

Robust Gravitational Wave Analysis at the Catalog Scale

Thesis by
Rhiannon Pollard Udall

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy in Physics

The Caltech logo is displayed in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended May 21st, 2025

© 2025

Rhiannon Pollard Udall
ORCID: 0000-0001-6877-3278

Some rights reserved. This thesis is distributed under a “Creative Commons
Attribution License”

ACKNOWLEDGEMENTS

This thesis is dedicated in honor of my Grammer, who passed away in my first year of graduate school. Her love, joy, and commitment to kindness and justice continues to inspire me, and I so dearly wish I could once again have a long conversation with her about anything and everything.

A great deal can happen in the space of five years, and many things can change, a fact I can attest to more than most. In that time, far too many people have given me far too much kindness to possibly enumerate it all. I will do my best, however, to acknowledge some of the special people who have helped me grow as a scientist and as a person.

When I started on the path towards being a scientist many years ago, it was with the help of many teachers throughout my early schooling, especially including Joel Blacher and Sergio Corvalan. Next were my undergraduate mentors, including Nikolaos Kidonakis, Deirdre Shoemaker, Laura Cadonati, James Clark, Karan Jani, Bhavesh Khamesra, and Richard O'Shaughnessy, who were patient and supportive even when I had no idea what I was doing. Many members of the LVK Collaboration have also given freely of their time and knowledge, including Gregory Ashton, Sharan Banagiri, Christopher Barry, Chad Hanna, Colm Talbot, and Daniel Williams. I also owe many thanks to my fellow graduate students and post-docs, with whom I have had so many edifying and entertaining conversations, including Jacob Golomb, Isaac Legred, Simona Miller, Sophie Hourihane, Ethan Payne, Alvin Li, Rico Lo, Virginia D'Emilio, Arianna Renzini, Lucy Thomas, Sophie Bini, Jane Glanzer, and Ryan Magee. I would especially like to thank Derek Davis, who saw me when I was adrift as a researcher and helped me find a path, and without whom much of this thesis could not exist. I similarly thank Jonah Kanner, whose has offered professional advice when I needed it, and kindness and perspective when I needed those even more. I am also thankful to Jess McIver, who has helped me expand my research horizons—and given me a job where I can do so!—and who serves as a model for guiding younger researchers with curiosity and compassion. I would also like to thank Katerina Chatziioannou, whose insights and advice have been immensely helpful in my growth as a researcher. Last but certainly not least, I am grateful to my advisor, Alan Weinstein, who has always encouraged my curiosity while offering grounding advice and lending an ear when I was struggling.

Separately, I also want to thank Sophie Hourihane, Simona Miller, Kyle Nelli, Emma Lenz, David Abramovich, Joshua Quinn Morgan, and all the other organizers of UAW 2478, who have worked tirelessly to make a more just Caltech for everyone.

Next comes my thanks to the many dear friends who have supported me and made my years in graduate school some of the happiest of my life. First I must thank Sydney, Lui, Katie, Greeny, Chris and Kate, who helped me make it through the strange and lonely covid year. I am grateful for the support, friendship, and joy offered by Ian, David, Emma, Elijah, Elina, Andy, and Yoonsoo, who have all helped to make my years here brighter. Sasha, an exemplar of dedication who nonetheless makes time to support his friends, has given me years of friendship. So to has Nicholas, whose friendship, advice, and willingness to push me to achieve greater things has aided me immensely over these years. Isaac, whose boundless curiosity and willingness to critically explore off-the-wall ideas are second to none, has also inspired me and offered me a great deal of friendship. Kyle, with his compassion, openness, and dedication to his friends, has consistently provided a helping hand when I am stressed and anxious. I would also like to thank Hannah, whose kindness, curiosity, and joy have inspired me, and whose friendship has been unwavering. I also must thank Ivy, who combines breadth of knowledge and critical insight with unflinching kindness, and who has been there for me when I need a shoulder to cry on. I wish also to thank Kiran, with their curiosity, determination, and exceptional compassion, as they have inspired and supported me. Finally, I want to thank Sarah, who has listened to my rants through many long hours and supported me unfailingly as I came to understand myself, while sharing many a bowl of soup with me—and with it, kindness and friendship—in the process. I could not have made it through the past five years without all of you, and I will treasure the time we have had together for the rest of my life, while looking forward to making many more memories in the future.

Lastly, I owe a great deal to my family, whose love has made everything I have accomplished possible. This includes my aunts Sue, Pam, Angie, and Betsy, my uncles Pete, James, and Jeff, and my cousins Tracy, Stanley, Jeffrey, Hannah Mae, and Paul, who have supported and grounded me from the very start. Next I wish to thank my Grammary, who has provided love and support when I needed them, and a push to be better when I needed that too. I owe so much to Peter, who has always been there when I needed him most, and whom I am so proud to call my brother. To my mom, who has given me so much love and so much support, a shoulder to

cry on and wisdom to help me live a better life, I am endlessly thankful and full of love for you. Finally, to my dad, who has given me unconditional love and who has been my biggest fan since my academic journey started, I could not have achieved this without you, and I love you fiercely. I love you all so much, and this thesis is for all of you.

I thank Alan Weinstein, Katerina Chatziioannou, David Reitze, Lee McCuller, and Gregg Hallinan for serving on my doctoral committee. I would also like to thank Jacob Golomb, Lucy Thomas, Derek Davis, David Reitze, and Alan Weinstein for helpful feedback on this thesis. This work was supported by National Science Foundation grant PHY-230920, and by the LIGO Laboratory at Caltech, of which I am a member. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the National Science Foundation, and operates under cooperative agreement PHY-2309200. I am grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459.

ABSTRACT

The rapid improvement in the sensitivity of ground based gravitational wave detectors has produced a huge variety of technical insights, but has also brought new challenges in gravitational wave data analysis. In this dissertation I address two of those challenges: the rapid increase in the number of detected events, and the need for robust astrophysical inferences in the presence of transient detector glitches. To manage the number of gravitational wave transients now regularly detected, I developed infrastructure for the LIGO-Virgo-KAGRA collaboration which monitors and collates the results of many disparate analyses in order to produce the final transient catalog. I implemented physically informed models for scattered light glitches into standard parameter estimation tools, and so that the potential realizations of these glitches can be marginalized over when performing astrophysical inference. This method was used to better understand GW191109, an event from the third observing run with potentially dynamical formation history. These tools were also applied to better understand the behavior of parameter estimation in the presence of glitches, and to search for statistical tests which can identify if parameter estimation is biased by the presence of a glitch.

LSC COLLABORATION MEMBER STATEMENT

This the work does not reflect the scientific opinion of the LSC and was not reviewed by the collaboration. Figure 9.1 is reproduced from an upcoming LIGO-Virgo-KAGRA collaboration paper, but does not represent the final, reviewed statement of the collaboration.

PUBLISHED CONTENT AND CONTRIBUTIONS

- ¹R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run”, *Phys. Rev. X* **13**, **I reviewed parameter estimation results from RIFT and bilby**, 041039 (2023) DOI:10.1103/PhysRevX.13.041039.
- ²R. Abbott et al. (LIGO Scientific, VIRGO), “GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run”, *Phys. Rev. D* **109**, **I reviewed parameter estimation results from RIFT and bilby**, 022001 (2024) DOI:10.1103/PhysRevD.109.022001.
- ³R. Abbott et al. (LIGO Scientific, VIRGO, KAGRA), “Search for intermediate-mass black hole binaries in the third observing run of Advanced LIGO and Advanced Virgo”, *Astron. Astrophys.* **659**, **I reviewed parameter estimation results from RIFT**, A84 (2022) DOI:10.1051/0004-6361/202141452.
- ⁴R. Abbott et al. (LIGO Scientific, Virgo), “GW190412: Observation of a Binary-Black-Hole Coalescence with Asymmetric Masses”, *Phys. Rev. D* **102**, **I analyzed the event with RIFT, providing one of the publication results**, 043015 (2020) DOI:10.1103/PhysRevD.102.043015.
- ⁵R. Abbott et al. (LIGO Scientific, Virgo), “Properties and Astrophysical Implications of the 150 M_{\odot} Binary Black Hole Merger GW190521”, *Astrophys. J. Lett.* **900**, **I analyzed the event with RIFT, providing one of the publication results**, L13 (2020) DOI:10.3847/2041-8213/aba493.
- ⁶R. Abbott et al. (LIGO Scientific, Virgo), “GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run”, *Phys. Rev. X* **11**, **I developed software for management of data from parameter estimation analyses**, 021053 (2021) DOI:10.1103/PhysRevX.11.021053.
- ⁷S. Soni et al. (LIGO), “LIGO Detector Characterization in the first half of the fourth Observing run”, *Class. Quant. Grav.* **42**, **I was responsible for the development of CBCFlow, used in event validation process**, 085016 (2025) DOI:10.1088/1361-6382/adc4b6.
- ⁸R. Udall, S. Hourihane, S. Miller, D. Davis, K. Chatziioannou, M. Isi, and H. Deshong, “Antialigned spin of GW191109: Glitch mitigation and its implications”, *Phys. Rev. D* **111**, *This article is included as a chapter in this thesis; I led the development and implementation of the study, including conceptualization, investigation, methodology, analysis, writing, and visualization*, 024046 (2025) DOI:10.1103/PhysRevD.111.024046.

- ⁹R. Magee, R. Sharma, A. Agrawal, and R. Udall, “Mitigating the impact of noise transients in gravitational-wave searches using reduced basis timeseries and convolutional neural networks”, **I created a test dataset for slow scattering glitches** (2024) DOI:10.48550/arXiv.2410.15513.
- ¹⁰R. Udall and D. Davis, “Bayesian modeling of scattered light in the LIGO interferometers”, Appl. Phys. Lett. **122**, *This article is included as a chapter in this thesis*; **I led the development and implementation of the study, including analysis, investigation, methodology, visualization, and writing**, 094103 (2023) DOI:10.1063/5.0136896.
- ¹¹D. Williams, J. Veitch, M. L. Chiofalo, P. Schmidt, R. P. Udall, A. Vajpeji, and C. Hoy, “Asimov: A framework for coordinating parameter estimation workflows”, J. Open Source Softw. **8**, **developed software for the RIFT interface**, 4170 (2023) DOI:10.21105/joss.04170.
- ¹²E. Payne, S. Hourihane, J. Golomb, R. Udall, D. Davis, and K. Chatziioannou, “Curious case of GW200129: Interplay between spin-precession inference and data-quality issues”, Phys. Rev. D **106**, **I participated in initial investigations during collaboration analysis**, 104017 (2022) DOI:10.1103/PhysRevD.106.104017.
- ¹³R. Udall, J. Brandt, G. Manchanda, A. Arulanandan, J. Clark, J. Lange, R. O’Shaughnessy, and L. Cadonati, “RUNMON-RIFT: Adaptive configuration and healing for large-scale parameter inference”, Astron. Comput. **42**, *This article is included as a chapter in this thesis*; **I led the development and implementation of the study, including conceptualization, methodology, software development, writing, and visualization**, 100664 (2022) DOI:10.1016/j.ascom.2022.100664.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	vi
LSC Collaboration Member Statement	vii
Published Content and Contributions	viii
Table of Contents	ix
List of Illustrations	xii
List of Tables	xv
I Astrophysics with Compact Binary Coalescences	1
Chapter I: Introduction	2
1.1 Overview of this Thesis' Contents	3
Chapter II: Background on Gravitational Wave Astrophysics	8
2.1 General Relativity and Gravitational Waves	8
2.2 Gravitational waves sourced by compact binary coalescences	10
2.3 The Astrophysical Population of Compact Object Binaries	13
2.4 Ground Based Gravitational Wave Detectors	14
2.5 Gravitational Wave Detector Glitches	29
Chapter III: Parameter Estimation for Gravitational Wave Transients	38
3.1 Bayesian inference in gravitational wave analysis	38
3.2 Parameter Estimation Pipelines	43
Chapter IV: Automatic Healing for RIFT Analyses	48
4.1 Introduction	48
4.2 Methods	50
4.3 Results	58
4.4 Conclusion	59
II Glitch Modeling	64
Chapter V: Glitch Modeling with Bilby	65
5.1 Glitch subtraction	65
5.2 Scattered light glitches	66
5.3 Glitch modeling with BILBY	71
Chapter VI: Bayesian Modeling for Slow Scattering	76
6.1 Introduction	76
6.2 Scattering Model	78
6.3 Bayesian analysis	79
6.4 Injection tests and single arches	83
6.5 Applications to real data	85

6.6	Conclusions	86
Chapter VII: The Anti-Aligned Spin of GW191109		93
7.1	Introduction	93
7.2	Modeling signals and glitches	96
7.3	Understanding the GW191109 Inference	103
7.4	Glitch-Marginalized Inference	109
7.5	Conclusions	116
7.6	Scattered light glitches in LHO	117
7.7	Detailed analysis settings	118
7.8	IMRPHENOMXPHM Analyses with BILBY	121
7.9	Alternate Slow Scattering Glitch Priors	122
7.10	Frequency Bin χ^2 Test	122
Chapter VIII: Other Glitch Investigations		131
8.1	Glitch residuals	131
8.2	Effects of time and phase	133
8.3	Impact of glitches with varying SNR	138
8.4	The frequency-bin χ^2 test in parameter estimation	140
III Gravitational Wave Transient Catalogs and Conclusion		147
Chapter IX: LVK Data Management		148
9.1	Transient searches	148
9.2	Detector characterization	152
9.3	Parameter estimation	152
Chapter X: Technical Infrastructure of Gravitational Wave Transient Catalogs:		
The Fourth Observing Run and Future Outlook		155
10.1	CBCFLOW	155
10.2	GWTC_PIPE and the catalog data product	158
10.3	Outlook for future gravitational wave catalog infrastructure	162
Chapter XI: Conclusion		166
11.1	Future plans for gravitational wave catalog infrastructure	166
11.2	Future plans for gravitational wave detector characterization	167

LIST OF ILLUSTRATIONS

<i>Number</i>	<i>Page</i>
2.1 An illustration of the idealized mirror	16
2.2 An illustration of scalar wave propagation through space	17
2.3 An illustration of the idealized two mirror system	17
2.4 The frequency dependent power enhancement of an example Fabry- Pérot interferometer	19
2.5 A schematics of optical fields in a simple Michelson interferometer .	20
2.6 A schematic of the prototypical Michelson interferometer	21
2.7 A schematic of sidebands in a Michelson inteferometer	22
2.8 A schematic of the propagation of the carrier beam and the signal sidebands in one interferometer arm	24
2.9 Shot noise sensitivity limit for $P_0 = 1 W$ and $L = 1 km$	27
2.10 Sensitivity limits due to both shot noise and radiation pressure	27
2.11 Breakdown of noise sources in aLigo according to the GWINC model	28
2.12 The layout of an interferometer with a power recycling cavity added .	29
2.13 The layout of an interferometer with arm cavities added	30
2.14 The layout of an interferometer with a signal recycling cavity added .	31
2.15 Sideband amplitude for various tunings of the SRC	31
2.16 Commonly occurring glitch morphologies	32
4.1 An example of automated configuration adjustment in case of railing	55
4.2 Sample statistics for automated blocking of unwanted computing nodes.	57
5.1 A schematic of scattered light	67
5.2 Slow scattering example	70
5.3 Fast scattering example	71
6.1 A diagram of the test mass and reaction chain	78
6.2 An Example of Scattering	80
6.3 Subtraction of real and simulated slow scattering glitches	82
6.4 Inferred amplitudes for real and simulated slow scattering glitches . .	91
6.5 Subtraction of slow scattering glitches in the data around GW191109	92
7.1 Witness prediction of slow scattering in LLO	98
7.2 Tracing χ_{eff} over frequency and time	105
7.3 Shifts in χ_{eff} for simulated signals	108

7.4	Joint inference of CBC and slow scattering glitch model	110
7.5	Joint inference of CBC and wavelet glitch model	112
7.6	Joint inference of wavelet signal model and wavelet glitch model . . .	113
7.7	Comparison of glitch reconstructions in LLO for differing models . .	115
7.8	Witness prediction of slow scattering in LHO	118
7.9	Frequency bin χ^2 test results for GW191109	123
8.1	Residual SNR when subtracting point realizations of a slow scattering glitch, as inferred by PE.	132
8.2	Posteriors for a GW191109 like signal with moderate χ_{eff} and a wavelet injected at varying times and frequencies.	134
8.3	Posteriors for a GW191109 like signal with extreme χ_{eff} and with a wavelet injected at varying times and frequencies.	135
8.4	Posteriors for a GW200129 like signal with moderate χ_p and with a wavelet injected at varying times and frequencies.	136
8.5	Posteriors for a GW200129 like signal with extreme χ_p and a wavelet injected at varying times and frequencies.	137
8.6	Posteriors and JS divergences for runs which injected a GW191109- like configuration with moderate χ_{eff} alongside a glitch scaled to varying SNR.	139
8.7	Posteriors and JS divergences for runs which injected a GW191109- like configuration with extreme χ_{eff} alongside a glitch scaled to vary- ing SNR.	140
8.8	Posteriors and JS divergences for runs which injected a GW200129- like configuration with moderate χ_p alongside a glitch scaled to varying SNR.	141
8.9	Posteriors and JS divergences for runs which injected a GW200129- like configuration with extreme χ_p alongside a glitch scaled to varying SNR.	142
8.10	Example binning for a χ^2 test	143
8.11	χ^2 statistic distributions alongside χ_{eff} posteriors for a wavelet of varying SNR	145
8.12	χ^2 statistic distributions alongside q posteriors for a wavelet of vary- ing SNR	146
9.1	A high level rendering of the flow of data involved in the production of the GWTC-4 catalog. Reproduced from the upcoming GWTC-4 Methods paper [1].	149

- 10.1 A diagram of the process by which search information is compiled
into the catalog data product. 159
- 10.2 A diagram of the process by which PE information is compiled into
the catalog data product. 161

LIST OF TABLES

<i>Number</i>		<i>Page</i>
4.1	Examples of common errors	54
7.1	Settings and properties for GW191109 analyses	121

Part I

Astrophysics with Compact Binary Coalescences

Chapter 1

INTRODUCTION

The detection of gravitational waves (GWs) on September 14th, 2015 [1] was one of the most monumental scientific and engineering accomplishments in human history. Requiring precision equivalent to measuring the distance to Alpha Centauri to within a width of a human hair, this detection required the efforts of thousands of scientists and engineers over decades in order to be realized. In the decade since this discovery the field of GW astrophysics has blossomed, offering insights into everything from the lives and deaths of stars [2–5] to the nature of extremely dense matter [6, 7], while also searching for exotica such as gravitationally lensed GWs [8] and violations of the theory of general relativity [9]. All of this has been driven by the steady improvement in the design and operation of the LIGO and Virgo detectors.

Improvements in sensitivity have also brought technical challenges, which are now some of the most pressing problems facing the field. The number of sources detected has ballooned, with the LIGO-Virgo-KAGRA (LVK) collaboration making more than two hundred detections in the fourth observing run as of March 20th [10]. This growth has necessitated rapid advancement of the tools used to perform analyses. Furthermore, as the noise floor has lowered, new sources of transient terrestrial noise, known as “glitches”, have also been revealed [11]. Not only do these glitches mimic GWs, they may contaminate the observed signals of true GWs, corrupting the astrophysical conclusions which may be drawn from them.

In my graduate studies, I have sought to address these two challenges. To accommodate the growing number of GW sources, I have led the development of infrastructure which monitors the many disparate analyses required to fully understand each GW event, as well as tools which process those analyses to produce final data products for public consumption. These tools are critical components of the fourth observing run Gravitational Wave Transient Catalog (GWTC) project, and are foundational to the catalog papers which will be released over the next two years.

To improve our analyses in the presence of glitches, I have developed methods which allow the modeling of certain types of glitches within the framework of astrophysical analyses. These methods permit statistically robust conclusions to be drawn about

the astrophysical sources which generated the GW signals which are contaminated by these glitches.

1.1 Overview of this Thesis' Contents

This thesis is separated into three parts. Part I covers introductory material for GW astrophysics and parameter estimation (PE), as well as work I did on the automation of one PE pipeline. Part II covers my work on the mitigation of glitches—specifically scattered light glitches—as well as the way in which those glitches impact the inference of astrophysical source parameters for incident GWs. Finally, Part III covers the process of creating the LVK transient catalogs, and the work I have done to build infrastructure for that project.

In Chapter 2 I discuss some common background for GW astrophysics. This includes a light overview of the derivation of GWs within the linearized GR, as well as a brief discussion of GWs originating from CBCs, and the source populations of those CBCs. Furthermore, it includes a discussion of some of the technical facets of ground based GW detectors, and the origin of some of their noise sources. Finally, it introduces glitches, which will be the focus of Part II.

In Chapter 3 I introduce the details of PE, which is the primary tool of analysis I used throughout my work. This includes a discussion of the standard frequency domain likelihood, common analysis priors, and inference pipelines which are used. I also briefly introduce the three major waveform families, and discuss their relative merits from the perspective of an analyst.

Chapter 4 contains published work from [12] on the `RIFT_RUNMONITOR` package, which was automation infrastructure I developed to assist in performing PE. `RIFT` is a PE pipeline which had a frustrating propensity for routine technical failures¹, and so it is common for analysts using it to spend large amounts of time monitoring and manually resubmitting analysis jobs. `RIFT_RUNMONITOR` alleviates this burden by parsing output logs for common failure modes and automatically resubmitting jobs if their failures are caused by hardware mismatches instead of mis-configuration. As part of this process, it is able to modify the `HTCONDOR` [13] submission settings in order to avoid cluster nodes with hardware mis-matches in the future, and stores those configuration updates for use across all other jobs which it is managing. It also automatically identifies when analyses are “railing”—that is, when the system has

¹Anecdotally it still has this propensity. I no longer work with this pipeline precisely because I do not want to find out for myself.

support for source parameters outside of the configured priors—and re-configures the priors of the analysis in order to compensate.

Chapter 5 is also a background chapter, introducing the procedure for performing PE with glitches which is used throughout Part II. First I derive the parameterized models for slow and fast scattering, which enable inference with these models. I also show the procedure for incorporating a glitch model into the standard frequency domain likelihood, including modifications to enable distance marginalization—an effective computational accelerant—in the joint glitch-CBC configuration.

Chapter 6 contains published work from [14] which introduced the modeling of slow scattering in `BILBY`. In addition to the derivation of the slow scattering model, this also includes demonstrations that this is an appropriate model to describe these glitches as they appear in real data. Furthermore, we showed that the model is capable of modeling arches which are not visible in spectrograms. Since spectrograms are the primary point of reference for whether data is “clean,” this means the model is able to indicate whether data may be corrupted by a glitch to a greater degree than an event validator may be able to do by visual inspection alone. Finally, we applied this model to GW191109, a notable GW event affected by slow scattering which we went on to investigate more deeply in subsequent work.

Chapter 7 is that subsequent published work [15]. We investigate the nature of GW191109, which is most notable for strong indications of anti-aligned spins, a potential indicator of dynamical formation. We first expanded upon previous analyses to identify which time segments and frequency bands were responsible for the measurement of anti-aligned spins, eventually narrowing our focus to a range of 30 – 40 Hz approximately 0.06 – 0.08 seconds before the merger. We also performed injection studies to identify the likelihood that such a small segment of data would play such a significant role, concluding that it was somewhat unlikely but not impossible to have such significant deviations when a small frequency band was removed from analysis in one detector. Next, we performed joint glitch-CBC inference using two glitch models: the parameterized slow scattering model I have introduced into `BILBY`, and the `BAYESWAVE` wavelets model. We found that while both models find glitch power in the relevant region of the Livingston data, the astrophysical parameters were not substantially altered by the modeling of this glitch power. Meanwhile `BAYESWAVE` finds a roughly one-in-three chance that the astrophysical parameters were substantially affected by glitch power, and that they should instead include high *aligned* spins. We conclude by noting that the only glitch

power known to be present in Livingston was slow scattering power reasonably well described by the parameterized model—which would thus favor the anti-aligned spin scenario, and hence dynamical formation—but that the data are not conclusive.

The final chapter of Part II, Chapter 8, includes a series of other analyses into the impact of glitches on PE. The first section shows that even when standard glitch subtraction methods are applied, statistical uncertainties mean residual glitch power will always remain. The second and third sections investigate how glitches can impact PE, with a focus on the role that the time and phase of a glitch, as well as its SNR, can play in determining whether astrophysical posteriors will be biased, and if so in what direction. The final section investigates the application of a frequency-bin χ^2 —first proposed in [16]—to PE results, following on work in Appendix 7.10 of Udall et al [15]. In particular, this section demonstrates that this statistic responds to progressive biasing of astrophysical results, even for glitches which would normally be too quiet to detect.

In Chapter 9 I give an overview of the procedure by which GW catalogs are produced. This includes some information about each of the core analysis stages for a given event—detection, initial parameter estimation, detector characterization, and production parameter estimation—and how these are coupled together in a larger workflow. I also reference how these are incorporated into the technical infrastructure I have developed.

Chapter 10 details that technical infrastructure, which broadly includes two projects: `CBCFLOW` and `GWTC_PIPE`. The first is the internal database and monitoring system which connects different parts of the catalog infrastructure into a cohesive whole. This includes monitoring `GRACEDB` for newly detected GWs, scraping parameter estimation results, and serving as a repository of configuration recommendations from detector characterization experts. The second is the pipeline which generates the publicly released data products. To do this, `GWTC_PIPE` parses through `CBCFLOW` databases to determine which results are from production analyses, then collates their contents and produces summary information about them. This process has allowed for iterating on catalog data products internally before ultimately releasing their finalized versions to the public, substantially improving timelines for downstream analyses when compared to the progression of analysis in the third observing run.

Finally, in Chapter 11 I summarize the results of this thesis, and explain future directions I plan to follow. These include extensions on glitch modeling work I have done on ground based detectors, and the application of similar techniques to the

analysis of the space-based LISA mission. They also include my plans to overhaul the analysis infrastructure of the LVK in preparation for the fifth observing run, to increase integration, robustness, and user-friendliness.

References

- ¹B. P. Abbott et al. (LIGO Scientific, Virgo), “Observation of Gravitational Waves from a Binary Black Hole Merger”, *Phys. Rev. Lett.* **116**, 061102 (2016) DOI:10.1103/PhysRevLett.116.061102.
- ²B. P. Abbott et al. (LIGO Scientific, Virgo), “GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs”, *Phys. Rev. X* **9**, 031040 (2019) DOI:10.1103/PhysRevX.9.031040.
- ³R. Abbott et al. (LIGO Scientific, VIRGO), “GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run”, *Phys. Rev. D* **109**, 022001 (2024) DOI:10.1103/PhysRevD.109.022001.
- ⁴R. Abbott et al. (LIGO Scientific, Virgo), “GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run”, *Phys. Rev. X* **11**, 021053 (2021) DOI:10.1103/PhysRevX.11.021053.
- ⁵R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), “Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3”, *Phys. Rev. X* **13**, 011048 (2023) DOI:10.1103/PhysRevX.13.011048.
- ⁶B. P. Abbott et al. (LIGO Scientific, Virgo), “Properties of the binary neutron star merger GW170817”, *Phys. Rev. X* **9**, 011001 (2019) DOI:10.1103/PhysRevX.9.011001.
- ⁷B. P. Abbott et al. (LIGO Scientific, Virgo), “GW170817: Measurements of neutron star radii and equation of state”, *Phys. Rev. Lett.* **121**, 161101 (2018) DOI:10.1103/PhysRevLett.121.161101.
- ⁸R. Abbott et al. (LIGO Scientific, KAGRA, VIRGO), “Search for Gravitational-lensing Signatures in the Full Third Observing Run of the LIGO–Virgo Network”, *Astrophys. J.* **970**, 191 (2024) DOI:10.3847/1538-4357/ad3e83.
- ⁹R. Abbott et al. (LIGO Scientific, VIRGO, KAGRA), “Tests of General Relativity with GWTC-3”, (2021) DOI:10.48550/arXiv.2112.06861.
- ¹⁰K. Burtnyk (“LIGO Laboratory”), Ligo-virgo-kagra announce the 200th gravitational wave detection of Press Release, https://www.ligo.caltech.edu/news/ligo20250320_2025.
- ¹¹S. Soni et al. (LIGO), “LIGO Detector Characterization in the first half of the fourth Observing run”, *Class. Quant. Grav.* **42**, 085016 (2025) DOI:10.1088/1361-6382/adc4b6.

- ¹²R. Udall, J. Brandt, G. Manchanda, A. Arulanandan, J. Clark, J. Lange, R. O’Shaughnessy, and L. Cadonati, “RUNMON-RIFT: Adaptive configuration and healing for large-scale parameter inference”, *Astron. Comput.* **42**, 100664 (2022) DOI:10.1016/j.ascom.2022.100664.
- ¹³E. M. Fajardo, J. M. Dost, B. Holzman, et al., “How much higher can htcondor fly?”, *Journal of Physics: Conference Series* **664**, 062014 (2015).
- ¹⁴R. Udall and D. Davis, “Bayesian modeling of scattered light in the LIGO interferometers”, *Appl. Phys. Lett.* **122**, 094103 (2023) DOI:10.1063/5.0136896.
- ¹⁵R. Udall, S. Hourihane, S. Miller, D. Davis, K. Chatziioannou, M. Isi, and H. Deshong, “Antialigned spin of GW191109: Glitch mitigation and its implications”, *Phys. Rev. D* **111**, 024046 (2025) DOI:10.1103/PhysRevD.111.024046.
- ¹⁶B. Allen, “ χ^2 time-frequency discriminator for gravitational wave detection”, *Phys. Rev. D* **71**, 062001 (2005) DOI:10.1103/PhysRevD.71.062001.

Chapter 2

BACKGROUND ON GRAVITATIONAL WAVE ASTROPHYSICS

In this chapter, I will discuss many of the fundamental background concepts underlying this thesis. For the sake of conciseness I will not derive formulae in detail—in dedicated background chapters such as Chapter 3 and Chapter 5 I will go into more detail in relevant analysis methods—but will provide an overview of some important results. I will begin with basic results on the nature of GWs and their modeling. Next, I will describe some of the astrophysical populations which are observed with ground base GW detectors. I will then describe those detectors in brief, and end with an introduction to terrestrial noise transients known as “glitches” which will be considered in detail in Part II.

2.1 General Relativity and Gravitational Waves

While the contents of my thesis do not directly utilize GR, a discussion of GWs should begin with some background on the nature of GWs, and the conventions and notation which will be used throughout this thesis. In this section I will provide this background, following closely the results and notation from Carroll [1] for the first part and from Cutler and Flanagan [2] for the second part. This section and Section 2.2 will use geometrized units $G = c = 1$.

Gravitational radiation in general relativity

In regions with weak gravitational fields, the metric may be written as a perturbation on the flat Minkowski metric

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad (2.1)$$

where the weakness of the field means that $|h_{\mu\nu}| \ll 1$.¹ This perturbation may be decomposed [1] into time (with Φ and w_i being the Gravitational scalar and vector potential) and spatial components, with the spatial components further decomposed into traceful and traceless components:

$$h_{00} = -2\Phi \quad (2.2)$$

$$h_{0i} = w_i \quad (2.3)$$

$$h_{ij} = 2s_{ij} - 2\Psi\delta_{ij} \quad (2.4)$$

¹Carroll also discusses the finer points of what “weak” means in the context of diffeomorphisms.

where i and j run over the spatial indices. In this Ψ is traceful and s_{ij} traceless:

$$\Psi = -\frac{1}{6}\delta^{ij}h_{ij} \quad (2.5)$$

$$s_{ij} = \frac{1}{2}(h_{ij} - \frac{1}{3}\delta^{kl}h_{kl}\delta_{ij}). \quad (2.6)$$

These expressions may be written in many gauges, but for GWs the relevant one is the transverse gauge

$$\partial_i s^{ij} = 0, \quad (2.7)$$

in which Einstein's equations become [1]

$$G_{00} = 2\nabla^2\Psi = 8\pi GT_{00} \quad (2.8)$$

$$G_{0j} = -\frac{1}{2}\nabla^2 w_j + 2\partial_0\partial_j\Psi = 8\pi GT_{0j} \quad (2.9)$$

$$G_{ij} = (\delta_{ij}\nabla^2 - \partial_i\partial_j) - \partial_0\partial_{(i}w_{j)} + 2\delta_{ij}\partial_0^2\Psi - \square s_{ij} = 8\pi GT_{ij}. \quad (2.10)$$

Gravitational waves are freely-propagating in empty space, and hence GW solutions may be found by setting $T_{\mu\nu} = 0$ in the above equations, which implies $\Psi = w_j = \Phi = 0$. Only s_{ij} is non-zero, and it takes on a wave equation

$$\square s_{ij} = 0 \quad (2.11)$$

where from this point it is typical to choose a gauge such that all of the other degrees of freedom are set to 0, leaving the transverse traceless gauge

$$h_{\mu\nu}^{\text{TT}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & 2s_{ij} & \\ 0 & & & \end{pmatrix}. \quad (2.12)$$

Then taking the plane-wave solution of the wave equation gives

$$h_{\mu\nu}^{\text{TT}} = C_{\mu\nu}e^{ik_\sigma x^\sigma}. \quad (2.13)$$

It is convenient to choose k^μ in a single direction, which using the condition that $k^\mu k_\mu = 0$ leaves $k^\mu = (\omega, 0, 0, \omega)$. This then forces the components of $C_{\mu\nu}$ to be

$$C_{\mu\nu}^{\text{TT}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & C_{11} & C_{12} & 0 \\ 0 & C_{12} & -C_{11} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (2.14)$$

which means the wave is characterized by two numbers, the plus and cross polarizations

$$h_+ = C_{11} \quad (2.15)$$

$$h_\times = C_{12}. \quad (2.16)$$

For the trace reversed strain $\bar{h}_{ij} = h_{\mu\nu} - \frac{1}{2}h\eta_{\mu\nu}$ the generation of gravitational waves is given to the lowest (quadrupole) order by [1]

$$\bar{h}_{ij}(t, \mathbf{x}) = \frac{2G}{r} \frac{d^2 I_{ij}}{dt^2}(t_r) \quad (2.17)$$

for an observer at distance r , where I_{ij} is the traceless quadrupole moment tensor. This formula is the source of many of the most important points of intuition about GWs. They are sourced by changes in the *quadrupole* moment, as opposed to the dipole moments which source EM radiation. Furthermore, because we observe strain as opposed to power fluxes, our signals only scale with the inverse of the distance, as opposed to the inverse square. This makes GW observatories exceptional probes of transients at large distances. Finally, the prefactor—for the traceless mass quadrupole in de-geometrized units [3]—of $\frac{G}{c^4 r}$ is *very* small for astrophysical values of r , for example $2.7 \times 10^{-69} \text{J}^{-1}$ at 100Mpc, which means that the mass quadrupole must be very large and changing very fast in order to generate observable GW strains. This requirement limits the sources which will be observable, and for ground based GW detectors this has so far meant only the mergers of compact objects—black holes and neutron stars—have been observed [4]. In the next section the details of these objects are explored in more detail.

2.2 Gravitational waves sourced by compact binary coalescences

This thesis is concerned with GWs which are generated by compact binary coalescences (CBCs), and so in this section I will describe some of the basic physics of these systems and introduce conventions for parameters which are used subsequently. To begin, one should note that binary black hole (BBH) systems may be characterized intrinsically by the masses of the two black holes m_i and the spin vectors of each \vec{S}_i , yielding a total of eight intrinsic parameters. To these are added seven extrinsic parameters which characterize the system's location and orientation in space-time with respect to the observing detector. Neutron stars add to these parameters which describe the response of their constituent matter to tidal effects, parameterized as tidal deformabilities, but for the purpose of this thesis neutron stars will not feature any further.

For completeness, the extrinsic parameters of a CBC are the following [5]:

1. The luminosity distance d_L is the distance to the source, which appears in Equation 2.23.
2. The time of the coalescence t_c , also discussed in different detector frames (i.e. t_L) or in the geocenter frame t_{gc} .
3. The inclination of the source ι with respect to us as observers, which for precessing systems is usually parameterized by the angle between the total angular momentum and the observer θ_{JN} .
4. The polarization angle of the GW, ψ .
5. The phase of the system at coalescence ϕ_c . For ϕ_c and t_c , the time of coalescence is taken to be the maximum of the amplitude envelope for the gravitational wave.
6. The location on the sky in right ascension α and declination δ , which for analysis efficiency may be replaced by the azimuth ϵ and zenith κ in a detector based sky parameterization [5].

This simplest formulation of these GWs is the first post-Newtonian order derivation [2, 3]. Using the gravitational wave flux computed from the quadrupole formula above

$$\frac{dE_{orb}}{dt} = -\frac{32}{5}\mu^2 r^4 \Omega^6, \quad (2.18)$$

Kepler's law

$$\Omega^2 = \frac{M}{R^3}, \quad (2.19)$$

and the orbital energy of a quasi-circular Keplerian system

$$E_{orb} = -\frac{M\mu}{2r}, \quad (2.20)$$

one may compute under an adiabatic assumption the evolution of that system due to gravitational waves

$$r(t) = \left(\frac{256}{5}\mu M^2\right)^{1/4} (t_c - t)^{1/4}, \quad (2.21)$$

where $M = m_1 + m_2$ is the total mass of the system and $\mu = \frac{m_1 m_2}{M}$ is the reduced mass of the system. Then one finds that the signal frequency evolution is

$$\frac{df}{dt} = \frac{96}{5}(\pi)^{8/3} \mathcal{M}^{5/3} f^{11/3}, \quad (2.22)$$

where the $\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} = \eta^{3/5} M$ is the chirp mass and $\eta = \frac{m_1 m_2}{M^2}$ is the symmetric mass ratio. This $f^{11/3}$ evolution is what gives CBCs the ‘‘chirp’’ moniker.

Applying the stationary phase approximation, the strain in the frequency domain at first post-Newtonian order is [2]

$$\tilde{h}(f) = \frac{Q(\alpha, \delta, \psi, \iota)}{d_L} \mathcal{M}^{5/6} f^{-7/6} \exp[i\Psi(f)], \quad (2.23)$$

where $Q(\alpha, \delta, \psi, \iota)$ encodes the detector response and $\Psi(f)$ is the phase

$$\Psi(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{3}{4}(8\pi \mathcal{M} f)^{-5/3}. \quad (2.24)$$

For GWs with many cycles in the detection band, constraints of phase evolution are vastly better than those of the amplitude evolution. Accordingly, the prominent role chirp mass plays in Equation 2.24 means that it is by far the best measured parameter for such systems.

Higher post-Newtonian orders are computed by including higher multipole moments, adding correction to the amplitude of each multipole component, and correcting the orbital energy and gravitational wave flux for strong field GR effects. Given the flux and orbital energy, one may then solve for the frequency evolution, and in turn the phase evolution. For a post-Newtonian expansion parameter $x = (\pi \mathcal{M} f)^{-11/3}$, the 1.5 PN correction to $\Psi(f)$ is [2]

$$\Psi(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{3}{4}(8\pi \mathcal{M} f)^{-5/3} \times \left[1 + \frac{20}{9} \left(\frac{743}{336} + \frac{11\mu}{4M} \right) x + (4\beta - 16\pi) x^{3/2} \right], \quad (2.25)$$

where β incorporates the spins in the direction of the orbital angular momentum, chosen to be \hat{z} , and is defined as

$$\beta = M^{-2} \left[\left(\frac{113}{12} + \frac{25}{4} \frac{m_2}{m_1} \right) s_{1z} + \left(\frac{113}{12} + \frac{25}{4} \frac{m_1}{m_2} \right) s_{2z} \right], \quad (2.26)$$

where s_{1z} and s_{2z} are the component spins in the direction of the orbital angular momentum. Thus at the 1 PN order measurements of the constituent masses become possible, while at 1.5 PN measurements of the system spin becomes possible. While β is the best measured spin parameter, it is more typical to discuss mass weighted effective aligned spin

$$\chi_{\text{eff}} = \frac{s_{1z} + q s_{2z}}{1 + q}, \quad (2.27)$$

where $q = \frac{m_2}{m_1}$, which is more readily interpretable, and which is conserved at the 2 PN order [6].

In Chapter 3, I will discuss how we use the more sophisticated descendants of this model to infer the astrophysical properties of observed GW systems. These inferences inherit the mathematical structure of these waveform models, and hence features such as the preferred parameterizations noted above play a major role in our understanding of these systems.

2.3 The Astrophysical Population of Compact Object Binaries

As mentioned in Section 2.1, the only sources detected so far by ground based GW detectors are CBCs [4]. To be more precise, these are the mergers of CBCs with masses in the range $O(M/M_\odot) \sim 1 - 10^2$ (where M_\odot is a solar mass). The constituents of these CBCs are expected to be the products of stellar evolution—either directly or by hierarchical formation from previous mergers—though primordial black holes in this mass range are possible [7]. Accordingly, it is worth briefly discussing the formation channels for these mergers, to motivate the astrophysics we hope to accomplish with ground based detectors.

Black holes and neutron stars are generated by the collapse of sufficiently large stars [8], and so the properties and physics of these stars is intimately linked with the resulting objects. Furthermore, the rate of orbital decay due to gravitational waves is quite slow at larger separations, necessitating some process to bring the objects close enough that this orbital decay can produce a merger within a Hubble time [9]. Thus, the population of mergers we observe will not be the population of all black holes and neutron stars, but the sub-population that has been subject to such a process, and the nature of this process will imprint itself upon the observed system.

Tautologically, if two compact objects merge then they either were gravitationally bound to each other at the time that they experienced collapse and became compact objects, or they were not. This divides the possible formation channels into two broad categories: isolated formation channels in which a binary stellar system evolves into a binary system of compact objects which goes on to merge, or dynamical channels in which two compact objects that evolved separately are brought into close proximity by some dynamical process [9–11]. These have certain characteristics which may distinguish the systems they generate, and each also have distinct sub-channels which may encode further information into the system.

Isolated formation channels which generate GW source systems operate under two general constraints: they must have two constituents which collapse to form a compact object, and some process must bring those constituents close enough together that they will merge within a Hubble time. The first condition sets a theoretical upper limit on the mass of $M \approx 40 - 50M_{\odot}$ for black holes which may occur in such channels, since models of the pair-instability super nova (PISN) process predict partial or total disruption of larger stars [8, 12]. The second leads to a few distinct sub-channels which may bring the two objects close enough together: a common envelope phase [9, 11], stable mass-transfer [13, 14], and chemically homogeneous evolution [15, 16]. All produce masses in the range $\mathcal{M} \approx 5 - 40M_{\odot}$, though with some internal variation in probabilities. They also all produce spins with relatively low magnitude which are preferentially aligned with the orbital angular momentum, though for high kicks there may be some systems with significant anti-aligned spins [14].

Dynamical formation channels do not necessarily feature the first constraint described above, but still must resolve the second. Dynamical processes which bring two compact objects into close proximity are much more likely to occur in very dense environments such as dense stellar clusters [17–19] or AGN disks [20, 21]. The constituents in these interactions may be first generation—that is, formed by the collapse of a star—or they may be second or greater generation, the product of an earlier black hole or neutron star merger, and accordingly, their masses may supersede the theoretical PISN mass limit [10, 12]. Second generation or higher objects will also have higher spins than those generated by collapse, since they retain the orbital angular momentum of their progenitor system as spin. In dense stellar clusters, objects brought together will have isotropic spin distributions, and hence these systems may also have significant anti-aligned spins or precessing spins [10, 22]. Finally, dynamical formation channels can happen on short enough time scales to produce mergers with non-trivial eccentricity [23]. Accordingly, systems with high masses and high spins not in the direction of the orbital angular momentum, and/or with measurable orbital eccentricity, are expected to be the best candidates for clear differentiation of isolated and dynamical formation channels.

2.4 Ground Based Gravitational Wave Detectors

This section closely follows Bond et al [24], and is derived from notes I made for an Astrophysics-Relativity-Cosmology (ARC) seminar—a graduate student led seminar series at Caltech—on this subject [25].

In this section I will offer a brief overview of the operating principles of modern ground based GW detectors. This overview will have significant omissions, including both controls systems and mode cleaning, which are absolutely critical elements of the detector. However, attention will be given to the use of cavities to enhance the detector, the actual behavior of the detection scheme, and some of the fundamental noise sources in the detector.

The observable associated with GWs is the strain which was derived in Section 2.1. The two polarizations h_+ and h_\times will each interact with the detector to cause relative differences in the length of orthogonal arms. For Michelson interferometers, these combine as:

$$h = \frac{L_X - L_Y}{L} = F_+ h_+ + F_\times h_\times, \quad (2.28)$$

where L_X , L_Y are the lengths of respective arms, and L is the typical detector length. $L_X - L_Y$ is the differential arm length DARM, and is used for sensing; the opposite combination $L_X + L_Y$ is called CARM and used for controlling the detector. The detector response functions F_+ and F_\times are governed by the detector geometry and the location and orientation of the source relative to each detector.

Building Blocks: Cavities and Michelson Interferometers

In this section, I will derive in broad terms the properties of two of the core components of modern GW detectors: the Fabry-Perot Cavity and the Michelson Interferometer.

Fabry-Perot Cavities

It is a standard result of electromagnetism that light can be characterized by its electric field, which—ignoring polarization and assume propagation along the z-axis—gives

$$E = \mathcal{R}(E'_0 \exp(i(\omega t - kz))), \quad (2.29)$$

where E'_0 is a complex number that folds in the initial phase, and ω and k are the standard angular frequency and angular wavenumber. For this discussion the instantaneous field at $t = 0$ and the complex amplitudes are of interest, so

$$E = E_0 \exp(-ikz) \quad (2.30)$$

where E_0 is a real number. Next, it is necessary to consider how this light will behave when propagating through space or interacting with a mirror.

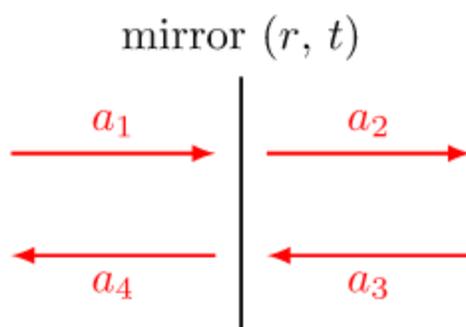


Figure 2.1: An illustration of the idealized mirror, reproduced from Bond et al [24].

Figure 2.1 shows schematically the interaction of light with mirrors, described by the system of equations:

$$a_2 = ita_1 + ra_3 \quad (2.31)$$

$$a_4 = ita_3 + ra_1. \quad (2.32)$$

Now, one may assume that the transmission introduces a phase shift of $\pi/2$ [24] and that these are lossless mirrors, so that $r^2 + t^2 = 1$. Rearranging the equations

$$a_1 = \frac{i}{t}(ra_3 - a_2) \quad (2.33)$$

$$a_4 = \frac{ri}{t}(ra_3 - a_2) + ita_3 = \frac{i}{t}(a_3 - ra_2), \quad (2.34)$$

this can be reframed this as a matrix equation:

$$\begin{pmatrix} a_1 \\ a_4 \end{pmatrix} = \frac{i}{t} \begin{pmatrix} -1 & r \\ -r & 1 \end{pmatrix} \begin{pmatrix} a_2 \\ a_3 \end{pmatrix}. \quad (2.35)$$

Figure 2.2 illustrates the case for propagation through space, corresponding to the system of equations

$$a_2 = a_1 \exp(-ikD) \quad (2.36)$$

$$a_4 = a_3 \exp(-ikD), \quad (2.37)$$

which has the matrix form

$$\begin{pmatrix} a_1 \\ a_4 \end{pmatrix} = \begin{pmatrix} \exp(ikD) & 0 \\ 0 & \exp(-ikD) \end{pmatrix} \begin{pmatrix} a_2 \\ a_3 \end{pmatrix}. \quad (2.38)$$

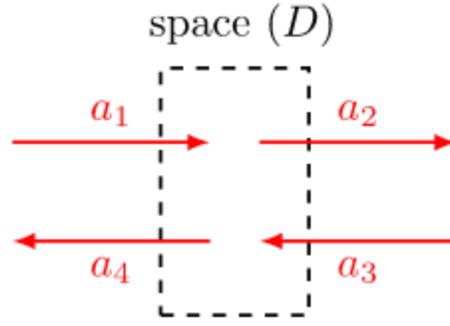


Figure 2.2: An illustration of scalar wave propagation through space, reproduced from Bond et al [24].

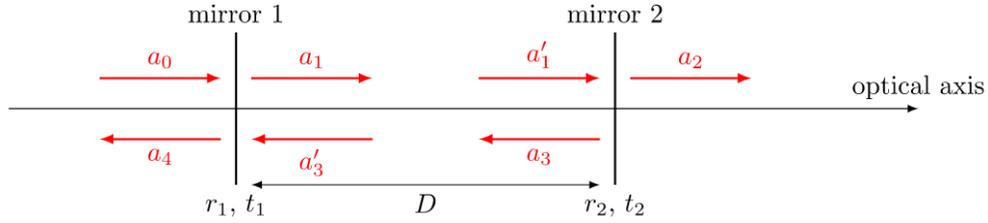


Figure 2.3: An illustration of the idealized two mirror system, reproduced from Bond et al [24].

Figure 2.3 illustrates the two mirror system, which combines the above cases. Here the second mirror has no field incoming from the right side, so $a_3 = 0$. Solving the matrix equation gives

$$\begin{pmatrix} a_0 \\ a_4 \end{pmatrix} = \frac{i}{t_1} \begin{pmatrix} -1 & r_1 \\ -r_1 & 1 \end{pmatrix} \begin{pmatrix} \exp(ikD) & 0 \\ 0 & \exp(-ikD) \end{pmatrix} \frac{i}{t_2} \begin{pmatrix} -1 & r_2 \\ -r_2 & 1 \end{pmatrix} \begin{pmatrix} a_2 \\ 0 \end{pmatrix}, \quad (2.39)$$

which reduces to

$$\begin{pmatrix} a_0 \\ a_4 \end{pmatrix} = \frac{-1}{t_1 t_2} \begin{pmatrix} \exp(ikD) - r_1 r_2 \exp(-ikD) & -r_2 \exp(ikD) + r_1 \exp(-ikD) \\ -r_2 \exp(ikD) + r_1 \exp(-ikD) & \exp(-ikD) - r_1 r_2 \exp(ikD) \end{pmatrix} \begin{pmatrix} a_2 \\ 0 \end{pmatrix}. \quad (2.40)$$

This leaves the expression for the ratio of the transmitted amplitude to the incoming amplitude:

$$\frac{a_2}{a_0} = \frac{-t_1 t_2 \exp(-ikD)}{1 - r_1 r_2 \exp(-i2kD)}. \quad (2.41)$$

The two mirror system is a Fabry-Pérot interferometer, one of the essential components of the modern GW detector. For a cavity of length L such that $kL = N\pi$ for

integer $N \in \mathbb{Z}$, the cavity is at a resonance. If instead kL is a half integer multiple of π , the cavity is at an anti-resonance. The frequency offset between peaks is called the free spectral range, or FSR, and may be found by

$$\text{FSR} = \frac{c}{2L}. \quad (2.42)$$

Similarly, the line width, or frequency width at half maximum (FWHM), is

$$\text{FWHM} = \frac{2\text{FSR}}{\pi} \arcsin\left(\frac{1 - r_1 r_2}{2\sqrt{r_1 r_2}}\right). \quad (2.43)$$

Together, these define the finesse of the cavity:

$$F = \frac{\text{FSR}}{\text{FWHM}} = \frac{\pi}{2 \arcsin\left(\frac{1 - r_1 r_2}{2\sqrt{r_1 r_2}}\right)}, \quad (2.44)$$

which for r_1, r_2 near 1 (a high finesse cavity) approximates to

$$F \approx \frac{\pi}{1 - r_1 r_2}. \quad (2.45)$$

As one may note, when a Fabry-Pérot cavity is on resonance, the circulating power in the cavity is much higher than the input power (see Figure 2.4). This method allows the “stepping up” of laser power, at the cost of having added degrees of freedom which must be controlled.

Michelson Interferometers

GW detectors are, at their most basic, Michelson interferometers, and so deriving basic properties of the simple case of a Michelson interferometer is necessary both to understand the system of GW detection, and the noise properties of the detector.

Figure 2.5 is a schematic of the fields in a simple Michelson interferometer. The expression for the most relevant field element is [24]

$$E_6 = E_0 r t \left[\exp\left(i(\phi_t + \phi_{r_1} + \Phi_1)\right) + \exp\left(i(\phi_t + \phi_{r_2} + \Phi_2)\right) \right]. \quad (2.46)$$

One may choose a convention of $\phi_t = \pi/2$, $\phi_{r_1} = \phi_{r_2} = 0$, equivalent to the convention of introducing a factor of i during transmission which was used above. Φ_1 and Φ_2 are the phase difference picked up along the Y and X arms respectively, so simplifying and combining the common and differential elements gives

$$E_6 = i r t E_0 \exp\left(i\frac{\Phi_1 + \Phi_2}{2}\right) 2 \cos\left(\frac{\Phi_1 - \Phi_2}{2}\right). \quad (2.47)$$

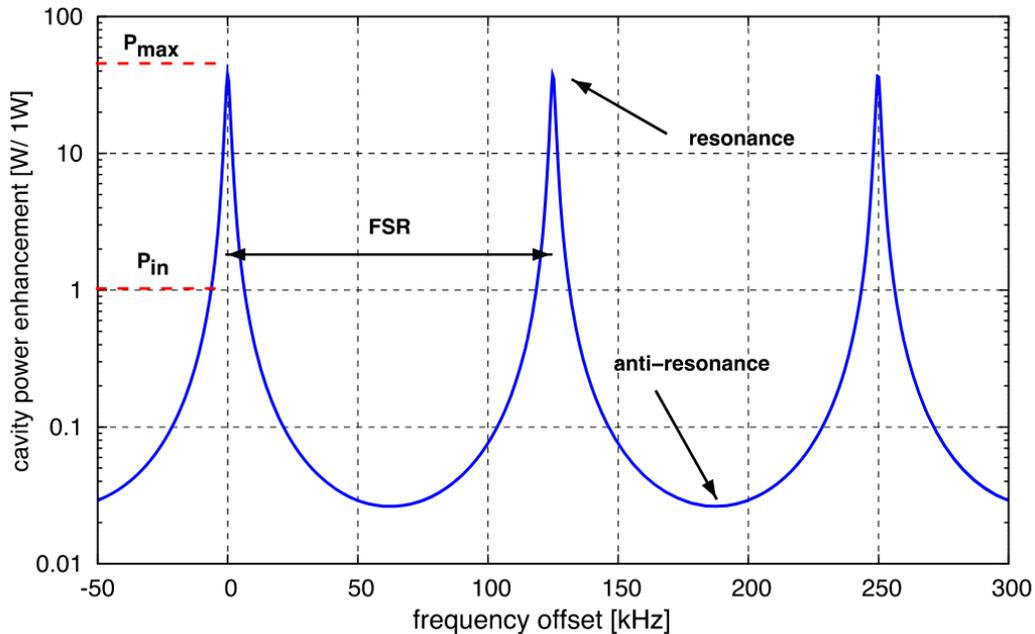


Figure 2.4: The frequency dependent power enhancement of an example Fabry-Pérot interferometer, reproduced from Bond et al [24].

