Topics in Gravitational Wave Physics: Lensing, Detection with Astrometry and Dark Siren Hubble Measurement

Thesis by
Yijun Wang

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2024
Defended May 24 2024

Yijun Wang
ORCID: 0000-0002-5581-2001

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my research advisors Yanbei Chen, Tzu-Ching Chang and Olivier Doré for their mentorship and guidance. Yanbei inspired me with his creative thoughts and breadth of curiosity. His attention to details and acute physics intuition help me stay focused and driven to understand my research work deeply and thoroughly. The freedom he gave me to work independently was one of the times that I grew the fastest as a researcher. I want to thank Tzu-Ching and Olivier for warmly welcoming me to Caltech; I always remember our first conversations in Lauritsen that started our research journey together, and they introduced me to the fascinating and complex world of cosmology and astronomy observation. During my most confused time in career planning, they took the time to share life experiences and lessons with me; you told me that the world is your oyster, and I take great courage in that.

I am immensely grateful for Kris Pardo, who was my postdoc mentor and continued working with me as a faculty in USC; aside from guiding my research, he also helped me with great patience to improve in science writing and communication. His optimistic outlook in research and life gives me strength in every low-SNR situation.

I greatly appreciate my candidacy and thesis committee members, Katerina Chatzi-ioannou and Rana Adhikari for their interest in my research and for offering ideas and suggestions.

I would like to thank my collaborators, Rico K.L. Lo, Alvin K.Y. Li, Ron Tso, Hang Yu, Brian Seymour, Luke Bouma and Benjamin Zhang. They introduced me to research subjects, data analysis methods and different perspectives that greatly enriched my learning experience.

I am grateful for all my friends at Caltech for contributing to my work-life-balance and embarking on colorful adventures and experiences.

I would like to thank my friends back home; Pinyan Liu, for knowing and loving me for more than twenty years; you inspire me by staying the same kind and optimistic person after school and work and everything; Nanzhi Luo, for keeping in touch and motivating me (unknowingly) to stay grounded and work hard; Jinyi Jiang, for being my companion in all the pretty little things in life:)

I also want to thank my partner, Tianzhe Zheng. Thank you for smoothing my ruffled feathers, for being retrospective and receptive, for washing the dishes, for trying your best to understand me, for growing with me and for making me happy. After all these years, I still really like you:)

I want to thank my parents, Zhaohui He and Yongning Wang, for supporting and trusting my choices, for giving me the courage to brave the world and the freedom to just be your child.

ABSTRACT

In this thesis, we study several subjects in gravitational wave (GW) physics, including gravitational wave lensing, detection with astrometry data and dark siren measurement of cosmological parameters.

We investigate various lensing features and their detection prospects in third-generation gravitational-wave networks. Firstly, we focus on type II lensed images which are Hilbert transforms of regular images. We compute the waveform mismatch and quantify the distinguishable fraction given Bayes factor thresholds over a range of binary mass ratio and redshifted mass. We make forecast on the detectable and distinguishable type II images in aLIGO Voyager, Cosmic Explorer and Einstein Telescope. This work shows that a significant number of type II images can be distinguished from waveforms alone, and this strategy can contribute to future pipelines for more accurate GW event inference.

We further model relativistic lensing in a large-inclination hierarchical triple system with a central Kerr supermassive black hole. We combine the elliptical integral formalism and optical scalar formalism to study image location and magnification. By analyzing the repeated lensing signature observed by the Decihertz Gravitational-wave Observatory, we examine the importance of relativistic images in detecting the presence of lensing or specifically the lens spin. We compute the detectable effective volume and estimate the upper limit for expected number of such events. This work demonstrates that lensing with relativistic images is a fruitful avenue where decihertz observation contributes to studies on intermediate-mass binary black holes and their galactic environment.

GW detection with astrometry was proposed as an alternative strategy that uses stellar astrometry data for GW measurement with flexible frequency coverage. We point out that surveys providing relative astrometry only can also be sensitive to GWs. We apply this method to the Roman Space Telescope Galactic Bulge Time Domain survey and make sensitivity forecast for both monochromatic GWs from supermassive binary black holes and stochastic GW background. We clarify the survey requirements and technical challenges for GW detection, and show that Roman will enable microhertz GW measurement for local sources. We also present on-going work to develop a data-processing pipeline to use *Kepler* archival data to search for GWs.

With increasing number of events in GW catalog, the GW source population offers a unique perspective into cosmology and astrophysics. In the last chapter, we use a Fisher information formalism to quantify the astrophysical model error tolerance of GW dark siren measurement on cosmological parameters. We generate galaxy catalog based on realistic survey and population parameters, and we apply expected GW uncertainties in third-generation ground-based networks. Based on simulation results, we study dominating error factors and make suggestions to dark siren selection strategy given different total error requirements.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Yijun Wang, Rico K. L. Lo, Alvin K. Y. Li, and Yanbei Chen. Identifying type II strongly lensed gravitational-wave images in third-generation gravitational-wave detectors. *Phys. Rev. D*, 103(10):104055, May 2021. doi: 10.1103/PhysRevD.103.104055.
YW participated in the idea conception, performed analysis and wrote the majority of the manuscript.

[2] Yijun Wang, Kris Pardo, Tzu-Ching Chang, and Olivier Doré. Gravitational wave detection with photometric surveys. *Phys. Rev. D*, 103(8):084007, April 2021. doi: 10.1103/PhysRevD.103.084007.
YW participated in the idea conception, performed analysis and wrote the manuscript.

[3] Yijun Wang, Kris Pardo, Tzu-Ching Chang, and Olivier Doré. Constraining the stochastic gravitational wave background with photometric surveys. *Phys. Rev. D*, 106(8):084006, October 2022. doi: 10.1103/PhysRevD.106.084006.
YW participated in the idea conception, performed analysis and wrote the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

The successful detection of gravitational-waves (GW) from compact binary mergers by the advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO) [91] and Virgo [10] Collaboration heralded a new window of astronomy observation. Since the first detection of GW150914 [3], subsequent events have greatly enriched our understanding of gravity and many aspects of astrophysics. To maximize the science output from this avenue, immense research efforts have been dedicated to understanding how to improve GW detection, interpret data and broaden its application.

This thesis features exploration in these three aspects through a number of projects. In Chapter 2 and 3, we focus on modeling waveform modifications from propagation effects and their detectability. Chapter 4 and 5 explores GW detection using relative astrometry and makes sensitivity forecast. In Chapter 6, we discuss the bias tolerance of the Hubble constant statistical inference using GW events. In the following section, we provide a brief overview of these topics and present summaries of each main-text chapters.

1.1 GW Spectrum

From the perspective of the GW spectrum, the tremendous success of current ground-based detectors only hints at the wealth of information we can get by unlocking access to GWs of full range of frequencies. Since the GW frequency is tied to the source mass scale, a wide sensitivity band allows us to explore from stellar-mass black holes and neutron stars to supermassive black holes at the center of galaxies. For similar groups of GW sources, observing them at different frequencies also gives us "snapshots" of their evolution. In this section, we briefly review existing and in-planning observing strategies over different GW frequency band.

Ground-based detectors cover the high frequency end between $O(10 - 10\text{k})$ Hz, ideal for detecting solar-mass binary mergers. Existing detectors include aLIGO [2], Virgo [10] and KAGRA [15]. Future upgrades include the aLIGO Voyager [11] and the addition of IndiGO [78]. In the upcoming decades, a new network of third-generation GW detectors are expected to bring orders-of-magnitude sensitivity

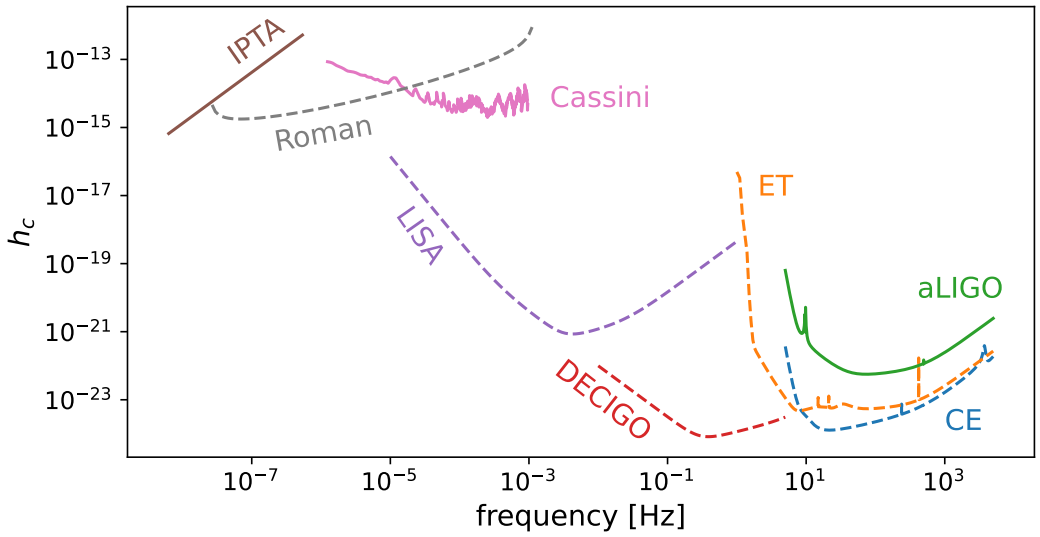

Figure 1.1: Detector power spectral density at different frequencies in the GW spectrum. Current detectors are shown in solid lines and planned ones are showed in dashed lines. See text for detector noise references.

improvement by extending detector baseline (Cosmic Explorer (CE) [121], Einstein Telescope (ET) [125]) [55].

The decihertz band is expected to be covered by space-based detectors. Several proposed detectors are Fabry-Perot Michelson interferometers with the same working principle as ground-based ones, and they reach lower frequency by avoiding Earth-borne noises (e.g. seismic contributions [94]). Observatories in planning include DeciGO [82] and TianGO [85], targeting intermediate-mass black hole binaries between $O(10^2)\ M_\odot$ and $O(10^4)\ M_\odot$ [85].

The millihertz range will be covered by space-based detectors, including the Laser Interferometer Space Antenna (LISA) [17] and TianQin [92]. Different from the decihertz detectors, LISA and TianQin are both transponder-type interferometers, where the mirrors will be in free fall according to their local gravitational potential [17, 82]. These detectors can probe massive black hole mergers (MBHMs) at high redshifts, which is instrumental for mapping black hole evolution history [134].

The lower frequency range is observed with a different set of strategies. Aside from continuous waves from massive black hole binaries (MBHBs) and supermassive black hole binaries (SMBHBs) ($10^5 \sim 10^9\ M_\odot$) [129], the stochastic GW background is another important target. Such background can come from a superposition of SMBHB radiation [42, 114], inflation or cosmic strings [35, 46, 73, 88, 131].

The most recent GW measurement in the microhertz comes from precision Doppler tracking of the *Cassini* Mission, and the isotropic GW background is constrained below $h_c < 2 \times 10^{-15}$ at about 0.3 millihertz [21]. Another detection method is by observing GW-induced apparent oscillation of stars [27, 68, 98, 99, 118, 119]. This method requires measuring astrometry of a large number of stars with high cadence, which suggests that datasets intended for other objectives, such as astrometry precision measurement and exoplanets microlensing, can be repurposed into GW probes. In Ref. [84, 102], the authors consider *Gaia* [117] and present GW sensitivity estimates, as the special instrument design and observing schedule of *Gaia* provides exquisite astrometry quality and numerous targets. Other promising candidates include the Roman Space Telescope (specifically the Galactic Bulge Time Domain survey) [1, 66, 138] and the *Kepler* Mission [28, 100]. Other detection variations have also been proposed; for example, GWs leave imprints on asteroid acceleration [63] and the phase of continuous GWs from galactic sources [79]; by monitoring these "carrier" waves, we can infer the presence of GWs. Interferometer-type detectors similar to LISA have also been introduced [129].

In the nanohertz range, pulsar timing strategy has been extremely fruitful with more than a decade of observing efforts. Pulsar Timing Arrays (PTA) measure the time of arrival (TOA) of pulse signals and look for modification from GWs. Across multiple pulsars, GWs manifest as a common red signal that demonstrates characteristic spatial correlation [22, 54, 101]. In the recent North American Nanohertz Observatory for Gravitational Waves (NANOGrav) [120] 15-year data release [12], it was found that the timing residual correlation between pulsars is consistent with the Hellings-Downs curve expected for stochastic GWs with a 3σ confidence [13, 76]. The background strain amplitude is measured to be $2.4^{+0.7}_{-0.6} \times 10^{-15}$ at $f = 1/\mathrm{yr}^{-1}$[13]. Although the nature of this signal was not established due to suspicious inferred astrophysical model parameters [14], more collected data is expected to fully clarify. Other PTA networks include the European Pulsar Timing Array (EPTA) [64], the Parkes Pulsar Timing Array (PPTA) [83] and the International Pulsar Timing Array (IPTA) [20].

In Figure 1.1, we show detector noise characteristic amplitude at different frequency ranges. Existing detectors are plotted with solid lines, while proposed strategies are shown in dashed lines. The amplitude of *IPTA* is calculated from analytical expression in Ref. [135], and assumes observing 20 pulsars every 17 days for 12.5 years with a timing uncertainty of 250 nanoseconds. The astrometry method

sensitivity forecast for Roman Space Telescope is described in Chapter 5. The *Cassini* constraint is pulled from Ref. [21] via `WebPlotDigitizer`[1]. The noise amplitude for LISA, DeciGO, ET, CE and aLIGO are given in Ref. [55, 124, 139].

In Chapter 4 and Chapter 5, we make sensitivity forecast on detecting continuous waves and stochastic background with the Roman Space Telescope. In the appendix, we discuss ongoing work using the *Kepler* archival data.

1.2 GW Lensing

As the GW travels through the universe, the waveform could pick up imprints from its propagation environment. Identifying these waveform distortions is crucial for properly extracting physical information from data. On one hand, GW waveform distortion from propagation effects can be confused with a shift in the intrinsic source parameters, and, if unaccounted for, can introduce bias in the source parameter estimation and the following analysis. On the other hand, the specific form of the distortion sheds light on properties of the objects the GWs interacted with, offering excellent opportunities to probe the cosmos [7, 87, 97, 112, 113, 137].

Lensing is one such common and important propagation effects. Depending on the impact parameter, GW wavelength and lens mass, lensing can be categorized into strong lensing, weak lensing and microlensing. Within strongly lensed images, we further differentiate relativistic images, when the lens is sufficiently strong (e.g. a supermassive black hole (SMBH)) and the GW path wraps around the lens multiple times before reach the observer.

When GW passes close to the lens and the wavelength is much shorter compared to the mass scale of the lens, the GW is bent strongly enough by the gravitational potential and splits into multiple images [126]. Galaxies are common lenses for GWs in the ground-based detector band [25, 50, 89, 110, 140]. Since the rate of lensed images directly depends on the foreground matter, GW lensing is informative on galaxy number density models [89, 90, 109]. As another example, since the lens solution is shared by both GWs and electromagnetic waves (EM) in general relativity (GR), observing both lensed GWs and EM counterparts gives a constraint on GW speed [41, 59].

For a strong lens such as a SMBH, the gravitational potential is sufficient for wrapping the propagation path around the black hole multiple times, creating relativistic images [31, 53]. Literature for relativistic images typically focuses on EM emis-

[1] `https://apps.automeris.io/wpd/`

sion, and their image position and magnification have been thoroughly investigated [18, 30, 31, 53]. In particular, analytical solutions for Schwarzschild strong lens and special cases for Kerr strong lens with symmetry are provided in the geometric optics limit [29, 31]. These study finds great application in the recent Event Horizon Telescope (EHT) observation [56, 57] on SMBHs at the center of galaxies. Relativistic images form the light ring, which in turn reveals information such as the SMBH spin magnitude, spin axis, accretion disk structure, and even strong-field gravity [26, 69, 70, 74]. The application of relativistic images in GW was discussed in earlier theoretical works [19, 95]. However, the low magnification of relativistic images dictates that only lens system with special geometry can possibly produce detectable signature by next generation GW detectors [77, 81, 86, 92]. In Ref. [52, 141], the authors explore a hierarchical triple lens system where the GW source is in orbit around the SMBH. Successful detection of such images can shed light on SMBH characteristics and astrophysics on the accretion disk in the galactic nuclei [23, 37, 38, 96, 111, 128].

Weak lensing occurs when the GW wavelength and lens mass scale becomes comparable or when the impact parameter is large. Depending on the context, the emphasized effect is either a small amplitude magnification or diffraction [107, 108, 132]. In the former case, GW weak lensing is typically used statistically as a probe of foreground cosmic matter. In combination with galaxy weak lensing and number density information, it is possible to probe deviation from general relativity, such as GW strain damping and massive graviton [103, 104]. In the latter case, diffraction phase modulation from weak lensing is typically used for individual events to probe small-scale foreground matter, e.g. dark matter subhalos [67]. GW microlensing refers to even smaller lens such as stars, and adds finer resolution features to, e.g., galaxy strong lenses [40, 49, 97].

Differentiating lensed GWs from unlensed ones poses great challenge to data analysis. For weakly lensed GWs, diffraction patterns are typically not degenerate with GW source parameters (e.g. [67]). This is not so for strongly lensed images, where, to leading order, the magnification is fully degenerate with a shift in luminosity distance and intrinsic mass [44]. Existing strategies include looking for multiple GW events with consistent intrinsic properties and sky location [33, 45, 75]. However, the false alarm rate (FAR) for this matching (i.e., event pairs with coincidentally similar parameters) increases with total event number, N, as N^2, while the true lensed GWs increase with N; this highlight the potential issue of FAR overwhelming true

lensed GWs [34], especially in light of the expected increase in event number from next generation GW detectors, LIGO Voyager [11], Einstein Telescope [125] and LIGO Cosmic Explorer [121]. Another identification strategy relies on the structure of GW source intrinsic mass distribution and considers outliers as lensing candidates [32, 109]. Nonetheless, not all strongly lensed GWs are degenerate; in Ref. [43, 93], the authors discuss lifting the degeneracy by considering higher GW modes, orbital eccentricity and spin procession.

In Chapter 2, we explore the detectability of type II strongly lensed images. In Chapter 3, we study the waveform and detection prospect of relativistic images in Kerr black hole hierarchical triple systems.

1.3 GW Dark Siren

GW events are tracers of matter structure in the universe. As the GW catalog expands, we expect not only to learn more about the GW source property and population, but also to shed light on cosmology and other astrophysical subjects. One application that has drawn great attention is to use GWs to measure the Hubble constant and clarify on the Hubble tension. The Hubble constant is an important cosmological parameter that describes the expansion of the universe; however, different measurement methods have yielded strongly contradicting conclusions about its value [80, 136]. The *Planck* observation on the Cosmic Microwave Background (CMB) yields $H_0 = 67.4 \pm 0.5$ km/s/Mpc [116]; but "standard candles" [47, 48, 122], e.g. Cepheids in Type Ia supernovae hosts, suggest a value of 73.2 ± 1.3 km/s/Mpc [123]. In light of this contradiction, a third independent test is in great need to shed light on the Hubble tension resolution.

Analogous to the standard candles, GWs contribute to this topic by offering an independent measurement on the luminosity distance. For standard candles, the intrinsic magnitude is known, and the luminosity distance is inferred from apparent magnitude [51]. For GWs, the luminosity distance is directly obtained from the waveform. If the source redshift is also known, the Hubble constant can be directly computed given specific cosmology models [8, 36, 127].

The redshift can be known in two different ways. Firstly, the redshift can be measured from a EM signal generated by the same GW source, which is called the "bright siren" case. The only example to date is GW170817, where the gamma-ray burst is detected by the Fermi Gamma-ray Burst Monitor [4, 5, 9]. Although individual bright siren can produce a Hubble measurement alone, one of the compact bodies

in the binary needs to be a neutron star to generate any EM signal. This population is significantly less numerous than binary black holes (BBHs) [6, 24].

It is also possible to know redshift statistically. In this "dark siren" methodology, each GW is not assigned a EM counterpart or a known host galaxy; rather, we require that the distribution of GW population should be consistent with some theoretical model. If the main focus is the intrinsic mass distribution, the dark siren can be further named "spectral siren" [39, 58, 62]. Models on intrinsic mass distribution can be motivated by astrophysical details on black hole formation, e.g., black hole mass gap due to pair instability [61]. In practice, these models can be captured by parameterized phenomenological models to allow efficiency and to accommodate model uncertainty, e.g. `TRUNCATED POWER LAW`, `BROKEN POWER LAW` and `POWER LAW + PEAK` [133]. The models can also be nonparametric to allow high degree of flexibility against model uncertainties, e.g. using Gaussian Process [60].

If the main focus is the redshift distribution of GW sources, we assume that GWs are tracers of their host galaxy, and we measure the Hubble constant by requiring consistency between the GW catalog luminosity distance distribution and the galaxy redshift profile [127]. This method has been applied with various modeling techniques and GW detector networks. As GW detector sensitivity improves, it is hoped that the dark siren method can constrain the Hubble constant to a percent level [16, 36, 65, 71, 72, 130, 143].

Another flavor of the dark siren focus on the clustering of GWs. Fundamental to this method, it is assumed that both GW sources and galaxies are dark-matter tracers, and the GW source clustering scale should be consistent with that of galaxies at the same redshift, e.g., Ref. [105, 106].

One challenge that the dark siren method, including the spectral sirens, ought to address is astrophysical model uncertainty. In Ref. [115], the authors show that using incorrect models with spectral sirens can give bias up to 3σ. For clustering analysis, a phenomenological form of galaxy bias and GW bias (relative to dark matter distribution) is typically assumed [105]. Similarly, redshift-direction-focused dark siren can also suffer from incorrect galaxy mass function. Since magnitude-limited galaxy catalogs are always incomplete to certain degrees, the full galaxy redshift distribution is a mixture between observed catalog and theoretical rate model. Given incorrect model assumption, the Hubble constant inference is biased. Understanding the dark siren inference's tolerance to these model error is critical, since the statistical error and systematic bias make up the total error budget. Whether the final error is

dominated by the statistics or the bias, clarifying the bottleneck factors is essential for proper usage of the dark siren method. A number of works have explored how changes in the theoretical model could affect the Hubble constant inference [8, 142].

In Chapter 6, we quantify the galaxy model error tolerance in third-generation ground-based GW detector networks and discuss dark siren source selection strategy.

1.4 Chapters Summary

Chapter 2: Identifying type II strongly lensed Gravitational-Wave Images in Third-Generation Gravitational-Wave Detectors

Identifying strongly lensed GW images is challenging as waveform amplitude magnification is degenerate with a shift in the source intrinsic mass and redshift. However, even in the geometric-optics limit, type II strongly lensed images cannot be fully matched by type I (or unlensed) waveform templates, especially with large binary mass ratios and orbital inclination angles. We propose to use this mismatch to distinguish individual type II images. Using planned noise spectra of Cosmic Explorer, Einstein Telescope and LIGO Voyager, we show that a significant fraction of type II images can be distinguished from unlensed sources, given sufficient SNR (~ 30). Incorporating models on GW source population and lens population, we predict the yearly detection rate of lensed GW sources with detectable type II images. We conclude that such distinguishable events are likely to appear in the third-generation detector catalog. Our method will significantly supplement existing strong lensing search strategies.

Chapter 3: Detection Prospects for Gravitational Wave Lensing in Hierarchical Triple Systems with Kerr Black Hole

In many situations, the leading standard lensing images, where the source lies behind the lens and the GW is only weakly deflected, are sufficient, as subsequent relativistic images are negligibly faint. However, next-order images can acquire significant magnification if the lens and source can be highly aligned. A hierarchical triple system, where a stellar-mass binary is in orbit around a supermassive black hole (SMBH), satisfies this condition if the observer inclination is large. Moreover, suitable orbital conditions allow the lensing signatures to be observed multiple times during the detector lifetime. In this work, we solve for lensed waveforms in hierarchical triples with a Kerr central SMBH up to the leading order relativistic images, adopting both the elliptical integral formulation and the optical scalar formalism. We consider the detection prospects of lensing features for the Decihertz Gravita-

tional wave Observatory. We also compute the effective volume where spin-induced lensing features can be distinguished from regular images. This work shows that lensing with relativistic images is a fruitful avenue where decihertz GW observation can contribute to studies on intermediate mass binary black holes and their galactic environment.

Chapter 4: Gravitational Wave Detection with Photometric Surveys

We study a GW detection technique based on astrometric solutions from photometric surveys and demonstrate that it offers a highly flexible frequency range that can uniquely complement existing detection methods. We emphasize that this method can be applied widely to any photometric surveys relying on relative astrometric measurements, in addition to surveys designed to measure absolute astrometry, such as *Gaia*. We illustrate how high-cadence observations of the galactic bulge, such as offered by the Roman Space Telescope's Galactic Bulge Time Domain survey, have the potential to probe GWs with complementary frequency range to *Gaia*, pulsar timing arrays (PTAs), and the Laser Interferometer Space Antenna (LISA). We propose several improvement strategies, including recovering the mean astrometric deflection and increasing astrometric accuracy, number of observed stars, field-of-view size and observational cadence. We also discuss how other existing and planned photometric surveys could contribute.

Chapter 5: Constraining the Stochastic Gravitational Wave Background with Photometric Surveys

The detection of the Stochastic Gravitational Wave Background (SGWB) is essential for understanding black hole population. In this chapter, we investigate the relative astrometric GW detection method in the microhertz range. We consider the Roman Space Telescope and *Gaia* as candidates and quantitatively discuss the survey sensitivity in both the frequency and spatial domains. We emphasize the importance of survey specific constraints on performance estimates by considering mean field-of-view (FoV) signal subtraction and angular power spectrum binning. We conclude that if the SGWB is at a similar level as in PTA estimates, both Roman and *Gaia* have the potential to detect this frequency-domain power excess. However, both Roman and *Gaia* are subject to FoV limitations, and are unlikely to be sensitive to the characteristic spatial pattern of the SGWB.

Chapter 6: Tolerance to Astrophysical Model Uncertainty in Dark Siren Hubble Measurement with Third-generation Gravitational-wave Detectors

In this work, we adopt a Fisher information formalism and study the maximum model error tolerance given specific total error budget. We adopt a mixture model with realistic galaxy survey and population parameters, as well as expected GW measurement uncertainties. We find that, to achieve a total error budget of 1% in the Hubble constant, the galaxy mass function should be known to $O(1\%)$. We identify galaxy redshift uncertainty, magnitude limit and GW angular localization error to be important error factors. In our simulation, we investigate the effective bright siren scenario and highlight that the dark siren selection strategy should be catered to measurement uncertainty and the target total error budget.

References

[1] Roman's Key Science Components, 2020. URL https://www.stsci.edu/roman/about/science-themes.

[2] J. Aasi et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7): 074001, April 2015. doi: 10.1088/0264-9381/32/7/074001.

[3] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, LIGO Scientific Collaboration, and Virgo Collaboration. GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. *Phys. Rev. D*, 93(12):122003, June 2016. doi: 10.1103/PhysRevD.93.122003.

[4] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, B. Agarwal, M. Agathos, (INTEGRAL, et al. Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A. *ApJ*, 848(2):L13, October 2017. doi: 10.3847/2041-8213/aa920c.

[5] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, B. Agarwal, M. Agathos, et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551(7678):85–88, November 2017. doi: 10.1038/nature24471.

[6] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, Ligo Scientific Collaboration Kagra Collaboration, VIRGO Collaboration, et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced

Virgo and KAGRA. *Living Reviews in Relativity*, 21(1):3, April 2018. doi: 10.1007/s41114-018-0012-9.

[7] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, others, LIGO Scientific Collaboration, and Virgo Collaboration. Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1. *Phys. Rev. D*, 100(10):104036, November 2019. doi: 10.1103/PhysRevD.100.104036.

[8] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, LIGO Scientific Collaboration, Virgo Collaboration, et al. A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo. *ApJ*, 909(2):218, March 2021. doi: 10.3847/1538-4357/abdcb7.

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

[10] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2): 024001, January 2015. doi: 10.1088/0264-9381/32/2/024001.

[11] R. X. Adhikari, K. Arai, A. F. Brooks, C. Wipf, O. Aguiar, P. Altin, B. Barr, L. Barsotti, R. Bassiri, A. Bell, et al. A cryogenic silicon interferometer for gravitational-wave detection. *Classical and Quantum Gravity*, 37(16): 165003, August 2020. doi: 10.1088/1361-6382/ab9143.

[12] Gabriella Agazie, Md Faisal Alam, Akash Anumarlapudi, Anne M. Archibald, et al. The NANOGrav 15 yr Data Set: Observations and Timing of 68 Millisecond Pulsars. *ApJ*, 951(1):L9, July 2023. doi: 10.3847/2041-8213/acda9a.

[13] Gabriella Agazie, Akash Anumarlapudi, Anne M. Archibald, Zaven Arzoumanian, et al. The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background. *ApJ*, 951(1):L8, July 2023. doi: 10.3847/2041-8213/acdac6.

[14] Gabriella Agazie, Akash Anumarlapudi, Anne M. Archibald, Paul T. Baker, et al. The NANOGrav 15 yr Data Set: Constraints on Supermassive Black Hole Binaries from the Gravitational-wave Background. *ApJ*, 952(2):L37, August 2023. doi: 10.3847/2041-8213/ace18b.

[15] T. Akutsu, M. Ando, K. Arai, Y. Arai, S. Araki, A. Araya, N. Aritomi, et al. Overview of KAGRA: Detector design and construction history. *Progress of Theoretical and Experimental Physics*, 2021(5):05A101, May 2021. doi: 10.1093/ptep/ptaa125.

[16] Viviane Alfradique, Miguel Quartin, Luca Amendola, Tiago Castro, and Alexandre Toubiana. The lure of sirens: joint distance and velocity measurements with third-generation detectors. *MNRAS*, 517(4):5449–5462, December 2022. doi: 10.1093/mnras/stac2920.

[17] Pau Amaro-Seoane, Heather Audley, Stanislav Babak, John Baker, Enrico Barausse, et al. Laser Interferometer Space Antenna. *arXiv e-prints*, art. arXiv:1702.00786, February 2017.

[18] Paolo Amore, Mayra Cervantes, Arturo de Pace, and Francisco M. Fernández. Gravitational lensing from compact bodies: Analytical results for strong and weak deflection limits. *Phys. Rev. D*, 75(8):083005, April 2007. doi: 10.1103/PhysRevD.75.083005.

[19] Nils Andersson. Scattering of massless scalar waves by a schwarzschild black hole: A phase-integral study. *Phys. Rev. D*, 52:1808–1820, Aug 1995. doi: 10.1103/PhysRevD.52.1808. URL https://link.aps.org/doi/10.1103/PhysRevD.52.1808.

[20] J. Antoniadis, Z. Arzoumanian, S. Babak, M. Bailes, A. S. Bak Nielsen, P. T. Baker, C. G. Bassa, B. Bécsy, et al. The International Pulsar Timing Array second data release: Search for an isotropic gravitational wave background. *MNRAS*, 510(4):4873–4887, March 2022. doi: 10.1093/mnras/stab3418.

[21] J. W. Armstrong, L. Iess, P. Tortora, and B. Bertotti. Stochastic Gravitational Wave Background: Upper Limits in the 10^{-6} to 10^{-3} Hz Band. *ApJ*, 599(2):806–813, December 2003. doi: 10.1086/379505.

[22] Stanislav Babak and Alberto Sesana. Resolving multiple supermassive black hole binaries with pulsar timing arrays. *Phys. Rev. D*, 85(4):044034, February 2012. doi: 10.1103/PhysRevD.85.044034.

[23] Jillian M. Bellovary, Mordecai-Mark Mac Low, Barry McKernan, and K. E. Saavik Ford. Migration Traps in Disks around Supermassive Black Holes. *ApJ*, 819(2):L17, March 2016. doi: 10.3847/2041-8205/819/2/L17.

[24] Marek Biesiada, Xuheng Ding, Aleksandra Piórkowska, and Zong-Hong Zhu. Strong gravitational lensing of gravitational waves from double compact binaries-perspectives for the Einstein Telescope. *J. Cosmology Astropart. Phys.*, 2014(10):080–080, October 2014. doi: 10.1088/1475-7516/2014/10/080.

[25] Marek Biesiada, Xuheng Ding, Aleksandra Piórkowska, and Zong-Hong Zhu. Strong gravitational lensing of gravitational waves from double compact binaries—perspectives for the Einstein Telescope. *J. Cosmology Astropart. Phys.*, 2014(10):080, October 2014. doi: 10.1088/1475-7516/2014/10/080.

[26] Ezequiel F. Boero and Osvaldo M. Moreschi. Strong gravitational lens image of the M87 black hole with a simple accreting matter model. *MNRAS*, 507 (4):5974–5990, November 2021. doi: 10.1093/mnras/stab2336.

[27] Laura G. Book and Éanna É. Flanagan. Astrometric effects of a stochastic gravitational wave background. *Phys. Rev. D*, 83(2):024024, January 2011. doi: 10.1103/PhysRevD.83.024024.

[28] William J. Borucki. KEPLER Mission: development and overview. *Reports on Progress in Physics*, 79(3):036901, March 2016. doi: 10.1088/0034-4885/79/3/036901.

[29] V. Bozza. Quasiequatorial gravitational lensing by spinning black holes in the strong field limit. *Phys. Rev. D*, 67(10):103006, May 2003. doi: 10.1103/PhysRevD.67.103006.

[30] V. Bozza. Comparison of approximate gravitational lens equations and a proposal for an improved new one. *Phys. Rev. D*, 78(10):103005, November 2008. doi: 10.1103/PhysRevD.78.103005.

[31] Valerio Bozza. Gravitational lensing by black holes. *General Relativity and Gravitation*, 42(9):2269–2300, September 2010. doi: 10.1007/s10714-010-0988-2.

[32] Tom Broadhurst, Jose M. Diego, and III Smoot, George. Reinterpreting Low Frequency LIGO/Virgo Events as Magnified Stellar-Mass Black Holes at Cosmological Distances. *arXiv e-prints*, art. arXiv:1802.05273, February 2018.

[33] Tom Broadhurst, Jose M. Diego, and III Smoot, George F. Twin LIGO/Virgo Detections of a Viable Gravitationally-Lensed Black Hole Merger. *arXiv e-prints*, art. arXiv:1901.03190, January 2019.

[34] Mesut Caliskan, Jose Maria Ezquiaga, Otto Hannuksela, and Daniel Holz. Lensing or luck? False alarm probabilities for gravitational lensing of gravitational waves. In *APS April Meeting Abstracts*, volume 2022 of *APS Meeting Abstracts*, page Y15.006, April 2022.

[35] Chia-Feng Chang and Yanou Cui. Stochastic gravitational wave background from global cosmic strings. *Physics of the Dark Universe*, 29:100604, September 2020. doi: 10.1016/j.dark.2020.100604.

[36] Hsin-Yu Chen, Maya Fishbach, and Daniel E. Holz. A two per cent Hubble constant measurement from standard sirens within five years. *Nature*, 562 (7728):545–547, October 2018. doi: 10.1038/s41586-018-0606-0.

[37] Xian Chen and Wen-Biao Han. Extreme-mass-ratio inspirals produced by tidal capture of binary black holes. *Communications Physics*, 1(1):53, September 2018. doi: 10.1038/s42005-018-0053-0.

[38] Xian Chen, Shuo Li, and Zhoujian Cao. Mass-redshift degeneracy for the gravitational-wave sources in the vicinity of supermassive black holes. *MNRAS*, 485(1):L141–L145, May 2019. doi: 10.1093/mnrasl/slz046.

[39] David F. Chernoff and Lee S. Finn. Gravitational Radiation, Inspiraling Binaries, and Cosmology. *ApJ*, 411:L5, July 1993. doi: 10.1086/186898.

[40] Mark H. Y. Cheung, Joseph Gais, Otto A. Hannuksela, and Tjonnie G. F. Li. Stellar-mass microlensing of gravitational waves. *arXiv e-prints*, art. arXiv:2012.07800, December 2020.

[41] Thomas E. Collett and David Bacon. Testing the Speed of Gravitational Waves over Cosmological Distances with Strong Gravitational Lensing. *Phys. Rev. Lett.*, 118(9):091101, March 2017. doi: 10.1103/PhysRevLett.118.091101.

[42] Curt Cutler and Eanna 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.

[43] Liang Dai and Tejaswi Venumadhav. On the waveforms of gravitationally lensed gravitational waves. *arXiv e-prints*, art. arXiv:1702.04724, February 2017.

[44] Liang Dai, Tejaswi Venumadhav, and Kris Sigurdson. Effect of lensing magnification on the apparent distribution of black hole mergers. *Phys. Rev. D*, 95(4):044011, February 2017. doi: 10.1103/PhysRevD.95.044011.

[45] Liang Dai, Barak Zackay, Tejaswi Venumadhav, Javier Roulet, and Matias Zaldarriaga. Search for Lensed Gravitational Waves Including Morse Phase Information: An Intriguing Candidate in O2. *arXiv e-prints*, art. arXiv:2007.12709, July 2020.

[46] Thibault Damour and Alexander Vilenkin. Gravitational radiation from cosmic (super)strings: Bursts, stochastic background, and observational windows. *Phys. Rev. D*, 71:063510, 2005. doi: 10.1103/PhysRevD.71.063510.

[47] T. de Jaeger, L. Galbany, S. González-Gaitán, R. Kessler, A. V. Filippenko, F. Förster, M. Hamuy, P. J. Brown, DES Collaboration, et al. Studying Type II supernovae as cosmological standard candles using the Dark Energy Survey. *MNRAS*, 495(4):4860–4892, July 2020. doi: 10.1093/mnras/staa1402.

[48] Suhail Dhawan, Saurabh W. Jha, and Bruno Leibundgut. Measuring the Hubble constant with Type Ia supernovae as near-infrared standard candles. *A&A*, 609:A72, January 2018. doi: 10.1051/0004-6361/201731501.

[49] J. M. Diego, O. A. Hannuksela, P. L. Kelly, G. Pagano, T. Broadhurst, K. Kim, T. G. F. Li, and G. F. Smoot. Observational signatures of microlensing in gravitational waves at LIGO/Virgo frequencies. *A&A*, 627:A130, July 2019. doi: 10.1051/0004-6361/201935490.

[50] Xuheng Ding, Marek Biesiada, and Zong-Hong Zhu. Strongly lensed gravitational waves from intrinsically faint double compact binaries—prediction for the Einstein Telescope. *J. Cosmology Astropart. Phys.*, 2015(12):006, December 2015. doi: 10.1088/1475-7516/2015/12/006.

[51] Scott Dodelson. *Modern Cosmology*. 2003.

[52] Daniel J. D'Orazio and Abraham Loeb. Repeated gravitational lensing of gravitational waves in hierarchical black hole triples. *Phys. Rev. D*, 101(8): 083031, April 2020. doi: 10.1103/PhysRevD.101.083031.

[53] Ernesto F. Eiroa and Diego F. Torres. Strong field limit analysis of gravitational retrolensing. *Phys. Rev. D*, 69(6):063004, March 2004. doi: 10.1103/PhysRevD.69.063004.

[54] J. A. Ellis, X. Siemens, and J. D. E. Creighton. Optimal Strategies for Continuous Gravitational Wave Detection in Pulsar Timing Arrays. *ApJ*, 756 (2):175, September 2012. doi: 10.1088/0004-637X/756/2/175.

[55] M. Evans, R. Sturani, S. Vitale, and E. Hall. Unofficial sensitivity curves (ASD) for aLIGO, Kagra, Virgo, Voyager, Cosmic Explorer and ET. Technical Report LIGO-T1500293-v11, September 2018. URL https://dcc.ligo.org/LIGO-T1500293-v11/public.

[56] Event Horizon Telescope Collaboration, Kazunori Akiyama, Antxon Alberdi, Walter Alef, Keiichi Asada, Rebecca Azulay, Anne-Kathrin Baczko, et al. First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. *ApJ*, 875(1):L1, April 2019. doi: 10.3847/2041-8213/ab0ec7.

[57] Event Horizon Telescope Collaboration, Kazunori Akiyama, Antxon Alberdi, Walter Alef, Juan Carlos Algaba, Richard Anantua, Keiichi Asada, et al. First Sagittarius A* Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole in the Center of the Milky Way. *ApJ*, 930(2):L12, May 2022. doi: 10.3847/2041-8213/ac6674.

[58] Jose María Ezquiaga and Daniel E. Holz. Spectral Sirens: Cosmology from the Full Mass Distribution of Compact Binaries. *Phys. Rev. Lett.*, 129(6): 061102, August 2022. doi: 10.1103/PhysRevLett.129.061102.

[59] Xi-Long Fan, Kai Liao, Marek Biesiada, Aleksandra Piórkowska-Kurpas, and Zong-Hong Zhu. Speed of gravitational waves from strongly lensed gravitational waves and electromagnetic signals. *Physical Review Letters*, 118(9):091102, 2017.

[60] Amanda M. Farah, Thomas A. Callister, Jose María Ezquiaga, Michael Zevin, and Daniel E. Holz. No need to know: astrophysics-free gravitational-wave cosmology. *arXiv e-prints*, art. arXiv:2404.02210, April 2024. doi: 10.48550/arXiv.2404.02210.

[61] R. 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 Supernova Black Hole Mass Gap. *ApJ*, 887(1):53, December 2019. doi: 10.3847/1538-4357/ab518b.

[62] Will M. Farr, Maya Fishbach, Jiani Ye, and Daniel E. Holz. A Future Percent-level Measurement of the Hubble Expansion at Redshift 0.8 with Advanced LIGO. *ApJ*, 883(2):L42, October 2019. doi: 10.3847/2041-8213/ab4284.

[63] Michael A. Fedderke, Peter W. Graham, and Surjeet Rajendran. Asteroids for microhertz gravitational-wave detection. *arXiv e-prints*, art. arXiv:2112.11431, December 2021.

[64] R. D. Ferdman, R. van Haasteren, C. G. Bassa, M. Burgay, I. Cognard, A. Corongiu, et al. The European Pulsar Timing Array: current efforts and a LEAP toward the future. *Classical and Quantum Gravity*, 27(8):084014, April 2010. doi: 10.1088/0264-9381/27/8/084014.

[65] Andreas Finke, Stefano Foffa, Francesco Iacovelli, Michele Maggiore, and Michele Mancarella. Cosmology with LIGO/Virgo dark sirens: Hubble parameter and modified gravitational wave propagation. *J. Cosmology Astropart. Phys.*, 2021(8):026, August 2021. doi: 10.1088/1475-7516/2021/08/026.

[66] B. Scott Gaudi, Rachel Akeson, Jay Anderson, Etienne Bachelet, David P. Bennett, Aparna Bhattacharya, Valerio Bozza, et al. 'Auxiliary' Science with the WFIRST Microlensing Survey. *BAAS*, 51(3):211, May 2019.

[67] Han Gil Choi, Chanung Park, and Sunghoon Jung. Small-scale shear: Peeling off diffuse subhalos with gravitational waves. *arXiv e-prints*, art. arXiv:2103.08618, March 2021. doi: 10.48550/arXiv.2103.08618.

