_{\rm dyn}(M_\odot) = m_2^{\rm bar}\left(a_1\left(\frac{m_1}{m_2}\right)^{n_1}\frac{1 - 2C_2}{C_2} - a_2\left(\frac{m_1}{m_2}\right)^{n_2}\frac{r_{\rm ISCO}}{m_1} + a_4\right) \tag{4.11}$$

The coefficients are: $a_1 = 0.007116$, $a_2 = 0.001436$, $a_4 = -0.02762$, $n_1 = 0.8636$, and $n_2 = 1.6840$.

Again for NSBH, the $m_{\rm ej}^{\rm wind}$ fit is defined by [114] here in Eq. 4.12.

$$M_{\rm disk}(M_\odot) = m_2^{\rm bar}\max\left(0, \alpha\frac{1-2C_2}{\eta^{1/3}} - \beta r_{\rm ISCO}\frac{C_2}{\eta} + \gamma\right)^{\delta} \tag{4.12}$$

In Eq. 4.12, $m_2^{\rm bar}$ refers to the baryonic mass of the secondary mass, η is the reduced mass and is defined as $\eta = (m_1 m_2)/(m_1 + m_2)$, and $r_{\rm ISCO}$ is the innermost stable circular orbit of the binary. From the fit, the coefficients are: $\alpha = 0.4064$, $\beta = 0.1388$, $\gamma = 0.2551$, and $\delta = 1.7612$.

These ejecta formulae are based on fits to numerical relativity simulations that can have high uncertainties in certain regions of the parameter space and do not take into account all the microphysics involved, such as neutrino heating and magnetic fields impacting the resulting remnant and ejecta [147]. Ejecta predictions outside of the 10^{-3}–$10^{-1}M_\odot$ range, as well as those for mergers with highly asymmetric mass ratios, are expected to have particularly large uncertainties [94, 147]. This is one of the reasons we define our quantity `HasEjecta` as the fraction of total ejecta values greater than $10^{-3}M_\odot$. We also use an upper limit of $10^{-1}M_\odot$ for each ejecta component, $m_{\rm ej}^{\rm wind}$ and $m_{\rm ej}^{\rm dyn}$, capping values above this at $10^{-1}M_\odot$. These two cutoffs ensure the ejecta values passed to our light curve models are reasonable and within the 10^{-3}–$10^{-1}M_\odot$ range, as those models are trained on values within that same range.

Predicted distributions of $m_{\rm ej}$, $m_{\rm ej}^{\rm wind}$, and $m_{\rm ej}^{\rm dyn}$ using these fits and marginalizing over EoS can be seen in Fig. 4.3 for an MDC event with injected masses of $m_1 = 1.40M_\odot$, $m_2 = 1.34M_\odot$. Further discussion found in Sec. 4.4.

Inclination

The mass ejected from a neutron star is not thought to be perfectly uniform and spherical [126, 219], which means the EM emission from a kilonova will not be isotropic. We expect to observe brighter emissions for face-on events, meaning $\theta \simeq 0°$. In the case of kilonovae, this inclination angle, θ, also known as the viewing angle, ranges from 0 to 90° as it is equivalent from 90 to 180° due to symmetry; in the GW case, face-on vs. face-off can be differentiated, and therefore measurements are reported from 0 to 180°. Observationally, we would expect to see an θ distribution that is uniformly distributed in the projection along the line of

sight, ($\cos\theta$). However, because GWs from face-on events have higher amplitudes, we are biased towards those that are face-on as they are easier to detect. Taking both these factors into account, we will get a distribution that peaks at lower θ values and tails off toward higher θ values, as seen in Fig. 4.4.

In the workflow, we have two means for drawing θ. The first is shown by the black line in Fig. 4.4. This shows the θ distribution resulting from observing scenarios simulations [84, 139] based on GWTC-3 [22] catalog fits. These are generated using `Bayestar` [216], by first simulating Gaussian noise and the accompanying matched filtering process in order to mimic realistic GW detections. Fig. 4.4 shows the KDE that estimates the θ distribution based on those simulations, and we can sample from this distribution. The second method uses standardized draws to mimic that same distribution, but instead of providing one random value, this method covers a range of θ values. For each sample, we cover the values 0, 15, 30, 45, 60, and 75 with probabilities proportional to what is seen in Fig. 4.4. This method removes the random nature of the draws which can be useful for small sample sizes. These two methods converge for large sample sizes and can be chosen based on the sample size and any time or resource constraints.

Kilonova light curve model

Our kilonova light curve model is based of the time-dependent, three-dimensional Monte Carlo radiative transfer code `POSSIS` [56], which can simulate kilonova light curves for a range of input parameters. A grid of these simulated kilonovae is then used to train a model in `NMMA` in order to generalize predictions to arbitrary parameters. We use separate BNS and NSBH model grids for training the BNS model `Bu2019lm` NSBH model `Bu2019nsbh`, respectively. Theses models take parameters of inclination angle (θ), half-opening angle (ϕ), dynamical ejecta $m_{\rm ej}^{\rm dyn}$, and wind ejecta $m_{\rm ej}^{\rm wind}$, and outputs light curves in nine *ugrizy* and *HJK* absolute magnitude (M_{AB}) bands. For this analysis, we have fixed this half-opening angle value to 30°, which was shown to be a reasonable fit to GW170817 [184].

Fig. 4.3 shows a peak r band M_{AB} light curve predictions from a single simulated event with $m_1 = 1.40M_\odot$ and $m_2 = 1.34M_\odot$, and the associated $m_{\rm ej}$ distributions.

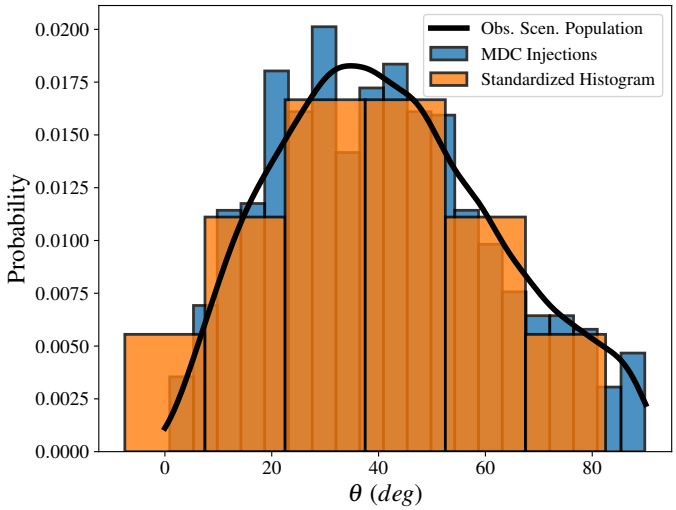

Figure 4.4: The two methods available for drawing θ: the observing scenarios population distribution [84] (black) and a standardized histogram (orange), compared to the histogram of θ from MDC injection samples with predicted $m_{\rm ej} \geq 10^{-3} M_{\odot}$ (blue). When using the observing scenarios distribution we draw a sample from the KDE for each initial component mass sample provided. When using the standardized histogram method, each initial component mass sample is run with $\theta = [0, 15, 30, 45, 60, 75]$ with counts proportional to the histogram, meaning each sample is run with the same θ values for consistency even with small sample sizes.

Fig. 4.5 shows all light curves across nine bands for that same event, as well as the median and 90% credible interval for $m_{\rm ej}$ and M_{AB}. The right panel of Fig. 4.7 shows peak r-band M_{AB} across a range of component masses consistent with BNS and NSBH mergers. The points show the median value across EoS realizations. In general, we find BNS events with a pair of low mass NSs produce the brightest kilonovae, and M_{AB} is strongly correlated with $m_{\rm ej}$. The light curve data products will be covered in detail and analyzed in Sec. 4.4.

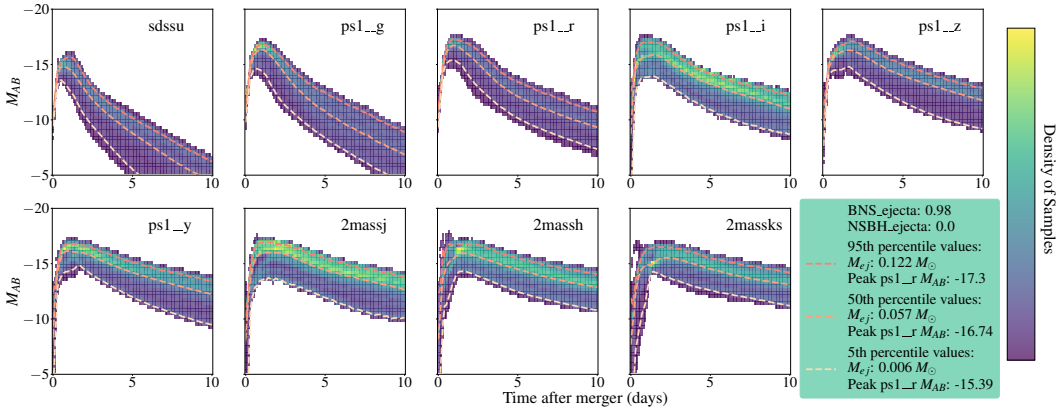

Figure 4.5: Heat map of light curves produced from EoS marginalization on parameter estimation of an MDC event injected with a primary mass of $1.40M_\odot$, and a secondary mass of $1.34M_\odot$. The 5th, 50th, and 95th percentile predictions are shown by the colored dotted lines and their corresponding $m_{\rm ej}$ values can be found in the legend.

4.4 Analysis of data products

Proposed data products

The first of our data products consists of three categories, which sum to a probability of 1: `BNS_ejecta`, `NSBH_ejecta`, and `no_ejecta`. `BNS_ejecta` refers to the probability that there will be significant $m_{\rm ej}$, defined as greater than $10^{-3}M_\odot$, produced by a BNS merger. Similarly, `NSBH_ejecta` refers to the probability that there will be significant $m_{\rm ej}$ produced by a NSBH merger; finally, `no_ejecta` refers to the probability that there will be $m_{\rm ej} < 10^{-3}M_\odot$. Each of these quantities assume the event is astrophysical in nature, and `no_ejecta` is indiscriminate of the type of merger. Due to EoS and parameter uncertainties, it is possible for predictions to contain non-zero values for both `BNS_ejecta` and `NSBH_ejecta`. In the usual case where the ejecta and light curve predictions are run on posterior samples from parameter estimation, each sample will be run with a number of different EoS realizations, and these realizations define the maximum NS mass. Not only can the range of parameter estimation posterior samples cover both BNS and NSBH regimes, a sample component mass near the NS max mass can be classified as a NS by one EoS realization, and as a BH by another. The sum of `BNS_ejecta` and `NSBH_ejecta`

will then provide the probability that a given event has significant ejecta, i.e. greater than $10^{-3}M_{\odot}$. We denote this quantity `HasEjecta`, and expect this to be useful metric for astronomers. These quantities specify likelihood of significant $m_{\rm ej}$, and are directly correlated to the likelihood a detectable kilonova exists.

In addition to the above data products, we provide $m_{\rm ej}$ and peak magnitude estimates in the form of three percentile values that constitute a median value and 90% credible interval: the 5th, 50th, and 95th percentiles. We also create probability density maps of the light curves produced, showing the range of possible outcomes in a informative way, along with the percentile estimates as seen in Fig. 4.5.

These complete $m_{\rm ej}$ and light curve predictions can be seen for an example MDC event in Fig. 4.5. Shown are probability density maps for light curves found by starting with parameter estimation and marginalizing over EoS. This event is entirely dominated by samples consistent with a BNS: Fig. 4.5 shows an event with q near one ($m_1 = 1.40M_{\odot}$, $m_2 = 1.34M_{\odot}$). We also show the median and 90% credible interval predictions for both $m_{\rm ej}$ and the peak r band magnitude. These light curve predictions across *ugrizy* and *HJK* bands are intended for astronomers to help with EM follow-up by providing an estimate of whether we can expect to see a kilonova, and how bright the kilonova may be.

Predictions compared to GW170817 + AT2017gfo

As a sanity check for our light curve predictions, we compare to AT2017gfo [7, 86, 217] the kilonova resulting from the BNS GW170817 [9]. We expect the observed light curves to fall within the 90% credible interval of our prediction. To make predictions, we start with parameter estimation from GW170817, and run the predictions workflow including EoS marginalization, producing a range of light curves, as seen in Fig. 4.6. We focus on the r band M_{AB} for the sake of this plot, and we find that the observed data does in fact fall within our 90% credible interval of predictions as expected.

Predictions across a grid of component masses

In order to provide an idea of what regimes in the parameter space are most likely to produce an observable kilonova, we make predictions across a wide grid of

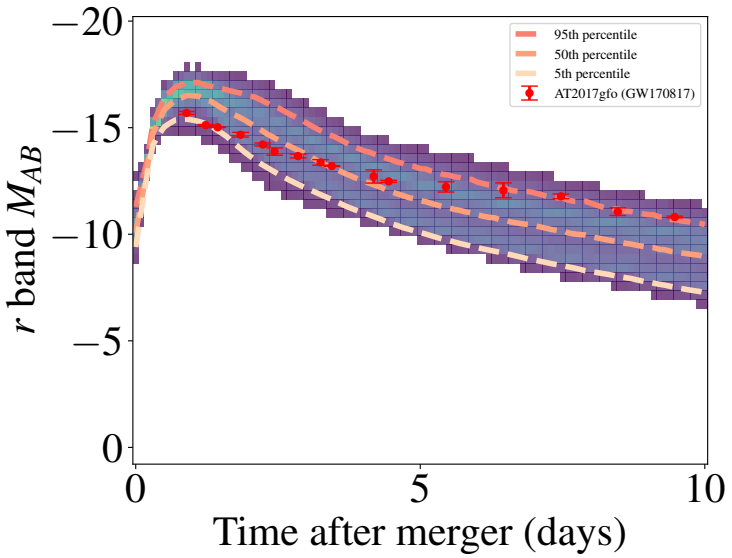

Figure 4.6: Comparison of light curve predictions using GW170817 posterior samples to AT2017gfo light curves. The observed r band light curves fall within the 90% credible interval of the predicted light curves.

component masses spanning BNS and NSBH events. For the consistency of Fig. 4.7 we assume zero spin for all objects. The left panel of Fig. 4.7 shows the median $m_{\rm ej}$ predictions across EoS realization for a grid of BNS and NSBH mergers. The dotted line between the BNS regime and NSBH regime denotes where the two source classifications dominate. Each sample is classified as a BNS or NSBH based on the maximum NS mass of the EoS realization, and uses the corresponding BNS or NSBH ejecta fits and light curve model. The difference in fits and models between the BNS and NSBH regimes can clearly be seen by the discontinuity near dotted line in Fig 4.7. We find the the low mass BNS events have the largest $m_{\rm ej}$, and that NSBH events tend to have less $m_{\rm ej}$. The same trends follow for peak r band M_{AB}, with a sharp drop off as we move from the BNS to the NSBH light curve model, again pointing to the fact BNS mergers are most likely to produce an observable kilonovae. It is worth mentioning, however, that larger spin values in the NSBH case are expected to produce more ejecta and brighter light curves than their non-spinning equivalents [113].

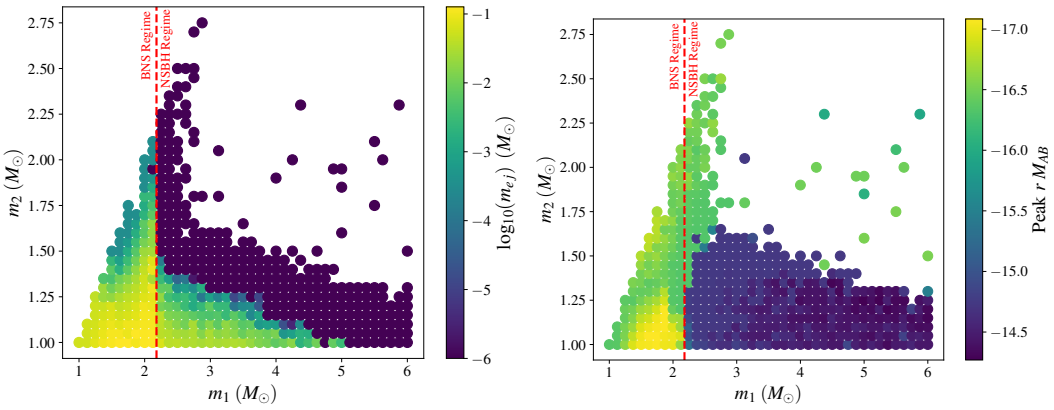

Figure 4.7: *Left:* A scatter colored by the median $m_{\rm ej}$ produced for a given component mass pair, marginalized over a number of EoS realizations. This plot only shows points that have some samples greater than $10^{-3}M_\odot$ and the low end of the color bar is capped at $10^{-6}M_\odot$. The BNS and NSBH regimes are indicated by a dotted line, loosely defined by a maximum NS mass of $\simeq 2.15\ M_\odot$. We find the low mass BNS regime produces the largest $m_{\rm ej}$, and see a significant drop off for the NSBH regime. *Right:* A the scatter colored by the median r band M_{AB} for a given component mass pair. The distribution is similar to the left panel due to the correlation of $m_{\rm ej}$ with the brightness of the kilonova, and with fainter light curves in the NSBH regime due to the different light curve model. For the consistency of the plot, all spins are set to zero and θ is drawn from the histogram method in Sec. 4.3.

Predictions from Mock Data Challenge events

In order to validate our proposed data products, we use MDC [69, 169] events. The MDC is a real-time simulation campaign where CBC waveforms are injected into O3 strain data. CBC searches are carried out and event candidates are uploaded internally to `GraceDB`[4], including downstream data products such as parameter estimation and sky localizations. This provides us with a set of parameter estimation posterior samples for merger events, and also allows us to refer back to the original simulated quantities for comparison.

We produced our data products for a set of MDC events, including both BNS and NSBH, for which we predict the likelihood of generating $m_{\rm ej}$, as well as the

[4] `https://gracedb.ligo.org/`

brightness of the resulting light curves. As described in the workflow above, we use the masses and spins generated by the parameter estimation, while marginalizing over EoS, to make the predictions. In order to compare our predictions to the injections and to sanity check our results, we run a set of MDC events with our EoS marginalized predictions, and compare those to predictions made using the injections and the SLy EoS.

Fig. 4.3 shows a detailed view of how $m_{\rm ej}^{\rm dyn}$, $m_{\rm ej}^{\rm wind}$, and $m_{\rm ej}$ vary across parameter estimation posterior samples from a single simulated event. We find all four quantities are highly correlated, with the highest $m_{\rm ej}$ and brightest r band M_{AB} values corresponding to small secondary NS mass in this case.

Further, Fig. 4.8 shows a violin plot of how the EoS marginalized predictions for parameter of estimation of MDC events of increasing total mass compare to one another and their injected source properties run with the SLy EoS. These events are meant to roughly cover a range component mass pairs that may be capable of producing a kilonova. As with all predictions in this paper, the distributions shown here are only for samples with $m_{\rm ej} \geq 10^{-3} M_{\odot}$. The curves surrounding the colored regions are KDE approximations of the samples, with a box and whisker plot enclosed and the median represented by the white dot. We find that the EoS marginalized distributions and the 90% credible intervals are generally consistent with the injections run with SLy.