Michelson interferometers measure differential arm length (DARM) by splitting light into two perpendicular beams, sending them along a set of arms, then recombining them. If the beams have travelled different lengths (modulo the light's wavelength) then they will pick up a relative phase, and hence destructively interfere with each other when recombined. For an idealized interferometer with a monochromatic laser and perfect 50:50 beam splitter (see Figure 2.6), the field E_6 at the detection photodiode (the anti-symmetric port) is

$$E_S = E_0 \frac{i}{2} \left(\exp(i2kL_Y) + \exp(i2kL_X) \right). \quad (2.48)$$

Defining common and differential arm lengths as

$$\bar{L} = \frac{L_Y + L_X}{2} \quad (2.49)$$

and

$$\Delta L = L_Y - L_X \quad (2.50)$$

respectively, and noting that

$$\Phi_i = 2kL_i, \quad (2.51)$$

this reduces to

$$E_S = E_0 i \exp(i2k\bar{L}) \cos(k\Delta L), \quad (2.52)$$

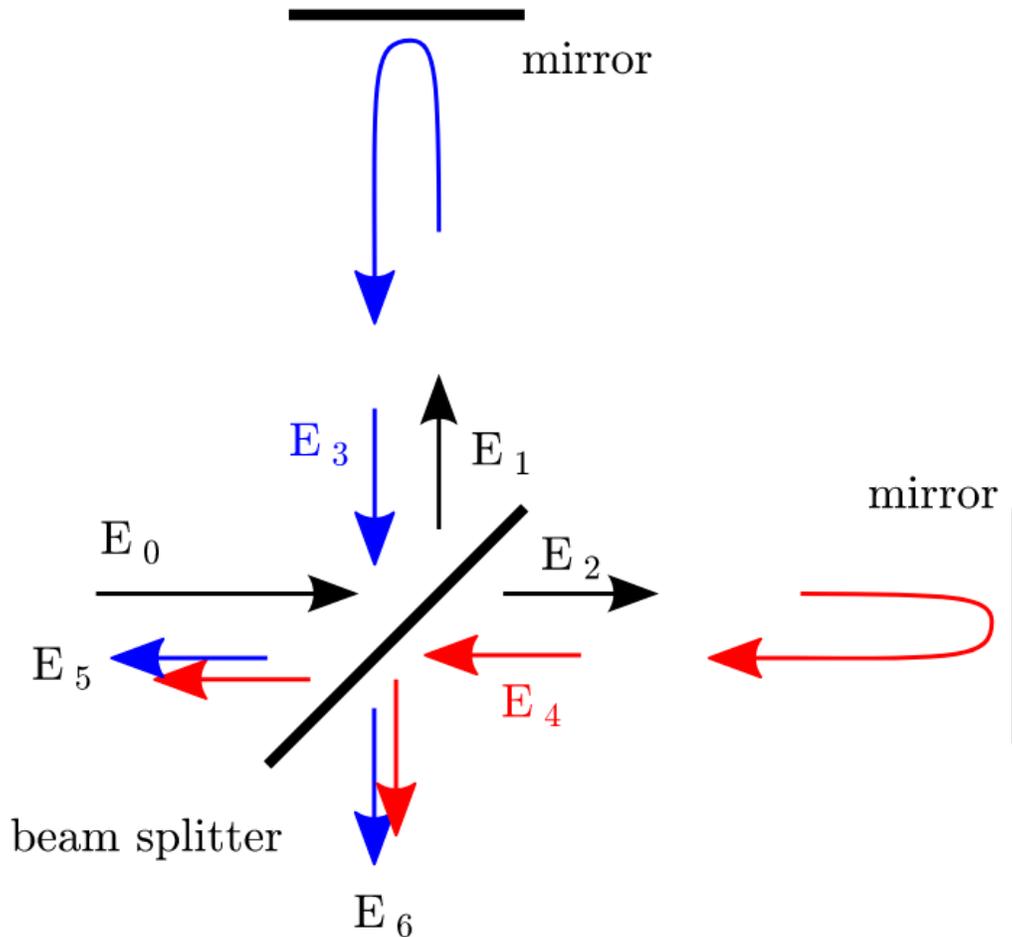


Figure 2.5: A schematics of optical fields in a simple Michelson interferometer, reproduced from Bond et al [24].

so that the intensity of the signal will be

$$P_0 \cos^2(2\pi\Delta L/\lambda). \quad (2.53)$$

When this is nearly 0 then the interferometer is on a dark fringe, and when it is nearly P_0 the interferometer is on a bright fringe. Modern interferometers are operated at or near the dark fringe, as will now be discussed.

Gravitational Wave Signal Detection

When a GW passes through the detector, the light in the arms is modulated in phase. This modulation produces sidebands, and these are the observational target of the various detection schemes which may be used.

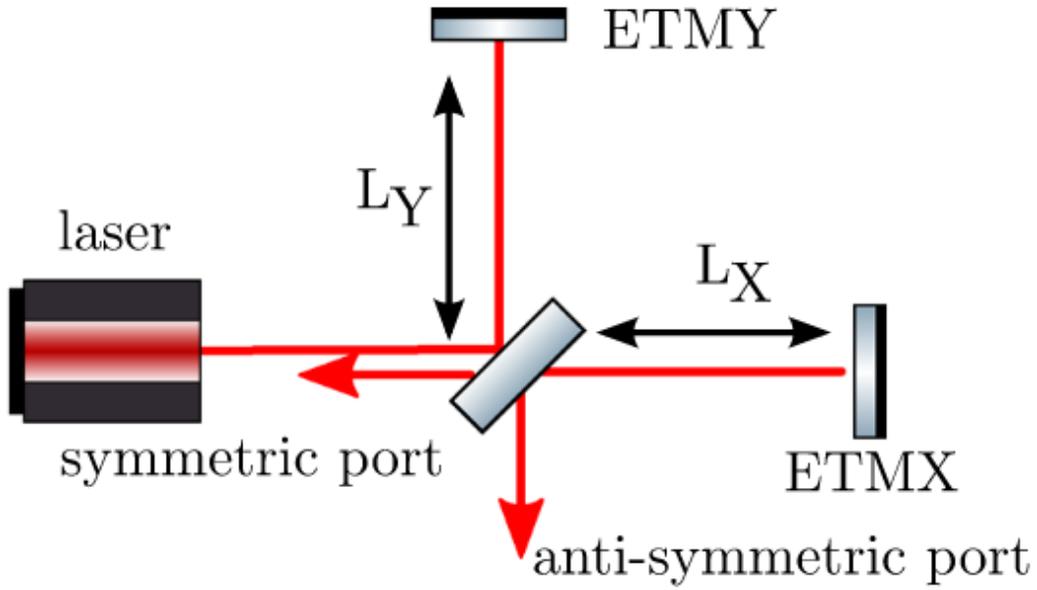


Figure 2.6: A schematic of the prototypical Michelson interferometer, reproduced from Bond et al [24].

Modulation and Sidebands

A phase modulated signal when a carrier field $E_{in} = E_0 \exp(i\omega_0 t)$ is modulated to

$$E = E_0 \exp\left(i(\omega_0 t + m \cos(\Omega t))\right), \quad (2.54)$$

where m is the amplitude of the modulating signal—in this case a GW—and ω is its angular frequency. Bessel functions of the first kind satisfy the identity

$$\exp(iz \cos(\phi)) = \sum_{k=-\infty}^{\infty} i^k J_k(z) \exp(ik\phi), \quad (2.55)$$

and for small modulation indices phase modulation may be approximated by Bessel functions

$$J_k(m) = \left(\frac{m}{2}\right) \sum_{n=0}^{\infty} \frac{-m^{2n}}{4^n n! (k+n)!} \approx \frac{1}{k!} \left(\frac{m}{2}\right)^k. \quad (2.56)$$

Making use of the identity $J_{-k}(m) = (-1)^k J_k(m)$ gives a second order expression for the phase modulated field

$$E \approx E_0 \exp(i\omega_0 t) \left[1 - \frac{m^2}{4} + i \frac{m}{2} \left(\exp(-i\Omega t) + \exp(i\Omega t) \right) \right]. \quad (2.57)$$

Notably, this field now has sidebands at $\omega_0 \pm \Omega$.

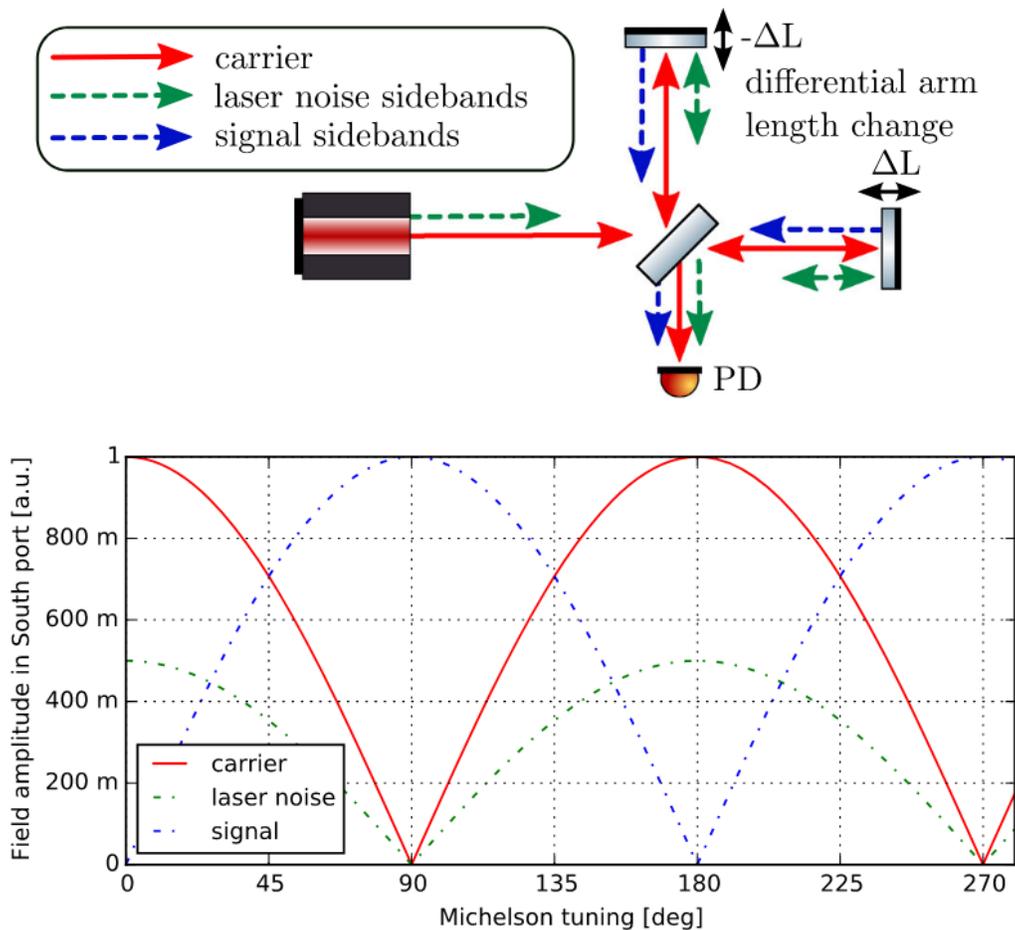


Figure 2.7: A schematic of sidebands in a Michelson interferometer, and the effect of tuning on the carrier and the sidebands, reproduced from Bond et al [24].

Figure 2.7 illustrates the behavior of these sidebands. Per the equation derived above, at 90 and 270 degree tunings the carrier power is minimized, and so too is the laser noise, which is a common sideband source across all legs. What is interesting is that the signal sidebands are maximized. Per the phase convention, transmission through a mirror introduces a phase of $\pi/2$, while reflection does not. Accordingly, the carrier and laser noise side bands will get a phase shift when entering the X arm, and when leaving the Y arm, such that they end up in the same phase. By contrast, the signal sideband will only experience a phase shift in the Y arm, so if no tuning is applied it will be perfectly out of phase and destructively interfere, whereas if there is a $\pi/2$ phase applied it will suddenly be perfectly in phase and constructively interfere.

Figure 2.7 shows that noise is minimized and signal maximized at the dark fringe,

but there is a complication: the absolute magnitude of the signal sidebands is very small, going quadratically with the strain, whereas if the sidebands beat against an oscillator signal the beat note will be linear in the strain. There are a number of solutions to this problem; currently LIGO utilizes a homodyne scheme, and in future upgrades will switch to a balanced homodyne scheme [26]. Essentially, the Michelson degrees of the interferometer are held very slightly off of the dark fringe, so that

$$\Delta L = \frac{\pi}{2k_0} + \delta_{off} \quad (2.58)$$

for $k_0 = \omega_0/c$ —the wavenumber of the carrier field—while the DC offset δ_{off} is held at a value much smaller than the wavelength of the carrier. In combination with Equation 2.47 and Equation 2.57 this gives

$$\begin{aligned} E &= irtE_0 \exp(i2k_0\bar{L}) \exp(i\omega_0 t) (2 \cos(k_0\Delta L) + s^+ + s^-) \\ &= irtE_0 \exp(i2k_0\bar{L}) \exp(i\omega_0 t) (2 \sin(k_0\delta_{off}) + s^+ + s^-), \end{aligned} \quad (2.59)$$

where s^\pm are the complex amplitudes of the upper and lower sidebands at the output port, yielding transmitted power

$$P = EE^* = TR|E_0|^2 \left(4 \sin^2(k_0\delta_{off}) + 2 \sin(k_0\delta_{off})(s^+ + s^-) + O(s^2) \right). \quad (2.60)$$

Response to a GW

The phase modulation due to an incident GW is given by

$$\phi = -k_0L \mp \frac{\omega_0}{2} \int_{t-L/c}^t h(t) = -k_0L \mp \delta\phi. \quad (2.61)$$

Considering a very simple gravitational wave

$$h(t) = h_0 \cos(\omega_{gw}t + \phi_{gw}), \quad (2.62)$$

one may compute

$$\delta\phi = \frac{\omega_0 h_0}{\omega_{gw}} \cos\left(\omega_{gw} + \phi_{gw} - \frac{\omega_{gw}L}{2c}\right) \sin\left(\frac{\omega_{gw}L}{2c}\right). \quad (2.63)$$

Going back to our phase modulation equation Equation 2.54 gives modulation index

$$m_{gw} = -\frac{\omega_0 h_0}{\omega_{gw}} \sin\left(\frac{k_{gw}L}{c}\right) \quad (2.64)$$

and phase

$$\Phi = -\frac{k_{gw}L}{2} + \phi_{gw} \quad (2.65)$$

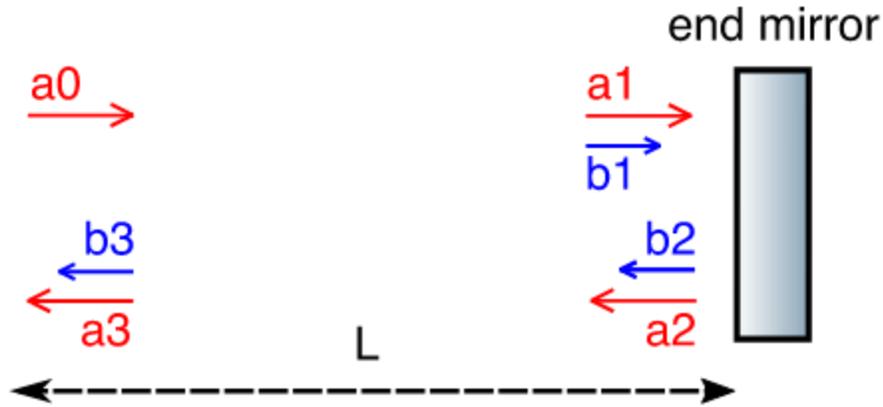


Figure 2.8: A schematic of the propagation of the carrier beam and the signal sidebands in one interferometer arm, reproduced from Bond et al [24].

where $k_{gw} = \omega_{gw}/c$.

Proceeding:

$$A_{gw} = \frac{m_{gw}}{2} = -\frac{\omega_0 h_0}{2\omega_{gw}} \sin\left(\frac{k_{gw}L}{c}\right) \quad (2.66)$$

$$\Phi_{gw}^{\pm} = \frac{\pi}{2} - Lk_0 \pm \left(-\frac{k_{gw}L}{2} + \phi_{gw}\right) \quad (2.67)$$

$$\alpha_{gw}^{\pm} = A_{gw} \exp(i\Phi_{gw}^{\pm}) \exp(\pm i\omega_{gw}t). \quad (2.68)$$

Looking to the schematic in Figure 2.8, the expressions for each optical propagation element are

$$a_3 = a_2 \exp(-ik_0L) \quad (2.69)$$

$$a_2 = r_{etm}a_1 \quad (2.70)$$

$$a_1 = a_0 \exp(-ik_0L) \quad (2.71)$$

and

$$b_1^\pm = a_0 \alpha_{gw}^\pm \quad (2.72)$$

$$b_2^\pm = r_{etm} b_1^\pm \quad (2.73)$$

$$b_3^\pm = b_2^\pm \exp(-i(k_0 \pm k_{gw})L) + a_2 \alpha_{gw}^\pm \quad (2.74)$$

$$= 2r_{etm} a_0 \alpha_{gw}^\pm \exp\left(-ik_0 L \mp i \frac{k_{gw} L}{2}\right) \cos\left(\mp i \frac{k_{gw} L}{2}\right) \quad (2.75)$$

$$= -i \frac{r_{etm} a_0 \omega_0 h_0}{2\omega_{gw}} \sin(k_{gw} L) \exp(-i2k_0 L) \exp(\pm i(\omega_{gw} t - k_{gw} L + \phi_{gw})). \quad (2.76)$$

For each arm respectively

$$b_X^\pm = b_3^\pm(L = L_X, a_0 = itE_0) \quad (2.77)$$

and

$$b_Y^\pm = b_3^\pm(L = L_Y, a_0 = rE_0) \quad (2.78)$$

so that

$$E_{out} = i2rtE_0 \cos(k_0 \Delta L) + b_X^+ + b_X^- + b_Y^+ + b_Y^-. \quad (2.79)$$

After recombining in the output port, these go to

$$b_Y^3 = \frac{rtE_0 \omega_0 h_0}{2\omega_{gw}} \sin(k_{gw} L_Y) \exp(-i2k_0 L_Y) \exp(\pm i(\omega_{gw} t - k_{gw} L_Y + \phi_{gw})) \quad (2.80)$$

$$b_X^3 = -\frac{rtE_0 \omega_0 h_0}{2\omega_{gw}} \sin(k_{gw} L_X) \exp(-i2k_0 L_X) \exp(\pm i(\omega_{gw} t - k_{gw} L_X + \phi_{gw})). \quad (2.81)$$

The negative sign on the X arm is a kludgy implementation of the antenna response to this polarization. Now, switch this to CARM and DARM, and assume a perfect 50:50 splitter. LIGO detectors are set to $k_{GW} \gg \Delta L$, such that $k_{gw}(\bar{L} + \Delta L/2) \approx k_{gw}\bar{L}$. Taking all this and summing—also using the fact that for our homodyne scheme $\sin(2k_0 \Delta L) \approx 2k_0 \delta_{off}$ —gives the field

$$E_{out} = iE_0 \cos(k_0 \Delta L) - i \frac{2k_0 \delta_{off} E_0 \omega_0 h_0}{\omega_{gw}} \sin(k_{gw} \bar{L}) \cos(\omega_{gw} t - k_{gw} \bar{L} + \phi_{gw}), \quad (2.82)$$

and the power due to the gravitational wave is

$$P_{gw} \approx 2k_0 \delta_{off} \frac{\omega_0 h_0}{\omega_{gw}} |E_0|^2 \sin(k_{gw} L) \cos(\omega_{gw} t - k_{gw} \bar{L} + \phi_{gw}). \quad (2.83)$$

Being proportional to strain h_0 this power is intrinsically very small, and hence needs amplification from a very large E_0 to be detectable. This also gives the transfer function

$$T_{gw \rightarrow P}(\omega_{gw}) \approx k_0 \delta_{off} |E_0|^2 \frac{\omega_0}{\omega_{gw}} \sin(k_{gw} \bar{L}) \exp(-ik_{gw} \bar{L}). \quad (2.84)$$

Noise Sources

For the analysts practical purposes, the most important implication of detector design is in the sensitivity of the detector at various frequency bands. This depends on the noise sources which pollute those frequency bands, and so it is valuable to discuss some of them in detail.

Shot Noise

Shot noise is a fundamental source of noise due to vacuum fluctuations of the electric fields; it can never be removed, but increasing the circulating power in the detector mitigates it. The noise PSD due to shot noise is given by

$$S_{P,DC} \approx 2P_0 (k_0 \delta_{off})^2 \hbar \omega_0. \quad (2.85)$$

Combined with the noise transfer function given in equation 2.84, this gives the noise-to-signal ratio

$$NSR = \frac{\sqrt{S_{P,DC}}}{T_{gw \rightarrow P}} = \sqrt{\frac{2\hbar}{P_0 \omega_0} \frac{\omega_{gw}}{\sin(\omega_{gw} \bar{L}/c)} \frac{h}{\sqrt{Hz}}}. \quad (2.86)$$

Importantly, this decreases with the square root of carrier power, and has a sharp feature at the free spectral range of the arm, though this doesn't come up very often in the LIGO detectors, since noise is already too high for detectability in that frequency range. Figure 2.9 shows what this looks like.

Radiation Pressure

Radiation pressure is another fundamental noise source, caused by displacement noise due to vacuum fluctuations against the mirrors, and results in a power spectral density

$$S_{\phi,RP} = \frac{8\hbar P_0^3 \omega_0 k_0^2}{M^2 c^2 \Omega^4}. \quad (2.87)$$

The NSR due to this will be

$$\sqrt{\frac{8\hbar P_0}{\omega_0} \frac{1}{\delta_{off} M c \omega_{gw}^2 \sin(\omega_{gw} \bar{L}/c)}}. \quad (2.88)$$

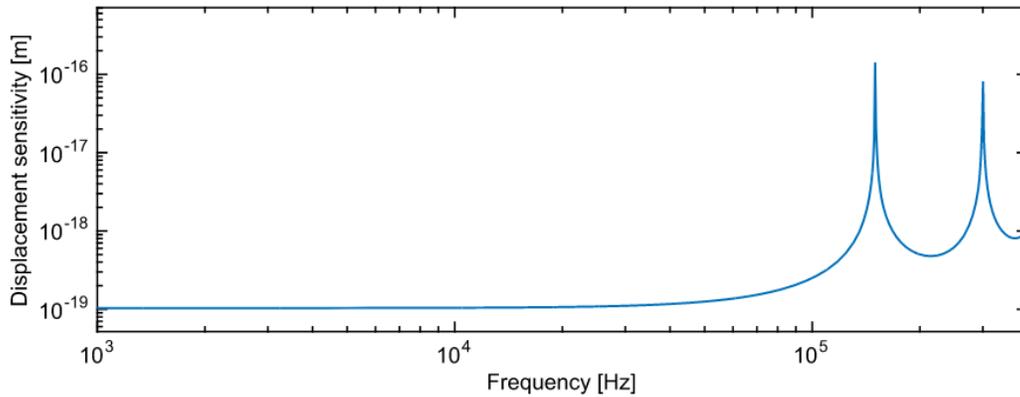


Figure 2.9: Shot noise sensitivity limit for $P_0 = 1\text{ W}$ and $L = 1\text{ km}$, reproduced from Bond et al [24]

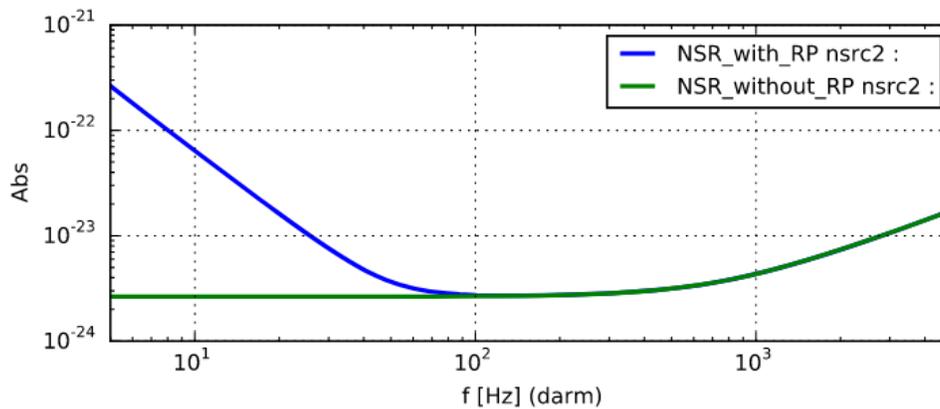


Figure 2.10: Sensitivity limits due to both shot noise and radiation pressure, reproduced from Bond et al [24].

Notably this increases with the square root of power, but it decreases with the square of the frequency, and so is principally a low frequency noise source. The combined effect of radiation pressure and shot noise can be seen in Figure 2.10. The full noise breakdown of the aLIGO design [27]—not the same as what actually exists—is seen in Figure 2.11.

Modern Interferometer Design Elements

Finally, I will briefly note some of the core design elements of modern GW detectors, which have made their present sensitivity possible.

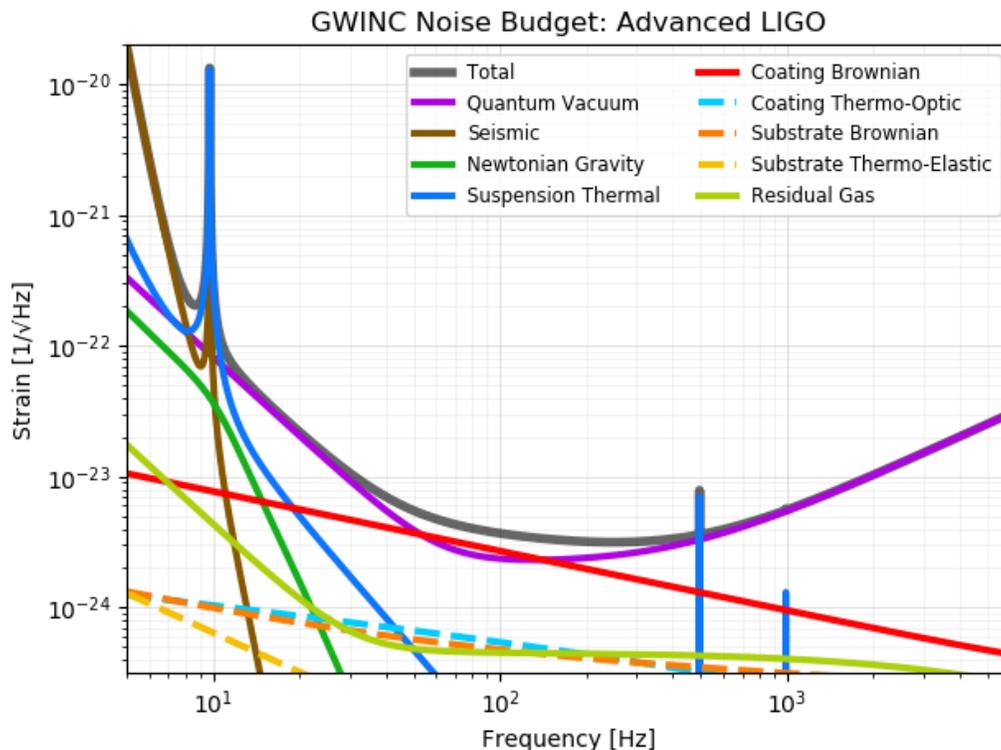


Figure 2.11: Breakdown of noise sources in aLigo according to the GWINC model, Reproduced from Rollins et al [28].

Power Recycling

As can be seen in the above discussion of shot noise, to decrease the noise floor it is necessary to increase circulating power. However, increasing the raw laser power past a certain point begins to sacrifice laser stability to an unacceptable degree, and so it is desirable to instead amplify the laser by optical methods. This is the purpose of the power recycling cavity, which essentially “catches” power leaking from the symmetric port and puts it back into the detector at resonance, a process requiring careful controls. This provides a gain in power proportional to the finesse of the PRC

$$G_{PR} \approx \frac{\mathcal{F}}{\pi}. \quad (2.89)$$

Figure 2.12 shows what this looks like schematically.

Arm Cavities

Adding Fabry-Pérot cavities to the arms of the detector also allows an increase in power, both in the carrier and, to an even greater extent, the signal sideband.

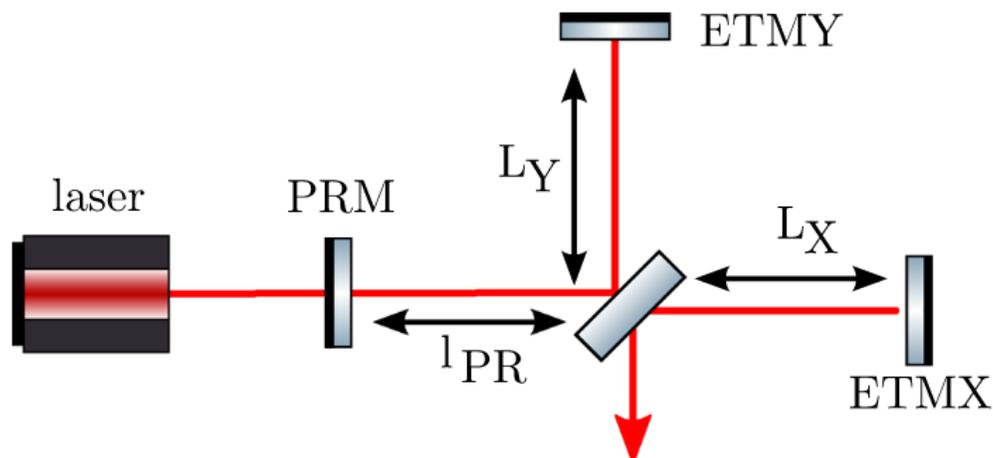


Figure 2.12: The layout of an interferometer with a power recycling cavity added, reproduced from Bond et al [24].

Equivalently, they increase the effective length of the arms, since photon lifetimes within the arms are significantly increased. Figure 2.13 shows the schematic of what this looks like.

Signal Recycling

A signal recycling cavity is a Fabry-Pérot cavity placed at the output port, as illustrated in Figure 2.14, which allows for resonant sideband extraction [24]. If the finesse of arm cavities above is very high, then a very sharp resonant feature is developed, narrowing the bandwidth. The SRC is tuned to an anti-resonant operating point, which increases the bandwidth of the detector. The effect of tuning on the sideband sensitivity can be seen in Figure 2.15.

2.5 Gravitational Wave Detector Glitches

Transient noise sources, known as “glitches”, will be discussed in detail in Part II of this thesis, but while discussing GW detectors some background on glitches can be valuable. GW detectors are typically characterized by their underlying noise process, which is assumed to be a Gaussian Markovian noise process which varies slowly over time, such that it may be treated as stationary over durations of minutes. This allows for clean decompositions in the frequency domain, where each frequency bin is assumed to be independent. However, these assumptions may be broken if rapidly varying—that is over the space of seconds—noise sources appear.

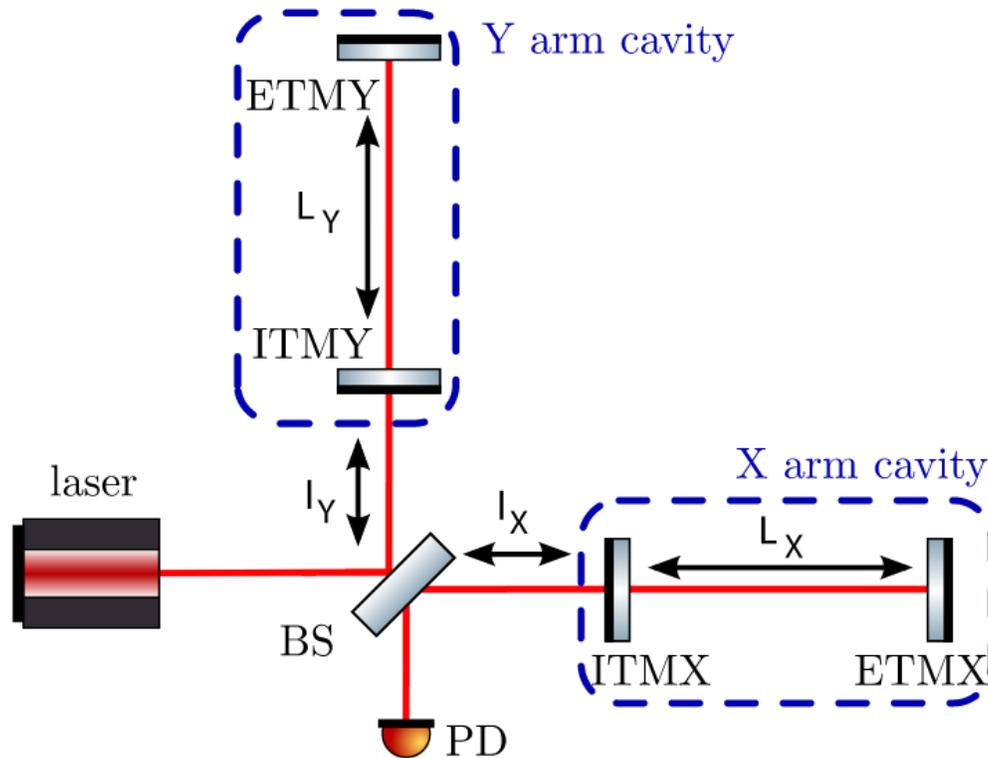


Figure 2.13: The layout of an interferometer with arm cavities added, reproduced from Bond et al [24].

The morphologies of these glitches can vary widely. Commonly occurring examples are shown in Figure 2.16, including blips, extremely loud glitches, slow scattering, and fast scattering, the last of which will be dealt with in detail in Part II. Furthermore, the processes which generate glitches are as varied as their morphologies. Some are well understood, such as scattered light glitches, while others have no clearly identified origin [29].²

The study of glitches and their mitigation is one of the central goals of detector characterization, and spans a wide variety of tasks. Some studies focus on modeling and testing the behavior of the detector itself, searching for environmental noise correlations and methods to correct them [31, 32]. Others studies seek to classify glitches, or to construct search methods which may identify the presence of glitch power [33–36]. Finally, some studies seek to mitigate the impact of glitches which are concurrent with GW observations, removing identified glitch power by one of a

²This is often reflected in the name of the glitch classification. For example, scattered light is named after its physical mechanism, while tomtes are named after a type of Scandinavian fae.

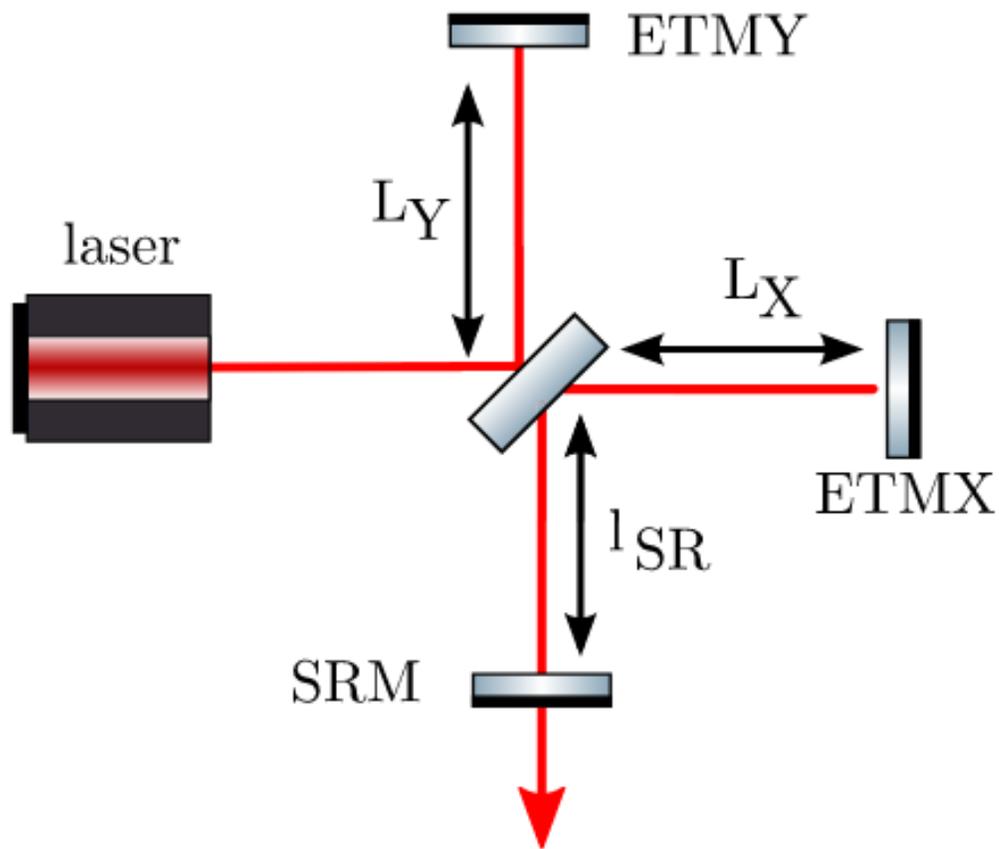


Figure 2.14: The layout of an interferometer with a signal recycling cavity added, reproduced from Bond et al [24].

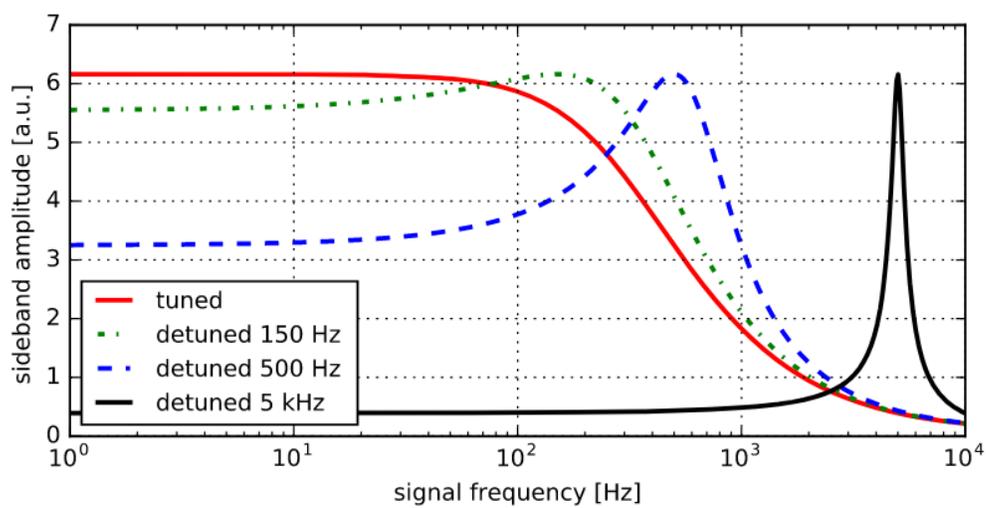


Figure 2.15: Sideband amplitude for various tunings of the SRC, reproduced from Bond et al [24].

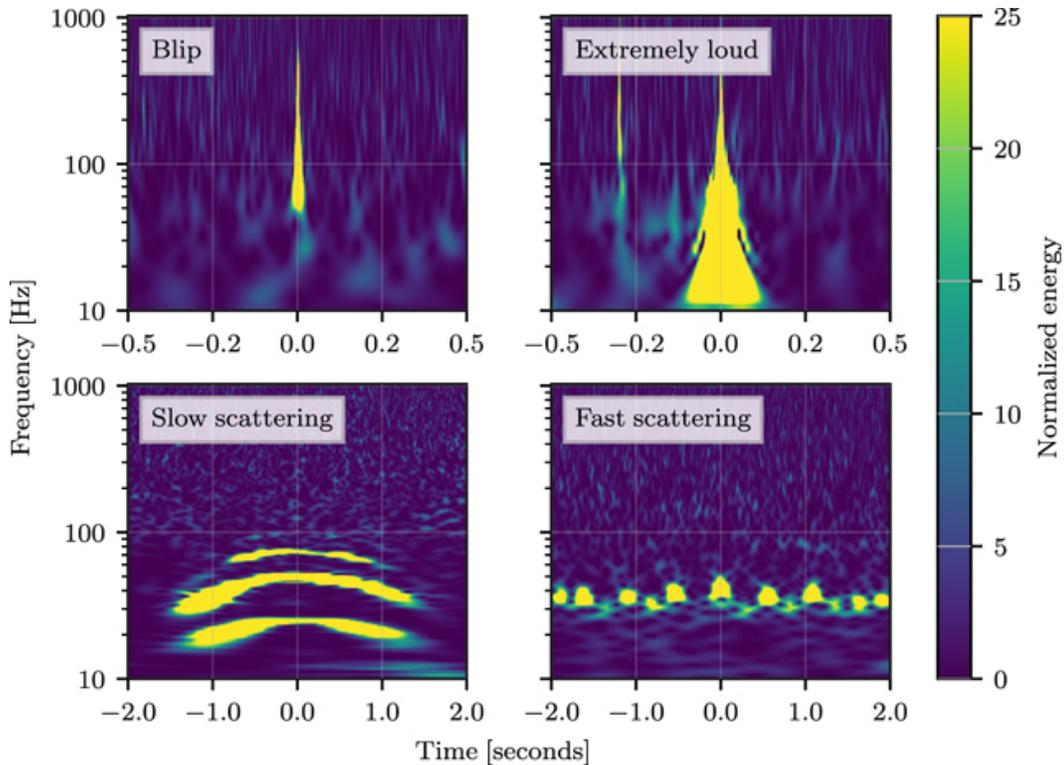


Figure 2.16: Commonly occurring glitch morphologies in ground based GW detectors. These include blips, “extremely loud” glitches, and both slow and fast scattering. Reproduced from Davis et al [30].

variety of methods [37–43]. The contents of this thesis focus on this final aspect of detector characterization, but it is useful to briefly discuss some of the more notable tools used by the detector characterization group to identify and classify glitches.

In addition to the main strain measurement, GW detectors also have thousands of auxiliary sensors which monitor many aspects of the detector. Accordingly, glitches are identified both from measured strain and from correlations between that strain and these auxiliary sensors. Algorithms to do each of these vary, from brute force correlators and excess power statistics to machine learning based classifiers. Though there are many such tools, a few will be of note for this thesis.

OMICRON

The first step for identifying many glitches consists of identifying excess power in the strain channel more generally, which can include both noise transients and genuine GWs. OMICRON [36] is commonly used for this purpose. This algorithm works by computing Q-transforms of data on a rolling basis, computing the power

in each tile, and identifying tiles with exceptionally high power. OMICRON is accordingly agnostic to the source of the power, and may produce triggers for both glitches and GWs, but the absence of matched filter search triggers may be used to identify the remainder as glitches. This process is useful in and of itself, and also provides triggers for downstream glitch analysis pipelines including Gravity Spy and GWDETFCHAR-SCATTERING.

Gravity Spy

Gravity Spy [34, 35] is a popular algorithm for the classification of detected glitches. Glitch triggers are produced by separate algorithms such as OMICRON, but these typically do not identify the source class of the glitch. Gravity Spy approaches the problem of doing so using a machine learning classification algorithm, based on spectrogram images produced of the offending glitches.

Gravity Spy is not generally used in the process of glitch mitigation within the collaboration, but is commonly used a source of training data for subsequent studies of glitch behavior. For example, in a search for the source of fast scattering glitches [31] due to the activity of trains passing by the Livingston site, Gravity Spy was used to identify instances of the glitch class of interest to allow for further study and correlation with environmental behavior. It is also commonly used as a source of training data for other machine learning algorithms targeting glitch behavior, such as GENGLI [44] which uses Gravity Spy to build training data for a glitch reconstruction algorithm.

Scattering

Scattered light glitches, which will be described in detail in Section 5.2, are an example of a glitch class which is sourced by known environmental noise sources[45]. Accordingly, it may be identified by correlating glitch power with the behavior of witness channels which are known to correspond with common culprits of scattering. The software package GWDETFCHAR-SCATTERING is used for this purpose, and identifies time segments which have increased activity in a scattering channel[46]. In doing so, this algorithm is able to identify examples of scattering, and indicate which surface within the detector may be responsible for the scattering in question.

References

- ¹S. M. Carroll, *Spacetime and Geometry: An Introduction to General Relativity* (Cambridge University Press, July 2019), doi:10.1017/9781108770385.
- ²C. Cutler and E. E. Flanagan, “Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral wave form?”, *Phys. Rev. D* **49**, 2658–2697 (1994) doi:10.1103/PhysRevD.49.2658.
- ³B. P. Abbott et al. (LIGO Scientific, Virgo), “The basic physics of the binary black hole merger GW150914”, *Annalen Phys.* **529**, 1600209 (2017) doi:10.1002/andp.201600209.
- ⁴R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run”, *Phys. Rev. X* **13**, 041039 (2023) doi:10.1103/PhysRevX.13.041039.
- ⁵I. 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**, 3295–3319 (2020) doi:10.1093/mnras/staa2850.
- ⁶B. P. Abbott et al. (LIGO Scientific, Virgo), “Observation of Gravitational Waves from a Binary Black Hole Merger”, *Phys. Rev. Lett.* **116**, 061102 (2016) doi:10.1103/PhysRevLett.116.061102.
- ⁷M. Andrés-Carcasona, A. J. Iovino, V. Vaskonen, H. Veermäe, M. Martínez, O. Pujolàs, and L. M. Mir, “Constraints on primordial black holes from LIGO–Virgo–KAGRA O3 events”, *Phys. Rev. D* **110**, 023040 (2024) doi:10.1103/PhysRevD.110.023040.
- ⁸R. Farmer, M. Renzo, S. E. de Mink, P. Marchant, and S. Justham, “Mind the gap: The location of the lower edge of the pair instability supernovae black hole mass gap”, *Astrophys. J.* **887**, 53 (2019) doi:10.3847/1538-4357/ab518b.
- ⁹A. Olejak, K. Belczynski, and N. Ivanova, “Impact of common envelope development criteria on the formation of LIGO/Virgo sources”, *Astron. Astrophys.* **651**, A100 (2021) doi:10.1051/0004-6361/202140520.
- ¹⁰R. C. Zhang, G. Fragione, C. Kimball, and V. Kalogera, “On the Likely Dynamical Origin of GW191109 and Binary Black Hole Mergers with Negative Effective Spin”, *Astrophys. J.* **954**, 23 (2023) doi:10.3847/1538-4357/ace4c1.
- ¹¹S. S. Bavera, T. Fragos, Y. Qin, E. Zapartas, C. J. Neijssel, I. Mandel, A. Batta, S. M. Gaebel, C. Kimball, and S. Stevenson, “The origin of spin in binary black holes: Predicting the distributions of the main observables of Advanced LIGO”, *Astron. Astrophys.* **635**, A97 (2020) doi:10.1051/0004-6361/201936204.

- ¹²J. Golomb, M. Isi, and W. M. Farr, “Physical Models for the Astrophysical Population of Black Holes: Application to the Bump in the Mass Distribution of Gravitational-wave Sources”, *Astrophys. J.* **976**, 121 (2024) doi:10.3847/1538-4357/ad8572.
- ¹³S. S. Bavera et al., “The impact of mass-transfer physics on the observable properties of field binary black hole populations”, *Astron. Astrophys.* **647**, A153 (2021) doi:10.1051/0004-6361/202039804.
- ¹⁴A. Olejak, J. Klencki, X.-T. Xu, C. Wang, K. Belczynski, and J.-P. Lasota, “Unequal-mass, highly-spinning binary black hole mergers in the stable mass transfer formation channel”, *Astron. Astrophys.* **689**, A305 (2024) doi:10.1051/0004-6361/202450480.
- ¹⁵I. Mandel and S. E. de Mink, “Merging binary black holes formed through chemically homogeneous evolution in short-period stellar binaries”, *Mon. Not. Roy. Astron. Soc.* **458**, 2634–2647 (2016) doi:10.1093/mnras/stw379.
- ¹⁶S. E. de Mink and I. Mandel, “The chemically homogeneous evolutionary channel for binary black hole mergers: rates and properties of gravitational-wave events detectable by advanced LIGO”, *Mon. Not. Roy. Astron. Soc.* **460**, 3545–3553 (2016) doi:10.1093/mnras/stw1219.
- ¹⁷K. Kremer, C. S. Ye, N. Z. Rui, N. C. Weatherford, S. Chatterjee, G. Fragione, C. L. Rodriguez, M. Spera, and F. A. Rasio, “Modeling Dense Star Clusters in the Milky Way and Beyond with the CMC Cluster Catalog”, *Astrophys. J. Suppl.* **247**, 48 (2020) doi:10.3847/1538-4365/ab7919.
- ¹⁸G. Fragione, A. Loeb, and F. A. Rasio, “On the Origin of GW190521-like events from repeated black hole mergers in star clusters”, *Astrophys. J. Lett.* **902**, L26 (2020) doi:10.3847/2041-8213/abbc0a.
- ¹⁹G. Fragione and S. Banerjee, “Binary black hole mergers from young massive and open clusters: comparison to GWTC-2 gravitational wave data”, (2021) doi:10.3847/2041-8213/ac00a7.
- ²⁰H. Tagawa, Z. Haiman, and B. Kocsis, “Formation and Evolution of Compact Object Binaries in AGN Disks”, *Astrophys. J.* **898**, 25 (2020) doi:10.3847/1538-4357/ab9b8c.
- ²¹V. Gayathri, D. Wysocki, Y. Yang, V. Delfavero, R. O. Shaughnessy, Z. Haiman, H. Tagawa, and I. Bartos, “Gravitational Wave Source Populations: Disentangling an AGN Component”, *Astrophys. J. Lett.* **945**, L29 (2023) doi:10.3847/2041-8213/acbfb8.
- ²²E. Payne, K. Kremer, and M. Zevin, “Spin Doctors: How to Diagnose a Hierarchical Merger Origin”, *Astrophys. J. Lett.* **966**, L16 (2024) doi:10.3847/2041-8213/ad3e82.

- ²³I. M. Romero-Shaw, P. D. Lasky, and E. Thrane, “Signs of Eccentricity in Two Gravitational-wave Signals May Indicate a Subpopulation of Dynamically Assembled Binary Black Holes”, *Astrophys. J. Lett.* **921**, L31 (2021) DOI:10.3847/2041-8213/ac3138.
- ²⁴C. Bond, D. Brown, A. Freise, and K. A. Strain, “Interferometer techniques for gravitational-wave detection”, *Living Rev in Relativ* **19** (2016) DOI:10.1007/s41114-016-0002-8.
- ²⁵R. Udall, *A sketch of the ligo detectors*, ARC Seminar Talk, <https://nicholasrui.com/arc-seminar/>, 2023.
- ²⁶E. Capote et al., “Advanced LIGO detector performance in the fourth observing run”, *Phys. Rev. D* **111**, 062002 (2025) DOI:10.1103/PhysRevD.111.062002.
- ²⁷J. Aasi, B. Abbott, R. Abbott, et al. (LIGO Scientific Collaboration), “Advanced LIGO”, *Class. Quantum Grav.* **32**, 074001 (2015) DOI:10.1088/0264-9381/32/7/074001.
- ²⁸J. G. Rollins, E. Hall, C. Wipf, and L. McCuller, “pygwinc: Gravitational Wave Interferometer Noise Calculator”, ascl:2007.020, <https://git.ligo.org/gwinc/pygwinc>, ascl:2007.020 (2020).
- ²⁹J. Glanzer et al., “Data quality up to the third observing run of advanced LIGO: Gravity Spy glitch classifications”, *Class. Quant. Grav.* **40**, 065004 (2023) DOI:10.1088/1361-6382/acb633.
- ³⁰D. Davis et al. (LIGO), “LIGO detector characterization in the second and third observing runs”, *Class. Quant. Grav.* **38**, 135014 (2021) DOI:10.1088/1361-6382/abfd85.
- ³¹J. Glanzer, S. Soni, J. Spoon, A. Effler, and G. González, “Noise in the LIGO livingston gravitational wave observatory due to trains”, *Class. Quant. Grav.* **40**, 195015 (2023) DOI:10.1088/1361-6382/acf01f.
- ³²A. Helmling-Cornell, P. Nguyen, R. Schofield, and R. Frey, “Automated evaluation of environmental coupling for Advanced LIGO gravitational wave detections”, *Class. Quant. Grav.* **41**, 145003 (2024) DOI:10.1088/1361-6382/ad5139.
- ³³A. E. Tolley, G. S. Cabourn Davies, I. W. Harry, and A. P. Lundgren, “ArchEnemy: removing scattered-light glitches from gravitational wave data”, *Class. Quant. Grav.* **40**, 165005 (2023) DOI:10.1088/1361-6382/ace22f.
- ³⁴M. Zevin et al., “Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science”, *Class. Quant. Grav.* **34**, 064003 (2017) DOI:10.1088/1361-6382/aa5cea.
- ³⁵M. Zevin et al., “Gravity Spy: lessons learned and a path forward”, *Eur. Phys. J. Plus* **139**, 100 (2024) DOI:10.1140/epjp/s13360-023-04795-4.

- ³⁶F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, “Omicron: a tool to characterize transient noise in gravitational-wave detectors”, *SoftwareX* **12**, 100620 (2020) DOI:10.1016/j.softx.2020.100620.
- ³⁷N. J. Cornish and T. B. Littenberg, “BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches”, *Class. Quant. Grav.* **32**, 135012 (2015) DOI:10.1088/0264-9381/32/13/135012.
- ³⁸S. Hourihane, K. Chatziioannou, M. Wijngaarden, D. Davis, T. Littenberg, and N. Cornish, “Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data”, *Phys. Rev. D* **106**, 042006 (2022) DOI:10.1103/PhysRevD.106.042006.
- ³⁹D. Davis, T. J. Massinger, A. P. Lundgren, J. C. Driggers, A. L. Urban, and L. K. Nuttall, “Improving the Sensitivity of Advanced LIGO Using Noise Subtraction”, *Class. Quant. Grav.* **36**, 055011 (2019) DOI:10.1088/1361-6382/ab01c5.
- ⁴⁰A.-K. Malz and J. Veitch, “Joint inference for gravitational wave signals and glitches using a data-informed glitch model”, (2025) DOI:10.48550/arXiv.2505.00657.
- ⁴¹G. Ashton, “Gaussian processes for glitch-robust gravitational-wave astronomy”, *Mon. Not. Roy. Astron. Soc.* **520**, 2983–2994 (2023) DOI:10.1093/mnras/stad341.
- ⁴²R. Udall and D. Davis, “Bayesian modeling of scattered light in the LIGO interferometers”, *Appl. Phys. Lett.* **122**, 094103 (2023) DOI:10.1063/5.0136896.
- ⁴³R. Udall, S. Hourihane, S. Miller, D. Davis, K. Chatziioannou, M. Isi, and H. Deshong, “Antialigned spin of GW191109: Glitch mitigation and its implications”, *Phys. Rev. D* **111**, 024046 (2025) DOI:10.1103/PhysRevD.111.024046.
- ⁴⁴M. Lopez, V. Boudart, S. Schmidt, and S. Caudill, “Simulating Transient Noise Bursts in LIGO with gengli”, (2022) DOI:10.48550/arXiv.2205.09204.
- ⁴⁵T. Accadia et al., “Noise from scattered light in Virgo’s second science run data”, *Class. Quant. Grav.* **27**, edited by F. Ricci, 194011 (2010) DOI:10.1088/0264-9381/27/19/194011.
- ⁴⁶S. Soni, C. Austin, A. Effler, et al. (LIGO), “Reducing scattered light in LIGO’s third observing run”, *Class. Quant. Grav.* **38**, 025016 (2020) DOI:10.1088/1361-6382/abc906.