[68] Sebastian Golat and Carlo R. Contaldi. All-sky analysis of astrochronometric signals induced by gravitational waves. *arXiv e-prints*, art. arXiv:2201.03903, January 2022.

[69] Samuel E. Gralla and Alexandru Lupsasca. Lensing by Kerr black holes. *Phys. Rev. D*, 101(4):044031, February 2020. doi: 10.1103/PhysRevD.101.044031.

[70] Samuel E. Gralla, Alexandru Lupsasca, and Daniel P. Marrone. The shape of the black hole photon ring: A precise test of strong-field general relativity. *Phys. Rev. D*, 102(12):124004, December 2020. doi: 10.1103/PhysRevD.102.124004.

[71] R. Gray, C. Messenger, and J. Veitch. A pixelated approach to galaxy catalogue incompleteness: improving the dark siren measurement of the Hubble constant. *MNRAS*, 512(1):1127–1140, May 2022. doi: 10.1093/mnras/stac366.

[72] Rachel Gray, Ignacio Magaña Hernandez, Hong Qi, Ankan Sur, Patrick R. Brady, Hsin-Yu Chen, Will M. Farr, Maya Fishbach, Jonathan R. Gair, Archisman Ghosh, Daniel E. Holz, Simone Mastrogiovanni, Christopher Messenger, Danièle A. Steer, and John Veitch. Cosmological inference using gravitational wave standard sirens: A mock data analysis. *Phys. Rev. D*, 101(12):122001, June 2020. doi: 10.1103/PhysRevD.101.122001.

[73] Leonid P. Grishchuk. REVIEWS OF TOPICAL PROBLEMS: Relic gravitational waves and cosmology. *Physics Uspekhi*, 48(12):1235–1247, December 2005. doi: 10.1070/PU2005v048n12ABEH005795.

[74] Shahar Hadar, Michael D. Johnson, Alexandru Lupsasca, and George N. Wong. Photon ring autocorrelations. *Phys. Rev. D*, 103(10):104038, May 2021. doi: 10.1103/PhysRevD.103.104038.

[75] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith. Search for Gravitational Lensing Signatures in LIGO-Virgo Binary Black Hole Events. *ApJ*, 874(1):L2, March 2019. doi: 10.3847/2041-8213/ab0c0f.

[76] R. W. Hellings and G. S. Downs. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. *ApJ*, 265:L39–L42, February 1983. doi: 10.1086/183954.

[77] Wen-Rui Hu and Yue-Liang Wu. The Taiji Program in Space for gravitational wave physics and the nature of gravity. *National Science Review*, 4(5): 685–686, 10 2017. ISSN 2095-5138. doi: 10.1093/nsr/nwx116. URL https://doi.org/10.1093/nsr/nwx116.

[78] B. Iyer, T. Souradeep, C. Unnikrishnan, S. Dhurandhar, S. Raja, and A. Sengupta. LIGO-India, Proposal of the Consortium for Indian Initiative in Gravitational-wave Observations (IndIGO),[Public]. Technical Report LIGO-M1100296-v2, November 2011. URL https://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=75988.

[79] María José Bustamante-Rosell, Joel Meyers, Noah Pearson, Cynthia Trendafilova, and Aaron Zimmerman. Gravitational Wave Timing Array. *arXiv e-prints*, art. arXiv:2107.02788, July 2021.

[80] Marc Kamionkowski and Adam G. Riess. The Hubble Tension and Early Dark Energy. *arXiv e-prints*, art. arXiv:2211.04492, November 2022. doi: 10.48550/arXiv.2211.04492.

[81] Seiji Kawamura, Masaki Ando, Naoki Seto, Shuichi Sato, Takashi Nakamura, Kimio Tsubono, et al. The Japanese space gravitational wave antenna: DECIGO. *Classical and Quantum Gravity*, 28(9):094011, May 2011. doi: 10.1088/0264-9381/28/9/094011.

[82] Seiji Kawamura, Takashi Nakamura, Masaki Ando, Naoki Seto, Tomotada Akutsu, et al. Space gravitational-wave antennas DECIGO and B-DECIGO. *International Journal of Modern Physics D*, 28(12):1845001, January 2019. doi: 10.1142/S0218271818450013.

[83] Matthew Kerr, Daniel J. Reardon, George Hobbs, Ryan M. Shannon, Richard N. Manchester, Shi Dai, Christopher J. Russell, Songbo Zhang, et al. The Parkes Pulsar Timing Array project: second data release. *PASA*, 37:e020, June 2020. doi: 10.1017/pasa.2020.11.

[84] Sergei A. Klioner. Gaia-like astrometry and gravitational waves. *Classical and Quantum Gravity*, 35(4):045005, February 2018. doi: 10.1088/1361-6382/aa9f57.

[85] Kevin A. Kuns, Hang Yu, Yanbei Chen, and Rana X Adhikari. Astrophysics and cosmology with a deci-hertz gravitational-wave detector: TianGO. *arXiv e-prints*, art. arXiv:1908.06004, August 2019.

[86] Kevin A. Kuns, Hang Yu, Yanbei Chen, and Rana X. Adhikari. Astrophysics and cosmology with a decihertz gravitational-wave detector: Tiango. *Phys. Rev. D*, 102:043001, Aug 2020. doi: 10.1103/PhysRevD.102.043001. URL https://link.aps.org/doi/10.1103/PhysRevD.102.043001.

[87] Kwun-Hang Lai, Otto A. Hannuksela, Antonio Herrera-Martín, Jose M. Diego, Tom Broadhurst, and Tjonnie G. F. Li. Discovering intermediate-mass black hole lenses through gravitational wave lensing. *Phys. Rev. D*, 98 (8):083005, October 2018. doi: 10.1103/PhysRevD.98.083005.

[88] L. Lentati, S. R. Taylor, C. M. F. Mingarelli, A. Sesana, S. A. Sanidas, A. Vecchio, et al. European Pulsar Timing Array limits on an isotropic stochastic gravitational-wave background. *MNRAS*, 453(3):2576–2598, November 2015. doi: 10.1093/mnras/stv1538.

[89] Shun-Sheng Li, Shude Mao, Yuetong Zhao, and Youjun Lu. Gravitational lensing of gravitational waves: a statistical perspective. *MNRAS*, 476(2): 2220–2229, May 2018. doi: 10.1093/mnras/sty411.

[90] Kai Liao, Xi-Long Fan, Xuheng Ding, Marek Biesiada, and Zong-Hong Zhu. Precision cosmology from future lensed gravitational wave and electromagnetic signals. *Nature Communications*, 8:1148, October 2017. doi: 10.1038/s41467-017-01152-9.

[91] LIGO Scientific Collaboration, J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, April 2015. doi: 10.1088/0264-9381/32/7/074001.

[92] Jun Luo, Li-Sheng Chen, Hui-Zong Duan, Yun-Gui Gong, Shoucun Hu, Jianghui Ji, Qi Liu, et al. TianQin: a space-borne gravitational wave detector. *Classical and Quantum Gravity*, 33(3):035010, February 2016. doi: 10.1088/0264-9381/33/3/035010.

[93] Jose María Ezquiaga, Daniel E. Holz, Wayne Hu, Macarena Lagos, and Robert M. Wald. Phase effects from strong gravitational lensing of gravitational waves. *arXiv e-prints*, art. arXiv:2008.12814, August 2020.

[94] F. Matichard, B. Lantz, R. Mittleman, K. Mason, J. Kissel, B. Abbott, S. Biscans, et al. Seismic isolation of Advanced LIGO: Review of strategy, instrumentation and performance. *Classical and Quantum Gravity*, 32(18): 185003, September 2015. doi: 10.1088/0264-9381/32/18/185003.

[95] Richard A. Matzner, Cécile DeWitte-Morette, Bruce Nelson, and Tian-Rong Zhang. Glory scattering by black holes. *Phys. Rev. D*, 31:1869–1878, Apr 1985. doi: 10.1103/PhysRevD.31.1869. URL https://link.aps.org/doi/10.1103/PhysRevD.31.1869.

[96] Barry McKernan, K. E. Saavik Ford, J. Bellovary, N. W. C. Leigh, Z. Haiman, B. Kocsis, W. Lyra, M. M. Mac Low, B. Metzger, M. O'Dowd, S. Endlich, and D. J. Rosen. Constraining Stellar-mass Black Hole Mergers in AGN Disks Detectable with LIGO. *ApJ*, 866(1):66, October 2018. doi: 10.3847/1538-4357/aadae5.

[97] Ashish Kumar Meena and Jasjeet Singh Bagla. Gravitational lensing of gravitational waves: wave nature and prospects for detection. *MNRAS*, 492 (1):1127–1134, February 2020. doi: 10.1093/mnras/stz3509.

[98] Deyan P. Mihaylov, Christopher J. Moore, Jonathan R. Gair, Anthony Lasenby, and Gerard Gilmore. Astrometric Effects of Gravitational Wave Backgrounds with non-Einsteinian Polarizations. *Phys. Rev. D*, 97(12): 124058, 2018. doi: 10.1103/PhysRevD.97.124058.

[99] Deyan P. Mihaylov, Christopher J. Moore, Jonathan R. Gair, Anthony Lasenby, and Gerard Gilmore. Astrometric effects of gravitational wave backgrounds with nonluminal propagation speeds. *Phys. Rev. D*, 101(2): 024038, January 2020. doi: 10.1103/PhysRevD.101.024038.

[100] David G. Monet, Jon M. Jenkins, Edward W. Dunham, Stephen T. Bryson, Ronald L. Gilliland, David W. Latham, William J. Borucki, and David G. Koch. Preliminary Astrometric Results from Kepler. *arXiv e-prints*, art. arXiv:1001.0305, January 2010. doi: 10.48550/arXiv.1001.0305.

[101] C. J. Moore, S. R. Taylor, and J. R. Gair. Estimating the sensitivity of pulsar timing arrays. *Classical and Quantum Gravity*, 32(5):055004, March 2015. doi: 10.1088/0264-9381/32/5/055004.

[102] Christopher J. Moore, Deyan P. Mihaylov, Anthony Lasenby, and Gerard Gilmore. Astrometric Search Method for Individually Resolvable Gravitational Wave Sources with Gaia. *Phys. Rev. Lett.*, 119(26):261102, December 2017. doi: 10.1103/PhysRevLett.119.261102.

[103] Suvodip Mukherjee, Benjamin D. Wandelt, and Joseph Silk. Probing the theory of gravity with gravitational lensing of gravitational waves and galaxy surveys. *MNRAS*, 494(2):1956–1970, May 2020. doi: 10.1093/mnras/staa827.

[104] Suvodip Mukherjee, Benjamin D. Wandelt, and Joseph Silk. Multimessenger tests of gravity with weakly lensed gravitational waves. *Phys. Rev. D*, 101 (10):103509, May 2020. doi: 10.1103/PhysRevD.101.103509.

[105] Suvodip Mukherjee, Benjamin D. Wandelt, Samaya M. Nissanke, and Alessandra Silvestri. Accurate precision cosmology with redshift unknown gravitational wave sources. *Phys. Rev. D*, 103(4):043520, February 2021. doi: 10.1103/PhysRevD.103.043520.

[106] Remya Nair, Sukanta Bose, and Tarun Deep Saini. Measuring the Hubble constant: Gravitational wave observations meet galaxy clustering. *Phys. Rev. D*, 98(2):023502, July 2018. doi: 10.1103/PhysRevD.98.023502.

[107] T. T. Nakamura and S. Deguchi. Wave Optics in Gravitational Lensing. *Progress of Theoretical Physics Supplement*, 133:137–153, January 1999. doi: 10.1143/PTPS.133.137.

[108] Takahiro T. Nakamura. Gravitational Lensing of Gravitational Waves from Inspiraling Binaries by a Point Mass Lens. *Phys. Rev. Lett.*, 80(6):1138–1141, February 1998. doi: 10.1103/PhysRevLett.80.1138.

[109] Ken K. Y. Ng, Kaze W. K. Wong, Tom Broadhurst, and Tjonnie G. F. Li. Precise LIGO lensing rate predictions for binary black holes. *Phys. Rev. D*, 97(2):023012, January 2018. doi: 10.1103/PhysRevD.97.023012.

[110] Masamune Oguri. Effect of gravitational lensing on the distribution of gravitational waves from distant binary black hole mergers. *MNRAS*, 480(3): 3842–3855, November 2018. doi: 10.1093/mnras/sty2145.

[111] Ryan M. O'Leary, Bence Kocsis, and Abraham Loeb. Gravitational waves from scattering of stellar-mass black holes in galactic nuclei. *MNRAS*, 395 (4):2127–2146, June 2009. doi: 10.1111/j.1365-2966.2009.14653.x.

[112] Kris Pardo, Maya Fishbach, Daniel E. Holz, and David N. Spergel. Limits on the number of spacetime dimensions from GW170817. *J. Cosmology*

Astropart. Phys., 2018(7):048, July 2018. doi: 10.1088/1475-7516/2018/07/048.

[113] Scott Perkins and Nicolás Yunes. Probing screening and the graviton mass with gravitational waves. *Classical and Quantum Gravity*, 36(5):055013, March 2019. doi: 10.1088/1361-6382/aafce6.

[114] E. S. Phinney. A Practical Theorem on Gravitational Wave Backgrounds. *arXiv e-prints*, art. astro-ph/0108028, August 2001.

[115] Grégoire Pierra, Simone Mastrogiovanni, Stéphane Perriès, and Michela Mapelli. Study of systematics on the cosmological inference of the Hubble constant from gravitational wave standard sirens. *Phys. Rev. D*, 109(8): 083504, April 2024. doi: 10.1103/PhysRevD.109.083504.

[116] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, et al. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, September 2020. doi: 10.1051/0004-6361/201833910.

[117] T. Prusti et al. The Gaia mission. *A&A*, 595:A1, November 2016. doi: 10.1051/0004-6361/201629272.

[118] Ted Pyne, Carl R. Gwinn, Mark Birkinshaw, T. Marshall Eubanks, and Demetrios N. Matsakis. Gravitational Radiation and Very Long Baseline Interferometry. *ApJ*, 465:566, July 1996. doi: 10.1086/177443.

[119] Wenzer Qin, Kimberly K. Boddy, Marc Kamionkowski, and Liang Dai. Pulsar-timing arrays, astrometry, and gravitational waves. *Phys. Rev. D*, 99 (6):063002, March 2019. doi: 10.1103/PhysRevD.99.063002.

[120] Scott Ransom, A. Brazier, S. Chatterjee, T. Cohen, J. M. Cordes, M. E. DeCesar, P. B. Demorest, J. S. Hazboun, M. T. Lam, R. S. Lynch, M. A. McLaughlin, S. M. Ransom, X. Siemens, S. R. Taylor, and S. J. Vigeland. The NANOGrav Program for Gravitational Waves and Fundamental Physics. In *Bulletin of the American Astronomical Society*, volume 51, page 195, September 2019.

[121] David Reitze, Rana X. Adhikari, Stefan Ballmer, Barry Barish, Lisa Barsotti, GariLynn Billingsley, Duncan A. Brown, Yanbei Chen, et al. Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. In *Bulletin of the American Astronomical Society*, volume 51, page 35, September 2019.

[122] Adam G. Riess, William H. Press, and Robert P. Kirshner. A Precise Distance Indicator: Type IA Supernova Multicolor Light-Curve Shapes. *ApJ*, 473:88, December 1996. doi: 10.1086/178129.

[123] Adam G. Riess, Stefano Casertano, Wenlong Yuan, J. Bradley Bowers, Lucas Macri, Joel C. Zinn, and Dan Scolnic. Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM. *ApJ*, 908(1): L6, February 2021. doi: 10.3847/2041-8213/abdbaf.

[124] Travis Robson, Neil J. Cornish, and Chang Liu. The construction and use of LISA sensitivity curves. *Classical and Quantum Gravity*, 36(10):105011, May 2019. doi: 10.1088/1361-6382/ab1101.

[125] B. Sathyaprakash, M. Abernathy, F. Acernese, P. Amaro-Seoane, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker, N. Beveridge, S. Birindelli, S. Bose, L. Bosi, S. Braccini, C. Bradaschia, T. Bulik, et al. Scientific Potential of Einstein Telescope. *arXiv e-prints*, art. arXiv:1108.1423, August 2011.

[126] Peter Schneider, Jürgen Ehlers, and Emilio E. Falco. *Gravitational Lenses*. 1992. doi: 10.1007/978-3-662-03758-4.

[127] B. F. Schutz. Determining the Hubble constant from gravitational wave observations. *Nature*, 323(6086):310–311, September 1986. doi: 10.1038/323310a0.

[128] Amy Secunda, Jillian Bellovary, Mordecai-Mark Mac Low, K. E. Saavik Ford, Barry McKernan, Nathan W. C. Leigh, Wladimir Lyra, and Zsolt Sándor. Orbital Migration of Interacting Stellar Mass Black Holes in Disks around Supermassive Black Holes. *ApJ*, 878(2):85, June 2019. doi: 10.3847/1538-4357/ab20ca.

[129] Alberto Sesana, Natalia Korsakova, Manuel Arca Sedda, Vishal Baibhav, Enrico Barausse, Simon Barke, Emanuele Berti, et al. Unveiling the gravitational universe at μ-Hz frequencies. *Experimental Astronomy*, 51(3):1333–1383, June 2021. doi: 10.1007/s10686-021-09709-9.

[130] Arman Shafieloo, Ryan E. Keeley, and Eric V. Linder. Will cosmic gravitational wave sirens determine the Hubble constant? *J. Cosmology Astropart. Phys.*, 2020(3):019, March 2020. doi: 10.1088/1475-7516/2020/03/019.

[131] Xavier Siemens, Jeffrey Hazboun, Paul T. Baker, Sarah Burke-Spolaor, Dustin R. Madison, Chiara Mingarelli, Joseph Simon, and Tristan Smith. Physics Beyond the Standard Model With Pulsar Timing Arrays. *BAAS*, 51 (3):437, May 2019.

[132] Ryuichi Takahashi and Takashi Nakamura. Wave Effects in the Gravitational Lensing of Gravitational Waves from Chirping Binaries. *ApJ*, 595(2):1039–1051, October 2003. doi: 10.1086/377430.

[133] The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, R. Abbott, H. Abe, F. Acernese, K. Ackley, N. Adhikari, R. X. Adhikari, V. K. Adkins, et al. Constraints on the cosmic expansion history from GWTC-3. *arXiv e-prints*, art. arXiv:2111.03604, November 2021. doi: 10.48550/arXiv.2111.03604.

[134] James Ira Thorpe, John Ziemer, Ira Thorpe, Jeff Livas, John W. Conklin, Robert Caldwell, Emanuele Berti, Sean T. McWilliams, Robin Stebbins, David Shoemaker, Elizabeth C. Ferrara, Shane L. Larson, et al. The Laser Interferometer Space Antenna: Unveiling the Millihertz Gravitational Wave Sky. In *BAAS*, volume 51, page 77, September 2019.

[135] Eric Thrane and Joseph D. Romano. Sensitivity curves for searches for gravitational-wave backgrounds. *Phys. Rev. D*, 88(12):124032, December 2013. doi: 10.1103/PhysRevD.88.124032.

[136] Licia Verde, Tommaso Treu, and Adam G. Riess. Tensions between the early and late Universe. *Nature Astronomy*, 3:891–895, September 2019. doi: 10.1038/s41550-019-0902-0.

[137] Aditya Vijaykumar, M. V. S. Saketh, Sumit Kumar, Parameswaran Ajith, and Tirthankar Roy Choudhury. Probing the large scale structure using gravitational-wave observations of binary black holes. *arXiv e-prints*, art. arXiv:2005.01111, May 2020.

[138] WFIRST Astrometry Working Group, Robyn E. Sanderson, Andrea Bellini, Stefano Casertano, Jessica R. Lu, Peter Melchior, Mattia Libralato, David Bennett, Michael Shao, Jason Rhodes, Sangmo Tony Sohn, Sangeeta Malhotra, Scott Gaudi, S. Michael Fall, Ed Nelan, Puragra Guhathakurta, Jay Anderson, and Shirley Ho. Astrometry with the Wide-Field Infrared Space Telescope. *Journal of Astronomical Telescopes, Instruments, and Systems*, 5: 044005, October 2019. doi: 10.1117/1.JATIS.5.4.044005.

[139] Kent Yagi and Naoki Seto. Detector configuration of DECIGO/BBO and identification of cosmological neutron-star binaries. *Phys. Rev. D*, 83(4): 044011, February 2011. doi: 10.1103/PhysRevD.83.044011.

[140] Lilan Yang, Xuheng Ding, Marek Biesiada, Kai Liao, and Zong-Hong Zhu. How Does the Earth's Rotation Affect Predictions of Gravitational Wave Strong Lensing Rates? *ApJ*, 874(2):139, April 2019. doi: 10.3847/1538-4357/ab095c.

[141] Hang Yu, Yijun Wang, Brian Seymour, and Yanbei Chen. Detecting gravitational lensing in hierarchical triples in galactic nuclei with space-borne gravitational-wave observatories. *Phys. Rev. D*, 104(10):103011, November 2021. doi: 10.1103/PhysRevD.104.103011.

[142] Hang Yu, Brian Seymour, Yijun Wang, and Yanbei Chen. Uncertainty and Bias of Cosmology and Astrophysical Population Model from Statistical Dark Sirens. *ApJ*, 941(2):174, December 2022. doi: 10.3847/1538-4357/ac9da0.

[143] Liang-Gui Zhu, Yi-Ming Hu, Hai-Tian Wang, Jian-dong Zhang, Xiao-Dong Li, Martin Hendry, and Jianwei Mei. Constraining the cosmological parameters using gravitational wave observations of massive black hole binaries and statistical redshift information. *Physical Review Research*, 4(1):013247, March 2022. doi: 10.1103/PhysRevResearch.4.013247.

Chapter 2

IDENTIFYING TYPE II STRONGLY LENSED GRAVITATIONAL-WAVE IMAGES IN THIRD-GENERATION GRAVITATIONAL-WAVE DETECTORS

Wang, Y., Lo. R. K. L., Li. A. K. Y. and Chen, Y. (2021). "Identifying type II strongly lensed gravitational-wave images in third-generation gravitational-wave detectors". In: *Phys.Rev.D*, 103(10):104055. DOI:10.1103/PhysRevD.103.104055.

2.1 Introduction

Successful detection of gravitational wave (GW) signals from compact binary mergers by the Advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO) and Virgo collaboration has greatly enriched our understanding of gravity and many aspects of astrophysics [see, e.g. 5, 70]. To extract physical information from detector data, proper signal interpretation is crucial. For this purpose, it is important to study changes in the waveform as it propagates through the universe, since, if unaccounted for, propagation effects can be confused with intrinsic GW features and introduce bias in subsequent analysis. On the other hand, results of propagation effects depend on properties both of the GW and the objects along its path that it interacts with [see, e.g., 3, 41, 48, 58, 59, 75]. Therefore, identifying such signatures also maximizes the scientific output of GW detection.

One GW propagation effect is strong gravitational lensing, in which the rays of a GW are bent strongly enough by a gravitational potential and form multiple images with different magnifications. Gravitational lensing of gravitational waves has attracted enormous interest. It has been estimated that third-generation detectors can detect up to hundreds of strongly lensed events [11, 43, 57, 77]; such events can then be used to study cosmological structures [22, 27, 43, 44, 55] and fundamental physics [19, 29, 51, 52].

However, identifying strongly lensed images is challenging since the predominant effect of strong lensing, namely the amplitude magnification by $\sqrt{\mu}$, is degenerate with scaling down the luminosity distance, D_L, by $\sqrt{\mu}$ and keeping the redshifted mass, $M_\bullet(1+z_s)$, constant [see, e.g., 23], where $M_\bullet$ is the total mass of the binary and z_s is the redshift of this GW source. This degeneracy stems from the fact that

General Relativity is a scale-free geometric theory, and that GW frequency evolution is unaffected by strong lensing [23].

Current search strategies typically look for multiple events in a catalog that are consistent in intrinsic properties and sky locations, and have orbital phase related in characteristic ways [see, e.g., 24, 35]. For example, this has been used to study the series of events GW170104, GW170814 and a sub-threshold trigger, GWC170620, as potential candidates for lensed images [24]. In Ref. [14], the event GW170814 and GW170809 are analyzed as potential strongly lensed companion images using a similar consistency test.

It is also proposed that a sharp transition in the inferred source intrinsic mass distribution at high mass values could single out strongly lensed images [13, 55]. This mass distribution anomaly argument, however, must be made in reference to an expected GW source distribution. Currently, such source population models are subject to considerable uncertainties.

The above strategies share two other drawbacks: (1) without prior knowledge of the lensed source parameters, all pairs of cataloged events must be searched over to find strongly lensed candidates. As detector sensitivity improves and next-generation detectors start observing, the computational cost of such analysis will surge with the increased number of detected events; (2) it is also required that more than one lensed images are detected. If all but one of the images are missed, the methods described above cannot ascertain if a GW image is strongly lensed.

For the above reasons, an intrinsic waveform distortion in a lensed image can be both a more definitive and efficient indicator of strong lensing. If such a lensed image is found, its estimated parameters help narrowing down the search space in the more general pair-wise search method mentioned above. An example is the frequency-specific GW diffraction in weak lensing [54, 69]. Diffraction signature was searched for in current detected events, but it has yet to be found [35]. It is also predicted that, for GWs within the frequency range of LIGO, diffraction becomes important when the lens mass ranges from $1 \sim 100\ M_\odot$ [17, 26, 48]. Strong lensing by such small lenses requires a small impact parameter, which places stringent requirement on the alignment of the GW source, the lens and the observer [see, e.g., 64]. Consequently, we expect such events to be rare.

Though diffraction is negligible for strong lensing (within a similar frequency range as LIGO), waveform distortion does occur in the geometric-optics limit when an

image originates from a saddle-point solution to the lens equation [see, e.g., 64]. Such images are called type II images, their waveforms are the Hilbert transforms of the corresponding unlensed waveforms. By contrast, waveforms of type I and type III images are identical to the unlensed waveform, up to a rescaling—and, for type III images, a sign flip. In Ref. [22], it is pointed out that type II images are degenerate with type I images with an azimuthal angle shift of $\pi/4$, if only the dominant (2,±2) modes are considered. For highly eccentric orbits, this degeneracy is partially lifted. For quasi-circular binaries, the degeneracy is also broken if higher multipoles are considered. Recently, Ref. [46] has systematically examined the type II images of a wide range of GW sources, including the effects of spin precession and orbital eccentricity. It was found that the type I/II waveform difference is still small upon tuning the azimuthal angle, the polarization angle and relative phases between GW modes. In this paper, we build upon the work in Ref. [46] by exploring whether such type II images can be distinguished from regular images despite the small waveform mismatch.

For third-generation detectors, such as the LIGO Voyager [7], the Einstein Telescope[1] [63] (ET) and LIGO Cosmic Explorer[2] [61] (CE) with current models, we expect to be able to detect a non-trivial number of such distinguishable type II events thanks to the expected high Signal-to-Noise Ratio (SNR).

This paper is organized as follows. In Section 2.2 we review the geometric optics theory for GW lensing. In Section 2.3, we calculate the best-match overlap between type II and type I waveforms over a range of detector-frame binary mass, mass ratio and orbital inclinations. We briefly discuss the implication of waveform mismatch for detection triggering in the current LIGO pipeline framework. In Section 2.4 we discuss the distinguishability of type II images in the high-SNR regime by comparing the log likelihoods under type I and type II image hypothesis. Based on this, we compute the fraction of distinguishable type II images. In Section 2.5, we incorporate population models on GW sources and lensing galaxies, and predict the expected number of events with distinguishable type II images for LIGO Voyager, ET and CE. We then discuss the results and draw the conclusion.

Throughout this work, we assume a ΛCDM universe with $(\Omega_M, \Omega_\Lambda) = (0.3, 0.7)$ and a Hubble Constant of $H_0 = 70$ km $s^{-1} Mpc^{-1}$.

[1] http://www.et-gw.eu/

[2] https://cosmicexplorer.org/

2.2 Lens Theory and Image Type

The geometric optics treatment of gravitational lensing is thoroughly investigated and well established by many authors [see, e.g., 53, 54, 64]. In this section, we summarize and discuss scenarios where type II images are distinctive from type I counterparts. We closely follow the discussion in Ref. [53] and keep mostly consistent notations.

Thin gravitational lens: geometric-optics limit

We adopt the thin lens model, in which the line-of-sight lens dimension is much smaller than separations between the GW source, the lens and the observer. The source plane and the lens plane are defined by the GW source and the lens center, and both planes are perpendicular to the optical axis connecting the lens center and the observer. All the lens mass is projected onto the lens plane. Lensing deflection to GW paths occurs only on the lens plane.

On each plane, the origin is established as its intersection with the optical axis. The source position has the dimensionless coordinate $\vec{y} = \vec{\eta} D_d/(r_* D_s)$ and the GW path intersects the lens plane at $\vec{x} = \vec{\xi}/r_*$. $\vec{\eta}, \vec{\xi}$ are coordinates with physical units of length, r_* is the lens' Einstein radius, while D_d, D_s are the observer's angular diameter distance to the lens and the source.

The amplitude of the observed image is then expressed as a Kirchhoff integral over the lens plane [see also, e.g., 64, 69],

$$F(\omega, \vec{y}) = \frac{\omega}{2\pi i} \int d^2\vec{x} e^{i\omega t(\vec{x},\vec{y})} \ , \tag{2.1}$$

where ω is the source-frame GW frequency and $t\,(\vec{x}, \vec{y})$ is the GW travel time difference between lensed paths and the unlensed path,

$$t\,(\vec{x}, \vec{y}) \approx \frac{1}{2}|\vec{x} - \vec{y}|^2 + t_\Phi \ , \tag{2.2}$$

where the first term accounts for the geometrical extra path length in the small deflection limit and the second term, t_Φ, is the Shapiro time delay inside the lens' gravitational potential. In the geometric optics limit, only paths very close to the stationary points of t contribute to the integral, and we may Taylor-expand the time delay around the j-th stationary point,

$$t\,(\vec{x}, \vec{y}) = t\,(\vec{x}_j, \vec{y}) + \frac{1}{2} dx^a dx^b T_{,ab}\,(\vec{x}_j, \vec{y}) + O(|d\vec{x}|^3) \ , \tag{2.3}$$

where dx^a is a component of the two-dimensional vector $d\vec{x} \equiv \vec{x} - \vec{x}_j$ on the lens plane, and $|,|$ denotes partial derivatives and repeated upper and lower indices imply

summation. The integral in Eq. (2.1) then reduces to two Gaussian integrals after diagonalizing the time delay Jacobian, $T_{,ab}$.

When $\det(T_{,ab}) > 0$, phase shifts from both the ω/i prefactor and the two Gaussian integrals depend on the sign of ω. When $\mathrm{Tr}(T_{,ab}) > 0$, the phase factor is 1, giving type I images. When $\mathrm{Tr}(T_{,ab}) < 0$, the phase shift is $-\mathrm{sgn}(\omega)\pi$, where the function sgn returns the sign of its argument. This phase shift gives type III images, which differ from type I by an overall phase of π. (Note that $\pm\pi$ phases are equivalent.)

When $\det(T_{,ab}) < 0$, the two Gaussian integrals give opposite phase shifts regardless the sign of ω, and no longer contribute to the overall phase of $F(\omega, \vec{y})$. The overall phase shift is then $-\mathrm{sgn}(\omega)\pi/2$, giving type II images which are equivalent to a Hilbert transform of type I images.

Gravitational waves from circular, non-spinning binaries

For compact binaries, the complex GW strain at infinity can be written as

$$h = h_+ - ih_\times = \sum_{l,m} {}_{-2}Y_{lm}(\iota, \phi) h_{lm} \ , \tag{2.4}$$

where the subscripts $+, \times$ denote plus and cross polarizations, and ${}_{-2}Y_{lm}(\iota, \phi)$ is the $s = -2$ spin-weighted spherical harmonics. For non-spinning binaries with quasi-circular orbits, we choose the coordinate system such that the orbital angular momentum is along the z axis. In this way, arguments ι and ϕ of the spin-weighted spherical harmonic also corresponds to the orbital inclination angle and the azimuthal angle, respectively.

Let us start out by considering $m \neq 0$ modes. The contribution from modes with $m = \pm m_0$, where m_0 is a positive integer, is

$$\tilde{h}_{\mathrm{I},m_0} = \sum_l \sum_{m=\pm m_0} {}_{-2}Y_{lm}(\iota, \phi) \tilde{h}_{\mathrm{I},lm} \ , \tag{2.5}$$

where the subscript I denotes the regular type I waveforms. The quantity $\tilde{h}_{\mathrm{I},lm}$ is the Fourier transform of h_{lm} in Eq. (2.4) via

$$\tilde{h}_{\mathrm{I},lm}(f) = \int_{-\infty}^{\infty} h_{lm}(t) e^{-2\pi i f t} dt \ . \tag{2.6}$$

We note that ϕ appears only in the factor of $\exp(im\phi)$ in ${}_{-2}Y_{lm}(\iota, \phi)$. Furthermore, for non-spinning, circular binaries, with orbital angular momentum along the z axis, in frequency domain, $m > 0$ modes only have negative frequency components and

the inverse is true for $m < 0$ modes. Therefore, the Hilbert transform of $\tilde{h}_{\mathrm{I},m_0}$, $\tilde{h}_{\mathrm{II},m_0}$ is written as

$$\begin{aligned}\tilde{h}_{\mathrm{II},m_0}(\iota,\phi) &= -i\ \mathrm{sgn}(f)\ \tilde{h}_{\mathrm{I},m_0}(\iota,\phi) \\ &= \tilde{h}_{\mathrm{I},m_0}\left(\iota,\phi+\frac{\pi}{2m_0}\right)\ .\end{aligned} \tag{2.7}$$

Therefore, for each subset of GW modes with $m = \pm m_0$, the Hilbert transform is degenerate with an additional orbital azimuthal angle $\Delta\phi = \pi/(2m_0)$. For example, the required angle change is $\Delta\phi = \pi/4$, provided that only the $(l, \pm 2)$ GW modes are considered. Modes with different $|m|$ require different angle changes to compensate for the Hilbert transform (e.g., the $(l, \pm 3)$ modes require $\Delta\phi = \pi/6$). This difference in the compensation requirements breaks the degeneracy between Hilbert-transformed signals and orbital azimuthal angle change.

Physically, $|m| \neq 2$ modes can be significant when the orbit is significantly eccentric [22]. For binaries with significant mass ratios and inclination angles, the $(3, \pm 3)$ modes become significant, breaking degeneracy. Figure 2.1 is analogous to Figure 2 in Ref. [22] and plots example type I/II waveforms from a binary with a detector-frame mass $\tilde{M} = 150\ M_\odot$, a mass ratio $q = 2.2$ and an orbital inclination angle $\iota = 80\,\mathrm{deg}$. The binary is non-spinning in a quasi-circular orbit, and all multipoles with $l \leq 4$ are included. The top two panels show that the type II image is not degenerate with the type I image with an additional time shift. The bottom panels show that, when we include only the $m = \pm m_0$ modes, the Hilbert transform is degenerate with the original waveform with $\Delta\phi = \pi/(2m_0)$.

For $m = 0$ modes, h is independent from ϕ, and one cannot recover its Hilbert transform via shifting ϕ. This in principle further breaks the degeneracy, although $m = 0$ modes are generally weak for non-spinning binaries in circular orbits. However, note that these are where the GW memory effects take place [30, 31, 38].

In this paper, we systematically explore GW sources which are non-spinning binary black holes in quasi-circular orbits. The distinguishable signature of type II images will be due to higher order GW modes, which is related to binary mass ratio, q, and orbital inclination, ι.

2.3 Waveform Mismatch

In this section, we quantify the mismatch between type I/II waveforms for non-spinning binaries, in preparation for discussion on their distinguishability in the next section. We also discuss the implication of this mismatch for the GW signal

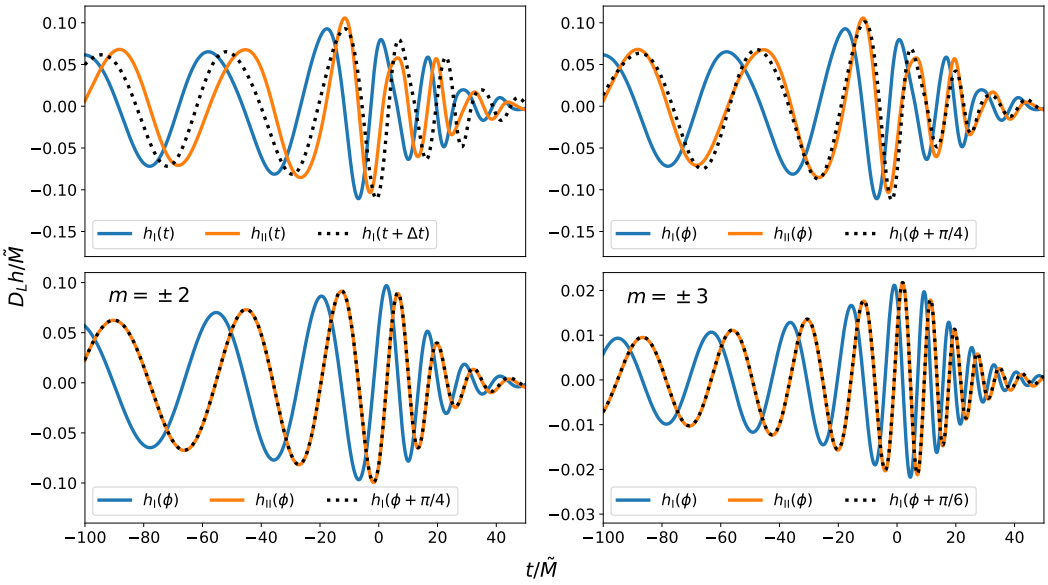

Figure 2.1: type I/II `NRSur7dq4` surrogate model waveforms from a binary with $\tilde{M} = 150\ M_\odot, q = 2.2, \iota = 80$ deg. The binary is non-spinning in a quasi-circular orbit. The black dotted line shows the type I waveform with a $\pi/4$ shift in the orbital azimuthal angle, and the shifted waveform is completely degenerate with the type II image. The orange dotted line shows the type II waveform, such that its peak overlaps with that of the type I waveform. We observe that the type I/II waveform offset cannot be compensated by a time shift.

veto process, namely, whether the mismatch leads to type II signal rejection in the current LIGO data analysis pipeline.

Best-match Overlap

In this section, we describe the procedure to compute the type I/II waveform difference over a large parameter space. We model only non-spinning binaries in quasi-circular orbits. Highly spinning binaries or those with highly eccentric orbits are expected to be fewer than the population we consider [see, e.g., 2, 78]. Since, the optical depth for type II images is also small, on the order of $10^{-3} \sim 10^{-4}$ [23, 43, 57], we exclude these less frequent sources from our analysis.

For this source population, frequency-domain GW strain is given by the Fourier transform of Eq. (2.4),

$$\tilde{h}_{\rm I}(f) = \sum_{l,m} {}_{-2}Y_{lm}(\iota,\phi)\frac{\tilde{H}_{{\rm I},lm}(\tilde{M},q,f)}{D_L}e^{-2\pi i f t_0 - i\Phi}\ , \tag{2.8}$$

where $\tilde{H}_{{\rm I},lm}(\tilde{M},q,f)/D_L$ is equal to $\tilde{h}_{{\rm I},lm}(f)$ in Eq. (2.6), with the dependence on

D_L explicitly shown. The waveform is a function of the detector-frame mass (or equivalently, the redshifted mass), $\tilde{M} = (1+z)M$ (where M is the intrinsic mass), the mass ratio, $q \equiv \tilde{M}_1/\tilde{M}_2 \geq 1$ ($\tilde{M}_1 + \tilde{M}_2 = \tilde{M}$), and the luminosity distance, D_L. The polarization angle, Φ, and signal time-of-arrival, t_0, add additional phase shifts to the signal.

For any two waveforms, $\tilde{h}_1, \tilde{h}_2$, we define the overlap by

$$\text{overlap} = \frac{\mathfrak{Re}(\langle \tilde{h}_1 | \tilde{h}_2 \rangle)}{\sqrt{\langle \tilde{h}_1 | \tilde{h}_1 \rangle \langle \tilde{h}_2 | \tilde{h}_2 \rangle}} = 1 - \epsilon \,, \tag{2.9}$$

where ϵ is the mismatch and $\langle \cdot \rangle$ denotes inner product given by

$$\langle a | b \rangle = \int_{-\infty}^{\infty} \frac{a^*(f) b(f)}{S_n(f)} df \,, \tag{2.10}$$

where $S_n(f)$ is the two-sided noise power spectral density. By applying the optimal matched filter, the SNR of $\tilde{h}$, ρ, is given by $\sqrt{\langle \tilde{h} | \tilde{h} \rangle}$.

Throughout this paper, we use Roman numeral subscripts to denote the image types and Arabic numeral subscripts to represent any individual waveform. We also adopt the simplifying assumption that both GW polarizations can be independently detected, i.e., the time-domain waveform is taken to be complex, as in Eq. 2.4. In Section 2.6, we discuss in more detail the validity of this assumption.

To obtain highly accurate models for $\tilde{h}$, we adopt the time-domain Numerical Relativity surrogate waveform model, `NRSur7dq4`, [74] extracted through the Python package, `gwsurrogate` [32]. This surrogate model provides all $l \leq 4$ mode waveforms, $h_{\mathrm{I},lm}(t)$, through the inspiral, merger and ringdown phases.

To avoid spurious edge effects due to the finite-length of surrogate waveforms, we apply a time-domain kaiser window function from `numpy.kaiser` [73] with $\beta = 4$. The window is centered at the waveform amplitude peak to maximally preserve waveform features. The signal is zero-padded prior to the Fourier transform to ensure sufficiently smooth transformed waveform.

To maximize the overlap, Ref. [46] separately tunes the azimuthal angle, polarization angle, and the relative phases between each of the GW multipoles. We adopt a different approach by tuning the intrinsic parameters of the GW sources. We adopt a nested search method. We first generate a type II signal template, $\tilde{h}_{\mathrm{II},0}$, with $(\tilde{M}_0, q_0, \iota_0, \phi_0)$. Since the waveform amplitude scaling does not contribute to the overlap, we fix $D_L = 3$ Gpc for all waveforms. We make a $(\tilde{M}, q)$ grid, with the

mass range centered on $\tilde{M}_0$ and mass ratio between 1 and 4 (the range of q used to train the surrogate model). At each grid point, we construct the type I template and use the Python module `scipy.optimize.dual_annealing` [76] to find the (ι, ϕ) that maximize the overlap between the type I template and the type II target. The spin-weighted spherical harmonics are computed using the Python package `spherical_functions`[3] and `quaternion`[4]. To implicitly maximize over t_0 and Φ, we take the Fourier transform of the integrand in Eq. (2.9) and pick the element with the largest absolute value [see, e.g., 50]:

$$
\begin{aligned}
(1-\epsilon)_{\max} &= \max_{t_0} \left| \frac{1}{a} \int_{-\infty}^{\infty} df \frac{\tilde{h}_{\rm I}^*(f)\tilde{h}_{II}(f)}{S_n(f)} e^{-2\pi i f t_0} \right| , \\
a &= \sqrt{\langle \tilde{h}_{\rm I}|\tilde{h}_{\rm I}\rangle \langle \tilde{h}_{\rm II}|\tilde{h}_{\rm II}\rangle} \ .
\end{aligned}
\tag{2.11}
$$

Figure 2.2 shows an example maximization result contour plot for a type II signal with $\tilde{M} = 150\ M_\odot$, $q = 1.7$ and $\iota = 70$ deg with CE noise curve. Due to the waveform mismatch, the best-match template has different parameter values from those of the true signal, with a maximal overlap of 99.06%.