Moving from left to right with increasing total mass, we find the median peak r band M_{AB} values follow a general decreasing trend, consistent with what is seen in Fig. 4.7. The red dotted line shows an approximate break between the BNS and NSBH regimes, where separate ejecta fits and light curve models dominate. This is the reason we see the bimodality of the violin plots, with the BNS samples mostly concentrated on top, and the NSBH samples on bottom. This is also evident for the SLy injected values, as there is a significant drop off between the first six events (classified as BNS by SLy), and the last two (classified by NSBH by SLy), as we cross the SLy maximum NS mass of $\simeq 2.1 M_{\odot}$. A single event can have samples in both the BNS and NSBH regime as parameter estimation covers a range of mass pairs, mostly consistent with a well constrained $\mathcal{M}$ but with varying q, and that

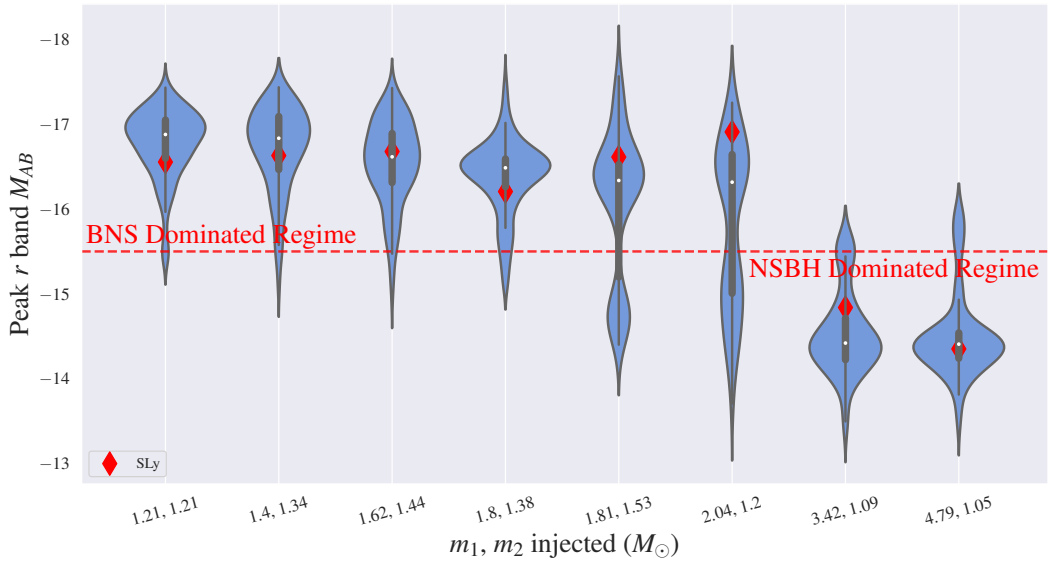

Figure 4.8: Violin plot showing peak r band M_{AB} predictions for 8 MDC events of increasing total mass. The range of predictions found from parameter estimation samples are compared to the injected parameters run with the SLy EoS. The curves show a KDE of the samples while inside the colored region there is a box and whisker plot. A dotted line is used to indicate the region dominated by BNS and NSBH samples, respectively.

our range of EoSs that we marginalize over each define their own maximum NS mass. In this way, we show our range of predictions including these uncertainties, and across source classifications. This plot shows BNS are most likely to produce bright, observable kilonovae based on our current understanding of the underlying factors.

Finally, we note that these representations may be an additional way of viewing the overall distribution, and could be produced for GW candidates.

4.5 Conclusion

In this paper, we propose new kilonova light curve and ejecta mass data products to help inform electromagnetic follow-up of GW candidates from CBC searches. To make predictions, we marginalize over EoS uncertainty, and use ejecta fits and light curve models to produce our estimates. Our data products include the

probabilities `HasEjecta`, which describes the probability of an candidate event having a $m_{\rm ej} \geq 10^{-3} M_{\odot}$, and `BNS_ejecta` and `NSBH_ejecta`, which describe the probability of BNS and NSBH mergers having $m_{\rm ej} \geq 10^{-3} M_{\odot}$. We also produce 90% percent credible interval estimates for $m_{\rm ej}$, and *ugrizy* and *HJK* M_{AB} bands.

We find that our predictions are not only consistent with GW170817 and AT2017gfo in Fig. 4.6, but also with MDC injections run with the SLy EoS in Fig. 4.8. Additionally, Fig. 4.8 demonstrates BNS are most likely to produce a bright, observable kilonova, and points to an inverse correlation between median peak *r* band M_{AB} total mass.

With the intention of informing follow-up decisions and helping enable the detection of future kilonovae, we will advocate for the public release of the aforementioned ejecta and light curve data products during the second half of O4. Our data products will not only help determine the likelihood that a GW event could produce a kilonova, but also estimate the amount of $m_{\rm ej}$ and the light curves produced, the direct observables astronomers need when making follow-up decisions. In this way, the data products are meant to be complementary to, and an extension of, existing data products such as `HasNS` and `HasRemnant` that are already widely used for follow-up decisions. Upon additional multi-messenger observations, we hope to improve our estimates by reducing uncertainties in our models and EoS set, further increasing the efficacy of our predictions.

Chapter 5

THE STANDARD SIREN APPROACH USING AN END-TO-END SIMULATION OF LOW-LATENCY GRAVITATIONAL-WAVE SEARCHES IN THE MOCK DATA CHALLENGE

Work reproduced from:

Megan Averill, Andrew Toivonen, et al. The Standard Siren approach using an end-to-end simulation of low-latency Gravitational-Wave Searches in the Mock Data Challenge. 2025.
A.T. advised M.A. as an undergraduate researcher for this work, it will soon be on arXiv and submitted to PASP.

5.1 Motivation

Multi-messenger observations of NS mergers and their associated kilonovae can provide standard siren measurements of H_0. While there are large uncertainties associated with a single standard siren measurement, combining H_0 posteriors from multiple events can provide a more precise measurement. Here, we use MDC events consisting of significant GW triggers provided by low-latency search pipelines during a real-time replay of a GW observing run to simulate a realistic set of BNS mergers and accompanying multi-messenger observations, then apply a framework for measuring H_0 using the entire set of events. With the set of 28 events combined with GW170817, and when considering a 1 σ ($\sim 68\%$) credible interval, H_0 can be measured down to a precision of $\sim 5\%$ and consistent with our injected value. We intend for this framework to be used in combination with future multi-messenger observations of NS mergers.

5.2 Standard Sirens: A Measure of the Hubble Constant

The joint detection of GWs from the BNS merger GW170817 [9] and the associated EM counterpart, the kilonova AT2017gfo [7, 86, 217] and GRB170817A [10, 118, 208], has led to great interest in multi-messenger searches due to the

far-reaching scientific insights the observations provided. Multi-messenger observations of kilonovae, the short-lived astrophysical transients that may accompany BNS or NSBH mergers, are highly desired for their insights across nuclear astrophysics and beyond. GW170817 has led to advances in nuclear astrophysics [135, 137, 166, 217, 234], cosmology [8, 58, 83, 94, 128], and tests of general relativity [45, 88, 104]. Specifically, the presence of heavy elements that power the kilonova emission revealed that r-process nucleosynthesis is taking place in these mergers [135, 151, 155, 171].

O4 began on May 23, 2023[1], and is currently in progress as the search for GW events and their associated counterparts [5] continues. Kilonovae are short lived, may be relatively faint, and can span large sky localizations on the order of $\approx 100 - 10,000$ deg^2 [53, 77, 101, 105, 106, 123, 142, 202, 212, 214, 235]; locating and identifying a transient as soon as possible is ideal in the hope of observing the peak of emissions.

Currently, there are a number of wide-field survey telescopes in use for multi-messenger efforts, such as the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) [174], Asteroid Terrestrial-impact Last Alert System (ATLAS) [226], the Zwicky Transient Facility (ZTF) [29, 51, 120]. In the near future, BlackGEM [54], the Vera Rubin Observatory, formerly known as the Large Synoptic Survey Telescope (LSST) [132], the Nancy Grace Roman Space Telescope [35], ULTRASAT [211], and Ultraviolet Explorer (UVEX) [148] will be joining the search. It is also possible to identify kilonovae associated with sGRBs [39, 133, 224], long GRBs [154, 193, 227], or even serendipitously in survey operations [31, 33, 34].

Source classification efforts [52, 67], which are used to determine the origin of GW events, are useful for informing these searches, and in the future additional data products specifically focused on remnants from neutron star mergers and their associated kilonovae may be available [49, 191, 225, 229]. Additional machine learning-based searches [168] and rapid parameter estimation techniques capable of producing samples within seconds are also in development [68, 90].

[1] `https://observing.docs.ligo.org/plan`

Multi-messenger observations of neutron star mergers can be used to provide an independent measurement of H_0 by combining distance information derived from the GW signal with redshift information from the host galaxy, which can be determined with detection of the EM counterpart [209]. With a well localized GW event, however, it can be possible to determine the host galaxy, or marginalize over multiple host galaxies, and estimate H_0 from the GW detection alone [183]. These GW events are often referred to as a "Standard Sirens" when an associated EM counterpart is detected, or "Dark Sirens" when not. In addition, there has been work to enable simultaneous inference of the NS EoS and H_0 for future merger events [117, 163].

Using the standard siren method with observations from GW170817, the measurement $H_0 = 70.0^{+12.0}_{-8.0} \mathrm{km\,s^{-1}\,Mpc^{-1}}$ was made [8]. Due to its large uncertainties, this value is consistent both with measures from The Planck collaboration, which finds $H_0 = 67.4 \pm 0.5 \ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ by use of cosmic microwave background (CMB) observations of the early universe and ΛCDM cosmology [28], and with the SH0ES collaboration which finds $H_0 = 74.03 \pm 1.42 \ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ by measuring the local expansion of the universe from observations of Cepheid variables and Type Ia supernovae [198]. This discrepancy between the precise measurements of Planck and SH0ES based on early and late time observations of the universe is often referred to as the "Hubble Tension". Currently, standard siren measurements are subject to a myriad of uncertainties, including the precision of our distance measurements, as well as spectroscopic or photometric redshift determination. In addition to these, more subtle effects such as selection biases and waveform systematics [149] can also play a minor role when we have multiple future detections.

In the future, we hope additional next-generation GW detectors such as Laser Interferometer Space Antenna (LISA) [32], Cosmic Explorer [102, 196], and Einstein Telescope [3] will help provide numerous multi-messenger observations of neutron star mergers that can be used to provide a much more precise measure of H_0 and help resolve this tension.

5.3 Method: A Simplified Example

In this paper, we start with a simplified example to demonstrate our framework in Sec. 5.3 using events we with `NMMA` [94, 184], then follow with a realistic end-to-end simulation of a set of multi-messenger events. The end-to-end simulation uses MDC [69] events which are realistic real-time simulations part of a simulation campaign used to stress-test the LLAI and evaluate performance of GW searches and their related data products. We will use BNS events exclusively in this study, however, a similar study could be done for NSBH events which are also capable of provided precise standard siren measurements [108]

Simulation Setup

We begin with a simplified example to demonstrate our framework, before proceeding to a realistic end-to-end simulation including selection effects. We use a set of BNS events simulated using `NMMA` [94, 184]. The simulation set was generated from the O4 observing scenarios outlined in [139] and consistent with GWTC catalog events [22]. We simulate a set of 28 binary neutron star events and accompanying GW and EM observations, uniform in source-frame volume and with SNR values greater than 10.

The parameter estimation is then done with Jim [97, 238, 239], a toolkit to perform Bayesian parameter estimation of GWs with Markov chain Monte Carlo and using normalizing flows [145, 185] as proposals to enhance sampling efficiency. The methodology was first described in [115], and we use the implementation from [237].

Estimating H0

The value of H_0 can be obtained from the redshift and distance measurements. For nearby, low-redshift events, the local approximation for H_0 can be used:

$$H_0 = \frac{cz}{D}\,. \tag{5.1}$$

However, to account for cosmological factors, we will also use the more accurate empirical fit [82, 197]:

$$D = \frac{cz}{H_0}\left(1 + \frac{1}{2}(1 - q_0)z - \frac{1}{6}(1 - q_0 - 3q_0^2 + j_0)z^2 + O(z^3)\right), \tag{5.2}$$

where q_0 is the present acceleration and j_0 is the prior deceleration. Both have values obtained from fit to cosmological models covered in [82, 197].

As Eq. 5.2 shows, the Hubble constant requires information on the distance and redshift of the source. In the case of a real multi-messenger detection, such as GW170817 [8], a distance posterior can be produced from parameter estimation based on the observed waveform. The redshift can be obtained by locating the EM counterpart and determining its host galaxy, for which galaxy catalogs can provide the redshift. We must also account for the initial prior on how the injections are distributed in space, shown in Eq. 5.6, which is uniform in comoving volume and source frame time:

$$p(z) \propto \frac{1}{1+z}\frac{dV_c}{dz}. \tag{5.3}$$

Redshift Measurement

BNS mergers may provide us with an EM counterpart in the form of a kilonova or GRB, which we can use to measure the redshift of the event by determination of the host galaxy. When the GW and EM observations are combined to estimate H_0, these events are referred to as standard sirens. Without the additional EM counterpart observations, it is much more difficult to reliably place the event in a host galaxy. These measurements may marginalize over the uncertainties of being located in multiple galaxies. Therefore, these dark sirens result in a larger measure of uncertainty when compared to standard sirens.

For the purpose of this study, we simulate standard siren observations of BNS events with an identified optical counterpart. When simulating the EM detection, we use the injected distance and assume the Planck18 cosmology [28] to determine the corresponding redshift. We use the injected redshift as the "true" redshift and draw Gaussian uncertainties for each measurement, within ±5% for events under 200 Mpc and ±10% for events over 200 Mpc. We expect the majority of redshifts outside 200

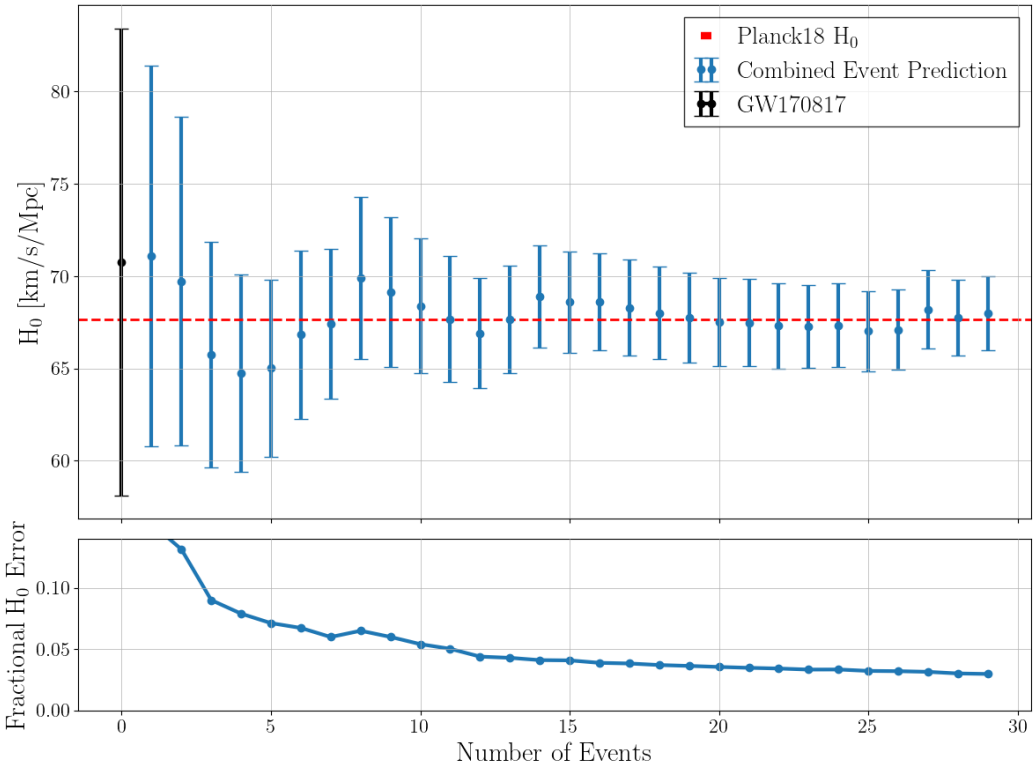

Figure 5.1: *Top:* H_0 posteriors for a sequential combination of our simulated events covered in Sec. 5.3. Shown are the mean values with 1σ ($\sim 68\%$ credible interval) error bars. *Bottom:* The fractional error, found as the ratio of the 1σ error to the mean H_0 value, for each sequential combination of the same events.

Mpc to be photometric, and therefore include the larger uncertainties ranging up to 10%.

Combining H_0 posteriors

Taking the distance and redshift information, we find H_0 posteriors for each simulated event. We then sequentially combine these event-by-event into a posterior for the entire set of events to precisely estimate H_0 as follows. Given a dataset X_i, the single-event posterior of H_0 is given by Bayes' theorem:

$$p(H_0|X_i) = \frac{p(X_i|H_0)p(H_0)}{p(X_i)}\,, \tag{5.4}$$

where $p(X_i|H_0)$ is our data, specifically posterior samples on distance. To combine the posterior distributions obtained from two such datasets, X_1 and X_2, with the same prior and that are conditionally independent, i.e.,

$$p(X_1, X_2|H_0) = p(X_1|H_0)p(X_2|H_0)\,, \tag{5.5}$$

we can write the posterior for two events as

$$p(H_0|X_1, X_2) \propto \frac{p(H_0|X_1)p(H_0|X_2)}{p(H_0)}\,. \tag{5.6}$$

Then, generalizing to n events, as:

$$p(H_0|X_1...X_n) \propto \frac{\prod_{n=1}^{n} p(H_0|X_n)}{p(H_0)^{n-1}}\,. \tag{5.7}$$

The resulting posterior is proportional to the product of the probability densities of each event, divided by the prior density $p(H_0)$. In practice, we use KDE to estimate the probability density, then sequentially combine events by multiplying probability densities, dividing by the prior density, and normalizing the final distribution [179].

A Single Realization

With GW170817 as the initial event and using our simulated events mentioned in Sec. 5.3, we apply the framework outlined later in the same section to sequentially combine each individual event posterior into one posterior on H_0 encompassing our set of 28 multi-messenger BNS events. The H_0 posteriors for each sequential combination can be seen in Fig. 5.1, where the mean posterior values are shown with 1σ ($\sim$ 68% credible interval) error bars. This demonstrates how the error will decrease with each sequential detection, but also how our estimates are easily impacted by uncertainties in a single detection when considering a small sample size. The fluctuations are mostly caused by over- or under-estimation of the injected distance by parameter estimation. For the first few events, the mean varies wildly, but after a handful of events, the mean values stabilize around the injected Planck18 value.

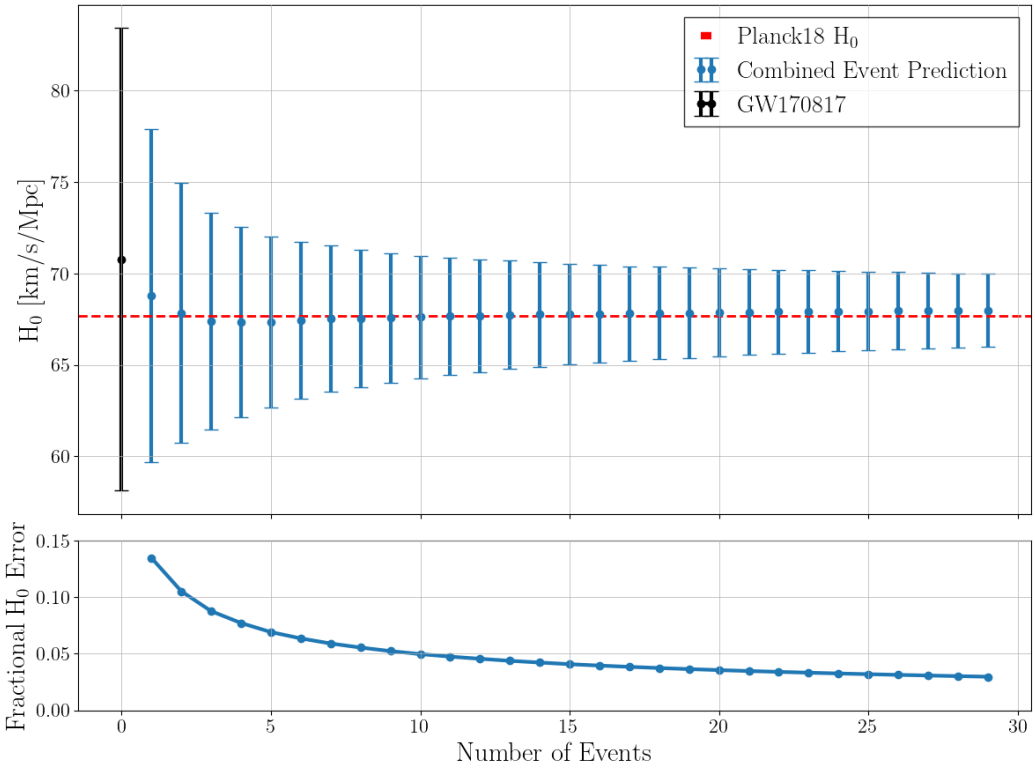

Figure 5.2: *Top:* H_0 posteriors averaged over 1000 realizations of the sequential combination of our simulated events covered in Sec. 5.3. Shown are the mean values with 1σ ($\sim$ 68% credible interval) error bars. *Bottom:* The fractional error, found as the ratio of the 1σ error to the mean H_0 value, for the same 1000 realizations above.