Chapter 3

PARAMETER ESTIMATION FOR GRAVITATIONAL WAVE TRANSIENTS

3.1 Bayesian inference in gravitational wave analysis

After a GW is detected, much of the analysis which will follow falls under the purview of parameter estimation (PE). PE is the process by which one infers the nature of the source system which produced the observed GW. Because we have generally only a single observation of any given GW¹, Bayesian statistics is used to assess our confidence in the parameters of the system which generated it, rather than making frequentist statements about its likelihood. Accordingly, to compute the posterior probability distribution given the data and our assumed model $p(\boldsymbol{\theta}|\mathbf{d}, \mathcal{M})$ it is necessary to introduce Bayes' Theorem [1, 2]

$$p(\boldsymbol{\theta}|\mathbf{d}, \mathcal{M}) = \frac{\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}, \mathcal{M})\pi(\boldsymbol{\theta}|\mathcal{M})}{\mathcal{Z}(\mathbf{d}|\mathcal{M})}, \quad (3.1)$$

where $\boldsymbol{\theta}$ are the binary source parameters, \mathbf{d} are the observed data, \mathcal{M} is the model—which includes assumed waveform physics, assumed noise properties, assumed priors, and more—for the analysis, $\mathcal{L}(\mathbf{d}|\boldsymbol{\theta}, \mathcal{M})$ is the likelihood of observing these data given the model and parameters in question, $\pi(\boldsymbol{\theta}|\mathcal{M})$ are the priors which encode our beliefs about the parameters $\boldsymbol{\theta}$ before the observation of \mathbf{d} , and $\mathcal{Z}(\mathbf{d}|\mathcal{M})$ is the integrated evidence

$$\mathcal{Z}(\mathbf{d}|\mathcal{M}) = \int_{\boldsymbol{\Theta}} \mathcal{L}(\mathbf{d}|\boldsymbol{\theta}, \mathcal{M})\pi(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}. \quad (3.2)$$

Equation 3.1 lays out the elements which are necessary for conducting Bayesian inference of CBC source parameters. One needs a likelihood function, the elements of a model, and a choice of priors to define the posterior. Furthermore, if the dimensionality of the parameter spaces becomes large ($N \gtrsim 5$) grid based methods of evaluation will become prohibitively costly, and so typically stochastic sampling algorithms such as Markov Chain Monte Carlo or nested sampling will be necessary to efficiently sample from this distribution. The remainder of this chapter will discuss these components in turn.

¹Potential lensed events notwithstanding.

The frequency domain likelihood

The likelihood is to be understood as the probability—in a frequentist sense—that the given model could produce data which we have observed. For now I will assume—though this assumption will be challenged in Part II of this thesis—that the data may be represented as the sum of a signal $h(t|\theta_0, \mathcal{M})$ and stationary Gaussian noise $n(t)$

$$d(t) = h(t|\theta_0, \mathcal{M}) + n(t). \quad (3.3)$$

By choosing a segment of data sufficiently long to contain the whole observable signal, then applying a tapered window to the ends, one may compute the frequency domain data, which is the representation preferred for most purposes

$$\tilde{d}(f) = \tilde{h}(f|\theta, \mathcal{M}) + \tilde{n}(f). \quad (3.4)$$

The models $\tilde{h}(f|\theta, \mathcal{M})$ are deterministic, and so the probabilistic elements of the likelihood will derive entirely from the properties of $\tilde{n}(f)$. The noise-weighted inner product in data space is [3]

$$(a|b) = 2 \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)}{S_n(f)} df, \quad (3.5)$$

or equivalently

$$(a|b) = 4\text{Re} \int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} df. \quad (3.6)$$

This equation also has the discretization

$$(a|b) = 4 \sum_{k=0}^N \left\{ \frac{[a_k b_k]^2}{\sigma_k^2} \right\}, \quad (3.7)$$

where a_k, b_k are the discretized component of templates a and b and σ_k^2 is the same for the power spectral density This formulation is used for computational purposes.

Under this definition the probability of a given noise distribution n may be represented as

$$p(n) \propto e^{-(n|n)}. \quad (3.8)$$

Accordingly, if the parameters being tested are true, and our model is correct, it will be the case that

$$n(t) = d(t) - h(t|\theta, \mathcal{M}), \quad (3.9)$$

and so the likelihood of these data under these parameters is (suppressing the model dependence for clarity)

$$\mathcal{L}(d|\theta, \mathcal{M}) \propto e^{-\frac{(d-h(\theta)|d-h(\theta))}{2}}. \quad (3.10)$$

This may also be discretized with the help of a Fast Fourier Transform to produce the Whittle likelihood used in GW analysis [1]

$$\ln \mathcal{L}(d|\boldsymbol{\theta}) = -\frac{1}{2} \sum_{k=0}^N \left\{ \frac{[d_k - h_k(\boldsymbol{\theta})]^2}{\sigma_k^2} + \ln(2\pi\sigma_k^2) \right\}, \quad (3.11)$$

where d_k , h_k are the discretized frequency components of the data and waveform respectively. The term $\ln(2\pi\sigma_k^2)$ is a constant term, and so sampling is typically performed using the likelihood ratio which drops this term.

Waveform approximants

As referenced in Section 3.1, one of the most important assumptions made as part of analysis is the model $\tilde{h}(f|\boldsymbol{\theta}, \mathcal{M})$. Because GR is a highly non-linear theory, computationally intensive numerical relativity (NR) simulations are necessary to fully solve for the GW waveform. These computational demands make it infeasible to perform NR as part of the PE process, and so waveform approximants have been developed to—as the name implies—approximate the GW waveform. These may be categorized into families, which share common methods of derivation and have their respective strengths and weaknesses. I will briefly discuss three of the most notable BBH waveform families—the phenomenological inspiral-merger-ringdown family, the effective-one-body family, and the NR surrogate family—and their use in LVK analysis.

Inspiral-Merger-Ringdown Phenomenological family waveform approximants

The Inspiral-Merger-Ringdown Phenomenological (IMRPhenom) family of waveform approximants are frequency domain approximants which, as the name implies utilize phenomenological methods of three regions of waveform evolution to capture the behavior of the true GW [4–6]. In each case, some ansatz is developed to describe the behavior in that region of both the amplitude and phase evolution, with free parameters that are then calibrated to reference waveforms. These reference waveforms are hybridizations of an effective-one-body model with a NR surrogate [4].

For the low frequency inspiral regime, the waveform is modeled using the post-Newtonian formalism [7] with added correction terms. For the high frequency ringdown regime, a Lorentzian with added background terms is used [4]. Finally, the intermediate merger regime is modeled phenomenologically to bridge the gap between low and high frequencies. Between these regimes hundreds of free calibration

parameters are introduced, and an optimization algorithm is used to minimize the mismatch with respect to the reference waveforms. For precessing-spin IMRPhenom models, a “twist-up” procedure is used to map the aligned spin waveforms to precessing frames [5].

From an analyst’s perspective, the IMRPhenom family provides a number of advantages. Firstly, computation is much faster—by multiple orders of magnitude, depending on the configuration [8, 9]—than competitor waveform approximants. IMRPhenomXPHM also allows a broad range of input parameters, including mass ratios up to $q = 1000$ [4], and a wide range of system total masses [4]. However, IMRPhenomXPHM also shows the lowest fidelities of the major waveforms for higher spinning events, making its results less reliable in events with more exotic configurations [6]. Accordingly, IMRPhenom waveforms are used by the LVK for initial, rapid followup, being supplemented by more computationally intensive results afterwards.

Effective-One-Body family waveform approximants

The spinning effective-one-body NR calibrated (SEOBNR) family is a time-domain family of waveform approximants. These approximants use a Hamiltonian—constructed with an effective-one-body formalism, hence the name—to evolve the system’s trajectory and model the GW waveform [9, 10]. As with the IMRPhenom family, these waveform approximants use post-Newtonian theory and NR to calibrate their models. For SEOBNRv5PHM, precessing-spin waveforms also use a “twist-up” procedure, though it is somewhat modified [9].

SEOBNR family waveforms span a similar range of input parameters to IMRPhenom family waveforms, and are expected to get *more* accurate a higher mass ratios due to the increasing fidelity of the effective-one-body approximation when one mass is substantially larger than the other. They also provide generally higher fidelity than do the IMRPhenom family of waveforms [6], and differences in the modeling used allow it to serve as a valuable check for waveform systematics [11]. The downside, however, is that this family of waveform approximants is generally the slowest, requiring in some cases specialized inference pipelines such as DINGO [12] and RIFT [13] to make analysis computationally tractable [9].

NRSurrogate family waveform approximants

The final family of waveform approximants which I will focus on are the NR surrogates, particularly NRSur7dq4 [14]. These approximants are constructed by

the direct interpolation of NR waveforms and so have the highest fidelities to NR of any approximant. However, this also restricts their region of validity to portions of parameter space for which there are sufficient NR waveforms to perform robust interpolation. For NRSur7dq4 this region is $q \leq 4$ and $|a_i| \leq 0.8$, though the waveform can also operate within the extrapolation region of $q \leq 6$ or $|a_i| \leq 0.99$ [14].

Due to the limitations in length of NR waveforms, they are also necessarily restricted in the number of cycles they may model, such that there is a minimum total system mass which they may model. Hybridized waveforms like NRHybSur3dq8 can resolve this challenge by stitching the interpolation region to an inspiral region modeled by PN, but these are presently limited to aligned spin configurations and are slower to evaluate [15]. As a result of these features, NRSur7dq4 is considered the best available model for systems with high masses and high spins. In particular, it has served as an important check on exceptionally interesting events which display significant spin effects including GW200129 [16–18] and GW191109 [18, 19].

Common CBC analysis priors

The choice of priors is an important one in the procedure of Bayesian inference, and so the choice of priors in PE for CBCs is fairly standardized. These are described in [20], but I will also recapitulate them briefly here and note certain common exceptions.

Nine of the fifteen CBC parameters are angles. Of these, four are polar angles—inclination θ_{JN} , spin tilt angles θ_1 and θ_2 , and declination δ —and so the first three have sinusoidal priors over $[0, \pi]$ while the fourth has a cosinusoidal prior over $[-\pi/2, \pi/2]$ due to the way it is defined. The other five angles—right ascension α , spin azimuthal angles ϕ_{12} and ϕ_{JL} , phase ϕ and polarization angle ψ —are all azimuthal angles, and so have uniform priors over $[0, 2\pi]$ with periodic boundary conditions.²

The other six parameters are two masses m_1 and m_2 ³, two dimensionless spin magnitudes a_1 and a_2 , the time of coalescence t_c , and the luminosity distance D_L . Spin magnitude priors are set to be uniform over the allowable bounds $a_i < 1$,

²Since ψ enters as 2ψ in the argument of sin and cos functions it is equivalently sampled over $[0, \pi]$ for efficiency.

³The masses which are directly measured—and hence the analysis priors used for the mass—are in the detector frame but, in combination with redshift z inferred from the luminosity distance, they may be converted to source frame masses m_1^{src} and m_2^{src} .

and time of coalescence is also uniform about the approximate time at which the GW trigger was observed. Mass priors may be set to be uniform in the component masses, or in equivalent representations such as chirp mass/mass ratio \mathcal{M}/q or total mass/mass ratio M/q . Occasionally they may be sampled in one parameterization under a prior which sets them to be uniform in the components [21]. Finally, luminosity distance is generally sampled from a cosmologically informed uniform in source frame prior, which sets the prior under an assumption of uniform rate per source frame time and comoving volume [20].

3.2 Parameter Estimation Pipelines

Having established the ingredients of computing a posterior distribution with Bayes' theorem, it remains to actually perform this computation. Traditionally this is performed with stochastic samplers [2, 22], though there is an increasing spectrum of other methods being attempted [12, 23]. Here I will discuss in some detail the two PE pipelines which are used directly in this thesis—`BILBY` and `RIFT`—and their positive and negative qualities.

BILBY

Since its introduction in 2019, `BILBY` has become the predominant PE pipeline in GW science. `BILBY` is designed as a flexible, modular, and user friendly implementation of Bayesian inference [22], and is designed to be broadly useful even outside the GW use case, though that remains its primary use case. Accordingly, `BILBY` uses an object-oriented paradigm, including classes for core elements of Bayesian inference such as the `Prior`, `Likelihood`, and so on. In Chapter 5 this flexibility is used extensively to implement an add-on package for the inference of glitch models.

Another important aspect of `BILBY`'s flexibility is that it implements multiple sampling methods, most prominently including the `DYNesty` nested sampler [24] and a custom Markov Chain Monte Carlo algorithm `bilby-mcmc` [25]. In practice `DYNesty` has found to be the most robust option for standard GW inference, and so is the LVK's workhorse algorithm, but this flexibility makes `BILBY` adaptable to future improvements in these algorithms.

`BILBY` also allows for the optional use of a variety of features which improve the computational efficiency of sampling and of likelihood computations. In Section 5.3 I discuss in detail the case of distance marginalization, but methods are also available for marginalization over time and phase parameters [20]. `BILBY` also implements methods for heterodyned likelihoods [26, 27], multibanded likelihoods [28, 29], and

reduced-order-quadrature (ROQ) likelihoods [21]. The last in particular have led to orders-of-magnitude improvements in the speed of analysis for low mass BNS signals which may have electromagnetic counterparts [30]. Separately, `BILBY` implements improvements to the `DYNESTY` sampling algorithm which produce significant improvements in the efficiency of sampling.

RIFT

`RIFT`, which will be discussed in some detail in Chapter 4, is an alternate PE inference pipeline with more use-case specific implementations of Bayesian inference. Born out of efforts to efficiently compute likelihoods using individual NR simulations, `RIFT` separates parameters into intrinsic parameters λ —masses and spins—and extrinsic parameters θ —everything else—and marginalizes over the intrinsic parameters. Using a combination of mode-by-mode decomposition and Monte Carlo integration, this marginalized likelihood may be computed efficiently for a given intrinsic parameter configuration [13, 31–33]⁴

$$\mathcal{L}_{\text{marg}}(\lambda) = \int_{\Theta} \mathcal{L}(\lambda, \theta) p(\theta) d\theta. \quad (3.12)$$

A key feature of this procedure is that the waveform computation itself is performed using a single set of intrinsic parameters, and variation over the extrinsic parameters happens without recomputation. For expensive waveforms such as those in the `SEOBNR` family this is highly advantageous, but this benefit is much less dramatic if the waveform itself is relatively inexpensive such as for the `IMRPhenom` family.

The other dramatic difference between `RIFT` and other PE pipelines is that posteriors are constructed by a fitting feature, then resampled. Samples are drawn from the current working distribution in intrinsic parameter space, their marginalized likelihoods are computed, then a fit over the likelihoods is computed using an algorithm such as a Gaussian process [31, 32]. This is initially done with the prior distribution, but the computed fit is then used as the new working distribution and this process proceeds recursively for a fixed number of iterations. This two-step process alternates between an embarrassingly parallelizable procedure which may be run on GPUs [13, 31] and a single fitting process which runs on a CPU core. Unfortunately, it turns out that this diversity of resource usage and high number of parallel jobs makes `RIFT` exceptionally susceptible to cluster issues, which result in routine job failures [34]. This issue is addressed in Chapter 4.

⁴Unless the job fails or hangs for some reason

References

- ¹G. Ashton et al., “BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy”, *Astrophys. J. Suppl.* **241**, 27 (2019) DOI:10.3847/1538-4365/ab06fc.
- ²J. Veitch et al., “Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library”, *Phys. Rev. D* **91**, 042003 (2015) DOI:10.1103/PhysRevD.91.042003.
- ³C. Cutler and E. E. Flanagan, “Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral wave form?”, *Phys. Rev. D* **49**, 2658–2697 (1994) DOI:10.1103/PhysRevD.49.2658.
- ⁴G. Pratten, S. Husa, C. Garcia-Quiros, M. Colleoni, A. Ramos-Buades, H. Estelles, and R. Jaume, “Setting the cornerstone for a family of models for gravitational waves from compact binaries: The dominant harmonic for nonprecessing quasicircular black holes”, *Phys. Rev. D* **102**, 064001 (2020) DOI:10.1103/PhysRevD.102.064001.
- ⁵G. Pratten et al., “Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes”, *Phys. Rev. D* **103**, 104056 (2021) DOI:10.1103/PhysRevD.103.104056.
- ⁶J. Mac Uilliam, S. Akcay, and J. E. Thompson, “Survey of four precessing waveform models for binary black hole systems”, *Phys. Rev. D* **109**, 084077 (2024) DOI:10.1103/PhysRevD.109.084077.
- ⁷L. Blanchet, “Gravitational Radiation from Post-Newtonian Sources and Inspiral Compact Binaries”, *Living Rev. Rel.* **17**, 2 (2014) DOI:10.12942/lrr-2014-2.
- ⁸C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, “Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries”, *Phys. Rev. D* **102**, 064002 (2020) DOI:10.1103/PhysRevD.102.064002.
- ⁹A. Ramos-Buades, A. Buonanno, H. Estellés, M. Khalil, D. P. Mihaylov, S. Ossokine, L. Pompili, and M. Shiferaw, “Next generation of accurate and efficient multipolar precessing-spin effective-one-body waveforms for binary black holes”, *Phys. Rev. D* **108**, 124037 (2023) DOI:10.1103/PhysRevD.108.124037.
- ¹⁰L. Pompili et al., “Laying the foundation of the effective-one-body waveform models SEOBNRv5: Improved accuracy and efficiency for spinning nonprecessing binary black holes”, *Phys. Rev. D* **108**, 124035 (2023) DOI:10.1103/PhysRevD.108.124035.
- ¹¹R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run”, *Phys. Rev. X* **13**, 041039 (2023) DOI:10.1103/PhysRevX.13.041039.

- ¹²M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, “Real-Time Gravitational Wave Science with Neural Posterior Estimation”, *Phys. Rev. Lett.* **127**, 241103 (2021) doi:10.1103/PhysRevLett.127.241103.
- ¹³D. Wysocki, R. O’Shaughnessy, J. Lange, and Y.-L. L. Fang, “Accelerating parameter inference with graphics processing units”, *Phys. Rev. D* **99**, 084026 (2019) doi:10.1103/PhysRevD.99.084026.
- ¹⁴V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, “Surrogate models for precessing binary black hole simulations with unequal masses”, *Phys. Rev. Research.* **1**, 033015 (2019) doi:10.1103/PhysRevResearch.1.033015.
- ¹⁵V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, “Surrogate model of hybridized numerical relativity binary black hole waveforms”, *Phys. Rev. D* **99**, 064045 (2019) doi:10.1103/PhysRevD.99.064045.
- ¹⁶M. Hannam et al., “General-relativistic precession in a black-hole binary”, *Nature* **610**, 652–655 (2022) doi:10.1038/s41586-022-05212-z.
- ¹⁷E. Payne, S. Hourihane, J. Golomb, R. Udall, D. Davis, and K. Chatziioannou, “Curious case of GW200129: Interplay between spin-precession inference and data-quality issues”, *Phys. Rev. D* **106**, 104017 (2022) doi:10.1103/PhysRevD.106.104017.
- ¹⁸T. Islam, A. Vajpeyi, F. H. Shaik, C.-J. Haster, V. Varma, S. E. Field, J. Lange, R. O’Shaughnessy, and R. Smith, “Analysis of GWTC-3 with fully precessing numerical relativity surrogate models”, (2023) doi:10.48550/arXiv.2309.14473.
- ¹⁹R. Udall, S. Hourihane, S. Miller, D. Davis, K. Chatziioannou, M. Isi, and H. Deshong, “Antialigned spin of GW191109: Glitch mitigation and its implications”, *Phys. Rev. D* **111**, 024046 (2025) doi:10.1103/PhysRevD.111.024046.
- ²⁰I. 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**, 3295–3319 (2020) doi:10.1093/mnras/staa2850.
- ²¹S. Morisaki and V. Raymond, “Rapid Parameter Estimation of Gravitational Waves from Binary Neutron Star Coalescence using Focused Reduced Order Quadrature”, *Phys. Rev. D* **102**, 104020 (2020) doi:10.1103/PhysRevD.102.104020.
- ²²G. Ashton, M. Huebner, P. Lasky, et al., “BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy”, *Astrophys. J. Suppl.* **241**, 27 (2019) doi:10.3847/1538-4365/ab06fc.
- ²³A. H. Nitz, “Robust, Rapid, and Simple Gravitational-wave Parameter Estimation”, (2024) doi:10.48550/arXiv.2410.05190.

- ²⁴J. S. Speagle, “dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences”, *Mon. Not. Roy. Astron. Soc.* **493**, 3132–3158 (2020) DOI:10.1093/mnras/staa278.
- ²⁵G. Ashton and C. Talbot, “Bilby-MCMC: an MCMC sampler for gravitational-wave inference”, *Mon. Not. Roy. Astron. Soc.* **507**, 2037–2051 (2021) DOI:10.1093/mnras/stab2236.
- ²⁶N. J. Cornish, “Heterodyned likelihood for rapid gravitational wave parameter inference”, *Phys. Rev. D* **104**, 104054 (2021) DOI:10.1103/PhysRevD.104.104054.
- ²⁷K. Krishna, A. Vijaykumar, A. Ganguly, C. Talbot, S. Biscoveanu, R. N. George, N. Williams, and A. Zimmerman, “Accelerated parameter estimation in Bilby with relative binning”, (2023) DOI:10.48550/arXiv.2312.06009.
- ²⁸S. Morisaki, “Accelerating parameter estimation of gravitational waves from compact binary coalescence using adaptive frequency resolutions”, *Phys. Rev. D* **104**, 044062 (2021) DOI:10.1103/PhysRevD.104.044062.
- ²⁹S. Vinciguerra, J. Veitch, and I. Mandel, “Accelerating gravitational wave parameter estimation with multi-band template interpolation”, *Class. Quant. Grav.* **34**, 115006 (2017) DOI:10.1088/1361-6382/aa6d44.
- ³⁰S. S. Chaudhary 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**, e2316474121 (2024) DOI:10.1073/pnas.2316474121.
- ³¹J. Wofford et al., “Expanding RIFT: Improving performance for GW parameter inference”, (2022) DOI:10.48550/arXiv.2210.07912.
- ³²J. Lange, R. O’Shaughnessy, and M. Rizzo, “Rapid and accurate parameter inference for coalescing, precessing compact binaries”, <https://arxiv.org/abs/1805.10457> (2018).
- ³³C. Pankow, P. Brady, E. Ochsner, and R. O’Shaughnessy, “Novel scheme for rapid parallel parameter estimation of gravitational waves from compact binary coalescences”, *Phys. Rev. D* **92**, 023002 (2015) DOI:10.1103/PhysRevD.92.023002.
- ³⁴R. Udall, J. Brandt, G. Manchanda, A. Arulanandan, J. Clark, J. Lange, R. O’Shaughnessy, and L. Cadonati, “RUNMON-RIFT: Adaptive configuration and healing for large-scale parameter inference”, *Astron. Comput.* **42**, 100664 (2022) DOI:10.1016/j.ascom.2022.100664.

Chapter 4

AUTOMATIC HEALING FOR RIFT ANALYSES

This chapter contains contents from the published work

R. Udall, J. Brandt, G. Manchanda, A. Arulanandan, J. Clark, J. Lange, R. O’Shaughnessy, and L. Cadonati, “RUNMON-RIFT: Adaptive configuration and healing for large-scale parameter inference”, *Astron. Comput.* **42**, 100664 (2022)
DOI:10.1016/j.ascom.2022.100664.

My contributions to this work include the development of the software used, analyses for Figure 4.1, writing, and editing.

4.1 Introduction

Since the first gravitational wave detection GW150914 [1], the Advanced Laser Interferometer Gravitational Wave Observatory (LIGO) [2] and Virgo [3, 4] detectors have continued to discover gravitational waves (GW) from coalescing binary black holes (BBHs) and neutron stars [5–12]. From the point at which data is collected, many computational analyses are required to render it into information of astrophysical interest, including detector characterization [13], calibration and data cleaning [14, 15], candidate identification [16–18], noise estimation [19], and parameter estimation [20–22]. For the small number of observations reported through GWTC-3 (approximately 90 over 3 observing runs), these analyses could be monitored by individual humans to identify and remedy any problems that can occur. However, as detector sensitivity improves the number of observations and thus inferences increases (potentially hundreds in O4 alone [23]), saturating the ability of individual humans to carefully curate each analysis individually, such that automation will be necessary to correct common problems in future observing runs. This problem is especially acute for parameter inference, which this chapter will focus on, though automation schemes have also been implemented for other types of analysis, (see for example [24–27]). The most salient comparison to this software is Asimov [28], which has been developed for LVK analyses using RIFT and other parameter inference pipelines, and was used in GWTC-2.1 and GWTC-3 [7, 8]. This software has features in common with RUNMON-RIFT, most importantly monitoring software for project level management and automatic resubmission. However, notable

features of our work—most importantly railing-correction and node exclusion—are not presently implemented in Asimov.

Parameter inference for gravitational waves is generally done within the Bayesian analysis framework. For many possible configurations of parameters which contribute to the gravitational wave (see Section 4.2 for details) likelihood values are computed—in this case using approximate models of waveform behavior [29, 30]—and are combined algorithmically with prior expectations to generate posterior distributions which describe the probability of various configurations. For this analysis to be robust, it generally requires at least $O(10^6)$ likelihood evaluations, which may be computationally expensive. Various methods exist to sample these distributions efficiently, but all are of substantial complexity, and are run primarily on supercomputing clusters. In turn, this complexity allows for many potential issues, both in the settings of the algorithm and the operation of the software, which may drastically reduce the pace of analysis.

In this paper, we discuss a newly developed Python package, RUNMON-RIFT¹, which seeks to address a number of such problems in inferences performed using RIFT [31], one of the parameter estimation (PE) pipelines to interpret events in GWTC-1 [5], GWTC-2 [6], GWTC-2.1 [7], and GWTC-3 [8], as well as many individual events [9–12]. RUNMON-RIFT (and RIFT more generally) is geared primarily towards use on the LIGO Data Grid, a collection of independently operated computing clusters running HTCondor [32–34] with a common software environment and identity access management system, though RUNMON-RIFT also sees some use on the Open Science Grid [35, 36] via the LDG interface to it.

The challenges faced by software in scientific computing are highly context dependent, and parameter estimation software is no exception. However, some issues are common in many gravitational wave inference pipelines, and we shall focus on discussing these, with solutions tailored to the specific circumstances of RIFT. Large-scale parameter inference is frequently bedeviled by computing issues which are, as an individual matter, relatively straightforward to address, but which are, taken together, very difficult to resolve systematically. Notable examples of this behavior include misconfigured nodes (e.g. the node’s hardware is incompatible with the system software distributions), transient issues (e.g. a node loses contact with the filesystem for a period while a job is attempting to run), and system wide configuration errors (e.g. expired authentication tokens, or filesystem errors).

¹Available at <https://pypi.org/project/runmonitor-RIFT/>

We introduce tools for managing such issues, both by immediately continuing the progress of a job, and by providing infrastructure to proactively avoid them and provide information about their origin to cluster administrators.

Another common issue with parameter inference is “railing”: an artificially narrow prior range that constrains the extent of the posterior distribution in a physical parameter, significantly skewing the final result. For many practical reasons, PE inference pipelines adopt narrow prior ranges based on expectations informed both by experience and by any additional information, such as the output of a detection pipeline which identified the event originally. This is imperfect, however, especially when analysis is being done in bulk and available person hours to identify optimal settings are limited. We implement a mechanism for correcting this automatically; this algorithm works best with RIFT for reasons which will be described in more detail in Section 4.2, but could be broadly adapted for any PE software.

This chapter is organized as follows. In Section 4.2 we review the RIFT parameter inference engine. We begin with the core functionalities of RUNMON-RIFT, including its logging and tools it implements which dramatically decrease the amount of person-hours required to ensure a workflow’s completion. This includes a discussion of common error modes, and of a prototypical computing issue which RUNMON-RIFT helped overcome. We then describe how we identify potential ‘railing’ in our posterior distribution, associated with artificially narrow boundaries, and we introduce an adaptive method to extend these parameter-space boundaries. Finally, we discuss a toy model to demonstrate how RUNMON-RIFT can go beyond reactive workflow management, and proactively ensure that the computational pool used by a workflow is less likely to contain transient computing issues. Section 4.3 demonstrates the automated healing of these runs in both a stereotypical case of railing and for our computing issues toy model.

4.2 Methods

RIFT Review

A coalescing compact binary in a quasicircular orbit can be completely characterized by its intrinsic and extrinsic parameters. By intrinsic parameters we refer to the binary’s masses m_i , spins, and any quantities characterizing matter in the system. For simplicity and reduced computational overhead, in this work we provide examples of parameter inference which assume all compact object spins are aligned with the orbital angular momentum; however, the techniques introduced in our study are

not specific to any specific set of parameters or dimension. By extrinsic parameters we refer to the seven numbers needed to characterize its spacetime location and orientation. We will express masses in solar mass units (M_\odot), and dimensionless nonprecessing spins in terms of cartesian components aligned with the orbital angular momentum $S_{i,z}$. We will use $\boldsymbol{\lambda}, \boldsymbol{\theta}$ to refer to intrinsic and extrinsic parameters, respectively.

RIFT [31] consists of a two-stage iterative process to interpret gravitational wave data d via comparison to predicted gravitational wave signals $h(\boldsymbol{\lambda}, \boldsymbol{\theta})$. In the first stage, denoted by ILE (Integrate Likelihood over Extrinsic parameters), for each λ_α from some proposed “grid” $\alpha = 1, 2, \dots, N$ of candidate parameters, RIFT computes a marginal likelihood

$$\mathcal{L}(\boldsymbol{\lambda}) \equiv \int \mathcal{L}_{full}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (4.1)$$

from the likelihood $\mathcal{L}_{full}(\boldsymbol{\lambda}, \boldsymbol{\theta})$ of the gravitational wave signal in the multi-detector network, accounting for detector response, and extrinsic parameters prior $\pi(\boldsymbol{\theta})$; see the RIFT chapter for a more detailed specification. In the second stage, denoted by CIP (Construct Intrinsic Posterior), RIFT performs two tasks. First, it generates an approximation to $\mathcal{L}(\boldsymbol{\lambda})$ based on its accumulated archived knowledge of marginal likelihood evaluations $(\lambda_\alpha, \mathcal{L}_\alpha)$. This approximation can be generated by Gaussian processes, random forests, or other suitable approximation techniques. Second, using this approximation, it generates the (detector-frame) posterior distribution

$$p(\boldsymbol{\lambda}) = \frac{\mathcal{L}(\boldsymbol{\lambda}) \pi(\boldsymbol{\lambda})}{\int d\boldsymbol{\lambda} \mathcal{L}(\boldsymbol{\lambda}) \pi(\boldsymbol{\lambda})}, \quad (4.2)$$

where prior $\pi(\boldsymbol{\lambda})$ is the prior on intrinsic parameters like mass and spin. The posterior is produced by performing a Monte Carlo integral: the evaluation points and weights in that integral are weighted posterior samples, which are fairly resampled to generate conventional independent, identically-distributed “posterior samples.” For further details on RIFT’s technical underpinnings and performance, see [20, 31, 37].

Parameter inference analyses generally require many configuration details, notably including prior assumptions and the amount of data to be analyzed. Most relevant to this work is the fact that, for computational efficiency, the priors adopted are generally targeted to cover a limited range of mass and luminosity distance most likely to enclose the true source parameters, with initial ranges chosen motivated by search results. A second critical setting is the starting frequency of the waveform’s

dominant quadrupole mode. For an inspiralling binary at early times, this frequency is twice the orbital frequency. Because the orbital frequency at the last stable orbit decreases with mass, for binaries with a large detector-frame mass a conventional starting frequency like $f_{lower} = 20\text{Hz}$ is too high: the waveform model doesn't permit it. Furthermore, generation of waveforms with higher modes must start at a reduced frequency ($f_{min} = 2f_{lower}/L_{max}$), in order that no mode's initial frequency is above the fiducial starting frequency (i.e., no mode starts in band). A third critical setting is the amount of data to analyze, or "segment length". As the relevant starting frequency or mass decreases, the amount of data needed to be analyzed increases. Misidentification of any of these settings can cause a cascade of changes. For example, a mis-adapted mass prior might artificially exclude low masses, requiring a re-evaluation of the relevant segment length.

RUNMON-RIFT Introduction

RUNMON-RIFT is a Python package to monitor and manage runs, including correcting common failure modes encountered. At its core, RUNMON-RIFT consists of tools to assemble and manage a lightweight run tracking log. In addition to monitoring the queuing system (HTCondor) logs, RUNMON-RIFT includes generic tools to parse, query, and even edit RIFT's internal files. This allows it to interpret log outputs for identification of pathological behaviors, and edit configuration files according to custom algorithms. A daemon will periodically use these tools to update status on each job under its purview, and, using the archived run logs, the RUNMON-RIFT suite can quickly assemble reports on run status, including measures of convergence. Moreover, being aware of the workflow's status and being able to edit the workflow and even RIFT settings, RUNMON-RIFT can adapt to issues arising with the host cluster, or individual machines upon it², in a fashion that's minimally disruptive to ongoing analysis. RUNMON-RIFT's "healing" functions provide unique capability to handle ubiquitous challenges arising in large-scale parameter inference calculations. In this work, we will illustrate three such operations. First, we'll consider healing parameter "railing", a generic issue associated with user mis-specified priors. Second, we'll demonstrate how RUNMON-RIFT can respond to a transient cluster issue, here exemplified by problems with GPU use. Finally, we'll show how RUNMON-RIFT can efficiently block use of undesired computing

²A given cluster may have many different machines, with varying behaviors due to hardware architecture, utilization protocols, and the like. We adopt the terminology "machine" to reflect HTCondor's internal attribute designation, but they may also be variously known as nodes or sites—machines on the primary LDG cluster, for example, have the naming convention `node###.cluster.lidas.cit`.

machines (e.g., identified by job failures or even slow past performance).

Managing Jobs

In its simplest manifestation, RUNMON-RIFT implements a run index with operational metadata. LDG clusters feature a web-facing directory which may be accessed from a browser, and in this directory a file structure is generated, in which the user's run are organized by what event they are running on. For each run, there are a series of text files containing information about the run such as the name, location, number of completed iterations, and convergence details; RUNMON-RIFT includes a set of utilities which allow the user to parse these conveniently. A daemon is used to automatically analyze the workflows which are registered to its database, and updates the aforementioned metadata accordingly. Thus, we have operational information on all ongoing runs, allowing us to quickly identify potential problems and characterize overall progress, both critically important when working with many often heterogeneous analyses simultaneously.

RUNMON-RIFT can provide fixes for some of the many other issues which can prevent progress on a run. These issues can be conveniently flagged by the code, by the use of specific return values from the two key routines(ILE and CIP). Alternatively, RUNMON-RIFT can parse the codes' output and HTCondor logs, to identify and characterize issues that can cause the run to fail. Quite frequently, these issues are transient in the sense that they are not caused by the structure of the analysis itself, but rather by incompatibilities which occur only in certain parts of a heterogeneous computing pool.

A prototypical example is GPU utilization: RIFT uses GPUs to improve efficiency, but a given cluster may include many separate machines, often varying dramatically in age. Updates in some standard computing environments to the software library used by RIFT (CUDA[38], called via the Python library CUPY[39]) rendered it incompatible with some machines on a popular cluster, which in turn led to an extremely high failure rate, forcing the user to resubmit repeatedly until a job would be lucky enough to land on a compatible machine. This example motivated the introduction of automated resubmission within RUNMON-RIFT, so that up time for runs could be maintained with minimal user intervention, and during times when users would not be available. Furthermore, it inspired the machine exclusion algorithm described in Section 4.2. Ultimately, the root of the problem was identified after a number of months, and usage of software libraries was altered to remove the issue

at its source for runs on shared IGWN filesystems, but the resubmission mechanics remain necessary for the highly heterogeneous OSG pool. Table 4.1 displays several additional errors which RUNMON solves in an analogous manner. Many result from instabilities in cluster filesystems which change frequently and are unavoidable for the end user. When a transient is sufficiently common and results in a consistent error message in the Python runtime, RIFT is edited to provide standard error codes for these errors, such that RUNMON-RIFT may more easily identify and cope with them.

Error Description	Recognition Method	Machine Excludable?	Fixed at Origin?
CUDA Compute Incompatibility	Custom Error Code	Yes	Yes for IGWN Clusters
Interpreter Runtime Error	Standard HTCondor Error Code	No	No
Interpreter Not Found Error	Standard HTCondor Error Code	Yes	Yes
XLAL File Transient	Output Parsing	No	Yes

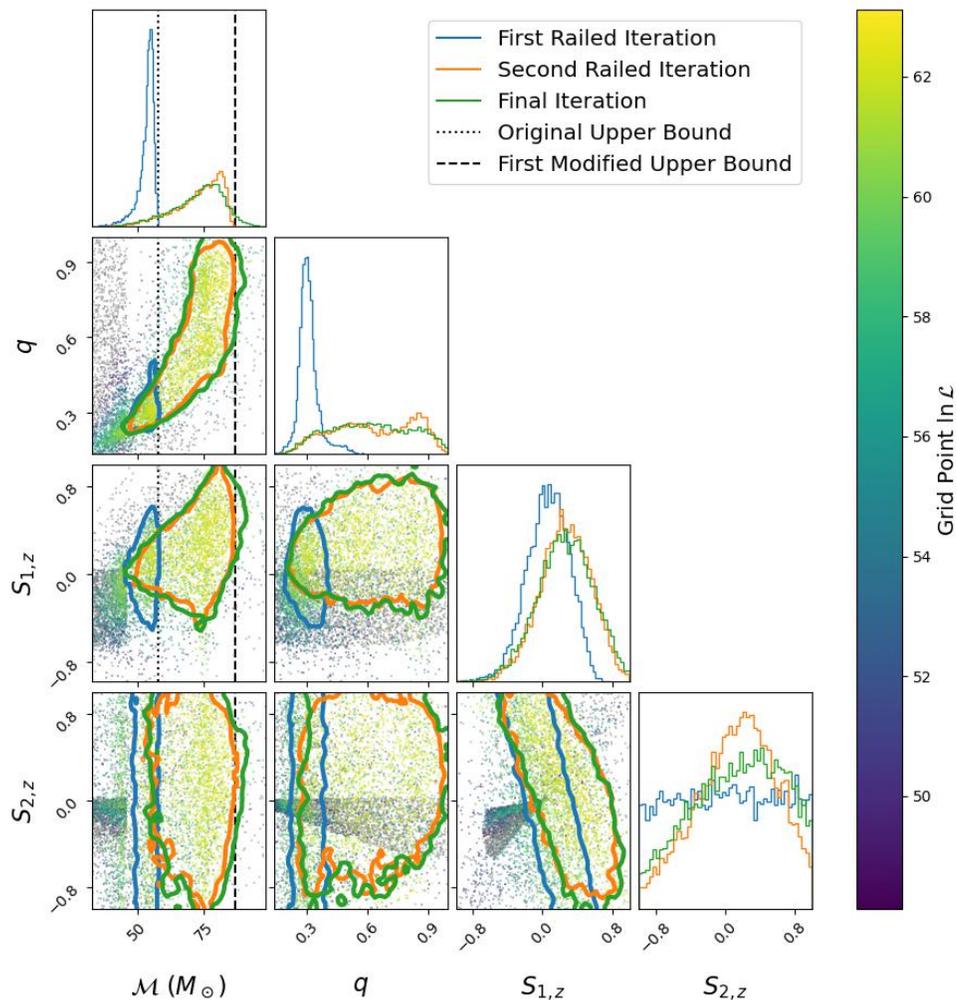
Table 4.1: Examples of common errors

Healing Parameter Railing

The priors $\pi(\theta)$, $\pi(\lambda)$ over extrinsic and intrinsic parameters are usually proportional to some a priori separable function. In each variable, the range and normalization of the prior is over some finite range. Sometimes the boundaries are physical and absolute, for example when integrating over phase or sky location. However, for variables like luminosity distance or mass, the user usually adopts upper and lower bounds for computational convenience, to bound the overall time to solution, centered on a weakly-informed guess. When performing large-scale inference, these arbitrary bounds are not-infrequently mis-specified, and the posterior is artificially constrained, “railing” against one or more boundaries.

Railing can be identified by having significant posterior support immediately adjacent to one of the arbitrary prior bounds. The blue curve in Figure 4.1 shows an example of a railed posterior. Quantitatively we identify it as follows. Suppose for parameter x we have a sampled posterior $\hat{p}(x)$. We divide the prior range $\pi(x)$ into bins of equal width, with 20 bins as the default. We then determine the density of the posterior in the first and last bin. If this density exceeds some threshold—we

Figure 4.1: Analysis for the GW190602_175927 [40], an event where standard parsing of internal low-latency estimates results in initially incorrect boundaries in \mathcal{M}_c . Contours are shown for iterations which triggered RUNMON-RIFT’s railing test, as well as the final result, and vertical lines show the boundaries at the iterations where railing was found. The final boundary occurs substantially to the right of the plot’s extent in \mathcal{M}_c . Colored points are individual points on the grid, with the heat map corresponding to likelihood.



use 0.03 by default—in either bin, we evaluate this bin to be railed, and extend the corresponding boundary accordingly. We emphasize this definition applies only to parameters with user-specified boundaries; parameters which have absolute limits, like the mass ratio $q = m_2/m_1$, do not rail against those limits, as more extreme values are unphysical.

RIFT’s intrinsic boundaries only impact the second (CIP) phase [Eq. (4.2)] of

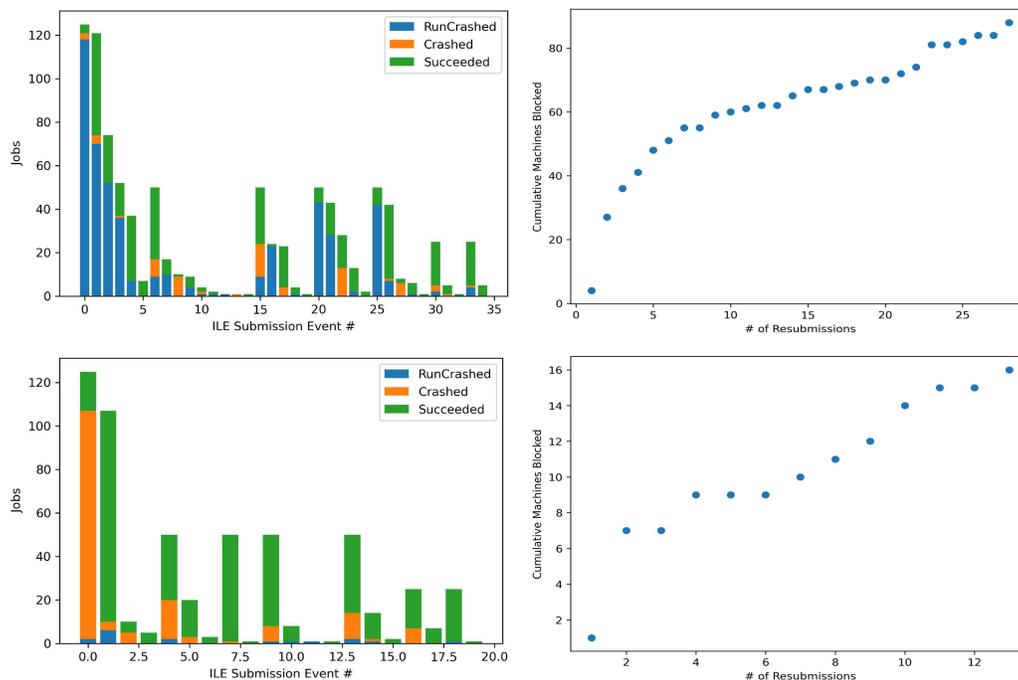
the cycle, not the phase in which the likelihood values are computed [Eq. (4.1)]. RUNMON-RIFT’s daemon reads the output of each CIP phase (these take the form of sampled posteriors over the intrinsic parameters $\hat{p}(\lambda)$), and applies our algorithm to identify railing if appropriate. If it is detected, the job is removed from the cluster, and RUNMON-RIFT changes the boundaries which are found to be railing. If lower bound railing is detected, R_{lower} is mapped to $(1 - m)R_{lower}$ and if upper bound railing is detected R_{upper} is mapped to $(1 + m)R_{upper}$, where m is 0.5 by default. The same job may then be resumed, without having to create a new workflow³. At least two iterations are required before the extended range can be fully explored, but since railing is normally identified early in the analysis, the job almost always has the ability to explore the extended range. In cases where a job does fail to explore this range, methods exist for creating new jobs which effectively continue the runs, though this does require manual intervention. Other parameter estimation methods—notably MCMC and nested sampling methods—require more complex methods of intervention and continuation to achieve similar results, such as the initialization of helper analyses, though it is also potentially feasible to automate these.

Problematic Machine Exclusion

Computing clusters frequently suffer from transient errors, usually triggered by some change in the computing environment, which take the form of everything from failed software dependencies to difficulties with file transfers. Since the specific conditions required to trigger these transients may only occur for certain computational tasks, on certain machines, or with specific settings, they may be difficult to track and address. Also problematic is the phenomena of “black hole” machines: a colloquial term referring to when a machine will accept a job, but that job will quickly fail due to something inherent to the machine (such as the aforementioned GPU incompatibility). Specifically, if large numbers of jobs are submitted in parallel (as is the case for high-throughput computing tasks, such as the ILE stage of RIFT), the scheduler will attempt to assign them in bulk. If the available computational resources are limited, then some fraction will be assigned and the rest will occupy the next spots in the queue. Accordingly, a single machine experiencing some transient error may fail immediately, then accept another job from this queue. In sufficiently low resource situations, this may result in the entire parallel content of an high-throughput computing job failing on a single machine.

³In the language of HTCCondor, we resubmit the dag

Figure 4.2: The behavior of ILE jobs and number of machines blocked as a function of the associated ILE submission batch, for high and low error rate scenarios. Left panels: Histogram of the number of failed jobs versus submission attempt. Jobs labelled Succeeded complete normally; jobs noted as RunCrashed have intentionally failed, due to landing on a set of pre-selected target hosts; and jobs labelled Crashed fail for other reasons, not infrequently associated with problematic or misconfigured host machines. Right panel: Cumulative number of blocked machines as a function of rescue attempt.



To mitigate the impact of problematic machines, RUNMON-RIFT allows for tracking machines associated with known transient errors, and provides a tool for instructing the relevant HTCondor jobs to exclude these machines from matchmaking consideration. The information about which machines should be excluded is shared across all jobs managed by a given daemon, and thus is propagated quickly for all of a user's active jobs. Logging of these machines and their associated failure modes also offers a collated set of data to provide system administrators when troubleshooting issues, such that the root problem may be identified and addressed, at which point it is straightforward to remove the restrictions the daemon imposed upon the pool.

4.3 Results

Healing

Figure 4.1 depicts a prototypical example of railing, along with the correction produced by RUNMON-RIFT. The pipeline constructor for the event in question, GW190602[40], produced a railed prior boundary in chirp mass \mathcal{M}_c when taking the metadata of the event’s initial detection as input. Accordingly, it required careful and tedious human intervention, lest any run be completely ruined. The use of RUNMON-RIFT may be seen to alleviate this in the progression of the results seen in Figure 4.1. The plot in question is an example of a corner plot—displaying both one-dimensional histograms of individual parameters, as well as their two-dimensional joint parameters, so that correlation may be understood and diagnosed. RIFT corner plots also include colored points to show the likelihood values of the underlying grid, with the yellow spectrum colors corresponding to the highest likelihood points, and the purple spectrum colors corresponding to the lowest likelihood points. Here the posterior after the first iteration of the workflow (the blue curve) may be seen to rail at the upper boundary in \mathcal{M}_c which was set by the pipeline constructor (the dotted black line). Notably, this distribution also has an erroneous posterior distribution in mass ratio q , due to the correlation of this parameter with the erroneous chirp mass. RUNMON-RIFT then automatically increased the upper boundary to the value seen in the dashed black line, and continued the sampling process. After a number of iterations, the posterior had shifted to the orange curve, which may be seen to also rail (though to a lesser degree) against the modified upper bound, and so RUNMON-RIFT modified the upper bound once more, well past limits of the corner plot ($\mathcal{M}_c \approx 135M_\odot$). Final sampling then brought the posterior to an unraild distribution (the green curve), which agrees with the results presented for this event in [7].

Runcrasher

To demonstrate the principle of machine exclusion, we construct an artificial scenario with known parameters and behavior which mimics the transient errors known to occur on computational clusters. In particular, we insert a step into the standard ILE portion of the workflow which tests the machine upon which the job lands, and produces a failure if that machine’s name satisfies certain constraints (e.g. if the last digit is 5). RUNMON-RIFT included this failure code as one of the known transient values, and the exclusion system was triggered accordingly. We conducted tests under various constraints, reflecting the varying incidence rate of transients. This

construct also naturally results in the aforementioned “black hole” machines when submission incidentally occurs at a time of high resource usage.

The results of this artificial scenario and corresponding intervention are shown in Figure 2. The bar charts indicate the behavior of individual machines under high and low transient incidence rates respectively. Transients are separated into two types: those which are caused by the runcrasher, which behave in a predictable manner and are subject to machine exclusion, and those which are caused by miscellaneous other transients, which are not well characterized and not subject to machine exclusion (for the runs in question these transients primarily involve accessing certain public files). The scatter plots show how many machines are actively excluded for each of these corresponding submission batches.

A number of features may be noted in these plots. Firstly, the submission events for which the total number of jobs increase are those submissions which occur at the beginning of a new iteration. The total duration of iterations for which the same numbers of jobs are submitted decrease correspondingly in the high-incidence case (the number of jobs submitted per iteration varies over the duration of the underlying workflow to improve its efficiency). Similarly, the relative proportion of errors which are due to the unmodeled transients increases. The low incidence case shows somewhat similar behavior, though it is also more strongly subject to low number statistics, as it is relatively rare to hit a failure machine in the first place.

When including analysis of the number of machines submitted, one may also see the expected trend: initial jobs result in substantially more blocked machines, while later jobs run in a cleaner pool, and thus are less likely to simultaneously interact with many error-triggering machines. One may also note that there are submission events (submission event 20, for example) for which most jobs fail but very few machines are blocked—this is an example of the aforementioned black-hole machine phenomenon. These unfortunately take longer to root out, since it gets progressively harder to filter through the pool when it is already mostly successful, but integration of these error lists across multiple runs mean that in a high usage context (if one has 10 runs simultaneously, for example) it is very feasible to fully eliminate problematic machines.

4.4 Conclusion

We have presented our Runmonitor for RIFT (RUNMON-RIFT), a utility which greatly aids in the operation of this inference pipeline. Gravitational wave science

is in an exciting time, with a rapid pace of discovery and exponentially increasing data to analyze. In this context, it is critical that the time required to complete a parameter estimation task, and the time the user spends actively monitoring and intervening in that task, be minimized. By introducing centralized diagnostic tools, RUNMON-RIFT makes it much easier for a user to check the status of their active jobs. Automated resubmission for known transient issues greatly decreases the amount of time a user spends actively engaging with the workflow (in particularly hostile computing environments this decrease may be up to an order of magnitude), and machine exclusion allows one to tailor the pool utilized towards the machines which will actually work consistently, decreasing restarts and improving efficiency for all cluster users. Monitoring of railing allows for aggressive (and hence efficient) initial settings, while also reducing the need for producing new workflows during exploratory phases of analysis.

References

- ¹B. Abbott et al. (The LIGO Scientific Collaboration and the Virgo Collaboration), “Observation of Gravitational Waves from a Binary Black Hole Merger”, *Physical Review Letters* **116**, 061102–+ (2016) DOI:10.1103/PhysRevLett.116.061102.
- ²LIGO Scientific Collaboration, J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, and et al., “Advanced LIGO”, *Classical and Quantum Gravity* **32**, 074001, 074001 (2015) DOI:10.1088/0264-9381/32/7/074001.
- ³T. Accadia and et al, “Virgo: a laser interferometer to detect gravitational waves”, *Journal of Instrumentation* **7**, P03012 (2012).
- ⁴F. Acernese et al. (VIRGO), “Advanced Virgo: a second-generation interferometric gravitational wave detector”, *Class. Quant. Grav.* **32**, 024001 (2015) DOI:10.1088/0264-9381/32/2/024001.
- ⁵B. P. Abbott et al. (LIGO Scientific, Virgo), “GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs”, *Phys. Rev. X* **9**, 031040 (2019) DOI:10.1103/PhysRevX.9.031040.
- ⁶R. Abbott et al. (LIGO Scientific, Virgo), “GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run”, *Phys. Rev. X* **11**, 021053 (2021) DOI:10.1103/PhysRevX.11.021053.
- ⁷R. Abbott et al. (LIGO Scientific, VIRGO), “GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half

- of the third observing run”, *Phys. Rev. D* **109**, 022001 (2024) doi:10.1103/PhysRevD.109.022001.
- ⁸R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run”, *Phys. Rev. X* **13**, 041039 (2023) doi:10.1103/PhysRevX.13.041039.
- ⁹B. P. Abbott et al. (LIGO Scientific, Virgo), “Properties of the binary neutron star merger GW170817”, *Phys. Rev. X* **9**, 011001 (2019) doi:10.1103/PhysRevX.9.011001.
- ¹⁰R. Abbott et al. (LIGO Scientific, Virgo), “Properties and Astrophysical Implications of the 150 M_{\odot} Binary Black Hole Merger GW190521”, *Astrophys. J. Lett.* **900**, L13 (2020) doi:10.3847/2041-8213/aba493.
- ¹¹R. Abbott et al. (LIGO Scientific, Virgo), “GW190412: Observation of a Binary-Black-Hole Coalescence with Asymmetric Masses”, *Phys. Rev. D* **102**, 043015 (2020) doi:10.1103/PhysRevD.102.043015.
- ¹²R. Abbott et al. (LIGO Scientific, KAGRA, VIRGO), “Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences”, *Astrophys. J. Lett.* **915**, L5 (2021) doi:10.3847/2041-8213/ac082e.
- ¹³D. Davis et al. (LIGO), “LIGO detector characterization in the second and third observing runs”, *Class. Quant. Grav.* **38**, 135014 (2021) doi:10.1088/1361-6382/abfd85.
- ¹⁴L. Sun et al., “Characterization of systematic error in Advanced LIGO calibration in the second half of O3”, (2021) doi:10.48550/arXiv.2107.00129.
- ¹⁵F. Acernese et al. (VIRGO), “Calibration of advanced Virgo and reconstruction of the detector strain $h(t)$ during the observing run O3”, *Class. Quant. Grav.* **39**, 045006 (2022) doi:10.1088/1361-6382/ac3c8e.
- ¹⁶S. A. Usman et al., “The PyCBC search for gravitational waves from compact binary coalescence”, *Class. Quant. Grav.* **33**, 215004 (2016) doi:10.1088/0264-9381/33/21/215004.
- ¹⁷K. Cannon et al., “GstLAL: A software framework for gravitational wave discovery”, *SoftwareX* **14**, 100680 (2021) doi:10.1016/j.softx.2021.100680.
- ¹⁸T. Adams, D. Buskulic, V. Germain, et al., “Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era”, *Class. Quant. Grav.* **33**, 175012 (2016) doi:10.1088/0264-9381/33/17/175012.
- ¹⁹T. B. Littenberg and N. J. Cornish, “Bayesian inference for spectral estimation of gravitational wave detector noise”, *Phys. Rev. D* **91**, 084034 (2015) doi:10.1103/PhysRevD.91.084034.