We calculate the overlap for sample points on the grid $X(\tilde{M}) \bigotimes Y(q) \bigotimes Z(\iota)$, with

$$
\begin{aligned}
X(\tilde{M}) &= \{60, 80, 100, 150, 200, 230, 260, \\
&\qquad 300, 400, 500, 600, 700, 800\}[M_\odot] \ , \\
Y(q) &= \{1.2, 1.7, 2.2, 2.7, 3.2\} \ , \\
Z(\iota) &= \{15, 30, 40, 50, 60, 70, 80\}[\text{deg}] \ .
\end{aligned}
$$

We then interpolate between the samples using the `scipy.interpolate` module [76] to construct a function $\epsilon(\tilde{M}, q, \iota)$. For non-spinning binaries, the interpolated function ensures that 90 deg $< \iota <$ 180 deg is symmetric to 0 deg $< \iota <$ 90 deg. We perform the same analyses for CE, ET and LIGO Voyager with their respective noise power spectral density (PSD) [7, 28, 61, 63].

Figure 2.3 shows the amplitude spectral density of CE, ET and LIGO Voyager [7, 28, 61, 63], as well as the waveform of a binary with $\tilde{M} = 200\ M_\odot$, $q = 2.2$, $\iota =$ 80 deg at $D_L = 1$ Gpc as an example. The low-frequency amplitude loss of the surrogate waveform is due to the finite length of the `NRSur7dq4` waveforms. For

[3] `https://github.com/moble/spherical_functions`
[4] `https://github.com/moble/quaternion`

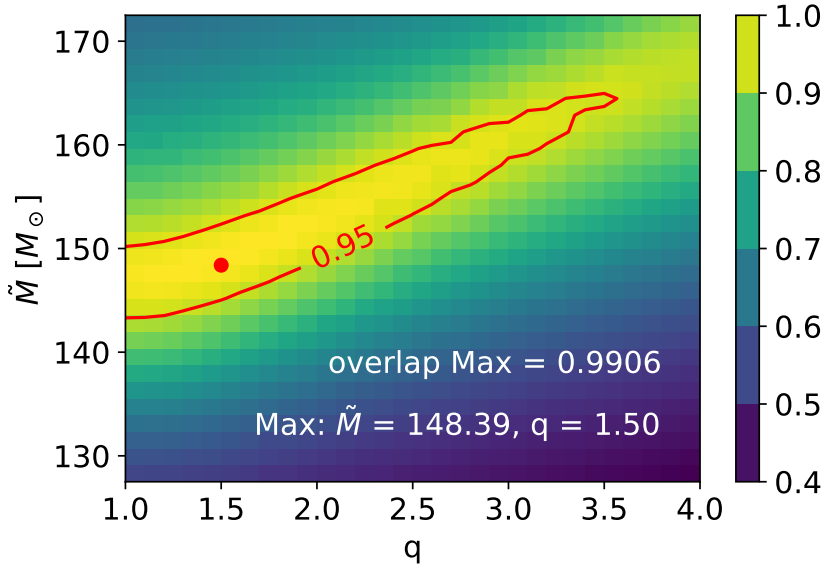

Figure 2.2: Contour plot for maximized overlap for a type II waveform with $\tilde{M} = 150\ M_\odot$, $q = 1.7$ and $\iota = 70$ deg. Grid point with the maximum overlap is shown with the red dot at $\tilde{M} = 148.39\ M_\odot$, $q = 1.50$. 95% overlap contour is shown in red.

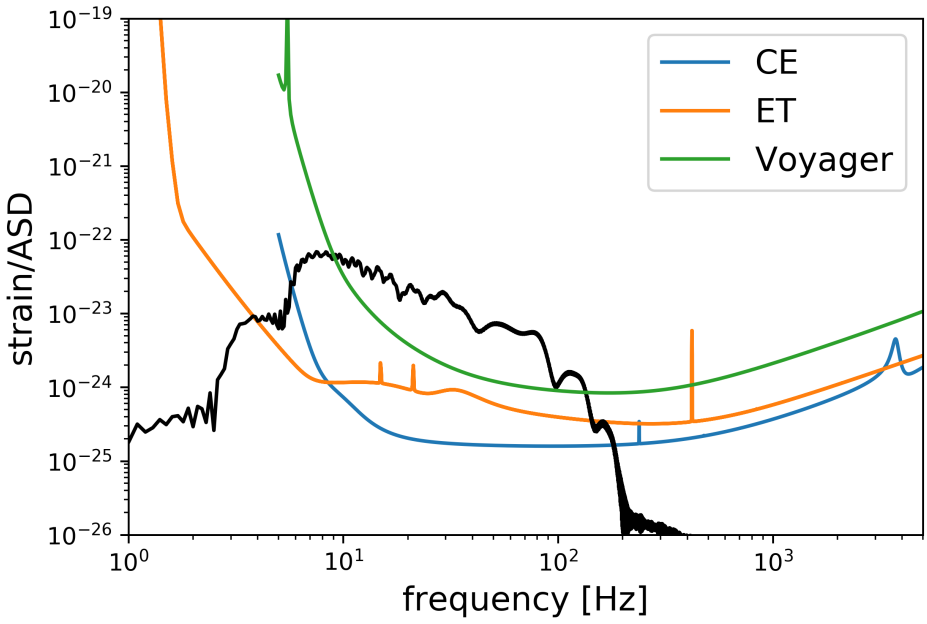

Figure 2.3: Positive frequency band waveform for a binary with $\tilde{M} = 200\ M_\odot$, $q = 2.2$, $\iota = 80$ deg and $D_L = 1$ Gpc, plotted in black. The amplitude spectral densities (ASDs) for CE, ET and LIGO Voyager are plotted with colored traces. Note that ASDs for CE and LIGO Voyager are available starting from 3 Hz.

less massive binaries, this effect results in significant loss of ρ, especially in the case of CE, where the low-frequency sensitivity degrades slower.

To estimate how this ρ loss affects the overlap values, we compute the maximum overlap for a ($\tilde{M}_0 = 60\ M_\odot, q = 3, \iota = 80$ deg) binary with the CE PSD, filtering all frequency components below 30 Hz, where the loss of ρ becomes significant. Compared with the unfiltered case ($(1-\epsilon)_{\max} = 0.981$), the overlap decreases only by 3.3×10^{-3}. Since $\tilde{M} = 60\ M_\odot$, $q = 3$ and $\iota = 80$ deg are roughly the smallest redshifted mass, largest mass ratio and inclination we consider, other binaries within our parameter space should have a smaller loss of the overlap. Considering the small size of the difference, we do not filter signals in subsequent analysis.

Figure 2.4 shows the best-match overlap for GW waveforms with a redshifted mass of 150 $M_\odot$ for the three GW detectors at selected mass ratio values. Maximization data points are shown with solid dots, and the interpolation functions are shown as smooth curves. The right axis shows the required ρ to distinguish type I/II waveforms with a log Bayes factor of 10 at the corresponding overlap values on the left axis. See discussion in Section 2.4. Consistent with intuition, the best-match overlap is the lowest for high mass-ratio signals at large inclinations. Over our parameter space, the mismatch value for such signals is typically on the order of 2%. We note that the same type II waveforms have the largest mismatch with type I waveforms in LIGO Voyager, as the LIGO Voyager PSD emphasizes high-frequency waveform components, where the Hilbert transform effect is more pronounced.

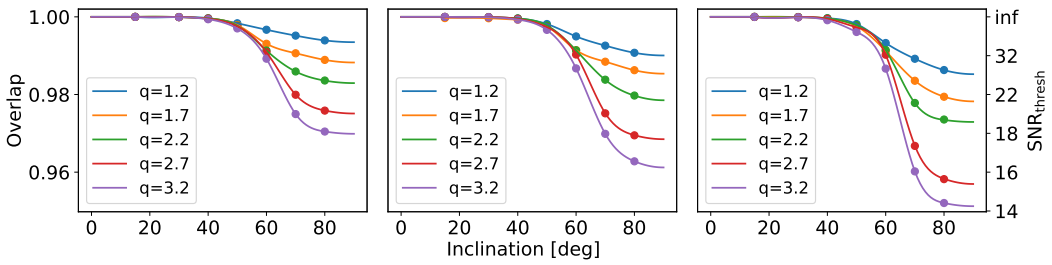

Figure 2.4: Overlap between type I and type II waveforms for $\tilde{M} = 150\ M_\odot$ at selected mass ratio values. The axis on the right shows the threshold ρ to distinguish such type II images from type I counterparts by a log Bayes factor of 10, for the corresponding waveform overlap value. See Section 2.4 for details. Panels from left to right are overlaps for CE, ET and LIGO Voyager respectively. In all panels, data points are shown with dots, and the interpolated overlap functions are shown in smooth curves.

Signal Veto

An ensuing concern from the mismatch is whether the difference in waveforms could lead to type II signals vetoed or assigned a lower significance value during observing runs. For the current GW data analysis pipelines, once a threshold ρ is reached, the data typically go through a χ^2 veto test to screen out spurious signals. In this section, we calculate the non-central parameter in the χ^2 statistic distribution from using type I templates to match type II signals.

The χ^2 veto was described in detail in Ref. [8]. This test characterizes the distribution of ρ over frequency bins and vetoes detector "glitches", or loud bursts of non-Gaussian noise that might have a high ρ, but have a frequency distribution very different from that of a genuine GW signal.

Suppose the best-match template to the signal, $\tilde{n} + \tilde{h}_0$ is $\tilde{h}_T$, where $\tilde{n}$ is noise and $\tilde{h}_0$ is the embedded waveform. We divide the detector sensitive frequency range into p disjoint sub-bands, Δf_j, such that the template ρ in each bin is $1/p$ of its total ρ,

$$\rho_{T,j} = \int_{-\Delta f_j, \Delta f_j} \frac{|\tilde{h}_T|^2}{S_n(f)} df = \frac{1}{p} \int_{\infty}^{\infty} \frac{|\tilde{h}_T|^2}{S_n(f)} df \ . \tag{2.12}$$

We then calculate the signal ρ in each frequency bin as

$$s_j \equiv \frac{1}{\sqrt{\langle \tilde{h}_T | \tilde{h}_T \rangle}} \int_{-\Delta f_j, \Delta f_j} \frac{\tilde{h}_T^*(\tilde{n} + \tilde{h}_0)}{S_n(f)} df \ . \tag{2.13}$$

We then define the χ^2statistic as

$$\chi^2 \equiv p \sum_{i=1}^{p} (s_i - s/p)^2, \ s \equiv \sum_{j=1}^{p} s_j \ . \tag{2.14}$$

In the case where the best-match template in the template bank does not exactly match the embedded waveform, the distribution of χ^2 over many Gaussian noise realizations is a classical χ^2 distribution with a non-central parameter,

$$\langle \chi^2 \rangle = p - 1 + \kappa \langle s \rangle^2 \ , \tag{2.15}$$

where $\langle \cdot \rangle$ denotes the average over noise realizations. The factor, κ, in the non-central parameter is bound by

$$0 < \kappa < \frac{1}{(1-\epsilon)^2} - 1 \approx 2\epsilon \ , \tag{2.16}$$

where ϵ is the minimized mismatch between the template and the underlying waveform, as is defined in Eq. 2.11. The approximate equality is satisfied when $\epsilon \ll 1$. This bound is agnostic of the specific waveform of the signal and templates. Consequently, the non-central parameter introduced by using type I templates on type II signals is smaller than $0.12\langle s\rangle^2$ in most cases, if we take the largest mismatch to be 6%. If such a non-central parameter lies within the χ^2 threshold during detection, type II images are unlikely to be vetoed.

type II Signal Recovery

There have been ongoing efforts to look for possible weaker (sub-threshold) strongly lensed counterparts of confirmed GW detections, assuming the latter being strongly lensed signals themselves [42, 47]. One method is to simulate lensed injections of a super-threshold GW event, then use a generic template bank to search for these injections through an injection run, and produce a targeted template bank for searching possible lensed counterparts of the target event by retaining only templates that can find the injections.

However, only type I lensed images have been considered for current searches. The question we would like to investigate is: Should type II lensed images be present in the data, would a type I template bank be able to find them? The answer to this question may be a crucial step for us to identify possible lensed GWs that we might have already detected but still not being discovered.

As a preliminary test to this question, we apply the search method detailed in Ref. [42] to the high-mass-ratio compact binary coalescence event GW190814 [6]. Using the waveform approximant `IMRPhenomXPHM` [60], we generate a set of simulated lensed injections for GW190814. They are then injected into real LIGO-Virgo data in two ways: (1) by treating them as type I images, and (2) by treating them as type II images, i.e. applying Hilbert transform to the waveform in the frequency domain as discussed previously. Through the GW CBC search pipeline GstLAL [49], we apply the previously generated type I image target bank to search for these injections in both tests, and finally we compare the number of missed injections to roughly estimate the effectiveness of a type I image bank to look for type II images.

As discussed in Ref. [42, 49], each GW candidate found in the GstLAL search will be assigned a log likelihood ratio statistic $\ln \mathcal{L}$ to measure its significance. The False-Alarm-Rate (FAR) can be calculated accordingly, which corresponds to how often noise will produce a trigger with ranking statistic $\ln \mathcal{L}$ larger or equal to the

ranking statistic $\ln \mathcal{L}^*$ of the trigger we are considering. In the search, an injection is said to be found if its FAR passes the usual threshold 1 in 30 days, as usual for a generic gravitational-wave search [4].

In both tests, we have injected a total of 8036 simulated lensed injections. We assume that the injected events are registered by both detectors in the aLIGO network and the Virgo detector. In test **A**, we apply a type I image bank to look for injected type II images. For test **B**, we use the same image bank and look for the type I counterpart of the injections in test **A**. In test **A**, 638 injections are missed, whereas in test **B** the missed count is 536. We observe that the number of missed injections increases when the injections were treated as type II images, indicating that the current search method for sub-threshold lensed GWs may be missing possible type II lensed signals.

However, it is important to remark that our current results are inconclusive since: (1) we have only been testing on one particular GW event, and (2) the exact reason for the extra number of injections to be missed are yet to be investigated. Nevertheless, our results indicate there could be improvements to the current search method for sub-threshold lensed GW signals, and further investigation will be done as future work.

2.4 Distinguishing Type II Events

While we have systematically examined the type I/II waveform mismatch, whether it enables us to distinguish type II images in actual GW experiments deserves further discussion. In this section, we use the waveform overlap and quantify the fraction of strongly lensed GW sources that have distinguishable type II images.

Bayes Factor

Using a Bayesian model (or equivalently hypothesis) selection framework, we quantify the distinguishability between a type I image and a type II image by computing the Bayes factor $\mathcal{B}$, which is the ratio of the probability of observing the data $\vec{d}$ under the hypothesis that the signal is of type II over that under the hypothesis that the signal is of type I, namely

$$\begin{aligned}\mathcal{B} &= \frac{p(\vec{d}|\text{ type II image})}{p(\vec{d}|\text{ type I image})} \\ &= \frac{\int d\vec{\theta}\ \mathcal{L}(\vec{\theta}|\text{ type II image})\pi(\vec{\theta}|\text{ type II image})}{\int d\vec{\theta}\ \mathcal{L}(\vec{\theta}|\text{ type I image})\pi(\vec{\theta}|\text{ type I image})},\end{aligned} \tag{2.17}$$

where $\mathcal{L}(\vec{\theta})$ is the (Whittle) likelihood as a function of the waveform parameters $\vec{\theta}$, and $\pi(\vec{\theta})$ is the prior distribution, which is different under the two hypotheses. The log likelihood function, up to a normalization constant, is given by

$$\begin{aligned} \ln \mathcal{L}_i(\vec{\theta}) &\propto -\frac{1}{2}\langle d - h_i(\vec{\theta}) | d - h_i(\vec{\theta})\rangle \\ &\propto -\frac{1}{2}\langle d|d\rangle + \langle d|h_i\rangle - \frac{1}{2}\langle h_i|h_i\rangle, \end{aligned} \tag{2.18}$$

where the subscript $i = \mathrm{I}, \mathrm{II}$ denotes the assumed image type. In an actual inference analysis, we do not know a priori the 'true' waveform parameters. Therefore, we usually evaluate the integrals in Eq. (2.17) using a sampling algorithm that explores the parameter space spanned by $\vec{\theta}$ stochastically.

Still, we can give an analytical approximate of the Bayes factor for distinguishing a type II image from a type I image using only the SNR ρ and the mismatch ϵ we calculated in Sec. 2.3. Following the treatment in Ref. [20, 25, 72], with the Laplace approximation we can write the log Bayes factor as

$$\ln \mathcal{B} \approx \ln \left[\frac{\mathcal{L}_{\mathrm{II}}(\vec{\theta}_{\mathrm{MLE}})}{\mathcal{L}_{\mathrm{I}}(\vec{\theta}_{\mathrm{MLE}})}\right] + \ln\left(\frac{\sigma_{\mathrm{II}}^{\mathrm{posterior}}}{\sigma_{\mathrm{I}}^{\mathrm{posterior}}}\right), \tag{2.19}$$

where $\sigma_i^{\mathrm{posterior}}$ is the posterior (uncertainty) volume assuming that the lensed GW is of type-i. The log likelihood ratio in Eq. (2.19) can be shown [25], in the high SNR limit, to scale as

$$\ln \left[\frac{\mathcal{L}_{\mathrm{II}}(\vec{\theta}_{\mathrm{MLE}})}{\mathcal{L}_{\mathrm{I}}(\vec{\theta}_{\mathrm{MLE}})}\right] \approx \epsilon\, \rho^2, \tag{2.20}$$

when the (minimized) mismatch $\epsilon \ll 1$. If we ignore the correlation between the parameters, we can estimate the posterior volume $\sigma_i^{\mathrm{posterior}}$ roughly as

$$\sigma_i^{\mathrm{posterior}} \approx \prod_{j=1}^{N} \sqrt{2\pi}\Delta\theta_i^{j,\,\mathrm{posterior}}, \tag{2.21}$$

with j loops over the N-dimensional vector $\vec{\theta}$ and $\Delta\theta_i^{j,\,\mathrm{posterior}}$ is the uncertainty of the 1D marginal posterior distribution for θ_j assuming that the image is of type-i. Note that here we assumed that identical prior was used when calculating the Bayes factor, except for the image type. The posterior volume ratio also scales with the the mismatch, actually. Since $\langle h_{\mathrm{I}}|h_{\mathrm{I}}\rangle \approx (1-\epsilon)^2\ \langle h_{\mathrm{II}}|h_{\mathrm{II}}\rangle$ and that $\Delta\theta_i^{j,\,\mathrm{posterior}} \propto 1/\sqrt{\langle h_i|h_i\rangle}$, therefore we have

$$\ln\left(\frac{\sigma_{\mathrm{II}}^{\mathrm{posterior}}}{\sigma_{\mathrm{I}}^{\mathrm{posterior}}}\right) \approx -N\epsilon. \tag{2.22}$$

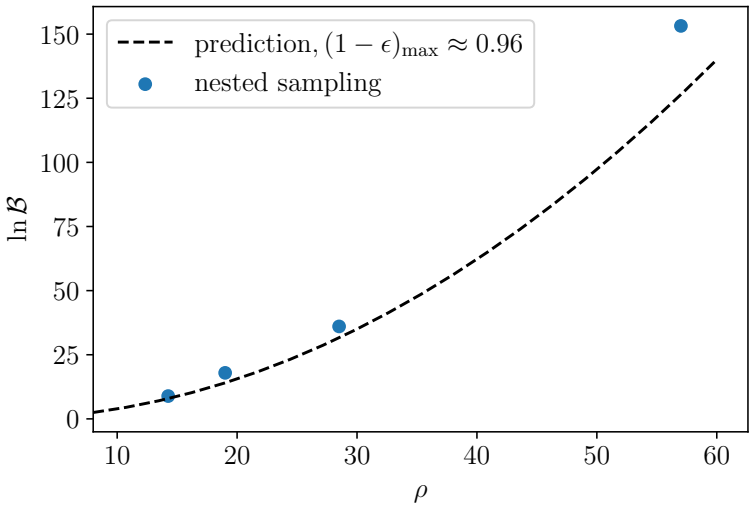

Figure 2.5: The log Bayes factor $\ln \mathcal{B}$ as a function of the SNR ρ of the injections with different luminosity distances and fixed mismatch ϵ, computed using Eq. (2.17) with nested sampling and Eq. (2.23). We see that the simulation results roughly follow the expected quadratic scaling with the SNR.

Indeed, in the high SNR limit, the first term in Eq. (2.19) is much larger than the second term as $N \sim 10$ and $\epsilon \ll 1$. Hence, we will ignore the contribution from the log posterior volume ratio in this paper. Therefore, we can estimate the log Bayes factor simply as [5]

$$\ln \mathcal{B} \approx \epsilon\, \rho^2. \tag{2.23}$$

Figure 2.5 shows the log Bayes factor as a function of the SNR ρ using nested sampling with the help of the library `bilby` [9] and `dynesty` [66] as in Eq. (2.17), as well as its approximate using only the optimal SNR and the mismatch using Eq. (2.23). Here we use the `IMRPhenomXHM` waveform model [34] for both the simulated signals and the inference. All simulated signals have a redshifted total mass of $\tilde{M} = 150 M_\odot, q = 3.2$ viewing at an inclination angle of $\iota = 80\,\text{deg}$ with different luminosity distances to adjust the optimal SNR. We see that the simulation results roughly follow the expected quadratic scaling with the optimal SNR. Indeed, by performing a least-squares fit we found that the exponent is 2.02 ± 0.07.

Since $\ln \mathcal{B}$ scales as SNR^2, even a small type I/II mismatch could lead to significant $\ln \mathcal{B}$ in the high-SNR regime. For instance, for a mismatch of 3%, an SNR of 20 would yield a log Bayes factor larger than 10, favoring the type II waveform

[5] Note that posterior volume also depends on dependences of h_I and h_II on θ_j

hypothesis, thereby identifying this event as a strongly lensed image regardless whether other images are detected. The right axis in Figure 2.4 shows the required SNR to produce $\ln \mathcal{B} = 10$ for the corresponding type I/II overlap values. While such SNR is high for the current aLIGO, for third-generation detectors, it occurs frequently. For example, an equal-mass binary with a detector frame total mass of 100 $M_\odot$ at $D_L = 8$ Gpc has an $\rho = 30$ for LIGO Voyager. The same source with $D_L = 17$ Gpc has $\rho = 131$ for CE.

Threshold Inclination

In this section, we find the range of parameters, $(\tilde{M}, q, z_s)$, where type II images can be distinguished via the log Bayes factor test. We choose $\ln \mathcal{B}_{\rm thresh} = 10$ as the criterion for distinguishability.

We begin by computing the distinguishable threshold inclination, ι, for sources with certain redshifted mass, mass ratio and redshift. Since both ρ and ϵ in Eq. (2.23) depend on ι, it is more straightforward to first fix $\tilde{M}, q$ and ι to obtain ϵ, and then scale ρ via D_L to achieve the $\ln \mathcal{B}_{\rm thresh}$ condition. Inverting $D_{L,\rm thresh}(\tilde{M}, q, \iota)$ yields $\iota_{\rm thresh}(\tilde{M}, q, D_L(z_s))$, where z_s is the GW source redshift.

To calculate ρ, we assume both GW polarizations can be detected, and the total amplitude is $\sqrt{h_+^2 + h_\times^2}$. In the Discussion Section, we further discuss the justifications for this assumption in the context of third-generation GW detectors. However, the finite length of the surrogate model waveform can lead to significant loss in ρ, even though the effect on waveform overlap is negligible, as demonstrated in Section 2.3. For a binary with $\tilde{M} = 60\ M_\odot, q = 3$, approximately 15% of ρ is lost in the case of CE. For LIGO Voyager, the noise increase starts earlier and steeper towards lower frequencies; consequently, the ρ loss for the same binary is only $\sim 5\%$. To accurately estimate ρ, we supplement the surrogate model waveform with analytical inspiral stage waveform, whose amplitude scales as $f^{-7/6}$ [21]. The inspiral amplitude is matched to the surrogate waveform amplitude at 0.5 $f_{\rm ISCO}$, where $f_{\rm ISCO}$ is the Innermost Stable Circular Orbit frequency, approximated as [see, e.g., 65]

$$f_{\rm ISCO} = \frac{c^3}{6^{3/2}\pi G \tilde{M}} \ . \tag{2.24}$$

We note that, by compensating for the lost ρ, our result is optimistic in estimating the distinguishability; while the early inspiral phase contributes significantly to ρ, the type I/II waveform mismatch is less pronounced. For the same ρ, the compensated inspiral waveform does not offer as much information as the higher frequency GW

phases for distinguishing type II images. Nonetheless, this overestimate is significant only for systems towards the low mass limit, where the expected detectable number of events is low due to the small ρ.

The mismatch ϵ is available from the interpolation function in Section 2.3. We do not consider binaries with best-match overlap larger than 0.999, i.e., we consider such mismatch a result of systematic errors and does not reflect actual waveform difference. As discussed in Section 2.3, the truncated surrogate waveform leads to errors in the best-match overlap, though for high mass systems, the error will be much smaller than 3.3×10^{-3} for the $\tilde{M} = 60\ M_\odot, q = 3, \iota = 80$ deg example binary. For computational cost concerns, we also limit the grid density in the nested maximization process. If the actual best-match binary is not on the grid points, the maximization result will deviate from the true value, and the size of the deviation depends on the distance between the true best-match and its closest grid point. Aside from systematic errors in the waveform and overlap optimization process, interpolation for ϵ also introduces errors. In particular, the cubic spline fit may introduce spurious trace curves to guarantee smoothness when connecting the limited number of samples. Especially in the case of CE, ρ can be very large, thus exaggerating the physical significance of such a small mismatch. The exact value of this threshold is tuned to exclude spurious interpolation function results. In the next subsection, we discuss our choice of the upper limit value for the best-match overlap and assess the impact of this mismatch resolution in the next subsection.

Figure 2.6 shows the threshold inclination as a function of source redshift assuming CE sensitivity. The left panel shows threshold inclination with fixed redshifted mass $\tilde{M} = 150\ M_\odot$. We observe that the mass ratio becomes an increasingly important factor at high inclinations. At low redshift, the threshold inclination is constrained primarily by the mismatch ϵ; at higher redshift (e.g., $z_s \sim 4.2$ for $q = 1.73$), the high inclination regions start to be excluded despite the large mismatch value, as ρ becomes too small. Beyond a certain redshift (e.g., $z_s \sim 5.2$ for $q = 1.73$), no combination of ρ and ϵ meets the $\ln \mathcal{B}_{\rm threshold}$ condition, and no more type II images can be distinguishable. The right panel shows similar threshold cures fixing the mass ratio to be 2.67. We observe a similar curve shape, although lighter binaries have smaller ρ and consequently a larger threshold inclination.

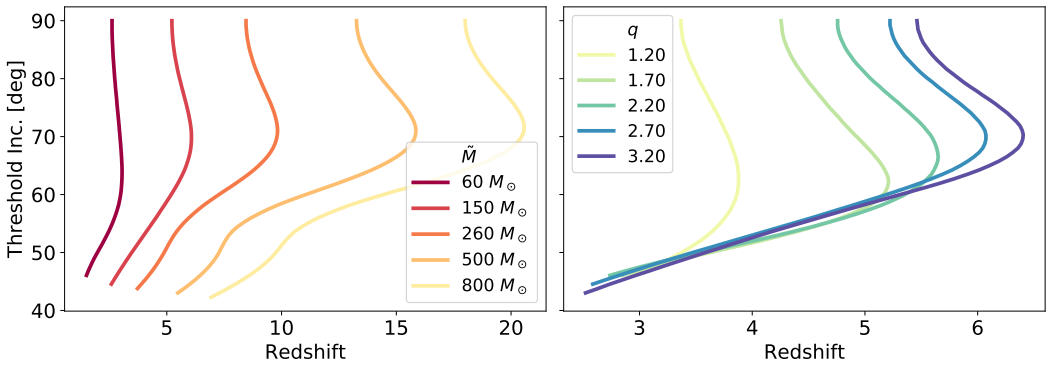

Figure 2.6: Inclination threshold curves for distinguishable type II sources as a function of redshift assuming CE sensitivity. *Left:* inclination threshold curves for fixed mass ratio $q = 2.67$ at selected redshifted mass values. *Right:* inclination threshold curves for binaries with redshifted mass $\tilde{M} = 150\ M_\odot$ with selected mass ratio values. The curve-crossing at low inclination values is due to systematic errors; see text for discussion.

Distinguishable Image Fraction

From the threshold inclination, we can further calculate the fraction of GW sources with distinguishable type II images, $\mathrm{fr}(\tilde{M}, q, z_s)$. For simplicity, we assume that GW sources and type II images are isotropically distributed, therefore the fraction of distinguishable type II images scales as the area of the celestial sphere within the ι threshold limits. The differential fraction is then proportional to $\sin\iota$. Figure 2.7 shows the distinguishable fraction of type II sources for the same binaries as in Figure 2.6. The cusps mark the redshift when high inclination regions start to be excluded due to smaller ρ. We observe that, for CE, large fractions of sources with type II images can be distinguished via the log Bayes factor test out to high redshift. Similar plots for ET and LIGO Voyager are shown in Figure 2.8 as dashed lines.

We have so far considered only the redshifted mass (detector-frame mass), $\tilde{M}$, as it is the direct input to the surrogate model, which assumes an asymptotically flat and stationary universe. The "apparent" total mass of the binary, M, is related to the redshifted mass by $M = \tilde{M}/(1+z_s)$. Due to lensing magnification, this inferred "apparent" total mass could be larger or smaller than the actual GW source total mass. We discuss magnification effects in Section 2.5. Therefore, the fraction of distinguishable type II sources with apparent mass M, mass ratio q at redshift z_s is given by

$$\mathrm{fr}_{\mathrm{app}}(M, q, z_s) = \mathrm{fr}(M(1+z_s), q, z_s)\ . \tag{2.25}$$

Figure 2.8 shows $\mathrm{fr}_{\mathrm{app}}(M, q, z_s)$ for selected apparent mass values in solid traces.

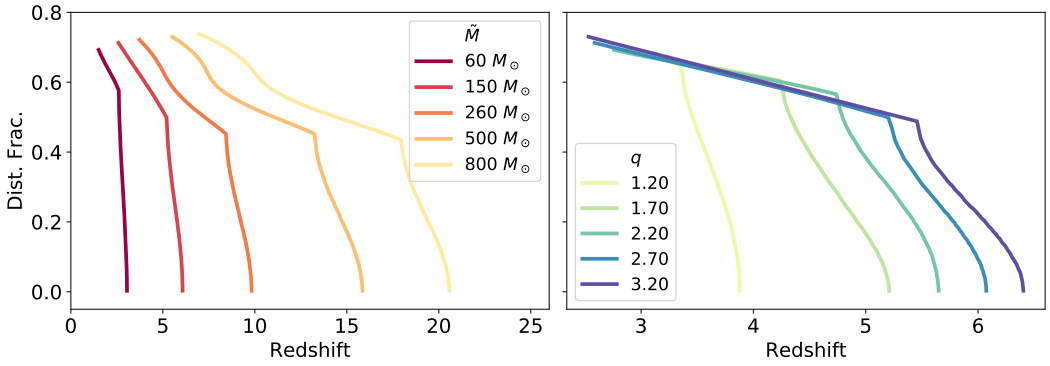

Figure 2.7: The fraction of distinguishable type II images as a function of redshift for CE sensitivity. *Right:* distinguishable fraction, $\mathrm{fr}(\tilde{M}, q, z_s)$, for constant $\tilde{M} = 150\ M_\odot$. *Left:* distinguishable fraction for constant $q = 2.67$. The cusps in the fraction correspond to the exclusion of high-inclination binaries with sub-threshold ρ. The fraction curves directly correspond to the threshold inclination curves in Figure 2.6.

The top row shows the fractions with a fixed mass ratio of 1.73, and the bottom row shows similar plots with mass ratio fixed at 2.67. The left, middle and right columns show results for CE, ET and LIGO Voyager, respectively. The distinguishable fractions for fixed redshifted mass, $\mathrm{fr}(\tilde{M}, q, z_s)$, are plotted for reference in dashed lines. The exact mass values are specified in the legend.

As expected, the resulting traces show similar trends and features as in Figure 2.7: at lower redshifts, the distinguishable fraction decreases with ρ. It then undergoes a cusp where the high inclination regions start to be excluded before continuing to decrease. The trace is jagged due to the finite spacing of the interpolation data points, rather than any physical jumps in the fraction.

In most cases, there is a significant fraction of GW sources with distinguishable type II images via the log Bayes factor test. As Figure 2.4 suggests, the mismatch value is not drastically different across the three detectors with different noise curve shapes. The redshift reach is rather primarily determined by ρ, related to the overall sensitivity level of different detectors. For example, for type II images with apparent mass $\tilde{M} = 60\ M_\odot$ and mass ratio $q = 1.76$, 60% can be distinguished in CE out to $z_s \sim 12.5$. For ET, 60% of the same population can be distinguished out to $z_s \sim 2$. Due to the lower sensitivity of LIGO Voyager, a similar fraction of such type II images can be identified only out to $z_s \sim 0.5$. However, for type II images with a higher apparent mass of $100\ M_\odot$, 50% can still be registered out to $z_s \sim 1$.

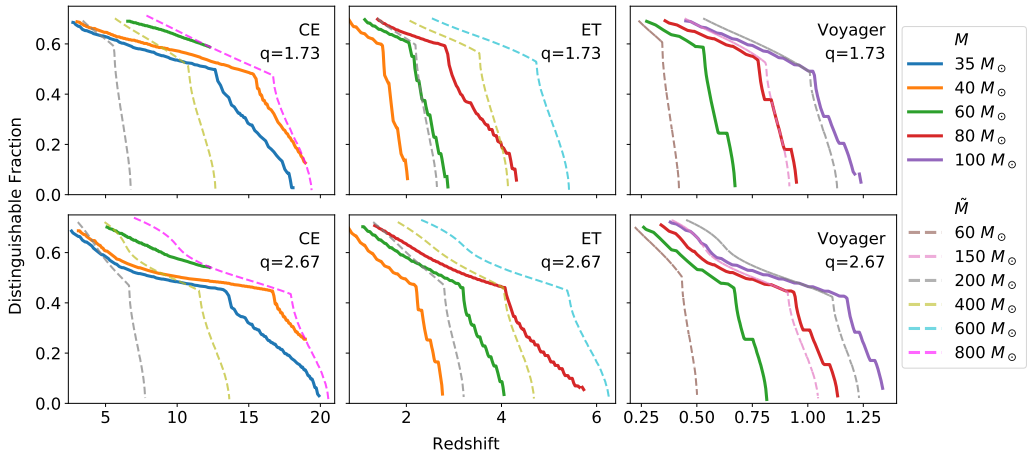

Figure 2.8: The fraction of distinguishable type II images as a function of redshift. *Top Row:* GW sources have constant mass ratio of 1.73. *Bottom Row:* GW sources have constant mass ratio of 2.67. The *Left, Middle,* and *Right* columns show distinguishable fractions assuming CE, ET and LIGO Voyager sensitivity, respectively. The fractions for fixed redshifted mass, $\mathrm{fr}(\tilde{M}, q, z_s)$, are shown in dashed lines, and those for fixed apparent mass, $\mathrm{fr}_{\mathrm{app}}(M, q, z_s)$, are shown in solid lines. The mass values are shown in the legend. Note that the fraction curves are jagged due to interpolation errors and limited data density.

Finally, we assess the impact of mismatch resolution. Throughout this paper, we adopt a minimum mismatch value of $\epsilon = 0.001$. Figure 2.9 shows the changes in the distinguishable fraction of GW sources with type II images for CE. The mass ratio is fixed to be $q = 1.73$, and the solid lines from left to right represent $\tilde{M} = 100\ M_\odot,\ 200\ M_\odot,\ 260\ M_\odot,\ 400\ M_\odot,\ 600\ M_\odot$ and $800\ M_\odot$. The dashed horizontal traces show the largest distinguishable fraction as a function of the redshifted mass.

As the waveform mismatch resolution becomes coarser, the distinguishable fraction decreases significantly. For instance, 70% of all sources with type II images with redshifted mass $m = 100\ M_\odot$ and $q = 1.73$ have distinguishable type II images out to redshift $z_s \sim 2.5$ if a mismatch of 0.001 is resolvable, but the fraction drops to 30% if the mismatch resolution is 0.007. With a mismatch resolution of 0.016, no such type II images are distinguishable. This critical role of the minimum resolvable mismatch suggests that the distinguishability of type II images does not solely depend on the SNR. In the era of third-generation GW detectors, the possible scientific output from GW detection events is not solely determined by the noise level. As is discussed, the waveform template bank density in the matched filtering search limits the waveform difference resolution. In addition, the detector calibration must also be sufficiently

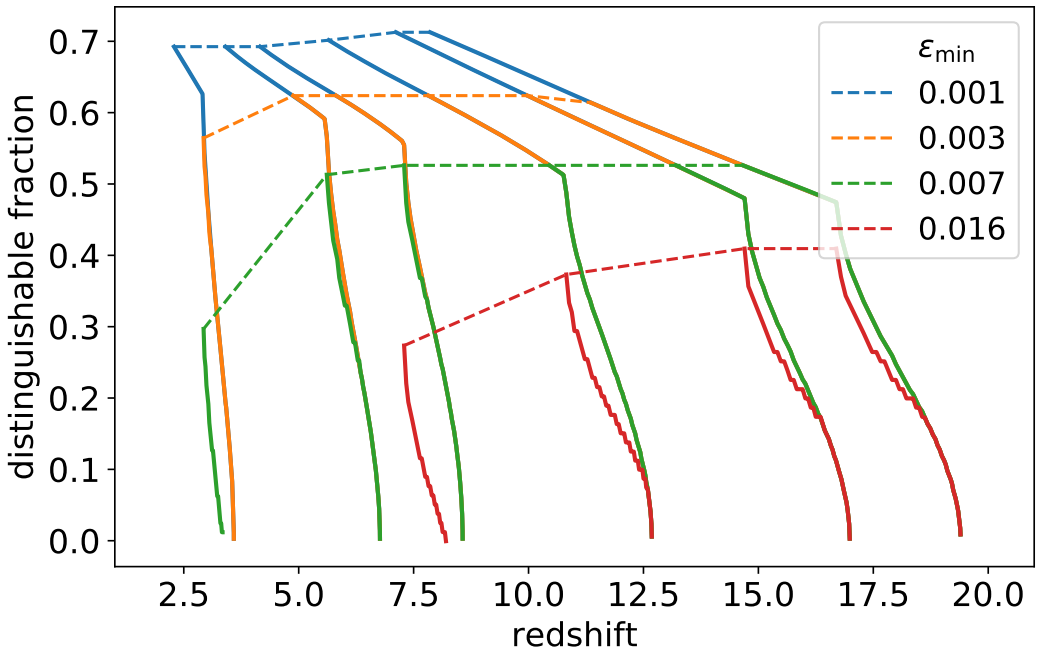

Figure 2.9: Distinguishable fractions of sources with type II images assuming a minimum resolvable mismatch of 0.001, 0.003, 0.007 and 0.016 with CE sensitivity. The solid lines show the distinguishable fractions for selected redshifted mass. The mass ratio is fixed to be 1.73. From left to right, the traces correspond to a redshifted mass of 100 $M_\odot$, 200 $M_\odot$, 260 $M_\odot$, 400 $M_\odot$, 600 $M_\odot$ and 800 $M_\odot$. Dashed lines show the maximum fraction for this range of redshifted mass. For a mismatch resolution of 0.016 (or 0.984 overlap), fraction traces for type II images with redshifted mass $\tilde{M} = 100, 200\ M_\odot$ (the leftmost two traces) are absent, since the resolution is larger than the maximum possible waveform mismatch for such GW sources.

accurate such that we can be confident that the mismatch from the data reflects a real signal difference, rather than an instrument systematic error. Otherwise, we cannot take full advantage of the large SNR offered by exquisite detector sensitivity. Since these factors during the third-generation GW detector era are still subject to much uncertainty, to our knowledge, we have chosen $\epsilon = 0.001$ as the fiducial value. The analysis should be refined as such information becomes available.

2.5 Detectable Population

The distinguishable fraction calculations depend only on the waveform mismatch, and do not assume astrophysical estimates on GW source and lens distributions. In this section, we describe how results of image distinguishability can be combined with astrophysical models to give a more detailed prediction of the detectable GW

events with type II images and those with distinguishable type II images. In the following subsections, we first offer an overview of the procedures for calculating the lensed population, followed by more detailed discussions on each ingredient.