Averaged Over 1000 Realizations

While a single realization gives a realistic representation of how H_0 estimates can be made for N events and with each sequential detection, there are often significant deviations between the intermediate estimates of realizations. Each realization will come to the same end result, however, after combining all events. For example, the estimates shown in Fig. 5.1 have an initial downward bias in H_0 as by chance this realization happened to start with multiple events where the distance was slightly overestimated by parameter estimation.

In addition to Fig. 5.1 where we show H_0 estimates for a series of BNS detections,

we also take the the same set of events, randomize the order of combination to create 1000 realizations, and then average over the results at each step to produce Fig. 5.2. GW170817 is kept as the first event in each realization. This provides a quantitative estimate of precision in the measurement expected for N combined events, irrespective of fluctuations in estimates that result from the order the events are combined. Deviations from the injected value may be the result of our limited sample size, or the inclusion of GW170817, which is consistent with, but does not necessarily correspond to the injected cosmology. While the final result for all events is unchanged, this gives a more accurate estimate of the precision H_0 can be estimated to for an intermediate number of events.

Averaging over 1000 realizations of this simplified example outlined above, after combining 10 events (including GW170817), we find $H_0 = 67.6 \pm 3.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Combining all 28 events plus GW170817 gives us $H_0 = 68.0 \pm 2.0\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. These values are both consistent with the injected Planck18 value of $H_0 = 67.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

5.4 End-to-End Simulation: Mock Data Challenge

For a realistic end-to-end simulation of a set of standard siren measurements, we use the MDC carried out prior to O4 [69], and covered in Chapter 2. The MDC is a real-time simulation campaign carried out by injecting simulated GW waveforms within a replay of detector data from O3. In this way, we have realistic compact binary signals injected onto a realistic noise background, that the GW search pipelines can run on. This MDC is used to stress-test and assess the performance of our low-latency alert infrastructure. We are able to cross-match injections with GW triggers, or events, uploaded by the pipelines, giving us both injected and recovered parameter values. In addition, just like for real GW events, parameter estimation is run, including the production of `BAYESTAR` [216, 232] sky maps and `Bilby` [42, 200] posterior samples. We use these sky maps to help simulate the detection of an EM counterpart, which can be used to provide a redshift measurement for calculating H_0.

In total, 1,489 BNS events were detected out of $\sim$ 20,000 BNS injections via

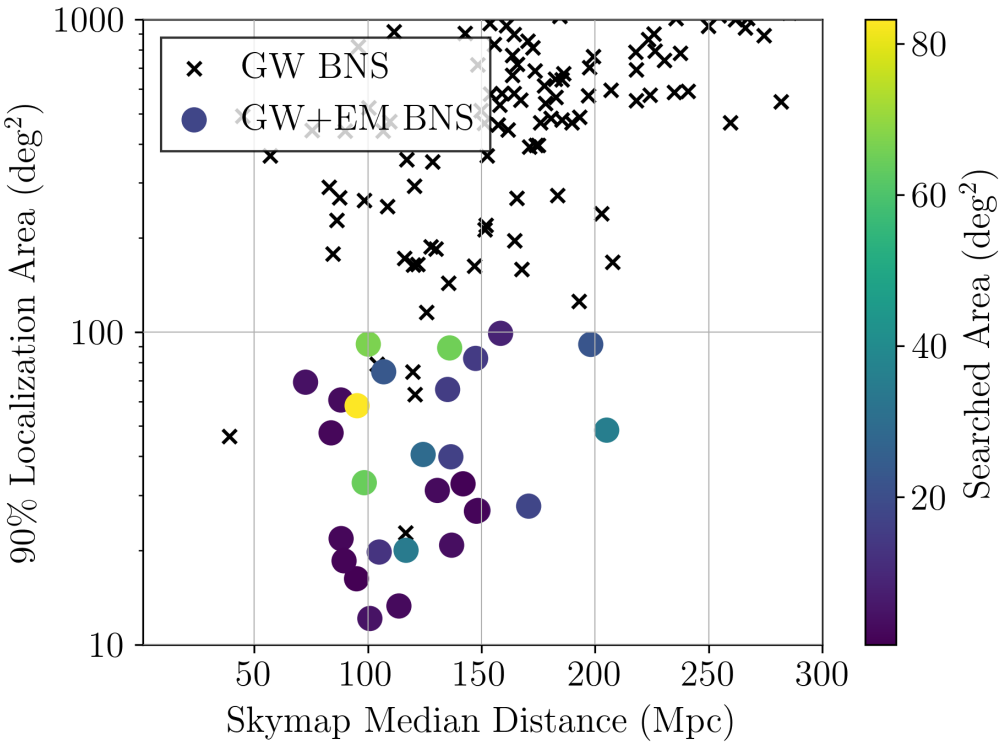

Figure 5.3: The 90% sky localization and median distance from the sky map, colored by the searched area for the 28 multi-messenger BNS events used from the MDC. In black are GW detections in the MDC which did not pass our thresholds for EM detection.

low-latency GW searches in a single 40-day MDC cycle. These detections are significant, i.e., they pass the significant FAR threshold defined at $\leq 3.9 \times 10^{-7}$ Hz (one per month). Up-to-date information on the current FAR threshold can be found in the IGWN userguide[2]. Out of these GW detections, the vast majority are too distant or poorly localized to reliably detect an EM counterpart if one was produced. To detect an EM counterpart, ToO searches must be able to cover enough of the sky localization and look deep enough to observe the likely faint and short-lived transient. We set conservative thresholds for EM detection, assuming BNS events within an injected luminosity distance of 200 Mpc and with a 90% localization area

[2]`https://emfollow.docs.ligo.org/userguide/`

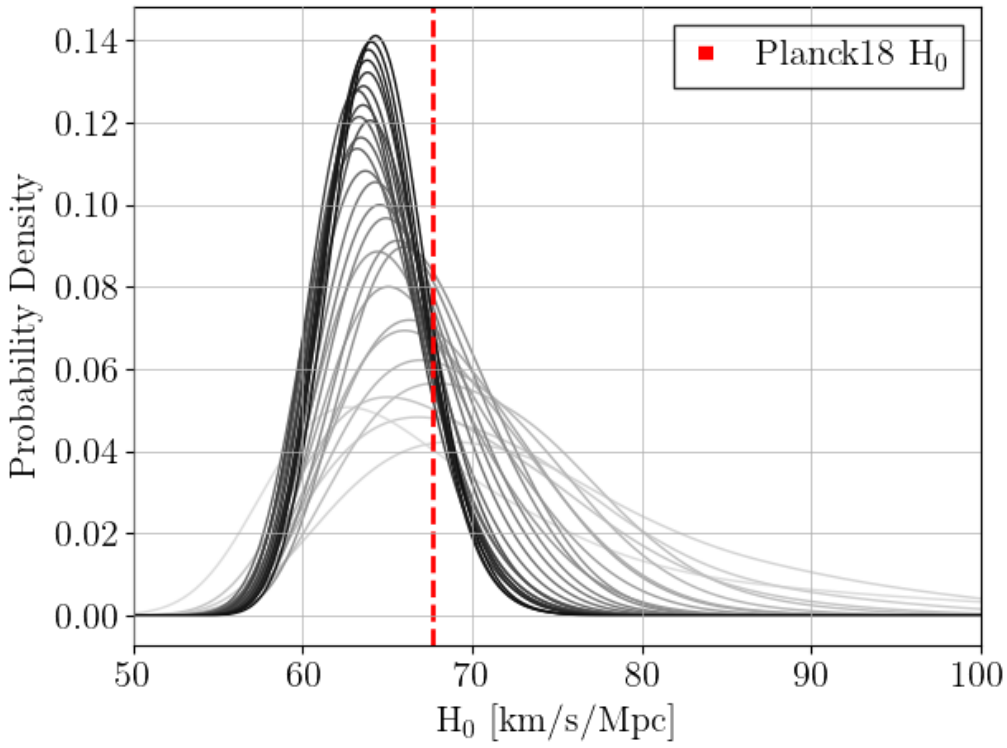

Figure 5.4: Posteriors for each successive combination of events, where darker lines include more events, without considering GW selection effects.

of less than 100 deg^2 have observed EM counterparts. These thresholds are meant to select events observable by both ZTF and Rubin, and are consistent with those assumed in [139]. This criteria narrows our sample down to just 29 BNS events, and one of those 29 is excluded from the rest of the study due to an issue with its corresponding parameter estimation, giving us a sample of 28 standard BNS events–plus GW170817. All of the MDC events are three detector events, including LIGO-Hanford, LIGO-Livingston [2], and Virgo [26] detectors. Fig. 5.3 shows the 90% localization area against the sky map median distance colored by searched area for the 28 multi-messenger BNS events used from the MDC. In black are shown adjacent GW events from the MDC which did not pass our thresholds for EM detection covered in this section.

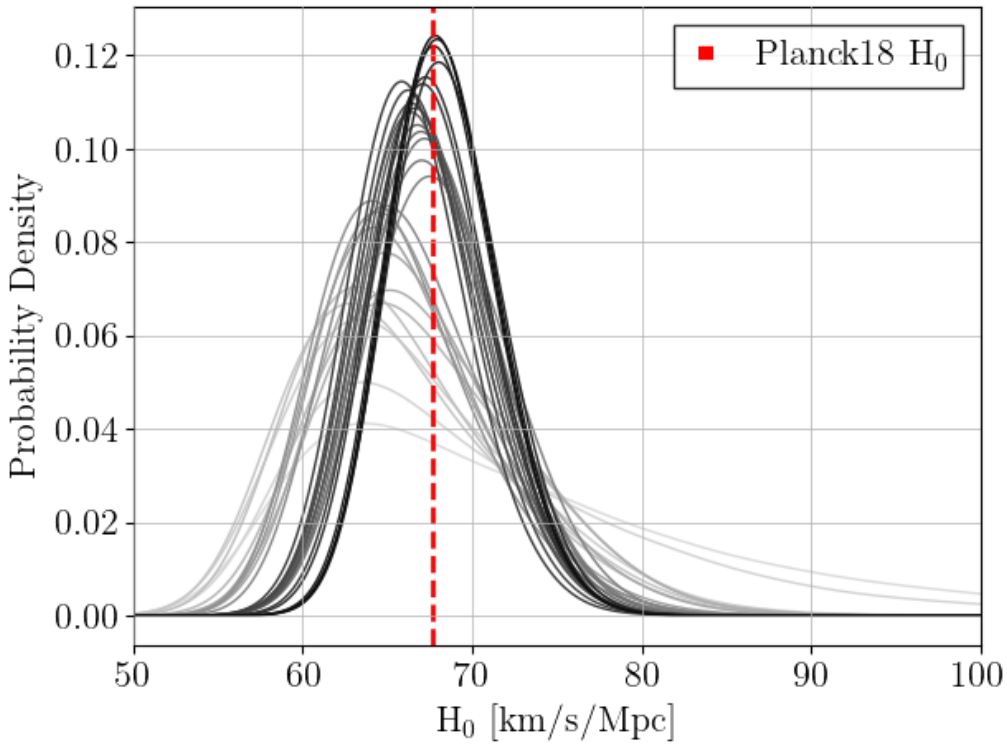

Figure 5.5: Posteriors for each successive combination of events, where darker lines include more events, including GW selection effects in Sec. 5.4.

Parameter Estimation

We use `Bilby` parameter estimation to provide our distance measurement and `BAYESTAR` for sky map statistics. The `Bilby` parameter estimation employs the `IMRPhenomD` waveform approximant [130, 138] to recover the observed signals, and assumes that dimensionless spins have norms less than 0.05 and are aligned with the orbital angular momentum. It noted the choice of waveform does not have a large impact or create significant bias when measuring H_0 [149]. To determine the redshift, we use the injected distance and assume Planck18 cosmology to find a redshift and associated uncertainty, as covered in Sec. 5.3. This is done to mimic an independent redshift measurement made by placing the EM source in a host galaxy.

We must also account for the prior used by `Bilby`, which is the same as shown in

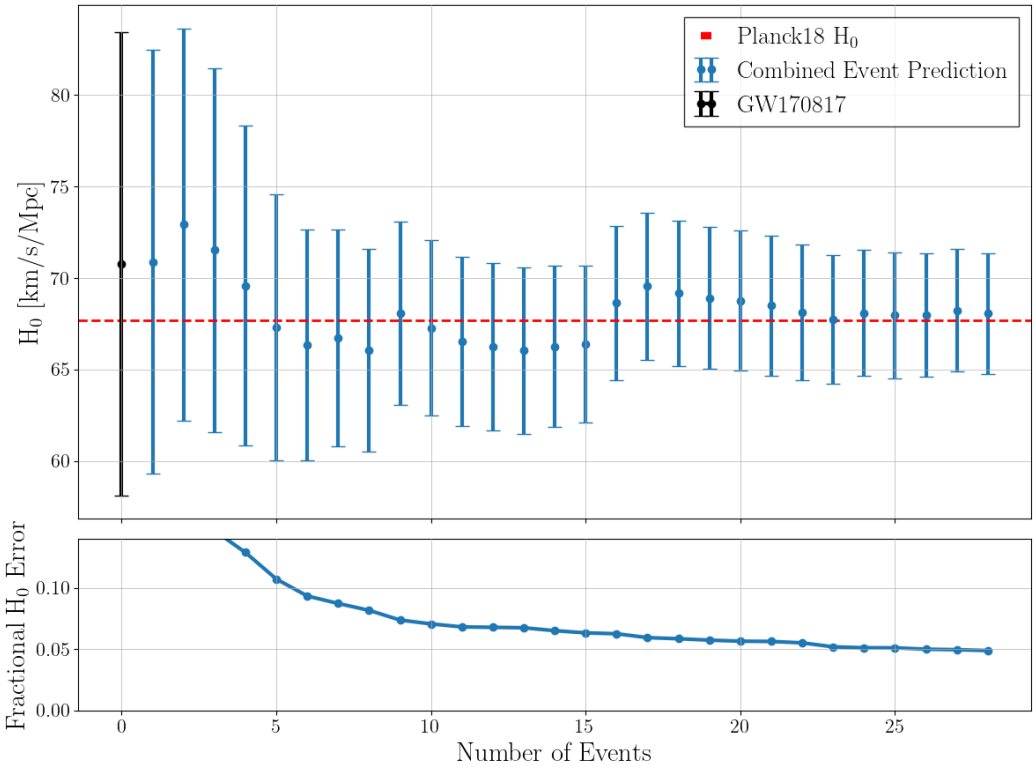

Figure 5.6: *Top:* H_0 posteriors for a sequential combination of the MDC events covered in Sec. 5.4. Shown are the mean values with 1σ ($\sim 68\%$ credible interval) error bars. *Bottom:* The fractional error, found as the ratio of the 1σ error to the mean H_0 value, for each sequential combination of the same events.

Eq. 5.6, a redshift prior which is uniform in comoving volume and source frame time.

GW Selection Effects

In our dataset, we use only significant events, which correspond to GW triggers which pass the significant FAR threshold defined at $\leq 3.9\times10^{-7}$ Hz (one per month) before the trials factor. Imposing a threshold creates a selection bias, however, as portions of the parameter space will be more likely to pass this threshold. To resolve this bias, we use Eq. (1) from [223]:

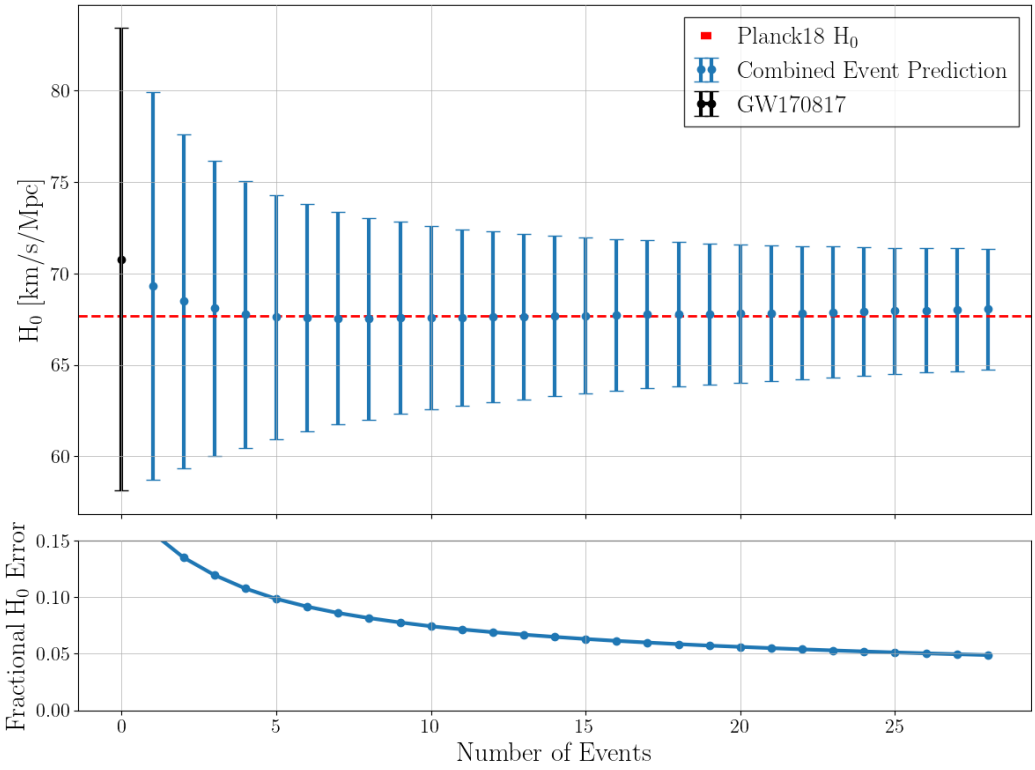

Figure 5.7: *Top:* H_0 posteriors averaged over 1000 realizations of the sequential combination of the MDC events covered in Sec. 5.4. Shown are the mean values with 1σ ($\sim$ 68% credible interval) error bars. *Bottom:* The fractional error, found as the ratio of the 1σ error to the mean H_0 value, for the same 1000 realizations above.

$$p_{det}(\theta) = \int_{p>p_{th}} d\mathbf{d}\, p(\mathbf{d}|\theta), \tag{5.8}$$

where p_{det} is found as the fraction of events that surpass the threshold p_{th}, in our case the significant FAR threshold, for a set of binary parameters θ. In order to find p_{det}, we assume that the selection bias is dominated by distance and SNR, and can be approximated from those alone. As we only consider BNS events which reside across a narrow range of masses and spins, we expect those biases to be minimal compared to distance and SNR.

Fig. 5.4 shows successive combination of events where the GW selection effects are not considered. Darker lines indicate the inclusion of more events. Fig. 5.5 shows the same event posteriors while accounting for the selection effects. As seen in Fig. 5.4, the selection effects lead to an underestimation H_0 via overestimation of distance, which is handled accurately in Fig. 5.5 by accounting for the biases on distance and SNR due to the significant FAR threshold. Averaging over 1000 realizations of the MDC events, after combining 10 events (including GW170817), we find $H_0 = 67.6 \pm 5.0\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$. Combining all 28 events plus GW170817 gives us $H_0 = 68.1 \pm 3.3\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$. Both these values are consistent with the injected Planck18 value of $H_0 = 67.4 \pm 0.5\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$.