- ²⁰D. Wysocki, R. O’Shaughnessy, J. Lange, et al., “Accelerating parameter inference with graphics processing units”, *Physical Review D* **99**, 084026, 084026 (2019) DOI:10.1103/PhysRevD.99.084026.
- ²¹J. Veitch, V. Raymond, B. Farr, et al., “Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library”, *Phys. Rev. D* **91**, 042003 (2015) DOI:10.1103/PhysRevD.91.042003.
- ²²G. Ashton, M. Hübner, P. D. Lasky, et al., “Bilby: a user-friendly bayesian inference library for gravitational-wave astronomy”, *The Astrophysical Journal Supplement Series* **241**, 27 (2019) DOI:10.3847/1538-4365/ab06fc.
- ²³B. Abbott et al (LIGO Scientific Collaboration and the Virgo Collaboration), “Prospects for Localization of Gravitational Wave Transients by the Advanced LIGO and Advanced Virgo Observatories”, *Living Reviews in Relativity* **19** (2016) DOI:10.1007/lrr-2016-1.
- ²⁴D. A. Brown, K. Vahi, M. Taufer, et al., “Reproducing gw150914: the first observation of gravitational waves from a binary black hole merger”, *Computing in Science & Engineering* **23**, 73–82 (2021) DOI:10.1109/mcse.2021.3059232.
- ²⁵K. Vahi, M. Rynge, G. Papadimitriou, D. A. Brown, R. Mayani, R. F. da Silva, E. Deelman, A. Mandal, E. Lyons, and M. Zink, “Custom Execution Environments with Containers in Pegasus-enabled Scientific Workflows”, (2019) DOI:10.48550/arXiv.1905.08204.
- ²⁶L. Singer et al., *Gwcelery*, <https://rtd.igwn.org/projects/gwcelery/en/latest/>, 2025.
- ²⁷R. Essick, P. Godwin, C. Hanna, et al., “iDQ: statistical inference of non-gaussian noise with auxiliary degrees of freedom in gravitational-wave detectors”, *Machine Learning: Science and Technology* **2**, 015004 (2020) DOI:10.1088/2632-2153/abab5f.
- ²⁸D. Williams, J. Veitch, M. L. Chiofalo, et al., “Asimov: A framework for coordinating parameter estimation workflows”, *J. Open Source Softw.* **8**, 4170 (2023) DOI:10.21105/joss.04170.
- ²⁹S. Ossokine et al., “Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation”, *Phys. Rev. D* **102**, 044055 (2020) DOI:10.1103/PhysRevD.102.044055.
- ³⁰C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, “Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries”, *Phys. Rev. D* **102**, 064002 (2020) DOI:10.1103/PhysRevD.102.064002.
- ³¹J. Lange, R. O’Shaughnessy, and M. Rizzo, “Rapid and accurate parameter inference for coalescing, precessing compact binaries”, (2018) DOI:10.48550/arXiv.1805.10457.

- ³²D. Thain, T. Tannenbaum, and M. Livny, “Distributed computing in practice: the Condor experience”, *Concurrency Comput. Pract. Exp.* **17**, <https://chtc.cs.wisc.edu/doc/condor-practice.pdf>, 323–356 (2005) DOI:10.1002/cpe.938.
- ³³E. M. Fajardo, J. M. Dost, B. Holzman, et al., “How much higher can htcondor fly?”, *Journal of Physics: Conference Series* **664**, 062014 (2015).
- ³⁴B. Bockelman, M. Livny, B. Lin, and F. Prezl, “Principles, technologies, and time: the translational journey of the htcondor-ce”, *Journal of Computational Science* (2020) DOI:10.1016/j.jocs.2020.101213.
- ³⁵R. Pordes, D. Petravick, B. Kramer, et al., “The open science grid”, in *J. phys. conf. ser.* Vol. 78, 78 (2007), p. 012057, DOI:10.1088/1742-6596/78/1/012057.
- ³⁶I. Sfiligoi, D. C. Bradley, B. Holzman, et al., “The pilot way to grid resources using glideinwms”, in *2009 wri world congress on computer science and information engineering*, Vol. 2, 2 (2009), pp. 428–432, DOI:10.1109/CSIE.2009.950.
- ³⁷J. Lange, “RIFT’ing the Wave: Developing and applying an algorithm to infer properties gravitational wave sources”, <https://repository.rit.edu/theses/10586/> (Rochester Institute of Technology, Aug. 2020).
- ³⁸S. Cook, Cuda programming: a developer’s guide to parallel computing with gpus, 1st (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2012), DOI:10.5555/2430671.
- ³⁹R. Okuta, Y. Unno, D. Nishino, et al., “Cupy: a numpy-compatible library for nvidia gpu calculations”, in *Proceedings of workshop on machine learning systems (learningsys) in the thirty-first annual conference on neural information processing systems (nips)* (2017).
- ⁴⁰B. Abbott et al. (The LIGO Scientific Collaboration), “Open data from the first and second observing runs of advanced ligo and advanced virgo”, *SoftwareX* **13**, 100658 (2021) DOI:10.1016/j.softx.2021.100658.

Part II

Glitch Modeling

GLITCH MODELING WITH BILBY

As previously mentioned in Section 2.5, GW detectors experience terrestrial transients known as “glitches,” at a rate of approximately one per minute [1, 2]. These affect our astrophysical analyses both by masquerading as GW transients themselves, and by biasing the inference of astrophysical properties when they overlap a true GW signal. Here I will describe the glitch mitigation methods currently deployed by the LVK, and the formalism which will be used in Chapters 6, 7, and 8 where I present improvements on those methods.

5.1 Glitch subtraction

When it is known that a glitch overlaps an astrophysical GW event, the glitch must be mitigated to prevent the biasing of the PE conducted. The standard method for this in the LVK is to model the glitch by itself, then subtract this model from the observed data, ideally leaving only Gaussian noise and the GW [3–5]. This does not account for any modeling uncertainties, and hence may introduce statistical biases even when the model itself is an accurate description of the glitch. To begin, I will describe two of the methods used by the LVK in O3 to perform glitch subtraction, before describing my work in performing more statistically robust mitigation for a class of glitches called scattered light glitches.

BAYESWAVE

The **BAYESWAVE** algorithm [6–8] models signals as a sum of Morlet-Gabor (sine-Gaussian) wavelets. These form an over-complete basis over the space of time domain functions, and so it is guaranteed that a sum of these may perfectly model any glitch morphology. To do so, however, may require a significant number of individual wavelets. To solve this problem, **BAYESWAVE** uses a transdimensional algorithm, in which the number of wavelets is included as a parameter within the sampling [6]. In order to be tractable, there is an Occam penalty applied for increasing numbers of wavelets, which in turn means that **BAYESWAVE** is best suited to glitches which are well characterized by a relatively small number of wavelets [6, 7].

There has also been work to incorporate CBC modeling into **BAYESWAVE** as well [8,

9]. This allows the inference of CBC parameters which have marginalized over the possible glitch realizations, and hence are robust to statistical uncertainties. However, because the wavelet model is flexible enough to capture signal power as well as glitch power, `BAYESWAVE` still relies on coherence between two or more detectors to confidently identify the astrophysical component of the signal. Accordingly, in cases where the data of astrophysical interest depend principally on a single detector, or where the glitch requires many wavelets to model properly, `BAYESWAVE` may not be the most effective available method.

GWSUBTRACT

`GWSUBTRACT` models a contribution to the strain data from some witness data channel, mediated by an unknown transfer function [5, 10]. This is appropriate when the nature of a glitch is well understood, and reflected in an identified witness channel for which a simple transfer function may be determined. To determine the transfer function, `GWSUBTRACT` divides the data into frequency bands and measures the cross power spectrum in each [10]. For glitches which are well witnessed, subtracting the convolution of the witness and the transfer function should give the correctly de-glitched data. However, this process does not incorporate any measurement uncertainties, and requires that there be significant amounts of off-source data in order to make an accurate measurement. This limits its application only to certain classes of glitches.

5.2 Scattered light glitches

One of the most prominent classes of glitch in ground based GW detectors are scattered light glitches. As the name implies, and as I have briefly referenced in 2.5 and Section 2.5, scattered light glitches are caused by light scattering within the detector. From the main beam, some fraction of the light will scatter off of reflective elements, then scatter again back into the main beam's path, hence interacting with optical elements. In doing so, the scattered light will follow a optical path with different length, and hence acquire a phase shift relative to the main beam [11]:

$$\phi_{sc}(t) = \frac{4\pi}{\lambda}(x_0 + \delta x_{opt}(t)) = \phi_0 + \delta\phi_{sc}(t) \quad (5.1)$$

Here x_0 is the path length between the reference position of the scatterer, $\delta x_{opt}(t)$ is the motion of the scatterer with respect to that path length, λ is the wavelength of the laser (1064 nm), and ϕ_0 and $\delta\phi_{sc}(t)$ are the phase shifts from the respective

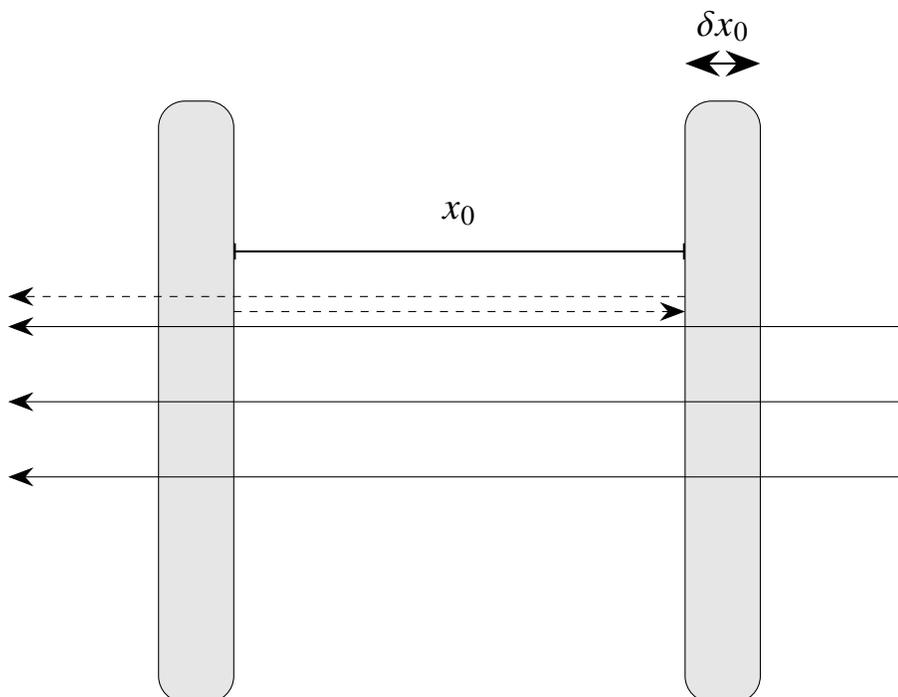


Figure 5.1: A schematic of scattered light. Solid lines represent the main beam path, while the dashed lines represent the path of the scattered light. This supposes two parallel planes, a configuration which corresponds to slow scattering, where the scattering is between the electrostatic drive on the reaction chain mass and the test mass itself [12]. Other scattering configurations scatterers that are not parallel also occur, such as in the case of scattering from baffles.

displacement terms. An example of this configuration is rendered in Section 5.1.

This phase offset results in a measured strain [11]:

$$h_{sc}(t) = G \sin\left(\frac{4\pi}{\lambda}(x_0 + \delta x_{opt}(t))\right), \quad (5.2)$$

where G captures the various factors which control the amplitude of the scattering. For the purposes of the following discussion this is an unknown constant, and I will show how it may be inferred implicitly in upcoming sections.

Now, the next stage in modeling scattering is to introduce some form for $\delta x_{opt}(t)$, which is the relevant term for determining observed scattering glitches. This will depend in turn upon the characteristics of the scatterer, and the nature of the driving force, and so I will discuss two prominent cases.

Slow scattering

Slow scattering is, as the name implies, scattering generated by a relatively slow driving force. To be more precise, slow scattering is driven by motion in the microseism frequency band of $f_{mod} \approx 0.1 - 0.3$ Hz, which is typically generated by ocean waves and sea storms [11–13]. Modeling this motion as a monochromatic driving force, and the scatterer as a simple harmonic oscillator, then the time dependent displacement is

$$\delta x_{opt}(t) = \delta x_0 \sin(2\pi f_{mod}(t - t_c)) \quad (5.3)$$

where δx_0 is the amplitude of the oscillations and t_c controls the phase of the oscillations, which is modeled as an offset in time for convenience. Substituting this and collecting constant terms gives

$$h_{sc}(t) = G \sin \left[\frac{4\pi}{\lambda} \left\{ x_0 + \delta x_0 \sin \left(2\pi f_{mod}(t - t_c) \right) \right\} \right] \quad (5.4)$$

$$= G \sin \left[C \sin \left(2\pi f_{mod}(t - t_c) \right) + \phi_0 \right], \quad (5.5)$$

where $C = \frac{4\pi\delta x_0}{\lambda}$ and $\phi_0 = \frac{4\pi x_0}{\lambda}$ is a constant phase offset.

From this, making a stationary phase approximation that

$$f(t) = \frac{1}{2\pi} \frac{d\phi}{dt} = \frac{4\pi\delta x_0 f_{mod}}{\lambda} \cos \left(2\pi f_{mod}(t - t_c) \right), \quad (5.6)$$

it is possible to predict the frequency of the resulting glitch as a function of time. Now connecting a bit with the operation of ground based GW detectors, astrophysical data are generally limited to 20 Hz and above, due to the rapid rise of technical noise sources as frequency decreases below this point [14]. Accordingly, for this scattered light glitch to affect observations, it must be the case that (taking $t = t_c$ to maximize, and 0.3 Hz as a reasonable microseism peak frequency)

$$\delta x_0 \geq \frac{(20 \text{ Hz})(1064 \text{ nm})}{4\pi(0.3 \text{ Hz})} \approx 5 \times 10^3 \text{ nm}, \quad (5.7)$$

but Soni et al [12] describes the ground motion on April 5, 2019 as “extremely high”, and it barely meets this threshold. So, then, how can slow scattering routinely create scattered light glitches in the astrophysical analysis frequency band?

The resolution to this is that slow scattering may see light which goes through multiple bounces before recombining with the main beam path, as described in Soni et al [12]. Each bounce introduces another δx_0 , and hence the strain noise goes to

$$h_{sc}(t) = G \sin \left[kC \sin \left(2\pi f_{mod}(t - t_c) \right) + \phi_0 \right] \quad (5.8)$$

for k bounces, and thus k arches at the peak of the sinusoid. The frequency also increases accordingly, with

$$f(t) = \frac{4\pi k \delta x_0 f_{mod}}{\lambda} \cos\left(2\pi f_{mod}(t - t_c)\right). \quad (5.9)$$

And so even for smaller amounts of ground motion it is possible that the glitch may enter the astrophysical analysis band.

Putting this all together, it is also useful to reparameterize in terms of the peak (harmonic) frequency

$$f_{harm} = \frac{4\pi \delta x_0 f_{mod}}{\lambda}, \quad (5.10)$$

such that the expression for the strain is now (summing over the number of bounces)

$$h_{sc}(t) = \sum_{k=0}^N G_k \sin\left[\frac{f_{harm} k}{f_{mod}} \sin\left(2\pi f_{mod}(t - t_c)\right) + \phi_k\right]. \quad (5.11)$$

One additional consideration is where to truncate the model of the glitch, since as written it continues infinitely. In practice slow scattering glitches will evolve over time, as the ground motion is not stationary over long time scales. To handle this, the phase evolution of the model is truncated at points where $f(t) = 0$, separating out an individual “stack” of arches. In the software implementation the user may control the number such stacks which will be modeled together, which is important for longer signals which may overlap multiple stacks over the course of many seconds. Additionally, after truncation the model will likely not have returned to the DC strain it began at. A tukey window is accordingly applied on each side of the modeled signal, to enable Fourier transforming (which requires periodicity at the boundaries).

Figure 5.2 demonstrates what this glitch looks like in a spectrogram of the data, including the multiple arches generated by repeated bounces off of the scattering surface. In Section 5.3 I will show how I have implemented this model into `BILBY` to allow the joint inference of the CBC along with the properties of these glitches. In Chapter 6 and Chapter 7 I will apply these methods in a variety of cases.

Fast scattering

Another potential form of scattering is fast scattering, in which the scatterer is driven at two frequencies: one in the microseism band, and the other the anthropogenic band of 1–5 Hz. Accordingly, into 5.2 we substitute

$$\delta x_{opt}(t) = \delta x_0 \sin(2\pi f_0(t - t_0)) + \delta x_1 \sin(2\pi f_1(t - t_1)), \quad (5.12)$$

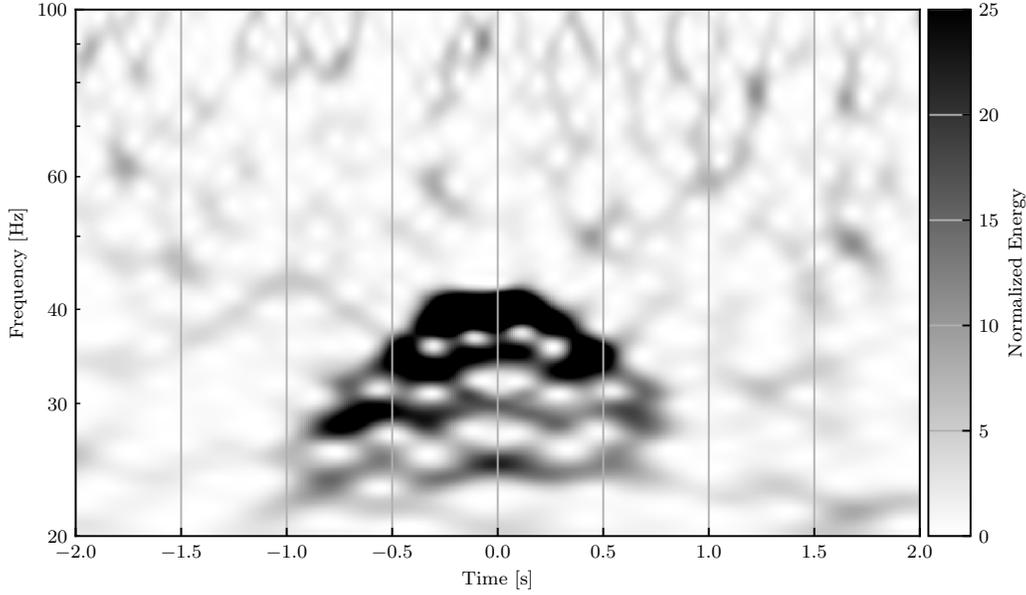


Figure 5.2: An example of slow scattering, as generated by the scattering model I have described. Of note are the distinct arches, each spaced by equal frequency intervals but otherwise sharing properties such as width and central time.

which gives a frequency expression of

$$f(t) = \frac{4\pi}{\lambda} \left[\delta x_0 f_0 \cos\left(2\pi f_0(t - t_0)\right) + \delta x_1 f_1 \cos\left(2\pi f_1(t - t_1)\right) \right] \quad (5.13)$$

$$= a_0 f_0 \cos\left(2\pi f_0(t - t_0)\right) + a_1 f_1 \cos\left(2\pi f_1(t - t_1)\right). \quad (5.14)$$

Fast scattering does not have multiple bounces, because the arm cavity baffles it is associated with do not form resonances as stable as those seen with slow scattering off of the electrostatic drive [13]. We can approximate the maximum frequency by setting $2\pi f_0(t - t_0) = 2\pi f_1(t - t_1) = 0 \pmod{2\pi^1}$, which yields $f_{max} = a_0 f_0 + a_1 f_1$. It is the combination of these two driving forces which allows periods of scatterer motion significant enough to bring this glitch into the astrophysical frequency band. The final expression for the strain due to fast scattering is:

$$h_{sc}(t) = A \sin \left[a_0 \sin \left(2\pi f_0(t - t_0) \right) + a_1 \sin \left(2\pi f_1(t - t_1) \right) + \phi \right]. \quad (5.15)$$

The process of truncation, while similar to that of slow scattering, is complicated by the lack of general analytic solutions for $f(t) = 0$. Instead, in the software

¹This is approximation because unless f_0 and f_1 are rational multiples there is no guarantee this will occur exactly within the duration of the scattering

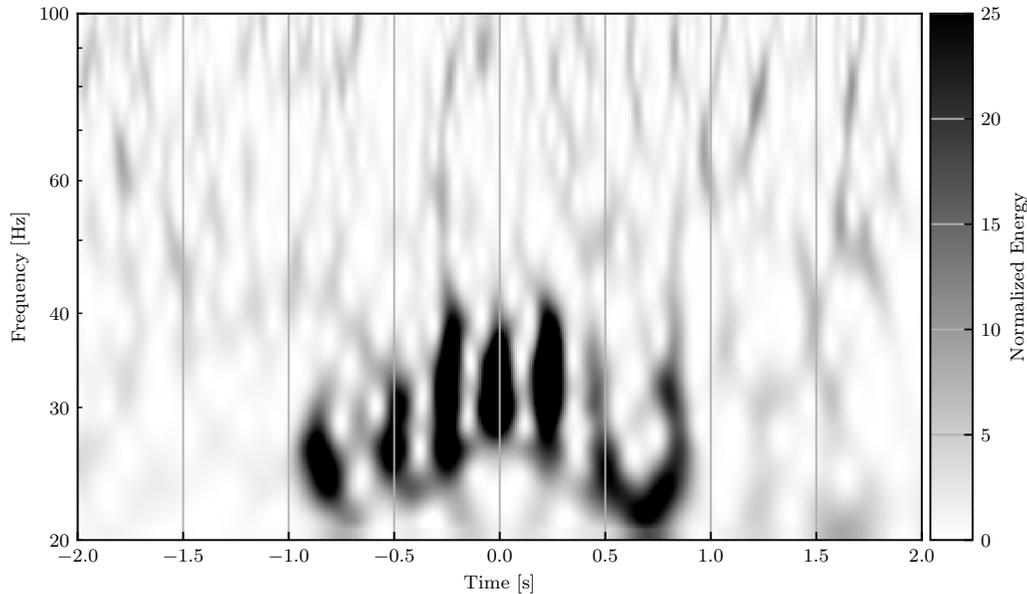


Figure 5.3: An example of fast scattering, as generated by the scattering model I have described. Unlike the slow scattering case, there are distinct peaks which recur at a higher frequency (roughly 4.2 Hz by construction in this case) while the overall peak frequencies are bounded by a sinusoidal envelope with width similar to that of slow scattering example in Figure 5.2.

implementation the user provides an approximate window, and the nearest zero crossings are determined analytically, at which point the truncation proceeds as described above.

Figure 5.3 demonstrates what this glitch looks like in a spectrogram of the data, including the multiple peaks bounded by a more slowly developing envelope, which are characteristic of fast scattering glitches.

5.3 Glitch modeling with **BILBY**

As I discussed in Section 3.2, **BILBY** [15] is a software package developed to perform Bayesian inference on GW transients. Accordingly, with some modification it can also accommodate the analysis of glitches. To do so I have developed a software package **BILBY_GLITCH** which facilitates the necessary conversions, and includes a number of glitch models including the two described above.

Joint glitch-CBC likelihoods

As was mentioned in Section 3.2, the standard gravitational wave likelihood (Equation 3.10) is

$$\ln \mathcal{L}(d|\theta) = -\frac{1}{2}(d - h(\theta)|d - h(\theta)). \quad (5.16)$$

It is straightforward to accommodate this to an added glitch model $g(\lambda)$

$$\ln \mathcal{L}(d|\theta, \lambda) = -\frac{1}{2}(d - h(\theta) - g(\lambda)|d - h(\theta) - g(\lambda)). \quad (5.17)$$

This adjusted likelihood assumes that the data is well described as some stationary Gaussian noise, our modeled terrestrial glitch $g(\lambda)$ with glitch parameters λ , and the astrophysical signal $h(\theta)$.

The use of a joint likelihood allows for the joint inference of λ and θ . Joint inference permits the exploration of any degeneracies between the glitch and CBC models. If left statistically unaccounted for—by the choice of a single point-realization—these degeneracies can result in the biasing of the CBC parameters. With joint inference, λ can instead be marginalized over, such that the CBC parameters account for any uncertainties introduced by the glitch model.

An interesting extension occurs in the use of distance marginalization, which is a computational technique for accelerating CBC inference [16–18]. Gravitational wave amplitudes scale inversely with distance, while for fixed detector frame mass the phase evolution is unchanged. For some reference distance D_0 , we have that

$$h(\theta, D_L) = h(\theta, D_0) \left(\frac{D_0}{D_L} \right), \quad (5.18)$$

where here θ is the normal binary parameters except D_0 . And so optimal SNR and the data-template inner product will satisfy scaling laws

$$(h(\theta, D_L)|h(\theta, D_L)) = \left(\frac{D_0}{D_L} \right)^2 (h(\theta, D_0)|h(\theta, D_0)) \quad (5.19)$$

and

$$(d|h(\theta, D_L)) = \left(\frac{D_0}{D_L} \right) (d|h(\theta, D_0)) \quad (5.20)$$

exactly. Following both the notation and procedure in Thrane and Talbot [16], I will notate $\kappa^2(D_L) = (d|h(\theta, D_L))$ and $\rho_{opt}^2(D_L) = (h(\theta, D_L)|h(\theta, D_L))$ for given reference waveform parameters θ , as well as \mathcal{Z}_N for the noise evidence $\frac{-\langle d|d \rangle}{2}$. The normal expression for log-likelihood from Section 3.1 can be expanded in scaling terms

$$\ln \mathcal{L}(\theta, D_L) = \frac{1}{2} \left[2 \left(\frac{D_0}{D_L} \right) \kappa^2(D_0) - \left(\frac{D_0}{D_L} \right)^2 \rho_{opt}^2(D_0) \right] + \ln \mathcal{Z}_N. \quad (5.21)$$

The likelihood may thus be marginalized over:

$$\begin{aligned} & \int dD_L \mathcal{L}(\theta, D_L) \pi(D_L) \\ &= \int dD_L \exp \left\{ \left(\frac{D_0}{D_L} \right) \kappa^2(D_0) - \frac{1}{2} \left(\frac{D_0}{D_L} \right)^2 \rho_{opt}(D_0)^2 + \mathcal{Z}_n \right\} \pi(D_L), \end{aligned} \quad (5.22)$$

so that

$$\ln \mathcal{L}_{marg} = \ln \mathcal{Z}_N + \ln \mathcal{L}_D(\kappa^2, \rho_{opt}), \quad (5.23)$$

where

$$\mathcal{L}_D(\kappa^2, \rho_{opt}) = \int dD_L \exp \left\{ \left(\frac{D_0}{D_L} \right) \kappa^2(D_0) - \frac{1}{2} \left(\frac{D_0}{D_L} \right)^2 \rho_{opt}(D_0)^2 \right\} \pi(D_L). \quad (5.24)$$

Because $\mathcal{L}_D(\kappa^2, \rho_{opt})$ depends only on these inner products, as opposed to the details of the waveform itself, one may numerically compute a lookup table over these variables for a given luminosity distance prior. This modified likelihood then takes these inner products as inputs, rather than the waveforms directly.

The above procedure is used for CBC only analysis, but in fact it also extends very naturally to joint inference of CBCs and glitches. Expanding out the joint likelihood gives

$$\begin{aligned} \ln \mathcal{L}(\theta, D_L, \lambda) &= -\frac{1}{2} \left(d - h(\theta, D_0) \left(\frac{D_0}{D_L} \right) - g(\lambda) \middle| d - h(\theta, D_0) \left(\frac{D_0}{D_L} \right) - g(\lambda) \right) \\ &= -\frac{1}{2} \left[(d|d) - \left(\frac{D_0}{D_L} \right) (d|h(\theta, D_0)) - (d|g(\lambda)) - \left(\frac{D_0}{D_L} \right) (h(\theta, D_0)|d) \right. \\ &\quad \left. + \left(\frac{D_0}{D_L} \right)^2 (h(\theta, D_0)|h(\theta, D_0)) + (h(\theta, D_0)|g(\lambda)) - (g(\lambda)|d) \right. \\ &\quad \left. + \left(\frac{D_0}{D_L} \right) (g(\lambda)|h(\theta, D_0)) + (g(\lambda)|g(\lambda)) \right], \end{aligned} \quad (5.26)$$

which is structurally similar to 5.22. As such, we can use the same lookup tables constructed for 5.24 by performing the substitution:

$$\kappa_g^2 = -(d|h(\theta, D_L)) + (g(\lambda)|h(\theta, D_L)). \quad (5.27)$$

Given this, the likelihood will have the full expression:

$$\ln \mathcal{L}_{marg} = \ln \mathcal{Z}_N + \mathcal{L}_D(\kappa^2, \rho_{opt}) - \frac{1}{2} (g(\lambda)|g(\lambda)) + (g(\lambda)|d). \quad (5.28)$$

Model implementation

Given this likelihood, `BILBY` can analyze any glitch which has a parametric model $g(\lambda)$ available. `BILBY_GLITCH` includes the two scattering models described above, as well as a single wavelet model equivalent to that used by `BAYESWAVE`. Unlike `BAYESWAVE`, transdimensional methods—for which the number of glitch parameters λ is allowed to vary—are not well developed in `BILBY`, although progress is being made [19].

I have also explored alternate models, including models which allow for the scatterer motion entering 5.2 to be inferred from a witness channel, or constructed from input wavelets. However, these models need further development to be fully robust, and so are left for future work.

References

- ¹S. Soni et al. (LIGO), “LIGO Detector Characterization in the first half of the fourth Observing run”, *Class. Quant. Grav.* **42**, 085016 (2025) doi:10.1088/1361-6382/adc4b6.
- ²D. Davis et al. (LIGO), “LIGO detector characterization in the second and third observing runs”, *Class. Quant. Grav.* **38**, 135014 (2021) doi:10.1088/1361-6382/abfd85.
- ³R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run”, *Phys. Rev. X* **13**, 041039 (2023) doi:10.1103/PhysRevX.13.041039.
- ⁴R. Abbott et al. (LIGO Scientific, VIRGO), “GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run”, *Phys. Rev. D* **109**, 022001 (2024) doi:10.1103/PhysRevD.109.022001.
- ⁵D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. Di Renzo, and G. Ashton, “Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run”, *Class. Quant. Grav.* **39**, 245013 (2022) doi:10.1088/1361-6382/aca238.
- ⁶N. J. Cornish and T. B. Littenberg, “BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches”, *Class. Quant. Grav.* **32**, 135012 (2015) doi:10.1088/0264-9381/32/13/135012.
- ⁷N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, “BayesWave analysis pipeline in the era of gravitational wave observations”, *Phys. Rev. D* **103**, 044006 (2021) doi:10.1103/PhysRevD.103.044006.

- ⁸S. Hourihane, K. Chatziioannou, M. Wijngaarden, D. Davis, T. Littenberg, and N. Cornish, “Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data”, *Phys. Rev. D* **106**, 042006 (2022) doi:10.1103/PhysRevD.106.042006.
- ⁹E. Payne, S. Hourihane, J. Golomb, R. Udall, D. Davis, and K. Chatziioannou, “Curious case of GW200129: Interplay between spin-precession inference and data-quality issues”, *Phys. Rev. D* **106**, 104017 (2022) doi:10.1103/PhysRevD.106.104017.
- ¹⁰D. Davis, T. J. Massinger, A. P. Lundgren, J. C. Driggers, A. L. Urban, and L. K. Nuttall, “Improving the Sensitivity of Advanced LIGO Using Noise Subtraction”, *Class. Quant. Grav.* **36**, 055011 (2019) doi:10.1088/1361-6382/ab01c5.
- ¹¹T. Accadia et al., “Noise from scattered light in Virgo’s second science run data”, *Class. Quant. Grav.* **27**, edited by F. Ricci, 194011 (2010) doi:10.1088/0264-9381/27/19/194011.
- ¹²S. Soni et al. (LIGO), “Reducing scattered light in LIGO’s third observing run”, *Class. Quant. Grav.* **38**, 025016 (2020) doi:10.1088/1361-6382/abc906.
- ¹³S. Soni, J. Glanzer, A. Effler, et al., “Modeling and reduction of high frequency scatter noise at LIGO Livingston”, *Class. Quant. Grav.* **41**, 135015 (2024) doi:10.1088/1361-6382/ad494a.
- ¹⁴E. Capote, L. Dartez, and D. Davis, “Technical noise, data quality, and calibration requirements for next-generation gravitational-wave science”, *Class. Quant. Grav.* **41**, 185001 (2024) doi:10.1088/1361-6382/ad694d.
- ¹⁵G. Ashton, M. Huebner, P. Lasky, et al., “BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy”, *Astrophys. J. Suppl.* **241**, 27 (2019) doi:10.3847/1538-4365/ab06fc.
- ¹⁶E. Thrane and C. Talbot, “An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models”, *Publ. Astron. Soc. Austral.* **36**, [Erratum: *Publ. Astron. Soc. Austral.* 37, e036 (2020)], e010 (2019) doi:10.1017/pasa.2019.2.
- ¹⁷L. P. Singer and L. R. Price, “Rapid Bayesian position reconstruction for gravitational-wave transients”, *Phys. Rev. D* **93**, 024013 (2016) doi:10.1103/PhysRevD.93.024013.
- ¹⁸G. Ashton et al., “BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy”, *Astrophys. J. Suppl.* **241**, 27 (2019) doi:10.3847/1538-4365/ab06fc.
- ¹⁹H. Tong et al., “Transdimensional Inference for Gravitational-wave Astronomy with Bilby”, *Astrophys. J. Suppl.* **276**, 50 (2025) doi:10.3847/1538-4365/ad9deb.

*Chapter 6***BAYESIAN MODELING FOR SLOW SCATTERING**

This chapter contains contents from the published work

R. Udall and D. Davis, “Bayesian modeling of scattered light in the LIGO interferometers”, *Appl. Phys. Lett.* **122**, 094103 (2023) DOI:10.1063/5.0136896.

My contributions to this work include writing and all BILBY analyses.

6.1 Introduction

The Laser Interferometer Gravitational-wave Observatory (LIGO) includes a pair of ground-based interferometers: LIGO Hanford and LIGO Livingston, designed to be sensitive to gravitational waves from astrophysical sources [1]. The LIGO detectors are dual-recycled Michelson interferometers with Fabry-Perot arm cavities 4 kilometers in length. Recent observations alongside the Virgo [2] and KAGRA [3] detectors have identified 90 gravitational-wave signals from the mergers of compact objects [4–6].

The sensitivity of LIGO is limited both by persistent noise sources, such as quantum noise and thermal noise [2, 7], as well as transient noise sources, such as earthquakes and thunder [8–11]. Transient noise is manifested in detector data as short-duration bursts of excess power, commonly referred to as “glitches.” Glitches can create challenges for gravitational-wave analyses by either mimicking gravitational-wave signals (preventing identification of signals [9, 12–16]) or by overlapping detected events (corrupting further analysis [17–21]).

One of the most common sources of transient noise in recent observing runs is from “scattered light” [22, 23]. Imperfections in the mirror surfaces can lead to light scattered off of the main beam path in the interferometer. If the light reflects off of a moving object and then rejoins the main beam path, this adds an additional time dependent phase shift to the observed light. Although scattered light may occur randomly, the impact of the source of reflection is well understood. If the scattering occurs from one of the test masses, the most likely source of reflection is the suspension system supporting the test mass or nearby optical components.

The most common source of reflection for scattered light in LIGO during the third recent observing run (O3) was from light reflected off of the reaction mass that is located directly behind the test masses at the end of each arm of the interferometer [22]. Figure 6.1 shows the optical path of the scattered light from this source of reflection. The reaction chain mass contains an electrostatic drive (ESD) that is used to control the motion of the main test mass. The highly reflective gold coatings used as part of the ESD allowed scattered light to reflect a large number of times, creating scattered light glitches with many arches. The rate of scattered light glitches from this source and others was so high that $\approx 20\%$ of gravitational-wave signals in O3 overlapped in time with such glitches [4, 5]. Improvements to this source of noise were made in the latter part of the third observing run by introducing reaction chain tracking, but these improvements were not able to eliminate scattered light completely.

In order to prevent glitches present in the data from biasing estimates of the source properties of gravitational-wave signals, any glitches that are nearby or directly overlap gravitational-wave signals are subtracted from the data [17, 21]. Using information from the motion of the suspension system, one can estimate the frequencies of scattered light glitches [24–29]. However, since these monitors only approximate the motion of the specific surface that is reflecting the light, it is difficult to precisely model individual scattered light glitches using these techniques. Other methods exist to subtract scattered light glitches, but these tools either use weakly modelled techniques [17, 30–33] or require that there is a relevant sensor which accurately witnesses the source of the glitch [21, 34, 35]. As scattered light glitches have a relatively long duration, rapidly changing frequency, and only witnesses that approximate the true source of the glitching, modelling and subtracting these glitches with these additional techniques is challenging. The high rate of these glitches, combined with a well-understood model for the relevant instrumental mechanism that creates the glitches, makes scattered light glitches ideal candidates to instead model using analytic methods.

In this chapter, we demonstrate how we can utilize the simple harmonic motion of the Advanced LIGO suspension system to model scattered light glitches. We then use Bayesian inference to estimate the parameters of individual scattered light glitches. This tool can be used to model the time dependent phase and amplitude of scattered light glitches for both detector characterization and to improve the Bayesian inference of the source properties of detected gravitational-wave events.

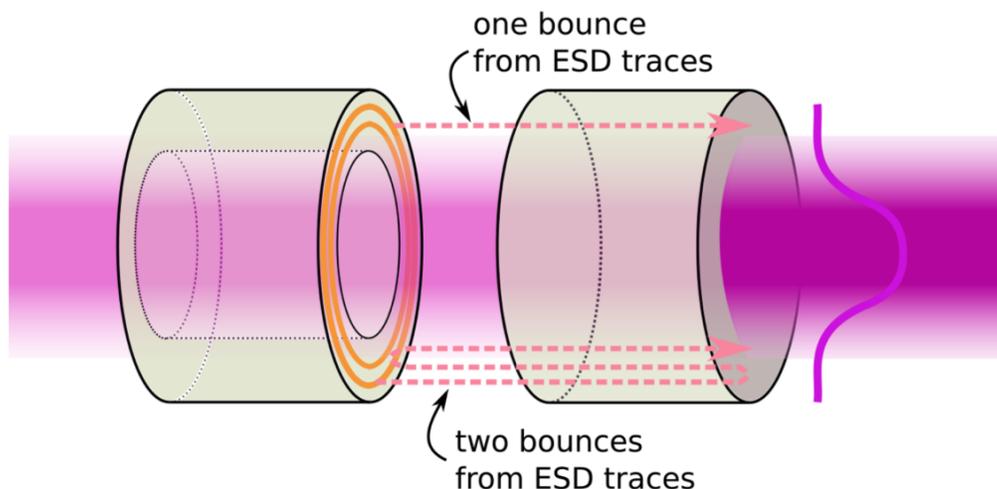


Figure 6.1: A diagram of the test mass and reaction chain mass showing the path that of scattered light. Some light is transmitted through the test mass and reflects off components of the electrostatic drive (ESD), illustrated as orange circles. As the direction of reflection is parallel with the main beam path, it is possible for the light to be reflected multiple times before rejoining the main beam path. Figure is reproduced from [Soni et al. \[22\]](#) ©IOP Publishing. Reproduced with permission. All rights reserved

6.2 Scattering Model

Generically, the excess strain noise, $h(t)$, produced by scattering of light with wavelength λ and of some amplitude A can be modeled based on the motion of the relevant surface, $x(t)$ [24]:

$$h(t) = A \sin \left[\frac{4\pi}{\lambda} x(t) + \phi \right]. \quad (6.1)$$

Therefore, the instantaneous frequency of this strain noise is

$$f(t) = \left| \frac{1}{2\pi} \frac{d}{dt} \left(\frac{4\pi}{\lambda} x(t) + \phi \right) \right| = \left| 2 \frac{v(t)}{\lambda} \right|. \quad (6.2)$$

This equation can be used to predict the frequency of the related scattered light glitches. If the scattered light is reflected N times, the scattered light glitch frequency is N times the frequency predicted by Equation 6.2. An example of LIGO data containing a scattered light glitch and the predicted glitch frequencies can be seen in Figure 6.2. This relationship has been used extensively to help identify the source of scattered light that is observed in gravitational-wave interferometers [24–29]; if the motion of a component in the detector predicts the same frequency evolution as is observed in the strain data, it is likely that the source is nearby this component.

If we assume that the surface that reflects the scattered light is a simple harmonic oscillator with frequency f_{mod} , then the motion of the optic is

$$x(t) = C \sin(2\pi f_{\text{mod}}t). \quad (6.3)$$

and then setting $f_{\text{harm}} = \max[f(t)] = \frac{4\pi C}{\lambda} f_{\text{mod}}$, we arrive at our scattering equation: [36]

$$h(t) = A \sin \left[\frac{f_{\text{harm}}}{f_{\text{mod}}} \sin(2\pi f_{\text{mod}}t) + \phi \right]. \quad (6.4)$$

Physically, f_{harm} corresponds to the maximum frequency of the scattered light glitch, $1/f_{\text{mod}}$ is twice the duration of each glitch, and A is the amplitude of the strain induced by the glitch. Typically, the relevant frequencies of seismic motion responsible for scattered light glitches are in the microseismic ($\approx 1/6$ Hz) band, resulting in a scattered light glitch every 3 seconds and $f_{\text{mod}} \approx 1/6$ Hz.

6.3 Bayesian analysis

To proceed, we now include the previously mentioned complication to this picture: scattering events consist of many harmonics, appearing on top of each other with increasing harmonic frequency and decreasing amplitude, and so accurately characterizing the glitch and removing it from the data stream requires inclusion of all of these harmonics. The frequencies of each harmonic are integer multiples of each other, such that the frequency separation between any two adjacent harmonics is nearly constant; we denote this separation as f_{diff} . Similarly, the modulation frequency and central time are almost exactly constant. Accordingly, we model a collection of N arches using these relations, with perturbation terms to account for small variations between harmonics:

$$h(t) = \sum_{k=0}^N A_k \sin \left[\frac{f_{\text{harm},k}}{f_{\text{mod},k}} \sin(2\pi f_{\text{mod},k}(t - t_{c,k}) + \phi_k) \right] \quad (6.5)$$

Where for $k > 0$

$$f_{\text{harm},k} = f_{\text{harm},0} + k f_{\text{diff}} + \delta f_{\text{harm},k} \quad (6.6)$$

$$f_{\text{mod},k} = f_{\text{mod},0} + \delta f_{\text{mod},k} \quad (6.7)$$

$$t_{c,k} = t_{c,0} + \delta t_{c,k} \quad (6.8)$$

This model greatly increases the sampling efficiency by explicitly including these relationships between harmonics, thus substantially reducing the configuration space of glitch morphology which must be searched. Moreover, the relations in modulation

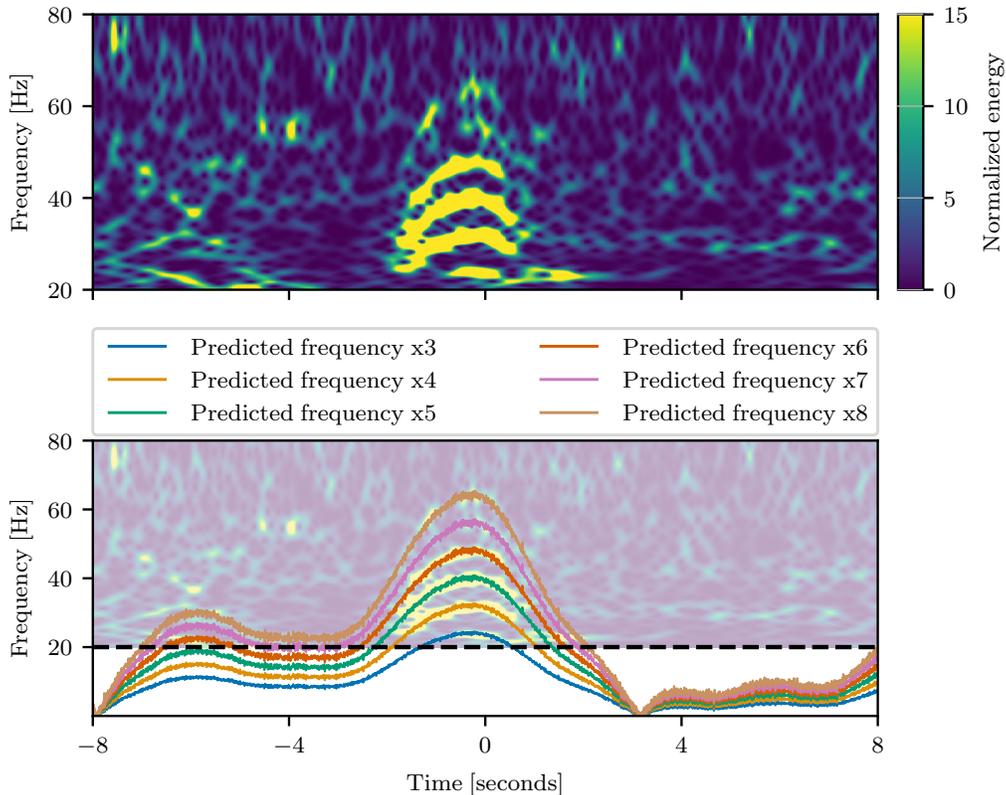


Figure 6.2: A comparison of data containing a scattered light glitch and the suspension motion data that is correlated with the glitch morphology. The top shows a spectrogram of data from LIGO Livingston during the second observing run showing an isolated scattered light glitch with many harmonics. The bottom shows the same spectrogram with lines overlaid showing the frequency of the scattering arches as predicted by the motion of the second stage of the suspension system holding one of the test masses. The dashed black line marks the minimum frequency of the spectrogram data. The suspension motion data has been multiplied by 1.25 to better match the observed arches. With this correction, there is clear agreement with the predicted and true frequencies of the scattering arch.

frequencies and central times are consistently almost exact, such that we can set $\delta f_{mod,k} = \delta t_{c,k} = 0$ without affecting the quality of the subtraction, further reducing the complexity of the sampling problem. While we have reasonably strong prior assumptions for the values of $f_{harm,0}$, $f_{mod,0}$, $t_{c,0}$, and f_{diff} , accurate subtraction requires sampling to produce the true posteriors. However, these prior assumptions allow us to set relatively tight prior bounds on these parameters, which improves the rate of sampling convergence. One limitation of this model is that the total number of harmonics must be fixed before beginning analysis of the data. When we

choose to fix the total number of harmonics to N , we describe this as the “ N arch model.” In a typical N arch inference, we wish to determine the values of $3N + 3$ parameters: the first arch is modeled with base parameters A_0 , $f_{mod,0}$, $f_{harm,0}$, $t_{c,0}$, and ϕ_0 , while each subsequent arch is modeled with the shared difference parameter f_{diff} , as well as fully independent amplitudes A_k and phases ϕ_k , and harmonic frequency perturbation $\delta f_{harm,k}$.

This formula describes an infinite series of arches, and so we apply a Tukey window [37] of width $1/2f_{mod}$, and windowing parameter $\alpha = 0.2$ centered about the arch sequence of interest.

Because this model has a large number of free parameters, we wish to infer posteriors on their true values, for which we turn to Bayesian inference [38]. Bayesian inference is built upon Bayes theorem, which for data d , parameters θ , and model \mathcal{M} is

$$p(\theta|d, \mathcal{M}) = \frac{p(d|\theta, \mathcal{M})p(\theta|\mathcal{M})}{p(d|\mathcal{M})} \quad (6.9)$$

The model \mathcal{M} consists of both the mapping from the parameters θ to the expected data, and the a priori beliefs about the parameters $p(\theta|\mathcal{M})$. More typically, we talk about the likelihood $\mathcal{L}(d|\theta) \propto p(d|\theta, \mathcal{M})$, and the priors $\pi(\theta)$. Following convention, we will always notate posteriors as $p(\theta)$ and priors as $\pi(\theta)$, for brevity. In this formulation we then have

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\pi(\theta)}{Z} \quad (6.10)$$

Where the evidence Z is the normalizing factor, such that

$$Z = \int \mathcal{L}(d|\theta)\pi(\theta)d\theta \quad (6.11)$$

The likelihood for transient behaviors in the data stream, both glitches and true gravitational-wave signals, is well known under the assumption of Gaussian stationary noise [39], and is given by

$$\ln \mathcal{L}(\theta) = -\frac{1}{2} \sum_k \left[\frac{|h_k(\theta) - d_k|^2}{\sigma_k^2} + \ln(2\pi\sigma_k^2) \right] \quad (6.12)$$

Where k denotes the frequency bin, $h_k(\theta)$ is the value of the template for θ at that frequency bin, d_k is the value of the data in that frequency bin, and σ_k is the value of the amplitude spectral density for the underlying noise in that frequency bin. Evaluation of the posterior must be done by numerical methods, typically by

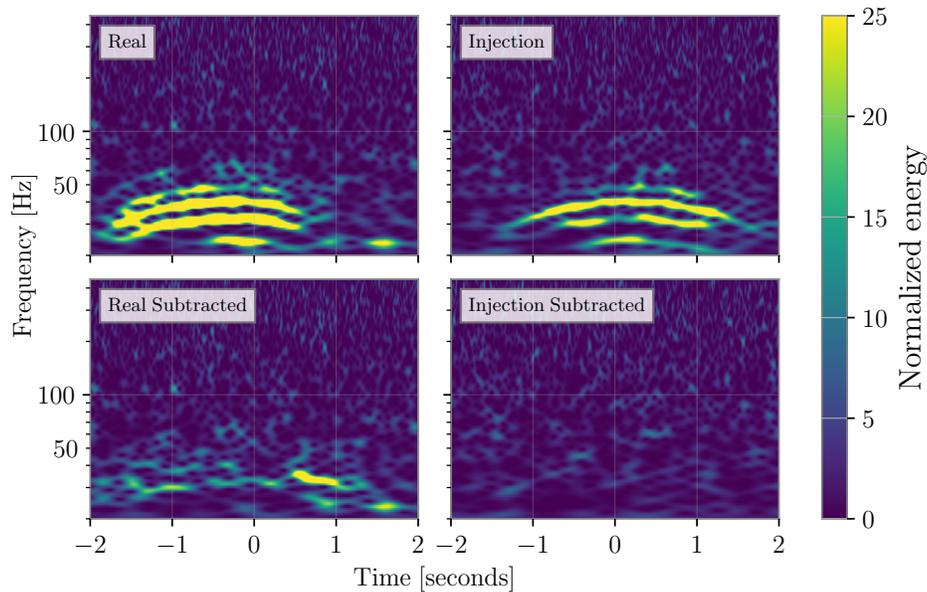


Figure 6.3: Here spectrograms are present for an example in data from LIGO Livingston during the second observing run (O2). The top left shows a real scattering event, corresponding to that plotted in Figure 6.2, which have six arches visible in the data. The bottom left shows the the subtraction of the reconstruction generated from maximum likelihood parameters for the inference of the scattering event, corresponding to the results plotted in the bottom left panel of Figure 6.4. Due to asymmetries in the scattering arch, as highlighted in Figure 6.2, this subtraction is not perfect. Next, on the top right, we inject these maximum likelihood parameters into Gaussian noise, generated according to the amplitude spectral density inferred for a stretch of data from the second observing run (O2). We then subtract the maximum likelihood parameters of the inference on this example, corresponding to the bottom right panel of Figure 6.4. Notably since the injection is perfectly symmetric, this subtraction is also highly effective.

sampling with either Markov Chain Monte Carlo (MCMC) [40] or nested sampling methods [41].

We perform sampling with the interface provided by Bilby [39], a library for Bayesian inference which is specially adapted for use with Gravitational Wave data. Bilby conveniently allows extension to custom models, and so we use this model in place of a standard gravitational-wave waveform template. Our sampling is done using the Dynesty sampling method [42], which is an implementation of the standard nested sampling method [41]. Briefly, nested sampling methods estimate the evidence by drawing points from the posterior, under a condition of monotonically increasing the minimum likelihood of the "live" set of points. The principal

challenge of nested sampling is to draw these points efficiently, since for models with more than a few dimensions the volume of the typical set is many orders of magnitude smaller than the volume of the prior space, and so naive sampling would be woefully inefficient. To do this more efficiently, we use differential evolution methods which draw further proposed points using a set of multiple proposals, each using information about the ensemble of live points at that time. Furthermore, it is desirable to choose priors which are sufficiently constrained as to not waste computational time, while also being sufficiently uninformed as to not bias the inference. Since the frequencies and central times may be roughly estimated by eye from spectrograms, we set uniform priors centered upon these estimates for $f_{harm,0}$, $f_{mod,0}$, f_{diff} , and $t_{c,0}$. Perturbations are centered on 0 with moderate width (for $\delta f_{harm,k}$, this width must be constrained to be less than f_{diff} , lest accidental degeneracies be introduced). Phases are estimated with a uniform prior over angles $[0, 2\pi)$, and amplitudes are estimated with log-uniform priors over 3 to 4 orders of magnitude, since their magnitudes are difficult to estimate from spectrograms.

The wall time of these analyses vary depending on the configuration, the amplitude of the glitch, and the complexity of the parameter space being explored, but are consistently less than one day under reasonable conditions. When run using a pool of 8 CPUs, the shortest run took approximately 20 minutes, while the longest took approximately 13 hours, which is typical for Bayesian inference problems with high dimensionality.

6.4 Injection tests and single arches

To begin model validation, we conducted four injection tests in Gaussian noise, drawn from the amplitude spectral density of the data surrounding the isolated scattering event analyzed below. These consisted of:

1. Injecting a scattered light glitch with three arches and recovering them with a three-arch model'
2. Injecting a scattered light glitch with three arches and recovering them with an eight-arch model;
3. Injecting a scattered light glitch with eight arches and recovering them with a three-arch model;
4. Injecting a scattered light glitch with eight arches and recovering them with an eight-arch model.