Overview

We define a GW event as a particular binary black hole (BBH) merger with possibly multiple images due to strong lensing. For our calculation, a GW image is detectable if its single-detector $\rho \geq 8$; we defer the detector network scenario to future studies. A GW event has a distinguishable type II image if this image satisfy the log Bayes factor threshold. The differential detectable and distinguishable merger rate is given by

$$\frac{\partial^3 \dot{N}_{\rm II,det}}{\partial M_\bullet \partial q \partial z_s} = \tau_{\rm II}(z_s) \frac{\partial^3 \dot{N}}{\partial M_\bullet \partial q \partial z_s} \int d\log_{10}\mu \left[\frac{\partial P_{\rm II}(\mu, z_s)}{\partial \log_{10}\mu} \int_0^{\pi/2} d\iota \ \sin\iota \, \Theta\left(\sqrt{\mu}\,\rho(M_\bullet, q, z_s, \iota) - 8\right) \right] , \tag{2.26}$$

$$\frac{\partial^3 \dot{N}_{\rm II,dis}}{\partial M_\bullet \partial q \partial z_s} = \tau_{\rm II}(z_s) \frac{\partial^3 \dot{N}}{\partial M_\bullet \partial q \partial z_s} \int d\log_{10}\mu \, \frac{\partial P_{\rm II}(\mu, z_s)}{\partial \log_{10}\mu} {\rm fr}\left(\tilde{M}, q, \tilde{z}_s\right) , \tag{2.27}$$

where $\partial^3 \dot{N}/\partial M_\bullet \partial q \partial z_s$ is GW event rate per intrinsic binary mass, $M_\bullet$, mass ratio, q, and GW source redshift, z_s, measured in the observer frame. The weighting factor $\sin\iota$ comes from the assumption that BBH mergers are distributed evenly on the sky. Θ is the Heaviside function. Multiplying with the optical depth $\tau_{\rm II}(z_s)$, we obtain the rate of events with at least one type II images. The quantity $\partial P_{\rm II}(\mu, z_s)/\partial \log_{10}\mu$ describes the distribution of magnification μ for type II images for sources at z_s, normalized such that

$$\int d\log_{10}\mu \, \frac{\partial P_{\rm II}(\mu, z_s)}{\partial \log_{10}\mu} = 1 \ . \tag{2.28}$$

Due to magnification, the source appears to have the same redshifted mass, but the inferred luminosity distance is different. Therefore,

$$\begin{aligned} \tilde{M} &= M_\bullet (1 + z_s) \\ D_L(\tilde{z}_s) &= D_L(z_s)/\sqrt{\mu} \ . \end{aligned} \tag{2.29}$$

The differential merger number per observer time is calculated as [see also 23]

$$\frac{\partial^3 \dot{N}}{\partial M_\bullet \partial q \partial z_s} = R_{\rm mrg}(M_\bullet, q, z_s) \frac{1}{1 + z_s} \frac{dV_c}{dz_s} , \tag{2.30}$$

where dV_c/dz_s is the differential comoving volume. The $1/(1 + z_s)$ factor accounts for the cosmological redshift and converts the source-frame merger rates into

detector-frame merger rates. In Figure 2.12, we plot this "modified" differential comoving volume and the total merger rates for reference. Since we have adopted the same population models, Figure 2.12 replicates Figure 1 in Ref. [43].

For fast calculation of $\rho(M_\bullet, q, z_s, \iota)$, we use the phenomenological model `IMRPhenomHM` [40], called from the Python package `pycbc.waveform` [56].

We note that for a type I/II waveform mismatch of 6%, the required ρ to be distinguishable is approximately 13, larger than the threshold SNR of 8, and none of the GW sources we consider have a larger waveform mismatch. Therefore, we may assume that the *distinguishable* images are all *detectable*, leading to the omission of the Heaviside function in Eq. (2.27). In addition, only 0.2% of all sample lens systems have a brighter type II image than the type I image. Considering errors from the lens-equation solution algorithm and the small number of events with distinguishable type II images, we may assume that the events with detectable or distinguishable type II images will most certainly have a detectable type I companion image.

In the following subsections, we compute the type II image optical depth and the magnification distribution. We then summarize procedures to calculate the total BBH merger rates. Detailed steps and adopted parameter values are presented in Appendix 2.7. We then make concrete detection population for CE, ET and LIGO Voyager and discuss results.

Optical Depth and Magnification

To obtain $\tau_{\rm II}(z_s)$ and $\partial P_{\rm II}(\mu, z_s)/\partial \log_{10}\mu$, we perform a Monte Carlo simulation. We consider elliptical galaxies as lenses, as they are expected to be the predominant lensing objects [57]. While the lens geometry and properties are expected to be more complex and varied in nature, studies showed that a simple lens model, such as the singular isothermal sphere model, is sufficient to capture most of the results from more sophisticated hydrodynamic simulations of the Universe [62]. In this study, we adopt the slightly more generalized singular isothermal ellipsoid model following the examples of Refs. [43, 57]. Refs. [11, 27] adopt the singular isothermal sphere lens model and predict the detectable strongly lensed events for ET. For one of the BBH evolutionary scenarios they investigate, it is predicted that 57.2 strongly lensed events can be detected out of the 2.08×10^5 total detectable BBH events per year, roughly a factor of three smaller than our prediction (see Table 2.1). In the future, we can adapt our analysis using different lens models and systematically study the

uncertainty in the strong-lensing population predictions.

We restrict the GW source redshift to $0.05 \leq z_s \leq 7$. In the low-redshift limit, GW sources in our local universe ($z_s \ll 0.05$) is unlikely to be strongly lensed, since lensing rates are expected to be low, and there are not sufficiently many massive galaxies in between to compensate. We set the upper limit of the galaxy redshift to $z_s = 7$, since such galaxies are faint and robust observational data is relatively scarce for developing a reliable phenomenological model of the mass function [33].

At each redshift, we generate samples of lenses, parameterized by surface velocity dispersion, σ_v, ellipticity, e, lens redshift, z_l, and the lens-plane angular coordinates of the lens, $\vec{\theta} = (x, y)$. For the number of lenses per unit σ_v per comoving volume, $\Psi(\sigma_v, z_l)$, we first adopt the modified Schechter function [18], which is calibrated to observation on galaxies in the solar neighborhood,

$$\Psi(\sigma_v, 0) = \phi_* \left(\frac{\sigma_v}{\sigma_*}\right)^{\alpha} \exp\left[-\left(\frac{\sigma_v}{\sigma_*}\right)^{\beta}\right] \frac{\beta}{\sigma_v \Gamma\left(\alpha/\beta\right)} , \tag{2.31}$$

where $\phi_* = 8.0 \times 10^{-3} h^3$ Mpc^{-3}, $\sigma_* = 161$ km/s, $\alpha = 2.32$ and $\beta = 2.67$. h is the Hubble parameter.

To account for the redshift dependence, we follow the prescription in Ref. [57], in which

$$\Psi(\sigma_v, z_l) = \Psi(\sigma_v, 0) \frac{\Psi_{\rm hyd}(\sigma_v, z_l)}{\Psi_{\rm hyd}(\sigma_v, 0)} , \tag{2.32}$$

where $\Psi_{\rm hyd}(\sigma_v, z_l)$ is the velocity dispersion function derived from hydrodynamical simulation in Ref. [71]. The redshift depedence of the galaxy comoving number density is shown on the right axis of Figure 2.10. We truncate σ_v at 50 km/s and 400 km/s to include the major part of the distribution in Eqn. (2.31). In general, the galaxy number density peaks around $z \sim 1, 2$ and decreases towards higher redshift as they have less time to form.

For galaxy ellipticity, we adopt the same Gaussian distribution as in Ref. [43], where the mean and standard deviation are 0.7 and 0.16, truncated at $e = 0.2$ and $e = 1$. The lens redshift is uniformly sampled from $[0, z_s]$. Since strong lensing occurs only when the angular separation between the lens and the source is small, we uniformly sample the lens positions within a square region centered at the source with the side length equal to four times the Einstein radius of the lens, given by [see 43]

$$\theta_E = 4\pi \left(\frac{\sigma_v}{c}\right)^2 \frac{D_{ls}}{D_s} , \tag{2.33}$$

where D_{ls} and D_s are the lens-source and observer-source separations, respectively.

For each sampled lens parameter set $(\sigma_v, e, z_l, \vec{\theta})$, we solve the lens equation with the Python package `lenstronomy`[6] [12] and obtain the number of images, image types and magnifications. Since our interest in distinguishable type II images is to identify strongly lensed GW sources, we compute the "source-based" optical depth, the fraction of GW sources with type II images, rather than the fraction of all images that are type II. Each sample with at least one type II image contributes to $\tau_{\rm II}$, while depending on the solution for μ, it contributes to $\partial P_{\rm II}/\partial \log_{10}\mu$ accordingly. To account for the lens population, each sample is weighted by the expected count of such a galaxy within the defined lens position range.

Figure 2.10 shows the type II image optical depth at various source redshifts on the left axis. Optical depths smaller than $\sim 10^{-5}$ are truncated, as they are too low to produce a possible lensed source. We observe that the optical depth is on the order of $10^{-3} \sim 10^{-4}$, consistent with results from ray-tracing studies using N-body simulations [see, e.g., 36, 37].

In the generated sample, the probability of a strongly lensed GW event (i.e., with multiple images) to have no type II images is smaller than 0.01% and therefore negligible. We conclude that the type II image optical depth is effectively identical to the strong lensing optical depth. We calculate that roughly 91.5% of all sources with multiple images have a type II image as the second "brightest" image, which suggests that if multiple images were to be detected, it is likely that at least one of the images may be a candidate for type II image distinction via the log Bayes factor test.

For larger redshifts, we do not extrapolate optical depth due to the lack of information on extremely high redshift galaxy velocity dispersion function from hydrodynamical simulations. Instead, we take the conservative limit and assume the optical depth to be constant beyond $z_s = 7$.

For the magnification distribution, we extract the type II images from the Monte Carlo simulation samples for each redshift. Figure 2.11 shows the rescaled image magnification distribution per $\log_{10}\mu$ at selected redshifts $z_s = 0.5, 0.8, 2, 6$. The *left* panel includes all images with the peak dominated by the slightly magnified

[6] `https://github.com/sibirrer/lenstronomy`

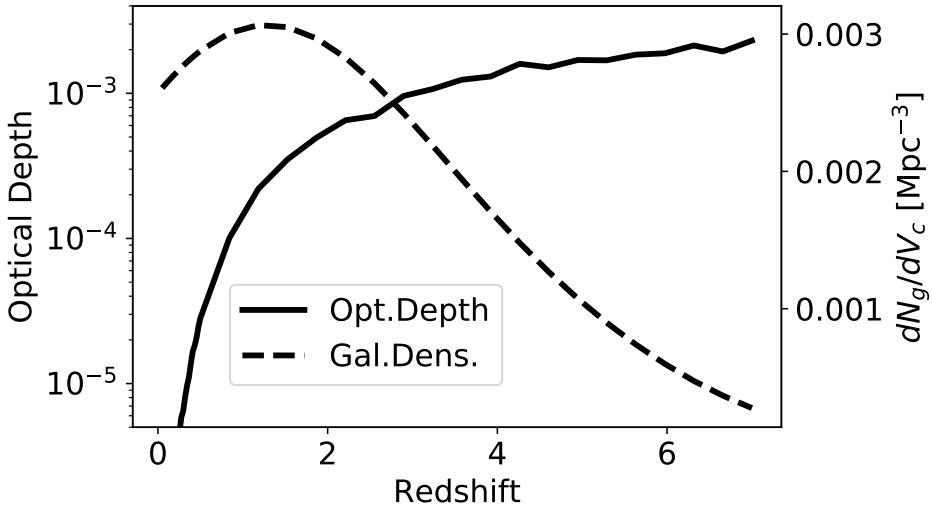

Figure 2.10: *Left* axis: optical depths, $\tau_{\rm II}(z_s)$, for GW sources with at least one type II image as a function of source redshift. Optical depths lower than $\sim 10^{-5}$ are omitted, as they are too low to predict an observable GW source at such redshifts with type II images in future detectors. *Right* axis: comoving number density of all galaxies modeled as lenses.

type I images. The *right* panel contains only type II images, which constitute the demagnified image population. The rescaling normalizes the highest image count in each case to 1. Since the magnification for all images peaks around 1, we ignore it when calculating the detectable strongly lensed GW events; instead, the detection rate of all strongly lensed events can be estimated by multiplying the detectable BBH merger rate under the no-lensing hypothesis by the strong lensing optical depth, which, as the Monte Carlo samples show, is effectively identical to the type II image optical depth.

We note that $\partial P_{\rm II}(\mu, z_s)/\partial \log_{10}\mu$ is independent from source redshift z_s by construction, as Figure 2.11 confirms. To explain this feature, we first note that the lens equation solution depends only on $\vec{\theta}/\theta_E$, where $\vec{\theta} = (x, y)$ and θ_E is defined in Eq. (2.33). Since the range of the possible lens angular positions, $\vec{\theta}$, is directly determined by θ_E, the image solution (image count, magnification, etc.) and its distribution remain constant under the scaling. The only remaining redshift-dependent quantity is the galaxy velocity dispersion function. However, Eq. (2.32) shows that only the overall magnitude of $\Psi(\sigma_v, z_l)$ changes with redshift. Consequently, we expect a universal normalized magnification distribution for all redshifts.

Finally, we fit $dP_{\rm II}(\mu)/d\log_{10}\mu$ by a log normal distribution with a mean of -0.35

and standard deviation of 0.57, truncated at $\log_{10}\mu = -2, 1$.

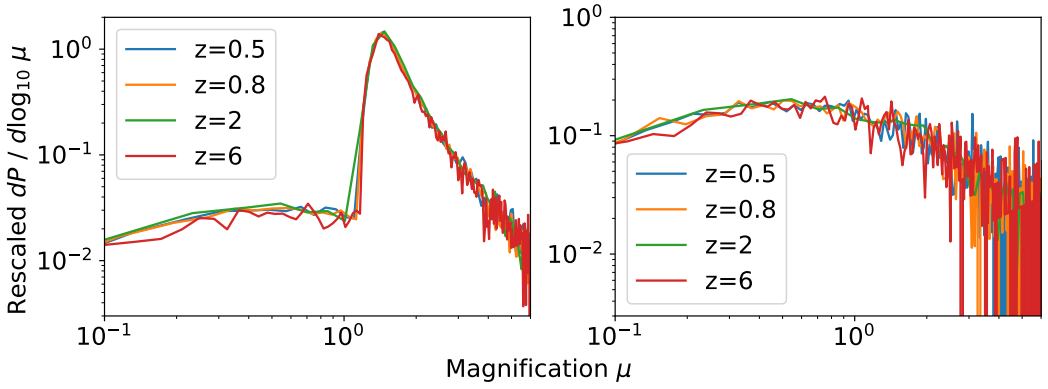

Figure 2.11: Rescaled magnification distribution at redshift $z = 0.5, 0.8, 2, 6$. The *left* panel shows $\partial P(\mu, z_s)/\partial \log_{10}\mu$, including all images from the Monte Carlo samples. The *right* panel shows the rescaled distribution of only type II images, $\partial P_{\rm II}(\mu, z_s)/\partial \log_{10}\mu$. The traces are rescaled such that the largest image count is normalized to 1.

GW Source Population

We adopt GW source population models provided by Ref. [15, 43]. In summary, we assume the merger rate of the primary black hole in a binary to be proportional to the formation rate of black holes and their progenitor stars. The merger rate is then calibrated to the observed BBH merger density in the local universe. We follow the prescription and the chosen astrophysical models in Ref. [15, 43], and we provide more details in Appendix 2.7 for reference. The BBH merger population is calibrated to a local merger rate of 103 $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ based on LIGO detection data up until GW170104 [1]. With the second LIGO-Virgo Gravitational-Wave Transient Catalog (GWTC-2), the local BBH merger rate is more tightly constrained to be $23.9^{+14.3}_{-8.6}$ $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ [1]. The data slightly favor that the merger rate increases with redshift, but remain statistically consistent with a non-evolving merger rate hypothesis [1]. This updated value suggests that our merger rate model could be an overestimate. However, the local rate difference is less than an order of magnitude, and the high-redshift merger rates are not constrained by LIGO data. As the local merger rate only acts as an overall scaling in the population model, our result can be easily scaled to reflect any differences. Consequently, our predictions serve as an adequate reference and can be easily adapted in light of new data and more accurate BBH population models.

Figure 2.12 replicates Figure 1 in Ref. [43] and shows the predicted BBH merger

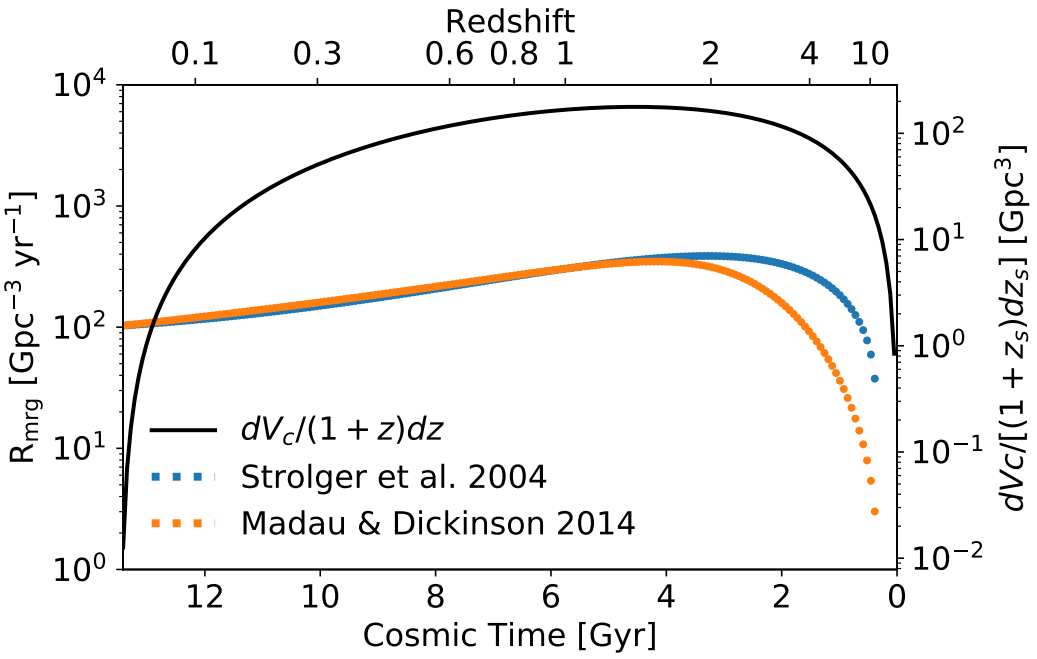

Figure 2.12: BBH merger rate density; see text for adopted SFR models. The blue over-arching trace plots the modified differential comoving volume and corresponds to values on the right axis.

rate density under the two Star Formation Rate (SFR) models in Ref. [45] and [68], respectively. Due to the intrinsic uncertainty in these analytical SFR models, we choose one, the more optimistic SFR in Ref. [68], for the following population estimates.

The total BBH merger rate per source redshift and the detectable merger rate are plotted in Figure 2.13. Since the strong lensing optical depth is in general smaller than 0.1% at the redshift with the most GW sources, we will neglect the magnification effect when calculating the total detectable GW events. To keep the detectable population estimate general, we assume the detectors to always be online. To incorporate the detector duty cycle, the detectable and distinguishable populations simply scale proportionally with the fraction of detector online time, since the type II waveforms can be identified by their own waveforms, and the duty cycle does not disproportionally affect particular GW image types. We estimate a total of 2.17×10^5 BBH mergers per year up to $z_s = 23$. The detectable total merger number is 2.17×10^5 for CE (99.96%), 1.96×10^5 for ET (90.3%) and 7.59×10^4 for LIGO Voyager (35.0%). We note that the detection rate is not only affected by the detector sensitivity, but also by the redshift distribution of BBH mergers and the comoving

volume. Even though ET has lower sensitivity than CE overall, it already covers the redshift range with peak GW source count ($z_s \sim 2$). At large redshift with $z_s \gg 7$, BBH mergers happen far less frequently due to a lack of black hole formation and the decreasing comoving volume per redshift. Consequently, the detection rate of ET is only slightly lower than CE. In the case of LIGO Voyager, the lower sensitivity excludes many sources from $z_s \sim 2$, leading to a larger loss in the detectable source fraction.

type II Image Rate

In this section, we combine lensing statistics and GW source population models to study the rate of detectable and distinguishable type II images in third generation GW detectors.

Figure 2.13 shows the differential event rate as a function of redshift for three detectors. In each panel, four different populations are shown. The total rate of BBH mergers are plotted as solid black curves. The dashed curves show the rate of detectable GW events. The dotted curves show the rate of events with a detectable type II image as in Eq. (2.26). The dot-dash curves show the rate of GW events with at least one distinguishable type II image.

As expected, the rate of BBH mergers in all three categories decreases with the detector sensitivity, especially at high redshifts. For LIGO Voyager, in particular, the rate of expected GW sources with distinguishable type II images drops quickly with redshift, consistent with the trend of the distinguishable fraction in Figure 2.8.

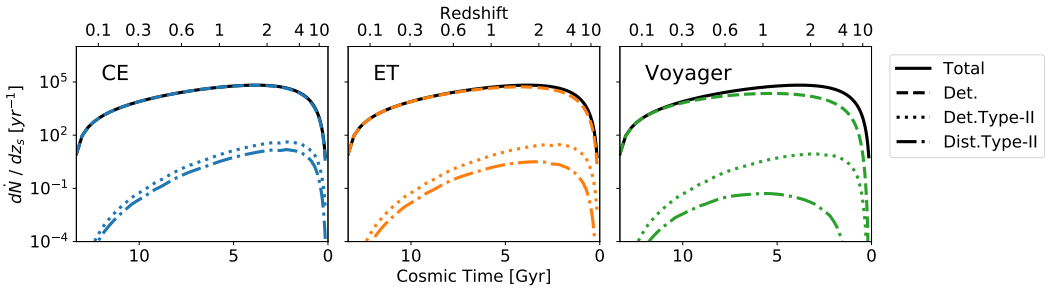

Figure 2.13: Yearly detected population per unit reshift prediction as a function of redshift. The panels from left to right show the detection population for CE, ET and LIGO Voyager. In all panels, the solid black line denotes the total BBH merger rate. The dashed curves show the rate of detectable GW sources (i.e., $\rho > 8$) when unlensed. The dotted curve shows the rate of GW sources with a detectable type II image. The dot-dash curve shows the event population with a distinguishable type II image. See text for total detection rates.

Figure 2.14 plots the same population prediction binned by the total mass of the BBH, with consistent line styles as in Figure 2.13. For all detectors, the detection rate decreases with increasing total mass, consistent with the underlying initial mass function. The detection rate of events with distinguishable type II images shows a cutoff at small total mass, which is primarily due to two factors. When the waveform mismatch for low-mass BBHs is smaller than the imposed mismatch resolution (i.e., $\epsilon < 0.001$), their type II images are considered indistinguishable from type I images. When the mismatch has just exceeded the resolution threshold, distinguishability requires very large SNRs, which may not be achievable depending on the detector sensitivity. Consequently, we observe a mass cutoff in all three detectors, which shifts to higher masses as the detector sensitivity decreases.

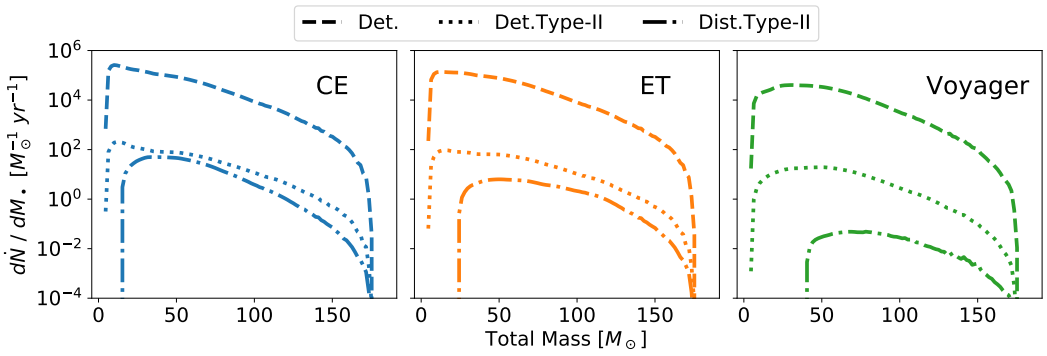

Figure 2.14: Detection rate as a function of BBH intrinsic mass. The panels from left to right shows the detection prediction for CE, ET and LIGO Voyager. The line styles are consistent with those in Figure 2.13.

Overall, we predict that CE will detect roughly 184.7 strongly lensed GW events per year, among which 172.2 have at least one detectable type II image. Among these strongly lensed GW sources, 56.9 per year have a type II image distinguishable via the log Bayes factor test. ET will be able to detect 157.1 strongly lensed events per year, and 118.2 of these have detectable type II images. However, due to reduced sensitivity, the number of sources with a distinguishable type II image drops to 8.6 per year. For LIGO Voyager, the yearly detection rate of GW events with detectable type II images is 27.4 per year out of the 38.4 strongly lensed events. The distinguishable type II image rate is 0.06 per year, which suggests that the possibility of observing a GW source with distinguishable type II images with LIGO Voyager is relatively slim. The detection rates are summarized in Table 2.1.

Discussion

In this section, we discuss the implication of the predicted detection rates for GW sources with distinguishable type II images. We re-examine assumptions in our analysis and explore how relaxing these assumptions lead to more an optimistic detection prediction.

As Figure 2.14 shows, the yearly detection rates of GW sources with distinguishable type II images are 56.9, 8.6 and 0.06 for CE, ET and LIGO Voyager, respectively. In particular, in the case of CE, more than 30% of all detectable strongly lensed sources will have distinguishable type II images. For such sources, detection of the type II image alone can confirm the existence of strongly lensed images, without pair-wise GW event inference on the strong lensing hypothesis. Once such images are identified, the inferred source parameter values can act as a prior during the subsequent and more elaborate catalog search for the other images.

For ET and LIGO Voyager, the expected detection rate is smaller, thus the distinguishable type II images will not be as powerful for confirming the strong lensing hypothesis as in the case of CE. However, we emphasize that if several of our conservative constraints can be relaxed, distinguishable type II images can still contribute to the identification of strong lensing.

The first condition we revisit is the waveform mismatch resolution. Throughout the analysis, we consistently adopt $\epsilon_{\rm min} = 0.001$, which excludes the binaries at small inclinations, and the distinguishable fraction is "saturated" at roughly 70% (see, e.g., Figure 2.7 and Figure 2.9). As Figure 2.9 suggests, the waveform mismatch resolution significantly affects the fraction of distinguishable type II images. If we can expect a better waveform resolution from third-generation GW detectors, the distinguishable fraction should increase considerably; as Figure 2.9 shows, the distinguishable fraction roughly doubles as the mismatch resolution improves from $O(1\%)$ to $O(0.1\%)$ assuming CE sensitivity. For CE and ET, this increase results in many more detectable sources at small redshifts ($z_s \sim 1, 2$), where the BBH population also peaks. This requirement has two implications for third-generation GW detector performance and data analysis process. As is discussed briefly in Section 2.3, the error in detector calibration should be much smaller, such that the waveform mismatch is not obscured by systematic uncertainties. In terms of the data analysis process, the density of the matching template bank should be such that the waveform difference is large compared with the template spacing. If such conditions are not satisfied, the high SNR detection offered by the third-generation

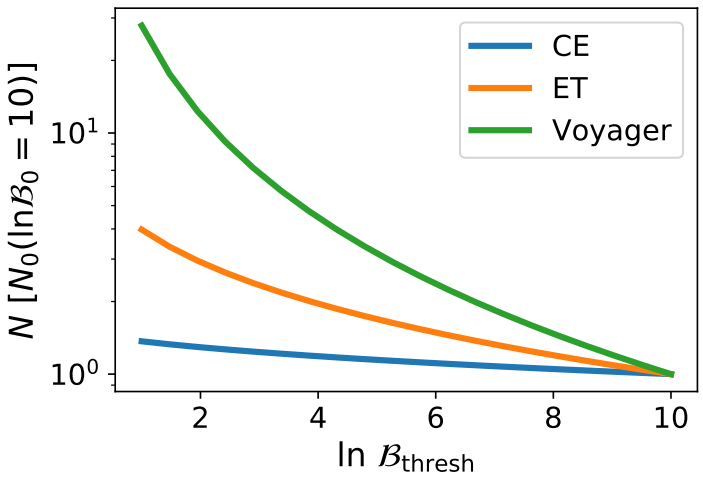

Figure 2.15: The number of GW sources with distinguishable type II images for different log Bayes factor threshold values, expressed as a fraction of the distinguishable number with the threshold value $\ln \mathcal{B}_{\rm thresh} = 10$.

detectors cannot be taken full advantage of to maximize the scientific output.

We have also taken a conservative estimate by setting the threshold log Bayes factor to be 10. Even for $\ln \mathcal{B}_{\rm thresh} = 5$, the type II image hypothesis is more than 100 times more likely than the type I image hypothesis, and an even smaller threshold value may be sufficient for realistic data analysis. Figure 2.15 shows the increase in the number of events with distinguishable type II images with a lower $\ln \mathcal{B}_{\rm thresh}$, normalized to the number when $\ln \mathcal{B}_{\rm thresh} = 10$. We observe that the increase is the most dramatic for LIGO Voyager, as a lower threshold extends the sensitive range to higher redshift ($z_s \sim 2$), where the GW source population peaks. For CE and ET, the increase is more modest, as they already detect most sources at $z_s \sim 2$ with high SNR. The extended range is then expected to add relatively fewer GW sources in comparison. Figure 2.16 shows the redshift distribution of the GW sources with distinguishable type II images with $\ln \mathcal{B}_{\rm thresh} = 2$, 5 and 10. As expected, the distinguishable rate increase is more significant at high redshift, and the effect is the strongest for LIGO Voyager; at $\ln \mathcal{B}_{\rm thresh} = 2$, 42.8% of all strongly lensed GW sources in CE are accompanied by at least one distinguishable type II image and 21.4% for ET. For LIGO Voyager, the distinguishable number is still small, but at $\sim 1/\mathrm{yr}$, it is more promising that such an event will appear in the LIGO Voyager catalog with a few years of observing run. The predicted detection rates are summarized in Table 2.1.

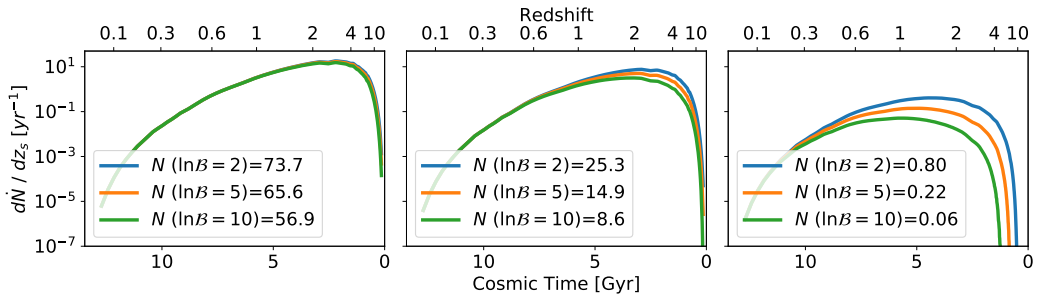

Figure 2.16: The expected number of GW sources with distinguishable type II images with $\ln \mathcal{B}_{\rm thresh} = 2,\ 5$, and 10. The expected yearly detection count for each threshold value is shown in the legend. The panels from left to right correspond to CE, ET and LIGO Voyager. As expected, the detection number increase is most significant at large redshift for all three detectors.

In addition, we have so far considered the single-detector scenario, and we estimate the advantage of a detector network via the simplifying assumption that both GW polarizations can be independently detected, i.e., the time-domain waveform for calculating the overlap and ρ is complex. If ET implements a triangular design, the detector itself is sufficient to capture the polarization content [63]. For LIGO Voyager and CE, the polarization content can be obtained if a concurrent detector network exists. In the upcoming decades, more GW observatories across the globe will start to observe, such as the expansion of the LIGO network to include IndiGO[7] [39]. This global network offers increased detector-networks ρ and an increased detection spatial resolution. On the other hand, the uncertainty in the polarization content from a realistic detector-network model may be partially degenerate with the Hilbert transform signal, thus subtracting away from the type I/II waveform difference and their distinguishability. A thorough investigation on realistic detector network effect is deferred to future studies.

2.6 Conclusions

In this paper, we study an intrinsic waveform signature of type II images of strongly lensed GW sources. For CE, ET and LIGO Voyager, we compute the best-match overlap between type I/II waveforms. We then calculate the required threshold orbit inclination to establish the type II waveform hypothesis by a favoring log Bayes factor of 10. The fraction of GW sources with distinguishable type II images is computed from the threshold inclination accordingly. For all three detectors, we

[7] `http://www.gw-indigo.org/tiki-index.php`

		CE	ET	Voyager
Det.		2.17×10^5	1.96×10^5	7.59×10^4
Det. SL		184.7	157.1	38.4
Det. type II		172.2	118.2	27.4
Dist. type II	$\ln \mathcal{B} \geq 10$	56.9 (33.1%)	8.6 (7.3%)	0.06 (0.22%)
	$\ln \mathcal{B} \geq 5$	65.6 (38.1%)	14.9 (12.6%)	0.22 (0.81%)
	$\ln \mathcal{B} \geq 2$	73.7 (42.8%)	25.3 (21.4%)	0.80 (2.93%)

Table 2.1: Predicted yearly detection rates. The columns show the detectable BBH merger rates, the rates for strongly lensed (SL) BBH mergers, the rates for BBH mergers with detectable type II images and the rates for BBH mergers with distinguishable type II images. The distinguishable event rates are given with $\ln \mathcal{B} \geq 10, 5, 2$. The fraction of events with detectable type II images that are also distinguishable is shown in the parenthesis.

find that significant fractions of type II images (e.g., 50 – 70%) of sufficiently high SNR GW events can be identified. In other words, if such a type II image is detected with reasonable SNR, it can likely be distinguished from regular type I images and used as the tell-tale evidence of strongly lensed events.

We also assess the effects of the type II signature in the context of the current LIGO data analysis process. We apply the targeted sub-threshold search method described in Ref. [42] on an example high-mass-ratio compact binary coalescence event GW190814 [6]. We generate a reduced template bank based on injection run results using simulated type I lensed injections of the target event. The resulting reduced bank is used, then, to search for the same set of simulated lensed injections in two different searches, in which they are injected as type I images (original waveforms) and type II images (Hilbert transform of the same waveforms) respectively.

Our preliminary result shows that there is a slight increase in the number of injections missed when they are treated as type II images. This hints at the possibility that the current search scheme may suffer from sensitivity loss without considering type II images. However, we remark that the current results in this study are only preliminary and will require further studies.

We then incorporate GW source population model and lensing probabilities to predict the expected number of GW sources with distinguishable type II images in CE, ET and LIGO Voyager respectively. For these three detectors, we predict the yearly detection rates are 56.9, 8.6 and 0.06 with a conservative threshold at

$\ln \mathcal{B}_{\rm thresh} = 10$. A relaxed log Bayes factor threshold boosts the expected detection rates, especially for LIGO Voyager; at $\ln \mathcal{B}_{\rm thresh} = 2$, the yearly detection rate for LIGO Voyager approaches $1/\mathrm{yr}$.

Such distinguishable type II images are "short-cuts" for identifying strongly lensed events, as they guarantee the existence of at least one other lensed image. They also improve the computational efficiency of searching for the companion images, as the estimated parameters, such as the redshifted mass, mass ratio and sky location, can inform a more comprehensive catalog search. As illustrated, this method will be most powerful with the unprecedented sensitivity offered by third-generation GW detectors.

Our work can be extended and refined in several directions. We can relax the constraints on GW source range by including spin and orbit eccentricity, as is studied in Ref. [46]. On one hand, the Hilbert transform of GWs from such sources may have a larger mismatch from the original waveform, favoring type II image distinguishability. On the other hand, the Hilbert transform may be partially degenerate with a parameter bias with the additional degrees of freedom. The effect of these competing factors warrants careful treatment.

We may consider realistic detector networks instead of assuming complete knowledge on both GW polarizations, which adds to the underlying waveform uncertainties. Similar to the hypothesized effect of binary spin and orbital eccentricity, uncertainty in the polarization may be partially degenerate with the Hilbert transform signature. However, a detector network yields larger signal SNR, which should promote the distinguishability of type II images.

It is also important to refine the lens modeling. While theoretical works on various lens types and their respective image characteristics abound, to our knowledge, the effect of the model choice on predictions for realistic detection has yet to be systematically investigated. Therefore, such a follow-up study is essential for understanding the uncertainty and robustness of this strongly-lensing detection forecast.

In conclusion, this study shows that the intrinsic waveform characteristics of type II images can be a powerful supplemental tool for hunting strongly lensed events in the catalog of third-generation GW detectors, when tens of such events may be identified.

2.7 Appendix: Binary Black Hole Merger Rate

In this appendix, we elaborate on the astrophysical models adopted to calculate the merger rate of binary black holes. In summary, we compute the BBH merger rate from population models on black hole progenitor stars and calibrate to the observed rate in the local universe.

Adapted from Eq. (B1) and (B2) in Ref. [43], the birth rate of individual black holes with mass $m_\bullet$ at redshift z_s, $R_{\rm birth}(m_\bullet, z_s)$ is given by

$$R_{\rm mrg}(m_\bullet, z_s) = \int dm_\star\, dt_d\, dZ\, \phi(m_\star)\dot{\psi}\left(t\left(z_s\right) - t_d\right) P(t_d) P\left(Z, t\left(z_s\right) - t_d\right) \delta\left[m_\star - g^{-1}(m_\bullet, Z)\right] , \tag{2.34}$$

where $m_\star$ is the mass of the progenitor star, Z is stellar metallicity and $t(z_s)$ is the cosmic time as a function of redshift. $g^{-1}(m_\bullet, Z)$ gives the stellar mass $m_\star$ with metallicity Z that leaves a black hole remnant with mass $m_\bullet$. The expression of remnant black hole mass as a function of stellar mass and metallicity is given in Ref. [67], with $20\ M_\odot < m_\star < 105\ M_\odot$ and $-5 < \log_{10} Z < -1.7$. $P(Z, t_z)$ is the redshift-dependent distribution of metallicity. The mean log metallicity at any redshift is given in Ref. [10]. At each redshift, the metallicity follows a log normal distribution [10].

t_d is the time delay between black hole formation and its merger with another black hole. $P(t_d)$ is the distribution of time delay, and we adopt the form $P(t_d) \propto t_d^{-1}$, truncated at $t_d = 50$ Myr and the Hubble time [43]. Note that we ignore the time delay between the formation of a star and the formation of its remnant. Since stellar evolution is on the order of Myr, which is negligibly small compare to the evolution time scale of galaxies and hence that of black holes, we can neglect it for model simplicity without incurring large errors.

The quantity $\phi(m_\star)$ is the initial mass function that describes the stellar mass distribution, which we assume to remain constant across redshift. Specifically, we adopt the Chabrier initial mass function [16] for $m_\star > 1\ M_\odot$, where $\phi(m_\star) \propto m_\star^{-2.3}$. The quantity $\dot{\psi}(t)$ is the Star Formation Rate (SFR) including all $m_\star$ at cosmic time t. We adopt the analytic SFR expression in Ref. [68]. We calibrate the merger rate at $z = 0$ to be 103 $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$, which is the expected local black hole merger rate given LIGO detection data up until GW170104 [1, 15]. See text for the effects of an updated local merger rate based on GWTC-2.

We note that $R_{\rm mrg}(m_\bullet, z_s)$ is the rates for black hole binary at z_s whose *primary*

black hole, i.e. the heavier one, is $m_\bullet$. We then assign a mass ratio value according to the distribution $P(q) \propto q$, with q truncated at 1.2 and 3.2. We can then directly convert the rates into $R_{\rm mrg}(M_\bullet, q, z_s)$ where $M_\bullet$ is the total binary mass.

References

[1] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, B. Agarwal, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, others, LIGO Scientific, and Virgo Collaboration. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Phys. Rev. Lett.*, 118(22):221101, June 2017. doi: 10.1103/PhysRevLett.118.221101.

[2] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, G. Allen, A. Allocca, others, LIGO Scientific Collaboration, and Virgo Collaboration. Search for intermediate mass black hole binaries in the first and second observing runs of the Advanced LIGO and Virgo network. *Phys. Rev. D*, 100(6):064064, September 2019. doi: 10.1103/PhysRevD.100.064064.

[3] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, others, LIGO Scientific Collaboration, and Virgo Collaboration. Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1. *Phys. Rev. D*, 100(10):104036, November 2019. doi: 10.1103/PhysRevD.100.104036.

[4] B. P. Abbott et al. 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(3):031040, 2019. doi: 10.1103/PhysRevX.9.031040.

[5] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen, others, LIGO Scientific Collaboration, and Virgo Collaboration. GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object. *ApJ*, 896(2):L44, June 2020. doi: 10.3847/2041-8213/ab960f.

[6] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, others, LIGO Scientific Collaboration, and Virgo Collaboration. GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar

Mass Compact Object. *ApJ*, 896(2):L44, June 2020. doi: 10.3847/2041-8213/ab960f.

[7] R. X. Adhikari, K. Arai, A. F. Brooks, C. Wipf, O. Aguiar, P. Altin, B. Barr, L. Barsotti, R. Bassiri, A. Bell, et al. A cryogenic silicon interferometer for gravitational-wave detection. *Classical and Quantum Gravity*, 37(16):165003, August 2020. doi: 10.1088/1361-6382/ab9143.

[8] Bruce Allen. χ^2 time-frequency discriminator for gravitational wave detection. *Phys. Rev. D*, 71(6):062001, March 2005. doi: 10.1103/PhysRevD.71.062001.

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

[10] Krzysztof Belczynski, Daniel E. Holz, Tomasz Bulik, and Richard O'Shaughnessy. The first gravitational-wave source from the isolated evolution of two stars in the 40-100 solar mass range. *Nature*, 534(7608):512–515, June 2016. doi: 10.1038/nature18322.

[11] Marek Biesiada, Xuheng Ding, Aleksandra Piórkowska, and Zong-Hong Zhu. Strong gravitational lensing of gravitational waves from double compact binaries—perspectives for the Einstein Telescope. *J. Cosmology Astropart. Phys.*, 2014(10):080, October 2014. doi: 10.1088/1475-7516/2014/10/080.

[12] Simon Birrer and Adam Amara. lenstronomy: Multi-purpose gravitational lens modelling software package. *Physics of the Dark Universe*, 22:189–201, December 2018. doi: 10.1016/j.dark.2018.11.002.

[13] Tom Broadhurst, Jose M. Diego, and III Smoot, George. Reinterpreting Low Frequency LIGO/Virgo Events as Magnified Stellar-Mass Black Holes at Cosmological Distances. *arXiv e-prints*, art. arXiv:1802.05273, February 2018.

[14] Tom Broadhurst, Jose M. Diego, and III Smoot, George F. Twin LIGO/Virgo Detections of a Viable Gravitationally-Lensed Black Hole Merger. *arXiv e-prints*, art. arXiv:1901.03190, January 2019.

[15] Liang Cao, Youjun Lu, and Yuetong Zhao. Host galaxy properties of mergers of stellar binary black holes and their implications for advanced LIGO gravitational wave sources. *MNRAS*, 474(4):4997–5007, March 2018. doi: 10.1093/mnras/stx3087.

[16] Gilles Chabrier. Galactic Stellar and Substellar Initial Mass Function. *PASP*, 115(809):763–795, July 2003. doi: 10.1086/376392.

[17] Mark H. Y. Cheung, Joseph Gais, Otto A. Hannuksela, and Tjonnie G. F. Li. Stellar-mass microlensing of gravitational waves. *arXiv e-prints*, art. arXiv:2012.07800, December 2020.

[18] Yun-Young Choi, Changbom Park, and Michael S. Vogeley. Internal and Collective Properties of Galaxies in the Sloan Digital Sky Survey. *ApJ*, 658 (2):884–897, April 2007. doi: 10.1086/511060.