5.5 Discussion

By applying a framework for combining H_0 posteriors from multiple events to a set of realistic multi-messenger BNS observations from the MDC, we demonstrate how H_0 can be estimated from a set of future multi-messenger observations and provide a quantitative estimate of the precision to which we can measure H_0 given a set number of detections.

We also quantify and mitigate the selection bias due GW searches by considering the fraction of MDC events detected within a given SNR and luminosity distance threshold in Sec. 5.4. This method could be applied to GW detections in a future observing run, and expanded beyond BNS events with the possible inclusion of additional merger quantities such as mass and spin. Our assumption of the selection bias being dominated by SNR and luminosity distance for BNS was proved effective for our dataset, but would have to be re-evaluated for masses and spins across all source classes. The selection bias in a future observing run could be accurately modeled using an MDC prior to the run with a comprehensive set of injections, just as was done for O4.

Unsurprisingly, the realistic MDC study in Sec. 5.4 had slightly higher uncertainty in the measurement of H_0 than the simplified example in Sec. 5.3. The MDC study had uncertainties of $\sim 7.5\%$ for after combining 10 events, and $\sim 5\%$ after all 29 events (including GW170817), compared to $\sim 5\%$ and $\sim 2.5\%$ for the simplified

example, as seen in Figs. 5.2, 5.7. We expect the MDC dataset to be a more realistic forecast of how uncertainties on H_0 for a set of standard siren detections, due to the inclusion of more realistic priors, GW selection effects, and search pipeline uncertainties. Averaging over 1000 realizations in Figs. 5.2, 5.7 also gives a better estimates of how the uncertainty decreases for each successive combination, free of the large statistical fluctuations for smaller numbers of events seen in Figs. 5.1, 5.6.

Using 1000 realizations of the MDC events, we find $H_0 = 67.6 \pm 5.0$ km s^{-1} Mpc^{-1} after combining 10 events, and $H_0 = 68.1 \pm 3.3$ km s^{-1} Mpc^{-1} after combining all 29 events. Both these estimations include GW170817 as one of the events, and are consistent with the injected Planck18 value of $H_0 = 67.4 \pm 0.5$ km s^{-1} Mpc^{-1}. Slight deviations from the injected H_0 value may be the result of our limited sample size, or the inclusion of GW170817, which is consistent with, but does not necessarily correspond to, the injected cosmology.

In the future, with next generation observatories and gravitational-wave detectors such as LISA [32], Cosmic Explorer [102, 196], and Einstein Telescope [3], we expect to see an increase in the number of BNS mergers we detect, making a set of numerous observations a possibility. With a set of standard siren BNS observations, we have shown that H_0 can be measured down to a few percent. These events will provide us with a means of measuring H_0 and shedding light on the Hubble Tension. We hope to use this framework for numerous multi-messenger detections in future observing runs, and that this framework can form the basis for a rapid pipeline for EM constraints alongside other efforts in `NMMA`, especially when considering promising machine learning-based parameter estimation techniques capable of running in seconds [68, 90].

Chapter 6

LOW-LATENCY CHIRP MASS ESTIMATES FOR GRAVITATIONAL-WAVE CANDIDATE EVENTS

6.1 Motivation

GW searches during O4 have yielded over 200 significant GW candidate events. These exciting discoveries enable scientific progress across a wide breadth of research areas, and are dependent on public GW data products and alerts. In this work, we introduce a new public GW data product that provides low-latency mass estimates for compact object mergers. These estimates are released as probabilities for a predetermined set of chirp mass ($\mathcal{M}$) bins for significant GW candidate events, and can be later updated by parameter estimation. The $\mathcal{M}$ is estimated through a combination of search pipeline information, sky maps, and parameter estimation. This data product will enable better interpretation of GW candidate events and aid multi-messenger ToO searches by providing mass estimates that can be used alongside source classification. We intend for this data product to be ingested by consumers of our alerts and enable downstream scientific efforts.

6.2 Public Chirp Mass ($\mathcal{M}$) Estimates

$\mathcal{M}$ is well measured from a GW signal, due to its relationship with the frequency evolution of the signal, so GW searches are well positioned to provide both low-latency $\mathcal{M}$ estimates. Low-latency GW searches have traditionally provided estimates of FAR, sky localization, and source classification, but no estimate of mass. We provide a new public data product for $\mathcal{M}$, currently being released for significant O4 candidate events. Chirp mass, $\mathcal{M}$, is defined as:

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}, \tag{6.1}$$

and is degenerate with the individual component masses, meaning a single $\mathcal{M}$ can correspond to multiple component mass pairs.

The LVK will provide binned estimates of $\mathcal{M}$ for CBC events, including cWB BBH [96, 143, 144] events. In solar masses, the $\mathcal{M}$ bins are:

$$[0.1, 0.87, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.7, 1.9, 2.1, 2.3, 3, 5.5, 11, 22, 44, 88, 1000]. \tag{6.2}$$

These estimates will be in the source frame, *except* in the case of the low-latency cWB BBH estimate, where there is no distance or redshift information available. In this case, we will report the "detector frame", or redshifted estimate. To shift to the source frame, we can simply relate the redshifted detector frame value $\mathcal{M}_z$ to $\mathcal{M}$ using the redshift z:

$$\mathcal{M}_z = (1 + z)\mathcal{M}. \tag{6.3}$$

The $\mathcal{M}$ information will be distributed by reporting the highest probability, or most likely, bin in the circular, and by making a `.json` file of the probabilities for each bin and a `.png` plot of the probabilities public on `GraceDB`, as seen in Fig. 6.1. The information will *not initially* be included in the alert packet `.xml`. This was done to ease integration for O4; inclusion is a possible goal for O5. Fig. 6.2 shows the low-latency and parameter estimation-based estimates run on MDC [69] events compared to their injected $\mathcal{M}$ values. We see that the distribution of $\mathcal{M}$ estimates closely match the injected values, particularly at low mass where there is large interest for multi-messenger follow-up.

The coarse $\mathcal{M}$ estimates are intended to help inform time-sensitive follow-up activities. This includes multi-messenger follow-up of events with at least one neutron star, as well as follow-up of high mass BBH events in the search for electromagnetic counterparts that could result from the flaring and disruption of AGN accretion disks during merger.

6.3 Bin Rationale

The bins were designed to be fine at lower masses where $\mathcal{M}$ is recovered very accurately, as evidenced in Fig. 6.2, and as these low masses are of particular interest do their potential of containing a NS. While specific bins do not exclusively

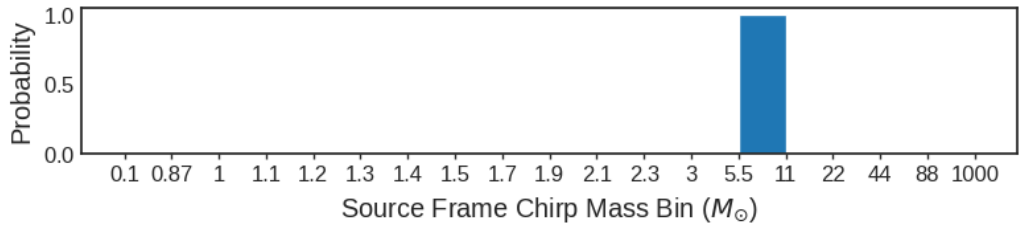

Figure 6.1: Example histogram of $\mathcal{M}$ estimate that will be available on `GraceDB` for significant events.

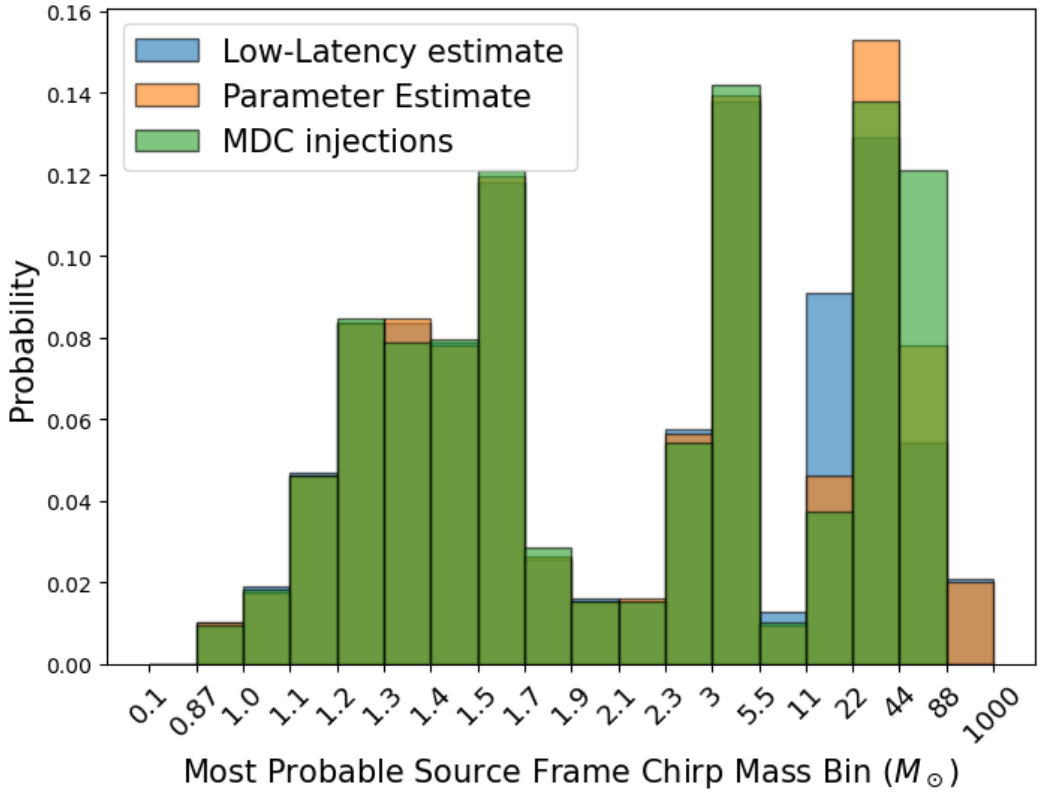

Figure 6.2: The low-latency and parameter estimation-based estimates run on MDC [69] events compared to their injected $\mathcal{M}$.

relate to a given source class, they were designed with the different regimes or classes in mind. As we cover the different regimes, keep in mind there are multiple component mass pairs for a given $\mathcal{M}$.

SSM Regime

This sub-solar mass (SSM) regime corresponds to the bin 0.1, 0.87 $M_\odot$, likely to contain an event with a component mass $< 1M_\odot$ Here, I highlight an example $\mathcal{M}$ of 0.87 $M_\odot$, which may correspond to a 1 $M_\odot$, 1 $M_\odot$ merger, demonstrating events with $\mathcal{M} < 0.87M_\odot$ will contain a component mass $< 1M_\odot$.

`HasNS` Regime

A reference to the `EM-Bright` [67] quantity `HasNS` which provides the probability a GW candidate event contains a NS, this regime roughly corresponds to $\mathcal{M}$ values with a strong possibility of containing a NS. The bins in the $\mathcal{M}$ range 0.87, 3 $M_\odot$ roughly fall within this regime, again with a reminded these regimes are not exclusive and are for illustrative purposes. Below are a pair of examples in this regime:

Example i): A $\mathcal{M}$ of 1.2 $M_\odot$ may correspond to a 1.4 $M_\odot$, 1.4 $M_\odot$ merger.

Example ii): A $\mathcal{M}$ of 3 $M_\odot$ may correspond to a 1.4 $M_\odot$, 10 $M_\odot$ merger.

BBH Regime

The coarse high mass bins are likely to correspond to BBH events. $\mathcal{M}$ bins above 3 $M_\odot$ have a strong chance of being a BBH. Below are a pair of examples in this regime:

Example i): A $\mathcal{M}$ of 8.7 $M_\odot$ may correspond to a 10 $M_\odot$, 10 $M_\odot$ merger.

Example ii): A $\mathcal{M}$ of 44 $M_\odot$ may correspond to a 50 $M_\odot$, 50 $M_\odot$ merger.

6.4 Implementation in alert infrastructure

Binned $\mathcal{M}$ estimates will come in two forms: (i) estimates based on information available in low latency (pipeline uploads, sky maps), and (ii) estimates based on `Bilby` parameter estimation [41, 200]. The former will be immediately available

in low latency, while the latter will accompany update alerts ~ hour(s) later when `Bilby` parameter estimation is complete. $\mathcal{M}$ estimates will only be provided for significant events. Similar to the procedure with other update information (EM-Bright, sky maps, etc), the updated $\mathcal{M}$ based on `Bilby` parameter estimation will be uploaded and identified as resulting from parameter estimation, without deleting or overwriting the original low-latency estimate. The estimates assume the candidate is astrophysical in origin.

I personally led the software development of the package `ligo-cgmi`, as well as the implementation of that package into the low-latency alert infrastructure, and continue to maintain the package and monitor the performance of the data product. `ligo-cgmi`, with the "CGMI" portion standing for Coarse-Grained Mass Information, contains the software that provides both the low-latency and parameter based estimates. This package was implemented into `GWCelery`, the orchestrator and task manager for the LLAI, and circular changes made in `ligo-followup-advocate` which provides circular templates for GW candidate events.

Low-latency $\mathcal{M}$

The low-latency $\mathcal{M}$ estimates will use the pipeline point estimate of the detector $\mathcal{M}$, combined with distance information provided by the `BAYESTAR` sky map [216], in order to correct the detector $\mathcal{M}$ value to the source frame. The result is binned and converted to an estimate of the probability corresponding to each bin. In the case of cWB BBH, there is no distance information, and therefore we assume a redshift of zero when reporting the source $\mathcal{M}$. The estimate will be uploaded in the form of a `.json` file named `mchirp_source.json`.

Parameter estimation-based $\mathcal{M}$

The $\mathcal{M}$ estimates based on `Bilby` parameter estimation will use the PE source frame values directly. The result is binned and converted to an estimate of the probability corresponding to each bin, and that estimate will be uploaded in the form of a `.json` file named `mchirp_source_PE.json`

6.5 Circular Changes

As previously mentioned, while not present in the alert packet, the $\mathcal{M}$ estimates will be highlighted in circulars for significant events. Specifically, the highest probability bin will be provided in the circular. The additional circular text is as follows.

Initial Circular

The format of the Initial Alert circular text is the following:

"The source chirp mass falls with highest probability in the bin (`left_bin, right_bin`) *solar masses, assuming the candidate is astrophysical in origin."*

Update Circular

For Update Circulars, the following text includes mention of the estimate coming from parameter estimation:

"The source chirp mass falls with highest probability in the bin (`left_bin, right_bin`) *solar masses* ***after parameter estimation [citation to Bilby],*** *assuming the candidate is astrophysical in origin."*

cWB BBH Circulars

For superevents where cWB BBH is the preferred event, the circular specifies that the redshift is assumed to be zero for the low-latency case:

"The ***redshifted (detector frame)*** *chirp mass falls with highest probability in the bin* (`left_bin, right_bin`) *solar masses, assuming the candidate is astrophysical in origin.* ***For cWB BBH events, no distance or redshift information is available in low latency.****"*

6.6 Multi-Messenger Applications

In this work we highlight two direct multi-messenger applications of the $\mathcal{M}$ estimates: i) inform follow-up of potential NS mergers by helping determine whether

the merger likely contained a NS, and how likely that event is to produce a kilonova, and ii) targeted follow-up of high mass BBH events in the hope of detecting an EM counterpart in the form of a flare from an AGN.

i) Neutron Star Merger Follow-up

ToO searches rely on low-latency data products and timely decision making when determining which GW candidate events are worth devoting telescope resources to. When searching for kilonovae, one main criteria astronomers consider is whether the candidate event likely contains a NS. Alongside quantities such as `HasNS` and p_{astro}, we expect the $\mathcal{M}$ estimates to help answer this question. In addition, not only do these quantities offer insight on the possible presence of a NS, they also provide an actual mass estimate. This is useful as we expect different masses to produce varying amounts of mass ejecta and consequently have varying likelihoods of producing a kilonova. As established in 4.7, we expect lower mass BNS events to produce both greater mass ejecta and a brighter kilonova peak. The $\mathcal{M}$ could also be used to predict the resulting remnant of the merger [165].

ii) BBH Follow-up

It is expected that high-mass BBH mergers that take place in the accretion disk of an AGN may produce an observable EM flare [121]. `BBHBot` is a automatic triggering pipeline designed to monitor GW triggers in real time, and schedule ToO observations with ZTF [50] to observe promising BBH candidates. `BBHBot` ingests public low-latency GW data products such as the new $\mathcal{M}$ estimates as of recently, in addition to the sky map and FAR in order to determine which events meet the triggering criteria. If candidate events pass cuts on mass and FAR as well as meet our observing criteria, including cuts on the 90% sky map area and air mass, an observing plan is automatically drafted and submitted to trigger ZTF.

Specifically, we are interested in events with $> 50\ M_\odot$ of total mass, and even more so those with $> 100\ M_\odot$ of total mass, so we filter for events most likely to fall within $\mathcal{M}$ bins $> 22\ M_\odot$. Keeping in mind the definition of $\mathcal{M}$ in Eq. 6.1, the merger of a pair of $25\ M_\odot$ BHs corresponds to a $\mathcal{M} \sim 22\ M_\odot$, while the merger of a pair of $50\ M_\odot$ BHs corresponds to a $\mathcal{M} \sim 44\ M_\odot$. Conveniently, there are $\mathcal{M}$ bin

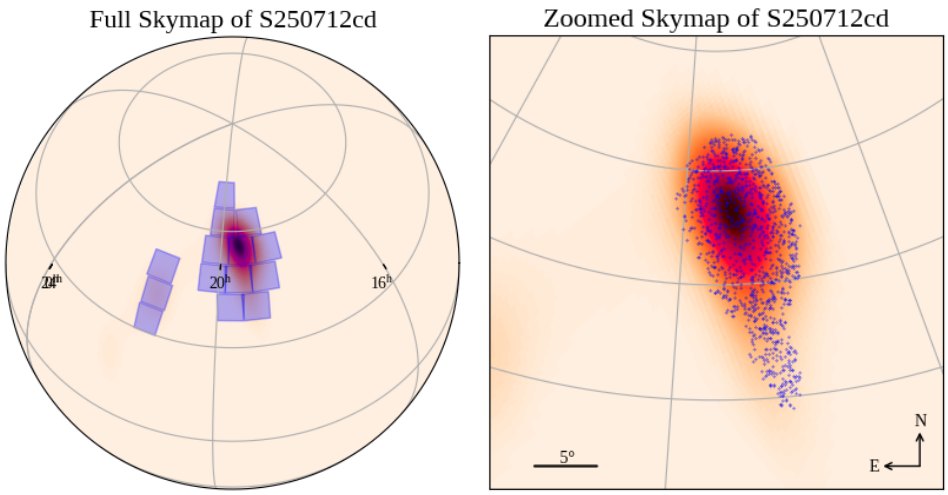

Figure 6.3: Targeted follow-up of the BBH GW candidate event S250712cd, displayed on the GW sky localization. ZTF telescope tilings are shown by the purple rectangles on the left, while on the right we see a zoomed version with known AGN marked as purple dots for reference.

edges at both 22 and 44 $M_\odot$.

Following the addition of $\mathcal{M}$ to the public GW data products in June 2025, `BBHBot` has used it in the real-time evaluation of more than a dozen candidate BBH mergers as of August 2025. Of these, nine exceeded the selection threshold of a $\mathcal{M} > 22\ M_\odot$. One candidate GW event, S250712cd[1], passed all selection criteria and was scheduled for automated observations by ZTF. The targeted follow-up of S250712cd is seen in Fig. 6.3, displayed on the GW sky localization. On the left, the ZTF telescope tilings are shown by the purple rectangles, while on the right we see a zoomed version of the sky map with known AGN marked as purple dots for reference. Those AGN are the sources that will be observed as candidates for potential flaring events.