The left column of Figure 6.3 shows the injection of case 4, and its residual after subtraction of the model with the maximum likelihood parameters. These tests should allow verification of a number of important points. Firstly, we should be able to accurately recover injected values, even when analyzing only a subset of the arches (case three). Secondly, we should be able to distinguish the absence or presence of arches correctly, in cases 2 and 4, respectively. Finally, we should be able to subtract the recovered parameters and arrive at nearly Gaussian noise. To test whether the data is consistent with Gaussian noise, we use the test introduced in [43]. We find that, after subtracting the best fit parameters from our analysis of the data with a scattered light injection, the data is consistent with Gaussian noise at a p-value of 0.97. Figure 6.4 allows for a visualization of the accuracy of our recoveries and the distinguishability of the null case. True (injected) values are accurately recovered for all injected arches. Furthermore, when only three arches are injected, recovery of the other five finds amplitudes that push toward zero, whereas when eight arches are injected, each arch except the last is recovered at distinctly nonzero values. Thus, when applied to true data, we have a meaningful test for model validity, when varying the number of arches being modeled. For the eight-arch case, this conclusion is further supported by the spectrograms in Figure 6.3, where the result of the subtraction is data which is visually indistinguishable from Gaussian noise.

Next, we performed analysis of an isolated scattering event in O2 data, under both the assumption that it included only three arches and the assumption that it included up to eight arches. The right panels of Figure 6.4 show the amplitudes recovered in each of these cases. Of note is the fact that for the eight-arch case all eight arches have amplitude posteriors inconsistent with amplitude of zero, which in conjunction with the injection validity tests shows that there is strong evidence that at least eight arches are present in the data, despite the fact that only six may be clearly seen in the spectrogram. Also of note is that for the subset of arches considered, the three-arch model is in agreement with the eight-arch model, such that while subtraction under the three-arch assumption may be incomplete, it is not biased. The subtraction in the spectrogram is somewhat less rosy; while fairly effective, artifacts do remain, especially on the right edge of the arch. This may be traced to asymmetry within the scattering mechanism, which causes a breakdown of the model's assumption and validity in this regime. Inclusion of extra freedom in our model will likely be necessary to handle this. Possibilities include modeling $f_{harm,0}$ and A_k as first or second order polynomials in time, but this will require significant testing to reach

maturation, and so is deferred to further work.

6.5 Applications to real data

As previously mentioned, due to their width and frequency, scattered light glitches are the most common glitches to overlap true gravitational-wave signals. Although their morphology is distinct from morphology of the types of gravitational-wave signals that have been detected so far, meaning they are less likely to be misidentified as true gravitational waves than other signals [9, 12–16], they nonetheless may bias parameter estimation due to the inclusion of excess power in the analysis band [19, 21]. Modeled subtraction of glitch power allows for the cleaning of data near or on top of the signal, without requiring assumptions about the relationship between the observed strain data and a sensor witnessing the source of the glitch, especially with regards to the amplitudes of the harmonics. Joint inference of the glitch model and the gravitational-wave source properties, which we hope to investigate in a future work, will be required for confident subtraction of arches coincident with signals, but inference with restricted priors may still allow for glitch subtraction without affecting the underlying signal. Particularly useful is the modelled test for the presence or absence of arches not seen in spectrograms. This allows for checks of the effect of cleaning procedures, ensuring that there is no residual power overlapping the signal being analyzed.

To consider this, we investigate the case of GW191109_010717, a signal in O3 [5]. GW191109_010717 was a high-mass system observed in LIGO Hanford and LIGO Livingston, which were both experiencing scattered light glitches at the time of the event. Scattered light glitches were subtracted from the data from both detectors using the Bayeswave algorithm [5, 31]. In LIGO Livingston the signal overlapped directly with a slow scattering arch, while in LIGO Hanford it did not, and so we focus on the case of Livingston. GW191109_010717 is also notable for the inference of a negative value of the effective aligned spin parameter χ_{eff} [44, 45], which has been tentatively connected to whether or not data that contains a scattered light glitch is used in the analysis [21]. For the data from the LIGO Livingston detector, we apply our model in two cases: a set of arches a few seconds before the signal but not overlapping it, and the arches which directly overlap the signal. In the off-signal case, we use an eight-arch model, while for inference on the signal we restrict ourselves to three-arch analysis. When doing analysis of an overlapping gravitational-wave signal and scattering event, overly broad priors may allow the gravitational-wave signal power to be conflated with glitch power. Targeted priors reduce the impact

of this, since the morphology of the arches and of the gravitational-wave signal - especially the values of f_{mod} which would be inferred for each - are inconsistent, but as a matter of safety we restrict ourselves to the region below 40 Hz, where the glitch has been shown to have the most effect [21]. Subtraction for each case may be seen in Figure 6.5. In the off-signal case, we are able to subtract excess power from the scattered light glitch, and notably only find support for the lower three arches which are visible. Using the test from [43], we find that the data few seconds before the signal is consistent with Gaussian noise at a p-value of 0.96 after subtraction of the best-fit parameters. Because the same scatterer generates the arches on-signal, this lends support to the possibility that there are only three arches in the on-signal case as well, but this cannot be confirmed for lack of joint inference. Meanwhile, we are able to cleanly subtract the three arches we do search for in the on-signal case, roughly at 25, 36, and 47 Hz. While it is interesting to investigate how the use of this method impacts estimates of the signal's source properties and if the data is consistent with Gaussian noise after subtraction of the glitch and the signal, we defer these questions to future studies.

6.6 Conclusions

Modelling scattered light glitches using Bayesian inferences and a physically motivated model as presented in this chapter has a number of benefits as compared to previous methods used to characterize and subtract these glitches. Firstly, modeled inference allows for robust subtraction of these glitches from detector data, including when coincident with true gravitational-wave signals. Secondly, we have shown that this modeled inference provides as meaningful test of the presence or absence of higher arch harmonics which may be present in the data, including arches which are not visible within spectrograms and thus might otherwise escape detection. Characterizing scattered light glitches in this way is independent of whether a source of the scattered light has been identified. Together, these two features have the potential to make this an important tool in mitigation of scattered light glitches in the future. Because our method uses the standard inference tools of gravitational wave parameter estimation, there are good prospects for future development of joint inference methods, allowing for disentanglement of scattered light glitches from true signals, and reducing the chances of misidentifying scattered light with features in these signals.

References

- ¹J. Aasi, B. Abbott, R. Abbott, et al. (LIGO Scientific Collaboration), “Advanced LIGO”, *Class. Quantum Grav.* **32**, 074001 (2015) doi:10.1088/0264-9381/32/7/074001.
- ²F. Acernese, M. Agathos, K. Agatsuma, et al. (VIRGO), “Advanced Virgo: a second-generation interferometric gravitational wave detector”, *Class. Quantum Grav.* **32**, 024001 (2015) doi:10.1088/0264-9381/32/2/024001.
- ³T. Akutsu, M. Ando, K. Arai, et al. (KAGRA), “KAGRA: 2.5 Generation Interferometric Gravitational Wave Detector”, *Nature Astron.* **3**, 35–40 (2019) doi:10.1038/s41550-018-0658-y.
- ⁴R. Abbott et al. (LIGO Scientific, Virgo), “GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run”, *Phys. Rev. X* **11**, 021053 (2021) doi:10.1103/PhysRevX.11.021053.
- ⁵R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run”, *Phys. Rev. X* **13**, 041039 (2023) doi:10.1103/PhysRevX.13.041039.
- ⁶R. Abbott et al. (LIGO Scientific, VIRGO), “GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run”, *Phys. Rev. D* **109**, 022001 (2024) doi:10.1103/PhysRevD.109.022001.
- ⁷A. Buikema, C. Cahillane, G. Mansell, et al. (aLIGO), “Sensitivity and performance of the Advanced LIGO detectors in the third observing run”, *Phys. Rev. D* **102**, 062003 (2020) doi:10.1103/PhysRevD.102.062003.
- ⁸B. P. Abbott, R. Abbott, T. D. Abbott, et al. (LIGO Scientific Collaboration, Virgo Collaboration), “Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914”, *Class. Quant. Grav.* **33**, 134001 (2016) doi:10.1088/0264-9381/33/13/134001.
- ⁹D. Davis et al. (LIGO), “LIGO detector characterization in the second and third observing runs”, *Class. Quant. Grav.* **38**, 135014 (2021) doi:10.1088/1361-6382/abfd85.
- ¹⁰F. Acernese et al. (Virgo), “Virgo Detector Characterization and Data Quality during the O3 run”, (2022) doi:10.48550/arXiv.2205.01555.
- ¹¹P. Nguyen, R. Schofield, A. Effler, et al. (AdvLIGO), “Environmental noise in advanced LIGO detectors”, *Class. Quant. Grav.* **38**, 145001 (2021) doi:10.1088/1361-6382/ac011a.
- ¹²B. P. Abbott, R. Abbott, T. Abbott, et al. (LIGO Scientific Collaboration, Virgo Collaboration), “Effects of data quality vetoes on a search for compact binary

- coalescences in Advanced LIGO’s first observing run”, *Class. Quant. Grav.* **35**, 065010 (2018) DOI:10.1088/1361-6382/aaaafa.
- ¹³A. H. Nitz, “Distinguishing short duration noise transients in LIGO data to improve the PyCBC search for gravitational waves from high mass binary black hole mergers”, *Class. Quant. Grav.* **35**, 035016 (2018) DOI:10.1088/1361-6382/aaa13d.
- ¹⁴M. Cabero, A. Lundgren, A. Nitz, et al., “Blip glitches in Advanced LIGO data”, *Class. Quant. Grav.* **36**, 15 (2019) DOI:10.1088/1361-6382/ab2e14.
- ¹⁵D. Davis, L. V. White, and P. R. Saulson, “Utilizing aLIGO Glitch Classifications to Validate Gravitational-Wave Candidates”, *Class. Quant. Grav.* **37**, 145001 (2020) DOI:10.1088/1361-6382/ab91e6.
- ¹⁶G. Ashton, S. Thiele, Y. Leconte, J. McIver, and L. K. Nuttall, “Parameterised population models of transient non-Gaussian noise in the LIGO gravitational-wave detectors”, *Class. Quant. Grav.* **39**, 175004 (2022) DOI:10.1088/1361-6382/ac8094.
- ¹⁷C. Pankow et al., “Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817”, *Phys. Rev. D* **98**, 084016 (2018) DOI:10.1103/PhysRevD.98.084016.
- ¹⁸J. Powell, “Parameter Estimation and Model Selection of Gravitational Wave Signals Contaminated by Transient Detector Noise Glitches”, *Class. Quant. Grav.* **35**, 155017 (2018) DOI:10.1088/1361-6382/aacf18.
- ¹⁹R. Macas, J. Pooley, L. K. Nuttall, D. Davis, M. J. Dyer, Y. Leconte, J. D. Lyman, J. McIver, and K. Rink, “Impact of noise transients on low latency gravitational-wave event localization”, *Phys. Rev. D* **105**, 103021 (2022) DOI:10.1103/PhysRevD.105.103021.
- ²⁰J. Y. L. Kwok, R. K. L. Lo, A. J. Weinstein, and T. G. F. Li, “Investigation of the effects of non-Gaussian noise transients and their mitigation in parameterized gravitational-wave tests of general relativity”, *Phys. Rev. D* **105**, 024066 (2022) DOI:10.1103/PhysRevD.105.024066.
- ²¹D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. Di Renzo, and G. Ashton, “Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run”, *Class. Quant. Grav.* **39**, 245013 (2022) DOI:10.1088/1361-6382/aca238.
- ²²S. Soni, C. Austin, A. Effler, et al. (LIGO), “Reducing scattered light in LIGO’s third observing run”, *Class. Quant. Grav.* **38**, 025016 (2020) DOI:10.1088/1361-6382/abc906.
- ²³S. Soni, C. P. L. Berry, S. B. Coughlin, et al., “Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning”, *Class. Quant. Grav.* **38**, 195016 (2021) DOI:10.1088/1361-6382/ac1ccb.

- ²⁴T. Accadia et al., “Noise from scattered light in Virgo’s second science run data”, *Class. Quant. Grav.* **27**, edited by F. Ricci, 194011 (2010) DOI:10.1088/0264-9381/27/19/194011.
- ²⁵T. Accadia, F. Acernese, P. Astone, et al. (VIRGO), “Characterization of the Virgo Seismic Environment”, *Class. Quant. Grav.* **29**, 025005 (2012) DOI:10.1088/0264-9381/29/2/025005.
- ²⁶G. Valdes, B. O’Reilly, and M. Diaz, “A Hilbert–Huang transform method for scattering identification in LIGO”, *Class. Quant. Grav.* **34**, 235009 (2017) DOI:10.1088/1361-6382/aa8e6b.
- ²⁷S. Bianchi, A. Longo, G. Valdes, G. González, and W. Plastino, “An automated pipeline for scattered light noise characterization”, *Class. Quant. Grav.* **39**, 195005 (2022) DOI:10.1088/1361-6382/ac88b0.
- ²⁸A. Longo, S. Bianchi, G. Valdes, N. Arnaud, and W. Plastino, “Daily monitoring of scattered light noise due to microseismic variability at the Virgo interferometer”, *Class. Quant. Grav.* **39**, 035001 (2022) DOI:10.1088/1361-6382/ac4117.
- ²⁹E. Goetz, “Gwdetchar/gwdetchar: 2.0.5”, (2021) DOI:10.5281/zenodo.5722057.
- ³⁰N. J. Cornish and T. B. Littenberg, “BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches”, *Class. Quant. Grav.* **32**, 135012 (2015) DOI:10.1088/0264-9381/32/13/135012.
- ³¹N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, “BayesWave analysis pipeline in the era of gravitational wave observations”, *Phys. Rev. D* **103**, 044006 (2021) DOI:10.1103/PhysRevD.103.044006.
- ³²S. Hourihane, K. Chatziioannou, M. Wijngaarden, D. Davis, T. Littenberg, and N. Cornish, “Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data”, *Phys. Rev. D* **106**, 042006 (2022) DOI:10.1103/PhysRevD.106.042006.
- ³³G. Ashton, “Gaussian processes for glitch-robust gravitational-wave astronomy”, *Mon. Not. Roy. Astron. Soc.* **520**, 2983–2994 (2023) DOI:10.1093/mnras/stad341.
- ³⁴D. Davis, T. J. Massinger, A. P. Lundgren, J. C. Driggers, A. L. Urban, and L. K. Nuttall, “Improving the sensitivity of advanced ligo using noise subtraction”, *Class. Quant. Grav.* **36**, 055011 (2019) DOI:10.1088/1361-6382/ab01c5.
- ³⁵K. Mogushi, “Reduction of transient noise artifacts in gravitational-wave data using deep learning”, (2021) DOI:10.48550/arXiv.2105.10522.
- ³⁶A. E. Tolley, G. S. Cabourn Davies, I. W. Harry, and A. P. Lundgren, “ArchEnemy: removing scattered-light glitches from gravitational wave data”, *Class. Quant. Grav.* **40**, 165005 (2023) DOI:10.1088/1361-6382/ace22f.

- ³⁷C. Talbot, E. Thrane, S. Biscoveanu, and R. Smith, “Inference with finite time series: Observing the gravitational Universe through windows”, *Phys. Rev. Res.* **3**, 043049 (2021) DOI:10.1103/PhysRevResearch.3.043049.
- ³⁸B. P. Abbott, R. Abbott, T. Abbott, et al. (LIGO Scientific, Virgo), “A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals”, *Class. Quant. Grav.* **37**, 055002 (2020) DOI:10.1088/1361-6382/ab685e.
- ³⁹G. Ashton, M. Huebner, P. Lasky, et al., “BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy”, *Astrophys. J. Suppl.* **241**, 27 (2019) DOI:10.3847/1538-4365/ab06fc.
- ⁴⁰N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines”, *Journal of Chemical Physics* **21**, 1087–1092 (1953) DOI:10.1063/1.1699114.
- ⁴¹J. Skilling, “Nested sampling for general Bayesian computation”, *Bayesian Analysis* **1**, 833–859 (2006) DOI:10.1214/06-BA127.
- ⁴²J. S. Speagle, “dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences”, *Mon. Not. Roy. Astron. Soc.* **493**, 3132–3158 (2020) DOI:10.1093/mnras/staa278.
- ⁴³L. Vazsonyi and D. Davis, “Identifying glitches near gravitational-wave signals from compact binary coalescences using the Q-transform”, *Class. Quant. Grav.* **40**, 035008 (2023) DOI:10.1088/1361-6382/acafd2.
- ⁴⁴P. Ajith et al., “Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins”, *Phys. Rev. Lett.* **106**, 241101 (2011) DOI:10.1103/PhysRevLett.106.241101.
- ⁴⁵L. Santamaria et al., “Matching post-Newtonian and numerical relativity waveforms: systematic errors and a new phenomenological model for non-precessing black hole binaries”, *Phys. Rev. D* **82**, 064016 (2010) DOI:10.1103/PhysRevD.82.064016.

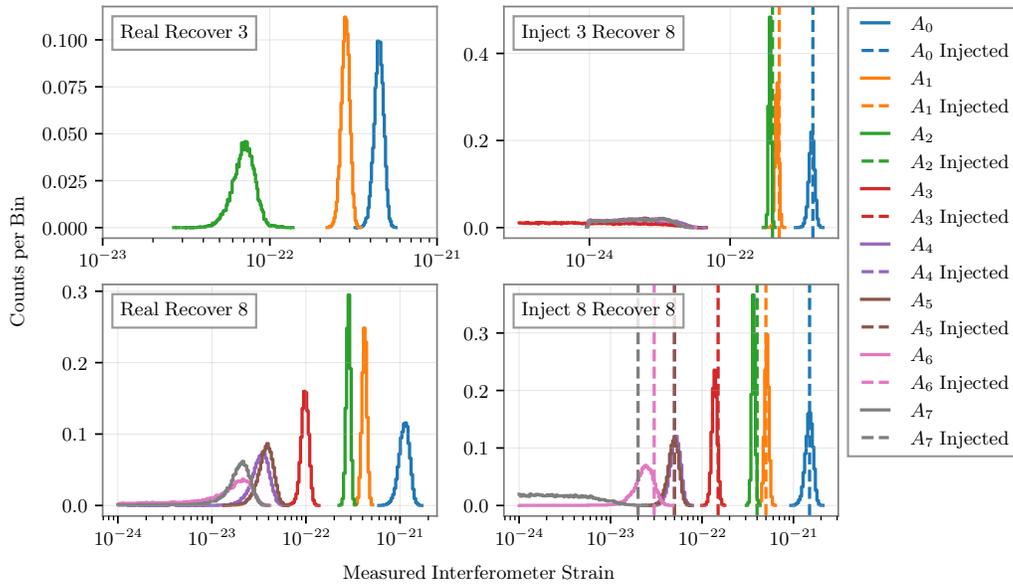


Figure 6.4: Amplitude posteriors for four cases with O2 data and O2-like data are plotted. The top left shows analysis of the O2 scattering example discussed, with inference done using a three-arch model. The bottom left shows the sample example, with inference done by an eight-arch model. Posteriors for arches recovered in both cases are shown to be consistent, while the eight-arch model also recovers amplitudes constrained away from zero for all eight modelled arches, despite only six arches being visible in the corresponding spectrogram (Figure 6.3). Subtraction of the maximum likelihood parameters for this panel may be seen in the bottom left panel of that figure. On the right, we perform consistency tests by injecting maximum likelihood parameters for the signal analyzed on the left into Gaussian noise drawn from the power spectral density of the example data segment. To test for the possibility of our model returning false positives, we first inject only three arches, while attempting to recover all eight, in the top right panel. Notably, the five not injected have amplitude posteriors consistent with zero, indicating that they are recovering true negatives. Finally, in the bottom left panel we inject eight arches and recover them with an eight-arch model. This actually returns a false negative for the eighth arch, which may be due to the specific noise realization, but does imply that the support for the eight-arch in the real case must be quite strong. This injection corresponds to the top right panel of Figure 6.3, while the subtraction of these maximum likelihood parameters may be seen in the bottom right panel of that figure.

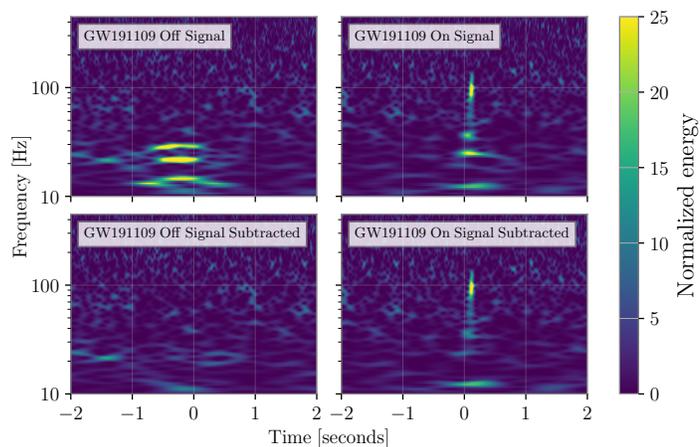


Figure 6.5: Spectrograms for the stretch of data around GW191109_010717 [5]. On the top left is data corresponding to a scattering event approximately three seconds before the gravitational wave signal. Inference was performed on this event, and the maximum likelihood parameters were subtracting, producing the residual plotted in the bottom left panel. Notably, though inference was done with a model which allowed up to eight arches, the three which were visible were the only ones with amplitudes constrained away from zero. The top right shows the stretch of data centered on the gravitational-wave signal, which overlapped with another scattering event. For this stretch of time we perform the analysis with a three-arch model, to prevent erroneous inclusion of gravitational-wave signal power in the glitch model. Subtraction of the maximum likelihood parameters for this three-arch model is shown in the bottom right panel.

THE ANTI-ALIGNED SPIN OF GW191109

This chapter contains contents from the published work

R. Udall, S. Hourihane, S. Miller, D. Davis, K. Chatziioannou, M. Isi, and H. Deshong, “Antialigned spin of GW191109: Glitch mitigation and its implications”, *Phys. Rev. D* **111**, 024046 (2025) DOI:10.1103/PhysRevD.111.024046.

My contributions to this work included all frequency domain analyses with BILBY (analyses 1-9, 11-19, and 22-221 in Table 7.1), writing, and editing.

7.1 Introduction

Reported in the third gravitational wave (GW) transient catalog (GWTC-3) [1], GW191109_010717 (more concisely GW191109) stands out among existing binary black hole (BBH) signals. With source-frame primary and secondary masses of $m_1 = 65_{-11}^{+11} M_\odot$ and $m_2 = 47_{-13}^{+15} M_\odot$ (90% symmetric credible intervals), it is among the most massive events. Furthermore, there is significant support for black hole (BH) spins anti-aligned with the orbital angular momentum: the mass-weighted effective spin [2–4] is $\chi_{\text{eff}} = -0.29_{-0.31}^{+0.42}$. For these reasons, as well as support for unequal masses, $q = m_2/m_1 = 0.73_{-0.24}^{+0.21}$, spin-precession, and hints of eccentricity [5, 6], the binary is potentially of dynamical and/or hierarchical origin [7, 8] and impacts population inference [9, 10].

Multiple GW191109 properties hint toward a dynamical origin. High masses, above the pair-instability supernova (PISN) limit of $45 - 70 M_\odot$ (depending on modeling assumptions) [11, 12], may require a hierarchical mechanism in order to form and merge. Asymmetric masses, in particular, might imply the merger of a second- and a first-generation BH [7]. Furthermore, population synthesis simulations of isolated formation scenarios typically find little support for spins anti-aligned with the orbital angular momentum, unless supernova kicks are exceptionally high [8, 13, 14]. Finally, eccentricity would also be challenging to explain except by dynamical processes [5, 15, 16], due to the rapid orbit circularization by GW emission [17].

Given their astrophysical implications, the inferred properties of GW191109 are worth scrutinizing. The first potential source of systematics is the waveform used

to model the signal. GWTC-3 employed the $\text{IMR}_{\text{PHENOMXPHM}}$ [18] and SEOB-NRv4PHM approximants [19], with inference performed by BILBY [20, 21] and RIFT [22] respectively. Both models include the physical effects of higher-order modes and spin-precession, and headline results (as quoted above) are their average. However, GW191109 is flagged for systematic differences between approximants [1], especially for the binary inclination (edge-on versus face-on/off respectively) and the longer $\chi_{\text{eff}} > 0$ tail with $\text{IMR}_{\text{PHENOMXPHM}}$. A third waveform, NRSUR7DQ4 [23], was employed in Ref. [24]. A direct surrogate of numerical relativity simulations, NRSUR7DQ4 is expected to be the most accurate available model for systems with high masses and spins [23–25]. These results bolster the evidence for dynamical origin, with a more negative spin, $\chi_{\text{eff}} = -0.38^{+0.21}_{-0.20}$, asymmetric masses, $q = 0.65^{+0.20}_{-0.19}$, and a precessing spin parameter [26] of $\chi_p = 0.59^{+0.26}_{-0.27}$. While waveform systematics remain relevant, the broad agreement between three waveforms (including a direct surrogate to numerical relativity) that $\chi_{\text{eff}} \lesssim 0$ to varying credibility, suggests that subsequent interpretations of its formation history remain valid.

A second potential source of systematics concerns modeling the detector noise. Around GW191109’s arrival, both LIGO [27] detectors experienced a terrestrial noise transient known as a scattered light glitch [1, 28, 29]. The Virgo detector [30] was offline at this time, and so only the LIGO detectors contributed to the observation. In LIGO Hanford (LHO), the glitch power was at a nadir while the event was in the detection band, making its impact on the inferred parameters negligible, see App. 7.6. As such, we ignore the LHO glitch going forward. By contrast, glitch power in the Livingston detector (LLO) was directly coincident in time and frequency with the signal, a circumstance which could bias astrophysical inference [31–35]. Specifically, glitch power extends up to $\sim 40\text{Hz}$, coincident with the signal; see Fig. 7.1. Spin parameters might be particularly susceptible to such data quality issues due to the relatively smaller imprint they leave on signals compared to, e.g., the BH masses. For example, GW200129 shows evidence of spin-precession [1, 25], but its significance depends on how the glitch that overlapped that signal is modeled [33, 36].

The headline GWTC-3 results were obtained after an estimate for the glitch had been subtracted from the data. The two-step process involved first modeling the signal and the glitch with a flexible sum of coherent and incoherent wavelets respectively with BAYESWAVE [37–39]. Second, a fair draw from the glitch posterior was subtracted

and the system parameters were inferred as quoted above. This procedure has been shown to generally lead to unbiased mass and (aligned) spin inference [31, 34]. However, uncertainties remain related to BAYESWAVE’s glitch model and in the fair draw chosen to be subtracted. These effects were investigated in Ref. [28], albeit with a simpler waveform model with single-spin precession and no higher-order modes, IMRPHENOMPv2 [40]. Glitch mitigation was found to affect the χ_{eff} inference by a similar amount as waveform systematics. Completely removing the glitch-affected data, i.e. all LLO data below 40 Hz, instead resulted in a dramatic shift of χ_{eff} to positive values $\chi_{\text{eff}} = 0.27^{+0.24}_{-0.48}$.

The stark impact of glitch-affected data on astrophysically-impactful spin inference motivates our study. In Sec. 7.3 we extend Ref. [28] to explore the manner in which the data inform the system parameters. Using NRSUR7DQ4 and a frequency-domain analysis, we find that the LLO data between 30 and 40 Hz are crucial for spin inference: excluding 30 – 40 Hz data shifts the probability of $\chi_{\text{eff}} < 0$ from 99.4% to 32.2%, effectively wiping out any preference for anti-aligned spins. A similar time-domain analysis [41] highlights the role of the data 0.1 – 0.04 s prior to merger. These data, which inform the $\chi_{\text{eff}} < 0$ measurement, coincide in time and frequency with excess power in LLO; see Fig. 7.2 and in particular the excess power at ~ 36 Hz. To check whether such dramatic shifts in support for $\chi_{\text{eff}} < 0$ are possible from Gaussian noise alone, we analyze 100 simulated signals consistent with GW191109. We find that shifts of this magnitude are unlikely but not impossible as 6% of the simulations experience a larger shift than GW191109.

In Sec. 7.4, we focus on the 36 Hz excess power and address the key question: is the excess power part of the signal (and hence $\chi_{\text{eff}} < 0$) or is it part of the glitch (and hence inference has been affected by systematics)? Rather than the two-step process of glitch fitting and subtraction, we perform a full analysis where we *simultaneously* model both the signal and the glitch. Using a physically motivated model for scattered light glitches [29] we find $\chi_{\text{eff}} < 0$ at the 99.9% level using NRSUR7DQ4. We attribute this to the fact that the 36 Hz power is more contained in time than expected for scattered light glitches that are characterized by extended arches in time-frequency. This analysis, therefore, attributes the 36 Hz power to the signal and thus prefers $\chi_{\text{eff}} < 0$. It is, however, possible that not all terrestrial power is due to scattered light or that the physical model of Ref. [29] does not capture all scattered light power. Instead, using a more flexible model for the glitch based on wavelets and BAYESWAVE and IMRPHENOMXPHM we obtain a bimodal solution

for the spin. One mode, preferred at the 70:30 level, attributes most of the 36 Hz power to the signal and results in $\chi_{\text{eff}} < 0$. The second mode attributes this power to the glitch and results in $\chi_{\text{eff}} > 0$. Given the low signal-to-noise ratio (SNR) of the 36 Hz power, these results are impacted by the priors of the glitch model parameters at the few percent level.

In Sec. 7.5 we summarize our conclusions. Physically grounded assumptions about the behavior of scattered light glitches lend support to $\chi_{\text{eff}} < 0$ for GW191109, and thus a dynamical origin. However, both systematic limitations on scattered light models and statistical uncertainty due to low SNR of the excess power and the impact of glitch priors prevent us from making that determination confidently. While the crucial 36 Hz power is not part of the scattered light glitch as modeled in Ref. [29], we cannot rule out glitch mismodeling or other types of terrestrial noise.

7.2 Modeling signals and glitches

The relevant data contain the GW191109 signal, glitch power, and Gaussian noise. In this section, we describe how we model the signal (Sec. 7.2), the glitch (Sec. 7.2), and methods for glitch mitigation (Sec. 7.2). We focus on the respective strengths and weaknesses of each approach and what unique information each supplies. All analyses model the Gaussian noise component with the power spectral densities (PSDs) from the GWTC-3 data release [42, 43]. Detailed settings and identification numbers for all analyses are given in Table 7.1 in App. 7.7.

Modeling the Compact Binary Signal

We use both time- and frequency-domain techniques to model the signal with either waveform approximants for compact binary signals or, more generically, with sine-Gaussian wavelets. All analyses consider data surrounding the nominal trigger time of GW191109, GPS time 1257296855.22, and employ a sampling rate of 1024 Hz, with the maximum analysis frequency set to 7/8 of the Nyquist frequency. Unless otherwise noted, analyses that model only the compact binary (and not the glitch) use a minimum frequency of 20 Hz in both detectors. We use standard compact-binary priors [21], notably uniform in detector-frame component masses and spin magnitude and orientation.

Frequency-domain inference

Frequency domain analyses with waveform approximants are based on BILBY [20, 21] with its implementation of the DYNASTY sampler [44] and BAYESWAVE [32], both

analyzing 4 s of data. The former models the signal with `NRSUR7DQ4` [23] and the latter with `IMRPHENOMXPHM` [18] (though for consistency we also perform checks with the former using `IMRPHENOMXPHM` in App. 7.8). `NRSUR7DQ4` supports a minimum mass ratio of 0.25 and minimum detector-frame chirp mass of $35 M_{\odot}$; neither restriction affects the analysis. We extend into the extrapolation region in spins, setting a maximum spin magnitude of 0.99. For comparison, we also perform analyses with `BAYESWAVE` where the signal is modeled as a flexible sum of coherent sine-Gaussian wavelets [37, 39]. Settings are similar to the glitch wavelet analysis described in Sec. 7.2, only here, the wavelets are coherently projected across the two detectors rather than being independent.

Time-domain inference

While GW inference is typically conducted in the frequency domain for computational efficiency, it can equivalently be conducted in the time domain [45–48]. Frequency domain analyses are non-local in time; to avoid non-trivial likelihood modifications [49], time-domain inference is necessary in order to isolate purely temporal features of the data. Below, we truncate the GW191109 data at different times around the 36 Hz excess power, and independently conduct inference on the pre- or post-cutoff-time data. For this, we use the time-domain inference code employed in Ref. [41] to study the GW190521 properties and which was based on time-domain implementations targeting post-merger data [46, 47]. All time-domain results are based on regions of 1 s of data around GW191109’s trigger time and employ `NRSUR7DQ4` [23]. The same PSDs are used in the time domain analyses are the same as those in the frequency domain analyses, i.e. from the GWTC-3 data release [42].

Modeling the Glitch

Both LHO and LLO experienced slow scattering noise around the time of GW191109. We use two models for the glitch power: a physically motivated model tailored to slow scattering, implemented in `BILBY`, and a more flexible wavelet model, implemented in `BAYESWAVE`.

Physically-parameterized scattering

As the name implies, scattered light glitches arise due to laser light that scatters off the main beam path, bounces off a surface, and recombines with the main beam [29, 50–53]. During periods of significant ground motion when the scattering surface

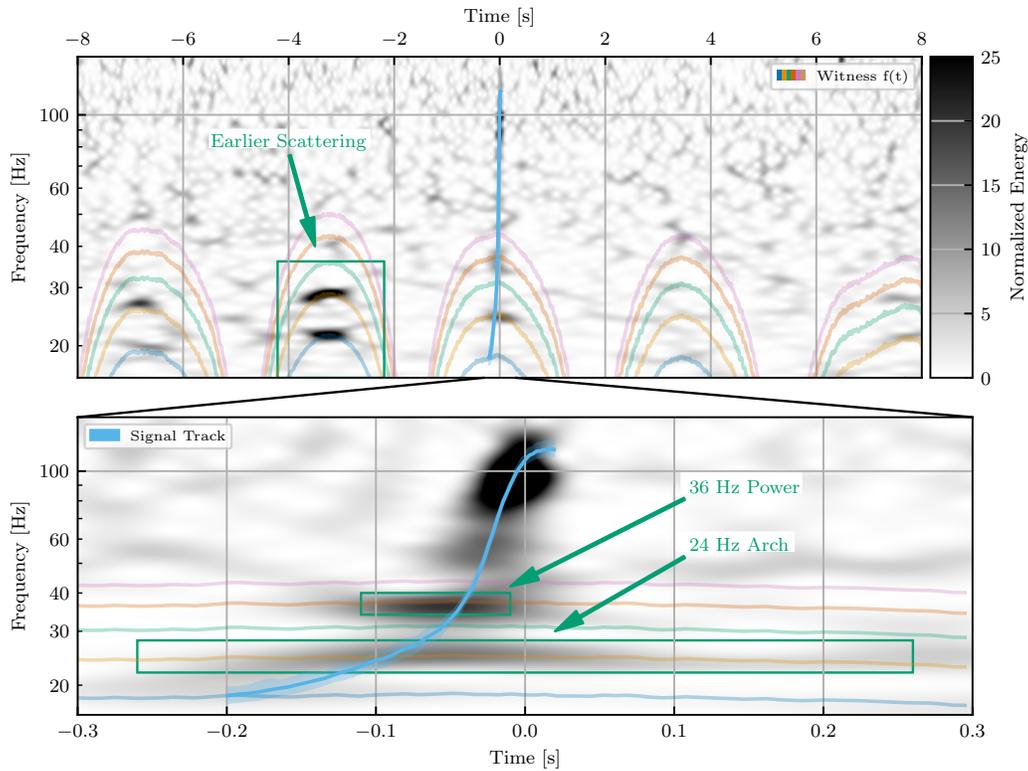


Figure 7.1: Spectrograms of the original (before glitch mitigation) data in LLO centered around the time of GW191109. The top panel shows ± 8 s of data, while the bottom panel zooms in around the event. Onto this, we plot the time-frequency tracks of the scattered light glitch, as predicted by the motion observed in the witness channel L1: SUS-ETMX_L2_WIT_L_DQ. This is the witness to the penultimate stage of the reaction chain pendulum for the X-arm end test mass. The scattering surface is the final stage of the reaction chain, and so this witness does not perfectly capture the motion of the scattering surface; to compensate, we apply a static coefficient of 1.38 to the predicted frequency, such that it is calibrated to the prominent scattering arches ~ 3 s before the event. We also plot the inferred signal from a NRSUR7DQ4 analysis of full-bandwidth data after glitch subtraction (Run 1 in Table 7.1). We annotate three regions of interest: the prominent scattering before the event (top panel), the long-duration excess power at 24 Hz (bottom panel), and the short-duration excess power at 36 Hz (bottom panel). Both the 24 Hz and the 36 Hz excess power coincide with expected glitch arches, however only the former has an arch-like shape.

moves, this light acquires a phase offset, resulting in excess noise. Figure 7.1 shows a spectrogram of the LLO data, along with the frequency tracks of the scattering excess noise as predicted by a witness data stream that captured the motion of the suspected scattering surface. The effect of scattering is most easily discernible 3 s before the signal, taking the form of a “stack” of arches, which is characteristic of slow scattering. Slow scattering results from low-frequency ground motion, $\sim 0.05\text{--}0.3$ Hz, driving slow movement of the scattering surface [51]. This induces phase noise with frequency [50]

$$f(t) = \left| \frac{2v_{sc}(t)}{\lambda} \right|, \quad (7.1)$$

with $v_{sc}(t)$ being the velocity of the scattering surface and $\lambda = 1064$ nm is the wavelength of the laser. In order for the glitch frequency to reach the analysis band, the scattered light must bounce multiple times, yielding a fixed frequency ratio between arches as the same amount of phase offset is accumulated with each successive bounce.

This picture forms the basis for a parametrized model for slow scattering that treats the scattering surface as a simple harmonic oscillator. We use the physically parameterized scattering model proposed in Ref. [29]. The model is a sum of frequency-modulated sinusoids with $2N + 4$ parameters, where N is the number of arches:

$$g(t) = \sum_{k=0}^N A_k \sin \left[\frac{f_{h,0} + k\delta f_h}{f_{mod}} \sin \left(2\pi f_{mod}(t - t_c) \right) + \phi_k \right]. \quad (7.2)$$

The peak frequency of the lowest arch is $f_{h,0}$ and the spacing in peak frequencies between adjacent arches is δf_h , such that the peak frequency of the k th arch is $f_{h,0} + k\delta f_h$.¹ The modulation frequency f_{mod} corresponds to the motion of the scattering surface (and hence the driving ground motion) and sets the width of the arch, while t_c is the time of peak frequency. Each arch k further has an independent amplitude A_k and phase ϕ_k .

Priors on these parameters reflect the physical slow scattering picture. For δf_h and $f_{h,0}$, we place uniform priors around the approximate values read from Fig. 7.1, $\delta f_h \sim \mathcal{U}(5, 8)$ Hz and $f_{h,0} \sim \mathcal{U}(18, 20)$ Hz, while for ϕ_k we set a uniform periodic prior, $\phi_k \sim \mathcal{U}(0, 2\pi)$. We employ two sets of priors on f_{mod} . “Physical” priors limit the modulation frequency to the microseism band $f_{mod} \sim \mathcal{U}(0.05 - 0.3)$ Hz [51].

¹Unlike Ref. [29], we fix the frequency ratio between arches to δf_h , thus eliminating $N - 1$ parameters.

“Targeted” priors further restrict the modulation based on the witness motion $f_{mod} \sim \mathcal{U}(0.05 - 0.15)$ Hz. While the former choice is more agnostic, the latter maximizes information from witness channels. Since the detector sensitivity varies by orders of magnitude in the frequency region spanned by the arches, we explore both a uniform and log-uniform amplitude for the amplitude A_k . We do not impose a relation between the arch amplitudes; while amplitudes might be expected to decrease with each arch, this is not universally the case [29].

The number of arches N is fixed and not a parameter of the model that is varied, unlike the flexible glitch model with `BAYESWAVE` discussed in Sec. 7.2. The choice of the number of arches, therefore, impacts the results, especially for the uniform amplitude prior. Motivated by Fig. 7.1 we set $N = 5$, a choice which we investigate in App. 7.9. All analyses that model the glitch with the slow scattering model further employ a reduced minimum frequency of 16 Hz in LLO. Though the signal SNR, ρ , is negligible between 16 and 20 Hz (0.16% of ρ^2 in LLO), this setting accommodates the ~ 18 Hz arch, which in turn informs the upper arches.

If the glitch overlapping GW191109 is consistent with the physical picture that motivates the slow scattering model, corresponding analyses provide the most sensitive results on the system properties. However, the model is also restricted to an interpretation of slow scattering and does not provide a means to test this assumption. If other non-Gaussian transient noise is present or if the physical picture does not fully capture the glitch morphology, biases might arise.

Wavelet glitch model

To mitigate against glitch modeling systematics, we also employ a more flexible approach with `BAYESWAVE` which models transient, non-Gaussian noise independently in each detector as sums of sine-Gaussian, Morlet-Gabor wavelets [37, 39]. Such wavelets are an overcomplete basis and any smooth function can be described with some linear combination of wavelets. Thus, this glitch model is flexible enough to fit a wide range of non-Gaussian transients without fine-tuning, including slow scattering [32, 34]. Unlike the parameterized scattering model, the `BAYESWAVE` glitch model is purely phenomenological, though motivated by the generic morphology of the LIGO glitches. Each wavelet is described by five parameters: central time t and frequency f , quality factor Q describing how quickly it is damped, amplitude A , and phase ϕ . We employ uniform priors over all parameters other than the amplitude, which is set through a prior on the wavelet SNR that peaks at 5 [37]. In addition

to these parameters, the number of wavelets in each detector is also a variable and sampled over with a uniform prior. Uniform prior bounds are wide enough so as to not affect the posterior.

Glitch Mitigation Approaches

We employ three approaches to mitigate and study the impact of the glitch on inference: (1) discarding the affected data, (2) subtracting an estimate for the glitch from the data, and (3) simultaneously modeling the signal and glitch and obtaining source parameters for the former by marginalizing over the latter.

Discarding Affected Data

The most straightforward way to mitigate the impact of a glitch is to discard the affected data, either by band-passing in the frequency domain or by analyzing limited segments in the time domain [28, 31, 33]. While straightforward to implement, such methods forego all information in the discarded data, making them suboptimal. We instead follow Refs. [28, 33] and discard glitch-affected data only as a consistency check and to study the impact of the glitch, or its residual, on inference. Such analyses confirm that mitigation is necessary and provide insights into the detailed behavior of the data.

Subtraction of a Glitch Estimate

GWTC-3 results on GW191109 were obtained after an estimate of the glitch was subtracted from the data [1]. In most cases, the estimate for the glitch is a fair draw from a previous analysis with `BAYESWAVE` [1, 54–56] but estimates generated from witness channels such as in `GWSUBTRACT` are also possible [28]. Glitch-subtracted data are then used for downstream source inference. This method retains all the data and information available and is, therefore, more suitable for production analyses. However, its efficacy hinges on the subtracted glitch estimate since the true morphology of the glitch cannot be perfectly known. In the fair draw case, the expected glitch residual SNR is non-zero due to statistical uncertainty [57]. In the witness channel case, the relevant transfer functions induce further systematic and/or statistical uncertainty [33]. Residual glitch power that could bias inference is therefore expected.

Marginalization Over Glitch Realizations

Since selecting a single glitch estimate to subtract results in residual glitch SNR, the final method is to marginalize over the glitch. This approach is the most robust, but it is also typically more difficult to implement. Given some parameterized glitch model $g(\phi)$, we can model the data as

$$d = n + h(\theta) + g(\phi), \quad (7.3)$$

From this, we may extend the typical likelihood in a single detector to include the glitch:

$$\ln \mathcal{L}(d|\theta, \phi) = -\frac{1}{2} \sum_k \left\{ \frac{[d_k - h_k(\theta) - \phi_k(\phi)]^2}{S_n(f_k)} + \ln(2\pi S_n(f_k)) \right\} \quad (7.4)$$

where k indexes the frequency bins being summed over, and $S_n(f_k)$ is the power spectral density in the k 'th frequency bin. In detectors without glitches this reduces to the standard CBC likelihood, and they combine in the usual way. Using this formulation, one may then sample over both $h(\theta)$ and $g(\phi)$ simultaneously. From these samples, one may then marginalize over ϕ to produce CBC posteriors which reflect uncertainties in the modeling of the glitch.

We perform three glitch-marginalized analyses on GW191109. First, using `BAYESWAVE`, we combine the signal model with `IMRPHENOMXPHM` described in Sec. 7.2 and the sine-Gaussian glitch model described in Sec. 7.2. Compared to previous relevant analyses [32–34] we have extended the signal model to support waveforms with spin-precession and higher-order modes. Second, again using `BAYESWAVE`, we combine the coherent wavelet signal model described in Sec. 7.2 and the incoherent wavelet glitch model described in Sec. 7.2 [35]. This analysis uses a more flexible—and thus less sensitive—model for the GW signal; it is thus used as an additional check. Even though `BAYESWAVE` has the capability to also marginalize over the Gaussian noise PSD [38, 58], we fix it for consistency with other analyses and since its effect on source inference is generally minimal [59]. Third, we implemented the physically-motivated scattered light glitch model of Sec. 7.2 in `BILBY`. This allows us to jointly use the slow scattering model and the `NRSUR7DQ4` approximant for the signal.

7.3 Understanding the GW191109 Inference

In this section, we explore the relation between the GW191109 inference, especially the $\chi_{\text{eff}} < 0$ measurement, and the glitch-affected data. In Fig. 7.1 we show spectrograms of the original data (without any glitch mitigation) in LLO at the time of the event.² Arch-like traces (multiple colors) show the glitch time-frequency tracks as predicted by a witness channel. The light blue track corresponds to GW191109 as inferred with NRSUR7DQ4 from data after the glitch was subtracted (Run 1 in Table 7.1). The upper panel presents 16 s of data; scattering arches are visible leading up to the event. In the bottom panel, we focus on the vicinity of the signal and highlight the intersection of the signal track with visible excess power along the projected scattering tracks. The first is at ~ 24 Hz and has the expected duration and morphology of a scattering arch. The second is at ~ 36 Hz and while it coincides with the glitch track predicted by the witness, the excess power duration is short and does not match the expected behavior of slow scattering. As noted in Ref. [28], this 36 Hz excess power is not included in the original BAYESWAVE glitch reconstruction and thus not subtracted in the GWTC-3 data.

We begin by confirming and extending the results of Ref. [28] with NRSUR7DQ4. Analyzing data from each detector separately (Runs 8 and 9 in Table 7.1) we confirm that the measurement is driven solely by LLO, which prefers $\chi_{\text{eff}} < 0$ at 99.6%, compared to 20.0% in LHO. Coherent analysis of both detectors (Run 1 in Table 7.1) tends to the LLO conclusion due to LLO's higher sensitivity in the relevant frequency range, shown below to be 20–40 Hz. Indeed, the maximum likelihood waveform from the coherent analysis accumulates 20% (8%) of its SNR squared in LLO (LHO) for frequencies below 40 Hz. This estimate further suggests that LHO data cannot aid in determining whether the critical ~ 36 Hz excess power is part of the signal or the glitch.

Similar differences in parameter inference per detector are present for other parameters as well, notably the detector-frame total mass M and luminosity distance D_L ; see footnote 3 for a discussion of the correlation between χ_{eff} and D_L . For example, in individual detector analyses (Runs 8 and 9 in Table 7.1) the detector-frame total mass is $M = 133^{+14}_{-14} M_{\odot}$ ($M = 162^{+21}_{-20} M_{\odot}$) in LLO (LHO), while the luminosity distance is $D_L = 1630^{+1360}_{-850}$ Mpc ($D_L = 2760^{+2300}_{-1570}$ Mpc) in LLO (LHO). The corresponding source-frame total mass remains the same as the increases in detector-frame mass

²A similar plot for the LHO data showing that the scattered light glitch does not overlap with the signal is given in App. 7.6.

and distance effectively “cancel out”. Though different, these estimates are still consistent with each other within statistical uncertainties so there is no indication of a discrepancy across detectors as was the case for GW200129 [33]. Moreover, these differences do not lead to diverging astrophysical interpretations like the χ_{eff} inference; we therefore focus on the latter in what follows.

Tracing inference across frequencies

To more precisely track the origin of the $\chi_{\text{eff}} < 0$ measurement across LLO data, we perform a series of coherent 2-detector analyses where we successively restrict the LLO frequencies, incrementing the minimum frequency f_L by 5 Hz from 20 to 45 Hz (Runs 1–5 in Table 7.1). We use the glitch-subtracted data where the 24 Hz arch from Fig. 7.1 has been subtracted, but the 36 Hz excess power has not [1]. A subset of these results are shown in Fig. 7.2 (pink shading). The top panel shows a spectrogram of the glitch-subtracted data; compared to Fig. 7.1, there is no excess noise at ~ 24 Hz.

Marginalized posteriors for χ_{eff} are shown in the bottom panel. The legend denotes the percentage of the total SNR squared ρ^2 (computed based on the maximum-likelihood full-band signal) that remains in the analysis window after each restriction. Removing data between 20–30 Hz (solid vs dashed horizontal lines in the top panel and histograms in the bottom panel) or 30–35 Hz (dashed vs dotted) removes 6% of ρ^2 but does not dramatically alter inference: $\chi_{\text{eff}} < 0$ is still preferred at 96.3% for $f_L = 35$ Hz. Such small shifts are likely consistent with the SNR reduction and regression to the prior (gray). Removing data 35 – 40 Hz (dotted vs dot-dashed) removes an additional 6% of ρ^2 and instead results in an abrupt shift in χ_{eff} , with $\chi_{\text{eff}} < 0$ now only at 32.2%, a moderate preference for positive values.³ Further bandwidth reduction does not modify the χ_{eff} posterior substantially (solid dark pink). These results indicate that it is the LLO data between 35 and 40 Hz that are crucial for measuring χ_{eff} , coinciding with the 36 Hz excess power visible in both the original data (Fig. 7.1) and the glitch subtracted data (Fig. 7.2).

The second panel from the top of Fig. 7.2 shows the whitened time-domain reconstructions. We compare signal reconstructions from two analyses with dramatically different χ_{eff} posteriors: the full bandwidth analysis that prefers $\chi_{\text{eff}} < 0$ against the

³Since the χ_{eff} prior is centered at zero, this shift to mildly positive values goes beyond regression to the prior. We attribute this to a mild $\chi_{\text{eff}} - D_L$ degeneracy that arises for merger-only signals. The uniform-in-volume prior favors larger D_L and results in larger χ_{eff} to compensate for the amplitude reduction. This degeneracy is less pronounced when the signal inspiral is visible, as then χ_{eff} is constrained by the inspiral phase evolution beyond just the merger amplitude.

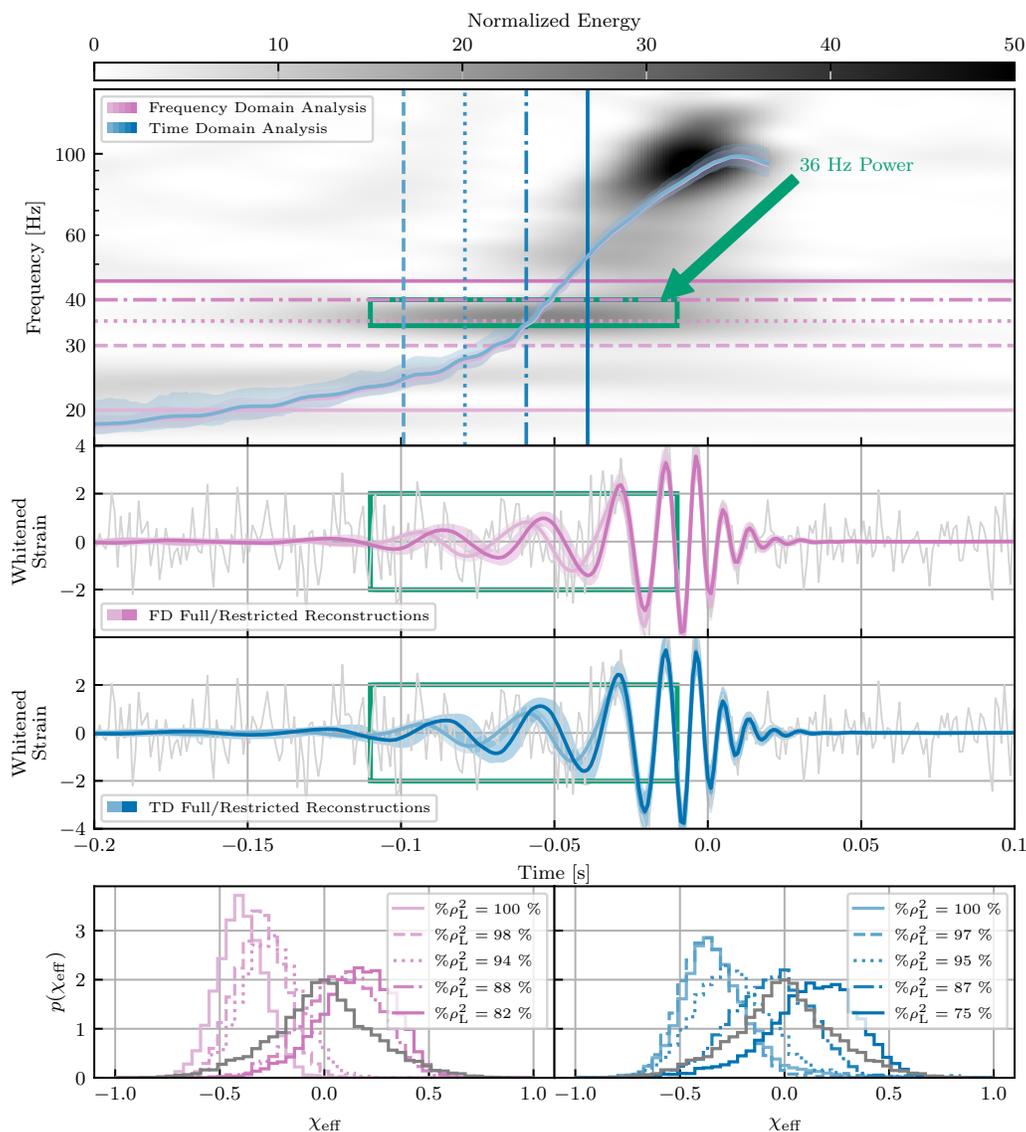


Figure 7.2: Tracing the χ_{eff} inference across frequencies and times. The top panel shows the spectrogram of the glitch-subtracted data around GW191109, with residual excess power at 36 Hz highlighted along with the signal track. We progressively remove data in the frequency domain (pink) and the time domain (blue) and re-analyze the restricted data. Vertical and horizontal lines in the top panel denote the time and frequency cuts, respectively; only data to the left or above these lines are analyzed. The two middle panels show the whitened time-domain data (grey) and signal reconstruction (pink and blue). Lighter colors correspond to the analyses of the full data, while darker colors correspond to the most restricted data (frequencies above 40 Hz and times from -0.04 s before merger onwards). The bottom row shows the χ_{eff} prior (gray) and χ_{eff} marginal posteriors from analyses with varying levels of data restriction, each corresponding to the lines on the top panel. The legend notes the SNR squared ρ^2 fraction in Livingston that remains in the analysis band after each data restriction.