[19] Thomas E. Collett and David Bacon. Testing the Speed of Gravitational Waves over Cosmological Distances with Strong Gravitational Lensing. *Phys. Rev. Lett.*, 118(9):091101, March 2017. doi: 10.1103/PhysRevLett.118.091101.

[20] Neil Cornish, Laura Sampson, Nicolas Yunes, and Frans Pretorius. Gravitational Wave Tests of General Relativity with the Parameterized Post-Einsteinian Framework. *Phys. Rev. D*, 84:062003, 2011. doi: 10.1103/PhysRevD.84.062003.

[21] Curt Cutler and Éanna 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(6):2658–2697, March 1994. doi: 10.1103/PhysRevD.49.2658.

[22] Liang Dai and Tejaswi Venumadhav. On the waveforms of gravitationally lensed gravitational waves. *arXiv e-prints*, art. arXiv:1702.04724, February 2017.

[23] Liang Dai, Tejaswi Venumadhav, and Kris Sigurdson. Effect of lensing magnification on the apparent distribution of black hole mergers. *Phys. Rev. D*, 95 (4):044011, February 2017. doi: 10.1103/PhysRevD.95.044011.

[24] Liang Dai, Barak Zackay, Tejaswi Venumadhav, Javier Roulet, and Matias Zaldarriaga. Search for Lensed Gravitational Waves Including Morse Phase Information: An Intriguing Candidate in O2. *arXiv e-prints*, art. arXiv:2007.12709, July 2020.

[25] Walter Del Pozzo, Katherine Grover, Ilya Mandel, and Alberto Vecchio. Testing general relativity with compact coalescing binaries: comparing exact and predictive methods to compute the Bayes factor. *Class. Quant. Grav.*, 31(20):205006, 2014. doi: 10.1088/0264-9381/31/20/205006.

[26] J. M. Diego, O. A. Hannuksela, P. L. Kelly, G. Pagano, T. Broadhurst, K. Kim, T. G. F. Li, and G. F. Smoot. Observational signatures of microlensing in gravitational waves at LIGO/Virgo frequencies. *A&A*, 627:A130, July 2019. doi: 10.1051/0004-6361/201935490.

[27] Xuheng Ding, Marek Biesiada, and Zong-Hong Zhu. Strongly lensed gravitational waves from intrinsically faint double compact binaries—prediction for the Einstein Telescope. *J. Cosmology Astropart. Phys.*, 2015(12):006, December 2015. doi: 10.1088/1475-7516/2015/12/006.

[28] M. Evans, R. Sturani, S. Vitale, and E. Hall. Unofficial sensitivity curves (ASD) for aLIGO, Kagra, Virgo, Voyager, Cosmic Explorer and ET. Technical Report LIGO-T1500293-v11, September 2018. URL https://dcc.ligo.org/LIGO-T1500293-v11/public.

[29] Xi-Long Fan, Kai Liao, Marek Biesiada, Aleksandra Piórkowska-Kurpas, and Zong-Hong Zhu. Speed of gravitational waves from strongly lensed gravitational waves and electromagnetic signals. *Physical Review Letters*, 118(9): 091102, 2017.

[30] Marc Favata. Post-Newtonian corrections to the gravitational-wave memory for quasicircular, inspiralling compact binaries. *Phys. Rev. D*, 80(2):024002, July 2009. doi: 10.1103/PhysRevD.80.024002.

[31] Marc Favata. The gravitational-wave memory effect. *Classical and Quantum Gravity*, 27(8):084036, April 2010. doi: 10.1088/0264-9381/27/8/084036.

[32] Scott Field, Chad Galley, and Jonathan Blackman. GWSurrogate: An easy-to-use interface to gravitational wave surrogate models. In *APS April Meeting Abstracts*, volume 2018 of *APS Meeting Abstracts*, page G14.005, January 2018.

[33] Steven L. Finkelstein. Observational Searches for Star-Forming Galaxies at z > 6. *PASA*, 33:e037, August 2016. doi: 10.1017/pasa.2016.26.

[34] Cecilio García-Quirós, Marta Colleoni, Sascha Husa, Héctor Estellés, Geraint Pratten, Antoni Ramos-Buades, Maite Mateu-Lucena, and Rafel Jaume. Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries. *Phys. Rev. D*, 102(6):064002, 2020. doi: 10.1103/PhysRevD.102.064002.

[35] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith. Search for Gravitational Lensing Signatures in LIGO-Virgo Binary Black Hole Events. *ApJ*, 874(1):L2, March 2019. doi: 10.3847/2041-8213/ab0c0f.

[36] Stefan Hilbert, Simon D. M. White, Jan Hartlap, and Peter Schneider. Strong lensing optical depths in a ΛCDM universe. *MNRAS*, 382(1):121–132, November 2007. doi: 10.1111/j.1365-2966.2007.12391.x.

[37] Stefan Hilbert, Simon D. M. White, Jan Hartlap, and Peter Schneider. Strong-lensing optical depths in a ΛCDM universe - II. The influence of the stellar mass in galaxies. *MNRAS*, 386(4):1845–1854, June 2008. doi: 10.1111/j.1365-2966.2008.13190.x.

[38] Moritz Hübner, Colm Talbot, Paul D. Lasky, and Eric Thrane. Measuring gravitational-wave memory in the first LIGO/Virgo gravitational-wave transient catalog. *Phys. Rev. D*, 101(2):023011, January 2020. doi: 10.1103/PhysRevD.101.023011.

[39] B. Iyer, T. Souradeep, C. Unnikrishnan, S. Dhurandhar, S. Raja, and A. Sengupta. LIGO-India, Proposal of the Consortium for Indian Initiative in Gravitational-wave Observations (IndIGO),[Public]. Technical Report LIGO-M1100296-v2, November 2011. URL https://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=75988.

[40] Chinmay Kalaghatgi, Mark Hannam, and Vivien Raymond. Parameter estimation with a spinning multimode waveform model. *Phys. Rev. D*, 101(10): 103004, May 2020. doi: 10.1103/PhysRevD.101.103004.

[41] Kwun-Hang Lai, Otto A. Hannuksela, Antonio Herrera-Martín, Jose M. Diego, Tom Broadhurst, and Tjonnie G. F. Li. Discovering intermediate-mass black hole lenses through gravitational wave lensing. *Phys. Rev. D*, 98(8):083005, October 2018. doi: 10.1103/PhysRevD.98.083005.

[42] Alvin K. Y. Li, Rico K. L. Lo, Surabhi Sachdev, C. L. Chan, E. T. Lin, Tjonnie G. F. Li, and Alan J. Weinstein. Targeted Sub-threshold Search for Strongly-lensed Gravitational-wave Events. *arXiv e-prints*, art. arXiv:1904.06020, April 2019.

[43] Shun-Sheng Li, Shude Mao, Yuetong Zhao, and Youjun Lu. Gravitational lensing of gravitational waves: a statistical perspective. *MNRAS*, 476(2): 2220–2229, May 2018. doi: 10.1093/mnras/sty411.

[44] Kai Liao, Xi-Long Fan, Xuheng Ding, Marek Biesiada, and Zong-Hong Zhu. Precision cosmology from future lensed gravitational wave and electromagnetic signals. *Nature Communications*, 8:1148, October 2017. doi: 10.1038/s41467-017-01152-9.

[45] Piero Madau and Mark Dickinson. Cosmic Star-Formation History. *ARA&A*, 52:415–486, August 2014. doi: 10.1146/annurev-astro-081811-125615.

[46] Jose María Ezquiaga, Daniel E. Holz, Wayne Hu, Macarena Lagos, and Robert M. Wald. Phase effects from strong gravitational lensing of gravitational waves. *arXiv e-prints*, art. arXiv:2008.12814, August 2020.

[47] Connor McIsaac, David Keitel, Thomas Collett, Ian Harry, Simone Mozzon, Oliver Edy, and David Bacon. Search for strongly lensed counterpart images of binary black hole mergers in the first two LIGO observing runs. *Phys. Rev. D*, 102(8):084031, October 2020. doi: 10.1103/PhysRevD.102.084031.

[48] Ashish Kumar Meena and Jasjeet Singh Bagla. Gravitational lensing of gravitational waves: wave nature and prospects for detection. *MNRAS*, 492(1): 1127–1134, February 2020. doi: 10.1093/mnras/stz3509.

[49] Cody Messick, Kent Blackburn, Patrick Brady, Patrick Brockill, Kipp Cannon, Romain Cariou, Sarah Caudill, Sydney J. Chamberlin, Jolien D. E.

Creighton, Ryan Everett, Chad Hanna, Drew Keppel, Ryan N. Lang, Tjonnie G. F. Li, Duncan Meacher, Alex Nielsen, Chris Pankow, Stephen Privitera, Hong Qi, Surabhi Sachdev, Laleh Sadeghian, Leo Singer, E. Gareth Thomas, Leslie Wade, Madeline Wade, Alan Weinstein, and Karsten Wiesner. Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. *Phys. Rev. D*, 95(4):042001, February 2017. doi: 10.1103/PhysRevD.95.042001.

[50] Blake Moore, Travis Robson, Nicholas Loutrel, and Nicolás Yunes. Towards a Fourier domain waveform for non-spinning binaries with arbitrary eccentricity. *Classical and Quantum Gravity*, 35(23):235006, December 2018. doi: 10.1088/1361-6382/aaea00.

[51] Suvodip Mukherjee, Benjamin D. Wandelt, and Joseph Silk. Probing the theory of gravity with gravitational lensing of gravitational waves and galaxy surveys. *MNRAS*, 494(2):1956–1970, May 2020. doi: 10.1093/mnras/staa827.

[52] Suvodip Mukherjee, Benjamin D. Wandelt, and Joseph Silk. Multimessenger tests of gravity with weakly lensed gravitational waves. *Phys. Rev. D*, 101(10): 103509, May 2020. doi: 10.1103/PhysRevD.101.103509.

[53] T. T. Nakamura and S. Deguchi. Wave Optics in Gravitational Lensing. *Progress of Theoretical Physics Supplement*, 133:137–153, January 1999. doi: 10.1143/PTPS.133.137.

[54] Takahiro T. Nakamura. Gravitational Lensing of Gravitational Waves from Inspiraling Binaries by a Point Mass Lens. *Phys. Rev. Lett.*, 80(6):1138–1141, February 1998. doi: 10.1103/PhysRevLett.80.1138.

[55] Ken K. Y. Ng, Kaze W. K. Wong, Tom Broadhurst, and Tjonnie G. F. Li. Precise LIGO lensing rate predictions for binary black holes. *Phys. Rev. D*, 97 (2):023012, January 2018. doi: 10.1103/PhysRevD.97.023012.

[56] Alex Nitz, Ian Harry, Duncan Brown, Christopher M. Biwer, Josh Willis, Tito Dal Canton, Collin Capano, Larne Pekowsky, Thomas Dent, Andrew R. Williamson, Gareth S Davies, Soumi De, Miriam Cabero, Bernd Machenschalk, Prayush Kumar, Steven Reyes, Duncan Macleod, Dfinstad, Francesco Pannarale, Thomas Massinger, Sumit Kumar, Márton Tápai, Leo Singer, Sebastian Khan, Stephen Fairhurst, Alex Nielsen, Shashwat Singh, and Shasvath. gwastro/pycbc: PyCBC release v1.16.11, October 2020.

[57] Masamune Oguri. Effect of gravitational lensing on the distribution of gravitational waves from distant binary black hole mergers. *MNRAS*, 480(3): 3842–3855, November 2018. doi: 10.1093/mnras/sty2145.

[58] Kris Pardo, Maya Fishbach, Daniel E. Holz, and David N. Spergel. Limits on the number of spacetime dimensions from GW170817. *J. Cosmology Astropart. Phys.*, 2018(7):048, July 2018. doi: 10.1088/1475-7516/2018/07/048.

[59] Scott Perkins and Nicolás Yunes. Probing screening and the graviton mass with gravitational waves. *Classical and Quantum Gravity*, 36(5):055013, March 2019. doi: 10.1088/1361-6382/aafce6.

[60] Geraint Pratten, Cecilio García-Quirós, Marta Colleoni, Antoni Ramos-Buades, Héctor Estellés, Maite Mateu-Lucena, Rafel Jaume, Maria Haney, David Keitel, Jonathan E. Thompson, and Sascha Husa. Let's twist again: computationally efficient models for the dominant and sub-dominant harmonic modes of precessing binary black holes. *arXiv e-prints*, art. arXiv:2004.06503, April 2020.

[61] David Reitze, Rana X. Adhikari, Stefan Ballmer, Barry Barish, Lisa Barsotti, GariLynn Billingsley, Duncan A. Brown, Yanbei Chen, et al. Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. In *Bulletin of the American Astronomical Society*, volume 51, page 35, September 2019.

[62] Andrew Robertson, Graham P. Smith, Richard Massey, Vincent Eke, Mathilde Jauzac, Matteo Bianconi, and Dan Ryczanowski. What does strong gravitational lensing? The mass and redshift distribution of high-magnification lenses. *MNRAS*, 495(4):3727–3739, July 2020. doi: 10.1093/mnras/staa1429.

[63] B. Sathyaprakash, M. Abernathy, F. Acernese, P. Amaro-Seoane, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker, N. Beveridge, S. Birindelli, S. Bose, L. Bosi, S. Braccini, C. Bradaschia, T. Bulik, et al. Scientific Potential of Einstein Telescope. *arXiv e-prints*, art. arXiv:1108.1423, August 2011.

[64] Peter Schneider, Jürgen Ehlers, and Emilio E. Falco. *Gravitational Lenses*. 1992. doi: 10.1007/978-3-662-03758-4.

[65] Alberto Sesana, Francesco Haardt, Piero Madau, and Marta Volonteri. Low-Frequency Gravitational Radiation from Coalescing Massive Black Hole Binaries in Hierarchical Cosmologies. *ApJ*, 611(2):623–632, August 2004. doi: 10.1086/422185.

[66] Joshua S. Speagle. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *MNRAS*, 493(3):3132–3158, April 2020. doi: 10.1093/mnras/staa278.

[67] Mario Spera, Michela Mapelli, and Alessandro Bressan. The mass spectrum of compact remnants from the PARSEC stellar evolution tracks. *MNRAS*, 451 (4):4086–4103, August 2015. doi: 10.1093/mnras/stv1161.

[68] Louis-Gregory Strolger, Adam G. Riess, Tomas Dahlen, Mario Livio, Nino Panagia, Peter Challis, John L. Tonry, Alexei V. Filippenko, Ryan Chornock, Henry Ferguson, Anton Koekemoer, Bahram Mobasher, Mark Dickinson, et al. The Hubble Higher z Supernova Search: Supernovae to z $\sim$1.6 and Constraints

on Type Ia Progenitor Models. *ApJ*, 613(1):200–223, September 2004. doi: 10.1086/422901.

[69] Ryuichi Takahashi and Takashi Nakamura. Wave Effects in the Gravitational Lensing of Gravitational Waves from Chirping Binaries. *ApJ*, 595(2):1039–1051, October 2003. doi: 10.1086/377430.

[70] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen, et al. Properties and astrophysical implications of the 150 Msun binary black hole merger GW190521. *arXiv e-prints*, art. arXiv:2009.01190, September 2020.

[71] Paul Torrey, Sarah Wellons, Francisco Machado, Brendan Griffen, Dylan Nelson, Vicente Rodriguez-Gomez, Ryan McKinnon, Annalisa Pillepich, Chung-Pei Ma, Mark Vogelsberger, Volker Springel, and Lars Hernquist. An analysis of the evolving comoving number density of galaxies in hydrodynamical simulations. *MNRAS*, 454(3):2770–2786, December 2015. doi: 10.1093/mnras/stv1986.

[72] Michele Vallisneri. Testing general relativity with gravitational waves: a reality check. *Phys. Rev. D*, 86:082001, 2012. doi: 10.1103/PhysRevD.86.082001.

[73] Stéfan van der Walt, S. Chris Colbert, and Gaël Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science and Engineering*, 13(2):22–30, March 2011. doi: 10.1109/MCSE.2011.37.

[74] Vijay Varma, Scott E. Field, Mark A. Scheel, Jonathan Blackman, Davide Gerosa, Leo C. Stein, Lawrence E. Kidder, and Harald P. Pfeiffer. Surrogate models for precessing binary black hole simulations with unequal masses. *Physical Review Research*, 1(3):033015, October 2019. doi: 10.1103/PhysRevResearch.1.033015.

[75] Aditya Vijaykumar, M. V. S. Saketh, Sumit Kumar, Parameswaran Ajith, and Tirthankar Roy Choudhury. Probing the large scale structure using gravitational-wave observations of binary black holes. *arXiv e-prints*, art. arXiv:2005.01111, May 2020.

[76] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17: 261–272, 2020. doi: https://doi.org/10.1038/s41592-019-0686-2.

[77] Lilan Yang, Xuheng Ding, Marek Biesiada, Kai Liao, and Zong-Hong Zhu. How Does the Earth's Rotation Affect Predictions of Gravitational Wave Strong Lensing Rates? *ApJ*, 874(2):139, April 2019. doi: 10.3847/1538-4357/ab095c.

[78] Matias Zaldarriaga, Doron Kushnir, and Juna A. Kollmeier. The expected spins of gravitational wave sources with isolated field binary progenitors. *MNRAS*, 473(3):4174–4178, January 2018. doi: 10.1093/mnras/stx2577.

Chapter 3

DETECTION PROSPECTS FOR GRAVITATIONAL-WAVE LENSING IN HIERARCHICAL TRIPLE SYSTEM WITH KERR BLACK HOLE

3.1 Introduction

The successful detection of Gravitational Wave (GW) events from compact binary mergers by the advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO) and Virgo collaboration heralds the era of gravitational wave astronomy and proves to be an invaluable source for our understanding of both relativity and astrophysics [1, 4, 5, 52]. Aside from being a target of interest itself, GWs carry information on their propagation environment [3, 40, 43, 45]; lensing is one such imprint that shed light on the GW source galaxy and other cosmological objects along its path [24, 36, 41, 53].

Much literature on GW lensing focuses on the leading standard lensing scenario, in which the lens, source and the observer are astronomical distances apart, and the wave is slightly deflected by the lens before reaching the observer [36, 53, 55]. For a strong lens, such as a supermassive black hole (SMBH), the wave can wind around the lens multiple times before reaching the observer, which produces higher order images, or relativistic images [15, 21]. The lowest order relativistic image is a "retrolensing" image [7, 38], which occurs when the source is in front of the lens, and the wave travels further back towards the lens before being deflected "back" towards the observer.

Relativistic images are thoroughly studied, typically in the context of electromagnetic (EM) signals. In Ref. [6, 13, 15, 21], the relativistic image position and magnification are comprehensively explored. Analytical solutions for Schwarzschild strong lens and special cases for Kerr strong lens have been given in the geometric optics limit [12, 15]. Interests for relativistic EM images were two-fold. First of all, the light ring, formed by tightly packed "rings" of relativistic EM images, is the primary target for the Event Horizon Telescope (EHT) [22, 23]. With successful observation of the shadow border around M87 and Sagittarius A*, the EHT has spurred yet greater interests in the shape of the shadow border and its science implications [11, 27, 28]. It is also studied whether individual relativistic images can

be discernible. In particular, Ref. [29] discusses the prospect of observing lensed relativistic images of the Sun by stellar-mass black holes at the edge of the solar system.

Relativistic lensing for GW images was discussed in earlier theoretical works in, e.g., Ref. [7, 38]. These works generally consider the leading order of such images, which is retrolensing. Specifically, their wave treatment puts retrolensing under the general framework of black hole scattering, and the spin difference between GWs and EM waves leads to qualitative differences beyond the geometric optics description. However, there is much less literature on the realistic detection prospects of higher order GW images. An important contributing factor is that higher order image magnification is exponentially suppressed [25] and generally lies beyond the sensitivity of the current GW detector network.

Nonetheless, detection of retrolensing GW images is not necessarily out of reach. Indeed, there are astrophysically motivated scenarios where the relativistic GW images have increased chances to be detectable, especially as the next generation GW detectors with superior sensitivity are planned and developed [30, 32, 35, 37]. An interesting example is a high inclination hierarchical triple system, where a stellar-mass binary black hole (BBH) emitting GWs is in orbit around a SMBH [20, 58]. Several possible formation channels for such hierarchical triple systems have been explored; for example, triples can form through gravitational capture in nuclear star clusters [42], migration traps in the Active Galactic Nucleus (AGN) disks [10, 39, 51] and tidal capture by the central SMBH to form binary extreme-mass-ratio inspirals [16, 17]. Direct detection of lensed GW signals from triples could, in turn, contribute to constraining and refining relevant theoretical models.

GW lensing in such triple system has been studied from several angles. In Ref. [20], the authors map out the various regimes for lensing in triple systems and estimate the probability and detectability of repeated lensing by the Laser Interferometer Space Antenna (LISA), which occurs when the stellar-mass binary completes multiple orbits around the SMBH during observation window. Repeated lensing is beneficial for detection purpose since it creates a distinctive waveform with several magnified peaks. In Ref. [58], the authors further include the leading relativistic image using analytical formula and discuss the detection prospects given TianGO sensitivity. In both works, it is assumed that the central black hole is a Schwarzschild black hole.

In this study, we explore the case where the central black hole is a Kerr black hole with arbitrary spin and a wide range of observer inclination angles. This

paper is organized as follows. In Section 3.2 and 3.3, we specify the geometry of the triple system and outline procedures for solving GW relativistic images and their magnification. In Section 3.4 and 3.5, we analyze the detection prospects of retrolensing images in the context of next-generation decihertz GW detectors. Finally, we make a forecast on the detection rate of lensed GWs and discuss paths for future work. Details on the simulation pipeline are enumerated in the Appendix Section 3.7 and 3.8.

3.2 Lensed Image Solution

In this section, we introduce the geometry of the hierarchical triple system and specify procedures for solving the GW lensed image positions. We start with the Boyer-Lindquist coordinate system, with the metric given by [14]

$$ds^2 = (1-\frac{2Mr}{\Sigma})dt^2-\frac{\Sigma}{\Delta}dr^2-\Sigma d\theta^2-\left[\Delta+\frac{2Mr(r^2+a^2)}{\Sigma}\right]\sin^2\theta d\phi^2+\frac{4aMr\sin^2\theta}{\Sigma}dtd\phi\ , \tag{3.1}$$

with

$$\begin{aligned}\Sigma &= r^2+a^2\cos^2\theta \\ \Delta &= r^2-2Mr+a^2\ .\end{aligned} \tag{3.2}$$

Throughout this work, we denote the mass of the central lensing black hole as M and the detector-frame mass of the stellar mass binary as $m_{\rm bbh}$. Taking advantage of the axial symmetry of Kerr spacetime, the azimuthal angle is chosen such that the projection of the line connecting the central black hole and the observer onto the equatorial plane is at $\phi = 0$. The GW source and the observer are denoted with the subscripts "s" and "o", and the polar angle is represented by $\mu \equiv \cos\theta$.

Since we are interested in the distant-observer case only, we treat the observer sky as a plane perpendicular to the line connecting the central black hole and the observer. The origin rests at the black hole, and the positive y direction is aligned with the projection of the black hole spin onto this plane[1].

Throughout this work, we assume the geometric optics limit is valid, in which the GW wavelength is much smaller compared to the length scale of the central black hole [50, 58]. For example, for a binary emitting decihertz GWs around a $10^8\ M_\odot$

[1]Such direction is not well-defined for a face-on central black hole. However, the individual relativistic images in this case are typically faint, which places this scenario outside the most interesting parameter space of this work. It is well-defined for the near edge-on case that we are interested in in this paper.

SMBH, the parameter $\omega = 8\pi M f$ is much larger than 1, placing us safely in the geometric optics regime. In this case, finding the lensed GW images reduces to finding null geodesics that pass through both the source and the observer location [19].

In Kerr spacetime, null geodesics are determined with two constants of motion, (J, Q); the geodesic equation can be written in the integrated form as [26]

$$\begin{aligned} I_r &= G_\theta \\ \phi_o - \phi_s &= I_\phi + J G_\phi \\ t_o - t_s &= I_t + a^2 G_t \ , \end{aligned} \tag{3.3}$$

with

$$\begin{aligned} I_r &= \int \frac{dr}{\pm\sqrt{R}} \\ I_\phi &= a \int \frac{r^2 + a^2 - aJ}{\Delta\sqrt{R}} dr - a \int \frac{dr}{\sqrt{R}} \\ I_t &= \int \frac{r^2\Delta + 2r(r^2 + a^2 - aJ)}{\pm\Delta\sqrt{R}} dr \\ G_\theta &= \int \frac{d\theta}{\pm\sqrt{\Theta}} \\ G_\phi &= \int \frac{\csc^2\theta}{\sqrt{\Theta}} d\theta \\ G_t &= \int \frac{\cos^2\theta}{\pm\sqrt{\Theta}} d\theta \\ \Theta &= Q + a^2\cos^2\theta - J^2\cot^2\theta \\ R &= r^4 + (a^2 - J^2 - Q)r^2 + 2r(Q + (J-a)^2) - a^2 Q \ . \end{aligned} \tag{3.4}$$

In the context of lensing and relativistic images, it is convenient to re-parameterize (J, Q) as (η, ψ) where, roughly speaking, η corresponds to the angular position of GW image on the observer sky, and the "radial" coordinate is related to ψ, which varies between $m - \frac{1}{2}$ and $m + \frac{1}{2}$ [14]. The parameter m is the number of angular inversion points of the geodesic, e.g., for unlensed, direct images, $m = 0$; for the leading lensed images, $m = 1$ and $m = 2$ for the first set of retrolensing images [14].

Since Kerr geodesics are completely integrable, the integrals have analytical antiderivatives, which greatly enhances the computational efficiency. For each set of $(r_o, \mu_o, r_s, \mu_s, \phi_s)$, we numerically find (η, ψ) that satisfy Eqn. (3.3) in *Mathematica* using the built-in `FindRoot` function. The analytical expressions are presented in the Appendix Section 3.7.

3.3 Image Magnification

After determining the constants of motion, we compute the magnification for each of the lensed images using the optical scalar formalism [19, 31, 46, 47].

In this approach, one imagines an infinitesimal null geodesic bundle, $x(\tau, s)$, where τ is an affine parameter and s specifies the deviation from the central geodesic [19]. The initial tangent vector is k^α. We can track the changes in the cross section of this infinitesimal bundle via the vector

$$\zeta^\alpha \equiv \partial x(\tau, s)/\partial s \ , \tag{3.5}$$

which can be shown to satisfy the geodesic deviation equation [19]. To describe this geodesic bundle, i.e., ζ^α, we construct an appropriate orthonormal tetrad, $(k^\alpha, p^\alpha, t^\alpha, t^{*\alpha})$ adapted to the tangent vector k^α (and hence (J, Q)) [46]. We also apply transformations such that $t^\alpha, t^{*\alpha}$ are purely spatial for observers at rest in the locally non-rotating frame (LNRF) [9], which makes them suitable basis vectors for quantities these observers can measure, such as the wave polarization and bundle cross section. The detailed construction process is presented in the Appendix Section 3.8.

The vector ζ^α can be decomposed in this tetrad as [19]

$$\zeta^\alpha = \zeta^* t^\alpha + \zeta t^{*\alpha} \ . \tag{3.6}$$

Note that the ζ without any indices is a complex number. The geodesic deviation equation then gives

$$\begin{aligned} \ddot{\zeta} &= \zeta^* R_{ktkt} + \zeta R_{ktkt^*} \\ &= -\Phi_{00}\zeta - \Psi_0 \zeta^*, \end{aligned} \tag{3.7}$$

where R_{abcd} is the Riemann tensor, e.g., $R_{ktkt} \equiv R_{abcd} k^a t^b k^c t^d$, and $\Phi_{00} = 0$ in Kerr spacetime. This equation is essentially Sachs' equations, and we make this connection explicit in the Appendix Section 3.8.

To describe an elliptical bundle cross section, it is convenient to rewrite the complex number as

$$\zeta = \xi e^{i\beta} + \nu e^{-i\beta} \ , \tag{3.8}$$

where ξ, ν are complex numbers, and β is a dummy variable ranging from 0 to 2π. It can be verified that the vector $\zeta^\alpha(\beta)$ traces out an ellipse with semimajor and semiminor axes as $|\xi| + |\nu|$ and $|\xi| - |\nu|$, and the ellipse area is proportional to

$|\xi|^2 - |\nu|^2$ [47]. Let x, y and g, h be the real and imaginary part of ξ, ν, respectively. Eqn. (3.7) then gives four coupled second-order ordinary differential equations (plus an additional first order equation that comes from the construction process of the null tetrad; see the Appendix for details). They can be written as nine first order ordinary differential equations, which we numerically solve using a fourth-order Runge-Kutta routine with step-doubling for error control [48]. For initial conditions, we consider a circular beam and set, e.g., $\dot{x} = 1$ and others to zero [31].

The magnification (in power) is then given by the ratio of the bundle cross section in the lensed case to that of the unlensed, flat-space cross section. Since we always consider the distant observer limit, we are interested in the asymptotic behavior of this magnification, where the growth rate of the cross section tends to a constant. Since the GW signal strength is proportional to its amplitude, its amplification is the square root of the geodesic bundle areal magnification [31],

$$F = \sqrt{|\mu|} = \frac{1}{\sqrt{||\dot{\xi}|^2 - |\dot{\nu}|^2|}} \,. \tag{3.9}$$

3.4 Waveform Prediction

In this section, we show an example of the lensing profile in a hierarchical triple system. The example triple has $r_s = 100\ M$, $r_o = 10^8\ M$ and $\mu_o = \cos 87°$. The corotating GW source is in the equatorial plane of the central black hole with $a = 0.6$. We choose r_o to be large enough that the magnification has approximately reached its asymptotic value [31].

As described in Section 3.2 and 3.3, we track the images as the binary orbits around the SMBH, keeping the leading order lensed images and the retrolensing images, i.e., $m = 1, 2$ [2]. This choice is motivated by the fact that image magnification decreases exponentially with increasing image order and we expect them to go below detector sensitivity [25].

The resulting amplitude magnification profile is shown in Figure 3.1. The color code represents images with different m values, or different numbers of angular inversion points. The gray dashed line are interpolated between the simulation data; in these regions, the corresponding image trajectory becomes almost equatorial. The top blue trace connects to the unlensed direct image; we observe that its magnification is roughly 1 already, and we assume its value is 1 beyond the simulated azimuthal

[2]The term "retrolensing image" is generally applicable to all images whose trajectory goes further back towards the lens before reaching the observer. In this work, we use the term mostly to denote the lowest order retrolensing image, i.e., $m = 2$.

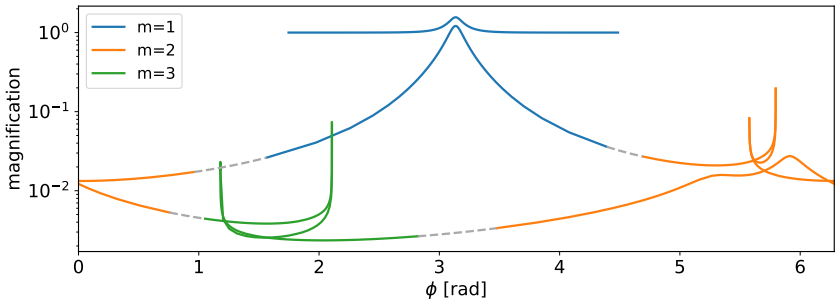

Figure 3.1: Magnification profile for an equatorial source throughout its orbit. The central black hole has a spin parameter of 0.6. The triple configuration is given as $r_s = 100\ M$, $r_o = 10^8\ M$ and $\mu_o = \cos 87°$. The legend shows images with different m values. The gray dashed line interpolates between the simulation data. The top blue trace connects to the direct, unlensed image, which we do not show in this figure. The set of three-inversion-point images ($m = 3$) includes an image that connects to $m = 4$ images, but we omit it since we do not include $m \geq 3$ images in later calculations. For both $m = 2, 3$ images, we observe similar structures where two new images appear and disappear with diverging magnification as the source goes through the caustic surfaces.

angle range. Similar to the $m = 2$ set of images, the $m = 3$ has an additional member that connects to the $m = 4$ set, which we do not show in this figure. As expected, the $m = 3$ images are roughly one order of magnitude fainter than $m = 2$ images; for this reason, we do not include it in the following batch simulations.

We observe that the leading lensing peaks (blue trace) is similar to when the lensing black hole is non-spinning; for higher orders, however, the peak splits into "double-peaks", as the GW source crosses the caustic surface of the Kerr black hole [14, 34, 50]. Upon crossing the caustic surface, two images with formally diverging magnification emerge or disappear. In optical scalar terms, the cross section of the infinitesimal geodesic bundle vanishes at the caustics [19, 50], thus giving formally infinitely bright images [3].

To further illustrate the retrolensing images, we show the image trajectory at representative orbital locations in Figure 3.2. For visual clarity, we decrease the orbital radius and the observer distance to $r_s = 20\ M$, $r_o = 30\ M$, keeping $a = 0.6$, $\mu_s = \cos 87°$. The magnification profile remains qualitatively similar to that shown in Figure 3.1. The panels share the same color code, representing

[3] In such cases, geometric optics is not sufficient, and diffraction curbs the magnification. Since this diverging region is sharp and doesn't accumulate large signal-to-noise, we do not further model diffraction to improve accuracy.

different images. Notably, the *Top Right* panel shows the image trajectories just after the GW source enters the caustics surface; two images emerge (red and yellow traces) with similar trajectories. The *Bottow Left*, on the other hand, shows the scenario just before the source exits the caustic surface; a different pair of two images (yellow and green traces) merge and disappear.

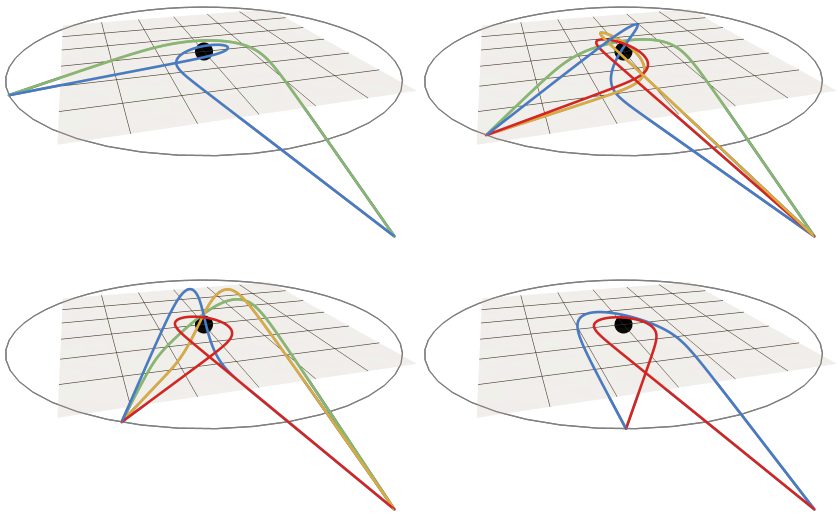

Figure 3.2: Retrolensing image geodesics for an equatorial ($r_s = 20\ M$) GW source in a triple system with an $a = 0.6$ central black hole. The observer is located at $\mu_s = \cos 87°$, $r_o = 30\ M$. The source azimuthal angles from top to bottom and from left to right are $\phi_s = 5,\ 5.48,\ 5.71$ and 6. Different images are marked with the same colors across the panels. From *Upper Left* to *Upper Right*, the source crosses the caustics surface, and two new images (the red and orange traces) emerge. From *Lower Left* to *Lower Right*, the source crosses the caustics again and two images (the green and yellow traces) merge and vanish. We show only images with $m = 2$.

In Figure 3.3, we show the retrolensing image magnification profile over a range of spin values and observer inclination angles. From left to right, the spin parameter is $a = (0.01, 0.5, 0.9)$; from top to bottom, the observer inclination is $\theta_o = (75.0°, 80.6°, 87.6°)$. For completeness, the edge of the standard lensing image is shown in blue trace in all panels. As spin and observer inclination increase, the orbit that intercepts with the caustic region grows, and we observe wider double-peaks. Such unique features can then be used to distinguish between triple systems with a spinning central black hole versus a non-spinning central black hole.

To compute the total lensing magnification to the GW waveform, we take the

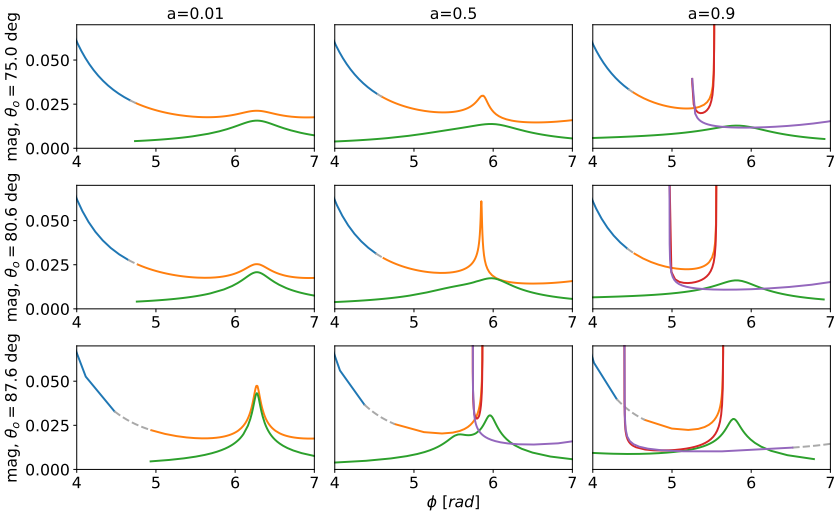

Figure 3.3: Magnification profile for retrolensing images for triple systems with various central black hole spin parameter and observer inclination. For all panels, $r_s = 100\ M$, $r_o = 10^8\ M$. The blue trace is the lensed $m = 1$ image, as is shown in Figure 3.1. From left to right, the spin values are $a = (0.01, 0.5, 0.9)$; from top to bottom, the observer inclination angle is $\theta_o = (75.0°, 80.6°, 87.6°)$. As the black hole spin and the observer inclination increase, the source orbit intersects with the caustic surface and the double-peak structure emerges. The green trace connects to $m = 3$ images (see Figure 3.1), which we do not include in the following waveform calculation and hence omit here.

conservative assumption that the lensing images add in power. Typically, the path difference between the images is on the same scale as the physical length scale of this geometric optical system, M. Since we work in the geometric optics regime, the GW wavelength should be much shorter than the lens mass scale, $1/f \ll M$. Moreover, the image path difference is also sensitive to the orbital location of the GW source. Consequently, the images have large and evolving phase differences, making them effectively incoherent. This description becomes inaccurate near the caustic crossing region, as two images share almost the same propagation paths and merge. However, as is seen in Figure 3.1 and 3.3, this region is very narrow and does not accumulate large signal-to-noise ratio (SNR). For this reason, we do not model it separately in the following forecast. See Section 3.6 for further discussion. Therefore, the total magnification is given as

$$F(\phi_s) = \sqrt{\sum_i F_i(\phi_s)^2}\ , \tag{3.10}$$

where i includes all images up to $m = 2$.

To obtain the lensing magnification in the frequency domain, we follow Ref. [57, 58] and adopt the Stationary Phase Approximation (SPA) treatment. Given any orbital angle at coalescence, we express the GW frequency as a function of time-to-coalescence and hence the orbital location, $f(\phi_s - \phi_{s,c})$. The subscript "c" denotes quantities at coalescence. Inverting the expression, we obtain the lensed waveform as

$$\begin{aligned} |\tilde{h}(f)| &= F(\phi_s(f, \phi_{s,c}))|\tilde{h}_c(f)| \\ &= F(\phi_s(f, \phi_{s,c}))\left(\frac{5}{96}\right)^{1/2} \frac{\mathcal{M}^{5/6}}{\pi^{2/3} D_L} f^{-7/6} , \end{aligned} \tag{3.11}$$

where $\mathcal{M}, D_L$ are the detector-frame chirp mass and luminosity distance, and $\tilde{h}_c(f)$ is the unmodified carrier waveform. In this work, we ignore other waveform modifications, such as Doppler shifts and Kozai-Lidov oscillations [57, 58]. Due to large uncertainty in future detector details, we make the optimistic assumption to detect all polarization content and do not include antenna patterns.

3.5 Detectability Horizon

In this section, we consider the lensed waveform from a representative hierarchical triple system and quantify its detectability prospect given realistic detector performance. First of all, we distinguish between lensed and unlensed signals. Then, we evaluate whether the spin of the central black hole gives a distinguishable lensing feature in the waveform compared with the Schwarzschild case, given the same orbital configuration.

The suitable GW frequency spectral range is subject to both the geometric optics limit requirement and the repeated lensing requirement. Since we target SMBH lenses, the GW frequency should be higher than the millihertz band. To accumulate sufficient SNR and to observe lensing signature multiple times [20], the source should stay in the detector sensitivity band for longer than the GW source orbital period, which suggests we should search below the ground-based detector frequency range, as these sources quickly coalesce. The two considerations indicate that the decihertz range can be especially promising. Within this band, we focus on the Decihertz Gravitational wave Observatory (DeciGO) [32], and the fitted noise power spectrum is given as [56]

$$S_n(f) = 7.05\times10^{-48}\left[1+\left(\frac{f}{f_p}\right)^2\right] + 4.8\times10^{-51}\left(\frac{f}{1\text{ Hz}}\right)^{-4}\left[1+\left(\frac{f}{f_p}\right)^2\right]^{-1} + 5.33\times10^{-52}\left(\frac{f}{1\text{ Hz}}\right)^{-4}\text{Hz}^{-1} , \tag{3.12}$$

where $f_p = 7.36$ Hz.

Since the waveform difference is typically small compared to whole waveform, we expect that the waveform difference is only distinguishable when the SNR is large. In this regime, the log Bayes factor, defined as the ratio between the likelihood of lensed images versus regular images, is given by [18, 55]

$$\begin{aligned} \ln \mathcal{B} \equiv \ln \frac{\mathcal{L}_L}{\mathcal{L}} &\approx \frac{1}{2}\epsilon\rho^2 \\ \epsilon &= \min\left(1 - \frac{\mathfrak{Re}(\langle \tilde{h}_L | \tilde{h} \rangle)}{\sqrt{\langle \tilde{h}_L | \tilde{h}_L \rangle \langle \tilde{h} | \tilde{h} \rangle}}\right)_{\phi_{u,c}} , \end{aligned} \tag{3.13}$$

where the mismatch is minimized over the orbital angle at coalescence of the unlensed waveform (subscript "u, c"). The inner product, $\langle \cdot \rangle$, is given by

$$\langle a | b \rangle = \int_{-\infty}^{\infty} \frac{a^*(f) b(f)}{S_n(f)} df \ , \tag{3.14}$$

where $S_n(f)$ is the two-sided detector noise power spectral density. In this work, we set the distinguishable threshold to $\ln \mathcal{B} = 1$.