[1] https://gcn.nasa.gov/circulars/41075

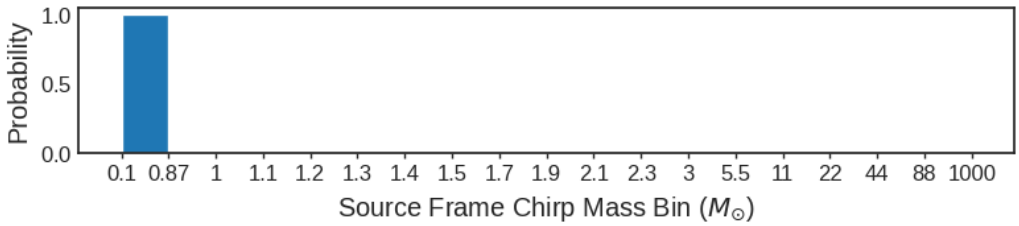

Figure 6.4: Histogram of $\mathcal{M}$ estimate for S250818k publicly available on `GraceDB`.

6.7 A timely example: S250818k

At the time of writing, S250818k is a low-significance GW candidate of interest. Given a possible external coincidence with an EM source[2], $\mathcal{M}$ estimates were provided for this GW event[3]. This event is of particular note given its low mass and high chance of containing a NS and possibly a sub-solar mass (SSM) object. In Fig. 6.4, we show the $\mathcal{M}$ estimate for this event, falling within the .1 to .87 $M_\odot$ bin. Both the low-latency and parameter estimation-based estimate agree on this bin. Should this candidate eventually make the GWTC, it would be significant as the first SSM event, and would be a nice demonstration of the usefulness of these $\mathcal{M}$ estimates.

6.8 Discussion

We provide public $\mathcal{M}$ estimates as a low-latency GW data product in the form of binned probabilities, which are plotted and uploaded to `GraceDB` as well as reported in the circular. With estimates using a combination of search pipeline information, sky maps, and parameter estimation, we strive to enable better interpretation of GW candidate events and inform multi-messenger ToO searches. We intend for this data product to be ingested by consumers of our alerts alongside source classification and enable downstream scientific efforts. Possible high-impact areas include multi-messenger follow-up of events with at least one neutron star, and searches for electromagnetic counterparts that could result from the flaring and disruption of AGN accretion disks during BBH mergers. We will continue to advocate for additional future data products with multi-messenger applications alongside these

[2]https://gcn.nasa.gov/circulars/41414
[3]https://gcn.nasa.gov/circulars/41440

mass estimates.

Chapter 7

DISCUSSION: THE ROAD AHEAD

There have been over 200 significant GW candidate events detected during O4 already. We have successfully sent out alerts for these events with a median time of $\sim$ 30 s after merger, an impressive feat enabling prompt follow-up of our events. In this era of multi-messenger astrophysics, just as astronomers rely on IGWN's low-latency alert products, we rely on their EM follow-up and complementary neutrino searches to accompany our GW detections and enrich our scientific discoveries. As we saw with GW170817, when working collaboratively across disciplines and collaborations, we can take monumental steps forward for modern astrophysics.

As O4 continues, we are still on the hunt for another joint, multi-messenger detection of a NS merger and associated EM counterparts such a kilonova or GRB. This only strengthens the demand for additional multi-messenger focused data products that will increase our chances of detection. Astronomers are eager for NS events. In this work, we cover multiple data products to that end: i) the multiple-pipeline ML classifier in Chapter 3, ii) the kilonova ejecta and light curve data products in Chapter 4, and iii) the $\mathcal{M}$ estimates in Chapter 6.

Our ML classifier is designed to complement p_{astro} and `GWSkyNet` values, by helping determine whether candidate events are astrophysical or terrestrial in nature. One major advantage of our classifier is that it utilizes information provided by multiple low-latency search pipelines its feature space. Events detected by multiple searches are much more likely to be astrophysical, while those only found by one search are much more likely to be terrestrial in origin. This is something that current data products do not account for, but our classifier would change that. We will advocate for the implementation and public release of our classifier's outputs in a future observing run, with the hope this classifier would be used to determine significance of future GW candidate events and inform follow-up efforts.

The kilonova data products estimate ejecta mass and peak light curve estimates for

GW candidate events in the form of a 90% credible interval, as well as a heat map plot for the range of expected light curves. We also define a quantity `HasEjecta`, to be used alongside the other `EM-Bright` quantities such as `HasNS` and `HasRemnant`, which estimates the likelihood the ejecta mass is $< 10^{-3} M_{\odot}$. We expect these data products to be a useful metric for astronomers in determining both whether a GW candidate event is capable of producing a kilonova, and if so, how bright it will be.

The $\mathcal{M}$ estimates, which are already implemented and being made public for significant GW events, provide probabilities for a predetermined set of $\mathcal{M}$ bins. These probabilities are available on `GraceDB` as well as reported in the circular. With this data product, we strive to enable better interpretation of GW candidate events by those outside the collaboration. We intend for this data product to be ingested by consumers of our alerts alongside source classification and enable downstream scientific efforts, such as multi-messenger follow-up of events with at least one neutron star and searches for EM counterparts that may result from BBH mergers that take place in AGN accretion disks.

In the future, next generation observatories and GW detectors such as Cosmic Explorer [102, 196], Einstein Telescope [3], and LISA [32] will lead to an increased rate of GW detections, including NS events capable of producing kilonovae and GRBs. Processing this higher rate of candidate events and identifying the most promising ones for EM follow-up will be challenging. I strive to enable multi-messenger searches in every way that I can, including advocating for additional future data products with multi-messenger applications and facilitation their release. As I conclude, I look forward with optimism and excitement for our field, and anticipate great stories of success in the years to come.

BIBLIOGRAPHY

[1] M. G. Aartsen et al. The IceCube Neutrino Observatory: Instrumentation and Online Systems. *JINST*, 12(03):P03012, 2017. doi: 10.1088/1748-0221/12/03/P03012. [Erratum: JINST 19, E05001 (2024)].

[2] Aasi et al. Advanced ligo. *Classical and Quantum Gravity*, 32(7):074001, 2015.

[3] Adrian Abac et al. The Science of the Einstein Telescope. 3 2025.

[4] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, and et al. Gw190425: Observation of a compact binary coalescence with total mass $\sim 3.4 m_{\odot}$. *The Astrophysical Journal*, 892(1):L3, Mar 2020. ISSN 2041-8213. doi: 10.3847/2041-8213/ab75f5. URL http://dx.doi.org/10.3847/2041-8213/ab75f5.

[5] B. P. Abbott et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Rel.*, 19:1, 2016. doi: 10.1007/s41114-020-00026-9.

[6] B. P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102, 2016. doi: 10.1103/PhysRevLett.116.061102.

[7] B. P. Abbott et al. Multi-messenger observations of a binary neutron star merger*. *The Astrophysical Journal Letters*, 848(2):L12, oct 2017. doi: 10.3847/2041-8213/aa91c9. URL https://dx.doi.org/10.3847/2041-8213/aa91c9.

[8] B. P. Abbott et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551(7678):85–88, 2017. doi: 10.1038/nature24471.

[9] B. P. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101, 2017. doi: 10.1103/PhysRevLett.119.161101.

[10] B. P. Abbott et al. Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A. *Astrophys. J. Lett.*, 848(2):L13, 2017. doi: 10.3847/2041-8213/aa920c.

[11] B. P. Abbott et al. GW170817: Measurements of neutron star radii and equation of state. *Phys. Rev. Lett.*, 121(16):161101, 2018. doi: 10.1103/PhysRevLett.121.161101.

[12] B. P. Abbott et al. Low-latency Gravitational-wave Alerts for Multimessenger Astronomy during the Second Advanced LIGO and Virgo Observing Run. *Astrophys. J.*, 875(2):161, 2019. doi: 10.3847/1538-4357/ab0e8f.

[13] B. P. Abbott et al. GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4 M_{\odot}$. *Astrophys. J. Lett.*, 892(1):L3, 2020. doi: 10.3847/2041-8213/ab75f5.

[14] Benjamin P. Abbott et al. Sensitivity of the Advanced LIGO detectors at the beginning of gravitational wave astronomy. *Phys. Rev. D*, 93(11):112004, 2016. doi: 10.1103/PhysRevD.93.112004. [Addendum: Phys.Rev.D 97, 059901 (2018)].

[15] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, and et al. Observation of gravitational waves from two neutron star–black hole coalescences. *The Astrophysical Journal Letters*, 915(1):L5, Jun 2021. ISSN 2041-8213. doi: 10.3847/2041-8213/ac082e. URL http://dx.doi.org/10.3847/2041-8213/ac082e.

[16] R. Abbott et al. All-sky search for long-duration gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run. *Phys. Rev. D*, 104(10):102001, 2021. doi: 10.1103/PhysRevD.104.102001.

[17] R. Abbott et al. All-sky search for short gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run. *Phys. Rev. D*, 104(12):122004, 2021. doi: 10.1103/PhysRevD.104.122004.

[18] R. Abbott et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. *arXiv e-prints*, 11 2021.

[19] R. Abbott et al. Tests of General Relativity with GWTC-3. 12 2021.

[20] R. Abbott et al. All-sky search for continuous gravitational waves from isolated neutron stars using Advanced LIGO and Advanced Virgo O3 data. *Phys. Rev. D*, 106(10):102008, 2022. doi: 10.1103/PhysRevD.106.102008.

[21] R. Abbott et al. Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3. *Phys. Rev. X*, 13(1):011048, 2023. doi: 10.1103/PhysRevX.13.011048.

[22] R. Abbott et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run. *Phys. Rev. X*, 13(4):041039, 2023. doi: 10.1103/PhysRevX.13.041039.

[23] Thomas C. Abbott, Eitan Buffaz, Nicholas Vieira, Miriam Cabero, Daryl Haggard, Ashish Mahabal, and Jess McIver. GWSkyNet-Multi: A Machine-learning Multiclass Classifier for LIGO-Virgo Public Alerts. *The Astrophysical Journal*, 927(2):232, March 2022. doi: 10.3847/1538-4357/ac5019.

[24] Abbott et al. Localization and broadband follow-up of the gravitational-wave transient GW150914. *Astrophys. J.*, 826(1):L13, 2016. doi: 10.3847/2041-8205/826/1/L13.

[25] Ernazar Abdikamalov, Giulia Pagliaroli, and David Radice. Gravitational Waves from Core-Collapse Supernovae. 10 2020. doi: 10.1007/978-981-15-4702-7_21-1.

[26] Acernese et al. Advanced Virgo. *Classical and Quantum Gravity*, 32(2): 024001, 2015.

[27] Michalis Agathos, Francesco Zappa, Sebastiano Bernuzzi, Albino Perego, Matteo Breschi, and David Radice. Inferring prompt black-hole formation in neutron star mergers from gravitational-wave data. *Physical Review D*, 101(4), feb 2020. doi: 10.1103/physrevd.101.044006. URL `https://doi.org/10.1103%2Fphysrevd.101.044006`.

[28] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020. doi: 10.1051/0004-6361/201833910. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[29] Tomás Ahumada et al. Searching for Gravitational Wave Optical Counterparts with the Zwicky Transient Facility: Summary of O4a. *Publ. Astron. Soc. Pac.*, 136(11):114201, 2024. doi: 10.1088/1538-3873/ad8265.

[30] T. Akutsu et al. Overview of KAGRA: Detector design and construction history. *PTEP*, 2021(5):05A101, 2021. doi: 10.1093/ptep/ptaa125.

[31] Mouza Almualla et al. Optimizing serendipitous detections of kilonovae: cadence and filter selection. *Mon. Not. Roy. Astron. Soc.*, 504(2):2822–2831, 2021. doi: 10.1093/mnras/stab1090.

[32] Pau Amaro-Seoane, Heather Audley, Stanislav Babak, John Baker, Enrico Barausse, Peter Bender, Emanuele Berti, Pierre Binetruy, Michael Born, Daniele Bortoluzzi, Jordan Camp, Chiara Caprini, Vitor Cardoso, Monica Colpi, John Conklin, Neil Cornish, Curt Cutler, Karsten Danzmann, Rita Dolesi, Luigi Ferraioli, Valerio Ferroni, Ewan Fitzsimons, Jonathan Gair, Lluis Gesa Bote, Domenico Giardini, Ferran Gibert, Catia Grimani, Hubert Halloin, Gerhard Heinzel, Thomas Hertog, Martin Hewitson, Kelly Holley-Bockelmann, Daniel Hollington, Mauro Hueller, Henri Inchauspe, Philippe Jetzer, Nikos Karnesis, Christian Killow, Antoine Klein, Bill Klipstein, Natalia Korsakova, Shane L Larson, Jeffrey Livas, Ivan Lloro, Nary Man, Davor Mance, Joseph Martino, Ignacio Mateos, Kirk McKenzie, Sean T McWilliams, Cole Miller, Guido Mueller, Germano Nardini, Gijs Nelemans, Miquel Nofrarias, Antoine Petiteau, Paolo Pivato, Eric Plagnol, Ed Porter, Jens Reiche, David Robertson, Norna Robertson, Elena Rossi, Giuliana Russano, Bernard Schutz, Alberto Sesana, David Shoemaker, Jacob Slutsky, Carlos F. Sopuerta, Tim Sumner, Nicola Tamanini, Ira Thorpe, Michael Troebs, Michele Vallisneri, Alberto Vecchio, Daniele Vetrugno, Stefano Vitale, Marta Volonteri, Gudrun Wanner, Harry Ward, Peter Wass, William Weber, John Ziemer, and Peter Zweifel. Laser Interferometer Space Antenna. *arXiv e-prints*, art. arXiv:1702.00786, February 2017. doi: 10.48550/arXiv.1702.00786.

[33] Cristina Andrade, Raiyah Alserkal, Luis Salazar Manzano, Emma Martin, Igor Andreoni, Michael W. Coughlin, Nidhal Guessoum, and Liliana Rivera Sandoval. The Effect of Vera C. Rubin Observatory Cadence Selections on Kilonova Detectability. 2 2025.

[34] Igor Andreoni et al. Fast-transient Searches in Real Time with ZTFReST: Identification of Three Optically Discovered Gamma-Ray Burst Afterglows and New Constraints on the Kilonova Rate. *Astrophys. J.*, 918(2):63, 2021. doi: 10.3847/1538-4357/ac0bc7.

[35] Igor Andreoni et al. Enabling kilonova science with Nancy Grace Roman Space Telescope. *Astropart. Phys.*, 155:102904, 2024. doi: 10.1016/j.astropartphys.2023.102904.

[36] Nicolas Andres et al. Assessing the compact-binary merger candidates reported by the MBTA pipeline in the LIGO–Virgo O3 run: probability of astrophysical origin, classification, and associated uncertainties. *Class. Quant. Grav.*, 39(5):055002, 2022. doi: 10.1088/1361-6382/ac482a.

[37] Eemeli Annala, Tyler Gorda, Aleksi Kurkela, and Aleksi Vuorinen. Gravitational-wave constraints on the neutron-star-matter equation of state.

Phys. Rev. Lett., 120:172703, Apr 2018. doi: 10.1103/PhysRevLett.120.172703. URL https://link.aps.org/doi/10.1103/PhysRevLett.120.172703.

[38] Nicolas Arnaud. LIGO and Virgo detector characterization and data quality: Contributions to the O3 run and preparation for O4. *Nucl. Instrum. Meth. A*, 1048:167945, 2023. doi: 10.1016/j.nima.2022.167945.

[39] Stefano Ascenzi, Michael W Coughlin, Tim Dietrich, Ryan J Foley, Enrico Ramirez-Ruiz, Silvia Piranomonte, Brenna Mockler, Ariadna Murguia-Berthier, Chris L Fryer, Nicole M Lloyd-Ronning, and Stephan Rosswog. A luminosity distribution for kilonovae based on short gamma-ray burst afterglows. *Monthly Notices of the Royal Astronomical Society*, 486(1): 672–690, 03 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz891. URL https://doi.org/10.1093/mnras/stz891.

[40] G. Ashton, E. Burns, T. Dal Canton, T. Dent, H. B Eggenstein, A. B. Nielsen, R. Prix, M. Was, and S. J. Zhu. Coincident detection significance in multimessenger astronomy. *Astrophys. J.*, 860(1):6, 2018. doi: 10.3847/1538-4357/aabfd2.

[41] Gregory Ashton, Moritz Hübner, Paul D. Lasky, Colm Talbot, Kendall Ackley, Sylvia Biscoveanu, Qi Chu, Atul Divakarla, Paul J. Easter, Boris Goncharov, Francisco Hernandez Vivanco, Jan Harms, Marcus E. Lower, Grant D. Meadors, Denyz Melchor, Ethan Payne, Matthew D. Pitkin, Jade Powell, Nikhil Sarin, Rory J. E. Smith, and Eric Thrane. BILBY: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy. *The Astrophysical Journal Supplement*, 241(2):27, April 2019. doi: 10.3847/1538-4365/ab06fc.

[42] Gregory Ashton et al. BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy. *Astrophys. J. Suppl.*, 241(2):27, 2019. doi: 10.3847/1538-4365/ab06fc.

[43] F. Aubin et al. The MBTA pipeline for detecting compact binary coalescences in the third LIGO–Virgo observing run. *Class. Quant. Grav.*, 38(9):095004, 2021. doi: 10.1088/1361-6382/abe913.

[44] John Baker et al. The Laser Interferometer Space Antenna: Unveiling the Millihertz Gravitational Wave Sky. 7 2019.

[45] T. Baker, E. Bellini, P. G. Ferreira, M. Lagos, J. Noller, and I. Sawicki. Strong constraints on cosmological gravity from GW170817 and GRB 170817A.

Phys. Rev. Lett., 119(25):251301, 2017. doi: 10.1103/PhysRevLett.119.251301.

[46] Sharan Banagiri, Christopher P. L. Berry, Gareth S. Cabourn Davies, Leo Tsukada, and Zoheyr Doctor. Unified pastro for gravitational waves: Consistently combining information from multiple search pipelines. *Phys. Rev. D*, 108(8):083043, 2023. doi: 10.1103/PhysRevD.108.083043.

[47] A. Bauswein, T. W. Baumgarte, and H.-T. Janka. Prompt merger collapse and the maximum mass of neutron stars. *Phys. Rev. Lett.*, 111:131101, Sep 2013. doi: 10.1103/PhysRevLett.111.131101. URL https://link.aps.org/doi/10.1103/PhysRevLett.111.131101.

[48] Andreas Bauswein, Oliver Just, Hans-Thomas Janka, and Nikolaos Stergioulas. Neutron-star Radius Constraints from GW170817 and Future Detections. *The Astrophysical Journal Letters*, 850(2):L34, December 2017. doi: 10.3847/2041-8213/aa9994.

[49] Andreas Bauswein, Oliver Just, Hans-Thomas Janka, and Nikolaos Stergioulas. Neutron-star radius constraints from GW170817 and future detections. *Astrophys. J. Lett.*, 850(2):L34, 2017. doi: 10.3847/2041-8213/aa9994.

[50] E. Bellm. The Zwicky Transient Facility. In P. R. Wozniak, M. J. Graham, A. A. Mahabal, and R. Seaman, editors, *The Third Hot-wiring the Transient Universe Workshop*, pages 27–33, 2014.