$f_L = 40$ Hz analysis with a mildly positive χ_{eff} . While the two analyses are conducted on different data subsets, we can still evaluate the waveforms across the same times and plot them together. The reconstructions are consistent during the merger (corresponding to high frequencies included in both analyses), but start diverging 2 – 3 cycles before merger. By eye, the full-band reconstruction better matches the data for $t \approx -0.6$ s, corresponding to the 36 Hz excess power. When that power is included in the analysis, the signal model absorbs it by setting $\chi_{\text{eff}} < 0$ and pushing the GW cycle to earlier times. If that power is not part of the analysis, χ_{eff} is no longer required to be negative and the 36 Hz excess power is left unaccounted for.

This conclusion raises the question of whether the 36 Hz excess power is part of the signal or part of a glitch that remained unsubtracted. Though the shift in the χ_{eff} posterior is suggestive of anomalous noise, it is possible that it is at least partly due to loss of information as 6% of ρ^2 in LLO is contained in the 35–40 Hz frequency band. In Sec. 7.3 we contextualize this χ_{eff} shift with simulated signals.

Tracing inference across times

Having identified the crucial frequencies for χ_{eff} inference, here we do the same across time with the time-domain analysis described in Sec. 7.2. When used on the full dataset, frequency- and time-domain analyses should yield equivalent results. Indeed, we find consistent posteriors for χ_{eff} when analyzing GW191109 in the frequency and time domains, as seen by the solid histograms in Fig. 7.2.⁴

However, as Fig. 7.1 shows, there is no 1-to-1 mapping between time and frequency for the glitch. Though not as apparent, the same is true for the signal beyond the inspiral regime or due to spin-precession and higher-order modes. Truncating the data in the time domain is, therefore, not equivalent to truncating in the frequency domain, as the former allows us to probe the effect of individual cycles (or parts of cycles) of the signal or the glitch.

Results from progressively excluding the earlier portion of the signal in the time-domain (Run 10 in Table 7.1) are shown in Fig. 7.2 (blue shading). We find broadly

⁴The time- and frequency-domain analyses employ different priors on masses, luminosity distance, and time. The time-domain inference uses priors which are uniform in detector-frame total mass, mass ratio, and luminosity distance; and are normally distributed in geocenter time, centered at 1257296855.2114642 with a width of 0.005 seconds. We confirm that the differences in time and the mass priors effect the posteriors minimally. Reweighting between the two luminosity distance priors proves difficult due to finite sampling and upweighting portions of parameter space with no support in the posterior. However, the luminosity distance posteriors from the time- and frequency-domain analyses are in high agreement when the full data is analyzed, despite using different priors.

similar results as the frequency-domain analysis: the full data yield preference for $\chi_{\text{eff}} < 0$. As the segment that contains the 36 Hz excess power is progressively removed (blue vertical lines in the top panel), the χ_{eff} posterior shifts to being principally positive (equivalent blue histograms in the bottom panel). Overall, the data 0.1–0.04 s before merger are crucial for $\chi_{\text{eff}} < 0$ inference. Compared to the frequency-domain results, the shift in the χ_{eff} posterior is more gradual, likely due to the fact that the 36 Hz power is more concentrated in frequency, hence no time “cut” abruptly completely excludes it. Waveform reconstructions from the time-domain analysis (third panel from the top in Fig. 7.2) yield consistent conclusions.

Simulated signals

We investigate the degree to which the abrupt shift in the χ_{eff} posterior in Fig. 7.2 is consistent with SNR loss from removing data with simulated signals. We simulate 100 signals drawn from the GW191109 full-band posterior (Run 1 in Table 7.1), add them to Gaussian noise drawn from the GW191109 PSDs in LLO and LHO, and analyze the full data versus the > 40 Hz data in LLO independently (Runs 22–221 in Table 7.1). Signals have true values $\chi_{\text{eff}} < 0$ but as data and signal SNR are removed when $f_L = 40$ Hz, we expect the posterior to become more prior-like and shift toward $\chi_{\text{eff}} = 0$. For each simulated signal, Fig. 7.3 shows the probability of $\chi_{\text{eff}} \leq 0$ from the full-data, $f_L = 20$ Hz, and the restricted-data, $f_L = 40$ Hz, analysis.

For almost all signals removing low-frequency LLO data results in a χ_{eff} posterior that shifts closer to the prior and positive values (lying below the diagonal) as expected for signals with true values of $\chi_{\text{eff}} < 0$. In most cases, this shift is marginal, and posteriors stay majority-negative, as evidenced by the high density (64% of all signals) in the top right quadrant (green axes). The next most likely outcome is the bottom right quadrant (orange axes), which contains 34% of the signals, including GW191109: here the χ_{eff} posterior shifts from favoring negative to positive values. Among these, GW191109 is one of the more extreme cases, exhibiting a shift more significant than 94% of the simulations. Therefore, we conclude that the χ_{eff} shift presented in Fig. 7.2 is *unlikely*, but not impossible, to be explained by a random Gaussian noise instantiation, i.e. without needing to invoke residual glitch power. In App. 7.10 we present further results based on a χ^2 test used in search algorithms that tracks how SNR is accumulated along the signal [60–62]. Consistent with Fig. 7.3, the test is inconclusive: the full-band analysis (Run 1 in Table 7.1) has behavior more extreme than most simulations, but it is not strongly

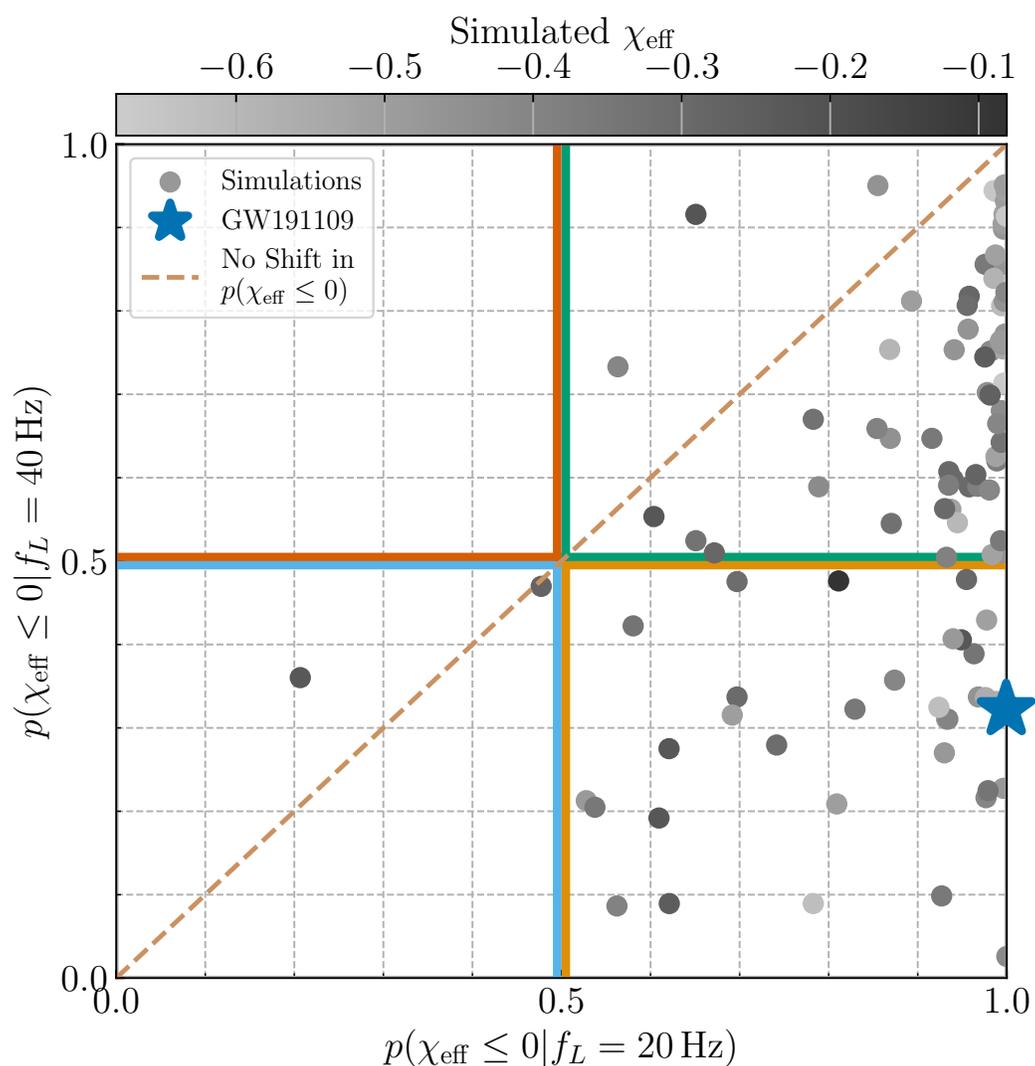


Figure 7.3: Shifts in the probability of $\chi_{\text{eff}} \leq 0$ for 100 simulated signals consistent with GW191109 in Gaussian noise (dots) and the real signal (cross). The x-axis corresponds to a full-band analysis, while the y-axis corresponds to a restricted-band analysis with $f_L = 40$ Hz. Going clockwise, the top left quadrant (red-orange axes) would contain cases where the posterior shifted from majority positive to negative (of which there were none), the top right quadrant (green axes) contains cases which were majority negative in both full- and restricted-band analyses, the bottom right quadrant (orange axes) contains cases which started majority negative and became majority positive (including GW191109), and the bottom left quadrant (blue axes) contains cases which were consistently majority positive. The $x = y$ line (dashed brown) corresponds to no shift in the probability for $\chi_{\text{eff}} \leq 0$.

inconsistent with them.

7.4 Glitch-Marginalized Inference

Having established that the 35–40 Hz data drive the negative χ_{eff} inference, we turn to the question of whether these data are meaningfully impacted by residual glitch power. We go beyond subtracting a single estimate for the glitch and simultaneously model both the signal and the glitch as described in Sec. 7.2. All analyses in this section use the original data in both detectors with no prior glitch mitigation. While this approach is robust against residual glitch power from subtracting a single glitch estimate, it is still impacted by modeling choices, specifically both the parametrized model (physical scattering model or wavelets) and the corresponding glitch parameter priors.

Since the 36 Hz excess power coincides in frequency with an arch predicted by the witness channel, Fig. 7.1, it is reasonable to expect it to be part of the scattering event and thus a prime target for the slow scattering model [29]. However, the time-frequency morphology of the 36 Hz excess power does not resemble scattering arches, which motivates the alternative wavelet-based glitch model. In principle, BAYESWAVE can fit any excess power by adding enough wavelets. Such a many-wavelet fit might be statistically disfavored, though, as it relies on a large number of parameters and a reduced posterior-to-prior volume. The exact quantitative impact of this Occam penalty is controlled by the wavelet parameter priors, which influence whether it is statistically favorable to add a wavelet to capture the excess power or instead attribute it to the signal. The most influential prior is likely the one for the wavelet amplitude, which —although broad— favors wavelets with SNR ~ 5 . The situation is further complicated by the the low LHO sensitivity in the relevant frequencies, which weakens its contribution to the likelihood, making the discrimination between glitch and signal even more dependent on the prior shape.

Slow scattering glitch model

We begin with the scattered light model in Fig. 7.4 (Run 13 in Table 7.1), which models five arches with a uniform amplitude prior and the “Targeted” modulation prior that is informed by the witness motion. The signal is modeled with NR-SUR7DQ4. The top panel shows a spectrogram of the data and the signal and glitch posteriors. The inferred glitch arches (multiple colors) match the witness prediction for the arch peak frequency spacing (~ 6 Hz) in the region of maximum glitch power. The optimal SNR ρ_{opt} posterior for each arch is shown in the bottom right

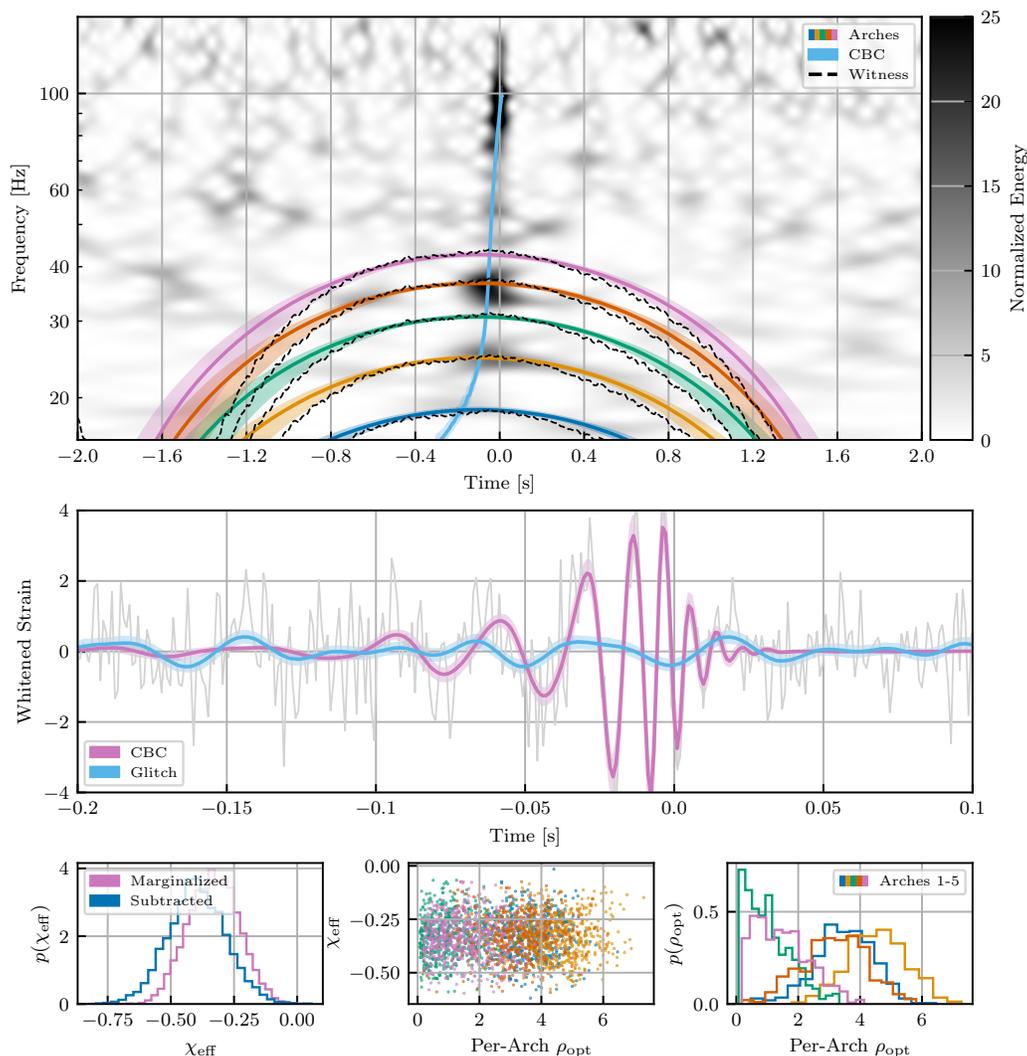


Figure 7.4: Jointly modeling the glitch with the physical slow scattering model and the signal with NRSUR7DQ4 (Run 13 in Table 7.1). In the top panel, we show a spectrogram of the data, along with the posterior for the glitch arches (median and 90% credible intervals; multiple colors), the signal track (blue), and the prediction of the witness channel (black dashed). In the middle panel, we show the whitened time-domain posterior reconstruction for the glitch (blue) and the signal (CBC; pink). In the bottom left panel, we show the marginalized χ_{eff} posterior from this analysis (pink), along with the equivalent result from glitch-subtracted data (Run 1 in Table 7.1; blue). In the bottom right panel, we show the marginalized posterior for the optimal SNR of each individual arch. Finally, in the bottom middle panel, we show a scatter plot of individual posterior samples in the $\rho_{\text{opt}} - \chi_{\text{eff}}$ plane for each arch, showing that no correlation exists.

panel, which reveals that three non-consecutive arches are confidently recovered with $\rho_{opt} > 0$: the first one at 18 Hz (blue), the second at 24 Hz (yellow), and the fourth at 36 Hz (orange). The third arch at 30 Hz has negligible SNR, $\rho_{opt} < 2$ at 88% credibility. Though seemingly surprising given the physical interpretation of scattered light based on bounces off of moving surfaces, a varying arch amplitude is commonly observed and the SNR further depends on the noise PSD that decreases with frequency in this range. The full glitch reconstruction in the time domain is plotted in the middle panel (blue) along with the signal (pink). As expected from the presence of multiple arches, the glitch does not have a constant frequency.

The χ_{eff} inference is presented in the bottom row. The bottom left panel shows the marginalized χ_{eff} posterior from this analysis (pink). For comparison, we also plot the posterior from the standard two-step analysis where the glitch has been pre-subtracted and only the signal is analyzed (Run 1 in Table 7.1; blue). Under glitch marginalization, χ_{eff} remains definitively negative at $\sim 100\%$ credibility, though the median increases from -0.40 to -0.33 . The glitch and χ_{eff} inference are uncorrelated, as shown in the bottom middle panel through a scatter plot for χ_{eff} and the optimal SNR of each arch.⁵ This suggests that even though there is a glitch arch at 36 Hz its time-frequency morphology does not match the 36 Hz excess power. Even when the signal and glitch are simultaneously modeled, most of the 36 Hz excess power is attributed to the signal and results in $\chi_{eff} < 0$. The time-domain reconstructions in the middle panel confirm this interpretation, with the signal reconstruction closely resembling those in Fig. 7.2, while a lower-amplitude glitch oscillation accounts for the remainder. We have verified that these χ_{eff} results are robust under alternative, yet reasonable, priors for the glitch: log-uniform in amplitude and the ‘‘Physical’’ modulation prior discussed in Sec. 7.2 (Runs 14, 15, and 16 in Table 7.1 for uniform amplitude with physical modulation, log-uniform amplitude with targeted modulation, and log-uniform amplitude with physical modulation respectively). We have also verified that other parameters, such as the binary total mass and mass ratio, remain consistent between glitch-subtracted and glitch-marginalized analyses.

To summarize, we conclude that the 36 Hz power is not exclusively due to the signal. Not only does the witness channel predict some glitch power, but also the

⁵Rather than the 36 Hz power, we attribute the small shift in the χ_{eff} median in Fig. 7.4 to the particular glitch realization that was subtracted for the GWTC-3 analysis. Indeed when we analyze the original data (no glitch mitigation) with only a signal (Run 12 in Table 7.1), we obtain a χ_{eff} posterior more similar to that of the marginalized analysis with a median χ_{eff} of -0.36 .

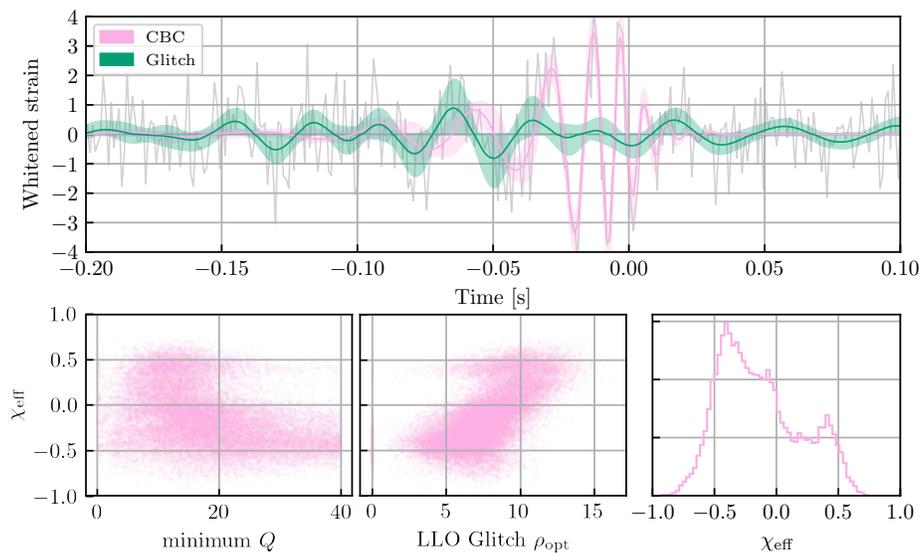


Figure 7.5: Jointly modeling the glitch with sine-Gaussian wavelets and the signal with IMRPHENOMXPHM (Run 21 in Table 7.1). The top panel shows the whitened time-domain data (grey) and median and 90% credible intervals for the glitch (green) and signal (CBC; pink). The bottom row displays marginalized posteriors. The right panel shows the glitch-marginalized χ_{eff} posterior, which displays a much larger spread than the results of Fig. 7.4, now being consistent with $\chi_{\text{eff}} = 0$. The left panel shows the scatter plot between χ_{eff} and the minimum quality factor Q among all wavelets of each posterior sample. Positive χ_{eff} is correlated with low Q . Scattered light is characterized by larger Q -values [34], confirming that $\chi_{\text{eff}} > 0$ only if the glitch does not match the expected scattered light morphology. The middle panel shows a scatter plot between χ_{eff} and the glitch SNR which are again correlated: higher glitch power leads to a more positive χ_{eff} .

slow scattering model places an $\rho_{\text{opt}} \approx 3$ arch, notably louder than its adjacent arches. However, the excess power is not entirely attributed to scattered light as it is morphologically inconsistent with a slow scattering arch.⁶ The $\chi_{\text{eff}} < 0$ inference, therefore, persists under the physical slow scattering interpretation of this glitch.

Wavelet glitch model

The physically-motivated slow scattering model finds some glitch power at 36 Hz but cannot account for the entire 36 Hz excess power. This might be because of modeling systematics, the presence of other (beyond slow-scattering) non-Gaussian

⁶In App. 7.9 we show that unphysical priors on the slow scattering parameters can indeed twist the model into fully absorbing the 36 Hz power and eliminating the $\chi_{\text{eff}} < 0$ inference. Such priors are, however, inconsistent with slow scattering, which forms the basis of the glitch model to begin with.

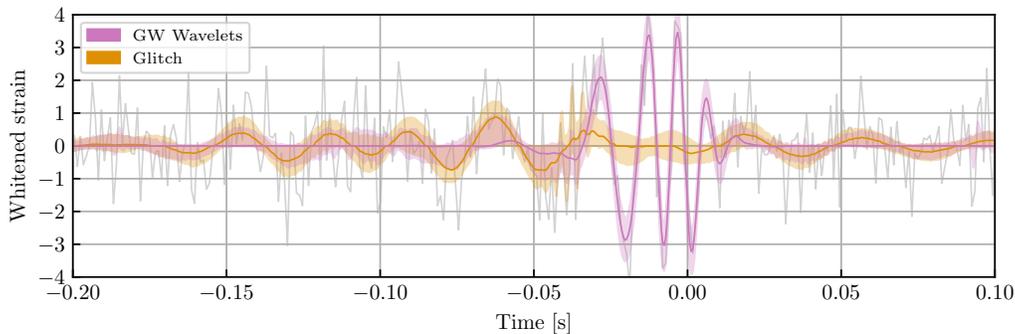


Figure 7.6: Jointly modeling both the glitch and the signal with sine-Gaussian wavelets (Run 20 in Table 7.1). We plot the whitened time-domain data (grey) and median and 90% credible intervals for the glitch (orange) and signal (GW Wavelets; purple). The 36 Hz excess power is consistent with originating from either the glitch or the signal at the 90% credible level.

noise, or simply because the 36 Hz excess power is indeed part of the signal. We explore these possibilities with `BAYESWAVE` and its more flexible wavelet-based glitch model as described in Sec. 7.2. We present two analyses: both marginalize over the glitch with wavelets but the GW signal is modeled with either the compact binary model `IMRPHENOMXPHM` or with coherent wavelets.

IMRPHENOMXPHM

In Fig. 7.5 we show results from the joint analysis with `IMRPHENOMXPHM` for the signal and wavelets for the glitch (Run 21 in Table 7.1). The top panel shows the whitened time-domain reconstructions. Compared to the reconstructions in Fig. 7.4 there is now increased uncertainty around the 36 Hz excess power, i.e. between times -0.09 and -0.04 s. This is due to the larger flexibility of the glitch model, which can now compete with the signal for the data around -0.06 s, leading to larger uncertainties for both models. The larger uncertainty is also reflected in the glitch-marginalized χ_{eff} posterior shown in the bottom right panel. Compared to Fig. 7.4, the χ_{eff} posterior is now much wider and entirely consistent with zero. It displays a broadly bimodal structure with one mode favoring $\chi_{\text{eff}} < 0$ and peaking at ~ -0.4 and the other favoring $\chi_{\text{eff}} > 0$ and peaking at ~ 0.4 . The antialigned mode is weakly favored at 70% of the posterior samples have $\chi_{\text{eff}} < 0$.

The increased χ_{eff} uncertainty is entirely due to the glitch and the competition between the signal and the glitch models. The bottom middle panel shows a posterior

scatter plot for χ_{eff} and the SNR of the glitch in LIGO Livingston.⁷ The glitch SNR is strongly correlated with χ_{eff} : a higher glitch power results in a more positive χ_{eff} . A small fraction of posterior samples, $\sim 6\%$, have vanishing glitch SNR (zero wavelets) and a strongly negative χ_{eff} , consistent with results from Fig. 7.4. Besides the glitch power, we examine the recovered glitch morphology in the bottom left panel, where we plot χ_{eff} against the minimum quality factor among wavelets in a particular posterior sample. The quality factor corresponds to the number of cycles in a wavelet, therefore scattering arches are characterized by larger values of Q [34]. This plot confirms the conclusions of Fig. 7.4: if the glitch is scattering-like (large Q), the model cannot capture the 36 Hz power, and χ_{eff} tends to be negative. Support for $\chi_{\text{eff}} > 0$ requires low values of Q which morphologically do not resemble scattering arches.

These results are qualitatively robust against different glitch priors. When using a prior for the amplitude of each wavelet that peaks at an SNR of 3 (instead of the default value of 5), we recover the same bimodal solution for χ_{eff} and the glitch SNR. However, the preference for the antialigned mode shifts from 70% to 60% suggesting that our quantitative results are impacted by the glitch prior at the few percent level. This shift is attributed to the fact that the updated prior makes it easier to low-SNR wavelets to be added to the posterior and thus capture the 36 Hz excess power away from the signal model. The impact of glitch priors is akin to the impact of compact-binary parameters on inference [63] and is expected to be more prominent for low-SNR glitches.

We perform a final sanity check by comparing the total (signal plus glitch) reconstructions of posterior samples with $\chi_{\text{eff}} > 0$ to those with $\chi_{\text{eff}} < 0$. Although the two posterior modes result in different interpretations of which parts of the data are signal and which are glitch, their sums are consistent with each other. This is expected as it is the total strain of signal-plus-glitch that is compared to the data to calculate the likelihood. So any solution must result in the same total strain. While we view this as a sanity check on the analysis convergence, it also suggests that there are two distinct ways to model the data, and this analysis does not strongly prefer one over the other.

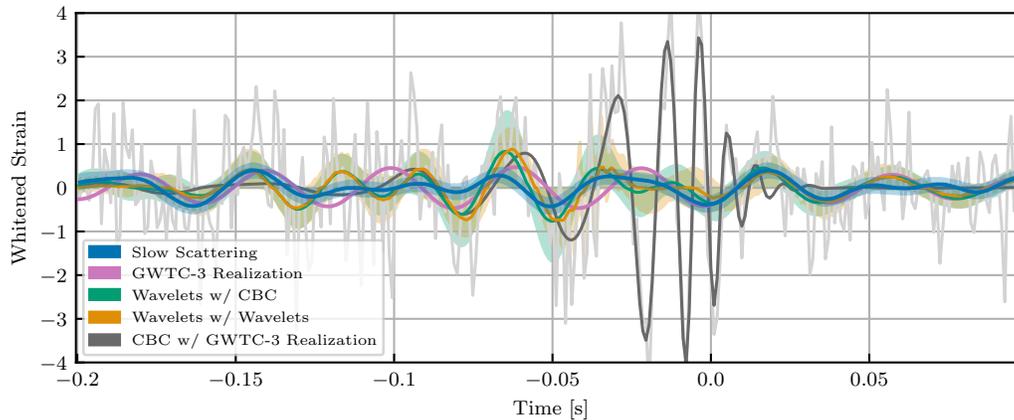


Figure 7.7: Comparison of reconstructions for the LIGO Livingston glitch that overlapped with GW191109 obtained by various analyses. The data are shown in grey, and for reference, we also show the maximum-likelihood GW reconstruction from the full-band analysis on the glitch-subtracted data in black (Run 1 in Table 7.1). The single realization subtracted for the GWTC-3 analysis is shown in pink [1]. The glitch inferred from the joint slow scattering and NRSUR7DQ4 analysis (Run 13 in Table 7.1) is shown in blue. The glitch inferred with wavelets is shown in green when the signal is modeled with IMRPHENOMXPHM (Run 21 in Table 7.1) and orange when the signal is also modeled with wavelets (Run 20 in Table 7.1).

Coherent wavelet model

For completeness, we present a final analysis where both the glitch and the GW are modeled with sums of wavelets [35] (Run 20 in Table 7.1). Since the signal model is now also phenomenological, we do not extract any binary parameters such as χ_{eff} which has thus far been guiding our conclusions. Instead, we directly interpret the time-domain reconstructions in Fig. 7.6. As expected, using more flexible models results in increased uncertainties. The 36 Hz (-0.06 s in the plot) power is still traded between the two models, and neither can rule out that it belongs to them at the 90% credible level. In contrast to the signal reconstructions thus far, Figs. 7.4 and 7.5, the coherent wavelet model is not able to confidently recover the signal inspiral between times -0.1 and -0.04 s. This is again due to the large flexibility of the wavelet signal model, which needs to extract each portion of the signal independently of the others [64] as opposed to the waveform model that coherently models the whole signal across inspiral and merger.

⁷This analysis allows for glitches in both detectors, but the Hanford data are consistent with no glitch power in the analysis window.

Comparing glitch reconstructions

Finally, we compare glitch reconstructions from the various glitch inferences considered in Fig. 7.7. The comparison includes the single glitch realization considered in GWTC-3 [1] and the three glitch-marginalized analyses presented in this study, Figs. 7.4, 7.5, and 7.6. The glitch reconstructions are largely consistent with each other, with the largest differences encountered in the crucial -0.06 s region. As expected, the wavelet-based reconstructions have a larger statistical uncertainty due to the larger model flexibility. This allows them to reach a larger amplitude at -0.06 s which is necessary in order to capture the 36 Hz excess power.

7.5 Conclusions

When seeking to interpret GW data in the presence of glitches, absolute confidence in all aspects of the analysis is impossible. Unlike compact-binary signals for which we have exact numerical relativity simulations to compare models against, glitch modeling does not have the luxury of a “ground truth” solution. Nonetheless, we have sought an understanding of GW191109, its astrophysically-influential $\chi_{\text{eff}} < 0$ inference, and the overlapping glitch within the limitations of imperfect glitch models and large statistical uncertainties.

We showed that the $\chi_{\text{eff}} < 0$ measurement is attributed to a segment of LIGO Livingston data occurring between 0.1 and 0.04 s before the merger, and between 30 and 40 Hz. These data are impacted by excess scattered light non-Gaussian noise, consistent with Ref. [28]. Simultaneously modeling the GW signal with compact-binary waveforms and the glitch yields results that depend on the glitch model. A physical glitch model tailored to slow scattering glitches cannot morphologically match the excess power observed in the 36 Hz range. Therefore the $\chi_{\text{eff}} < 0$ measurement still stands. A more flexible wavelet-based glitch model is instead able to fully account for the 36 Hz excess power and wipe out all support for $\chi_{\text{eff}} < 0$. Though witness channel information suggests that slow scattering was indeed what occurred during GW191109, we cannot rule out shortcomings of the slow scattering parametrized model or additional non-Gaussian noise.

Given this, we cannot make absolute statements about the properties of GW191109. If, as expected from witness channel information, the data contain Gaussian noise, a well-modeled slow scattering glitch, and a GW signal, then GW191109 likely had asymmetric masses and $\chi_{\text{eff}} < 0$, strongly implying a dynamical origin [8]. However, if other non-Gaussian noise was present in the data, or the glitch morphology varied

from classical slow scattering, spin inference becomes uninformative—though in any situation, GW191109 remains one of the heaviest observations to date. Distinguishing between these interpretations is challenging. Firstly, LIGO Hanford’s sensitivity in the relevant frequency range is diminished, it can therefore not contribute to the question of whether the crucial 36 Hz power is coherent (and thus part of the signal) or incoherent (and thus part of the glitch). Secondly, the overall low SNR of the glitch makes results depend on the glitch model priors, e.g. the `BAYESWAVE` glitch prior explored in Sec. 7.4.

Our analysis builds upon Refs. [28, 33] to propose a framework for in-depth analyses of glitch-afflicted data. The framework includes cross-detector comparisons, band- and time-limited analyses, simulated signals, marginalizing over the glitch, and exploring different glitch models (tailored to a specific glitch family or flexible) and prior assumptions.

As GW astronomy collects more data and seeks to constrain increasingly more subtle effects, mitigating systematics related to data quality presents a complementary challenge to waveform systematics. Similar to waveform systematics, data quality systematics can be particularly troublesome for spin inference, which typically leaves a subtle imprint on the data and is concentrated on a small (time or frequency) region of data. Studies such as the ones presented here and in Ref. [33] are based on targeted, intensive follow-up of selected events, hand-chosen for the astrophysically important inference. Data quality systematics aggregating over catalogs of detections require additional care to identify and mitigate in an automated way, e.g., [65]. Such efforts will be significantly aided by the work of experts in reducing the absolute rate of glitches, in characterizing the state of the detectors, and in developing efficient and statistically sound analyses in the presence of glitches. In this work we present techniques to help address these challenges moving forward.

7.6 Scattered light glitches in LHO

In Fig. 7.8 we show a spectrogram of the data in LHO at the time of GW191109 and the scattering tracks predicted by the witness channel. When GW191109 entered the LHO frequency band, the scattering surface motion was at a minimum, so that the signal and the glitch are disjoint in time and frequency. Accordingly, we expect source inference to be unaffected by the glitch. Reference [28] reached similar conclusions. We confirm this expectation by performing analyses which restrict the frequency band in LHO in a similar fashion to main-text LLO analyses. When

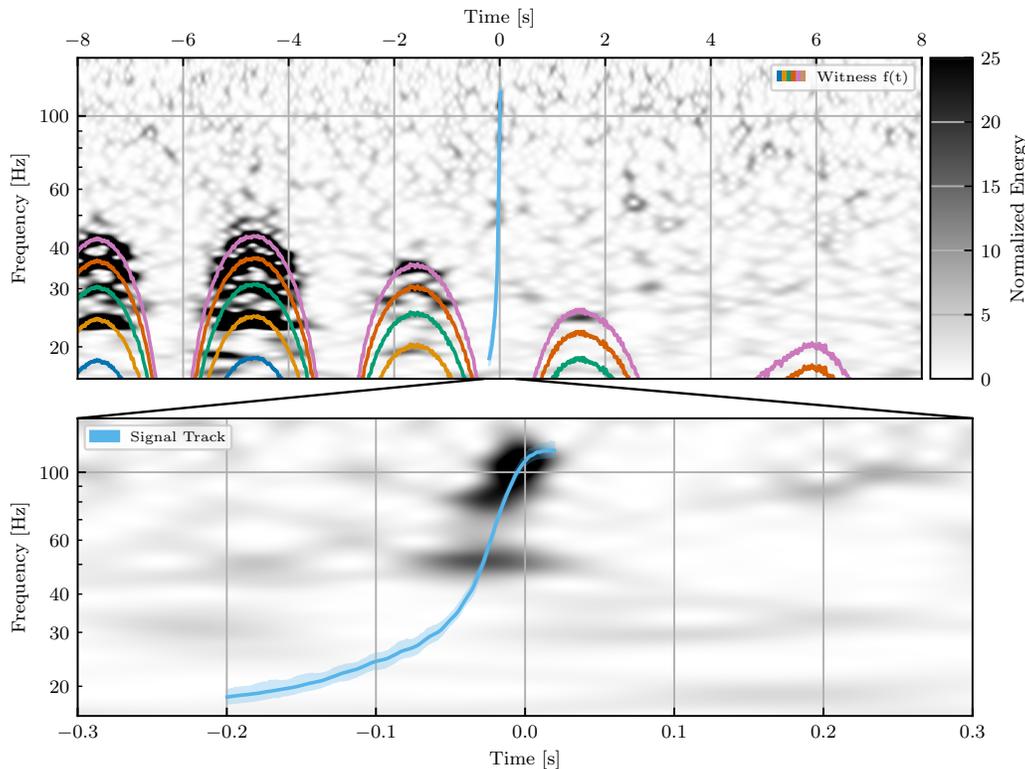


Figure 7.8: Similar to Fig. 7.1 but for LHO data at the time of GW191109, with the scattering tracks predicted by the motion of the witness channel H1:SUS-ETMX_L2_WIT_L_DQ. The absolute intensity of the slow scattering was significantly worse than in LLO, but the signal occurred at a minimum in the scattering, such that there is no overlap in time and frequency between the glitch arches and the GW191109 track (blue).

restricting to > 40 Hz in LHO (Run 6 in Table 7.1) but with no LLO restrictions, the χ_{eff} posterior remains almost entirely negative ($\chi_{\text{eff}} < 0$ at 99.9% credibility, the same as Run 1 in Table 7.1 that uses all data in both detectors). When removing sub-40 Hz data in both detectors (Run 7 in Table 7.1), we obtain a modestly positive result ($\chi_{\text{eff}} < 0$ at 33.3% credibility), but no more so than when we only restricting the LLO data (32.2%) (Run 4 in Table 7.1). While it is mildly surprising that removing so much low-frequency data in LHO has so little apparent effect on inference, we attribute this to the significant difference between LLO and LHO low-frequency sensitivity.

7.7 Detailed analysis settings

Run ID	Data	Glitch Model	Signal Model	Data Restrictions	Analysis Type
1	Subtracted	—	NRSUR7DQ4	—	BILBY-FD
2	Subtracted	—	NRSUR7DQ4	$f_L = 30$ Hz	BILBY-FD
3	Subtracted	—	NRSUR7DQ4	$f_L = 35$ Hz	BILBY-FD
4	Subtracted	—	NRSUR7DQ4	$f_L = 40$ Hz	BILBY-FD
5	Subtracted	—	NRSUR7DQ4	$f_L = 45$ Hz	BILBY-FD
6	Subtracted	—	NRSUR7DQ4	$f_H = 40$ Hz	BILBY-FD
7	Subtracted	—	NRSUR7DQ4	$f_L = f_H = 40$ Hz	BILBY-FD
8	Subtracted	—	NRSUR7DQ4	No LHO	BILBY-FD
9	Subtracted	—	NRSUR7DQ4	No LLO	BILBY-FD
10	Subtracted	—	NRSUR7DQ4	Various t_H, t_L	TD
11	Subtracted	—	IMRPHENOM- XPHM	—	BILBY-FD
12	Original	—	NRSUR7DQ4	—	BILBY-FD
13	Original	Slow Scattering (Uniform + Targeted)	NRSUR7DQ4	—	BILBY-FD
14	Original	Slow Scattering (Uniform + Physical)	NRSUR7DQ4	—	BILBY-FD

15	Original	Slow Scattering (Log-Uniform + Targeted)	NRSUR7DQ4	—	BILBY-FD
16	Original	Slow Scattering (Log Uniform + Physical)	NRSUR7DQ4	—	BILBY-FD
17	Original	Slow Scattering (Uniform + Targeted, N=4)	NRSUR7DQ4	—	BILBY-FD
18	Original	Slow Scattering (Uniform + Unphysical)	NRSUR7DQ4	—	BILBY-FD
19	Original	Slow Scattering (Uniform + Targeted)	IMRPHENOM-XPHM	—	BILBY-FD
20	Original	Wavelets	Wavelets	—	BAYESWAVE-FD
21	Original	Wavelets	IMRPHENOM-XPHM	—	BAYESWAVE-FD
22-121	Simulated	—	NRSUR7DQ4	—	BILBY-FD
122-221	Simulated	—	NRSUR7DQ4	$f_L = 40 \text{ Hz}$	BILBY-FD

Table 7.1: Settings and properties for all analyses presented in this work. We list from left to right: a unique run ID hyperlinked in the text, the type of data used (original or glitch-subtracted GWTC-3 data [42, 66]), how the glitch is modeled per Sec. 7.2, how the CBC signal is modeled per Sec. 7.2, frequency or time cuts on the data on top of the default settings, and the analysis type (software and data domain - FD for frequency and TD for time). Analyses based on glitch-subtracted data use the data provided by GWTC-3 [1], while analyses that marginalize over the glitch employ the original unmitigated data. Frequency bands are described by f_H and f_L designating the minimum frequency of analysis in LHO and LLO respectively. For runs which use the parameterized slow scattering model, the parenthetical descriptions correspond to the choice of amplitude prior and modulation frequency prior respectively for each run. All slow scattering analyses model five slow scattering arches, with the exception of Run 17.

In this appendix we provide details about the settings of all analyses presented in this study. Table 7.1 identifies all analyses with a unique index, referenced throughout the text. We also list the data analyzed, the relevant glitch and signal models, any restrictions applied to the data being analyzed, and the analysis type (both the software used and the data domain in which it operates).

7.8 IMRPHENOMXPHM Analyses with BILBY

To assess whether differences between BAYESWAVE results and BILBY results are due to waveform systematics, we also perform two analyses using BILBY and IMRPHENOMXPHM: one on subtracted data (Run 11 in Table 7.1), and one using the slow scattering glitch model (Run 19 in Table 7.1). The analysis on subtracted data found $\chi_{\text{eff}} \leq 0$ at 99.3% credibility, while the analysis marginalizing over the slow scattering model found $\chi_{\text{eff}} \leq 0$ at 99.9% credibility. From this we conclude that the observed differences between BILBY and BAYESWAVE are due to the choice of glitch model, rather than the choice of waveform approximant.

7.9 Alternate Slow Scattering Glitch Priors

To test our assumption that it is appropriate to use the slow scattering model with five scattering arches, we also perform a test using four scattering arches with uniform amplitude and targeted modulation frequency priors (Run 17 in Table 7.1). This result finds $\chi_{\text{eff}} \leq 0$ with $\sim 100\%$ credibility, indicating that the inclusion of an arch around 42 Hz does not alter the conclusions of this work.

The slow scattering model under the physically expected range of modulation frequencies $f_{\text{mod}} \sim \mathcal{U}(0.05 - 0.3)$ Hz results in arches that are too extended in time to match the 36 Hz excess power morphology. We explore what values of f_{mod} are required in order to impact χ_{eff} inference, with an analysis that employs a uniform amplitude prior and a maximum modulation frequency of 5 Hz (Run 18 in Table 7.1). We recover a tri-modal structure favoring $f_{\text{mod}} = 1.5$ Hz and a less negative χ_{eff} , with $\chi_{\text{eff}} < 0$ at 77.1% credibility. However, $f_{\text{mod}} = 1.5$ Hz is 10 times larger than the scattering surface motion witnessed by the channel L1:SUS-ETMX_L2_WIT_L_DQ. Such a result would presume the existence of some alternative source of frequency modulated phase noise, either due to another scattering surface driven at a different frequency, or some non-scattering mechanism, coincidentally aligned in time and frequency with the known scatterer. While we cannot rule out the existence of such a source, there is no physical motivation to presuppose its existence. We instead use this analysis to emphasize the conclusion from the BAYESWAVE study, namely that sufficiently flexible glitch models allow for a wider range of possibilities.

7.10 Frequency Bin χ^2 Test

Tests which assess the Gaussianity of data [67, 68] may be applied to residual data after glitch and signal subtraction, but these do not address whether the signal model is capturing any glitch power. In this appendix we instead consider the frequency bin χ^2 test as employed by search algorithms [60–62]. Qualitatively, it assesses tension between the signal waveform and the data over the entire frequency band, and hence measures deviations due both to model misspecification and to distribution of power not characteristic of a CBC, e.g., a glitch.

For each posterior waveform, we divide the frequency band into p bins of equal optimal SNR. If the data are consistent with the sum of the waveform in question and Gaussian noise, then the matched-filter SNR will also be evenly distributed over these bins. For the j th bin, the matched-filter SNR $\rho_{\text{mf},j}$ will deviate from the mean

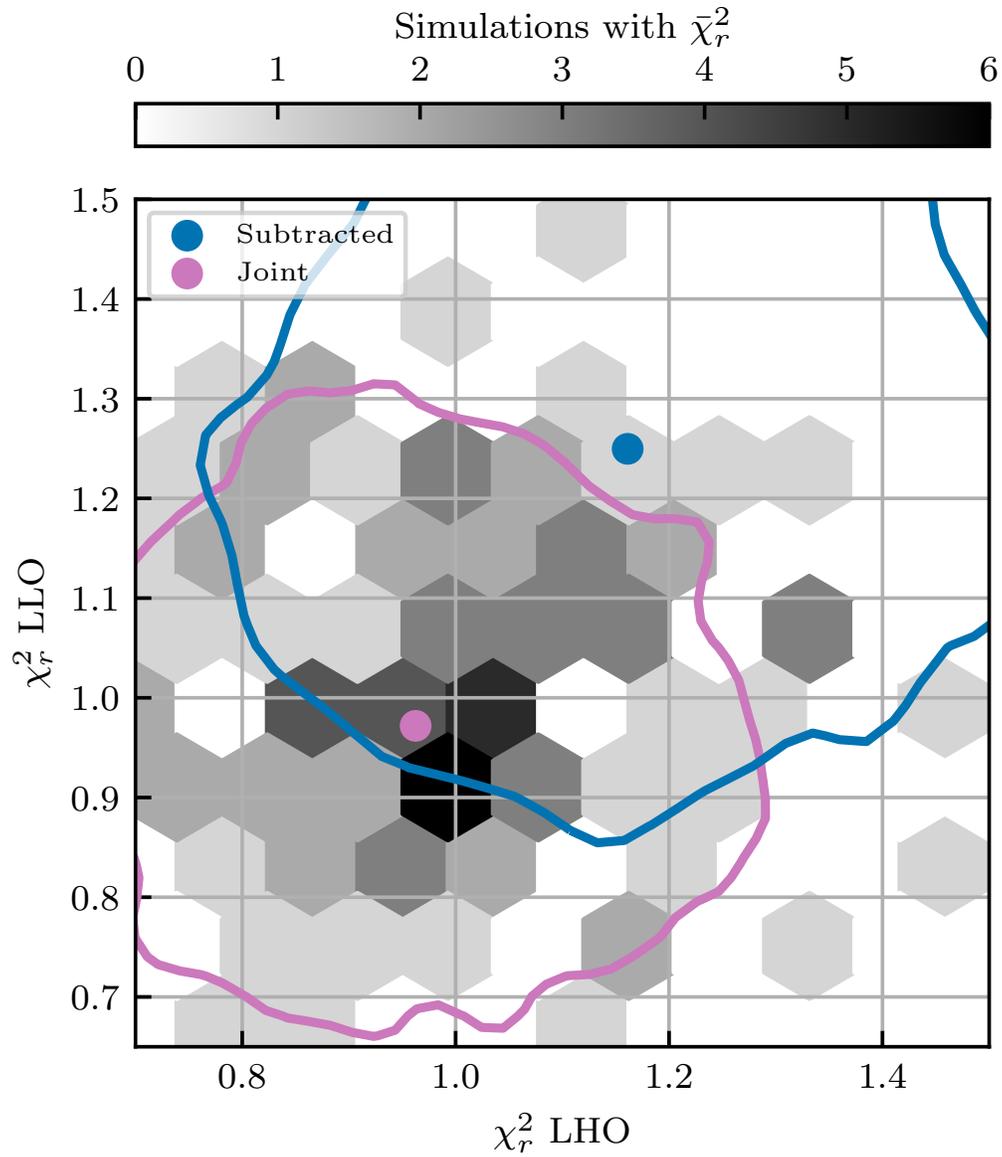


Figure 7.9: Distribution of χ_r^2 in both detectors for the glitch-subtracted analysis (blue; Run 1) and the glitch-marginalized analysis (purple; Run 13). The colormap corresponds to the distribution of $\bar{\chi}_r^2$ from simulated signals consistent with GW191109, Runs 22-121. Dots denote the distribution mean and contours denote the 90% level. For the reference distribution, we histogram the $\bar{\chi}_r^2$ values in LHO and LLO from each simulation.

ρ_{mf} by

$$\Delta\rho_{\text{mf},j} = \rho_{\text{mf},j} - \frac{\rho_{\text{mf}}}{p}. \quad (7.5)$$

The statistic

$$\chi^2 = p \sum_{j=1}^p |\Delta\rho_{\text{mf},j}|^2 \quad (7.6)$$

is distributed according to a χ^2 distribution with $2p - 2$ degrees of freedom under Gaussian noise[60].⁸ The normalized statistic

$$\chi_r^2 = \frac{\chi^2}{2p - 2} \quad (7.7)$$

will then have an expected value of 1. Deviations indicate that the data might not be solely described by the waveform plus Gaussian noise, likely due to a glitch. We compute χ_r^2 for each GW191109 signal posterior sample on data where the corresponding glitch posterior sample has been subtracted. We denote the mean statistic over posterior samples as $\bar{\chi}_r^2$. We then compare against corresponding results from the simulated signals of Sec. 7.3. The reason we compare against simulations rather than directly the frequentist expectation for Eq. (7.7) is that the distribution over the posterior samples is not equivalent to a distribution over many Gaussian noise realizations.

In Fig. 7.9 we plot the statistic distribution over posterior samples in both detectors for the glitch-subtracted analysis of Run 1 and the glitch-marginalized analysis with the slow-scattering model of Run 13. The colormap corresponds to results from simulated signals where we bin the mean statistic $\bar{\chi}_r^2$ for each simulated signal. Glitch-marginalization results in a statistic whose mean is more closely in accordance with the frequentist expectation value in both detectors ($\bar{\chi}_r^2 = 0.96$ and $\bar{\chi}_r^2 = 0.97$ in LHO and LLO respectively) than glitch-subtraction ($\bar{\chi}_r^2 = 1.16$ and $\bar{\chi}_r^2 = 1.24$ in LHO and LLO). Compared to the simulated signals, glitch-marginalization results in $\bar{\chi}_r^2$ more extreme than that of 41% (45%) of simulations in LHO (LLO), while the glitch-subtracted result has a $\bar{\chi}_r^2$ more extreme than 81% (90%) of simulations in LHO (LLO). To produce a meta-statistic, we use Fisher's method [69] to compute the likelihood of these statistics occurring together, assuming that the p-values are uncorrelated. This creates another χ^2 statistic, this time with two degrees of freedom per detector. For the glitch-marginalized result, we obtain 2.25, corresponding to a p-value of 0.69, while for the glitch-subtracted results we have 7.93, giving a

⁸Two degrees of freedom correspond to the real and imaginary components in each bin, while two are removed since deviations must sum to zero in each of the real and imaginary components.

p -value of 0.09. Consistent with expectations, glitch-marginalization results in residuals that are more consistent with Gaussian noise after removing the glitch and signal reconstruction.

References

- ¹R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), “GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run”, *Phys. Rev. X* **13**, 041039 (2023) doi:10.1103/PhysRevX.13.041039.
- ²E. Racine, “Analysis of spin precession in binary black hole systems including quadrupole-monopole interaction”, *Phys. Rev. D* **78**, 044021 (2008) doi:10.1103/PhysRevD.78.044021.
- ³P. Ajith et al., “Inspirals-mergers-ringdown waveforms for black-hole binaries with non-precessing spins”, *Phys. Rev. Lett.* **106**, 241101 (2011) doi:10.1103/PhysRevLett.106.241101.
- ⁴L. Santamaria et al., “Matching post-Newtonian and numerical relativity waveforms: systematic errors and a new phenomenological model for non-precessing black hole binaries”, *Phys. Rev. D* **82**, 064016 (2010) doi:10.1103/PhysRevD.82.064016.
- ⁵I. M. Romero-Shaw, P. D. Lasky, and E. Thrane, “Four Eccentric Mergers Increase the Evidence that LIGO–Virgo–KAGRA’s Binary Black Holes Form Dynamically”, *Astrophys. J.* **940**, 171 (2022) doi:10.3847/1538-4357/ac9798.
- ⁶N. Gupte et al., “Evidence for eccentricity in the population of binary black holes observed by LIGO-Virgo-KAGRA”, (2024) doi:10.48550/arXiv.2404.14286.
- ⁷A. Antonelli, K. Kritos, K. K. Y. Ng, R. Cotesta, and E. Berti, “Classifying the generation and formation channels of individual LIGO-Virgo-KAGRA observations from dynamically formed binaries”, *Phys. Rev. D* **108**, 084044 (2023) doi:10.1103/PhysRevD.108.084044.
- ⁸R. C. Zhang, G. Fragione, C. Kimball, and V. Kalogera, “On the Likely Dynamical Origin of GW191109 and Binary Black Hole Mergers with Negative Effective Spin”, *Astrophys. J.* **954**, 23 (2023) doi:10.3847/1538-4357/ace4c1.
- ⁹H. Tong, S. Galadage, and E. Thrane, “Population properties of spinning black holes using the gravitational-wave transient catalog 3”, *Phys. Rev. D* **106**, 103019 (2022) doi:10.1103/PhysRevD.106.103019.
- ¹⁰C. Adamcewicz, P. D. Lasky, and E. Thrane, “Evidence for a Correlation between Binary Black Hole Mass Ratio and Black Hole Spins”, *Astrophys. J.* **958**, 13 (2023) doi:10.3847/1538-4357/acf763.

- ¹¹R. Farmer, M. Renzo, S. E. de Mink, P. Marchant, and S. Justham, “Mind the gap: The location of the lower edge of the pair instability supernovae black hole mass gap”, *Astrophys. J.* **887**, 53 (2019) doi:10.3847/1538-4357/ab518b.
- ¹²S. E. Woosley and A. Heger, “The Pair-Instability Mass Gap for Black Holes”, *Astrophys. J. Lett.* **912**, L31 (2021) doi:10.3847/2041-8213/abf2c4.
- ¹³V. Kalogera, “Spin orbit misalignment in close binaries with two compact objects”, *Astrophys. J.* **541**, 319–328 (2000) doi:10.1086/309400.
- ¹⁴D. Gerosa, E. Berti, R. O’Shaughnessy, K. Belczynski, M. Kesden, D. Wysocki, and W. Gladysz, “Spin orientations of merging black holes formed from the evolution of stellar binaries”, *Phys. Rev. D* **98**, 084036 (2018) doi:10.1103/PhysRevD.98.084036.
- ¹⁵M. Zevin, I. M. Romero-Shaw, K. Kremer, E. Thrane, and P. D. Lasky, “Implications of Eccentric Observations on Binary Black Hole Formation Channels”, *Astrophys. J. Lett.* **921**, L43 (2021) doi:10.3847/2041-8213/ac32dc.
- ¹⁶I. M. Romero-Shaw, P. D. Lasky, and E. Thrane, “Signs of Eccentricity in Two Gravitational-wave Signals May Indicate a Subpopulation of Dynamically Assembled Binary Black Holes”, *Astrophys. J. Lett.* **921**, L31 (2021) doi:10.3847/2041-8213/ac3138.
- ¹⁷P. C. Peters, “Gravitational Radiation and the Motion of Two Point Masses”, *Phys. Rev.* **136**, B1224–B1232 (1964) doi:10.1103/PhysRev.136.B1224.
- ¹⁸G. Pratten et al., “Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes”, *Phys. Rev. D* **103**, 104056 (2021) doi:10.1103/PhysRevD.103.104056.
- ¹⁹S. Ossokine et al., “Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation”, *Phys. Rev. D* **102**, 044055 (2020) doi:10.1103/PhysRevD.102.044055.
- ²⁰G. Ashton et al., “BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy”, *Astrophys. J. Suppl.* **241**, 27 (2019) doi:10.3847/1538-4365/ab06fc.
- ²¹I. 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**, 3295–3319 (2020) doi:10.1093/mnras/staa2850.
- ²²J. Lange, R. O’Shaughnessy, and M. Rizzo, “Rapid and accurate parameter inference for coalescing, precessing compact binaries”, <https://arxiv.org/abs/1805.10457> (2018).
- ²³V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, “Surrogate models for precessing binary black hole simulations with unequal masses”, *Phys. Rev. Research.* **1**, 033015 (2019) doi:10.1103/PhysRevResearch.1.033015.