In this example, the GW source is an equatorial equal-mass binary with detector-frame mass $m_{\rm bbh} = 100\ M_\odot$. We assume a central black hole with $M = 10^8\ M_\odot$ and the orbital radius is $r_s = 100\ M$. Depending on the orbital angle at coalescence of the lensed waveform, the waveform mismatch is different. In the case where the standard lensing peak or the retrolensing peak occurs when the GW frequency is at the most sensitive band of the detector, the waveform mismatch is especially emphasized. We assume that the orbital angle at coalescence of the lensed waveform is evenly distributed over 2π and average the maximum distance over which the waveform difference is distinguishable.

In Figure 3.4, we show the maximal distance out to which repeatedly lensed GWs can be distinguished from unlensed ones with $\ln \mathcal{B} \geq 1$ for various observer inclination and central black hole spin value. The simulated cases inclination is evenly spaced between 75 deg and 89 deg, and the SMBH spin is given by $[0.01, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$. The search grid is marked out with gray dashed lines. Consistent with intuition, lensing becomes more distinguishable as the observer inclination increases and the standard lensing peak grows. While there seems to be some variation with spin, it is subdominant; this is also in part due to our conservative model that images add only in power, such that an image with $F_i(f) \ll 1$ contributes only $F_i(f)^2$ to the final image. As Figure 3.4 shows, lensing

can be distinguished out to 10 $\sim$ 60 Gpc in luminosity distance, or $z = 1.4 \sim 6.2$, assuming a cosmology with $h = 0.7, \Omega_m = 0.3$. This suggests that third-generation GW detectors are able to gather abundant instances of triple system lensing out to significant redshift in this high-inclination case.

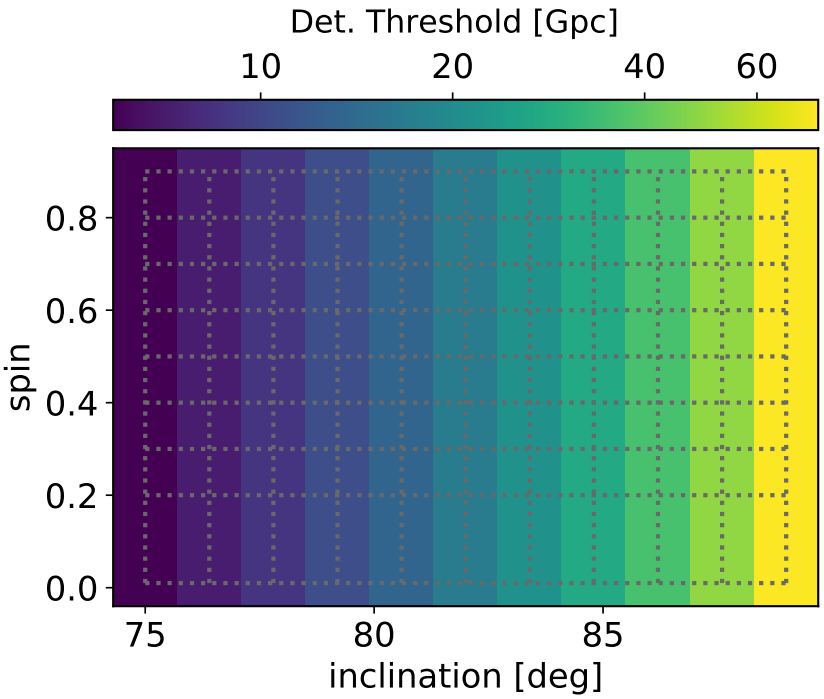

Figure 3.4: Maximal distance for distinguishing lensing signature in triple systems for various spin parameter and observer inclination values. We assume the detector sensitivity of DeciGO, and the GW source is an equal-mass binary with detector-frame mass of $m_{\rm bbh} = 100\ M_\odot$. The triple system orbital radius is the same as in Figure 3.3. Simulated cases are marked with the dashed gray grid. The dominant factor shows to be the orbital inclination, and lensing in high inclination triple systems can be distinguished out to cosmological distances.

In Figure 3.5, we show the maximal distance where lensing from spinning central black holes can be distinguished from non-spinning central black holes. Specifically, we compute the waveform difference between non-spinning and spinning central black hole with the same observer inclination angle. The non-spinning base case is approximated by setting $a = 0.01$ (the last row in the dashed gray grid in Figure 3.4). In this case, the difference is due to a shift in the standard lensing peak and the distinctive retro-lensing double-peak. We observe an overall trend that the waveforms become more distinctive as the spin and inclination increases, and lensing from spinning central black hole is distinguishable from non-spinning central black

hole within $1 \sim 10$ Gpc in luminosity distance. The irregularities on top of the trend, e.g. the $a = 0.7, \theta_o = 83.4$ deg case has a slightly further horizon than neighboring cells, can be attributed to the simulation accuracy margin; we note that the waveform difference here can be quite small, $O(10^{-8} \sim 10^{-4})$.

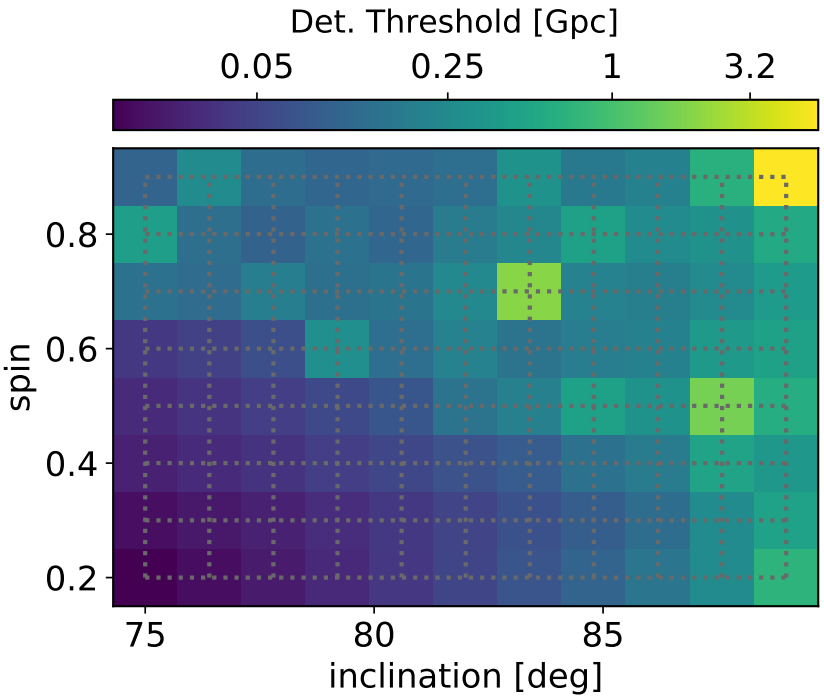

Figure 3.5: Maximal distance for distinguishing lensing with spinning central black hole. The triple parameters and the searched space is the same as in Figure 3.4. The non-spinning case is proxied by the $a = 0.01$ case. Simulated cases are marked with the dashed gray grid. As spin and observer inclination increases, spin features can be distinguished up to several Gpc. The irregularities (e.g., $a = 0.7, \theta_o = 83.4$ deg) above the overall trend is due to simulation accuracy; the waveform mismatch is typically quite small, $O(10^{-8} \sim 10^{-4})$.

We further address how the leading relativistic images contribute to the distinguishability of lensing signatures. The spin effects manifest through both the shapes and locations of the standard lensing peak and the retrolensing peaks. After identifying the best-fitting unlensed waveform or Schwarzschild-lensing waveform (with minimized waveform mismatch as is defined in Eqn. (3.13)), we compute the waveform mismatch in the two halves of the orbit, $\phi_s \in (-\pi/2, \pi/2), (\pi/2, 3\pi/2)$. While the transition between the standard lensing images ($m = 1$) and retrolensing images ($m = 2$) does not strictly follow this partition, the two halves contain the most prominent lensing peaks. As before, we convert the waveform mismatch of the two halves

into the maximum distinguishable distance. In Figure 3.6, we show the maximum distinguishable distance from the two halves as the orbital angle at coalescence changes. In this example, the system parameters are given by $a = 0.6, \theta_o = 87.6$ deg. The colored traces show mismatches computed from the retrolensing half ($\phi_s \in (-\pi/2, \pi/2)$) and the standard lensing half ($\phi_s \in (\pi/2, 3\pi/2)$). The mismatch from the total waveform is shown in the dashed black line.

As Figure 3.4 shows, the distinguishability of lensed waveform from unlensed ones is relatively insensitive to spin, albeit the large change in retrolensing waveforms. This suggests that the distinguishability is primarily driven by the standard lensing peak. In the *Upper* panel in Figure 3.6, the retrolensing contribution (blue trace) is diminishingly weak, consistent with previous observation. To distinguish lensing with spinning lenses, on the other hand, the standard lensing peak is not always dominant, since the its offset due to spin can be partially absorbed with a shift in the matching template outer orbital angle at when the stellar-mass binary coalesce. As the binary coalesces near $\phi_s = 0$, the retrolensing features are placed in the most sensitive range of DeciGO, and the waveform distinguishability is mainly driven by the retrolensing features. By taking the orbital average, the maximum distinguishable distance using standard lensing waveform only is about 2.23 times smaller than the maximum distance using the full waveform. To make an accurate prediction on the expected number of such lensed sources expected in decihertz GW detectors, it is thus important to model the relativistic images.

3.6 Discussion & Conclusion

In this work, we accurately model lensed GW signals from hierarchical triple systems with a Kerr supermassive black hole. We adopt both the elliptical integral formulation and the optical scalar formalism to numerically compute the lensing image paths and their respective magnification. Working in the high SNR regime, we make forecast on the distinguishability of lensing features for a range of central black hole spin value and observer inclination.

Using the predicted noise level of DeciGO, we find that such repeated lensing in high-inclination hierarchical triple systems can be distinguished from unlensed GW signals out to cosmological distances; depending on the spin value and observer inclination, the difference can be resolved for systems out to $z = 1.4 \sim 6.2$. Differentiating spinning black hole lens from non-spinning ones is more challenging, partly due to the small amplitude of the double-peak retrolensing pattern. Nonethe-

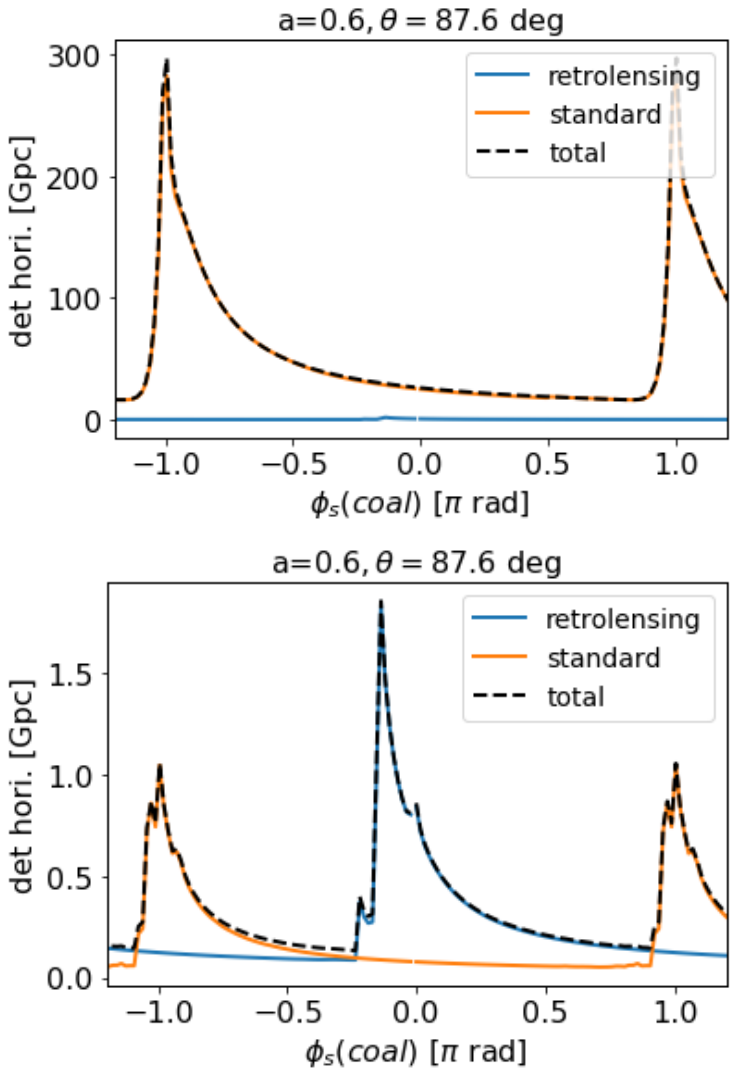

Figure 3.6: Maximal distance for distinguishing lensing with spinning central black hole with $a = 0.6, \theta_o = 87.6$ deg. The triple parameters and the searched space is the same as in Figure 3.4. The non-spinning case is proxied by the $a = 0.01$ case. The *Upper* panel shows the detectability threshold of the total lensing pattern. The *Lower* panel shows how distinguishable lensing from a spinning lens is compared to Schwarzschild lenses. In both panels, the standard lensing half and retrolensing half are shown in colored traces, and the total waveform scenario is shown as a black dashed line.

less, it is plausible to extract this difference if the triple system is within a couple of Gpc in luminosity distance. Furthermore, we show that the retrolensing peaks play an important role in distinguishing the spin-induced lensing features.

Alternatively, the detectable and distinguishable lensing signature horizon can be

converted into an effective comoving volume as

$$V_{\rm eff} = \int \Theta(D_c - D_{c,\rm max}(a,\theta)) \sin\theta dD_c d\theta d\phi \ , \tag{3.15}$$

where Θ is the Heaviside function and D_c is the comoving distance. The conversion from luminosity distance to comoving distance follows the ΛCDM model as before. The inclination is summed over our simulation range between 75 deg and 89 deg. This treatment is an underestimate of the effective volume, since all other inclinations are assumed to have maximum distinguishable distance of zero. Nonetheless, we do not expect this approximation to deviate strongly from the true value; the larger inclination angle does not represent significant volume, and the detectability is fairly weak at smaller inclination values; see, e.g. Figure 3.5. Thus, we provide a conservative estimate that the repeated lensing pattern should be distinguishable from unlensed ones in an effective comoving volume of 2.96 Gpc^3 (with a radius of $D_c = 0.89$ Gpc) at moderate central black hole spin of $a = 0.4$ (the effective volume has negligible variation with spin). In terms of the difference incurred by the central black hole spin, the distinguishable effective comoving volume is 0.14 Gpc^3 ($R = 327$ Mpc) at moderate spin ($a = 0.4$), and 0.24 Gpc^3 ($R = 387$ Mpc) at high spin ($a = 0.9$).

The concrete number of such GW candidates relies on astrophysical models on BBH distribution around SMBHs and the population of SMBHs themselves, and both are active research areas. As noted in Section 3.1, multiple hierarchical triple formation processes have been proposed in theory [10, 16, 17, 39, 42, 51], yet more direct GW data on these BBH mergers is needed to help constrain models. On the front of the SMBH population and spin properties, observational evidence, primarily coming from X-ray reflection spectroscopy, continues to shed light on SMBH spin distribution in the local universe and possible dependence on mass [44, 49, 54].

Despite the large degree of uncertainty, we can estimate the upper bound of the number of distinguishable events. The local BBH merger rate is constrained by the second LIGO-Virgo Gravitational-Wave Transient Catalog (GWTC-2) to be $23.9^{+14.3}_{-8.6}$ Gpc^{-3}yr^{-1} [2]. While decihertz detectors probe a different BBH population, this reference number suggests that $O(50)$ merger per year can have distinguishable lensing signature, and $O(1)$ could carry distinguishable imprint from the spinning central black hole.

Our work can be further developed in several directions. Firstly, we note that the model triple system always contains an equatorial GW source in a corotating

orbit around the central black hole. However, observational evidence suggests that SMBHs are not necessarily spin-aligned with the accretion disks, or the alignment process takes significant time [8, 33]. Given sufficiently large SNR, it may also be possible to infer orbital configuration from the lensing profile. A careful study on the resulting waveform modification and the effects on observation is deferred to future study.

Secondly, we note that the lensing images are added in power when we compute the total image magnification. In Section 3.4, we argue that this conservative choice reflects the fact that the pathlength difference between the images is on the order of M. Since this scale is longer than the GW wavelength by assumption, the images become effectively incoherent sources. While pairs of images get close in phase at caustic crossing, their phase coherence quickly decays away from the narrow caustic peaks. Due to the short duration of this in-phase period and the limit of detector sensitivity, we do not specifically model this. Nonetheless, the image pathlength difference can be readily obtained from our simulation, and the more exact waveform modifications can be computed.

We also exclude the polarization difference; relative to the stellar-mass binary orbital plane, the different images are emitted to different angles and contain different polarization content. While it is an interesting qualitative feature, the small wave amplitude implies that it has limited effect for realistic detection analysis. Nonetheless, since the image paths are already solved, it is straightforward to include this effect via the appropriate coordinate transformation.

Finally, it would be more thorough to maximize the waveform overlap over physical source parameters such as the observer inclination. In this work, we consider maximization over the orbital angle at coalescence, motivated by the fact that the spin effect on the standard lensing peak can be largely absorbed with a orbital phase change. Searching over the observer inclination alters the shape of the standard lensing peak, which can further reduce the waveform mismatch. However, the observer inclination change cannot fully account for the retrolensing peaks, especially as the double-peak develops. Since the retrolensing peaks contribute significantly to setting Kerr lenses and Schwarzschild lenses apart, it is likely that maximizing over an additional parameter does not qualitatively change the detection prospect. A full account relies on more extensive lensing simulation cases, which we defer to future studies.

Our study provides a platform for accurately modeling GW lensing in hierarchical

triple systems. The example case also suggests that such features will be visible in the data from next-generation decihertz detectors; with the extraordinary detector sensitivity, lensing offers a valuable tool for studying the environment of decihertz GW sources and can also be of interest for the broader astrophysics community.

3.7 Appendix: Lensed Image Geodesic Solution

In this appendix, we provide details on the geodesic solution process, combining the treatment in Ref. [14, 26].

Firstly, we consider unstable photon orbits near the Kerr black hole given by $R = 0, \partial R/\partial r = 0$, where R is the radial potential (last line in Eqn. (3.4). For an orbit with radial coordinate, r_m, a photon that escapes from this unstable orbit has an observer's sky position of

$$\begin{aligned}
\theta_{1,m} &= \frac{r_m^2(r_m-3)+a^2(r_m+1)}{r_o a(r_m-1)\sqrt{1-\mu_o^2}} \\
\theta_{2,m} &= \pm\frac{\sqrt{\Lambda(r_m)}}{r_o a(r_m-1)\sqrt{1-\mu_o^2}} \\
\Lambda(r_m) &= r_m^3\left[4a^2 - r_m(r_m-3)^2\right] \\
&\quad - 2a^2 r_m(2a^2 - 3r_m + r_m^3)\mu_o^2 - a^4(r_m-1)\mu_o^4
\end{aligned} \tag{3.16}$$

under the distant observer limit. A solution exists if $\Lambda(r_m) \geq 0$, which gives two extrema, $r_\pm$, corresponding to equatorial prograde and retrograde photons. The shadow border can be expressed with a dummy variable η as

$$r_m = \frac{1}{2}\left[r_+(1-\cos\eta) + r_-(1+\cos\eta)\right] \ , \tag{3.17}$$

with η ranging from $-\pi$ to π.

The observer sky coordinate of the lensed image can then be parameterized as

$$\begin{aligned}
\theta_1 &= \theta_{1,m}(\eta)(1+\epsilon) \\
\theta_2 &= \theta_{2,m}(\eta)(1+\epsilon)
\end{aligned} \tag{3.18}$$

with $\epsilon \in (0, \infty)$. Note that unlensed images that come directly from the source to the observe can be parameterized in a similar way, with $\epsilon \in [-1, \infty)$.

In the distant observer limit, the image position relates to the geodesic constants of motion (J, Q) via

$$\begin{aligned}
J &= -\theta_1 r_o\sqrt{1-\mu_o^2} \\
Q &= \theta_2^2 r_o^2 + \mu_o^2(\theta_1^2 r_o^2 - a^2) \ .
\end{aligned} \tag{3.19}$$

The parameters η, ϵ then determine the geodesic. We can represent ϵ with another parameter, ψ, defined as

$$2K(k)\psi = a\mu_- I_1(\epsilon, \eta) + \text{sgn}(\eta) F(\lambda_o, k) \ , \tag{3.20}$$

with

$$\begin{aligned} I_1 &= \int \frac{dr}{\pm\sqrt{R}} \\ \lambda_o &= \arcsin \frac{\mu_o}{\mu_+} \\ k &= -\frac{\mu_+^2}{\mu_-^2} \\ \mu_\pm^2 &= \frac{\sqrt{b_{JQ}^2 + 4a^2 Q} \pm b_{JQ}}{2a^2} \\ b_{JQ} &= a^2 - J - Q \ , \end{aligned} \tag{3.21}$$

and F, K are the incomplete and complete elliptical integral of the first kind. The parameter ψ enjoys the nice property of being confined to $[m - \frac{1}{2}, m + \frac{1}{2}]$ for the same image order, which greatly simplifies the solution process when we are interested in images of a particular order. Recall that m is the number of angular inversion points of the geodesic as it extends from the source to the observer.

We now quote closed form solution to several integrals in Eqn. (3.3) [26]. The antiderivatives for the angular integrals are given by

$$\begin{aligned} \mathcal{G}_\theta &= -\frac{1}{-u_- a^2} F\left(\arcsin\left(\frac{\cos\theta}{\sqrt{u_+}}\right) | \frac{u_+}{u_-}\right) \\ \mathcal{G}_\phi &= -\frac{1}{-u_- a^2} \Pi\left(u_+, \arcsin\left(\frac{\cos\theta}{\sqrt{u_+}}\right) | \frac{u_+}{u_-}\right) \\ \mathcal{G}_t &= \frac{2u_+}{-u_- a^2} E'\left(\arcsin\left(\frac{\cos\theta}{\sqrt{u_+}}\right) | \frac{u_+}{u_-}\right) \\ E'(\phi|k) &\equiv \frac{E(\phi|k) - F(\phi|k)}{2k} \\ u_\pm &= \Delta_\theta \pm \sqrt{\Delta_\theta^2 + \frac{Q}{a^2}} \\ \Delta_\theta &= \frac{1}{2}\left(1 - \frac{Q + J^2}{a^2}\right) \ , \end{aligned} \tag{3.22}$$

where E, Π are incomplete elliptical integrals of the second and third kind. The full integral is determined with the number of angular inversion points, m and the null tangent vector at the observer, i.e., η. For $x \in (t, \theta, \phi)$,

$$G_x = m\hat{G}_x + \text{sgn}(\eta)\left[(-1)^m \mathcal{G}_{x,o} - \mathcal{G}_{x,s}\right], \tag{3.23}$$

where $\hat{G}_x$ is the integral over one half-liberation between two angular inversion points. The subscripts "o, s" indicate that the antiderivatives are evaluated at the observer and the source, respectively.

The radial integrals can be written as

$$\begin{aligned} I_r &= I_0 \\ I_\phi &= \frac{2a}{r_+ - r_-}\left[\left(r_+ - \frac{aJ}{2}\right) I_+ - \left(r_- - \frac{aJ}{2}\right) I_-\right] \\ I_t &= \frac{4}{r_+ - r_-}\left[r_+\left(r_+ - \frac{aJ}{2}\right) I_+ - r_-\left(r_- - \frac{aJ}{2}\right) I_-\right] \\ &\quad + 4I_0 + 2I_1 + I_2 \ . \end{aligned} \tag{3.24}$$

We introduce the shorthand notation

$$r_{ij} = r_i - r_j, \ i, j \in (1, 2, 3, 4, +, -) \ , \tag{3.25}$$

where $r_1 < r_2 < r_3 < r_4$ are real roots to the radial potential R.

In our simulation, we are interested in the lensed images only, and the antiderivatives for the integrals are given as [4]

$$\begin{aligned} \mathcal{I}_0 &= F^{(2)}(r) \\ \mathcal{I}_1 &= r_3 F^{(2)}(r) + r_{43}\Pi_1^{(2)}(r) \\ \mathcal{I}_2 &= \frac{\sqrt{R(r)}}{r - r_3} - \frac{r_1 r_4 + r_2 r_3}{2} F^{(2)}(r) - E^{(2)}(r) \\ \mathcal{I}_\pm &= -\Pi_\pm^{(2)}(r) - \frac{F^{(2)}(r)}{r_{\pm 3}} \ , \end{aligned} \tag{3.26}$$

where

$$\begin{aligned} F^{(2)}(r) &= \frac{2}{\sqrt{r_{31} r_{42}}} F(\arcsin x_2(r)|\kappa) \\ E^{(2)}(r) &= \sqrt{r_{31} r_{42}} E(\arcsin x_2(r)|\kappa) \\ \Pi_1^{(2)}(r) &= \frac{2}{\sqrt{r_{31} r_{42}}} \Pi\left(\frac{r_{41}}{r_{31}}, \arcsin x_2(r)|\kappa\right) \\ \Pi_\pm^{(2)}(r) &= \frac{2}{\sqrt{r_{31} r_{42}}} \frac{r_{43}}{r_{\pm 3} r_{\pm 4}} \Pi\left(\frac{r_{\pm 3} r_{41}}{r_{\pm 4} r_{31}}, \arcsin x_2(r)|\kappa\right) \ . \end{aligned} \tag{3.27}$$

[4] The (2) superscript is inherited from Ref. [26], which discusses all null geodesics in Kerr spacetime. The lensed GW geodesic belongs to category 2.

The convenience functions are given by

$$\begin{aligned} I_x &= \mathcal{I}_{x,o} + \mathcal{I}_{x,s} \\ \kappa &= \frac{r_{32} r_{41}}{r_{31} r_{42}} \\ x_2(r) &\equiv \sqrt{\frac{r_{31}(r - r_4)}{r_{41}(r - r_3)}} \ , \end{aligned} \tag{3.28}$$

with $x \in (0, 1, 2, +, -)$.

3.8 Appendix: Geodesic Bundle

In this appendix, we supplement details on the optical scalar formalism to compute lensed image amplification. Firstly, we discuss the construction of the null tetrad, $(k^\alpha, p^\alpha, t^\alpha, t^{*\alpha})$ that satisfies the requirements in the main text. It turns out that the relevant quantities are easier to evaluate in the locally non-rotating frame (LNRF) [9, 46, 47], which relates to the Boyer-Lindquist coordinate via the one-forms

$$\begin{aligned} \omega^{(0)} = \sqrt{\frac{\Sigma\Delta}{A}} dt, \ \omega^{(1)} = \sqrt{\frac{\Sigma}{\Delta}} dr, \\ \omega^{(2)} = \sqrt{\Sigma} d\theta, \ \omega^{(3)} = d\phi - \frac{2ar}{A} dt \ , \end{aligned} \tag{3.29}$$

where $A = (r^2 + a^2)^2 - a^2 \Delta \sin^2\theta$. In this text we use indices with parenthesis to indicate quantities evaluated in this LNRF.

We first start with an arbitrary orthonormal null tetrad

$$\begin{aligned} l^{(a)} &= \frac{1}{\sqrt{2}}(1, -1, 0, 0) \\ n^{(b)} &= \frac{1}{\sqrt{2}}(1, 1, 0, 0) \\ m^{(a)} &= \frac{1}{\sqrt{2}}(0, 0, 1, i) \\ m^{*(a)} &= \frac{1}{\sqrt{2}}(0, 0, 1, -i) \ . \end{aligned} \tag{3.30}$$

We adapt this tetrad to the geodesic tangent vector, such that $l^{(a)} \to k^{(a)}$, and

$$m^{(a)} \to t_*^{(a)} = \frac{1}{\sqrt{2}}(-\Gamma, \Gamma, 1, i) \ , \tag{3.31}$$

where $\Gamma = k^{(a)} m_{(b)} / k^{(c)} l_{(c)}$.

We apply a null rotation such that t_* becomes purely spatial for an observer

$$t_*^{(a)} \longrightarrow t_+^{(a)} = t_*^{(a)} - \left(\frac{t_*^{(b)} u_{o,(b)}}{k^{(c)} u_{o,(c)}} \right) , \tag{3.32}$$

where u_o is the observer's four-velocity. In this work, we choose the set of observers at rest in the LNRF.

Finally, we apply a rotation such that the Sachs' equations are simplified, i.e., the Newman-Penrose scalar $\epsilon \equiv -\frac{1}{2} t_{\alpha;\beta} t^{*\alpha} k^{\beta} = 0$ [46] (semicolon indicates covariant differentiation),

$$t_+^{(a)} \longrightarrow t^{(a)} = t_+^{(a)} e^{-i\chi} , \tag{3.33}$$

where

$$\dot{\chi} = -2i\epsilon_+, \ \epsilon_+ = -\frac{1}{2} t_{+(a);(b)} t_+^{*(a)} k^{(b)} . \tag{3.34}$$

Note that, as in the main text, the overhead dot indicates derivative with respect to the affine parameter.

We now comment briefly on the connection between the formalism and the Sachs' equations with the optical scalars ρ, σ. Since partial derivatives commute, the vector ζ^{α} satisfies

$$k^{\alpha} \nabla_{\alpha} \zeta^{\beta} - \zeta^{\alpha} \nabla_{\alpha} k^{\beta} = 0 . \tag{3.35}$$

Plugging in the decomposition of ζ^{α}, we find

$$\begin{aligned} \dot{\zeta} &= t^a k_{a;b} t^{*b} \zeta + t^a k_{a;b} t^b \zeta^* \\ &\equiv -\rho(\tau)\zeta - \sigma(\tau)\zeta^* . \end{aligned} \tag{3.36}$$

Taking another derivative, we obtain the geodesic deviation equation Eqn. (3.7) with the form

$$\dot{\rho} = \rho^2 + \sigma\sigma^*, \ \dot{\sigma} = \sigma(\rho^* + \rho) + \Psi_0 . \tag{3.37}$$

References

[1] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116 (6):061102, February 2016. doi: 10.1103/PhysRevLett.116.061102.

[2] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough,

B. Agarwal, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, others, LIGO Scientific, and Virgo Collaboration. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Phys. Rev. Lett.*, 118(22):221101, June 2017. doi: 10.1103/PhysRevLett.118.221101.

[3] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, others, LIGO Scientific Collaboration, and Virgo Collaboration. Tests of general relativity with the binary black hole signals from the LIGO-Virgo catalog GWTC-1. *Phys. Rev. D*, 100(10):104036, November 2019. doi: 10.1103/PhysRevD.100.104036.

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

[5] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen, others, LIGO Scientific Collaboration, and Virgo Collaboration. GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object. *ApJ*, 896(2):L44, June 2020. doi: 10.3847/2041-8213/ab960f.

[6] Paolo Amore, Mayra Cervantes, Arturo de Pace, and Francisco M. Fernández. Gravitational lensing from compact bodies: Analytical results for strong and weak deflection limits. *Phys. Rev. D*, 75(8):083005, April 2007. doi: 10.1103/PhysRevD.75.083005.

[7] Nils Andersson. Scattering of massless scalar waves by a schwarzschild black hole: A phase-integral study. *Phys. Rev. D*, 52:1808–1820, Aug 1995. doi: 10.1103/PhysRevD.52.1808. URL https://link.aps.org/doi/10.1103/PhysRevD.52.1808.

[8] Srimanta Banerjee, Chandrachur Chakraborty, and Sudip Bhattacharyya. Alignment and precession of a black hole misaligned with its accretion disc: application to low-mass X-ray binaries. *MNRAS*, 487(3):3488–3504, August 2019. doi: 10.1093/mnras/stz1518.

[9] James M. Bardeen, William H. Press, and Saul A. Teukolsky. Rotating Black Holes: Locally Nonrotating Frames, Energy Extraction, and Scalar Synchrotron Radiation. *ApJ*, 178:347–370, December 1972. doi: 10.1086/151796.

[10] Jillian M. Bellovary, Mordecai-Mark Mac Low, Barry McKernan, and K. E. Saavik Ford. Migration Traps in Disks around Supermassive Black Holes. *ApJ*, 819(2):L17, March 2016. doi: 10.3847/2041-8205/819/2/L17.

[11] Ezequiel F. Boero and Osvaldo M. Moreschi. Strong gravitational lens image of the M87 black hole with a simple accreting matter model. *MNRAS*, 507(4): 5974–5990, November 2021. doi: 10.1093/mnras/stab2336.

[12] V. Bozza. Quasiequatorial gravitational lensing by spinning black holes in the strong field limit. *Phys. Rev. D*, 67(10):103006, May 2003. doi: 10.1103/PhysRevD.67.103006.

[13] V. Bozza. Comparison of approximate gravitational lens equations and a proposal for an improved new one. *Phys. Rev. D*, 78(10):103005, November 2008. doi: 10.1103/PhysRevD.78.103005.

[14] V. Bozza. Optical caustics of Kerr spacetime: The full structure. *Phys. Rev. D*, 78(6):063014, September 2008. doi: 10.1103/PhysRevD.78.063014.

[15] Valerio Bozza. Gravitational lensing by black holes. *General Relativity and Gravitation*, 42(9):2269–2300, September 2010. doi: 10.1007/s10714-010-0988-2.

[16] Xian Chen and Wen-Biao Han. Extreme-mass-ratio inspirals produced by tidal capture of binary black holes. *Communications Physics*, 1(1):53, September 2018. doi: 10.1038/s42005-018-0053-0.

[17] Xian Chen, Shuo Li, and Zhoujian Cao. Mass-redshift degeneracy for the gravitational-wave sources in the vicinity of supermassive black holes. *MNRAS*, 485(1):L141–L145, May 2019. doi: 10.1093/mnrasl/slz046.

[18] Walter Del Pozzo, Katherine Grover, Ilya Mandel, and Alberto Vecchio. Testing general relativity with compact coalescing binaries: comparing exact and predictive methods to compute the Bayes factor. *Classical and Quantum Gravity*, 31(20):205006, October 2014. doi: 10.1088/0264-9381/31/20/205006.

[19] Sam R Dolan. Geometrical optics for scalar, electromagnetic and gravitational waves in curved spacetime. *arXiv e-prints*, art. arXiv:1806.08617, June 2018.

[20] Daniel J. D'Orazio and Abraham Loeb. Repeated gravitational lensing of gravitational waves in hierarchical black hole triples. *Phys. Rev. D*, 101(8): 083031, April 2020. doi: 10.1103/PhysRevD.101.083031.

[21] Ernesto F. Eiroa and Diego F. Torres. Strong field limit analysis of gravitational retrolensing. *Phys. Rev. D*, 69(6):063004, March 2004. doi: 10.1103/PhysRevD.69.063004.

[22] Event Horizon Telescope Collaboration, Kazunori Akiyama, Antxon Alberdi, Walter Alef, Keiichi Asada, Rebecca Azulay, Anne-Kathrin Baczko, et al. First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. *ApJ*, 875(1):L1, April 2019. doi: 10.3847/2041-8213/ab0ec7.

[23] Event Horizon Telescope Collaboration, Kazunori Akiyama, Antxon Alberdi, Walter Alef, Juan Carlos Algaba, Richard Anantua, Keiichi Asada, et al. First Sagittarius A* Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole in the Center of the Milky Way. *ApJ*, 930(2):L12, May 2022. doi: 10.3847/2041-8213/ac6674.

[24] Han Gil Choi, Chanung Park, and Sunghoon Jung. Small-scale shear: Peeling off diffuse subhalos with gravitational waves. *arXiv e-prints*, art. arXiv:2103.08618, March 2021. doi: 10.48550/arXiv.2103.08618.

[25] Samuel E. Gralla and Alexandru Lupsasca. Lensing by Kerr black holes. *Phys. Rev. D*, 101(4):044031, February 2020. doi: 10.1103/PhysRevD.101.044031.

[26] Samuel E. Gralla and Alexandru Lupsasca. Null geodesics of the Kerr exterior. *Phys. Rev. D*, 101(4):044032, February 2020. doi: 10.1103/PhysRevD.101.044032.

[27] Samuel E. Gralla, Alexandru Lupsasca, and Daniel P. Marrone. The shape of the black hole photon ring: A precise test of strong-field general relativity. *Phys. Rev. D*, 102(12):124004, December 2020. doi: 10.1103/PhysRevD.102.124004.

[28] Shahar Hadar, Michael D. Johnson, Alexandru Lupsasca, and George N. Wong. Photon ring autocorrelations. *Phys. Rev. D*, 103(10):104038, May 2021. doi: 10.1103/PhysRevD.103.104038.

[29] Daniel E. Holz and John A. Wheeler. Retro-MACHOs: π in the Sky? *ApJ*, 578(1):330–334, October 2002. doi: 10.1086/342463.

[30] Wen-Rui Hu and Yue-Liang Wu. The Taiji Program in Space for gravitational wave physics and the nature of gravity. *National Science Review*, 4(5):685–686, 10 2017. ISSN 2095-5138. doi: 10.1093/nsr/nwx116. URL https://doi.org/10.1093/nsr/nwx116.

[31] Oliver James, Eugénie von Tunzelmann, Paul Franklin, and Kip S. Thorne. Gravitational lensing by spinning black holes in astrophysics, and in the movie Interstellar. *Classical and Quantum Gravity*, 32(6):065001, March 2015. doi: 10.1088/0264-9381/32/6/065001.

[32] Seiji Kawamura, Masaki Ando, Naoki Seto, Shuichi Sato, Takashi Nakamura, Kimio Tsubono, et al. The Japanese space gravitational wave antenna: DECIGO. *Classical and Quantum Gravity*, 28(9):094011, May 2011. doi: 10.1088/0264-9381/28/9/094011.

[33] A. R. King, S. H. Lubow, G. I. Ogilvie, and J. E. Pringle. Aligning spinning black holes and accretion discs. *MNRAS*, 363(1):49–56, October 2005. doi: 10.1111/j.1365-2966.2005.09378.x.

[34] R. Kormann, P. Schneider, and M. Bartelmann. Isothermal elliptical gravitational lens models. *A&A*, 284:285–299, April 1994.

[35] Kevin A. Kuns, Hang Yu, Yanbei Chen, and Rana X. Adhikari. Astrophysics and cosmology with a decihertz gravitational-wave detector: Tiango. *Phys. Rev. D*, 102:043001, Aug 2020. doi: 10.1103/PhysRevD.102.043001. URL https://link.aps.org/doi/10.1103/PhysRevD.102.043001.

[36] Kwun-Hang Lai, Otto A. Hannuksela, Antonio Herrera-Martín, Jose M. Diego, Tom Broadhurst, and Tjonnie G. F. Li. Discovering intermediate-mass black hole lenses through gravitational wave lensing. *Phys. Rev. D*, 98(8):083005, October 2018. doi: 10.1103/PhysRevD.98.083005.

[37] Jun Luo, Li-Sheng Chen, Hui-Zong Duan, Yun-Gui Gong, Shoucun Hu, Jianghui Ji, Qi Liu, et al. TianQin: a space-borne gravitational wave detector. *Classical and Quantum Gravity*, 33(3):035010, February 2016. doi: 10.1088/0264-9381/33/3/035010.

[38] Richard A. Matzner, Cécile DeWitte-Morette, Bruce Nelson, and Tian-Rong Zhang. Glory scattering by black holes. *Phys. Rev. D*, 31:1869–1878, Apr 1985. doi: 10.1103/PhysRevD.31.1869. URL https://link.aps.org/doi/10.1103/PhysRevD.31.1869.

[39] Barry McKernan, K. E. Saavik Ford, J. Bellovary, N. W. C. Leigh, Z. Haiman, B. Kocsis, W. Lyra, M. M. Mac Low, B. Metzger, M. O'Dowd, S. Endlich, and D. J. Rosen. Constraining Stellar-mass Black Hole Mergers in AGN Disks Detectable with LIGO. *ApJ*, 866(1):66, October 2018. doi: 10.3847/1538-4357/aadae5.

[40] Ashish Kumar Meena and Jasjeet Singh Bagla. Gravitational lensing of gravitational waves: wave nature and prospects for detection. *MNRAS*, 492(1): 1127–1134, February 2020. doi: 10.1093/mnras/stz3509.

[41] Suvodip Mukherjee, Benjamin D. Wandelt, and Joseph Silk. Probing the theory of gravity with gravitational lensing of gravitational waves and galaxy surveys. *MNRAS*, 494(2):1956–1970, May 2020. doi: 10.1093/mnras/staa827.

[42] Ryan M. O'Leary, Bence Kocsis, and Abraham Loeb. Gravitational waves from scattering of stellar-mass black holes in galactic nuclei. *MNRAS*, 395(4): 2127–2146, June 2009. doi: 10.1111/j.1365-2966.2009.14653.x.

[43] Kris Pardo, Maya Fishbach, Daniel E. Holz, and David N. Spergel. Limits on the number of spacetime dimensions from GW170817. *J. Cosmology Astropart. Phys.*, 2018(7):048, July 2018. doi: 10.1088/1475-7516/2018/07/048.

[44] A. R. Patrick, J. N. Reeves, D. Porquet, A. G. Markowitz, V. Braito, and A. P. Lobban. A Suzaku survey of Fe K lines in Seyfert 1 active galactic

nuclei. *MNRAS*, 426(3):2522–2565, November 2012. doi: 10.1111/j.1365-2966.2012.21868.x.

[45] Scott Perkins and Nicolás Yunes. Probing screening and the graviton mass with gravitational waves. *Classical and Quantum Gravity*, 36(5):055013, March 2019. doi: 10.1088/1361-6382/aafce6.

[46] Serge Pineault and R. C. Roeder. Applications of Geometrical Optics to the Kerr Metric. Analytical Results. *ApJ*, 212:541–549, March 1977. doi: 10.1086/155073.

[47] Serge Pineault and R. C. Roeder. Applications of Geometrical Optics to the Kerr Metric. 11. Numerical Results. *ApJ*, 213:548–557, April 1977. doi: 10.1086/155186.

[48] William H. Press and Saul A. Teukolsky. Adaptive stepsize runge-kutta integration. *Computers in Physics*, 6(2):188–191, 1992. doi: 10.1063/1.4823060. URL https://aip.scitation.org/doi/abs/10.1063/1.4823060.

[49] Christopher S. Reynolds. The spin of supermassive black holes. *Classical and Quantum Gravity*, 30(24):244004, December 2013. doi: 10.1088/0264-9381/30/24/244004.

[50] Peter Schneider, Jürgen Ehlers, and Emilio E. Falco. *Gravitational Lenses*. 1992. doi: 10.1007/978-3-662-03758-4.

[51] Amy Secunda, Jillian Bellovary, Mordecai-Mark Mac Low, K. E. Saavik Ford, Barry McKernan, Nathan W. C. Leigh, Wladimir Lyra, and Zsolt Sándor. Orbital Migration of Interacting Stellar Mass Black Holes in Disks around Supermassive Black Holes. *ApJ*, 878(2):85, June 2019. doi: 10.3847/1538-4357/ab20ca.

[52] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen, et al. Properties and astrophysical implications of the 150 Msun binary black hole merger GW190521. *arXiv e-prints*, art. arXiv:2009.01190, September 2020.