[51] E. C. Bellm, S. R. Kulkarni, M. J. Graham, R. Dekany, R. M. Smith, R. Riddle, F. J. Masci, G. Helou, T. A. Prince, S. M. Adams, C. Barbarino, T. Barlow, J. Bauer, R. Beck, J. Belicki, R. Biswas, N. Blagorodnova, D. Bodewits, B. Bolin, V. Brinnel, T. Brooke, B. Bue, M. Bulla, R. Burruss, S. B. Cenko, C.-K. Chang, A. Connolly, M. Coughlin, J. Cromer, V. Cunningham, K. De, A. Delacroix, V. Desai, D. A. Duev, G. Eadie, T. L. Farnham, M. Feeney, U. Feindt, D. Flynn, A. Franckowiak, S. Frederick, C. Fremling, A. Gal-Yam, S. Gezari, M. Giomi, D. A. Goldstein, V. Z. Golkhou, A. Goobar, S. Groom, E. Hacopians, D. Hale, J. Henning, A. Y. Q. Ho, D. Hover, J. Howell, T. Hung, D. Huppenkothen, D. Imel, W.-H. Ip, Ž. Ivezić, E. Jackson, L. Jones, M. Juric, M. M. Kasliwal, S. Kaspi, S. Kaye, M. S. P. Kelley, M. Kowalski, E. Kramer, T. Kupfer, W. Landry, R. R. Laher, C.-D. Lee, H. W. Lin, Z.-Y. Lin, R. Lunnan, M. Giomi, A. Mahabal, P. Mao, A. A. Miller, S. Monkewitz, P. Murphy, C.-C. Ngeow, J. Nordin, P. Nugent, E. Ofek, M. T. Patterson, B. Penprase, M. Porter, L. Rauch, U. Rebbapragada, D. Reiley, M. Rigault, H. Rodriguez, J. van Roestel, B. Rusholme, J. van Santen, S. Schulze, D. L. Shupe, L. P.

Singer, M. T. Soumagnac, R. Stein, J. Surace, J. Sollerman, P. Szkody, F. Taddia, S. Terek, A. Van Sistine, S. van Velzen, W. T. Vestrand, R. Walters, C. Ward, Q.-Z. Ye, P.-C. Yu, L. Yan, and J. Zolkower. The Zwicky Transient Facility: System Overview, Performance, and First Results. *Publications of the Astronomical Society of the Pacific*, 131(1):018002, January 2019.

[52] Marina Berbel, Miquel Miravet-Tenés, Sushant Sharma Chaudhary, Simone Albanesi, Marco Cavaglià, Lorena Magaña Zertuche, Dimitra Tseneklidou, Yanyan Zheng, Michael W. Coughlin, and Andrew Toivonen. Bayesian real-time classification of multi-messenger electromagnetic and gravitational-wave observations. *Class. Quant. Grav.*, 41(8):085012, 2024. doi: 10.1088/1361-6382/ad3279.

[53] C. P. L. Berry, I. Mandel, H. Middleton, et al. Parameter estimation for binary neutron-star coalescences with realistic noise during the Advanced LIGO era. *Astrophys. J.*, 804(2):114, 2015. doi: 10.1088/0004-637X/804/2/114.

[54] S. Bloemen, P. Groot, G. Nelemans, and M. Klein-Wolt. The BlackGEM Array: Searching for Gravitational Wave Source Counterparts to Study Ultra-Compact Binaries. In S. M. Rucinski, G. Torres, and M. Zejda, editors, *Living Together: Planets, Host Stars and Binaries*, volume 496 of *Astronomical Society of the Pacific Conference Series*, page 254, July 2015.

[55] Sukanta Bose, Archana Pai, and Sanjeev V. Dhurandhar. Detection of gravitational waves from inspiraling compact binaries using a network of interferometric detectors. *Int. J. Mod. Phys. D*, 9:325–329, 2000. doi: 10.1142/S0218271800000360.

[56] M. Bulla. POSSIS: predicting spectra, light curves, and polarization for multidimensional models of supernovae and kilonovae. *Monthly Notices of the Royal Astronomical Society*, 489(4):5037–5045, Nov 2019. doi: 10.1093/mnras/stz2495.

[57] Mattia Bulla. The critical role of nuclear heating rates, thermalization efficiencies, and opacities for kilonova modelling and parameter inference. *Mon. Not. Roy. Astron. Soc.*, 520(2):2558–2570, 2023. doi: 10.1093/mnras/stad232.

[58] Mattia Bulla, Michael W. Coughlin, Suhail Dhawan, and Tim Dietrich. Multi-Messenger Constraints on the Hubble Constant through Combination of Gravitational Waves, Gamma-Ray Bursts and Kilonovae from Neutron Star Mergers. *Universe*, 8(5):289, 2022. doi: 10.3390/universe8050289.

[59] Miriam Cabero, Ashish Mahabal, and Jess McIver. GWSkyNet: a real-time classifier for public gravitational-wave candidates. *Astrophys. J. Lett.*, 904 (1):L9, 2020. doi: 10.3847/2041-8213/abc5b5.

[60] T. A. Callister et al. Observing gravitational waves with a single detector. *Classical and Quantum Gravity*, 2017. doi: 10.1088/1361-6382/aa7a76.

[61] Priscilla Canizares, Scott E. Field, Jonathan Gair, Vivien Raymond, Rory Smith, and Manuel Tiglio. Accelerated gravitational-wave parameter estimation with reduced order modeling. *Phys. Rev. Lett.*, 114(7):071104, 2015. doi: 10.1103/PhysRevLett.114.071104.

[62] Kipp Cannon, Chad Hanna, and Jacob Peoples. Likelihood-Ratio Ranking Statistic for Compact Binary Coalescence Candidates with Rate Estimation, 4 2015.

[63] Kipp Cannon et al. Toward Early-Warning Detection of Gravitational Waves from Compact Binary Coalescence. *Astrophys. J.*, 748:136, 2012. doi: 10.1088/0004-637X/748/2/136.

[64] Celery contributors. Celery: Distributed task queue. URL `https://docs.celeryq.dev/`.

[65] E. Chabanat, P. Bonche, P. Haensel, J. Meyer, and R. Schaeffer. A Skyrme parametrization from subnuclear to neutron star densitiesPart II. Nuclei far from stabilities. *Nuclear Physics A*, 635(1-2):231–256, May 1998. doi: 10.1016/S0375-9474(98)00180-8.

[66] Man Leong Chan et al. GWSkyNet. II. A Refined Machine-learning Pipeline for Real-time Classification of Public Gravitational Wave Alerts. *Astrophys. J.*, 972(1):50, 2024. doi: 10.3847/1538-4357/ad496a.

[67] Deep Chatterjee, Shaon Ghosh, Patrick R. Brady, Shasvath J. Kapadia, Andrew L. Miller, Samaya Nissanke, and Francesco Pannarale. A Machine Learning Based Source Property Inference for Compact Binary Mergers. *Astrophys. J.*, 896(1):54, 2020. doi: 10.3847/1538-4357/ab8dbe.

[68] Deep Chatterjee et al. Rapid likelihood free inference of compact binary coalescences using accelerated hardware. *Mach. Learn. Sci. Tech.*, 5(4): 045030, 2024. doi: 10.1088/2632-2153/ad8982.

[69] Sushant Sharma Chaudhary, Andrew Toivonen, et al. Low-latency gravitational wave alert products and their performance at the time of the

fourth LIGO-Virgo-KAGRA observing run. *Proc. Nat. Acad. Sci.*, 121(18): e2316474121, 2024. doi: 10.1073/pnas.2316474121.

[70] Min-A Cho. *Low-Latency Searches for Gravitational Waves and Their Electromagnetic Counterparts with Advanced LIGO and Virgo*. PhD thesis, University of Maryland, 2019.

[71] Chornock et al. The electromagnetic counterpart of the binary neutron star merger ligo/virgo gw170817. iv. detection of near-infrared signatures of r -process nucleosynthesis with gemini-south. *The Astrophysical Journal Letters*, 848(2):L19, 2017. URL http://stacks.iop.org/2041-8205/848/i=2/a=L19.

[72] Nelson Christensen. Stochastic Gravitational Wave Backgrounds. *Rept. Prog. Phys.*, 82(1):016903, 2019. doi: 10.1088/1361-6633/aae6b5.

[73] I M Christie, A Lalakos, A Tchekhovskoy, R Ferná ndez, F Foucart, E Quataert, and D Kasen. The role of magnetic field geometry in the evolution of neutron star merger accretion discs. *Monthly Notices of the Royal Astronomical Society*, 490(4):4811–4825, sep 2019. doi: 10.1093/mnras/stz2552. URL https://doi.org/10.1093%2Fmnras%2Fstz2552.

[74] Qi Chu, Manoj Kovalam, Linqing Wen, Teresa Slaven-Blair, Joel Bosveld, Yanbei Chen, Patrick Clearwater, Alex Codoreanu, Zhihui Du, Xiangyu Guo, Xiaoyang Guo, Kyungmin Kim, Tjonnie G. F. Li, Victor Oloworaran, Fiona Panther, Jade Powell, Anand S. Sengupta, Karl Wette, and Xingjiang Zhu. Spiir online coherent pipeline to search for gravitational waves from compact binary coalescences. *Phys. Rev. D*, 105:024023, Jan 2022. doi: 10.1103/PhysRevD.105.024023. URL https://link.aps.org/doi/10.1103/PhysRevD.105.024023.

[75] Qi Chu et al. SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences. *Phys. Rev. D*, 105(2):024023, 2022. doi: 10.1103/PhysRevD.105.024023.

[76] Alberto Colombo, Om Sharan Salafia, Francesco Gabrielli, Giancarlo Ghirlanda, Bruno Giacomazzo, Albino Perego, and Monica Colpi. Multi-messenger Observations of Binary Neutron Star Mergers in the O4 Run. *Astrophys. J.*, 937(2):79, 2022. doi: 10.3847/1538-4357/ac8d00.

[77] Neil J Cornish and Tyson B Littenberg. Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches. *Classical and Quan-*

tum Gravity, 32(13):135012, 2015. URL http://stacks.iop.org/0264-9381/32/i=13/a=135012.

[78] Michael Coughlin, Tim Dietrich, Kyohei Kawaguchi, Stephen Smartt, Christopher Stubbs, and Maximiliano Ujevic. Toward Rapid Transient Identification and Characterization of Kilonovae. *Astrophys. J.*, 849(1):12, 2017. doi: 10.3847/1538-4357/aa9114.

[79] Michael W Coughlin, Tim Dietrich, Zoheyr Doctor, Daniel Kasen, Scott Coughlin, Anders Jerkstrand, Giorgos Leloudas, Owen McBrien, Brian D Metzger, Richard O'Shaughnessy, and Stephen J Smartt. Constraints on the neutron star equation of state from at2017gfo using radiative transfer simulations. *Monthly Notices of the Royal Astronomical Society*, 480(3): 3871–3878, 2018. doi: 10.1093/mnras/sty2174. URL http://dx.doi.org/10.1093/mnras/sty2174.

[80] Michael W Coughlin, Tim Dietrich, Sarah Antier, Mattia Bulla, Francois Foucart, Kenta Hotokezaka, Geert Raaijmakers, Tanja Hinderer, and Samaya Nissanke. Implications of the search for optical counterparts during the first six months of the Advanced LIGO's and Advanced Virgo's third observing run: possible limits on the ejecta mass and binary properties. *Monthly Notices of the Royal Astronomical Society*, 492(1):863–876, 12 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz3457. URL https://doi.org/10.1093/mnras/stz3457.

[81] Michael W Coughlin, Tim Dietrich, Ben Margalit, and Brian D Metzger. Multimessenger Bayesian parameter inference of a binary neutron star merger. *Monthly Notices of the Royal Astronomical Society: Letters*, 489(1):L91–L96, 08 2019. ISSN 1745-3925. doi: 10.1093/mnrasl/slz133. URL https://doi.org/10.1093/mnrasl/slz133.

[82] Michael W. Coughlin, Sarah Antier, Tim Dietrich, Ryan J. Foley, Jack Heinzel, Mattia Bulla, Nelson Christensen, David A. Coulter, Lina Issa, and Nandita Khetan. Measuring the Hubble Constant with a sample of kilonovae. *Nature Commun.*, 11(1):4129, 2020. doi: 10.1038/s41467-020-17998-5.

[83] Michael W. Coughlin, Tim Dietrich, Jack Heinzel, Nandita Khetan, Sarah Antier, Mattia Bulla, Nelson Christensen, David A. Coulter, and Ryan J. Foley. Standardizing kilonovae and their use as standard candles to measure the Hubble constant. *Phys. Rev. Res.*, 2(2):022006, 2020. doi: 10.1103/PhysRevResearch.2.022006.

[84] Michael W. Coughlin, Amanda Farah, Emily, Abigail, Leo P. Singer, and R. Weizmann. LIGO/Virgo/KAGRA Observing Capabilities: Simulated Detections and Localization for O4 and O5 (October 2022 edition), October 2022. URL https://doi.org/10.5281/zenodo.7026209.

[85] Michael W. Coughlin et al. Constraints on the neutron star equation of state from AT2017gfo using radiative transfer simulations. *Mon. Not. Roy. Astron. Soc.*, 480(3):3871–3878, 2018. doi: 10.1093/mnras/sty2174.

[86] Coulter et al. Swope supernova survey 2017a (sss17a), the optical counterpart to a gravitational wave source. *Science*, 358(6370):1556–1558, 2017. doi: 10.1126/science.aap9811.

[87] P. S. Cowperthwaite, E. Berger, V. A. Villar, et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. II. UV, Optical, and Near-infrared Light Curves and Comparison to Kilonova Models. *The Astrophysical Journal Letters*, 848:L17, October 2017. doi: 10.3847/2041-8213/aa8fc7.

[88] Paolo Creminelli and Filippo Vernizzi. Dark Energy after GW170817 and GRB170817A. *Phys. Rev. Lett.*, 119(25):251302, 2017. doi: 10.1103/PhysRevLett.119.251302.

[89] Tito Dal Canton, Alexander H. Nitz, Bhooshan Gadre, Gareth S. Cabourn Davies, Veronica Villa-Ortega, Thomas Dent, Ian Harry, and Liting Xiao. Real-time Search for Compact Binary Mergers in Advanced LIGO and Virgo's Third Observing Run Using PyCBC Live. *Astrophys. J.*, 923(2):254, 2021. doi: 10.3847/1538-4357/ac2f9a.

[90] Maximilian Dax, Stephen R. Green, Jonathan Gair, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Real-Time Gravitational Wave Science with Neural Posterior Estimation. *Phys. Rev. Lett.*, 127(24):241103, 2021. doi: 10.1103/PhysRevLett.127.241103.

[91] Thomas Dent. Pycbc live p_{astro} for o4. Technical Report LIGO-T2300168, https://dcc.ligo.org/T2300168/public, 2023.

[92] T. Dietrich and M. Ujevic. Modeling dynamical ejecta from binary neutron star mergers and implications for electromagnetic counterparts. *Class. Quant. Grav.*, 34(10):105014, 2017. doi: 10.1088/1361-6382/aa6bb0.

[93] Tim Dietrich, Michael W. Coughlin, Peter T. H. Pang, Mattia Bulla, Jack Heinzel, Lina Issa, Ingo Tews, and Sarah Antier. New constraints on the

supranuclear equation of state and the hubble constant from nuclear physics – multi-messenger astronomy. *Science*, 370:1450–1453, 2020. doi: 10.1126/science.abb4317.

[94] Tim Dietrich, Michael W. Coughlin, Peter T. H. Pang, Mattia Bulla, Jack Heinzel, Lina Issa, Ingo Tews, and Sarah Antier. Multimessenger constraints on the neutron-star equation of state and the hubble constant. *Science*, 370 (6523):1450–1453, dec 2020. doi: 10.1126/science.abb4317. URL https://doi.org/10.1126%2Fscience.abb4317.

[95] K. L. Dooley et al. GEO 600 and the GEO-HF upgrade program: successes and challenges. *Class. Quant. Grav.*, 33:075009, 2016. doi: 10.1088/0264-9381/33/7/075009.

[96] Marco Drago, Sergey Klimenko, Claudia Lazzaro, Edoardo Milotti, Guenakh Mitselmakher, Valentin Necula, Brendan O'Brian, Giovanni Andrea Prodi, Francesco Salemi, Marek Szczepanczyk, et al. Coherent waveburst, a pipeline for unmodeled gravitational-wave data analysis. *SoftwareX*, 14:100678, 2021.

[97] Thomas D. P. Edwards, Kaze W. K. Wong, Kelvin K. H. Lam, Adam Coogan, Daniel Foreman-Mackey, Maximiliano Isi, and Aaron Zimmerman. Differentiable and hardware-accelerated waveforms for gravitational wave data analysis. *Phys. Rev. D*, 110(6):064028, 2024. doi: 10.1103/PhysRevD.110.064028.

[98] Albert Einstein. Zur Allgemeinen Relativitätstheorie. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1915:778–786, 1915. [Addendum: Sitzungsber.Preuss.Akad.Wiss.Berlin (Math.Phys.) 1915, 799–801 (1915)].

[99] Albert Einstein. Approximative Integration of the Field Equations of Gravitation. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1916:688–696, 1916.

[100] Albert Einstein. The foundation of the general theory of relativity. *Annalen Phys.*, 49(7):769–822, 1916. doi: 10.1002/andp.19163540702.

[101] Reed Essick, Salvatore Vitale, Erik Katsavounidis, Gabriele Vedovato, and Sergey Klimenko. Localization of short duration gravitational-wave transients with the early advanced ligo and virgo detectors. *The Astrophysical Journal*, 800(2):81, 2015. URL http://stacks.iop.org/0004-637X/800/i=2/a=81.

[102] Matthew Evans et al. A Horizon Study for Cosmic Explorer: Science, Observatories, and Community. 9 2021.

[103] Becca Ewing et al. Performance of the low-latency GstLAL inspiral search towards LIGO, Virgo, and KAGRA's fourth observing run. "", 5 2023.

[104] Jose María Ezquiaga and Miguel Zumalacárregui. Dark Energy After GW170817: Dead Ends and the Road Ahead. *Phys. Rev. Lett.*, 119(25): 251304, 2017. doi: 10.1103/PhysRevLett.119.251304.

[105] S. Fairhurst. Triangulation of gravitational wave sources with a network of detectors. *New J. Phys.*, 11(12):123006, December 2009. doi: 10.1088/1367-2630/11/12/123006.

[106] S. Fairhurst. Source localization with an advanced gravitational wave detector network. *Class. Quant. Grav.*, 28:105021, 2011. doi: 10.1088/0264-9381/28/10/105021.

[107] Will M. Farr, Jonathan R. Gair, Ilya Mandel, and Curt Cutler. Counting and confusion: Bayesian rate estimation with multiple populations. *Phys. Rev. D*, 91:023005, Jan 2015. doi: 10.1103/PhysRevD.91.023005. URL https://link.aps.org/doi/10.1103/PhysRevD.91.023005.

[108] Stephen M. Feeney, Hiranya V. Peiris, Samaya M. Nissanke, and Daniel J. Mortlock. Prospects for Measuring the Hubble Constant with Neutron-Star–Black-Hole Mergers. *Phys. Rev. Lett.*, 126(17):171102, 2021. doi: 10.1103/PhysRevLett.126.171102.

[109] Rodrigo Fernández, Daniel Kasen, Brian D. Metzger, and Eliot Quataert. Outflows from accretion discs formed in neutron star mergers: effect of black hole spin. *Monthly Notices of the Royal Astronomical Society*, 446(1):750–758, nov 2014. doi: 10.1093/mnras/stu2112. URL https://doi.org/10.1093%2Fmnras%2Fstu2112.

[110] Rodrigo Fernández, Alexander Tchekhovskoy, Eliot Quataert, Francois Foucart, and Daniel Kasen. Long-term GRMHD simulations of neutron star merger accretion discs: implications for electromagnetic counterparts. *Monthly Notices of the Royal Astronomical Society*, 482(3):3373–3393, oct 2018. doi: 10.1093/mnras/sty2932. URL https://doi.org/10.1093%2Fmnras%2Fsty2932.

[111] P. Fernique, T. Boch, T. Donaldson, D. Durand, W. O'Mullane, M. Reinecke, and M. Taylor. MOC - HEALPix Multi-Order Coverage map Version 1.0. IVOA Recommendation 02 June 2014, June 2014.

[112] Heather Kin Yee Fong. From simulations to signals: Analyzing gravitational waves from compact binary coalescences, 2018.