- ²⁴T. Islam, A. Vajpeyi, F. H. Shaik, C.-J. Haster, V. Varma, S. E. Field, J. Lange, R. O’Shaughnessy, and R. Smith, “Analysis of GWTC-3 with fully precessing numerical relativity surrogate models”, (2023) DOI:10.48550/arXiv.2309.14473.
- ²⁵M. Hannam et al., “General-relativistic precession in a black-hole binary”, *Nature* **610**, 652–655 (2022) DOI:10.1038/s41586-022-05212-z.
- ²⁶P. Schmidt, F. Ohme, and M. Hannam, “Towards models of gravitational waveforms from generic binaries II: Modelling precession effects with a single effective precession parameter”, *Phys. Rev. D* **91**, 024043 (2015) DOI:10.1103/PhysRevD.91.024043.
- ²⁷J. Aasi et al. (LIGO Scientific), “Advanced LIGO”, *Class. Quant. Grav.* **32**, 074001 (2015) DOI:10.1088/0264-9381/32/7/074001.
- ²⁸D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. Di Renzo, and G. Ashton, “Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run”, *Class. Quant. Grav.* **39**, 245013 (2022) DOI:10.1088/1361-6382/aca238.
- ²⁹R. Udall and D. Davis, “Bayesian modeling of scattered light in the LIGO interferometers”, *Appl. Phys. Lett.* **122**, 094103 (2023) DOI:10.1063/5.0136896.
- ³⁰F. Acernese (Virgo), “The Advanced Virgo detector”, *J. Phys. Conf. Ser.* **610**, edited by G. Ciani, J. W. Conklin, and G. Mueller, 012014 (2015) DOI:10.1088/1742-6596/610/1/012014.
- ³¹C. Pankow et al., “Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817”, *Phys. Rev. D* **98**, 084016 (2018) DOI:10.1103/PhysRevD.98.084016.
- ³²K. Chatziioannou, N. Cornish, M. Wijngaarden, and T. B. Littenberg, “Modeling compact binary signals and instrumental glitches in gravitational wave data”, *Phys. Rev. D* **103**, 044013 (2021) DOI:10.1103/PhysRevD.103.044013.
- ³³E. Payne, S. Hourihane, J. Golomb, R. Udall, D. Davis, and K. Chatziioannou, “Curious case of GW200129: Interplay between spin-precession inference and data-quality issues”, *Phys. Rev. D* **106**, 104017 (2022) DOI:10.1103/PhysRevD.106.104017.
- ³⁴S. Hourihane, K. Chatziioannou, M. Wijngaarden, D. Davis, T. Littenberg, and N. Cornish, “Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data”, *Phys. Rev. D* **106**, 042006 (2022) DOI:10.1103/PhysRevD.106.042006.
- ³⁵S. Ghonge, J. Brandt, J. M. Sullivan, M. Millhouse, K. Chatziioannou, J. A. Clark, T. Littenberg, N. Cornish, S. Hourihane, and L. Cadonati, “Assessing and mitigating the impact of glitches on gravitational-wave parameter estimation: A model agnostic approach”, *Phys. Rev. D* **110**, 122002 (2024) DOI:10.1103/PhysRevD.110.122002.

- ³⁶R. Macas, A. Lundgren, and G. Ashton, “Revisiting the evidence for precession in GW200129 with machine learning noise mitigation”, *Phys. Rev. D* **109**, 062006 (2024) doi:10.1103/PhysRevD.109.062006.
- ³⁷N. J. Cornish and T. B. Littenberg, “BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches”, *Class. Quant. Grav.* **32**, 135012 (2015) doi:10.1088/0264-9381/32/13/135012.
- ³⁸T. B. Littenberg and N. J. Cornish, “Bayesian inference for spectral estimation of gravitational wave detector noise”, *Phys. Rev. D* **91**, 084034 (2015) doi:10.1103/PhysRevD.91.084034.
- ³⁹N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, “BayesWave analysis pipeline in the era of gravitational wave observations”, *Phys. Rev. D* **103**, 044006 (2021) doi:10.1103/PhysRevD.103.044006.
- ⁴⁰M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, “Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms”, *Phys. Rev. Lett.* **113**, 151101 (2014) doi:10.1103/PhysRevLett.113.151101.
- ⁴¹S. J. Miller, M. Isi, K. Chatziioannou, V. Varma, and I. Mandel, “GW190521: Tracing imprints of spin-precession on the most massive black hole binary”, *Phys. Rev. D* **109**, 024024 (2024) doi:10.1103/PhysRevD.109.024024.
- ⁴²R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), “Open Data from the Third Observing Run of LIGO, Virgo, KAGRA, and GEO”, *Astrophys. J. Suppl.* **267**, 29 (2023) doi:10.3847/1538-4365/acdc9f.
- ⁴³R. Abbot et al. (LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration), “Gwtc-3: compact binary coalescences observed by ligo and virgo during the second part of the third observing run — parameter estimation data release”, (2023) doi:10.5281/zenodo.8177023.
- ⁴⁴J. S. Speagle, “dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences”, *Mon. Not. Roy. Astron. Soc.* **493**, 3132–3158 (2020) doi:10.1093/mnras/staa278.
- ⁴⁵M. Isi and W. M. Farr, “Analyzing black-hole ringdowns”, (2021) doi:10.48550/arXiv.2107.05609.
- ⁴⁶M. Isi, W. M. Farr, M. Giesler, M. A. Scheel, and S. A. Teukolsky, “Testing the Black-Hole Area Law with GW150914”, *Phys. Rev. Lett.* **127**, 011103 (2021) doi:10.1103/PhysRevLett.127.011103.
- ⁴⁷M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky, “Testing the no-hair theorem with GW150914”, *Phys. Rev. Lett.* **123**, 111102 (2019) doi:10.1103/PhysRevLett.123.111102.

- ⁴⁸G. Carullo, W. Del Pozzo, and J. Veitch, “Observational Black Hole Spectroscopy: A time-domain multimode analysis of GW150914”, *Phys. Rev. D* **99**, [Erratum: *Phys.Rev.D* 100, 089903 (2019)], 123029 (2019) doi:10.1103/PhysRevD.99.123029.
- ⁴⁹C. D. Capano, M. Cabero, J. Westerweck, J. Abedi, S. Kastha, A. H. Nitz, Y.-F. Wang, A. B. Nielsen, and B. Krishnan, “Multimode Quasinormal Spectrum from a Perturbed Black Hole”, *Phys. Rev. Lett.* **131**, 221402 (2023) doi:10.1103/PhysRevLett.131.221402.
- ⁵⁰T. Accadia et al., “Noise from scattered light in Virgo’s second science run data”, *Class. Quant. Grav.* **27**, edited by F. Ricci, 194011 (2010) doi:10.1088/0264-9381/27/19/194011.
- ⁵¹S. Soni et al. (LIGO), “Reducing scattered light in LIGO’s third observing run”, *Class. Quant. Grav.* **38**, 025016 (2020) doi:10.1088/1361-6382/abc906.
- ⁵²D. Davis et al. (LIGO), “LIGO detector characterization in the second and third observing runs”, *Class. Quant. Grav.* **38**, 135014 (2021) doi:10.1088/1361-6382/abfd85.
- ⁵³A. E. Tolley, G. S. Cabourn Davies, I. W. Harry, and A. P. Lundgren, “ArchEnemy: removing scattered-light glitches from gravitational wave data”, *Class. Quant. Grav.* **40**, 165005 (2023) doi:10.1088/1361-6382/ace22f.
- ⁵⁴B. P. Abbott et al. (LIGO Scientific, Virgo), “GW170817: Measurements of neutron star radii and equation of state”, *Phys. Rev. Lett.* **121**, 161101 (2018) doi:10.1103/PhysRevLett.121.161101.
- ⁵⁵B. P. Abbott et al. (LIGO Scientific, Virgo), “GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs”, *Phys. Rev. X* **9**, 031040 (2019) doi:10.1103/PhysRevX.9.031040.
- ⁵⁶R. Abbott et al. (LIGO Scientific, Virgo), “GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run”, *Phys. Rev. X* **11**, 021053 (2021) doi:10.1103/PhysRevX.11.021053.
- ⁵⁷C. Cutler and J. Harms, “BBO and the neutron-star-binary subtraction problem”, *Phys. Rev. D* **73**, 042001 (2006) doi:10.1103/PhysRevD.73.042001.
- ⁵⁸K. Chatziioannou, C.-J. Haster, T. B. Littenberg, W. M. Farr, S. Ghonge, M. Millhouse, J. A. Clark, and N. Cornish, “Noise spectral estimation methods and their impact on gravitational wave measurement of compact binary mergers”, *Phys. Rev. D* **100**, 104004 (2019) doi:10.1103/PhysRevD.100.104004.
- ⁵⁹C. Plunkett, S. Hourihane, and K. Chatziioannou, “Concurrent estimation of noise and compact-binary signal parameters in gravitational-wave data”, *Phys. Rev. D* **106**, 104021 (2022) doi:10.1103/PhysRevD.106.104021.
- ⁶⁰B. Allen, “ χ^2 time-frequency discriminator for gravitational wave detection”, *Phys. Rev. D* **71**, 062001 (2005) doi:10.1103/PhysRevD.71.062001.

- ⁶¹S. A. Usman et al., “The PyCBC search for gravitational waves from compact binary coalescence”, *Class. Quant. Grav.* **33**, 215004 (2016) DOI:10.1088/0264-9381/33/21/215004.
- ⁶²D. Davis, L. V. White, and P. R. Saulson, “Utilizing aLIGO Glitch Classifications to Validate Gravitational-Wave Candidates”, *Class. Quant. Grav.* **37**, 145001 (2020) DOI:10.1088/1361-6382/ab91e6.
- ⁶³S. Vitale, D. Gerosa, C.-J. Haster, K. Chatziioannou, and A. Zimmerman, “Impact of Bayesian Priors on the Characterization of Binary Black Hole Coalescences”, *Phys. Rev. Lett.* **119**, 251103 (2017) DOI:10.1103/PhysRevLett.119.251103.
- ⁶⁴S. Ghonge, K. Chatziioannou, J. A. Clark, T. Littenberg, M. Millhouse, L. Cadonati, and N. Cornish, “Reconstructing gravitational wave signals from binary black hole mergers with minimal assumptions”, *Phys. Rev. D* **102**, 064056 (2020) DOI:10.1103/PhysRevD.102.064056.
- ⁶⁵J. Heinzl, C. Talbot, G. Ashton, and S. Vitale, “Inferring the astrophysical population of gravitational wave sources in the presence of noise transients”, *Mon. Not. Roy. Astron. Soc.* **523**, 5972–5984 (2023) DOI:10.1093/mnras/stad1823.
- ⁶⁶R. Abbot et al. (LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration), “Gwtc-3: compact binary coalescences observed by ligo and virgo during the second part of the third observing run — glitch modelling for events”, (2021) DOI:10.5281/zenodo.5546680.
- ⁶⁷R. Macas and A. Lundgren, “Sensitive test of non-Gaussianity in gravitational-wave detector data”, *Phys. Rev. D* **108**, 063016 (2023) DOI:10.1103/PhysRevD.108.063016.
- ⁶⁸S. Yamamura, H. Yuzurihara, T. Yamamoto, and T. Uchiyama, “Extension of the characterization of non-Gaussianity in gravitational wave detectors with a statistical hypothesis test”, *Class. Quant. Grav.* **41**, 205008 (2024) DOI:10.1088/1361-6382/ad7665.
- ⁶⁹R. A. Fisher, Statistical methods for research workers, 5 (Oliver and Boyd, 1928), DOI:10.1007/978-1-4612-4380-9_6.

OTHER GLITCH INVESTIGATIONS

This chapter contains analysis from two papers which are works in progress. The first, from which Section 8.1, Section 8.2, and Section 8.3 are derived, will focus on the residuals of glitch subtraction, and the biases which these may produce in subsequent parameter estimation. The second, from which Section 8.4 derives, will focus on statistical tests to assess whether parameter estimation results are affected by the presence of glitch, extending the work done in Section 7.10.

8.1 Glitch residuals

Just as the presence of noise introduces uncertainties in our estimates of astrophysical parameters, it also introduces uncertainty in our reconstruction of glitch morphology. In turn, this uncertainty means that for the subtraction of any inferred glitch realization, the SNR of the residual will be non-zero. This problem has been investigated in the context of the LISA global fit [1, 2]. In the linear signal regime, we may estimate the error of the best fit parameters [3]

$$\delta\theta^\alpha \equiv \hat{\theta}^\alpha - \theta_0^\alpha \approx (\Gamma^{-1})^{\alpha\beta} \langle \mathbf{n} | \partial_\beta \mathbf{h} \rangle, \quad (8.1)$$

where $\hat{\theta}^\alpha$ is the maximum likelihood estimate of the signal parameters θ^α , θ_0^α are the true signal parameters, and $(\Gamma^{-1})^{\alpha\beta}$ is the inverse Fisher matrix, where the Fisher matrix is defined as

$$\Gamma_{\alpha\beta} = \left(\frac{\partial h}{\partial \theta^\alpha} \middle| \frac{\partial h}{\partial \theta^\beta} \right). \quad (8.2)$$

From this also follows a linear approximation for the residual waveform

$$\delta \mathbf{h} \equiv \hat{\mathbf{h}} - \mathbf{h}_0 \approx \partial_\alpha \mathbf{h} \delta\theta^\alpha. \quad (8.3)$$

From this, one may compute the expected residual SNR

$$\mathbb{E}(\langle \delta \mathbf{h} | \delta \mathbf{h} \rangle) = \mathbb{E}(\langle \partial_\alpha \mathbf{h} \delta\theta^\alpha | \partial_\beta \mathbf{h} \delta\theta^\beta \rangle) \quad (8.4)$$

$$= \Gamma_{\alpha\beta} \mathbb{E}(\langle \delta\theta^\alpha | \delta\theta^\beta \rangle) \quad (8.5)$$

$$= \Gamma_{\alpha\beta} \mathbb{E}((\Gamma^{-1})^{\alpha\gamma} \langle \mathbf{n} | \partial_\gamma \mathbf{h} \rangle (\Gamma^{-1})^{\beta\epsilon} \langle \mathbf{n} | \partial_\epsilon \mathbf{h} \rangle) \quad (8.6)$$

$$= \Gamma_{\alpha\beta} (\Gamma^{-1})^{\alpha\gamma} (\Gamma^{-1})^{\beta\epsilon} \mathbb{E}(\langle \mathbf{n} | \partial_\gamma \mathbf{h} \rangle \langle \mathbf{n} | \partial_\epsilon \mathbf{h} \rangle) \quad (8.7)$$

$$= \Gamma_{\alpha\beta} (\Gamma^{-1})^{\alpha\gamma} (\Gamma^{-1})^{\beta\epsilon} \Gamma_{\gamma\epsilon} \quad (8.8)$$

$$= \Gamma_{\alpha\beta} (\Gamma^{-1})^{\alpha\beta} = N_p. \quad (8.9)$$

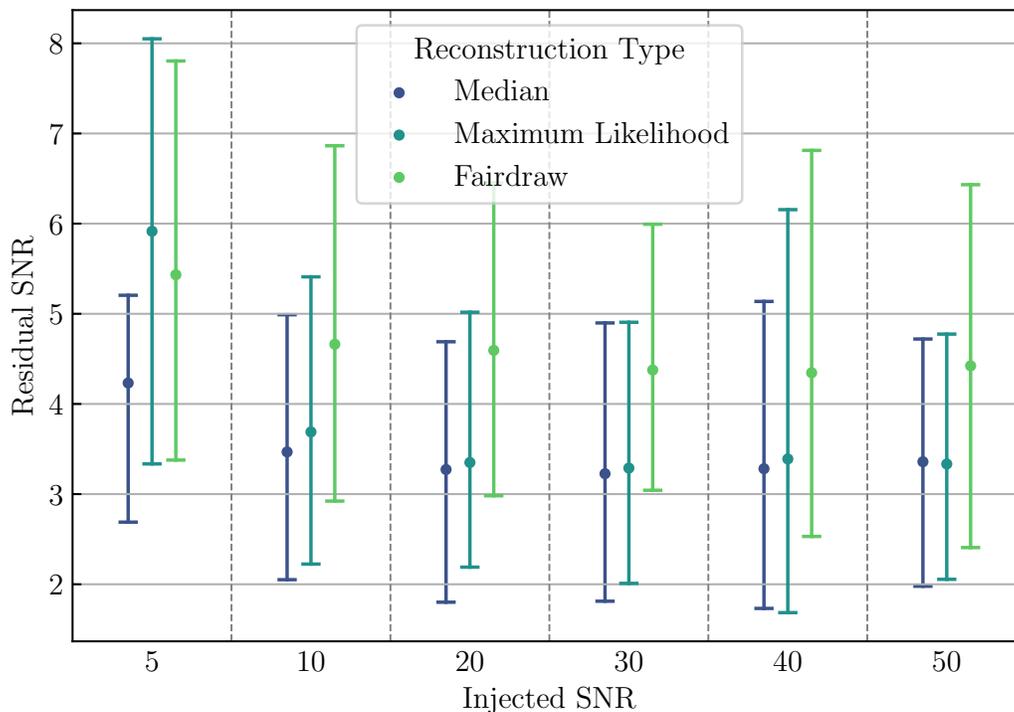


Figure 8.1: Residual SNR when subtracting point realizations of a slow scattering glitch, as inferred by PE.

where N_p is the number of parameters in the model. Importantly, this result holds *regardless of the SNR of the true signal*, and as a statistical error it is present even when there are no systematics due to mismodeling. Models with greater numbers of parameters, and hence more degrees of freedom, will have greater residual SNR since the added flexibility allows the model to more effectively capture some of the noise, leading it further astray.

We investigate this result empirically by simulating a large number of glitches at varying SNRs, then subtracting inferred glitch realizations. While the above result holds for the maximum likelihood point, we expect similar, though smaller, errors to occur for other methods of drawing realizations from the posterior. For this reason, we consider three methods of choosing a glitch realization to subtract: the maximum likelihood point, a fairdraw from the posterior, and the median glitch realization, computed point-wise in the frequency domain.

We test this with the parameterized slow scattering model, using a configuration consisting of three scattering arches in Gaussian noise. In Figure 8.1, we plot the results for each configuration at six values of injected SNR. This configuration has

$N_p = 10$, and so the expectation value for the maximum likelihood residual SNR is $\delta\rho \approx 3.16$. We see that this prediction holds for all injected SNR cases except the lowest ($\rho_{\text{inj}} = 5$). We also find that the fair draw produces a somewhat higher SNR, while the median realization produces a residual SNR which is less than or equal to that of the maximum likelihood residual in each case. Importantly we also find that, again with the exception of $\rho_{\text{inj}} = 5$, residual SNR is largely fixed regardless of the injected SNR, as predicted.

The case of $\rho_{\text{inj}} = 5$ is special, showing residual SNRs which are frequently *higher than the SNR of the glitch which was injected*. We interpret this to be a failure of the inference pipeline to identify the glitch at all when SNRs are so low. This is concerning, since it implies that for sufficiently low SNR glitches attempting subtraction may be doing more harm than good. As we shall see in Section 8.3, low SNR glitches can still impact parameter estimation results, making this result especially concerning.

8.2 Effects of time and phase

In merger dominated GW signals, individual cycles are critical to determining the properties of the system [4]. Accordingly, we would expect the details of interaction between a glitch and these cycles to substantially change the impacts of that glitch on PE [5, 6]. To test this, we simulate data containing a CBC signal as well as a single wavelet ($N_p = 5$), and vary the parameters of the wavelet in turn. Equation 8.1 lets us predict this error for the maximum likelihood point under a linear signal approximation:

$$\delta\theta^\alpha \approx (\Gamma^{-1})^{\alpha\beta} \langle \mathbf{n} + \mathbf{g} | \partial_\beta \mathbf{h} \rangle, \quad (8.10)$$

representing both the glitch \mathbf{g} and the Gaussian noise \mathbf{n} . In this section we will perform analyses with zero-noise, such that $\mathbf{n} = 0$ in Equation 8.10.

Notably, this formula predicts that the error depends on the overlap of the glitch with the *gradient* of the waveform model with respect to its parameters θ , rather than its realization. Intuitively, this is because glitches will impact the parameter estimation when, *combined with some alternate CBC configuration*, they are able to effectively mimic the true signal, which does not necessarily mean they resemble the true signal. For a glitch to bias measurements of spins, it only needs to resemble the *difference* between two configurations with dramatically different spins, a difference which is often quite subtle for more massive BBH systems.

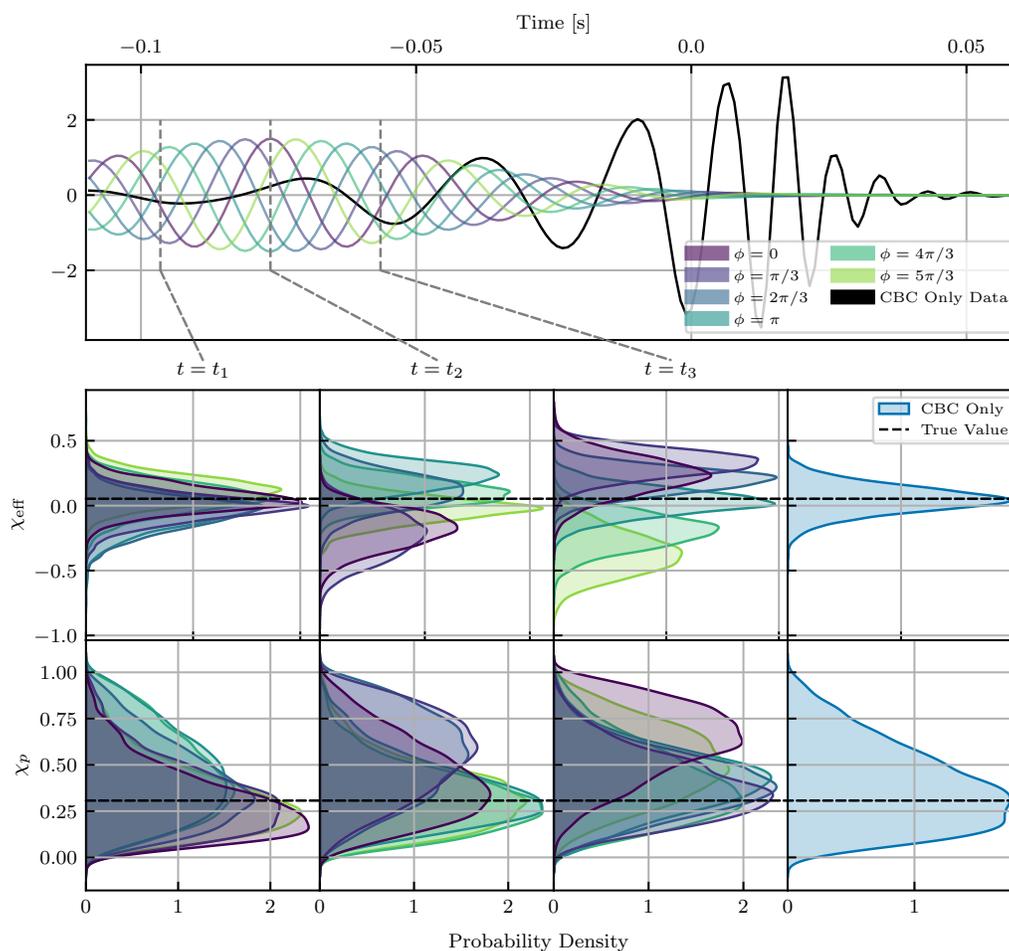


Figure 8.2: Posteriors for a GW191109 like signal with moderate χ_{eff} and a wavelet injected at varying times and frequencies. To aid in visualization, we include the wavelet at each phase, centered on t_2 .

The effect of the phase-time overlap

To explore the impact of phase and time on inferred parameters, we take reference wavelets from analyses of GW191109 and GW200129 and vary their phase and central time, while leaving their amplitude, quality factor, and central frequency unchanged. In Figures 8.2, 8.3, 8.4, and 8.5 we plot the inferred χ_{eff} and χ_p as a function of the wavelet phase, along with a reference case where no wavelet is added. We show the data into which these are injected, along with a visualization of how the wavelet morphology varies with respect to phase when it is injected at $t = t_2$ for the respective cases.

Figure 8.2 shows the case of a GW191109-like signal with moderate χ_{eff} , along with a wavelet adapted to the morphology of GW191109 (wavelet SNR $\rho_g = 7.51$).

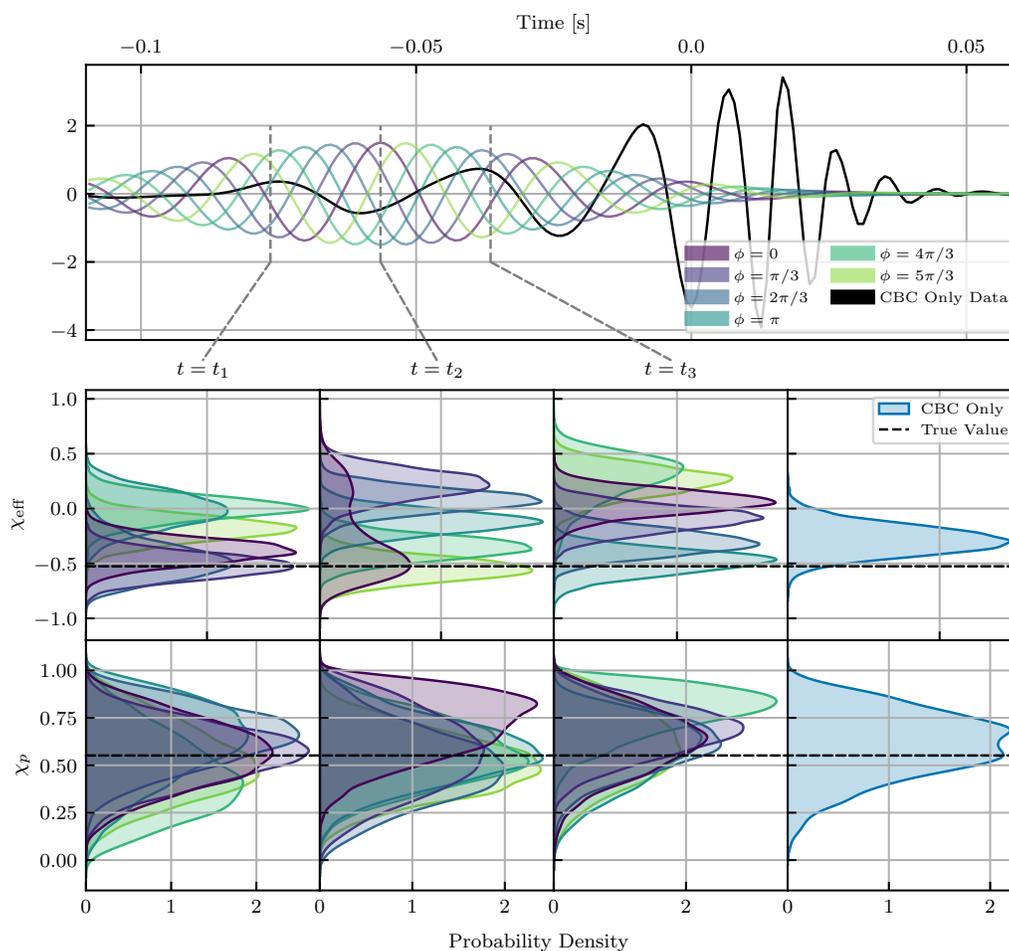


Figure 8.3: Posteriors for a GW191109 like signal with extreme χ_{eff} and with a wavelet injected at varying times and frequencies. To aid in visualization, we include the wavelet at each phase, centered on t_2 .

Here a very sharp dependence on both time and phase is visible. At the earliest time ($t = t_1$) posteriors only vary slightly from the reference posterior, and there is only marginal dependence on the phase. At $t = t_2$ and $t = t_3$ especially, however, there is a significant variation with respect to the phase of the wavelet, and the posteriors are substantially different than the reference posterior. One notable case is that of $\phi = 5\pi/3$ for $t = t_3$, where the posterior on χ_{eff} is almost entirely negative, despite the injected value being near zero. For this CBC configuration χ_p has less variation, though there is still one case ($\phi = 0$ at $t = t_3$) where measurement of χ_p is strongly peaked away from zero.

Figure 8.3 shows the case of a GW191109-like signal with an extreme χ_{eff} injected along with the same wavelet as before, and it tells a similar story. The reference

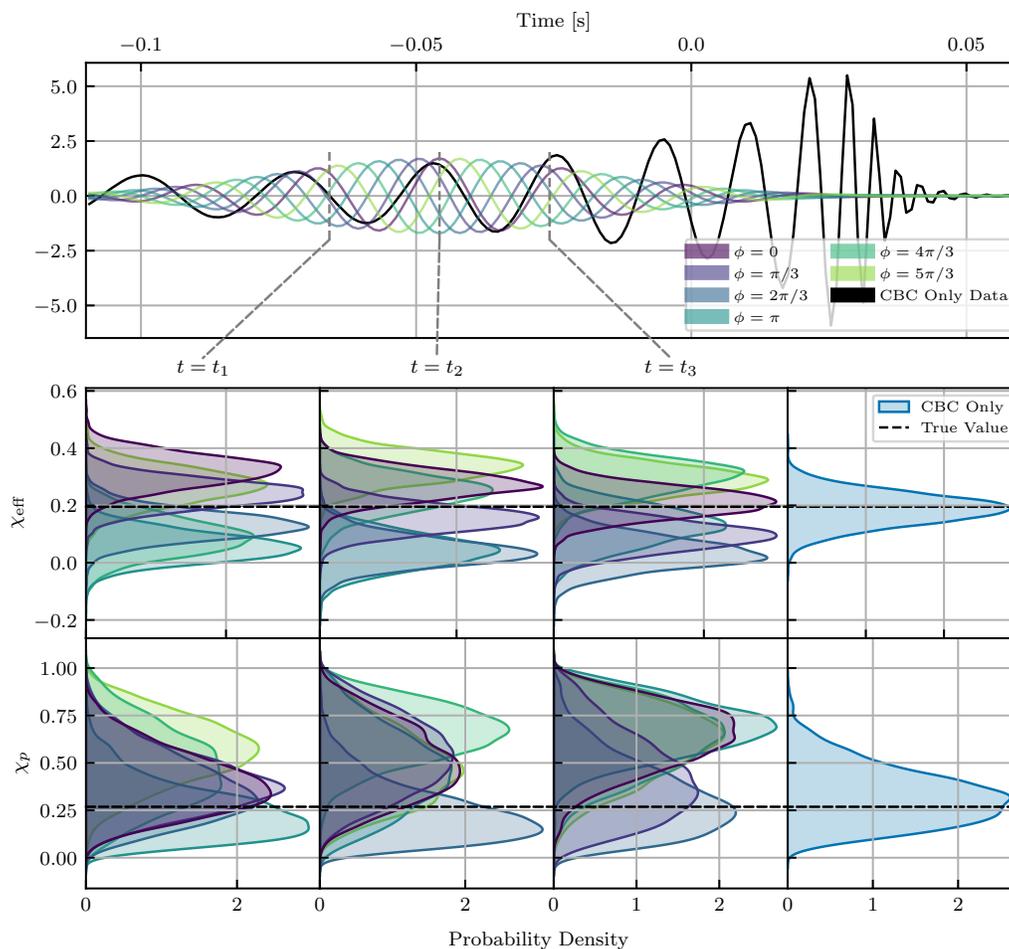


Figure 8.4: Posteriors for a GW200129 like signal with moderate χ_p and with a wavelet injected at varying times and frequencies. To aid in visualization, we include the wavelet at each phase, centered on t_2 .

distribution for χ_{eff} is peaked closer to 0 than the true injected value, due to the effects of prior. Comparatively, depending on its time and phase the wavelet may do anything from pushing the distribution of χ_{eff} to *more* negative values, to driving it to a mostly positive distribution. For these values of t_2 and t_3 it is interesting to note that the distributions show similar spreads with respect to ϕ , but at different values. For example, at $t = t_2$ the most positive distribution of χ_{eff} occurs at $\phi = \pi/3$, while for $t = t_3$ it occurs at $\phi = 4\pi/3$. This behavior reflects the intuition gained from the wavelets plotted in Figure 8.3, since different time-phase combinations can produce similar fluctuations at given cycles, which drive the distribution biases.

Figure 8.4 shows the case of a GW200129-like signal with moderate χ_p , with a wavelet adapted to the GW200129 morphology (wavelet SNR $\rho_g = 7.96$). Here

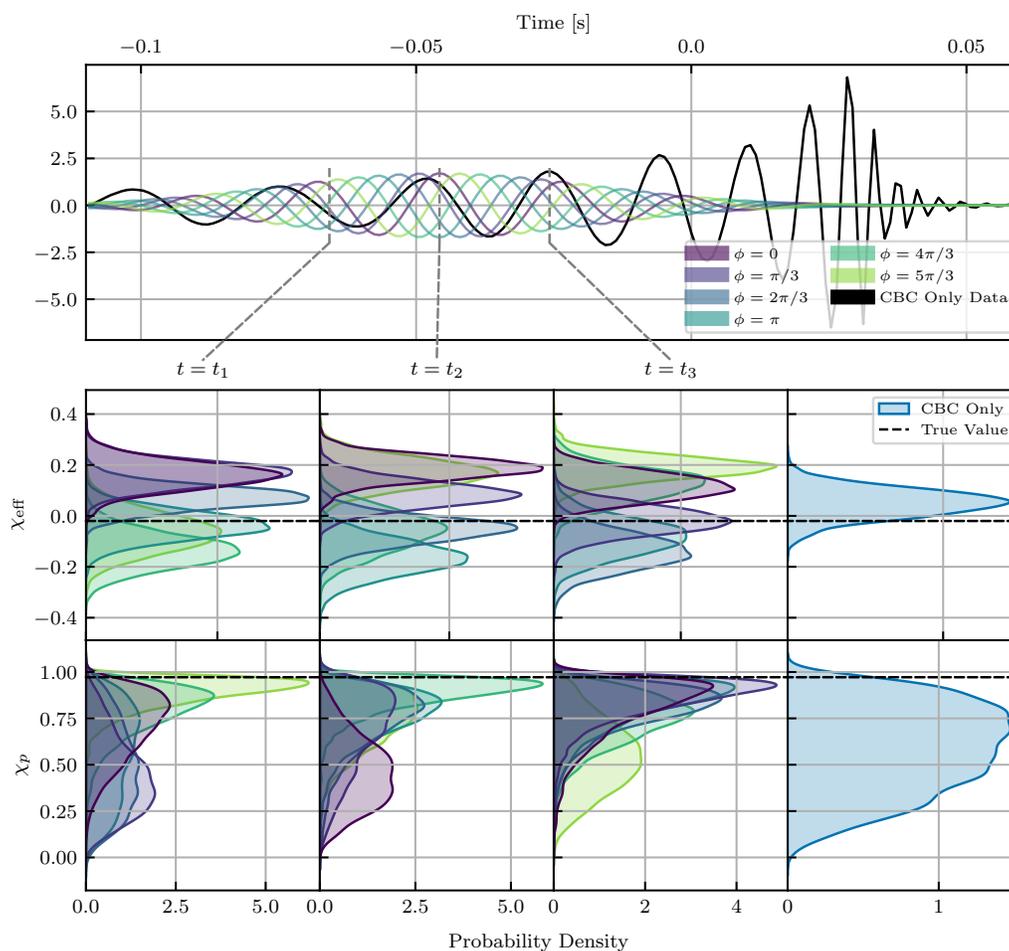


Figure 8.5: Posteriors for a GW200129 like signal with extreme χ_p and a wavelet injected at varying times and frequencies. To aid in visualization, we include the wavelet at each phase, centered on t_2 .

χ_{eff} shows variations in time and phase much as before, but the behavior of χ_p is significantly more interesting. In particular, for this moderate value of χ_p there exist some cases—for example, $\phi = 4\pi/3$ at $t = t_2$ —where the distribution fully excludes 0, and almost fully excludes the injected value. Here as before, time and phase combine to produce substantially different behaviors, though the phase values which result in given behaviors are more static over time, where for example $\phi = 4\pi/3$ produces significantly higher χ_p distributions at both $t = t_2$ and $t = t_3$.

Figure 8.5 shows the case of a GW200129-like signal with extreme χ_p , and the same reference wavelet as used in Figure 8.4. Here as in Figure 8.3 the injected value is extreme and hence disfavored by priors which have vanishing support at $\chi_p = 1$ in accordance with physical constraints, resulting in a reference distribution

which does not support such extreme measurements. In some cases—for example $\phi = 4\pi/3$ with $t = t_2$ —the distribution is pushed to actually include the injected value at high confidence, while for other values—such as $\phi = 0$ at $t = t_2$ —it is pushed further towards the prior than the reference distribution. Once again, the effects of time and phase produce widely varying astrophysical results.

8.3 Impact of glitches with varying SNR

Next we consider how glitches of varying SNRs can impact the astrophysical inference of events. Glitch rates generally increase at lower SNRs [7], though below a certain threshold they may not stand out from the noise sufficiently to be identified. If glitches at such low SNRs are able to impact PE posteriors, it raises the possibility that PE could be biased by glitches *without us even knowing that a glitch is present in the data*.

In order to investigate this question, we proceed much as in Section 8.2. For each case we choose a wavelet which had a qualitatively significant impact on the results of the PE. We then inject this wavelet, scaling the amplitude such that the SNR will be a fixed value between one and ten. We then plot the posteriors on χ_{eff} and χ_p , along with the Jensen-Shannon (JS) divergences [8]—a measure of the similarity between two distributions, being 0 if they are the same and 1 if they are entirely distinct—between them and the reference distribution. The LVK uses a threshold of $JS = 0.007$ to identify when two distributions are “significantly” different [9], and $JS = 0.002$ is approximately the amount of variation expected due to stochastic sampling [10].

Figure 8.6 shows the same moderate χ_{eff} GW191109-like used in Figure 8.2, injected along with the $\phi = \pi/3, t = t_3$ wavelet at varying SNRs. As would be expected from this case in Figure 8.2, χ_p does not change at any SNR, but χ_{eff} shifts dramatically towards positive values. The JS divergence of the χ_{eff} distributions against their reference also reflects the rapidity of this shift.

Figure 8.7 shows the same extreme χ_{eff} GW191109-like used in Figure 8.3, injected along with the $\phi = \pi/3, t = t_2$ wavelet at varying SNRs. In this case there are dramatic shifts in both χ_{eff} and χ_p . Although the influence of the prior means that the reference distribution already disfavors the true value of χ_{eff} , a wavelet of SNR 3 is enough for the posterior to completely reject the true value. Meanwhile, χ_p has the interesting property that it only begins to shift when the wavelet has an SNR of six, but after that shifts quite quickly, unlike the steady shift observed in χ_{eff} .

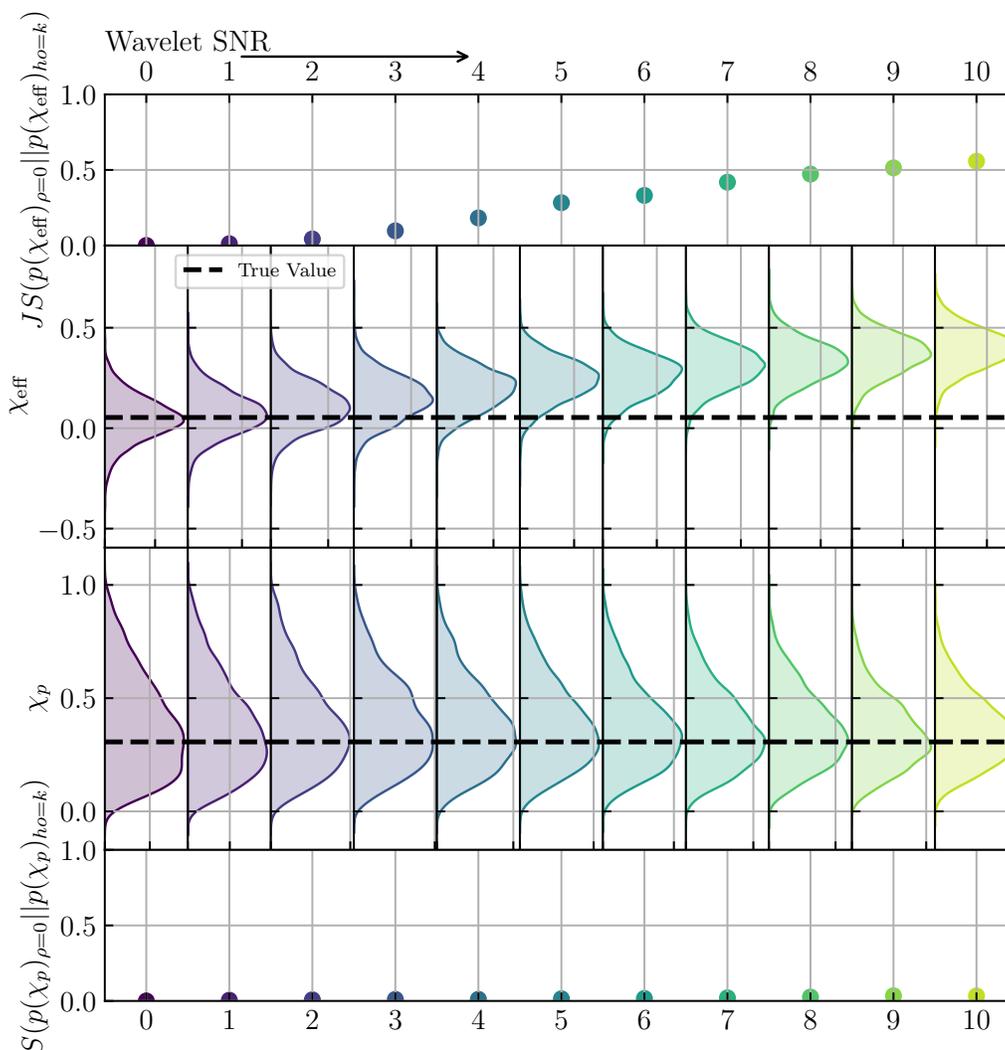


Figure 8.6: Posteriors and JS divergences for runs which injected a GW191109-like configuration with moderate χ_{eff} alongside a glitch scaled to varying SNR.

Figure 8.8 shows the moderate χ_p GW200129-like distribution used in Figure 8.4, injected along with the $\phi = \pi, t = t_3$ wavelet at varying SNRs. This case shows the most dramatic χ_p shifts, with the highest SNR cases strongly excluding the true value in preference of extreme χ_p values. It also shows some moderate shifts in χ_{eff} , but here these follow the pattern χ_p showed above, where they do not begin to shift until some threshold wavelet SNR is passed—in this case around an SNR of 7—at which point they shift fairly rapidly.

Finally, Figure 8.9 shows the extreme χ_p GW200129-like distribution used in Figure 8.5, injected alongside the $\phi = 0, t = t_2$ wavelet at varying SNRs. In this case χ_{eff} shifts significantly but steadily, while χ_p behaves in a unique manner. Here,

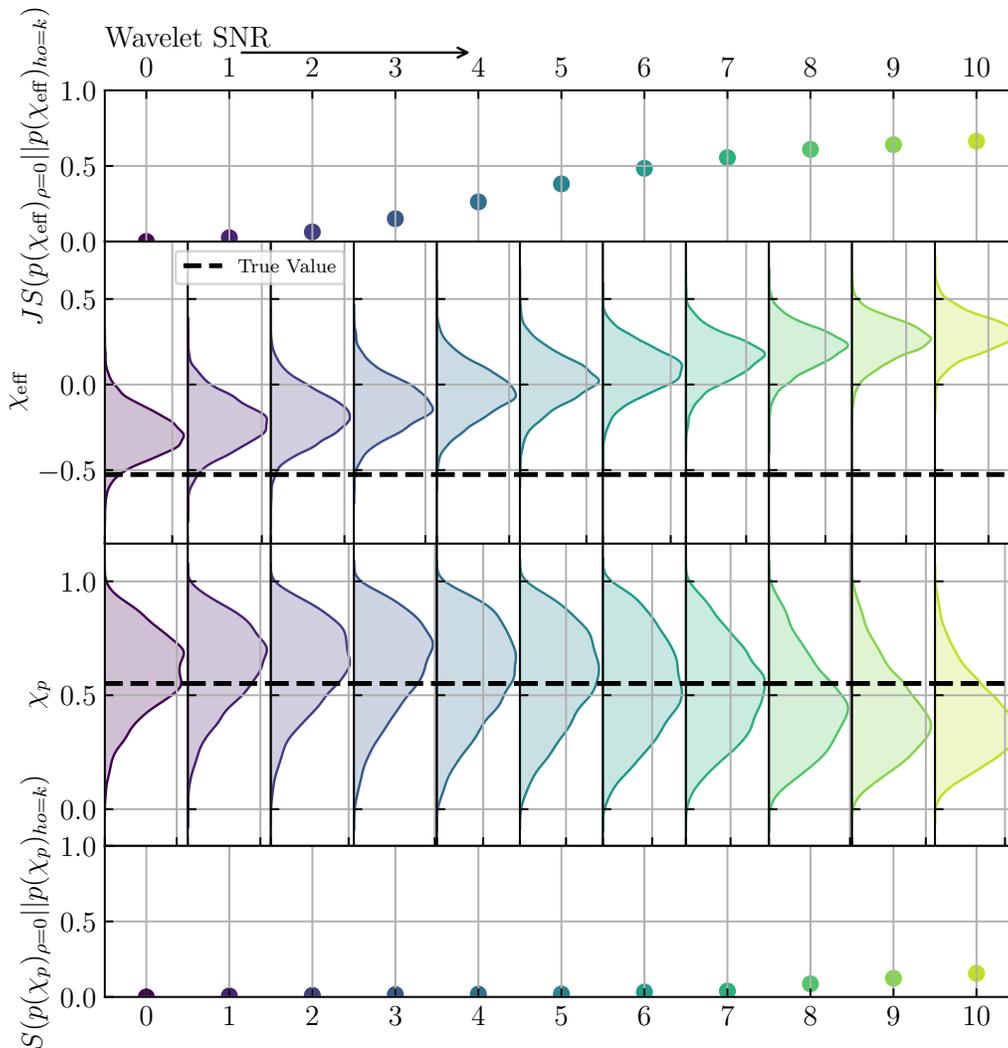


Figure 8.7: Posteriors and JS divergences for runs which injected a GW191109-like configuration with extreme χ_{eff} alongside a glitch scaled to varying SNR.

the posterior on χ_p shifts significantly towards less extreme values—away from the true value—before its progression reverses, such that the highest SNR case largely resembles the reference distribution. This is an interesting case to interpret, since it implies that the added wavelet power was mimicking a precessing signal *up to a point* before it became too significant and the inference “snapped back” to the original interpretation in χ_p , while maintaining the altered χ_{eff} posterior.

8.4 The frequency-bin χ^2 test in parameter estimation

The frequency-bin χ^2 test is a statistical test of whether some data are well described as the combination of a given template and Gaussian noise. Originally developed by Allen [11] and deployed in the Pycbc search pipeline [12], it is used to downrank

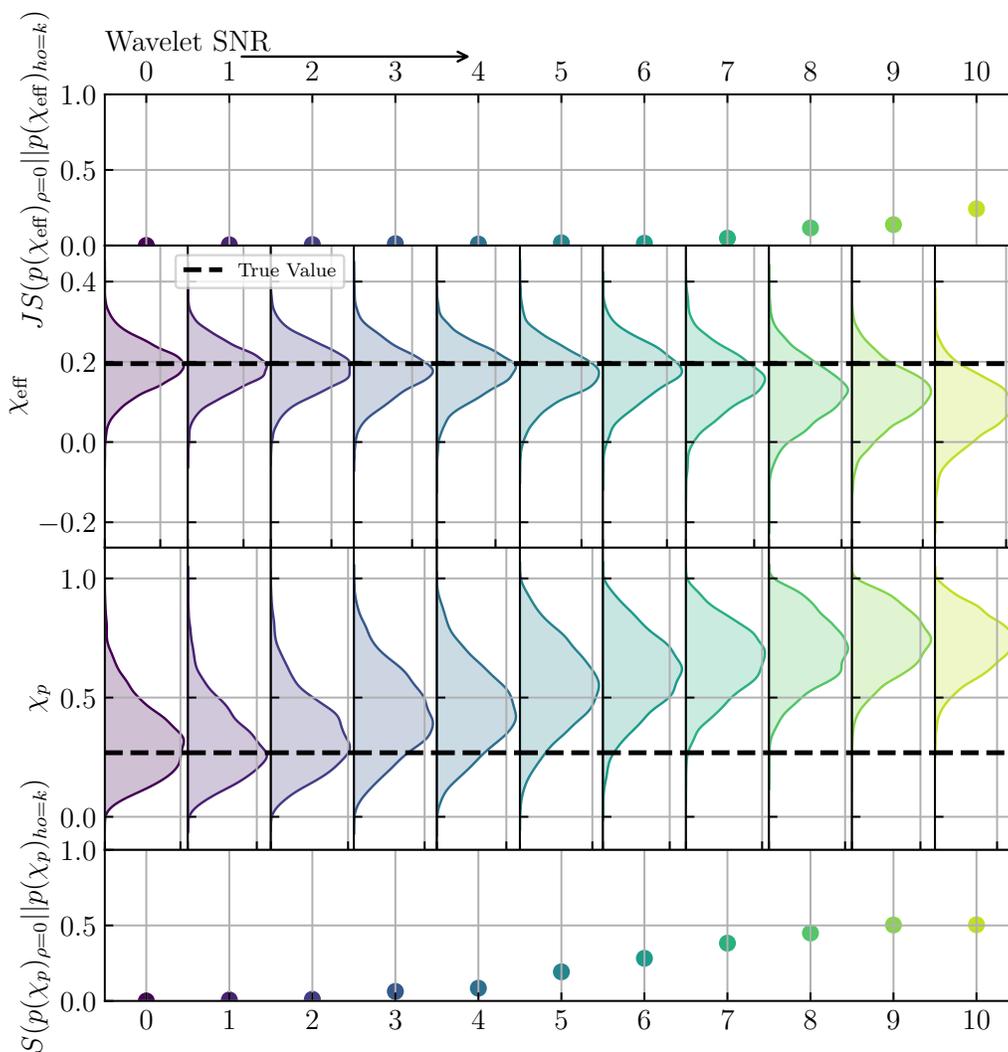


Figure 8.8: Posteriors and JS divergences for runs which injected a GW200129-like configuration with moderate χ_p alongside a glitch scaled to varying SNR.

triggers which have significant SNR but which do not match the template very well, a combination which could indicate that the trigger is a glitch. It was applied to PE for the first time in Udall et al [5]—an analysis which is reproduced in Section 7.10 of this thesis accordingly—in an attempt to assess whether different methods of glitch mitigation produce results which are less biased.