[53] Aditya Vijaykumar, M. V. S. Saketh, Sumit Kumar, Parameswaran Ajith, and Tirthankar Roy Choudhury. Probing the large scale structure using gravitational-wave observations of binary black holes. *arXiv e-prints*, art. arXiv:2005.01111, May 2020.

[54] D. J. Walton, E. Nardini, A. C. Fabian, L. C. Gallo, and R. C. Reis. Suzaku observations of 'bare' active galactic nuclei. *MNRAS*, 428(4):2901–2920, February 2013. doi: 10.1093/mnras/sts227.

[55] Yijun Wang, Rico K. L. Lo, Alvin K. Y. Li, and Yanbei Chen. Identifying type II strongly lensed gravitational-wave images in third-generation gravitational-wave detectors. *Phys. Rev. D*, 103(10):104055, May 2021. doi: 10.1103/PhysRevD.103.104055.

[56] Kent Yagi and Naoki Seto. Detector configuration of DECIGO/BBO and identification of cosmological neutron-star binaries. *Phys. Rev. D*, 83(4): 044011, February 2011. doi: 10.1103/PhysRevD.83.044011.

[57] Hang Yu and Yanbei Chen. Direct determination of supermassive black hole properties with gravitational-wave radiation from surrounding stellar-mass black hole binaries. *Phys. Rev. Lett.*, 126:021101, Jan 2021. doi: 10.1103/PhysRevLett.126.021101. URL https://link.aps.org/doi/10.1103/PhysRevLett.126.021101.

[58] Hang Yu, Yijun Wang, Brian Seymour, and Yanbei Chen. Detecting gravitational lensing in hierarchical triples in galactic nuclei with space-borne gravitational-wave observatories. *Phys. Rev. D*, 104(10):103011, November 2021. doi: 10.1103/PhysRevD.104.103011.

Chapter 4

GRAVITATIONAL-WAVE DETECTION WITH PHOTOMETRIC SURVEYS

Wang, Y., Pardo, K., Doré, O., and Chang, T.-C. (2021). "Gravitational-wave Detection with Photometric Surveys". In: *Phys.Rev.D*, 103(8):084007. DOI:10.1103/PhysRevD.103.084007.

4.1 Introduction

The successful detection of gravitational wave (GW) signals from binary mergers with the Advanced Laser Interferometer Gravitational-wave Observatory (Advanced LIGO) and Virgo collaboration (see e.g. [5, 54]) has spurred great interest in improving detection sensitivity and developing independent detection methods. For GW astronomy, it is crucial that we have access to GWs across as wide a frequency spectrum as possible, since different frequency bands are sensitive to their respective groups of GW sources. A continuous frequency band also allows for observation of the same GW source as it evolves to higher frequencies towards merger, allowing us to extract as much information as possible.

The Advanced LIGO is sensitive to GWs between 10 Hz and 7 kHz [4], ideal for detecting solar-mass binary mergers. The deci-hertz band will be covered by space-based detectors such as TianGo [33] and DECIGO [29], targeting intermediate-mass black hole binaries ($\sim 10^2 - 10^4\ M_\odot$ [33]). The milli-hertz band will be covered by the space-based Laser Interferometer Space Antenna (LISA) [7] and TianQin [36]. These interferometer-type detectors directly measure the GW-induced change in separation between either suspended or free falling mirrors. In this case, the detector frequency range is limited by noise factors, such as mirror position alignment error, quantum noise and thermal noise [see, e.g., 48]. The space-based detectors are sensitive to massive black hole mergers (MBHMs) at high redshifts (e.g. LISA can detect $10^5\ M_\odot$ mergers at $z \sim 15$ with an SNR of ~ 100 in the ringdown stage [7]). Observing MBHMs will be instrumental for modeling black hole evolution history and understanding strong-field gravity features [55].

At lower frequencies, interferometer-type detectors are no longer available and there is a frequency gap until the Pulsar Timing Array (PTA) detection method becomes

applicable. PTAs measure the Time of Arrival (TOA) of pulses from stable millihertz pulsars. Passing GWs modify the pulse frequency, which translates into a timing residual signal. By cross-correlating timing residuals from pairs of pulsars, GW parameters can be extracted [12, 21, 42]. The PTA frequency band is limited by mission lifetime as well as the observational cadence. For example, a 5-year survey with an observational frequency of 17 year^{-1} ($\sim 1/3$ week^{-1}) [53, 59] is sensitive to GWs from 6.3×10^{-9} Hz to 5.4×10^{-7} Hz. With longer signal integration time and more pulsar pairs, PTAs can detect the supermassive black hole merger background (SMBHMB) as well as individual supermassive black hole binaries (SMBHBs) with chirp mass between 10^4 $M_\odot$ and 10^{10} $M_\odot$ [49]. Recently, significant evidence for a signal with common amplitude and spectral slope across monitored pulsars was recovered from the 12.5-yr data by the North American Nanohertz Observatory for Gravitational Wave (NANOGrav). However, there was no statistically strong evidence for the quadrupolar spatial correlation expected from a GW background in the General Relativity framework, and it remains to be determined if the observed signal is indeed astrophysical [10].

We can also detect GWs via astrometry [17, 46]. Analogous to the theoretical basis for PTAs, passing GWs perturb photon trajectories as they travel from the observed stars to the detector. This perturbation leaves a GW-specific change to the apparent star positions. It is, in principle, possible to extract this change in position from high-precision astrometric data. Similar to the PTA method, the sensitive frequency range depends on both survey lifetime and observational cadence. Accordingly, using astrometric measurements as GW probes is a highly flexible technique since observational frequency is tunable, depending on mission design. Furthermore, we demonstrate in the paper that these GW measurements can be made with relative astrometry and do not require dedicated absolute astrometric missions.

A suitable photometric survey with cadence higher than PTAs can unlock the intermediate frequency band between PTAs and LISA. A survey with such sensitivity range would be able to detect massive black hole binaries from 10^5 $M_\odot$ to 10^9 $M_\odot$ during inspiral and close to merger. Examples of such sources include the highly eccentric binaries that go near coalescence in the sensitivity range of LISA [14, 15]. Detecting these GW sources will add invaluable data for constraining black hole evolution models.

Accessing this frequency range also opens up opportunities for joint analysis of a GW source population using various GW detectors targeting their respective frequency

range. Although studies of a variety of massive binary black hole assembly scenarios suggest that most GW sources in the nanohertz band show little frequency drift on the scale of 10 years [50, 51], some sources detectable by LISA start emitting GWs with a potentially detectable signal strength at frequencies lower than the LISA sensitivity limit, such as highly eccentric binaries ejected from stellar clusters due to natal kicks or dynamical processes [14, 15]. It is unlikely to observe one GW source migrate across the frequency spectrum, since the inspiral time for sources at the low frequency limit of LISA can be on the order of gigayears [15]. However, observing the same population at these different frequencies allows us to piece together ensemble source properties and their evolution.

This astrometric GW detection method in the context of *Gaia* has been studied in detail [31, 43]. *Gaia* as a GW probe is sensitive from $10^{-8.5}$ Hz to 10^{-6} Hz; at $f > 10^{-7.5}$ Hz, *Gaia* will outperform PTA efforts [43]. In this paper, we discuss how this analysis can in principle be done with astrometric data from any photometric surveys even though they may not provide absolute astrometric measurements as *Gaia* does. As a specific example, we forecast the GW detection sensitivity of the Nancy Grace Roman Space Telescope[1], NASA's next flagship observatory after the James Webb Space Telescope.

The Roman Space Telescope will observe billions of galaxies and thousands of supernovae to probe the time evolution of dark energy and large-scale structure [see, e.g., 6]. It will perform a micro-lensing survey on the inner Milky Way, as well as high contrast imaging and spectroscopic studies of individual close-by exoplanets [3]. For GW detection, its notional Exoplanet MicroLensing (EML) survey is particularly relevant. It is expected to observe 10^8 stars in 7 fields [23]. It operates in the near-IR with a $\sim 0.281\,\mathrm{deg}^2$ field of view (FoV), with an estimated single-exposure astrometric precision of 1.1 mas [62]. During its nominal lifetime of 5 years, it will survey a total area of $1.97\,\mathrm{deg}^2$ between Galactic longitudes of -0.5 deg and 1.5 deg, and Galactic latitudes between -0.5 deg and -2 deg. Observational time consists of six 72-day seasons. During each season, the Roman Space Telescope visits the seven fields sequentially and repeats this cycle every 15 minutes. This gives a maximum of $\sim 41,000$ exposures per source, making it "one of the deepest exposures of the sky ever taken" [23].

In this paper, we begin by reviewing the theory for GW-induced astrometric deflections and outlining the general strategy for using photometric surveys as GW

[1] `https://roman.gsfc.nasa.gov/`

probes. We then assess the potential of the Roman EML survey to detect individual binary signals. In Section 4.4, we discuss directions for performance improvement for photometric surveys similar to the Roman EML survey as GW probes. We then expand to other telescopes and surveys and discuss their potential for astrometrically detecting GWs.

All of the code used to produce the figures and analysis in this paper is available at: `https://github.com/kpardo/estoiles-public`.

4.2 Photometric Surveys as GW Probes

In this section, we first summarize how GW signatures manifest as observable variation in the astrometric solution. We then present estimates of the sensitivity of photometric surveys to GWs as well as their frequency resolution.

GW Signature in Astrometry

In short, a passing GW perturbs the spacetime along the photon trajectory as it travels from the observed star to the detector. This perturbation causes a shift in the stellar apparent position from its true position. Theoretical details are derived in [46] in the distant source limit and later generalized in [17]. Here we present a brief summary, closely following steps in [17].

We start with the model where the GW source and observer are stationary in Minkowski spacetime and the GW is a linear perturbation to flat spacetime. Throughout this paper, we use Greek alphabet to denote components of 4-vectors and Latin alphabet to denote the spatial dimensions. Indices that appear both as upper and lower indices imply summation over all dimensions. We also adopt the transverse-traceless gauge. Under this gauge condition, components of the perturbation tensor, $h_{\mu\nu}$, can be non-zero only when both indices are spatial, and the tensor trace is 0, i.e.,

$$h_{0\mu} = 0, \; h^{\mu}_{\mu} = 0 \; .$$

The metric can then be written as,

$$ds^2 = -dt^2 + (\delta_{ij} + h_{ij}) dx^i dx^j \; . \tag{4.1}$$

We can write the photon trajectory as,

$$x^{\alpha}(\lambda) = x^{\alpha}_{(0)}(\lambda) + x^{\alpha}_{(1)}(\lambda) \; , \tag{4.2}$$

where subscript (0) indicates quantities in unperturbed spacetime, and subscript (1) indicates first order corrections. λ is the associated affine parameter. We calculate

the Christoffel symbols in this metric and write the geodesic equation as,

$$\frac{d^2x^0_{(1)}}{d\lambda^2} = -\frac{\omega_0^2}{2}n^i n^j h_{ij,0} \tag{4.3}$$

$$\frac{d^2x^k_{(1)}}{d\lambda^2} = -\frac{\omega_0^2}{2}\Big[-2n^i h_{ki,0} + n^i n^j\left(h_{ki,j}+h_{kj,i}-h_{ij,k}\right)\Big], \tag{4.4}$$

where ω_0 is the photon frequency without GW perturbation. Integrating the geodesic equation with respect to λ gives the photon trajectory and 4-momentum.

We then compute the GW perturbation in the observer frame. We first construct an orthonormal tetrad, $e_{\hat{\alpha}}$ where $e_{\hat{0}} = \vec{u}$ and $\vec{u}$ is the observer's 4-velocity. We also require this tetrad to be parallel-transported along the observer worldline. Imposing the parallel-transport equation and the metric, we can express the observer tetrad in terms of the GW and the unperturbed basis vectors. The observed photon 4-momentum, $k^{\hat{\alpha}}$, can be found via a coordinate transformation, and its spatial part gives $n^{\hat{i}}$. Assuming small deflections, $dn^{\hat{i}} = n^{\hat{i}} - n^{\hat{i}}_{(0)}$.

It is oftentimes useful to assume monochromatic plane-wave GWs and a distant source, in which case the observed star is many GW wavelengths away from the observer. In the plane-wave model, the integral along geodesics can be done analytically, resulting in some geometrical constant factors and a phase in the form of $e^{-i2\pi f\omega_0(1+\mathbf{p}\cdot\mathbf{n})\lambda_s}$, where $2\pi f$ and ω_0 are frequencies of the GW and the photon, respectively. $\mathbf{p}$ is the GW propagation direction and $\mathbf{n}$ points towards the observed star. In the distant source limit (i.e., $\omega_0\lambda_s \gg c/2\pi f$), prefactors to the integral become negligibly small and we may ignore this term. Consequently, the leading order of the signal depends only on the GW amplitude at the observer. $dn^{\hat{i}}$ is thus much simplified and becomes [17, 46]:

$$dn^{\hat{i}}(t,\mathbf{n}) = \frac{n^i+p^i}{2\,(1+\mathbf{p}\cdot\mathbf{n})}h_{jk}(t,\mathbf{0})n^j n^k - \frac{1}{2}h^{ij}(t,\mathbf{0})n_j\,, \tag{4.5}$$

$$h_{ij}(t,\mathbf{x}) = \mathrm{Re}\left[\mathcal{H}_{ij}e^{-i2\pi f(t-\mathbf{p}\cdot\mathbf{x})}\right], \tag{4.6}$$

where $\mathcal{H}_{ij}$ denotes the plane wave amplitude tensor. The distant source approximation is also adopted in PTA analyses, where the integral from the pulsar to the observer is reduced to consideration about the two end points only (see, e.g.,[8]). For PTA analyses, an additional reason to drop the GW term at the pulsar is that

such a signal would be uncorrelated between different pulsars, whereas the GW perturbation at the detector is shared. When we consider the correlation between timing residuals, these pulsar perturbation terms can thus be treated as random noise [42]. In Section 4.3 we discuss the validity of this assumption in our work.

For small astrometric deflections, it suffices to consider the leading order of $\mathcal{H}_{ij}$ [see e.g. 16],

$$\mathcal{H}_{ij} = A H_{ij}(\mathbf{p}) \tag{4.7}$$

$$A_{(0)} = \frac{2G^{5/3}}{c^4}(\pi f)^{2/3}\frac{\mathcal{M}_c^{5/3}}{D_L} \sim f^{2/3}\mathcal{M}_s^{5/3} \ . \tag{4.8}$$

$H_{ij}(\mathbf{p})$ is the polarization tensor for GWs propagating along $\mathbf{p}$. $A_{(0)}$ is the leading term of the GW amplitude, A, which depends on the source frame GW frequency, f, the chirp mass, $\mathcal{M}_c$, and the luminosity distance, D_L. $\mathcal{M}_c$ is defined as $m\left(q/(1+q)^2\right)^{3/5}$, where m is the total mass of the binary and q is the mass ratio, m_1/m_2, assuming m_1 is the smaller mass. For GW sources not at cosmological distances (i.e., redshift $z \ll 1$), we may ignore the cosmological redshift to the wave frequency. Throughout this paper, we always assume such close-by sources and we do not differentiate between source frame and observer frame GW frequency. Our threshold GW source estimates validate this assumption. To this order, we note that scaling $\mathcal{M}_c$ by an arbitrary factor κ is completely degenerate with scaling D_L by $\kappa^{5/3}$. Therefore, it is convenient to define a scaled mass $M_s \equiv \mathcal{M}_c/D_L^{3/5}$, which represents all sources that give the same leading order GW signal, at a fixed frequency.

In Figure 4.1 we reproduce Figure 1 in Ref. [43] and illustrate the astrometric deflection pattern for a field of stars in the northern hemisphere in Galactic coordinates, due to a face-on GW source at zenith. It is clear that the deflection magnitude is largest on the Galactic plane. Deflections induced by the plus and cross polarizations are orthogonal, and the quadrupolar pattern is clear. The right panels show the astrometric deflection in a square Field of View (FoV), assuming the telescope is in the Galactic plane and points directly to the Galactic center. This FoV model has roughly the same area as the true FoV of the Roman Space Telescope but differs in shape. We adopt it nonetheless in our analysis for simplicity.

The bottom panel shows the total deflection pattern while the upper panel shows the deflection pattern after subtracting the mean deflection. This is expected to be the actual observed signal, as the pointing reconstruction strategy of the Roman Space Telescope will likely absorb deflections uniform across the FoV. A measure of

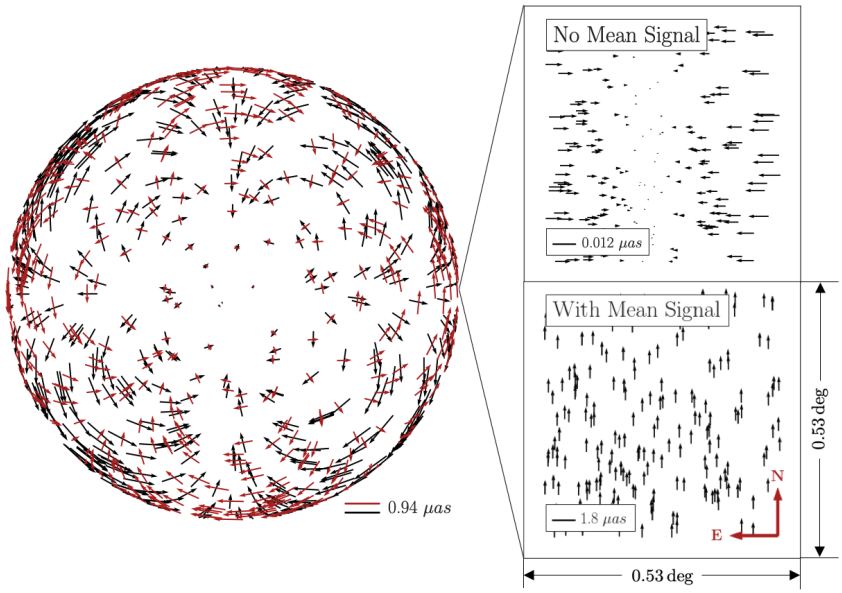

Figure 4.1: Illustration of the expected stellar astrometric deflections. *Left:* Orthographically projected dn for a subset of stars observed by *Gaia* in the northern hemisphere onto the galactic plane (inspired by a very similar plot in Ref. [43]). The North Galactic pole is at the center which is also the position of the GW source. Black arrows correspond to the real part of the waveform at GW phase $\phi = 0$ (plus polarization), and the red arrows correspond to that at $\phi = \pi/4$ (cross polarization). The source is a $10^9\ M_\odot$ equal-mass binary black hole at 1 Mpc at $(l = 90\,\text{deg}, b = 90\,\text{deg})$ in galactic coordinates, emitting GWs at 10^{-6} Hz. This inclination angle is set to $i = 0$ (i.e. face-on) and the polarization angle is $\psi = 0$. *Right:* Deflections within the Roman Space Telescope's FOV during the EML survey. The lower panel shows the total deflection, and the upper panel shows the deflection after subtracting the mean, since the mean is expected to be absorbed in the pointing reconstruction; for further discussion see Section 4.4. Star coordinates are selected from the Gaia Data Release 2 catalog, with brightness $0 < G < 9$ [18, 45]. Density of stars reflects only a subset of the true stellar density in the catalog.

the magnitude of the mean-subtracted deflections is the divergence of dn integrated across the FoV, since the mean deflection field has zero divergence. For the particular GW source position and telescope pointing in Figure 4.1, we compute the integrated divergence of the astrometric deflection in Eqn. (4.5) to be Al_{FoV}^2 assuming small FoV side length l_{FoV}, where A is the GW amplitude. From the top right panel in Figure 4.1 and the divergence theorem, the integrated divergence is proportional to $l_{\text{FoV}}\langle|dn_{\text{ms}}|\rangle$, where $\langle|dn_{\text{ms}}|\rangle$ is the average magnitude of the mean-subtracted deflections. This scaling relation is confirmed numerically using various l_{FoV}.

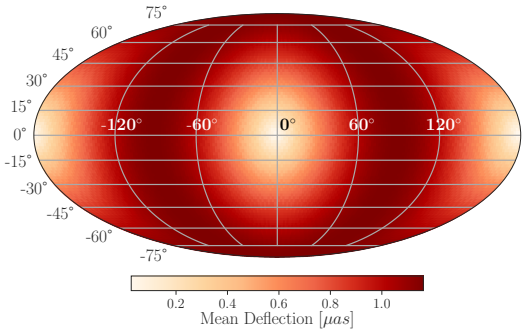

Figure 4.2: Mean maximal deflection due to GW sources at different Galactic coordinates. The FoV is fixed to point towards the Galactic center, i.e. $l = 0$ deg, $b = 0$ deg. The magnitude is calculated as that of the orthographic projection of dn in the FoV, averaged over all observed stars. The maximum occurs when the source position vector is perpendicular to the star position vectors, which is consistent with Eqn. (4.5). Aside from its coordinates, the GW source at each position has the same properties as that in Figure 4.1. In combination with Figure 4.1, we observe that the quadrupolar deflection pattern does not show up when we consider signal magnitude only.

For any small-FoV, relative-astrometry telescope, we may estimate the observable deflection signal by

$$\langle |dn_{\rm ms}| \rangle \approx \frac{l_{\rm FoV}}{l_{\rm FoV,RST}} \langle |dn_{\rm ms,RST}| \rangle \ , \tag{4.9}$$

where the RST subscript denotes parameter values in the case of the Roman Space Telescope. For example, in a FoV similar to the Hubble Space Telescope ($l_{\rm FoV} \approx$ 2.4 arcmin [13]), the mean magnitude of the mean-subtracted deflections is only 7.4% of that in a FoV similar to the Roman Space Telescope, which has a $l_{\rm FoV} \approx$ 32 arcmin. For further discussions see Section 4.4.

The magnitude of the astrometric deflection as a function of the GW source position on the sky is shown in Figure 4.2. We assume the telescope FoV points to the Galactic center. Properties of this GW source are the same as in Figure 4.1. For illustration purpose, we fix the polarization angle to be 0. The mean deflection is averaged over 1000 randomly distributed stars within the FoV (the number of stars is not representative of the actual stellar density; it is picked for clear visualization). The deflection is maximal when the GW source position is orthogonal to observed star positions, which is consistent with Eqn. (4.5) and Figure 4.1.

Sensitivity Curve Estimate

For single exposures, the astrometric accuracy, $\Delta\theta$, is determined by pixel size and pixel placement error [62]. Typically, astrometric deflections due to GWs are small compared to any realistic single-exposure resolution values, therefore they cannot be resolved from isolated measurements of individual stars. This limit, however, can be statistically improved by considering repeated observation of a vast collection of stars.

Firstly, within each exposure, we consider the correlated astrometric deflection between N_s stars, which improves the astrometric resolution by $\sqrt{N_s}$. Secondly, if the same stars are measured for N_m times throughout the survey, the measurement error is reduced by $\sqrt{N_m}$, which transforms the single-measurement resolution to the end-of-survey resolution. Assuming the same observational cadence throughout the survey, $N_m = T_{\rm obs}/\Delta t$, where $T_{\rm obs}$ is the total observation time and $1/\Delta t$ is the observational cadence. The minimum detectable GW amplitude is then

$$h = \frac{\Delta\theta}{\sqrt{N_s N_m}} \,. \tag{4.10}$$

These two statistical improvements are subject to survey-specific constraints. In Section 4.3, we provide further discussion on this limit in the context of the Roman EML survey. Throughout our analysis, we assume that, within certain limits, high frequency oscillations are sampled equally well as low frequency ones, i.e. without including discrete sampling effects. In reality, the discrete telescope schedule to visit a sky patch and mission duty cycle impose an upper limit on the maximum number of observable cycles, i.e., the GW frequency, before the deflections become poorly sampled.

We emphasize that this estimated sensitivity is valid only when entire deflection signals are observable. For telescopes taking relative astrometric measurements, such as the Roman Space Telescope, the observable is, in fact, a small fraction of this total signal, which lowers the sensitivity. See Section 4.3 for detailed discussions.

Frequency Resolution

In this subsection, we outline how to calculate the GW frequency resolution of a photometric survey. From the instrument perspective, the frequency resolution is determined by the exposure timing accuracy, σ_t. To calculate how σ_t translates into end-of-mission frequency resolution, Δf, we model the GW phase, or equivalently, the phase of the astrometric deflection signal, as a quadratic function of the time of

exposures,

$$\begin{aligned}\phi &= \phi_0 + \frac{d\phi}{dt}t + \frac{1}{2}\frac{d^2\phi}{dt^2}t^2 \\ &= \phi_0 + 2\pi f t + \pi \dot{f} t^2 \ .\end{aligned} \tag{4.11}$$

Using Fisher information theory, the uncertainty of the coefficients are

$$\begin{aligned}\sigma(f) &\approx \frac{\sqrt{12}}{T_{\rm obs}}\frac{f\sigma_t}{\sqrt{N_m}} \\ \sigma(\dot{f}) &\approx \frac{\sigma(f)}{T_{\rm obs}} \sim \frac{f\sigma_t}{T_{\rm obs}^2\sqrt{N_m}} \ ,\end{aligned} \tag{4.12}$$

where, again, $T_{\rm obs}$ and N_m is the total number of exposures.

The campaign frequency sensitivity is then estimated by

$$\Delta f = \sigma_{\dot{f}} T_{\rm obs} \ . \tag{4.13}$$

To determine whether this resolution is sufficient to capture frequency change of GWs within the detector frequency band, we compare Δf with the frequency evolution of observable sources. The intrinsic inspiral binary frequency is given by (see, e.g., [30])

$$f \sim \frac{1}{m\pi}\left(\frac{1}{4\eta^{1/4}}\left(1+\eta_1\Theta^{-1/4}\right)\right)^{3/2} \ , \tag{4.14}$$

where $\eta \equiv q/(1+q)^2$ and m and q are the total mass of the binary (in natural units where $G = c = 1$) and the mass ratio, as defined before. η_1 is defined as $743/4032 + 11/48\eta$ and Θ as $\eta\left(t_c - t\right)/5m$. Time to coalesce t_c is $5m/\left[\eta\left(8\pi m f\right)^{8/3}\right]$. For systems that remain in inspiral stage throughout survey time, we consider the difference of GW frequencies evaluated at the beginning and the end of the survey. For systems that merger within observational time, we take the final frequency to be the Innermost Stable Circular Orbit frequency, $f_{\rm ISCO}$, beyond which the systems quickly coalesce and Eqn. (4.14) no longer captures the actual frequency.

This characteristic frequency progression could help distinguishing the GW signal from other noise factors and provide additional information for GW source parameter estimation.

4.3 Detecting GWs with the Roman Space Telescope

In this section, we explore the potential of the Roman EML survey as a GW probe. We first discuss its sensitivity frequency range following the method outlined in Section 4.2. Since the procedure is general, we apply a parallel analysis on *Gaia* for comparison. We then describe a method to extract GW signals from photometric data via Bayesian inference. We apply this technique to the Roman Space Telescope and calculate its sensitivity curve.

Roman EML Survey Sensitivity Curve

Similarly to PTAs, the GW frequency band of photometric surveys is constrained by the observation time span and cadence. At the low frequency limit, the GW period should not be longer than the observation time. Signals with longer period are close to being linear over the observational window and thus are likely to be absorbed as telescope motion or proper motion in the astrometric solution. At the high frequency limit, the GW period should not be shorter than twice the observational cadence to satisfy the Nyquist-Shannon sampling condition.

In addition, the low frequency limit is subject to more detailed and survey-specific modifications. Firstly, the low frequency limit where $f_{\rm min} \approx 1/T_{\rm obs}$ can be technically relaxed to $f_{\rm min} \approx 1/2T_{\rm obs}$, since the former limit still produces an oscillatory signal that cannot be fully absorbed by any linear proper motion model [31]. However, we note that this only leads to a factor of two difference, and we ignore this factor when estimating the frequency range. Furthermore, the general guideline works best for uniform sampling, whereas actual surveys may have significant periods of downtime between observational windows. In this case, detecting low frequency GWs requires precisely piecing together high-cadence observation seasons that may be quite separated in time. Deflection change within each season is only a fraction of the total amplitude, and may well be approximated by linear proper motion. Considering the magnitude of the signal and uncertainties from long season-separation, this wave reconstruction process will likely introduce large errors that render further data analysis unfeasible. For a conservative limit, $f_{\rm min}$ is $1/2T_s$ where T_s is the length of one observational season.

Specifically for the Roman EML survey, we assume a 15-minute cadence with six 72-day observational seasons spread out over the nominal 5-year mission time. The previous constraints then give a conservative frequency range as,

$$7.7 \times 10^{-8}\ \text{Hz} < \Omega_{\rm Roman} < 5.6 \times 10^{-4}\ \text{Hz}\ .$$

We also assume a single-exposure astrometric accuracy of $\Delta\theta \sim 1.1$ mas, estimated for $H_{AB} = 21.6$ stars, [62] and a total of $N_s \sim 10^8$ stars with $W145_{AB} < 23$ [23]. We note that all GW signals within the Roman EML survey frequency range have wavelengths smaller than ~ 0.1 pc, which is much smaller than the distance to any stars Roman Space Telescope observes. Therefore we may safely use the distant source limit described in Section 4.2.

We now calculate the frequency resolution, following the procedure in Section 4.2. Taking a conservative timing accuracy of 1 s and $T_{\rm obs} = 6 \times 72$ days (i.e., assuming all seasons happen consecutively), we estimate Δf to be $\sim 10^{-14}$ Hz and $\sim 10^{-12}$ Hz for signals at the lower and upper frequency band limit, respectively. For light systems ($\log_{10} \mathcal{M}_c[M_\odot] = 5.7$), intrinsic frequency change of GWs during the inspiral ranges from 10^{-11} Hz to 10^{-2} Hz, depending on its frequency at the start of the observation. If such a system is initially observed to emit GW at $\sim 6 \times 10^{-5}$ Hz or higher, it coalesces within $T_{\rm obs}$. For heavy systems ($\log_{10} \mathcal{M}_c[M_\odot] = 9.7$), GW frequency change ranges from 10^{-8} Hz to 10^{-6} Hz. Such heavy systems coalesce within the observational time window if they emit GW at $\sim 2 \times 10^{-7}$ Hz at the start of the mission. In all cases, the Roman EML survey will be sensitive to the frequency evolution of detected GWs. We note that it should increase the sensitivity of the Roman EML survey to GWs; however, a full analysis of this effect is outside the scope of this work.

For *Gaia*, assuming 70 evenly-spaced visits of the same stars, uniformly spread out over the nominal 5-year mission time [45], the frequency range is,

$$6.3 \times 10^{-9}\ \text{Hz} < \Omega_{Gaia} < 4.5 \times 10^{-7}\ \text{Hz}\ .$$

This range differs from [31] at the upper limit, since they used the *Gaia* rotational period of $\sim$ 6h as the cadence. Since we are interested in the average sensitivity applicable to the majority of the observed stars, we adopt the more conservative cadence of 70/5-year. We adopt $\Delta\theta \sim 0.7$ mas, which is the parallax uncertainty for $G \sim 20$ stars in *Gaia* Data Release 2 [2][37]. This magnitude threshold value is picked for convenient comparison with the Roman EML survey, where relatively fainter stars could also be observed in the near IR. We assume $N_s \sim 10^9$ [18].

Applying Eqn. (4.10) to the Roman EML survey and *Gaia*, we show their strain sensitivity in Figure 4.3. For the Roman EML survey, the result of Eqn. (4.10) is

[2] In the recent *Gaia* Early Data Release 3, the standard uncertainty in declination at epoch J2016.0 for magnitude $G = 20$ stars is 0.382 mas [35]. The performance estimate we give here will not be drastically different.

shown as the dashed line; it is the sensitivity if the Roman EML survey can capture the mean astrometric deflection signal. The sensitivity with the mean-subtracted signals is shown as the black solid line, assuming an average sensitivity decrease of 100. We reiterate that the Roman Space Telescope takes relative astrometric measurements, recording only the relative positions of objects with each other and across the exposures with its nominal astrometric resolution. The absolute positions will be determined by the guiding stars and the telescope pointing with a larger uncertainty. Consequently, it is this mean-subtracted/relative measurements that constitute the data for extracting GW signals. See Figure 4.5 for an illustration of this scaling. Sensitivity curves for the International Pulsar Timing Array (IPTA) [see 53, 56, 59] and LISA [47] are shown for comparison. It is important to note that, due to the targeted signal type, the sensitivity is represented by different quantities for each detector. For IPTA and LISA, the sensitivity is represented by the characteristic noise strain amplitude, $\tilde{h}_n(f) \equiv \sqrt{f S_n(f)}$, which is a unitless quantity derived from the detector noise power spectral density ($|\tilde{}|$ denotes frequency-domain quantities). For the Roman EML survey and *Gaia*, the sensitivity is plotted as the minimum instantaneous (i.e., time-domain) GW strain amplitude, h, which is estimated based on scaling arguments in Eqn. (4.10). This choice of representation is motivated by the fact that within the frequency sensitivity range of the Roman EML survey and *Gaia*, we expect to see mostly monochromatic GWs. Finally, we reiterate that this estimate method ignores the telescope duty cycle and does not model the effect of having six separated observing seasons on signals at various frequencies.

In Figure 4.3, the colored blocks denote example sources within each detector frequency range. In the IPTA frequency range, the yellow block shows the characteristic strain of the expected supermassive black hole binary background [49], $\tilde{h}_c$(SMBHBB). In the frequency range of LISA, we plot the characteristic strain amplitude of an illustrative GW source with $\mathcal{M}_c = 10^6\ M_\odot, q = 1$ at $D_L = 25$ Gpc. At the low-frequency end, this signal is truncated arbitrarily at 3×10^{-4} Hz for visual clarity. At the high-frequency end, it is stopped at the corresponding $f_{\rm ISCO}$. The slope of the signal is $-1/6$, characteristic of the inspiral-stage of a black hole binary GW [20, 42]. In the frequency band of the Roman EML survey, we show in red and violet blocks the instantaneous GW strain amplitude with chirp masses $10^{9.7}\ M_\odot$ and $10^7\ M_\odot$ at 50 Mpc, respectively. The starting frequency is set arbitrarily to the low frequency limit of the Roman EML survey, and the signal is cut off at $f_{\rm ISCO}$. The instantaneous amplitude scales as $f^{3/2}$ [16, 41]. It should be noted that these individual binary source signals are merely illustrative; they do not represent the

expected GW source population from binary black hole formation theory.

We observe that were the Roman EML survey able to observe the mean deflection signal, it would outperformed *Gaia* at overlapping frequency ranges. This is mainly due to the high cadence observations. However, the mean-subtraction procedure considerably curbs its expected performance. Nonetheless, its high-cadence observations allow for the detection of 10^{-6} Hz – 10^{-5} Hz GWs, which are inaccessible by other dedicated GW observatories, such as PTAs and LISA. In this range, possible GW sources include SMBHBs with $\mathcal{M}_c \sim 10^8\ M_\odot - 10^9\ M_\odot$ at later stages of the inspiral. Due to the larger GW amplitude, such sources will be more detectable than the same GW population earlier in their inspirals, which are targets of PTAs. At the high frequency range, $10^5\ M_\odot$ massive black hole binaries and highly eccentric binaries are at the inspiral stage. LISA, on the other hand, will observe these systems much closer to their coalescence [14, 15]. Observing the different stages of this population offers invaluable data for piecing together their evolution process, emphasizing the potential of the astrometry GW detection method.

MCMC Sensitivity Threshold Analysis

GW signals in astrometric measurements can be extracted via Bayesian inference. This analysis framework is demonstrated in Ref. [43], where the authors implement a signal injection-retrieval study tailored for *Gaia*. Specifically, they consider a set of mock *Gaia* exposures and obtain posterior distributions for seven GW source parameters, plus and cross polarization amplitudes, $h_{+,\times}$, their respective initial phases, $\phi_{+,\times}$, GW frequency, f and two angles describing direction to the GW source, $\vec{q}$ (equivalent to $-\mathbf{p}$ in Section 4.2). In this paper, we focus on characterizing the intrinsic binary parameters that are detectable from the Roman EML survey data. For this purpose, we fix the extrinsic parameters (i.e., GW phase, polarization angle and source position) and derive limits on the binary chirp mass, $\mathcal{M}_c$, and luminosity distance, D_L, across the Roman EML survey frequency spectrum. Specifically, we set the wave phase, inclination angle, and polarization angle to 0. We also fix the GW source at the zenith position in the Galactic frame, as illustrated in Figure 4.1. The FoV is modeled as a 0.53 deg $\times$0.53 deg square centered on the galactic center. Fixing the contribution from phase and positional parameters, either by assigning specific representative values, as we do, or by numerically and analytically marginalizing over them, is also commonly adopted in PTA studies to reduce search space dimensions [see, e.g., 12, 21, 42].

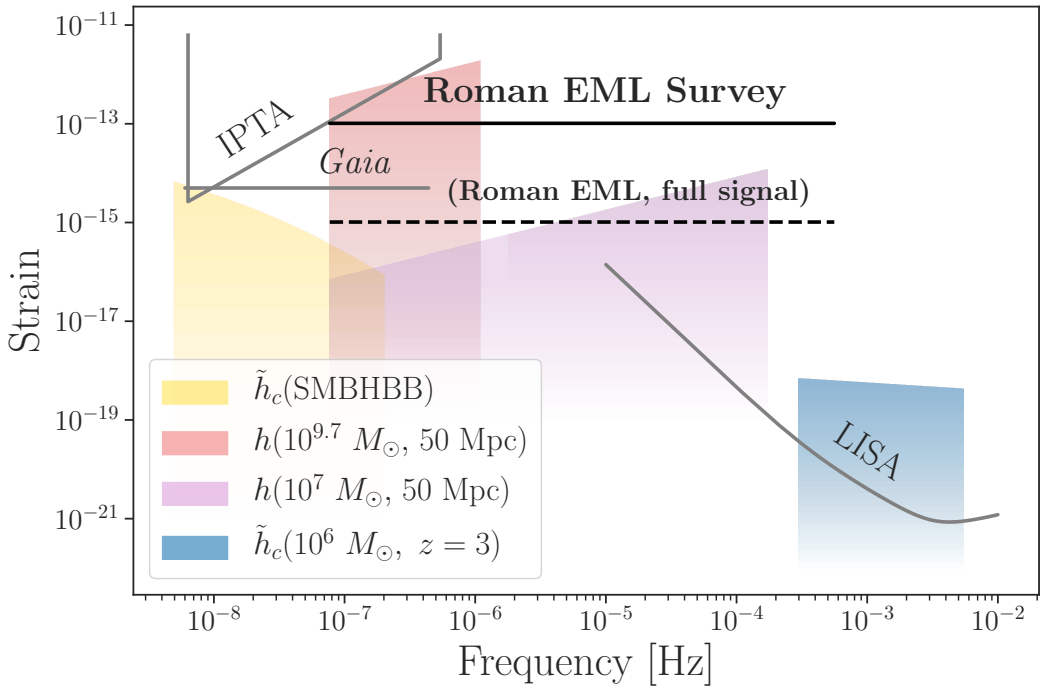

Figure 4.3: Strain sensitivity of various GW detectors and corresponding example signals. Note that the sensitivity for different detectors is not represented by the same quantity, in anticipation of the signal source types. The sensitivity for LISA and IPTA is represented by the dimensionless characteristic noise strain from the detector noise power spectral density, given by $\tilde{h}_n(f) = \sqrt{fS_n(f)}$. In the frequency range of IPTA, the yellow block shows the expected $\tilde{h}_c$ of the supermassive black hole binary background (SMBHBB). In the frequency range of LISA, the blue block shows the characteristic strain, $\tilde{h}_c(f) \equiv 2f\tilde{h}(f)$, of a fast-evolving $10^6\ M_\odot$ binary at $D_L = 25$ Gpc. The sensitivity of *Gaia* and Roman EML Survey is represented by the detectable instantaneous (time-domain) strain, h, of monochromatic GWs, assuming end-of-survey performance. For Roman EML Survey, the solid line shows the sensitivity under signal mean subtraction, and the dashed line shows the sensitivity if full astrometric deflections are detectable. The red and violet block show the time-domain wave amplitude of monochromatic GWs with chirp masses $10^{9.7}\ M_\odot$ and $10^7\ M_\odot$. The shown frequency range of these GWs are limited by the mission lifetime of Roman EML survey and f_{ISCO} at this chirp mass. Note that the example signals of Roman EML and LISA are illustrative; they do not reflect the GW source population expected from binary black hole formation theory. As is shown, the frequency band from roughly 5×10^{-7} Hz to 1×10^{-5} Hz is uniquely accessed by Roman EML survey.

Under our assumption, we consider the optimal case for detection. As the relative angle between the star position, $\mathbf{n}$, and the GW source position, $\vec{q}$, decreases, signal magnitude decreases accordingly and the detection threshold becomes more stringent. By fixing the GW phase to be 0, we simulate the *a posteriori* analysis, where, after observing at least one deflection cycle, we can determine the deflection amplitude from the entire data set.

Rather than calculating the full posterior distribution from mock data as in Ref. [43], we estimate the detection threshold by computing the likelihood of the signal-present hypothesis for various GW sources assuming we have observed the maximal astrometric deflection from the baseline. The astrometry measurement error is assumed to follow a Gaussian distribution with zero mean and no correlation across time. The standard deviation is the single-exposure, single-source astrometric accuracy. The signal is the mean-subtracted dn in Eqn. (4.5).

For simplicity and computational efficiency, we include only a subset of the expected number of observed stars and consider a single exposure at the maximal deflection. We then scale the results to approximate analysis outcomes with a full mock data set. Specifically, we randomly populate the FoV with 1000 stars. To account for the effect of the expected 10^8 observed stars, we scale down the astrometric resolution, σ, by $\sqrt{10^5}$. We make tests using several start counts ranging from 10^3 to 10^6, and observe no systematic bias. As Eqn. (4.10) suggests, we simulate the many exposures by scaling down σ by $\sqrt{N_m}$. See Appendix 4.6 for further justification.

We calculate the likelihood with Markov chain Monte Carlo (MCMC) simulations using the Python package `emcee` with no injection signal. We determine the 68%, 95% and 99.7% upper limits on M_s. We adopt a flat prior between

$$4.54 < \log_{10} M_s \ [M_\odot/\mathrm{Mpc}^{3/5}] < 11.54 \ ,$$

which is equivalent to flat priors between:

$$5.74 < \log_{10} \mathcal{M}_c \ \ [M_\odot] < 9.74$$

and

$$-3 < \log_{10} D_L \ \ [\mathrm{Mpc}] < 2 \ .$$

The upper bound of chirp mass is chosen such that the GW sources are realistic and have significant lifetime within the frequency band of the Roman EML survey. Parameter limits on the luminosity distance and the chirp mass are chosen to produce sufficiently strong signals in light of the theoretically calculated sensitivity curve.