[113] Francois Foucart, M. Brett Deaton, Matthew D. Duez, Lawrence E. Kidder, Ilana MacDonald, et al. Black hole-neutron star mergers at realistic mass ratios: Equation of state and spin orientation effects. *Phys.Rev.*, D87:084006, 2013. doi: 10.1103/PhysRevD.87.084006.

[114] Francois Foucart, Tanja Hinderer, and Samaya Nissanke. Remnant baryon mass in neutron star-black hole mergers: Predictions for binary neutron star mimickers and rapidly spinning black holes. *Phys. Rev. D*, 98(8):081501, 2018. doi: 10.1103/PhysRevD.98.081501.

[115] Marylou Gabrié, Grant M. Rotskoff, and Eric Vanden-Eijnden. Adaptive Monte Carlo augmented with normalizing flows. *Proc. Nat. Acad. Sci.*, 119 (10):e2109420119, 2022. doi: 10.1073/pnas.2109420119.

[116] Shaon Ghosh, Xiaoshu Liu, Jolien Creighton, Ignacio Magaña Hernandez, Wolfgang Kastaun, and Geraint Pratten. Rapid model comparison of equations of state from gravitational wave observation of binary neutron star coalescences. *Physical Review D*, 104(8), oct 2021. doi: 10.1103/physrevd.104.083003. URL `https://doi.org/10.1103/physrevd.104.083003`.

[117] Tathagata Ghosh, Bhaskar Biswas, and Sukanta Bose. Simultaneous inference of neutron star equation of state and the Hubble constant with a population of merging neutron stars. *Phys. Rev. D*, 106(12):123529, 2022. doi: 10.1103/PhysRevD.106.123529.

[118] A. Goldstein et al. An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A. *Astrophys. J. Lett.*, 848(2):L14, 2017. doi: 10.3847/2041-8213/aa8f41.

[119] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, 622:759–771, April 2005. doi: 10.1086/427976.

[120] Matthew J. Graham, S. R. Kulkarni, Eric C. Bellm, Scott M. Adams, Cristina Barbarino, Nadejda Blagorodnova, Dennis Bodewits, Bryce Bolin, Patrick R. Brady, S. Bradley Cenko, Chan-Kao Chang, Michael W. Coughlin, Kishalay De, Gwendolyn Eadie, Tony L. Farnham, Ulrich Feindt, Anna Franckowiak, Christoffer Fremling, Suvi Gezari, Shaon Ghosh, Daniel A. Goldstein, V. Zach Golkhou, Ariel Goobar, Anna Y. Q. Ho, Daniela Huppenkothen,

Željko Ivezić, R. Lynne Jones, Mario Juric, David L. Kaplan, Mansi M. Kasliwal, Michael S. P. Kelley, Thomas Kupfer, Chien-De Lee, Hsing Wen Lin, Ragnhild Lunnan, Ashish A. Mahabal, Adam A. Miller, Chow-Choong Ngeow, Peter Nugent, Eran O. Ofek, Thomas A. Prince, Ludwig Rauch, Jan van Roestel, Steve Schulze, Leo P. Singer, Jesper Sollerman, Francesco Taddia, Lin Yan, Quan-Zhi Ye, Po-Chieh Yu, Tom Barlow, James Bauer, Ron Beck, Justin Belicki, Rahul Biswas, Valery Brinnel, Tim Brooke, Brian Bue, Mattia Bulla, Rick Burruss, Andrew Connolly, John Cromer, Virginia Cunningham, Richard Dekany, Alex Delacroix, Vandana Desai, Dmitry A. Duev, Michael Feeney, David Flynn, Sara Frederick, Avishay Gal-Yam, Matteo Giomi, Steven Groom, Eugean Hacopians, David Hale, George Helou, John Henning, David Hover, Lynne A. Hillenbrand, Justin Howell, Tiara Hung, David Imel, Wing-Huen Ip, Edward Jackson, Shai Kaspi, Stephen Kaye, Marek Kowalski, Emily Kramer, Michael Kuhn, Walter Landry, Russ R. Laher, Peter Mao, Frank J. Masci, Serge Monkewitz, Patrick Murphy, Jakob Nordin, Maria T. Patterson, Bryan Penprase, Michael Porter, Umaa Rebbapragada, Dan Reiley, Reed Riddle, Mickael Rigault, Hector Rodriguez, Ben Rusholme, Jakob van Santen, David L. Shupe, Roger M. Smith, Maayane T. Soumagnac, Robert Stein, Jason Surace, Paula Szkody, Scott Terek, Angela Van Sistine, Sjoert van Velzen, W. Thomas Vestrand, Richard Walters, Charlotte Ward, Chaoran Zhang, and Jeffry Zolkower. The zwicky transient facility: Science objectives. *Publications of the Astronomical Society of the Pacific*, 131(1001):078001, may 2019. doi: 10.1088/1538-3873/ab006c.

[121] Matthew J. Graham et al. A Light in the Dark: Searching for Electromagnetic Counterparts to Black Hole–Black Hole Mergers in LIGO/Virgo O3 with the Zwicky Transient Facility. *Astrophys. J.*, 942(2):99, 2023. doi: 10.3847/1538-4357/aca480.

[122] P. J. Groot, S. Bloemen, P. M. Vreeswijk, J. C. J. van Roestel, P. G. Jonker, G. Nelemans, M. Klein-Wolt, R. Lepoole, D. L. A. Pieterse, M. Rodenhuis, W. Boland, M. Haverkorn, C. Aerts, R. Bakker, H. Balster, M. Bekema, E. Dijkstra, P. Dolron, E. Elswijk, A. van Elteren, A. Engels, M. Fokker, M. de Haan, F. Hahn, R. ter Horst, D. Lesman, J. Kragt, J. Morren, H. Nillissen, W. Pessemier, G. Raskin, A. de Rijke, L. H. A. Scheers, M. Schuil, S. T. Timmer, L. Antunes Amaral, E. Arancibia-Rojas, I. Arcavi, N. Blagorodnova, S. Biswas, R. P. Breton, H. Dawson, P. Dayal, S. De Wet, C. Duffy, S. Faris, M. Fausnaugh, A. Gal-Yam, S. Geier, A. Horesh, C. Johnston, G. Katusiime, C. Kelley, A. Kosakowski, T. Kupfer, G. Leloudas, A. Levan, D. Modiano, O. Mogawana, J. Munday, J. Paice, F. Patat, I. Pelisoli, G. Ramsay, P. T.

Ranaivomanana, R. Ruiz-Carmona, V. Schaffenroth, S. Scaringi, F. Stoppa, R. Street, H. Tranin, M. Uzundag, S. Valenti, M. Veresvarska, M. Vučković, H. C. I. Wichern, R. A. M. J. Wijers, R. A. D. Wijnands, and E. Zimmerman. The BlackGEM Telescope Array. I. Overview. *Publications of the Astronomical Society of the Pacific*, 136(11):115003, November 2024. doi: 10.1088/1538-3873/ad8b6a.

[123] K. Grover, S. Fairhurst, B. F. Farr, et al. Comparison of Gravitational Wave Detector Network Sky Localization Approximations. *Phys. Rev.*, D89(4): 042004, 2014. doi: 10.1103/PhysRevD.89.042004.

[124] Mark Hannam, Patricia Schmidt, Alejandro Bohé, Leïla Haegel, Sascha Husa, Frank Ohme, Geraint Pratten, and Michael Pürrer. Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms. *Phys. Rev. Lett.*, 113 (15):151101, 2014. doi: 10.1103/PhysRevLett.113.151101.

[125] I. W. Harry and S. Fairhurst. Targeted coherent search for gravitational waves from compact binary coalescences. *Physical Review D*, 83(8), apr 2011. doi: 10.1103/physrevd.83.084002. URL `https://doi.org/10.1103/physrevd.83.084002`.

[126] J. Heinzel, M. W. Coughlin, T. Dietrich, M. Bulla, S. Antier, N. Christensen, D. A. Coulter, R. J. Foley, L. Issa, and N. Khetan. Comparing inclination dependent analyses of kilonova transients. *Mon. Not. Roy. Astron. Soc.*, 502 (2):3057–3065, 2021. doi: 10.1093/mnras/stab221.

[127] Shaun Hooper, Shin Kee Chung, Jing Luan, David Blair, Yanbei Chen, and Linqing Wen. Summed Parallel Infinite Impulse Response (SPIIR) Filters For Low-Latency Gravitational Wave Detection. *Phys. Rev. D*, 86:024012, 2012. doi: 10.1103/PhysRevD.86.024012.

[128] Kenta Hotokezaka, Ehud Nakar, Ore Gottlieb, Samaya Nissanke, Kento Masuda, Gregg Hallinan, Kunal P. Mooley, and Adam. T. Deller. A Hubble constant measurement from superluminal motion of the jet in GW170817. *Nature Astron.*, 3(10):940–944, 2019. doi: 10.1038/s41550-019-0820-1.

[129] R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *The Astrophysical Journal Letters*, 195:L51–L53, January 1975. doi: 10.1086/181708.

[130] Sascha Husa, Sebastian Khan, Mark Hannam, Michael Pürrer, Frank Ohme, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms

and anatomy of the signal. *Phys. Rev. D*, 93(4):044006, 2016. doi: 10.1103/PhysRevD.93.044006.

[131] S. Huth et al. Constraining Neutron-Star Matter with Microscopic and Macroscopic Collisions. *Nature*, 606:276–280, 2022. doi: 10.1038/s41586-022-04750-w.

[132] Željko Ivezić et al. LSST: from Science Drivers to Reference Design and Anticipated Data Products. *Astrophys. J.*, 873(2):111, 2019. doi: 10.3847/1538-4357/ab042c.

[133] Zhi-Ping Jin, Stefano Covino, Neng-Hui Liao, Xiang Li, Paolo D'Avanzo, Yi-Zhong Fan, and Da-Ming Wei. A kilonova associated with GRB 070809. *Nature Astron.*, 4:77–82, 2020.

[134] Shasvath J. Kapadia et al. A self-consistent method to estimate the rate of compact binary coalescences with a Poisson mixture model. *Class. Quant. Grav.*, 37(4):045007, 2020. doi: 10.1088/1361-6382/ab5f2d.

[135] Daniel Kasen, Brian Metzger, Jennifer Barnes, Eliot Quataert, and Enrico Ramirez-Ruiz. Origin of the heavy elements in binary neutron-star mergers from a gravitational wave event. *Nature*, 2017. doi: 10.1038/nature24453. [Nature551,80(2017)].

[136] Mansi M Kasliwal, Daniel Kasen, Ryan M Lau, Daniel A Perley, Stephan Rosswog, Eran O Ofek, Kenta Hotokezaka, Ranga-Ram Chary, Jesper Sollerman, Ariel Goobar, and David L Kaplan. Spitzer Mid-Infrared Detections of Neutron Star Merger GW170817 Suggests Synthesis of the Heaviest Elements. *Monthly Notices of the Royal Astronomical Society: Letters*, 01 2019. ISSN 1745-3925. doi: 10.1093/mnrasl/slz007. URL `https://doi.org/10.1093/mnrasl/slz007`.

[137] Kasliwal et al. Illuminating gravitational waves: A concordant picture of photons from a neutron star merger. *Science*, 358(6370):1559–1565, 2017. ISSN 0036-8075. doi: 10.1126/science.aap9455. URL `http://science.sciencemag.org/content/358/6370/1559`.

[138] Sebastian Khan, Sascha Husa, Mark Hannam, Frank Ohme, Michael Pürrer, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D*, 93(4):044007, 2016. doi: 10.1103/PhysRevD.93.044007.

[139] R. Weizmann Kiendrebeogo et al. Updated Observing Scenarios and Multi-messenger Implications for the International Gravitational-wave Networks O4 and O5. *Astrophys. J.*, 958(2):158, 2023. doi: 10.3847/1538-4357/acfcb1.

[140] Ray W. Klebesadel, Ian B. Strong, and Roy A. Olson. Observations of Gamma-Ray Bursts of Cosmic Origin. *Astrophys. J. Lett.*, 182:L85–L88, 1973. doi: 10.1086/181225.

[141] S. Klimenko, G. Vedovato, M. Drago, G. Mazzolo, G. Mitselmakher, C. Pankow, G. Prodi, V. Re, F. Salemi, and I. Yakushin. Localization of gravitational wave sources with networks of advanced detectors. *Physical Review D*, 83(10):102001, May 2011. doi: 10.1103/PhysRevD.83.102001.

[142] S. Klimenko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. A. Prodi, C. Lazzaro, K. Ackley, S. Tiwari, C. F. Da Silva, and G. Mitselmakher. Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors. *Phys. Rev. D*, 93:042004, Feb 2016. doi: 10.1103/PhysRevD.93.042004. URL `http://link.aps.org/doi/10.1103/PhysRevD.93.042004`.

[143] S. Klimenko et al. Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors. *Phys. Rev. D*, 93(4):042004, 2016. doi: 10.1103/PhysRevD.93.042004.

[144] Sergey Klimenko, I Yakushin, A Mercer, and Guenakh Mitselmakher. A coherent method for detection of gravitational wave bursts. *Classical and Quantum Gravity*, 25(11):114029, 2008.

[145] Ivan Kobyzev, Simon J. D. Prince, and Marcus A. Brubaker. Normalizing Flows: An Introduction and Review of Current Methods. *IEEE Trans. Pattern Anal. Machine Intell.*, 43(11):3964–3979, 2021. doi: 10.1109/tpami.2020.2992934.

[146] Manoj Kovalam, Md Anwarul Kaium Patwary, Anala K. Sreekumar, Linqing Wen, Fiona H. Panther, and Qi Chu. Early Warnings of Binary Neutron Star Coalescence Using the SPIIR Search. *Astrophys. J. Lett.*, 927(1):L9, 2022. doi: 10.3847/2041-8213/ac5687.

[147] Christian Jürgen Krüger and Francois Foucart. Estimates for Disk and Ejecta Masses Produced in Compact Binary Mergers. *Phys. Rev. D*, 101(10):103002, 2020. doi: 10.1103/PhysRevD.101.103002.

[148] S. R. Kulkarni et al. Science with the Ultraviolet Explorer (UVEX). 11 2021.

[149] Nina Kunert, Jonathan Gair, Peter T. H. Pang, and Tim Dietrich. Impact of gravitational waveform model systematics on the measurement of the Hubble constant. *Phys. Rev. D*, 110(4):043520, 2024. doi: 10.1103/PhysRevD.110.043520.

[150] Xiaoyu Lai, Enping Zhou, and Renxin Xu. Strangeons constitute bulk strong matter: Test using gw 170817. *The European Physical Journal A*, 55(4): 60, Apr 2019. ISSN 1434-601X. doi: 10.1140/epja/i2019-12720-8. URL https://doi.org/10.1140/epja/i2019-12720-8.

[151] J. M. Lattimer and D. N. Schramm. Black-hole-neutron-star collisions. *The Astrophysical Journal Letters*, 192:L145–L147, September 1974. doi: 10.1086/181612.

[152] Isaac Legred, Katerina Chatziioannou, Reed Essick, Sophia Han, and Philippe Landry. Impact of the PSR J0740+6620 radius constraint on the properties of high-density matter. *Phys. Rev. D*, 104(6):063003, 2021. doi: 10.1103/PhysRevD.104.063003.

[153] Isaac Legred, Katerina Chatziioannou, Reed Essick, Sophia Han, and Philippe Landry. Impact of the PSR J0740+6620 radius constraint on the properties of high-density matter: Neutron star equation of state posterior samples, April 2022. URL https://doi.org/10.5281/zenodo.6502467.

[154] Andrew J. Levan et al. Heavy-element production in a compact object merger observed by JWST. *Nature*, 626(8000):737–741, 2024. doi: 10.1038/s41586-023-06759-1.

[155] Li-Xin Li and Bohdan Paczynski. Transient events from neutron star mergers. *The Astrophysical Journal Letters*, 507(1):L59, 1998. URL http://stacks.iop.org/1538-4357/507/i=1/a=L59.

[156] Jing Luan, Shaun Hooper, Linqing Wen, and Yanbei Chen. Towards low-latency real-time detection of gravitational waves from compact binary coalescences in the era of advanced detectors. *Phys. Rev. D*, 85:102002, 2012. doi: 10.1103/PhysRevD.85.102002.

[157] LV. Low-latency gravitational-wave alerts for multimessenger astronomy during the second advanced LIGO and virgo observing run. *The Astrophysical Journal*, 875(2):161, apr 2019. doi: 10.3847/1538-4357/ab0e8f. URL https://doi.org/10.3847/1538-4357/ab0e8f.

[158] LVK. GW170817: Measurements of neutron star radii and equation of state. *Physical Review Letters*, 121(16), oct 2018. doi: 10.1103/physrevlett.121.161101. URL https://doi.org/10.1103\/physrevlett.121.161101.

[159] LVK. All-sky search for short gravitational-wave bursts in the third advanced LIGO and advanced virgo run. *Physical Review D*, 104(12), dec 2021. doi: 10.1103/physrevd.104.122004. URL https://doi.org/10.1103\/physrevd.104.122004.

[160] LVK. Gwtc-3: Compact binary coalescences observed by ligo and virgo during the second part of the third observing run, 2021.

[161] Ryan Lynch, Salvatore Vitale, Reed Essick, Erik Katsavounidis, and Florent Robinet. Information-theoretic approach to the gravitational-wave burst detection problem. *Physical Review D*, 95(10), may 2017. doi: 10.1103/physrevd.95.104046. URL https://doi.org/10.1103/physrevd.95.104046.

[162] Ryan Magee, Deep Chatterjee, Leo P. Singer, Surabhi Sachdev, Manoj Kovalam, Geoffrey Mo, Stuart Anderson, Patrick Brady, Patrick Brockill, Kipp Cannon, Tito Dal Canton, Qi Chu, Patrick Clearwater, Alex Codoreanu, Marco Drago, Patrick Godwin, Shaon Ghosh, Giuseppe Greco, Chad Hanna, Shasvath J. Kapadia, Erik Katsavounidis, Victor Oloworaran, Alexander E. Pace, Fiona Panther, Anwarul Patwary, Roberto De Pietri, Brandon Piotrzkowski, Tanner Prestegard, Luca Rei, Anala K. Sreekumar, Marek J. Szczepańczyk, Vinaya Valsan, Aaron Viets, Madeline Wade, Linqing Wen, and John Zweizig. First demonstration of early warning gravitational-wave alerts. *The Astrophysical Journal Letters*, 910(2):L21, apr 2021. doi: 10.3847/2041-8213/abed54. URL https://doi.org/10.3847\/2041-8213/abed54.

[163] Spencer J. Magnall, Simon R. Goode, Nikhil Sarin, and Paul D. Lasky. Directly inferring cosmology and the neutron-star equation of state from gravitational-wave mergers. 10 2024.

[164] B. Margalit and B. Metzger. Constraining the maximum mass of neutron stars from multi-messenger observations of gw170817. *The Astrophysical Journal Letters*, 850(2), 2017. doi: 10.3847/2041-8213/aa991c.

[165] Ben Margalit and Brian D. Metzger. The Multi-Messenger Matrix: the Future of Neutron Star Merger Constraints on the Nuclear Equation of State. *Astrophys. J. Lett.*, 880(1):L15, 2019. doi: 10.3847/2041-8213/ab2ae2.

[166] Raffaella Margutti et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/VIRGO GW170817. V. Rising X-ray Emission from an Off-Axis Jet. *Astrophys. J.*, 848(2):L20, 2017. doi: 10.3847/2041-8213/aa9057.

[167] Ethan Marx, William Benoit, Trevor Blodgett, Deep Chatterjee, Emma de Bruin, Steven Henderson, Katrine Kompanets, Siddharth Soni, Michael Coughlin, Philip Harris, and Erik Katsavounidis. Machine learning-enabled search for binary black hole mergers in LIGO-Virgo-KAGRA's third observing run. *Physical Review D*, 112(4):043007, August 2025. doi: 10.1103/1v7r-bkzs.