Briefly, the test divides the frequency range of the analysis into p frequency bins, such that for the given template $\tilde{h}(f; \theta)$ the optimal SNR is equally distributed across each bin. Figure 8.10 illustrates this procedure. Mathematically, considering the

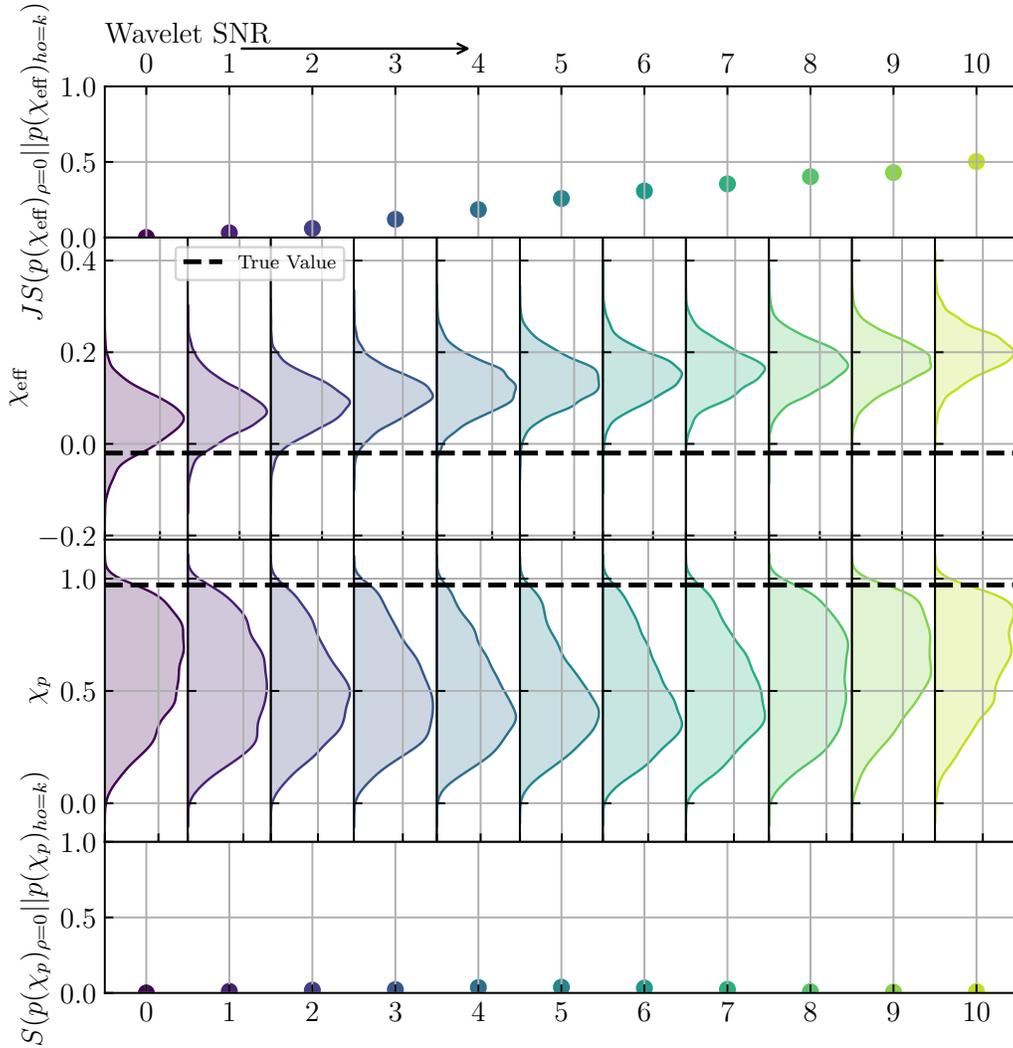


Figure 8.9: Posteriors and JS divergences for runs which injected a GW200129-like configuration with extreme χ_p alongside a glitch scaled to varying SNR.

matched filter SNR in each frequency bin

$$\rho_{mf,k} = 4 \int_{f_{k-1}}^{f_k} \frac{\tilde{d}(f)\tilde{h}^*(f;\theta)}{S_n(f)} df, \quad (8.11)$$

one may also compute the deviation from the expected per-bin matched filter SNR

$$\Delta\rho_{mf,k} = \rho_{mf,k} - \frac{\rho_{mf}}{p}, \quad (8.12)$$

where ρ_{mf} is the matched filter SNR computed over the entire frequency range. Allen shows [11] that this quantity then satisfies the property

$$\langle |\Delta\rho_{mf,k}|^2 \rangle = \frac{2}{p} \left(1 - \frac{1}{p}\right), \quad (8.13)$$

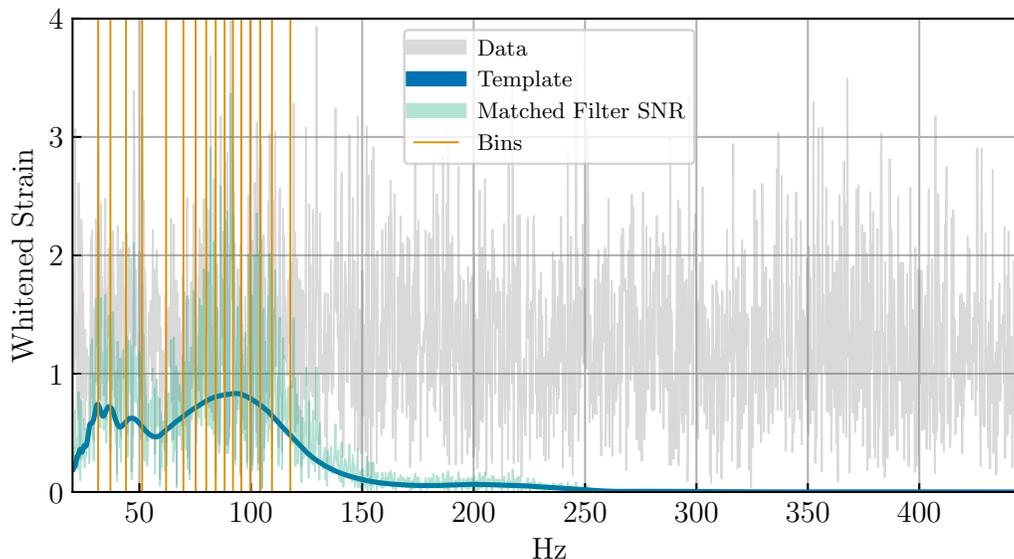


Figure 8.10: An example of bins constructed to evenly divide optimal SNR over the frequency band. This plot of whitened frequency domain strain is peculiar in its use of linear scale, but in these dimensions the area under the blue curve should be constant for each of the bins. The χ^2 test then measures whether the inner product between the data and the template, shown in green, is also distributed evenly across those bins, up to Gaussian noise fluctuations.

which allows the definition of the statistic

$$\langle \chi^2 \rangle = p \sum_{k=1}^p |\Delta \rho_{m,f,k}|^2. \quad (8.14)$$

Allen shows that this is distributed as a χ^2 distribution with $2p - 2$ degrees of freedom. Importantly, because this is a statistical statement about the true template in the presence of Gaussian noise, this derivation applies to the results of PE as well, although the distribution over the posterior will not generally speaking be a χ^2 distribution itself.

To apply this test to PE, we compute it for many configurations in the posterior of an analysis, and plot these together. This is then plotted in the form of the reduced χ^2 statistic

$$\bar{\chi}^2 = \frac{\chi^2}{2p - 2}, \quad (8.15)$$

such that if the values do not peak near one, there is reason to suspect that the PE may be biased by a glitch. Exactly *how* much suspicion one should possess is tricky, since it depends on how much this statistic may be expected to vary under normal

Gaussian noise fluctuations. Determining such a background is tricky. While it is possible to compute them by brute-force through hundreds of PE analyses, such as was done in Udall et al [5], this test must be computationally efficient if it is to be widely deployed. That being said, it is still interesting to ask whether the distribution will shift when there *is* a glitch, all other things being equal.

To test this, we apply a test similar to that used in Section 8.3 above, specifically with the configuration used in Figure 8.7. One important difference, however, is that for this test to be meaningful this analysis must now be performed in the presence of Gaussian noise. The results of this are plotted in Figures 8.11 and 8.12. The distribution of the χ^2 statistic in Livingston—where the wavelet was injected—increases rapidly even for low wavelet SNRs, while the statistic in Hanford is basically static. This mirrors the shift in χ_{eff} (Figure 8.11), and especially in q (Figure 8.12), since the shift in this statistic reflects cumulative error over the parameter space as opposed to the shift due to any given parameter error. This is a very promising indication that the statistic is fairly sensitive to errors from the presence of a glitch, although more work is still required to quantify its statistical strength.

References

- ¹T. Robson and N. Cornish, “Impact of galactic foreground characterization on a global analysis for the LISA gravitational wave observatory”, *Class. Quant. Grav.* **34**, 244002 (2017) DOI:10.1088/1361-6382/aa9601.
- ²C. Cutler and J. Harms, “BBO and the neutron-star-binary subtraction problem”, *Phys. Rev. D* **73**, 042001 (2006) DOI:10.1103/PhysRevD.73.042001.
- ³C. Cutler and E. E. Flanagan, “Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral wave form?”, *Phys. Rev. D* **49**, 2658–2697 (1994) DOI:10.1103/PhysRevD.49.2658.
- ⁴S. J. Miller, M. Isi, K. Chatziioannou, V. Varma, and I. Mandel, “GW190521: Tracing imprints of spin-precession on the most massive black hole binary”, *Phys. Rev. D* **109**, 024024 (2024) DOI:10.1103/PhysRevD.109.024024.
- ⁵R. Udall, S. Hourihane, S. Miller, D. Davis, K. Chatziioannou, M. Isi, and H. Deshong, “Antialigned spin of GW191109: Glitch mitigation and its implications”, *Phys. Rev. D* **111**, 024046 (2025) DOI:10.1103/PhysRevD.111.024046.
- ⁶E. Payne, S. Hourihane, J. Golomb, R. Udall, D. Davis, and K. Chatziioannou, “Curious case of GW200129: Interplay between spin-precession inference and data-quality issues”, *Phys. Rev. D* **106**, 104017 (2022) DOI:10.1103/PhysRevD.106.104017.

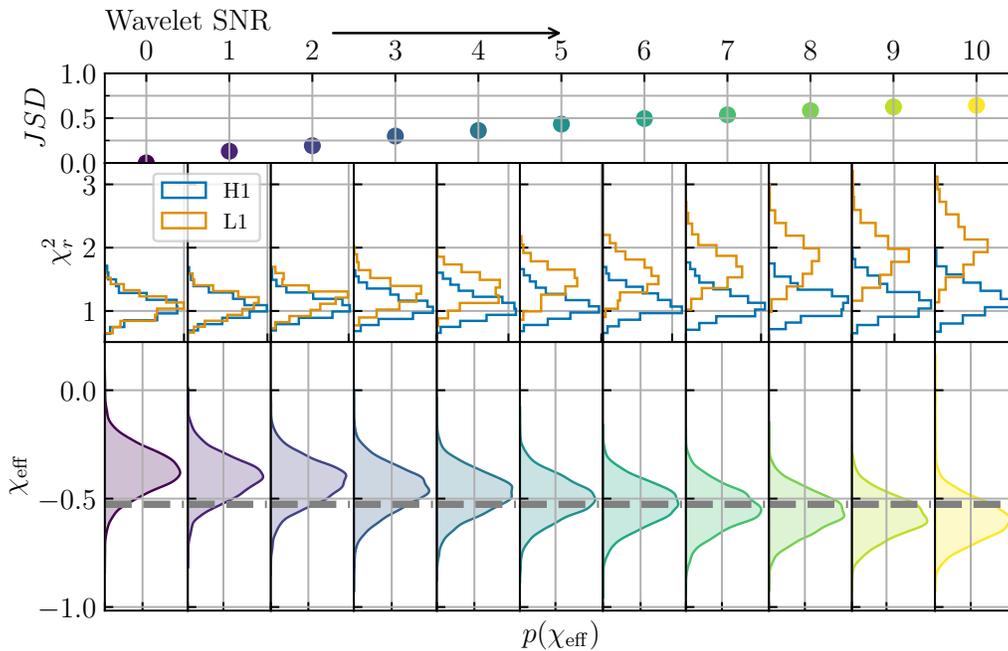


Figure 8.11: The posteriors on χ_{eff} for a simulated CBC system overlapping with a simulated wavelet in the Livingston detector, along with the JS divergence between each posterior and the reference distribution, and the distribution of the χ^2 statistic in each detector.

⁷S. Soni et al. (LIGO), “LIGO Detector Characterization in the first half of the fourth Observing run”, *Class. Quant. Grav.* **42**, 085016 (2025) DOI:10.1088/1361-6382/adc4b6.

⁸J. Lin, “Divergence measures based on the Shannon entropy”, *IEEE Trans. Info. Theor.* **37**, 145–151 (1991) DOI:10.1109/18.61115.

⁹R. Abbott et al. (LIGO Scientific, Virgo), “GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run”, *Phys. Rev. X* **11**, 021053 (2021) DOI:10.1103/PhysRevX.11.021053.

¹⁰I. 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**, 3295–3319 (2020) DOI:10.1093/mnras/staa2850.

¹¹B. Allen, “ χ^2 time-frequency discriminator for gravitational wave detection”, *Phys. Rev. D* **71**, 062001 (2005) DOI:10.1103/PhysRevD.71.062001.

¹²S. A. Usman et al., “The PyCBC search for gravitational waves from compact binary coalescence”, *Class. Quant. Grav.* **33**, 215004 (2016) DOI:10.1088/0264-9381/33/21/215004.

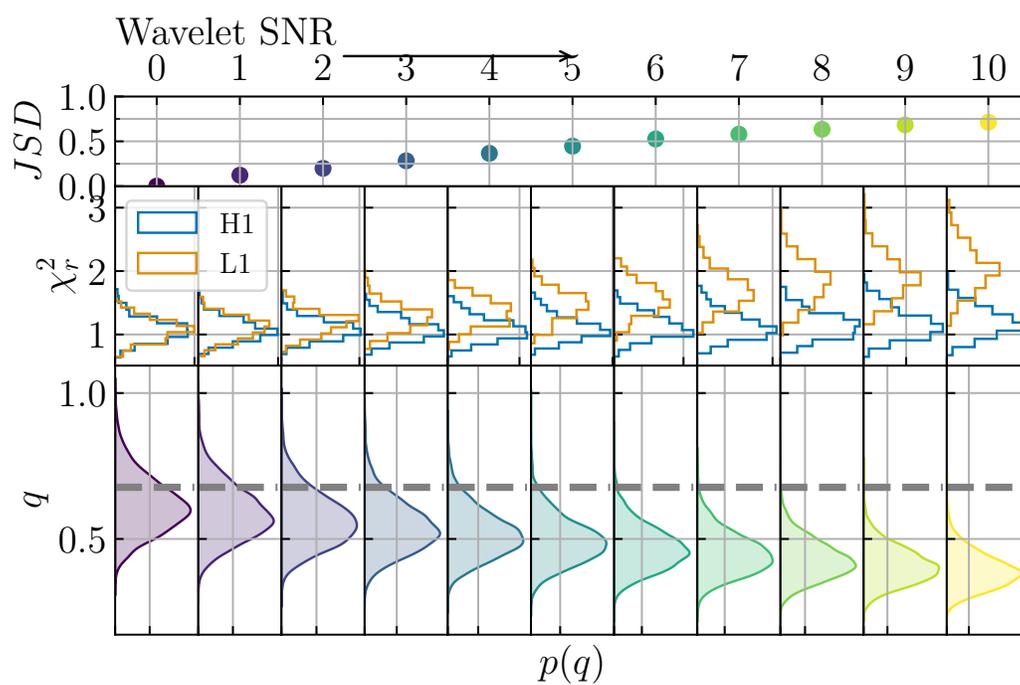


Figure 8.12: The posteriors on q for a simulated CBC system overlapping with a simulated wavelet in the Livingston detector, along with the JS divergence between each posterior and the reference distribution, and the distribution of the χ^2 statistic in each detector.

Part III

Gravitational Wave Transient Catalogs and Conclusion

Chapter 9

LVK DATA MANAGEMENT

The production of a GW transient catalog is exceedingly complex, involving hundreds of scientists performing dozens of individual analyses. The flow of data from analysis to analysis, and ultimately into a data product, is correspondingly complex, and errors at any stage may become magnified and result in significant delays of the final product. I will describe in general terms the most important phases of analysis, as well as the issues which occurred in the third observing run. In Chapter 10, I will describe the work I have done to solve these issues and allow for dramatic improvements in efficiency in the fourth observing run and in the upcoming gravitational wave transient catalogs.

Figure 9.1 displays graphically the flow of data for the O4a catalog GWTC-4—GWTC-5 and GWTC-6 will likely also use this scheme with minor modifications—between the elements of analysis which will be discussed. Strain data are filtered by search pipelines, initially in low-latency searches then during archival “offline” searches. These results are logged in GRACEDB, where they trigger subsequent followup. Initial PE runs are performed, and these results are logged into CBCFLOW, which will be discussed in Section 10.1.

At higher latency, detector characterization experts assess if any glitch mitigation efforts are required for observed events, and perform glitch subtraction accordingly. The initial PE results and recommendations for glitch mitigation are then used to configure production PE runs using ASIMOV, and launch them accordingly. The results of these production PE runs are then fed back into CBCFLOW. Finally GWTC_PIPE takes the production PE results and offline search results and collates them into a data product for internal and public release.

9.1 Transient searches

The first step to analyzing a GW event is to identify it, and for this purpose there exist a variety of transient searches. Broadly these may be grouped into two types: matched filter searches and unmodeled searches. For GW transients in Gaussian noise which are well described by template CBC waveforms in a pre-computed template bank, matched filtering is the optimal search method [2, 3]. However,

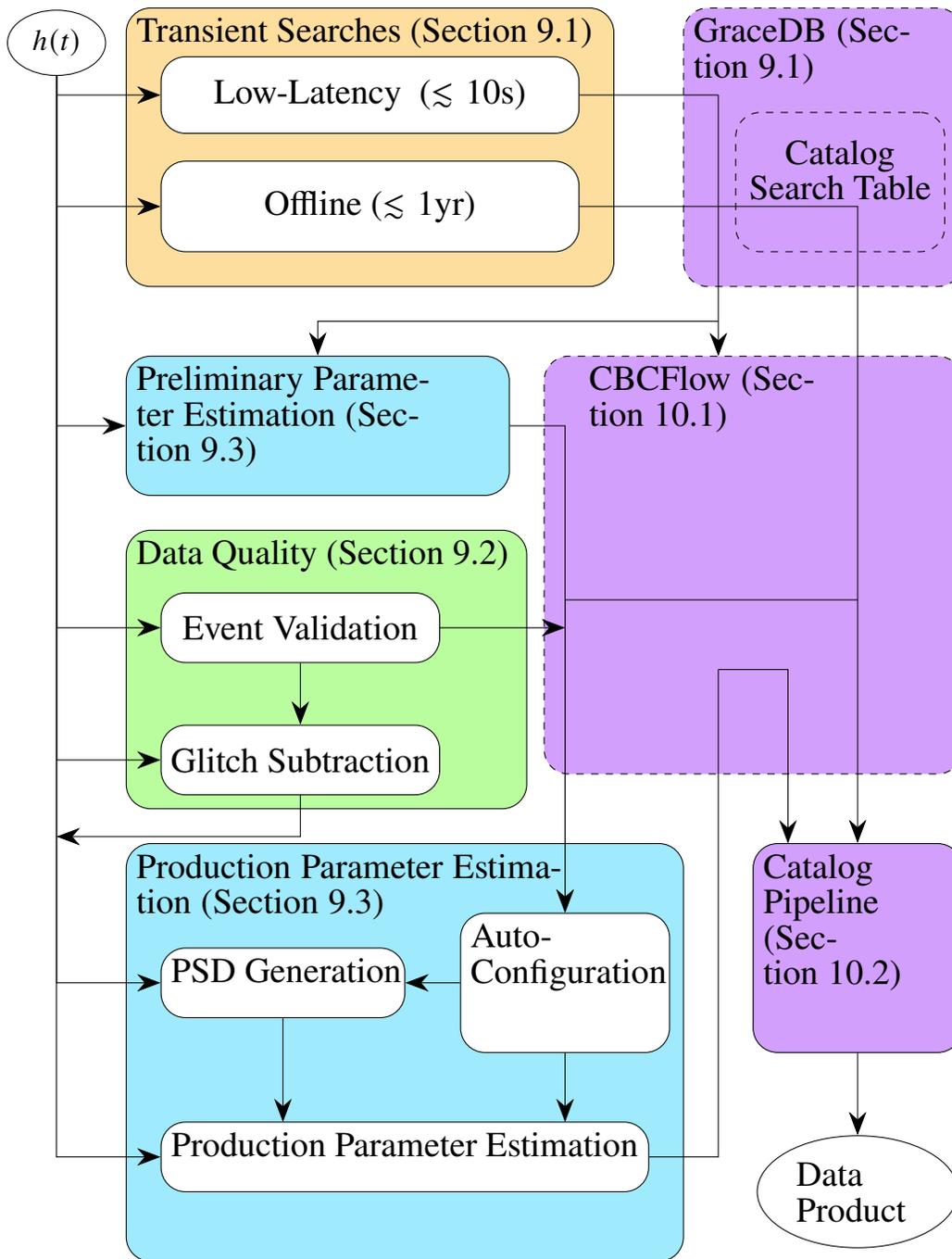


Figure 9.1: A high level rendering of the flow of data involved in the production of the GWTC-4 catalog. Reproduced from the upcoming GWTC-4 Methods paper [1].

in cases where the GW is not well described by any template waveform—due to incompleteness of the template bank, inaccuracy of the underlying waveform, or potentially even if the GW is not sourced by a CBC at all—unmodeled searches may be more successful, and so these are also used in LVK data analysis [4]. In this section I will briefly discuss the two varieties of search pipeline from the perspective of data management, as well as the event reporting infrastructure in which they operate.

All pipelines produce certain data in common. Firstly, they all report false alarm rates (FAR), which are treated as the primary detection statistic by the catalog, as well as event times and SNRs. In the GWTC-4, events with a FAR less than two per day are included in the search results, while PE is only run on events for which the FAR is less than one per year [1]. Secondly, they report astrophysical source classifications, most notably the p_{astro} statistic defined as [1, 5]:

$$p_{astro}(x) = \frac{R_{astro}f(x)}{R_{astro}f(x) + R_{noise}b(x)}, \quad (9.1)$$

where R_{astro} and R_{noise} are the rates of astrophysical and noise distribution events respectively, while $f(x)$ and $b(x)$ are the probability distribution of signal and noise events respectively at the ranking statistic x , which is typically FAR. Finally, the searches report skymaps, usually computed by BAYESTAR [6], which facilitate followup electromagnetic observations.

Matched filter pipelines

Over the course of the fourth observing run the LVK has operated four distinct matched filter pipelines—PYCBC [7], GSTLAL [8], MBTA [9], and SPIIR [10]—with the first three ultimately appearing in the O4a catalog. Generally speaking these compute an SNR timeseries [1, 11]

$$\rho(t) = 4 \left| \int_{f_{low}}^{f_{high}} \frac{h(f)T^*(f)}{S_n(f)} e^{2\pi i f t} df \right| \quad (9.2)$$

and search for peaks in this timeseries. These peaks are then ranked, usually with additional filtering to downweight potential glitches, and from these rankings FARs and other downstream data are computed.

While there are many technical differences, from the perspective of data management they are quite consistent. All provide a standardized set of data stored in files which share formatting standards, and all upload the same data products, including skymaps and source classifications. Importantly for the purposes of Chapter 10,

because these pipelines use templated searches, they provide preliminary estimates of the CBC source properties.

Unmodeled search pipelines

Unmodeled search pipelines, most prominently `cwb`, do not rely on templated match filtering to search for GW signals. Rather, they use more flexible parameterizations—prominently wavelets such as those used by `BAYESWAVE`—in order to search for signals which are coherent between detectors. `cwb` uses wavelet transforms to produce a time-frequency representation, then triggers on coherent peaks in this representation [12, 13]. `cwb` also includes a BBH focused sub-pipeline, which was used in the fourth observing run, which enables the computation of chirp mass [13].

From a data management perspective, the differences in `cwb` with respect to matched filter pipelines can sometimes make it difficult to manage. Most prominently, data is represented in a different file type than with matched filter searches, making parsing more of a challenge¹. It can also be difficult to identify which quantities best correspond to those computed in matched filter searches, such as which SNR is most closely analogous. Finally, though chirp mass is computed, individual mass components are not, which may leave gaps in datasets which represent this information such as `CBCFLOW`.

GraceDB and GWCelery

To manage the results of each search pipeline, an infrastructure has been developed for storing, aggregating, and annotating results across events and pipelines [14][15]. This allows for rapid release of results which are relevant to EM astronomers, including skymaps which are updated as more in-depth analyses complete [16]. When triggers are identified by search pipelines, `GWCELERY` [15] collates them into superevents and annotated with relevant information tags. These are uploaded to `GRACEDB` [14], which serves as a database for all search results. In the fourth observing run `GRACEDB` was further augmented with a catalog table, which version controlled offline results as they were uploaded, and sourced them into a single database which could be fetched for a verified list of catalog events.

¹I would approximate that 10-15% of all `CBCFLOW` bugs were in some way related to the parsing of `cwb` data.

9.2 Detector characterization

Detector characterization played a number of roles in the fourth observing run catalog. After initial detections, experts would use automatically generated data quality reports to interpret the possibility that observed triggers were due to terrestrial noise [17]. After events were verified and initial PE was performed, event validators would use these same reports to identify transient noise impacting identified GWs. When identified, transient noise was subtracted if possible using `BAYESWAVE` as discussed in Chapter 5, and if this proved impossible then recommendations were made for time and/or frequency segments to exclude when performing PE. These recommendations were then fed into `CBCFLOW`, which propagated them onwards to production PE analyses.

9.3 Parameter estimation

In Chapter 3 I extensively discussed the details of how PE is performed for CBCs. From a catalog data perspective, there are two important phases of PE: the initial exploratory analyses, and the production PE analyses. Online PE is launched automatically by `GWCELERY`, and any followup exploratory analysis is performed by a rota of analysts. Production PE, by contrast, is conducted by `ASIMOV`, and incorporates recommendations from detector characterization’s event validation experts, as well as the results of exploratory runs. Both of these feed information into `CBCFLOW`, which stores both information about the analysis process—such as the analysts involved and the status of the analysis review—as well as metadata about the analysis itself, like the location of result and configuration files and the waveform approximant used. Finally, results are read out of `CBCFLOW` using the catalog pipeline `GWTC_PIPE` to construct the final data products, as will be further discussed in Section 10.2.

References

- ¹A. G. Abac et al. (KAGRA, VIRGO, LIGO Scientific), “GWTC-4: Methods for identifying and characterizing gravitational-wave transients”, *Astrophys. J.* (In Prep).
- ²E. E. Flanagan and S. A. Hughes, “Measuring gravitational waves from binary black hole coalescences: 1. Signal-to-noise for inspiral, merger, and ringdown”, *Phys. Rev. D* **57**, 4535–4565 (1998) DOI:10.1103/PhysRevD.57.4535.
- ³C. Cutler and E. E. Flanagan, “Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral waveform?”, *Phys. Rev. D* **49**, 2658–2697 (1994) DOI:10.1103/PhysRevD.49.2658.

- ⁴T. 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**, 083018 (2022) doi:10.1103/PhysRevD.105.083018.
- ⁵B. P. Abbott et al. (LIGO Scientific, Virgo), “The Rate of Binary Black Hole Mergers Inferred from Advanced LIGO Observations Surrounding GW150914”, *Astrophys. J. Lett.* **833**, L1 (2016) doi:10.3847/2041-8205/833/1/L1.
- ⁶L. P. Singer and L. R. Price, “Rapid Bayesian position reconstruction for gravitational-wave transients”, *Phys. Rev. D* **93**, 024013 (2016) doi:10.1103/PhysRevD.93.024013.
- ⁷S. A. Usman et al., “The PyCBC search for gravitational waves from compact binary coalescence”, *Class. Quant. Grav.* **33**, 215004 (2016) doi:10.1088/0264-9381/33/21/215004.
- ⁸C. Messick et al., “Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data”, *Physical Review D* **95**, 042001 (2017) doi:10.1103/PhysRevD.95.042001.
- ⁹T. Adams, D. Buskulic, V. Germain, et al., “Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era”, *Class. Quant. Grav.* **33**, 175012 (2016) doi:10.1088/0264-9381/33/17/175012.
- ¹⁰Q. Chu et al., “SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences”, *Phys. Rev. D* **105**, 024023 (2022) doi:10.1103/PhysRevD.105.024023.
- ¹¹B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, “FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries”, *Phys. Rev. D* **85**, 122006 (2012) doi:10.1103/PhysRevD.85.122006.
- ¹²S. Klimenko et al., “Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors”, *Phys. Rev. D* **93**, 042004 (2016) doi:10.1103/PhysRevD.93.042004.
- ¹³V. Tiwari, S. Klimenko, V. Neula, and G. Mitselmakher, “Reconstruction of chirp mass in searches for gravitational wave transients”, *Class. Quant. Grav.* **33**, 01LT01 (2016) doi:10.1088/0264-9381/33/1/01LT01.
- ¹⁴A. Pace et al., *Gracedb*, <https://gracedb.ligo.org/>, 2025.
- ¹⁵L. Singer et al., *Gwcelery*, <https://rtd.lgw.org/projects/gwcelery/en/latest/>, 2025.
- ¹⁶S. S. Chaudhary 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**, e2316474121 (2024) doi:10.1073/pnas.2316474121.

¹⁷S. Soni et al. (LIGO), “LIGO Detector Characterization in the first half of the fourth Observing run”, *Class. Quant. Grav.* **42**, 085016 (2025) doi:10.1088/1361-6382/adc4b6.

TECHNICAL INFRASTRUCTURE OF GRAVITATIONAL WAVE TRANSIENT CATALOGS: THE FOURTH OBSERVING RUN AND FUTURE OUTLOOK

In this chapter I will discuss the work I have done on various aspects of the technical infrastructure used by the LVK for the production of GW transient catalogs in the fourth observing run. This principally consists of two major software projects for which I am the lead developer, which I will explain in detail. The first, `CBCFLOW` [1], manages the storage and communication of metadata between analyses which form the stages of GW transient analysis, as described in Chapter 9. The second, `GWTC_PIPE` [2], parses the contents of `CBCFLOW` and of `GraceDB` to distill the stored metadata into the appropriate set of production results, which are then packaged to form publicly released data products [3].

10.1 CBCFLOW

Simplistically, `CBCFLOW` as a project consists of two parts. First, it is a database system using JSON files stored in a `git` library, including both the files themselves and the software infrastructure to read and write them. Second, `CBCFLOW` also contains utilities to automatically fetch metadata from analyses as they progress, both by providing hooks to other analysis software, and by running a periodic monitor process to provide updates. In this chapter I will describe the design and implementation of each of these components.

CBCFLOW metadata

`CBCFLOW` tracks metadata within JSON files on a per-event basis, meaning that each event stored has a JSON file following a standard specification. This specification is set using a JSON schema [4] which was collaboratively constructed by the LVK CBC group. In particular, for each event tracked data will include information about the search triggers associated with it, the PE performed on it, the results of any detector characterization investigations, and the status of the event within the catalog data product discussed in Section 10.2. Strict adherence to this specification—algorithmically enforced—fixes the lack of standardization in internal metadata products which plagued analyses in the third observing run. The centralization of

information about an event into a single file allows for consistent reference to a single source of truth, correcting another significant issue which arose during the third observing run.

Beyond the specification of the JSON data stored in files, CBCFLOW also provides a robust python API for reading, updating, and writing this metadata. A `Metadata` class provides a representation of this data which makes it easy for users to load requested files and read included data. Moreover, this API provides methods to perform updates to metadata in a manner which guarantees continued adherence to the necessary specification. These updates have a number of special cases which must be accounted for: setting scalar values, adding and removing array elements, and acting recursively on array elements which are identified by a unique ID (UID). Accordingly, I developed a recursive tree-search to implement these changes, including a system of creating “negative images” which can be used to remove elements. Cumulatively, this API forms an implementation of standard CRUD operations (Create, Read, Update, and Delete).

CBCFLOW libraries

To manage the collected metadata for events over a period of time—in practice, the subdivisions of the observing run corresponding to each catalog release—CBCFLOW collects many metadata files into a single library. This library is a GIT [5] repository containing the metadata file themselves, along with configuration data and helper scripts. Being a GIT repository, this library may then be duplicated and branched in the normal ways to enable modification. The gitlab instance of this library is considered to be the source of truth, and acts as a clearing house for synchronizing any other instances, and which also operates a continuous integration (CI) pipeline to run helper scripts and data validation routines. The most critical downstream instance is that which exists on the CBC group namespace on CIT, and which is operated upon by the monitor discussed in Section 10.1 and which is kept in sync with the gitlab instance by that same monitor.

The configuration of the library has a few critical elements. Most obviously, since libraries are expected to correspond to fixed time frames, it sets the time boundaries in which events will be included. It also sets thresholds in the false alarm rate for which events will be included. During online operation, this threshold corresponds to the threshold for the initiation of online PE, $FAR < 2/\text{year}$. For offline analyses, it is then extended to the “deep” catalog threshold of $FAR < 2/\text{day}$. Finally,

libraries may be configured in either “gracedb” or “gwtc-gracedb” mode, determining whether events are sourced from the GraceDB online event lists, or from the catalog table discussed in Section 9.1.

Another key feature of CBCFLOW libraries is the use of a custom GIT merge driver. Since GIT is usually used for the management of text or text-like files such as source code, a typical invocation of `git merge` will perform a line-by-line comparison of a file’s contents in two different versions: the base file which current exists, and the head file which is being merged onto the base. However, executing line-by-line merge operations on two valid JSON files satisfying some schema does *not* guarantee that the resulting file will also be a valid JSON file satisfying that schema, due to issues including trailing commas and the splicing of arrays.

The solution to this problem is to redefine how `git merge` is executed. CBCFLOW uses a recursive tree search to merge two JSON files, traveling through each of them to identify discrepancies between the base and head files. A particular challenge is that of *removing* elements from arrays, as identifying when on branch has had an element removed requires a sense of its history. To this end, the CBCFLOW also includes information about the most-recent-common-ancestor commit, which is also traversed to determine if any elements have been removed.

Libraries also include a number of helper scripts and the configuration of a gitlab CI. Data validation is performed automatically, to verify that all files satisfy the necessary JSON specification before they can be merged. Public facing data is collected in a series of gitlab issues, which are automatically updated using another component of the CI. Finally, summary information for weekly teleconferences is collected into summary pages which are produced on a once-per-week basis.

The CBCFLOW monitor

When, in the course of online operations, a new GW event is detected, it is necessary to update CBCFLOW with its initial characteristics, and to continue updating as follow-up PE progresses. Once online operations are complete and the library transitions to catalog mode, it is necessary to update the library regularly as the offline search results are completed, such that CBCFLOW reflects the current state of the catalog. These tasks are all completed by a monitor process, typically running on the CBC group namespace. That process is run periodically, using either a CRONTAB, or the CRONDOR periodic job submission mechanism in HTCONDOR [6]. In practice, we found the CRONTAB implementation to be more stable, and so all standard monitors

run in this mode.

For each event which satisfies the library criterion—as enumerated in Section 10.1—metadata is scraped from GraceDB about each of the search results corresponding to the event. This information includes both the events detection statistics, detection time, and, in the case of matched filter results, the estimated source parameters. Next, the monitor will also scrape the metadata of all initial PE analyses which are logged in the PE group namespace. Using this, it provides the seed information for downstream analyses, as well as the initial contents of preliminary data product releases.

The CBCFLOW-ASIMOV interface

The final crucial component of CBCFLOW is the interface to other automated processes, of which ASIMOV [7] is the most notable. ASIMOV has two hooks within CBCFLOW. The first reads metadata, in particular initial PE results and the recommendations of detector characterization experts, to initialize the production PE for an event. The other updates CBCFLOW with metadata from that analysis, including its status, configuration, and results. Both of these are run by the ASIMOV monitor process, rather than that of CBCFLOW.

10.2 GWTC_PIPE and the catalog data product

Once production PE and offline searches are completed, the CBCFLOW library contains all of the metadata required to assemble a public data product for the catalog release. However, it *also* contains a large amount of other junk: failed analyses, redundant review statements, miscellaneous notes, and so on. Accordingly, direct release of the CBCFLOW library is not appropriate, and it must instead be distilled into a polished data product for public consumption, such as that release for the third observing run [3]. Accomplishing this task is the role of GWTC_PIPE.

GWTC_PIPE is split into two processes, one to collect and summarize PE results, the other to do the same for search results. They share common features—technically, they subclass the same parent class—including methods to load in the correct cbcflow library, determine the appropriate event lists, and produce descriptive information such as the unique hash for a given data release. While running, these processes also check the review status and finalization status of each of the analyses collected, and propagate these back to CBCFLOW for reference.

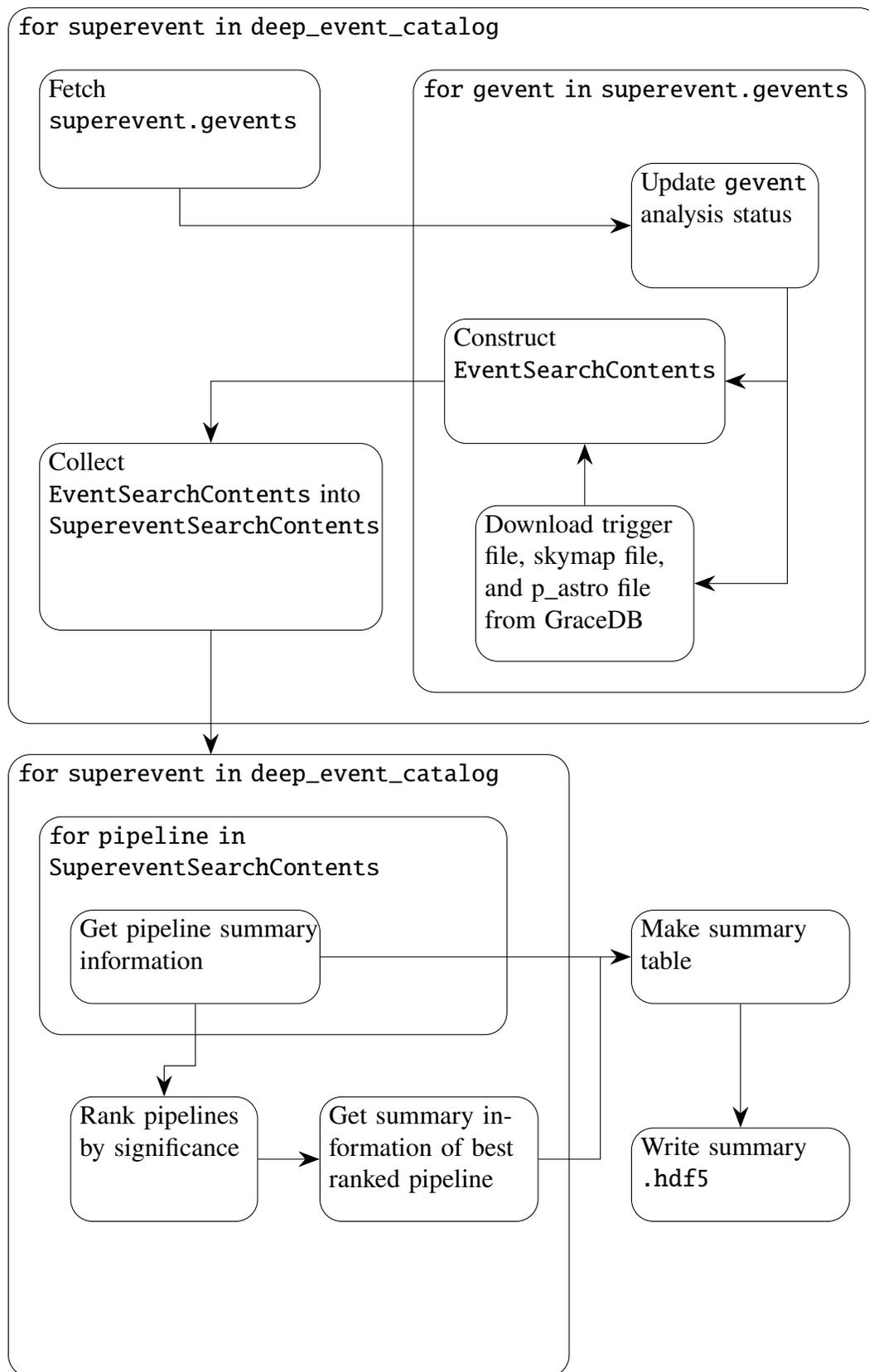


Figure 10.1: A diagram of the process by which search information is compiled into the catalog data product.

Producing the catalog search data product

The purpose of `GWTC_PIPE`'s search components is threefold. First, it determines the review and finalization status of each event given its chunk and search pipeline. Next, it must fetch search products—that is information on the triggers, along with their corresponding source classification and skymap information—from GraceDB for each event. Finally, it produces summary information about each event, and compiles these into summary tables written to an `.hdf5` file. Figure 10.1 illustrates this process.

Configuration for the search pipeline is relatively minimal, consisting of a series of files—one for each search pipeline—detailing the boundaries of the search chunks. Each of these is given a review and finalization status, configured by a search pipeline representative. Using `CBCFLOW`'s search results, which reflect only the search results stored in the most recent catalog table, performs the necessary downselection to production results.

Producing the catalog parameter estimation data product

The requirements for the preparation of the catalog PE data product are similar to those of the search data product. Results must be fetched and collated, the status of analyses must be tracked, and summary information about them must be generated. Unlike with search results, it is also necessary to perform some amount of downselection of the PE results, as all PE analyses are tracked in `CBCFLOW`, as opposed to just the most recent production results. Accordingly, `GWTC_PIPE` internally filters results, preferring production, reviewed, and non-deprecated results which use the approximant requested in the configuration. There are also separate branches for when combined results—`PESUMMARY` files which contain results from all of the finalized production analyses—which will substitute these finalized results instead while still tracking review statuses of each individual analysis. Figure 10.2 illustrates this procedure.

Configuration for the PE side of `GWTC_PIPE` is more involved than that of the search side. For each superevent, the set of “requested” analyses is configured, corresponding to the set of waveform approximants which are expected to be used in analysis. This configuration is built by parsing the `ASIMOV` ledger for the approximants used on each superevent. Also included is the set of parameters for which credible intervals are to be computed. These should generally correspond to the set of parameters used in the `GWOSC` event portal [8].

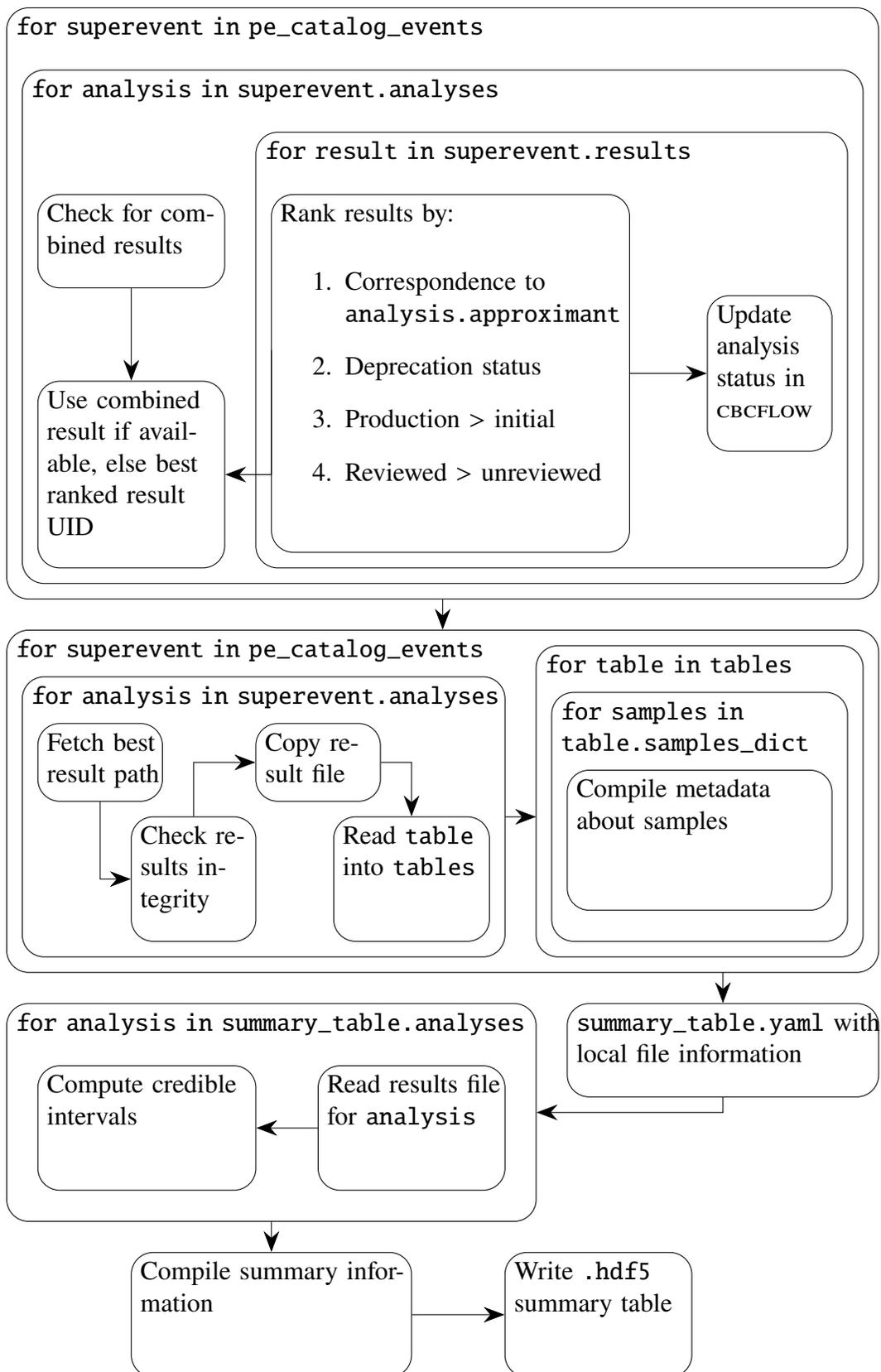


Figure 10.2: A diagram of the process by which PE information is compiled into the catalog data product.

10.3 Outlook for future gravitational wave catalog infrastructure

CBCFLOW is, to use the common parlance, very janky. While JSON specification is frequently used in a variety of roles, storage of data within JSON files is typically managed by more sophisticated infrastructure such as MongoDB [9]. The system of git repository management, and the complex infrastructure required to facilitate robust `git merge` operations, introduces vulnerabilities including race conditions, multiple-source-of-truth situations, slow data-access routines, and non-existent data compression. In the fifth observing run, the number of events is expected to once again increase substantially [10], and so more robust infrastructure must be developed in order to cope with the increased analysis burden. As part of planning for this future, we have also had significant discussions with a team of experts from the CI-Compass collaboration, who specialize in effective management of scientific data, and in this section I will rely heavily upon their insights.

Considerations in the choice of database structure

In database theory, there are broadly speaking two types of database: relational and non-relational [11], also commonly referred to as “SQL” and “NoSQL” databases. Relational databases store data in a series of tables, which reference each other in order to store the relationships between data. These databases typically require rigid schemas in order to function properly, but in turn they offer some flexibility, powerful query syntax, and reasonably efficient data storage. For many decades this paradigm was dominant, and for many applications it remains the standard. By contrast, non-relational databases have a variety forms, including columnar and key-value pairs. These are often more flexible, and can be useful for storing large arrays of numeric data, making them popular in big-data applications [12].

The formal name for CBCFLOW’s system of data management is a document-oriented database [9, 12], meaning that data is stored in a series of documents. These are a subset of non-relational databases, and can inherit some of the advantages: schemas can be more flexible, and representations are relatively intuitive. However, CBCFLOW actually gives up the flexibility advantage by its use of a strict schema, while still losing access to the sophisticated query tools which are available in relational databases. Furthermore, without an external implementation of standard CRUD operations CBCFLOW has substantial vulnerabilities.

In discussion with CI-compass, they recommended that a direct port of CBCFLOW into CLICKHOUSE would be the most effective approach [12]. CLICKHOUSE provides

a columnar (NoSQL) approach, while still maintaining much of the query syntax from SQL. This has the potential create a more convenient representation of the data we already have with minimal intervention. For future development, this may still be the best approach: it's relatively intuitive, easy to set up, and has robust tools for scaling as the amount of data increases, though for the foreseeable future we will remain far below the threshold at which relational databases begin to struggle. With time to perform significant restructuring, however, there is also a case to be made for more dramatic departures.

A future version of CBCFLOW which exists as a proper relational database would have many advantages. Since we have already gone through the effort to develop a robust schema, much of the overhead in developing a relational database approach is already complete. Furthermore, much of our data is also best understood in terms of its relations—especially the hierarchy of analyses performed on a given event—rather than columnar approaches per se. However, this approach would require buy-in from users, for whom this database structure may be less intuitive. To this end, the development of an effective Python API will be critical to further integration.

Choice of Python API

Most scientists within the LVK collaboration use PYTHON as their primary programming language, and this is accordingly the form in which they will generally access metadata. CBCFLOW offers a sophisticated PYTHON API for CRUD operations, but most users do not actually use it. Rather, most users do not bother to perform create, update, or delete operations at all, and only read data by simply loading in the JSON data themselves. While this ease of use does serve as an advantage, for scientific programmers a little freedom can often be a dangerous thing. Ideally, we would construct a PYTHON API which is flexible enough and convenient enough that collaboration members will use it by choice, and seek to increase the integration of their own software with it.

One possibility is to use PYDANTIC [13] with the SQL interface SQLMODEL [14]. PYDANTIC provides type validated data structures for python, with the PYTHON API being inseparable from the schema of the data. SQLMODEL extends this by connecting PYDANTIC models to databases, allowing for efficient representation of the data. Ideally, using these it would be possible to create a user-friendly representation of LVK analysis data, while maintaining the robust operations require to sustain

increasing numbers of automated analysis processes. More work needs to be done to address possible challenges and solutions for this approach.

References

- ¹R. Udall and G. Ashton, CBCFLOW, version 0.6.2, <https://git.ligo.org/cbc/projects/cbcflow>, Apr. 23, 2025.
- ²R. Udall and D. Davis, GWTC_PIPE, version 0.1.0, <https://git.ligo.org/rhiannon.udall/gwtc-pipeline>, Apr. 23, 2025.
- ³R. Abbot et al. (LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration), “Gwtc-3: compact binary coalescences observed by ligo and virgo during the second part of the third observing run — parameter estimation data release”, (2023) DOI:10.5281/zenodo.8177023.
- ⁴F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte, and D. Vrgoč, “Foundations of json schema”, in Proceedings of the 25th international conference on world wide web, WWW ’16 (2016), pp. 263–273, DOI:10.1145/2872427.2883029.
- ⁵D. Spinellis, “Git”, IEEE Software **29**, 100–101 (2012) DOI:10.1109/MS.2012.61.
- ⁶R. A. Erickson, M. N. Fienen, S. G. McCalla, E. L. Weiser, M. L. Bower, J. M. Knudson, and G. Thain, “Wrangling distributed computing for high-throughput environmental science: an introduction to htcondor”, PLoS computational biology **14**, e1006468 (2018) DOI:10.1371/journal.pcbi.1006468.
- ⁷D. Williams, J. Veitch, M. L. Chiofalo, et al., “Asimov: A framework for coordinating parameter estimation workflows”, J. Open Source Softw. **8**, 4170 (2023) DOI:10.21105/joss.04170.
- ⁸R. Abbott et al. (KAGRA, VIRGO, LIGO Scientific), “Open Data from the Third Observing Run of LIGO, Virgo, KAGRA, and GEO”, Astrophys. J. Suppl. **267**, 29 (2023) DOI:10.3847/1538-4365/acdc9f.
- ⁹C.-O. Truică, E.-S. Apostol, J. Darmont, and T. B. Pedersen, “The forgotten document-oriented database management systems: an overview and benchmark of native xml dodbmses in comparison with json dodbmses”, Big Data Research **25**, 100205 (2021) DOI:<https://doi.org/10.1016/j.bdr.2021.100205>.
- ¹⁰D. Shoemaker, N. Arnaud, and T. Sawada, “Observing Scenario timeline graphic, post-O3”, <https://dcc.ligo.org/LIGO-G2002127/public> (2025).
- ¹¹P. Beynon-Davies, Database systems (Palgrave Macmillan, Houndmills, Basingstoke, Hampshire RG21 6XS, 1998).
- ¹²V. Pascucci, S. Petruzza, et al., personal communication, Sept. 18, 2024.
- ¹³Pydantic Services, PYDANTIC, version 2.11.13, <https://github.com/pydantic/pydantic>, Apr. 8, 2025.

¹⁴S. Ramírez et al., SQLMODEL, version 0.0.24, <https://github.com/fastapi/sqlmodel>, Mar. 6, 2025.

Chapter 11

CONCLUSION

In this thesis, I have discussed my work to address two of the challenges which arise as the sensitivity of GW detectors increases: the rapid growth in the number of detected GWs, and the robust measurement of astrophysical source parameters in the presence of glitches. For the first, I have developed two critical components of the infrastructure used to produce GW transient catalogs in the fourth observing run. For the second, I have demonstrated the implementation of a parameterized scattered light model within the BILBY PE framework, and used it for joint glitch-CBC inference of GW191109, one of the most promising candidates in the third observing run for a dynamical formation history. I have also used this implementation to explore a number of questions about the impact of glitches on PE, including simulation studies and the development of a statistic to identify when PE is corrupted by the presence of a glitch. However, much work remains to be done on each of these topics, and I also intend to explore other related topics in the future.

11.1 Future plans for gravitational wave catalog infrastructure

In Chapter 10 I describe the infrastructure I have developed to facilitate the production of GW transient catalogs, but also note the issues that have been identified during its implementation. While the state of this infrastructure is a substantial improvement over the situation in the third observing run, and will suffice for the production of the fourth observing run catalogs, the fifth observing run will have far more GW events to track, and more intensive analyses to perform. As such, the next generation of infrastructure will need to correct many of these challenges to keep pace.

Firstly, we will adopt a proper database system and query language, to reduce technical debt and improve the efficiency of queries. Better API design can make this system more user-friendly, and improve the integration across all aspects of the catalog process. Moreover, institutional changes within the collaboration have and will continue to improve our ability to create a development team to support this work, and institutionalize the infrastructure more effectively going forward.

Separately, the work described in Chapter 10 has not yet had a methods paper

prepared to describe it, and this will be a task I take up in the short term. In addition to CBCFLOW and GWTC_PIPE, this paper will also cover the augmentations made to GRACEDB to support the catalog process.

11.2 Future plans for gravitational wave detector characterization

I intend to continue building upon my work described in this thesis, especially in Chapter 8. As noted in that chapter, I intend to prepare two publications containing some of the contents of that chapter. The first will be investigate the interaction between glitches and PE, while the second will explore statistical tests for PE which has been biased by the presence of a glitch.

In addition to the sections included in this thesis, the first of those publications will also include BAYESWAVE investigations, as well as a section studying how post-subtraction residuals may impact the PE results. In combination with the sections included in this thesis, these investigations have the potential to provide a great deal of job security to detector characterization experts.¹ If, as implied, there are a significant number of glitches below the threshold of detectability which are impacting PE, there is a need to radically rethink the methodology by which we perform PE in the presence of glitches, though some proposed methods may be suitable to such a regime [1, 2]. Furthermore, the results with respect to the subtraction of glitches imply that even when we know of the presence of glitches overlapping signals, the biases due to subtraction may be as bad or worse than those which would result from doing nothing at all, highlighting the need for marginalization techniques such as I have implemented in BILBY.

The second publication which I have planned would help to address these uncertainties, if indeed a strong statistic may be found which identifies whether a given distribution is biased due to a glitch. The principal challenge for this work will be characterizing the strength of the test; that is, finding an effective way to compute the null hypothesis for a given statistic without performing obscene amounts of PE.² Once progress on this front is made, it may be possible identify when the specters of bias described above may be dispelled, and when they must be confronted.

In the future, I will also be extending the methods which I have developed in this thesis to the study of glitches in the space based LISA mission [5]. While in some respects this will be unlike the work I have done previously—the nature of

¹In so far as such a thing could exist for an academic in 2025, which is to say not really at all.

²Or, alternately, accelerating that PE by orders of magnitude, though the work of an undergraduate student [3] attempting this with DINGO [4] has shown this to be easier said than done.

the detector and the glitches which appear in it will be entirely different—in other respects it will be familiar. Lessons about the value of marginalization over glitch realizations, of how valuable physical glitch models can be, and of the necessity for efficient sampling will all carry over to this new analysis scenario.

References

- ¹G. Ashton, “Gaussian processes for glitch-robust gravitational-wave astronomy”, *Mon. Not. Roy. Astron. Soc.* **520**, 2983–2994 (2023) DOI:10.1093/mnras/stad341.
- ²R. Legin, M. Isi, K. W. K. Wong, Y. Hezaveh, and L. Perreault-Levasseur, “Gravitational-Wave Parameter Estimation in non-Gaussian noise using Score-Based Likelihood Characterization”, (2024) DOI:10.48550/arXiv.2410.19956.
- ³S. Kalabalık, “Decoding the effects of varying frequency bands on gw source characterization”, <https://dcc.ligo.org/LIGO-T2400274> (2024).
- ⁴M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, “Real-Time Gravitational Wave Science with Neural Posterior Estimation”, *Phys. Rev. Lett.* **127**, 241103 (2021) DOI:10.1103/PhysRevLett.127.241103.
- ⁵A. Spadaro, R. Buscicchio, D. Vetrugno, A. Klein, S. Vitale, R. Dolesi, W. J. Weber, and M. Colpi, “Glitch systematics on the observation of massive black-hole binaries with LISA”, *Phys. Rev. D* **108**, 123029 (2023) DOI:10.1103/PhysRevD.108.123029.