[168] Ethan Marx et al. Machine-learning pipeline for real-time detection of gravitational waves from compact binary coalescences. *Phys. Rev. D*, 111(4): 042010, 2025. doi: 10.1103/PhysRevD.111.042010.

[169] Duncan Meacher, Michael Coughlin, Sean Morris, Tania Regimbau, Nelson Christensen, Shivaraj Kandhasamy, Vuk Mandic, Joseph D. Romano, and Eric Thrane. Mock data and science challenge for detecting an astrophysical stochastic gravitational-wave background with Advanced LIGO and Advanced Virgo. *Phys. Rev. D*, 92(6):063002, 2015. doi: 10.1103/PhysRevD.92.063002.

[170] Cody Messick et al. Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data. *Phys. Rev. D*, 95(4):042001, 2017. doi: 10.1103/PhysRevD.95.042001.

[171] B. D. Metzger, G. Martínez-Pinedo, S. Darbha, E. Quataert, A. Arcones, D. Kasen, R. Thomas, P. Nugent, I. V. Panov, and N. T. Zinner. Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei. *Monthly Notices of the Royal Astronomical Society*, 406:2650–2662, August 2010. doi: 10.1111/j.1365-2966.2010.16864.x.

[172] Brian D. Metzger. Kilonovae. *Living Rev. Rel.*, 23(1):1, 2020. doi: 10.1007/s41114-019-0024-0.

[173] T. Mishra et al. Search for binary black hole mergers in the third observing run of Advanced LIGO-Virgo using coherent WaveBurst enhanced with machine learning. *Phys. Rev. D*, 105(8):083018, 2022. doi: 10.1103/PhysRevD.105.083018.

[174] J. S. Morgan, N. Kaiser, V. Moreau, D. Anderson, and W. Burgett. Design Differences between the Pan-STARRS PS1 and PS2 Telescopes. *Proc. SPIE Int. Soc. Opt. Eng.*, 8444:0H, 2012. doi: 10.1117/12.926646.

[175] Soichiro Morisaki and Vivien Raymond. Rapid Parameter Estimation of Gravitational Waves from Binary Neutron Star Coalescence using Focused Reduced Order Quadrature. *Phys. Rev. D*, 102(10):104020, 2020. doi: 10.1103/PhysRevD.102.104020.

[176] Soichiro Morisaki, Rory Smith, Leo Tsukada, Surabhi Sachdev, Simon Stevenson, Colm Talbot, and Aaron Zimmerman. Rapid localization and inference on compact binary coalescences with the Advanced LIGO-Virgo-KAGRA gravitational-wave detector network. *arXiv e-prints*, art. arXiv:2307.13380, July 2023. doi: 10.48550/arXiv.2307.13380.

[177] Elias R. Most, Lukas R. Weih, Luciano Rezzolla, and Jürgen Schaffner-Bielich. New constraints on radii and tidal deformabilities of neutron stars from gw170817. *Phys. Rev. Lett.*, 120:261103, Jun 2018. doi: 10.1103/PhysRevLett.120.261103. URL https://link.aps.org/doi/10.1103/PhysRevLett.120.261103.

[178] V. Necula, S. Klimenko, and G. Mitselmakher. Transient analysis with fast Wilson-Daubechies time-frequency transform. *J. Phys. Conf. Ser.*, 363:012032, 2012. doi: 10.1088/1742-6596/363/1/012032.

[179] Willie Neiswanger, Chong Wang, and Eric Xing. Asymptotically Exact, Embarrassingly Parallel MCMC. *arXiv e-prints*, art. arXiv:1311.4780, November 2013. doi: 10.48550/arXiv.1311.4780.

[180] Alexander H. Nitz, Thomas Dent, Tito Dal Canton, Stephen Fairhurst, and Duncan A. Brown. Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search. *Astrophys. J.*, 849(2):118, 2017. doi: 10.3847/1538-4357/aa8f50.

[181] Alexander H. Nitz, Tito Dal Canton, Derek Davis, and Steven Reyes. Rapid detection of gravitational waves from compact binary mergers with PyCBC Live. *Phys. Rev. D*, 98(2):024050, 2018. doi: 10.1103/PhysRevD.98.024050.

[182] Alexander H. Nitz, Marlin Schäfer, and Tito Dal Canton. Gravitational-wave Merger Forecasting: Scenarios for the Early Detection and Localization of Compact-binary Mergers with Ground-based Observatories. *Astrophys. J. Lett.*, 902:L29, 2020. doi: 10.3847/2041-8213/abbc10.

[183] A. Palmese et al. A statistical standard siren measurement of the Hubble constant from the LIGO/Virgo gravitational wave compact object merger GW190814 and Dark Energy Survey galaxies. *Astrophys. J. Lett.*, 900(2): L33, 2020. doi: 10.3847/2041-8213/abaeff.

[184] Peter T. H. Pang et al. An updated nuclear-physics and multi-messenger astrophysics framework for binary neutron star mergers. *Nature Commun.*, 14(1):8352, 2023. doi: 10.1038/s41467-023-43932-6.

[185] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing Flows for Probabilistic Modeling and Inference. *J. Machine Learning Res.*, 22(1):2617–2680, 2021. doi: 10.5555/3546258.3546315.

[186] Polina Petrov, Leo P. Singer, Michael W. Coughlin, Vishwesh Kumar, Mouza Almualla, Shreya Anand, Mattia Bulla, Tim Dietrich, Francois Foucart, and Nidhal Guessoum. Data-driven Expectations for Electromagnetic Counterpart Searches Based on LIGO/Virgo Public Alerts. *Astrophys. J.*, 924(2):54, 2022. doi: 10.3847/1538-4357/ac366d.

[187] Pian et al. Spectroscopic identification of r-process nucleosynthesis in a double neutron-star merger. *Nature*, 551:67 EP –, 10 2017. URL `http://dx.doi.org/10.1038/nature24298`.

[188] Brandon Joseph Piotrzkowski. *Searching for Gravitational Wave Associations with High-Energy Astrophysical Transients*. PhD thesis, University of Wisconsin-Milwaukee, 2022.

[189] Tsvi Piran, Ehud Nakar, and Stephan Rosswog. The electromagnetic signals of compact binary mergers. *Monthly Notices of the Royal Astronomical Society*, 430(3):2121–2136, 2013. doi: 10.1093/mnras/stt037. URL `http://dx.doi.org/10.1093/mnras/stt037`.

[190] Geraint Pratten et al. Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes. *Phys. Rev. D*, 103(10):104056, 2021. doi: 10.1103/PhysRevD.103.104056.

[191] Anna Puecher and Tim Dietrich. Machine-learning classifier for the post-merger remnant of binary neutron stars. *Phys. Rev. D*, 110(12):123038, 2024. doi: 10.1103/PhysRevD.110.123038.

[192] David Radice, Albino Perego, Francesco Zappa, and Sebastiano Bernuzzi. Gw170817: Joint constraint on the neutron star equation of state from multimessenger observations. *The Astrophysical Journal Letters*, 852(2):L29, 2018. URL http://stacks.iop.org/2041-8205/852/i=2/a=L29.

[193] Jillian C. Rastinejad et al. A kilonova following a long-duration gamma-ray burst at 350 Mpc. *Nature*, 612(7939):223–227, 2022. doi: 10.1038/s41586-022-05390-w.

[194] Nayyer Raza, Man Leong Chan, Daryl Haggard, Ashish Mahabal, Jess McIver, Audrey Durand, Alexandre Larouche, and Hadi Moazen. GWSkyNet-Multi II: an updated deep learning model for rapid classification of gravitational-wave events. 1 2025.

[195] Nayyer Raza et al. Explaining the gwskynet-multi machine learning classifier predictions for gravitational-wave events. *The Astrophysical Journal*, 963 (98), 2024. doi: 10.3847/1538-4357/ad13ea.

[196] David Reitze et al. Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. *Bull. Am. Astron. Soc.*, 51(7):035, 2019.

[197] Adam G. Riess et al. A 2.4% Determination of the Local Value of the Hubble Constant. *Astrophys. J.*, 826(1):56, 2016. doi: 10.3847/0004-637X/826/1/56.

[198] Adam G. Riess et al. A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km s^{-1} Mpc^{-1} Uncertainty from the Hubble Space Telescope and the SH0ES Team. *Astrophys. J. Lett.*, 934(1):L7, 2022. doi: 10.3847/2041-8213/ac5c5b.

[199] Florent Robinet, Nicolas Arnaud, Nicolas Leroy, Andrew Lundgren, Duncan Macleod, and Jessica McIver. Omicron: A tool to characterize transient noise in gravitational-wave detectors. *SoftwareX*, 12:100620, jul 2020. doi: 10.1016/j.softx.2020.100620. URL https://doi.org/10.1016/j.softx.2020.100620.

[200] I. M. Romero-Shaw et al. Bayesian inference for compact binary coalescences with bilby: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue. *Mon. Not. Roy. Astron. Soc.*, 499(3):3295–3319, 2020. doi: 10.1093/mnras/staa2850.

[201] S. Rosswog, U. Feindt, O. Korobkin, et al. Detectability of compact binary merger macronovae. *Class. Quant. Grav.*, 34(10):104001, 2017. doi: 10.1088/1361-6382/aa68a9.

[202] C. Röver, R. Meyer, G. M. Guidi, A. Viceré, and N. Christensen. Coherent Bayesian analysis of inspiral signals. *Classical and Quantum Gravity*, 24: S607–S615, October 2007. doi: 10.1088/0264-9381/24/19/S23.

[203] Soumen Roy, Anand S. Sengupta, and Nilay Thakor. Hybrid geometric-random template-placement algorithm for gravitational wave searches from compact binary coalescences. *Phys. Rev. D*, 95(10):104045, 2017. doi: 10.1103/PhysRevD.95.104045.

[204] Soumen Roy, Anand S. Sengupta, and Parameswaran Ajith. Effectual template banks for upcoming compact binary searches in Advanced-LIGO and Virgo data. *Phys. Rev. D*, 99(2):024048, 2019. doi: 10.1103/PhysRevD.99.024048.

[205] Surabhi Sachdev, Ryan Magee, Chad Hanna, Kipp Cannon, Leo Singer, Javed Rana SK, Debnandini Mukherjee, Sarah Caudill, Chiwai Chan, Jolien D. E. Creighton, Becca Ewing, Heather Fong, Patrick Godwin, Rachael Huxford, Shasvath Kapadia, Alvin K. Y. Li, Rico Ka Lok Lo, Duncan Meacher, Cody Messick, Siddharth R. Mohite, Atsushi Nishizawa, Hiroaki Ohta, Alexander Pace, Amit Reza, B. S. Sathyaprakash, Minori Shikauchi, Divya Singh, Leo Tsukada, Daichi Tsuna, Takuya Tsutsui, and Koh Ueno. An early-warning system for electromagnetic follow-up of gravitational-wave events. *The Astrophysical Journal Letters*, 905(2):L25, dec 2020. doi: 10.3847/2041-8213/abc753. URL https://doi.org/10.3847%2F2041-8213%2Fabc753.

[206] Shio Sakon et al. Template bank for compact binary mergers in the fourth observing run of Advanced LIGO, Advanced Virgo, and KAGRA, 11 2022.

[207] M. Saleem et al. The science case for LIGO-India. *Class. Quant. Grav.*, 39 (2):025004, 2022. doi: 10.1088/1361-6382/ac3b99.

[208] V. Savchenko et al. INTEGRAL Detection of the First Prompt Gamma-Ray Signal Coincident with the Gravitational-wave Event GW170817. *Astrophys. J. Lett.*, 848(2):L15, 2017. doi: 10.3847/2041-8213/aa8f94.

[209] Bernard F. Schutz. Determining the Hubble Constant from Gravitational Wave Observations. *Nature*, 323:310–311, 1986. doi: 10.1038/323310a0.

[210] Masaru Shibata and Kenta Hotokezaka. Merger and mass ejection of neutron star binaries. *Annual Review of Nuclear and Particle Science*, 69(1):41–64, 2019. doi: 10.1146/annurev-nucl-101918-023625. URL https://doi.org/10.1146/annurev-nucl-101918-023625.

[211] Y. Shvartzvald et al. ULTRASAT: A Wide-field Time-domain UV Space Telescope. *Astrophys. J.*, 964(1):74, 2024. doi: 10.3847/1538-4357/ad2704.

[212] T. Sidery, B. Aylott, N. Christensen, et al. Reconstructing the sky location of gravitational-wave detected compact binary systems: methodology for testing and comparison. *Phys. Rev.*, D89(8):084060, 2014. doi: 10.1103/PhysRevD.89.084060.

[213] Daniel M. Siegel and Brian D. Metzger. Three-dimensional GRMHD simulations of neutrino-cooled accretion disks from neutron star mergers. *The Astrophysical Journal*, 858(1):52, may 2018. doi: 10.3847/1538-4357/aabaec. URL https://doi.org/10.3847%2F1538-4357%2Faabaec.

[214] L. P. Singer, L. R. Price, B. Farr, et al. The First Two Years of Electromagnetic Follow-Up with Advanced LIGO and Virgo. *Astrophys. J.*, 795(2):105, 2014. doi: 10.1088/0004-637X/795/2/105.

[215] Leo Singer and Judith Racusin. General Coordinates Network (GCN): Nasa's Next Generation Time-Domain and Multimessenger Astronomy Alert System. *Bulletin of the AAS*, 55(2), jan 31 2023. https://baas.aas.org/pub/2023n2i108p02.

[216] Leo P. Singer and Larry R. Price. Rapid bayesian position reconstruction for gravitational-wave transients. *Phys. Rev. D*, 93:024013, Jan 2016. doi: 10.1103/PhysRevD.93.024013. URL https://link.aps.org/doi/10.1103/PhysRevD.93.024013.

[217] S. J. Smartt et al. A kilonova as the electromagnetic counterpart to a gravitational-wave source. *Nature, 10.1038/nature24303*, 2017. doi: 10.1038/nature24303.

[218] Rory Smith, Scott E. Field, Kent Blackburn, Carl-Johan Haster, Michael Pürrer, Vivien Raymond, and Patricia Schmidt. Fast and accurate inference on gravitational waves from precessing compact binaries. *Phys. Rev.*, D94 (4):044031, 2016. doi: 10.1103/PhysRevD.94.044031.

[219] Albert Sneppen, Darach Watson, Andreas Bauswein, Oliver Just, Rubina Kotak, Ehud Nakar, Dovi Poznanski, and Stuart Sim. Spherical symmetry in the kilonova AT2017gfo/GW170817. *Nature*, 614(7948):436–439, 2023. doi: 10.1038/s41586-022-05616-x.

[220] Joshua S. Speagle. dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. Roy. Astron. Soc.*, 493(3): 3132–3158, 2020. doi: 10.1093/mnras/staa278.

[221] Cosmin Stachie, Michael W Coughlin, Tim Dietrich, Sarah Antier, Mattia Bulla, Nelson Christensen, Reed Essick, Philippe Landry, Benoit Mours, Federico Schianchi, and Andrew Toivonen. Predicting electromagnetic counterparts using low-latency gravitational-wave data products. *Monthly Notices of the Royal Astronomical Society*, 505(3):4235–4248, jun 2021. doi: 10.1093/mnras/stab1492. URL https://doi.org/10.1093%2Fmnras%2Fstab1492.

[222] Marek J. Szczepańczyk et al. Search for gravitational-wave bursts in the third Advanced LIGO-Virgo run with coherent WaveBurst enhanced by machine learning. *Phys. Rev. D*, 107(6):062002, 2023. doi: 10.1103/PhysRevD.107.062002.

[223] Colm Talbot and Eric Thrane. Flexible and Accurate Evaluation of Gravitational-wave Malmquist Bias with Machine Learning. *Astrophys. J.*, 927(1):76, 2022. doi: 10.3847/1538-4357/ac4bc0.

[224] N. R. Tanvir, A. J. Levan, A. S. Fruchter, J. Hjorth, K. Wiersema, R. Tunnicliffe, and A. de Ugarte Postigo. A "kilonova" associated with short-duration gamma-ray burst 130603B. *Nature*, 500:547, 2013. doi: 10.1038/nature12505.

[225] Andrew Toivonen et al. What to Expect: Kilonova Light Curve Predictions via Equation of State Marginalization. *Publ. Astron. Soc. Pac.*, 137(3):034506, 2025. doi: 10.1088/1538-3873/adbcd7.

[226] J. L. Tonry, L. Denneau, A. N. Heinze, B. Stalder, K. W. Smith, S. J. Smartt, C. W. Stubbs, H. J. Weiland, and A. Rest. Atlas: A high-cadence all-sky survey system. *Publications of the Astronomical Society of the Pacific*, 130 (988):064505, 2018. URL http://stacks.iop.org/1538-3873/130/i=988/a=064505.

[227] E. Troja et al. A nearby long gamma-ray burst from a merger of compact objects. *Nature*, 612(7939):228–231, 2022. doi: 10.1038/s41586-022-05327-3.

[228] Leo Tsukada et al. Improved ranking statistics of the GstLAL inspiral search for compact binary coalescences. *Phys. Rev. D*, 108(4):043004, 2023. doi: 10.1103/PhysRevD.108.043004.

[229] Seiya Tsukamoto, Andrew Toivonen, Holton Griffin, Avyukt Raghuvanshi, Megan Averill, Frank Kerkow, Michael W. Coughlin, Man Leong Chan, and Leo Singer. Astrophysical or Terrestrial: Machine learning classification

of gravitational-wave candidates using multiple-search information. *arXiv, https://doi.org/10.48550/arXiv.2508.14242*, 8 2025.

[230] A. L. Urban. *Monsters in the Dark: High Energy Signatures of Black Hole Formation with Multimessenger Astronomy*. PhD thesis, University of Wisconsin-Milwaukee, 2016.

[231] Samantha A. Usman et al. The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quant. Grav.*, 33(21):215004, 2016. doi: 10.1088/0264-9381/33/21/215004.

[232] J. Veitch and A. Vecchio. Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network. *Phys. Rev. D*, 81:062003, 2010. doi: 10.1103/PhysRevD.81.062003.

[233] Verónica Villa-Ortega, Thomas Dent, and Andrés Curiel Barroso. Astrophysical source classification and distance estimation for pycbc live, 2022.

[234] Darach Watson et al. Identification of strontium in the merger of two neutron stars. *Nature*, 574(7779):497–500, 2019. doi: 10.1038/s41586-019-1676-3.

[235] L. Wen and Y. Chen. Geometrical Expression for the Angular Resolution of a Network of Gravitational-Wave Detectors. *Phys. Rev.*, D81:082001, 2010. doi: 10.1103/PhysRevD.81.082001.

[236] Linqing Wen. DATA ANALYSIS OF GRAVITATIONAL WAVES USING a NETWORK OF DETECTORS. *International Journal of Modern Physics D*, 17(07):1095–1104, jul 2008. doi: 10.1142/s0218271808012723. URL `https://doi.org/10.1142/s0218271808012723`.

[237] Kaze W. k. Wong, Marylou Gabrié, and Daniel Foreman-Mackey. flowMC: Normalizing flow enhanced sampling package for probabilistic inference in JAX. *J. Open Source Softw.*, 8(83):5021, 2023. doi: 10.21105/joss.05021.

[238] Kaze W. K. Wong, Maximiliano Isi, and Thomas D. P. Edwards. Fast Gravitational-wave Parameter Estimation without Compromises. *Astrophys. J.*, 958(2):129, 2023. doi: 10.3847/1538-4357/acf5cd.

[239] Thibeau Wouters, Peter T. H. Pang, Tim Dietrich, and Chris Van Den Broeck. Robust parameter estimation within minutes on gravitational wave signals from binary neutron star inspirals. *Phys. Rev. D*, 110(8):083033, 2024. doi: 10.1103/PhysRevD.110.083033.

[240] Xian-Feng Zhao. The properties of the massive neutron star PSR J0348+0432. *Int. J. Mod. Phys. D*, 24(08):1550058, 2015. doi: 10.1142/S0218271815500583.