The detectable $\log_{10}\mathcal{M}_c - \log_{10} D_L$ parameter space at selected GW frequencies is show in Figure 4.4. The top row shows the detectable binaries when the mean signal is subtracted, and the bottom row shows those when the full signal can be registered. The columns represent the sensitivity at different GW frequencies. Systems that already reach the ISCO are excluded from the accessible parameter space, since they quickly coalesce afterwards, and our analytical waveform expression in Eqn. (4.8) for the inspiral stage no longer captures the actual GW waveform. Specifically, (see, e.g., [52])

$$f_{\rm ISCO} = \frac{c^3}{6^{3/2}\pi G}\frac{(q^2(1+q))^{3/10}}{\mathcal{M}_c}\ , \tag{4.15}$$

and

$$\mathcal{M}_c(f)_{\rm max} = \frac{c^3}{6^{3/2}\pi G}\frac{2^{3/10}}{f}\ , \tag{4.16}$$

where, since $\mathcal{M}_c(f)_{\rm max}$ is an increasing function of q, we set $q = 1$.

Figure. 4.4 shows that, at all frequencies, the detectable parameter space is reduced significantly by subtracting the mean signal, and for GW with frequencies larger than 1×10^{-6} Hz, the parameter space is increasingly affected by the ISCO limit.

We summarize in Figure 4.5 the detection threshold across the Roman EML survey frequency band by plotting the 95% upper limit on $\mathcal{M}_c$ at 1 Mpc and 10 Mpc. Detection thresholds assuming an astrometric accuracy of 0.11 mas or full astrometric deflection signal are also plotted. As expected, the range of the detectable GW sources is limited by the signal strength and intrinsic frequency limits (i.e., $f_{\rm ISCO}$). Between these two competing factors, the "sweet spot" frequency with the largest accessible parameter space in the $\log_{10}\mathcal{M}_c - \log_{10} D_L$ plane is roughly located at 10^{-6} Hz. With its current expected performance, the Roman EML survey is sensitive to GWs from massive black hole binaries with $\mathcal{M}_c > 10^{7.4}\ M_\odot$ up to $D_L \sim 1$ Mpc; up to $D_L \sim 10$ Mpc, binaries with $\mathcal{M}_c > 10^{8.3}\ M_\odot$. Although this threshold excludes many of the interesting GW sources we hope to detect, Figure 4.5 shows that such sources out to 100 Mpc could be observable if the Roman Space Telescope can achieve a 0.11 mas astrometric accuracy, which is a possible improvement over the currently estimated 1.1 mas.

4.4 Discussion

In this section, we elaborate on the Roman Space Telescope pointing reconstruction strategy and evaluate its impact on GW detection. We then propose recommendations for maximizing the seredipitous GW scientific output from photometric survey

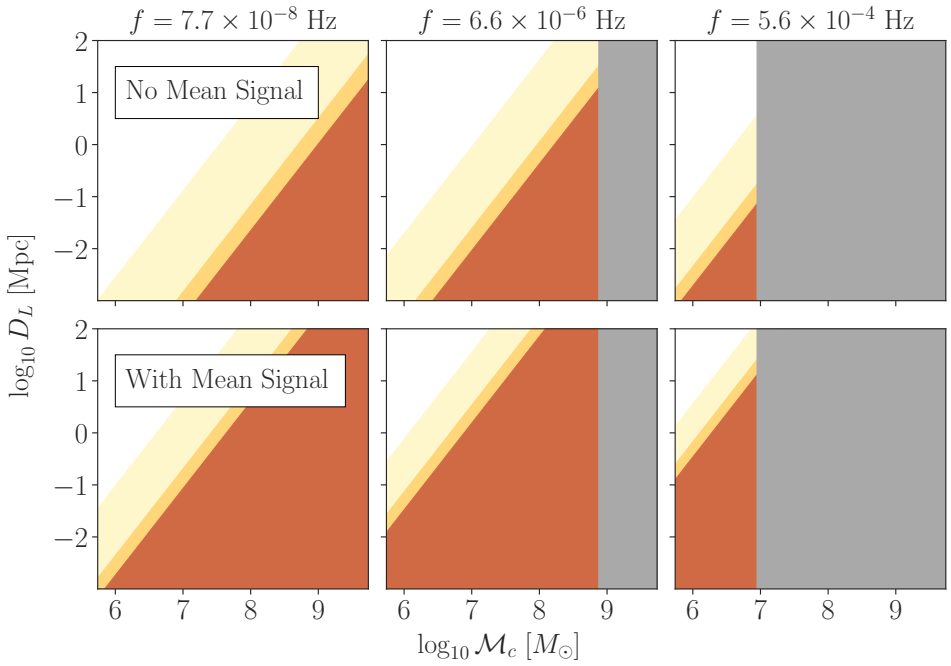

Figure 4.4: Detection sensitivity of the Roman EML Survey in $\log_{10} \mathcal{M}_c - \log_{10} D_L$ space at multiple fixed frequency. Colors indicate detection thresholds at different confidence levels. *Top Row:* sensitivity when signals are mean-subtracted. *Bottom Row:* Sensitivity when the full signal is observable. For the three columns, the GW frequencies are fixed to be 7.7×10^{-8} Hz, 6.6×10^{-6} Hz and 5.6×10^{-4} Hz, respectively. In all panels, GW sources that reach the ISCO at the specified frequency or lower are blocked out in gray.

instruments. Finally, we review some ongoing and planned surveys and discuss their merits and drawbacks as potential GW probes.

Roman Space Telescope Pointing Reconstruction and GS Selection

Here we expand on the mean subtraction technique discussed in Section 4.3 and assess its impact on the reach of the Roman EML survey as a GW probe.

Prior to launch, 4 to 18 guiding stars (GSs) will be selected in each observed field [62]. Of the 18 detectors of the Roman Space Telescope, each contains at most one guiding star. These stars are likely to be bright, and their precise absolute positions and proper motion will be available in external catalogs, e.g., in the *Gaia* catalog [62]. Their astrometric solution in the Roman Space Telescope operational epoch is extrapolated from the external catalog measurement (a similar procedure to study proper motions of galactic bulge stars is described in Ref.[19]). The absolute astrometry of all stars in the FoV is then obtained in post-processing by

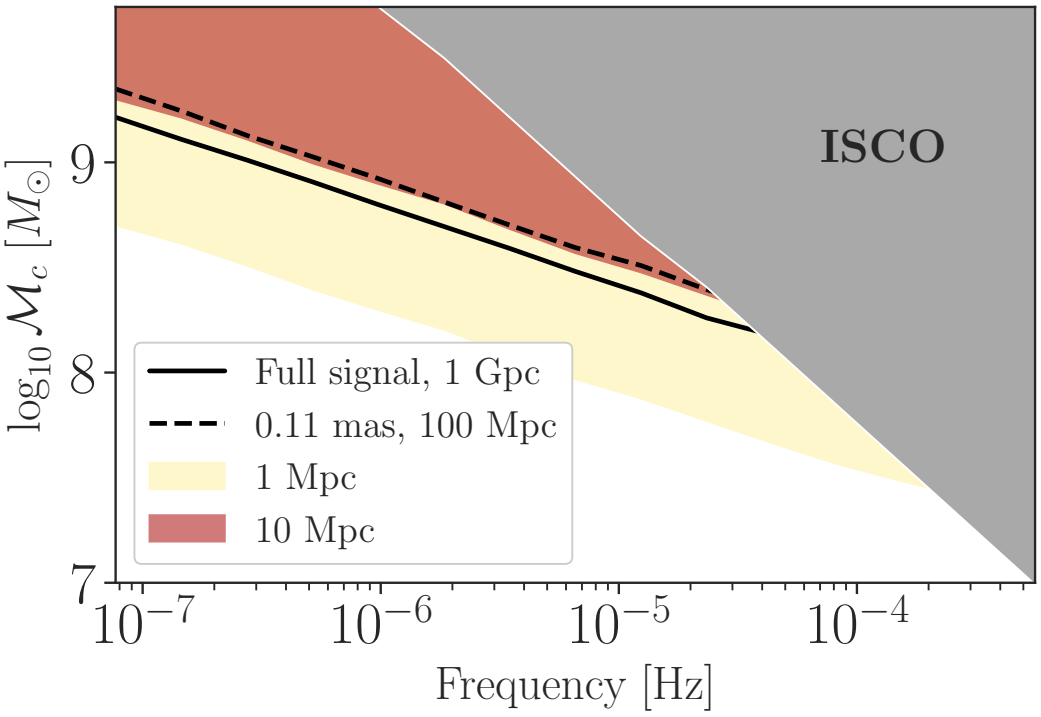

Figure 4.5: Sensitivity of the Roman EML Survey to $\log_{10} \mathcal{M}_c$. The upper right corner (shaded gray) excludes massive systems that reach the ISCO at each frequency or lower. Detection sensitivity threshold is represented by the detectable chirp mass at a 2σ confidence level at various luminosity distances. Yellow and red blocks show detectable mass ranges at 1 Mpc and 10 Mpc respectively, assuming the mean astrometric deflection is subtracted from the signal. The dashed line shows the detection threshold at 100 Mpc if the astrometric accuracy were to improve to 0.11 mas, equivalent to a factor of 10 improvement in the sensitivity. The solid line shows the sensitivity at 1 Gpc if the mean signal were observable, roughly comparable to a factor of 100 sensitivity improvement. See Section 4.4 for further discussions.

simultaneously fitting the GSs to their extrapolated positions.

As argued in Section 4.3, this tracking process will likely absorb a mean displacement signal within the FoV. Specifically for the Roman Space Telescope, this will be the mean deflections of the GSs. Though the choice of GSs is not yet available, we can gauge the effect of GS selection by repeating the MCMC study but subtracting only the mean of the GSs. For simplicity, we study two cases with 4 and 16 GSs. In each case, we model the detectors as square blocks that completely fill the FoV (i.e. no gaps, etc.) and place one GS in each of the square blocks. The position of the GS within each detector is then randomly chosen. We find that different GS choices only lead to $< 1\%$ variation in the upper limit confidence value, and having

fewer GSs gives larger variations. Thus, our strategy to subtract the mean of all stars serves as a good reference regardless of the mission specifics.

This mean-subtraction process significantly reduces the effective signal, and the sensitivity level is generally two orders of magnitude lower than the full-signal scenario. Figure 4.5 shows that the Roman EML survey is most sensitive to very massive binaries $\left(\sim 10^8\ M_\odot\right)$ at close distances ($\sim$ 1 Mpc). Since this is physically unlikely, the Roman EML survey with its current design will be limited as a GW probe. In fact, Figure 4.3 shows that the Roman EML survey would have better sensitivity than *Gaia* if the mean signal were to be detectable, in which case the accessible parameter space would be greatly expanded.

This prediction is different from that for *Gaia* in Ref. [43], since a full-signal analysis is assumed. In the case of *Gaia*, this treatment is warranted since *Gaia* simultaneously observes through two widely separated FoVs and does not need to perform mean subtraction [45]. The sampled GW deflection patterns are consequently distinct and cannot be absorbed by the same pointing calibration process. For essentially the same reason, *Gaia* can measure absolute parallax rather than relative parallax [34].

While this outlined strategy is specific to the Roman EML survey, we note that the loss of the mean astrometric deflection signal is a typical feature of photometric surveys. Even though this loss presents a challenge for resolving individual GWs, the sensitivity might be better for joint signals of several GWs. We expect the combined GWs to produce a deflection pattern richer in features, and thus easier to detect. Such signals would come from SMBHBs at the centers of galaxies in the local universe, and the astrometric measurements can be used to study their population statistics. This is analogous to using PTA measurements to constrain the energy density of the stochastic GW background produced by massive black hole mergers across all redshifts [see, e.g., 9, 46, 58]. It is estimated that $\sim$ 100 continuous GW sources in the PTA band exist within 225 Mpc [40]. We may then speculate that a significant number of SMBHBs within our frequency range exist in the local universe, and their joint signal may be above detection threshold for the astrometry method.

Optimizing Photometric Surveys for GW

In this section, we give specific recommendations to maximize the GW detection potential of photometric surveys within the GW frequency gap of existing detection

methods. We use the expected performance of the Roman EML survey as a reference point and quantitatively describe a model survey capable of detecting a fiducial target, a 10^7 $M_\odot$ binary at 50 Mpc.

To estimate the required sensitivity, we proceed directly from Figure 4.5; the detection threshold is lowered by the same order of magnitude as the increase in effective signal strength. Therefore, to claim a 2σ detection on GWs from this new fiducial target, the model survey is 100 times more sensitive than the Roman EML survey. In the following sections we discuss ways to achieve this sensitivity.

Mean-signal Recovery Fraction

As suggested by the previous subsection, pointing reconstruction strategies determine whether the mean astrometric deflection could be observable, which translates into approximately two orders of magnitude sensitivity difference. Though the mean-subtracted deflection pattern and the full signal pattern differ in both the deflection magnitude averaged over all stars within the FoV and the pattern shape, we only use the average deflection magnitude as an approximate metric to compare sensitivity.

For our model FoV configuration, the average deflection magnitude after subtracting the mean is roughly 100 times smaller than the average full signal (see, for example, Figure 4.1), thus the sensitivity is roughly 100 times worse. We may then define a mean-signal recovery fraction to roughly quantify the observable deflection relative to the full signal. For example, a mean-signal recovery fraction of 50% implies that the average magnitude of the observable deflections after the astrometry solution is half of the full signal magnitude. Consequently, the sensitivity would be roughly 50 times higher than what we obtained from the MCMC study assuming mean-subtracted signals. Ideally, the photometric survey retains nearly all of the mean signal, relaxing the detection threshold by roughly a factor of 100. Such a model survey, with all other parameters similar to the Roman EML survey, can already detect the fiducial GW source. In general, higher recovery fraction allows detection of intrinsically weaker GW sources, such as farther and lighter systems, or the same system but much earlier in its evolution track.

Astrometric Accuracy

As shown by Eqn. (4.10), the strain amplitude threshold is linearly proportional to the astrometric accuracy. All else equivalent, the model survey improves upon the

Roman Space Telescope accuracy by at least a factor of 100, giving a single-exposure single-source astrometric resolution better than 11 μas. In this work, we assume an astrometric accuracy of 1.1 mas, which is 1/100th of the detector pixel size [62]. We expect this accuracy to be routinely performed, but it is possible that 0.11 mas can be achieved [32]. In this case, the Roman EML survey will be ten times more sensitive and will already be able to detect binaries with $\mathcal{M}_c > 10^{8.3} M_\odot$ within 100 Mpc (see Figure 4.5).

For comparison, the astrometric accuracy of *Gaia* is $0.1 \sim 2$ mas (for $G = 17$ and $G = 21$ stars, respectively) [18]. The expected imaging resolution of the Square Kilometer Array (SKA)[3] [61] at 12.5 GHz is 0.04 arcsec [2]; assuming a fractional position error requirement smaller than 1% [1], the SKA can achieve an astrometric accuracy better than 0.4 mas.

Number of Stars

The statistical advantage of observing more stars ($\propto \sqrt{N_s}$) is stated in Eqn. (4.10). This number can be expressed as

$$\begin{aligned} N_s &= \int \alpha\left(\vec{r}\right) \rho\left(L, \vec{r}\right) dL \; d^3\vec{r}_{\rm obs} \\ &\approx A \int \alpha(\theta_0, \phi_0, r) \rho(L, \theta_0, \phi_0, r) r^2 dL \; dr \; , \end{aligned} \tag{4.17}$$

where $\alpha\left(\vec{r}\right)$ is the effective detectable fraction after photon loss during propagation (e.g., dust absorption, crowding effect, etc.), and $\rho\left(L, \vec{r}\right)$ is the population density of luminosity-L stars. A is the covered angular area. In the second equality, we assume small variation of the integrand in the angular directions. Since GW detection requires frequent visit to the same field, we assume a "deep survey" mode where the total surveyed angular area is small and this equality is satisfied.

Evidently, the observational efficiency of telescopes is greatly increased if they can penetrate to further distances per area (i.e., large α). Therefore, a telescope with near-infrared filters outperforms one operating in the visible band, as near-infrared photons suffer less absorption by galactic dust along propagation. The optimal choice for the filter wavelength should, however, be balanced between this low-absorption advantage and the large-diffraction effect for long wavelengths, which degrades the point spread function and thus the astrometric accuracy.

[3] https://www.skatelescope.org/the-ska-project/

The θ_0, ϕ_0 dependence suggests the importance of pointing directions. Specifically, surveys pointing toward the galactic center have larger ρ for fixed distance and luminosity. For magnitude limited surveys, this implies a larger number of observed stars. Conversely, surveys in high latitude regions are less advantageous since they observe fewer stars above certain magnitude limits. For comparison, the stellar density down to $H(AB) = 20$ mag at Galactic Latitude of 60 deg is approximately 3000 stars/deg^2 [39].

It is also intuitive that a larger FoV leads to more observed stars, all else equivalent. Therefore, the model survey will have comparable bands and pointing directions to the Roman Space Telescope during its EML survey, but with a 100 times larger ($\sim 200\,\mathrm{deg}^2$) survey area.

FoV Size

The effect of increasing the FoV size is partially degenerate with increasing the survey area, but it also allows for a larger deflection residual after mean subtraction. Specifically, the subtracted mean decreases as the variation across the FoV at each exposure becomes more significant. The scaling relation between the FoV sidelength, average deflection vector field divergence and the average deflection magnitude is described in Section 4.2.

The combination of changes both in signal magnitude and pattern will likely be a complex effect that, in general, enhance the sensitivity. In principle, FoV patches can be stitched together to provide a larger effective FoV to include more pattern variation. However, the field-switching process must be exquisitely controlled such that the absorbed mean for each field is approximately the global mean solution in the larger effective FoV. However, due to the very large scale over which the GW-induced deflection pattern varies (on the order of tens of degrees), it is unlikely that future surveys can outperform the Roman EML survey by a factor of 100 through this means alone.

Observational Cadence & Mission Length

The impact of observational cadence is two-fold: it determines the sensitive frequency range and contributes to the statistical improvement of sensitivity. To complement LISA, therefore, the upper limit frequency should be at least $\sim 10^{-5}$ Hz. It follows that an the model survey observes the same patch of sky at least once a day. For sensitivity improvement, the model survey has a longer effective observational

time than the six 72-day epochs of the Roman EML survey. For example, a 10-year survey with full duty cycle improves the sensitivity by a factor of 3.

Other Potential Photometric GW Probes

In this section, we further develop the guidelines for assessing photometric surveys as GW probes. We discuss ground-based and space-based telescopes in turn by pointing out their respective merits and drawbacks as potential GW probes. To the best of our knowledge, all the observatories discussed below would suffer from the limiting mean signal subtraction we discussed above.

A challenge with ground-based telescopes as GW probes lies in their relatively coarse astrometric resolution compared with space telescopes, due to atmospheric perturbation to the signal. For example, the Rubin Observatory[4] has a single-exposure astrometric accuracy of ~ 11 mas [27], an order of magnitude larger than that of the Roman Space Telescope.

This resolution drawback can be partially compensated by a large number of observed stars, large FoV, and great observational flexibility. For instance, the Rubin Observatory is expected to observe a total of $\sim 4 \times 10^9$ stars with a FoV size of $\sim 10\,\mathrm{deg}^2$. Each sky patch is visited ~ 100 times during its 10-year lifetime. By increasing its observational cadence by a factor of 5 (~ 1 week^{-1} on average), the Rubin Observatory would become sensitive to GWs with $f < 1.5 \times 10^{-6}$ Hz.

This astrometric method can also be applied to high-resolution radio telescopes, such as the SKA and the Next Generation Very Large Array (ngVLA)[5] [38]. As discussed, SKA can achieve an astrometric accuracy better than 0.4 mas. The ngVLA features a maximum baseline resolution as small as 0.17 mas at 41 GHz. It is also estimated that a large number of quasars can be observed in the radio band [see, e.g., $\sim 10^6$ in 28], which can serve as GW detectors instead of stars. Taking the SKA as an example, the relatively smaller number of observed quasars compared with stars observed by *Gaia* can potentially be compensated by a more frequent observation schedule to give similar performance at a higher frequency. Increasing the exposure time would also directly increase the number of detected quasars. Specifically, an SKA survey taking measurements every 40 minutes has a GW frequency band similar to that of the Roman EML survey ($f < 2 \times 10^{-4}$ Hz). To the authors' knowledge, there is currently no high-cadence survey planned.

[4] https://www.lsst.org/lsst/

[5] https://ngvla.nrao.edu/

Free from atmospheric effects, space-based telescopes can potentially observe a great number of stars to excellent precision. As an example, the ESA mission Euclid[6] is equipped with a near-infrared photometer with a $\sim$ 1 mas single-exposure astrometric accuracy [44], similar to the expected performance of the Roman Space Telescope. The HabEx Workhorse Camera (HWC) onboard the Habitable Exoplanet Observatory (HabEx)[7] is expected to have similar, if not better, angular resolution to the Roman Space Telescope, albeit with a much smaller FoV [24].

The deciding factors then become the observed fields and observational cadence. Unlike the Roman Space Telescope, HabEx is not designed as a survey instrument; instead, it focuses on characterizing a handful of targets in great detail, and thus will not be suitable for our purpose. While Euclid does feature a deep survey, with 40 $\deg^2$ of sky observed every 15 days [44], these fields are close to the ecliptic pole with low stellar density. However, a high-cadence survey in its extended mission lifetime, following the recommendations we outline, could contribute meaningfully to GW detection.

4.5 Concluding Remarks

In this paper, we show how to use a photometric survey as a GW probe that uniquely bridges the GW frequency spectrum gap between existing detection methods. It is also not required that such photometric surveys are designed specifically to provide absolute astrometric solutions. We demonstrate that relative astrometric deflections of observed stars within the FoV already allow for GW detection, albeit at a lower sensitivity level. We discuss key factors that determine sensitivity. We then assess the potential of the Roman EML survey in its current definition as a GW probe. In Section 4.4, we make recommendations for maximizing the GW scientific output of photometric surveys and quantify the desirable performance via a model survey. Finally, we review existing and planned photometric surveys, and discuss their relative strengths and drawbacks as potential GW probes.

We note that our analysis can be refined in several ways. For instance, we have yet to explicitly include stellar proper motion in our simulation, which can in theory be subtracted via quadratic fitting. Such proper motions may even be correlated across the FoV, if, for example, open clusters are present. However, we expect these motions to have limited impact on the GW sensitivity once we consider the signal

[6] https://sci.esa.int/web/euclid/home
[7] https://www.jpl.nasa.gov/habex/

variation over time. Especially for high-frequency GWs, their oscillatory nature leaves a distinct signature from physical proper motion over long timescales.

We could also model the seven fields of the Roman EML survey jointly. A combined analysis of the data from all fields might amount to having a larger effective FoV, should the temporal and pointing accuracy during the field-switching process allow. Incorporating the GW frequency evolution could also enhance sensitivity.

The recommendations in Section 4.3 should serve as a reference for maximizing GW science from future photometric surveys. The current expected performance of the Roman Space Telescope could make detecting individual GWs a challenge. However, with some luck and a novel pointing reconstruction strategy, we may yet detect individual GWs with the Roman Space Telescope.

Software: astropy [11], astroquery [25], emcee [22], matplotlib [26], numpy [57], scipy [60]

4.6 Appendix: Scaling Argument

In Section 4.3, we employ scaling arguments to show how the full survey sensitivity can be approximated by the MCMC results on a subset of stars in a single exposure. In this appendix we derive these scaling relations explicitly.

The log likelihood for a full dataset should be written as

$$\ln \mathcal{L} = \sum_{i=1}^{N_m} \sum_{j=1}^{N_s} \sum_{k=0,1} \frac{(ds_{i,j,k} - dn_{i,j,k}(\Psi))^2}{2\sigma^2} \,, \tag{4.18}$$

where the indices i, j, k represents the exposures, the number of stars and the two components of the deflection vector. ds is the data vector; in the null signal case, each data vector component follows a Gaussian distribution with zero mean and a standard deviation of σ. $dn(\Psi)$ is the predicted astrometric deflection given the GW source parameter Ψ. σ is the telescope single-exposure single-source astrometric resolution. On average,

$$\begin{aligned} \ln \mathcal{L} &\approx N_m N_s \sum_{k=0,1} \frac{\langle (ds_k - dn_k(\Psi))^2 \rangle}{2\sigma^2} \\ &= \sum_{k=0,1} \frac{\langle (ds_k - dn_k(\Psi))^2 \rangle}{2(\sigma/\sqrt{N_m N_s})^2} \,, \end{aligned} \tag{4.19}$$

where $\langle \cdot \rangle$ an average over the observed stars and the exposures. This is the same scaling relation in Section 4.2 and Section 4.3, up to a constant scaling factor on the order of 1.

We note that the sensitivity is independent from the frequency of the GWs. By modeling the signal as purely sinusoidal, we numerically calculate the variance of the signal parameters using the Fisher information matrix. Once we have observed a significant number of signal cycles, the standard deviation of the wave amplitude approaches a constant, $\sqrt{2\pi/T}$, where T is the total observation time.

References

[1] SKA1 level 0 science requirements. Technical Report SKA-TEL-SKO-0000007, October 2015. URL https://astronomers.skatelescope.org/wp-content/uploads/2015/11/SKA-TEL-SKO-0000007_SKA1_Level_0_Science_RequirementsRev02-part-1-signed.pdf.

[2] Technical Information - the Telescopes, August 2018. URL https://www.skatelescope.org/wp-content/uploads/2018/08/16231-factsheet-telescopes-v71.pdf.

[3] Roman's Key Science Components, 2020. URL https://www.stsci.edu/roman/about/science-themes.

[4] J. Aasi et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, April 2015. doi: 10.1088/0264-9381/32/7/074001.

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

[6] Aniket Agrawal, Teppei Okumura, and Toshifumi Futamase. Constraining neutrino mass and dark energy with peculiar velocities and lensing dispersions of Type Ia supernovae. *Phys. Rev. D*, 100(6):063534, September 2019. doi: 10.1103/PhysRevD.100.063534.

[7] Pau Amaro-Seoane, Heather Audley, Stanislav Babak, John Baker, Enrico Barausse, et al. Laser Interferometer Space Antenna. *arXiv e-prints*, art. arXiv:1702.00786, February 2017.

[8] Melissa Anholm, Stefan Ballmer, Jolien D. E. Creighton, Larry R. Price, and Xavier Siemens. Optimal strategies for gravitational wave stochastic background searches in pulsar timing data. *Phys. Rev. D*, 79(8):084030, April 2009. doi: 10.1103/PhysRevD.79.084030.

[9] Z. Arzoumanian, P. T. Baker, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, S. Chatterjee, B. Christy, J. M. Cordes, N. J. Cornish, F. Crawford, H. Thankful Cromartie, K. Crowter, M. DeCesar, P. B. Demorest, T. Dolch, J. A. Ellis, others, and NANOGrav Collaboration. The NANOGrav 11 Year Data Set: Pulsar-timing Constraints on the Stochastic Gravitational-wave Background. *ApJ*, 859(1):47, May 2018. doi: 10.3847/1538-4357/aabd3b.

[10] Zaven Arzoumanian, Paul T. Baker, Harsha Blumer, Bence Becsy, Adam Brazier, Paul R. Brook, Sarah Burke-Spolaor, Shami Chatterjee, Siyuan Chen, James M. Cordes, Neil J. Cornish, et al. The NANOGrav 12.5-year Data Set: Search For An Isotropic Stochastic Gravitational-Wave Background. *arXiv e-prints*, art. arXiv:2009.04496, September 2020.

[11] Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, A. Ginsburg, J. T. Vand erPlas, L. D. Bradley, D. Pérez-Suárez, M. de Val-Borro, T. L. Aldcroft, K. L. Cruz, T. P. Robitaille, E. J. Tollerud, C. Ardelean, T. Babej, others, and Astropy Contributors. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *AJ*, 156(3):123, September 2018. doi: 10.3847/1538-3881/aabc4f.

[12] Stanislav Babak and Alberto Sesana. Resolving multiple supermassive black hole binaries with pulsar timing arrays. *Phys. Rev. D*, 85(4):044034, February 2012. doi: 10.1103/PhysRevD.85.044034.

[13] S. M. Baggett, J. W. MacKenty, M. Robberto, R. A. Kimble, R. J. Hill, H. Bushouse, W. Baggett, T. Brown, G. Hartig, B. Hilbert, I. N. Reid, G. Delo, R. Foltz, S. D. Johnson, O. Lupie, E. Malumuth, T. Pham, A. M. Russell, A. Waczynski, Y. Wen, and WFC3 Team. The Wide Field Camera 3 Detectors. In *American Astronomical Society Meeting Abstracts*, volume 205 of *American Astronomical Society Meeting Abstracts*, page 05.02, December 2004.

[14] Sambaran Banerjee. LISA sources from young massive and open stellar clusters. *Phys. Rev. D*, 102(10):103002, November 2020. doi: 10.1103/PhysRevD.102.103002.

[15] Sambaran Banerjee. Stellar-mass black holes in young massive and open stellar clusters IV: updated stellar-evolutionary and black hole spin models and comparisons with the LIGO-Virgo O1/O2 merger-event data. *MNRAS*, September 2020. doi: 10.1093/mnras/staa2392.

[16] Luc Blanchet. Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries. *Living Reviews in Relativity*, 17(1):2, December 2014. doi: 10.12942/lrr-2014-2.

[17] Laura G. Book and Éanna É. Flanagan. Astrometric effects of a stochastic gravitational wave background. *Phys. Rev. D*, 83(2):024024, January 2011. doi: 10.1103/PhysRevD.83.024024.

[18] A. G. A. Brown et al. Gaia Data Release 2. Summary of the contents and survey properties. *A&A*, 616:A1, August 2018. doi: 10.1051/0004-6361/201833051.

[19] Will Clarkson, Kailash Sahu, Jay Anderson, T. Ed Smith, Thomas M. Brown, R. Michael Rich, Stefano Casertano, Howard E. Bond, Mario Livio, Dante Minniti, Nino Panagia, Alvio Renzini, Jeff Valenti, and Manuela Zoccali.

Stellar Proper Motions in the Galactic Bulge from Deep Hubble Space Telescope ACS WFC Photometry. *ApJ*, 684(2):1110–1142, September 2008. doi: 10.1086/590378.

[20] Curt Cutler and Éanna 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(6):2658–2697, March 1994. doi: 10.1103/PhysRevD.49.2658.

[21] J. A. Ellis, X. Siemens, and J. D. E. Creighton. Optimal Strategies for Continuous Gravitational Wave Detection in Pulsar Timing Arrays. *ApJ*, 756(2):175, September 2012. doi: 10.1088/0004-637X/756/2/175.

[22] Daniel Foreman-Mackey, Alex Conley, Will Meierjurgen Farr, David W. Hogg, Dustin Lang, Phil Marshall, Adrian Price-Whelan, Jeremy Sanders, and Joe Zuntz. emcee: The MCMC Hammer, March 2013.

[23] B. Scott Gaudi, Rachel Akeson, Jay Anderson, Etienne Bachelet, David P. Bennett, Aparna Bhattacharya, Valerio Bozza, Sebastiano Calchi Novati, Calen B. Henderson, Samson A. Johnson, Jeffrey Kruk, Jessica R. Lu, Shude Mao, Benjamin T. Montet, David M. Nataf, Matthew T. Penny, Radoslaw Poleski, Clément Ranc, Kailash Sahu, Yossi Shvartzvald, David N. Spergel, Daisuke Suzuki, Keivan G. Stassun, and Rachel A. Street. 'Auxiliary' Science with the WFIRST Microlensing Survey. *BAAS*, 51(3):211, May 2019.

[24] B. Scott Gaudi, Sara Seager, Bertrand Mennesson, Alina Kiessling, Keith Warfield, Kerri Cahoy, John T. Clarke, Shawn Domagal-Goldman, Lee Feinberg, Olivier Guyon, Jeremy Kasdin, Dimitri Mawet, Peter Plavchan, Tyler Robinson, Leslie Rogers, Paul Scowen, et al. The Habitable Exoplanet Observatory (HabEx) Mission Concept Study Final Report. *arXiv e-prints*, art. arXiv:2001.06683, January 2020.

[25] Adam Ginsburg, Brigitta M. Sipőcz, C. E. Brasseur, Philip S. Cowperthwaite, Matthew W. Craig, Christoph Deil, James Guillochon, Giannina Guzman, Simon Liedtke, Pey Lian Lim, Kelly E. Lockhart, Michael Mommert, Brett M. Morris, Henrik Norman, Madhura Parikh, Magnus V. Persson, Thomas P. Robitaille, Juan-Carlos Segovia, Leo P. Singer, Erik J. Tollerud, Miguel de Val-Borro, Ivan Valtchanov, Julien Woillez, Astroquery Collaboration, and a subset of astropy Collaboration. astroquery: An Astronomical Web-querying Package in Python. *AJ*, 157(3):98, March 2019. doi: 10.3847/1538-3881/aafc33.

[26] John D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering*, 9(3):90–95, May 2007. doi: 10.1109/MCSE.2007.55.

[27] Željko Ivezić, Steven M. Kahn, J. Anthony Tyson, Bob Abel, Emily Acosta, Robyn Allsman, David Alonso, Yusra AlSayyad, Scott F. Anderson, John

Andrew, James Roger P. Angel, George Z. Angeli, Reza Ansari, Pierre Antilogus, Constanza Araujo, Robert Armstrong, Kirk T. Arndt, Pierre Astier, et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *ApJ*, 873(2):111, March 2019. doi: 10.3847/1538-4357/ab042c.

[28] Andrew H. Jaffe. Observing gravitational radiation with QSO proper motions and the SKA. *New A Rev.*, 48(11-12):1483–1485, December 2004. doi: 10.1016/j.newar.2004.09.018.

[29] Seiji Kawamura, Takashi Nakamura, Masaki Ando, Naoki Seto, Tomotada Akutsu, et al. Space gravitational-wave antennas DECIGO and B-DECIGO. *International Journal of Modern Physics D*, 28(12):1845001, January 2019. doi: 10.1142/S0218271818450013.

[30] Lawrence E. Kidder. Coalescing binary systems of compact objects to (post)$^{5/2}$-Newtonian order. V. Spin effects. *Phys. Rev. D*, 52(2):821–847, July 1995. doi: 10.1103/PhysRevD.52.821.

[31] Sergei A. Klioner. Gaia-like astrometry and gravitational waves. *Classical and Quantum Gravity*, 35(4):045005, February 2018. doi: 10.1088/1361-6382/aa9f57.

[32] Jeff Kruk. personal communication.

[33] Kevin A. Kuns, Hang Yu, Yanbei Chen, and Rana X Adhikari. Astrophysics and cosmology with a deci-hertz gravitational-wave detector: TianGO. *arXiv e-prints*, art. arXiv:1908.06004, August 2019.

[34] L. Lindegren and U. Bastian. Basic principles of scanning space astrometry. In *EAS Publications Series*, volume 45 of *EAS Publications Series*, pages 109–114, January 2010. doi: 10.1051/eas/1045018.

[35] L. Lindegren, S. A. Klioner, J. Hernández, A. Bombrun, M. Ramos-Lerate, H. Steidelmüller, U. Bastian, M. Biermann, et al. Gaia Early Data Release 3: The astrometric solution. *arXiv e-prints*, art. arXiv:2012.03380, December 2020.

[36] Jun Luo, Li-Sheng Chen, Hui-Zong Duan, Yun-Gui Gong, Shoucun Hu, Jianghui Ji, Qi Liu, Jianwei Mei, Vadim Milyukov, Mikhail Sazhin, Cheng-Gang Shao, Viktor T. Toth, Hai-Bo Tu, Yamin Wang, Yan Wang, Hsien-Chi Yeh, Ming-Sheng Zhan, Yonghe Zhang, Vladimir Zharov, and Ze-Bing Zhou. TianQin: a space-borne gravitational wave detector. *Classical and Quantum Gravity*, 33(3):035010, February 2016. doi: 10.1088/0264-9381/33/3/035010.

[37] X. Luri, A. G. A. Brown, L. M. Sarro, F. Arenou, C. A. L. Bailer-Jones, A. Castro-Ginard, J. de Bruijne, T. Prusti, C. Babusiaux, and H. E. Delgado. Gaia Data Release 2. Using Gaia parallaxes. *A&A*, 616:A9, August 2018. doi: 10.1051/0004-6361/201832964.

[38] Mark McKinnon, Anthony Beasley, Eric Murphy, Rob Selina, Rick Farnsworth, and Alex Walter. ngVLA: The Next Generation Very Large Array. In *BAAS*, volume 51, page 81, September 2019.

[39] Y. Mellier. Euclid: Mapping the Geometry of the Dark Universe. In *Science from the Next Generation Imaging and Spectroscopic Surveys*, page 3, October 2012.

[40] Chiara M. F. Mingarelli, T. Joseph W. Lazio, Alberto Sesana, Jenny E. Greene, Justin A. Ellis, Chung-Pei Ma, Steve Croft, Sarah Burke-Spolaor, and Stephen R. Taylor. The local nanohertz gravitational-wave landscape from supermassive black hole binaries. *Nature Astronomy*, 1:886–892, November 2017. doi: 10.1038/s41550-017-0299-6.

[41] C. J. Moore, R. H. Cole, and C. P. L. Berry. Gravitational-wave sensitivity curves. *Classical and Quantum Gravity*, 32(1):015014, January 2015. doi: 10.1088/0264-9381/32/1/015014.

[42] C. J. Moore, S. R. Taylor, and J. R. Gair. Estimating the sensitivity of pulsar timing arrays. *Classical and Quantum Gravity*, 32(5):055004, March 2015. doi: 10.1088/0264-9381/32/5/055004.

[43] Christopher J. Moore, Deyan P. Mihaylov, Anthony Lasenby, and Gerard Gilmore. Astrometric Search Method for Individually Resolvable Gravitational Wave Sources with Gaia. *Phys. Rev. Lett.*, 119(26):261102, December 2017. doi: 10.1103/PhysRevLett.119.261102.

[44] Will Percival, Michael Balogh, Dick Bond, Jo Bovy, Raymond Carlberg, Scott Chapman, Patrick Cote, Nicolas Cowan, Sebastien Fabbro, Laura Ferrarese, Stephen Gwyn, Renee Hlozek, Michael Hudson, John Hutchings, JJ Kavelaars, Dustin Lang, Alan McConnachie, Adam Muzzin, Laura Parker, Chris Pritchet, Marcin Sawicki, David Schade, Douglas Scott, Kendrick Smith, Kristine Spekkens, James Taylor, and Chris Willott. The Euclid Mission. In *Canadian Long Range Plan for Astronony and Astrophysics White Papers*, volume 2020, page 20, October 2019. doi: 10.5281/zenodo.3758532.

[45] T. Prusti et al. The Gaia mission. *A&A*, 595:A1, November 2016. doi: 10.1051/0004-6361/201629272.

[46] Ted Pyne, Carl R. Gwinn, Mark Birkinshaw, T. Marshall Eubanks, and Demetrios N. Matsakis. Gravitational Radiation and Very Long Baseline Interferometry. *ApJ*, 465:566, July 1996. doi: 10.1086/177443.

[47] Travis Robson, Neil J. Cornish, and Chang Liu. The construction and use of LISA sensitivity curves. *Classical and Quantum Gravity*, 36(10):105011, May 2019. doi: 10.1088/1361-6382/ab1101.

[48] LIGO Science Instrument List, :, A. Buikema, C. Cahillane, G. L. Mansell, C. D. Blair, R. Abbott, C. Adams, R. X. Adhikari, A. Ananyeva, S. Appert, K. Arai, J. S. Areeda, Y. Asali, S. M. Aston, C. Austin, A. M. Baer, M. Ball, S. W. Ballmer, S. Banagiri, D. Barker, L. Barsotti, J. Bartlett, B. K. Berger, J. Betzwieser, D. Bhattacharjee, G. Billingsley, S. Biscans, et al. Sensitivity and Performance of the Advanced LIGO Detectors in the Third Observing Run. *arXiv e-prints*, art. arXiv:2008.01301, August 2020.

[49] A. Sesana, A. Vecchio, and C. N. Colacino. The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays. *MNRAS*, 390(1):192–209, October 2008. doi: 10.1111/j.1365-2966.2008.13682.x.

[50] A. Sesana, A. Vecchio, and M. Volonteri. Gravitational waves from resolvable massive black hole binary systems and observations with Pulsar Timing Arrays. *MNRAS*, 394(4):2255–2265, April 2009. doi: 10.1111/j.1365-2966.2009.14499.x.

[51] Alberto Sesana and Alberto Vecchio. Measuring the parameters of massive black hole binary systems with pulsar timing array observations of gravitational waves. *Phys. Rev. D*, 81(10):104008, May 2010. doi: 10.1103/PhysRevD.81.104008.

[52] Alberto Sesana, Francesco Haardt, Piero Madau, and Marta Volonteri. Low-Frequency Gravitational Radiation from Coalescing Massive Black Hole Binaries in Hierarchical Cosmologies. *ApJ*, 611(2):623–632, August 2004. doi: 10.1086/422185.

[53] S. R. Taylor, M. Vallisneri, J. A. Ellis, C. M. F. Mingarelli, T. J. W. Lazio, and R. van Haasteren. Are We There Yet? Time to Detection of Nanohertz Gravitational Waves Based on Pulsar-timing Array Limits. *ApJ*, 819(1):L6, March 2016. doi: 10.3847/2041-8205/819/1/L6.

[54] The LIGO Scientific Collaboration and the Virgo Collaboration. GW190412: Observation of a Binary-Black-Hole Coalescence with Asymmetric Masses. *arXiv e-prints*, art. arXiv:2004.08342, April 2020.

[55] James Ira Thorpe, John Ziemer, Ira Thorpe, Jeff Livas, John W. Conklin, Robert Caldwell, Emanuele Berti, Sean T. McWilliams, Robin Stebbins, David Shoemaker, Elizabeth C. Ferrara, Shane L. Larson, et al. The Laser Interferometer Space Antenna: Unveiling the Millihertz Gravitational Wave Sky. In *BAAS*, volume 51, page 77, September 2019.

[56] Eric Thrane and Joseph D. Romano. Sensitivity curves for searches for gravitational-wave backgrounds. *Phys. Rev. D*, 88(12):124032, December 2013. doi: 10.1103/PhysRevD.88.124032.

[57] Stéfan van der Walt, S. Chris Colbert, and Gaël Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science and Engineering*, 13(2):22–30, March 2011. doi: 10.1109/MCSE.2011.37.

[58] R. van Haasteren, Y. Levin, G. H. Janssen, K. Lazaridis, M. Kramer, B. W. Stappers, G. Desvignes, M. B. Purver, A. G. Lyne, R. D. Ferdman, A. Jessner, I. Cognard, G. Theureau, N. D'Amico, A. Possenti, M. Burgay, A. Corongiu, J. W. T. Hessels, R. Smits, and J. P. W. Verbiest. Placing limits on the stochastic gravitational-wave background using European Pulsar Timing Array data. *MNRAS*, 414(4):3117–3128, July 2011. doi: 10.1111/j.1365-2966.2011.18613.x.

[59] J. P. W. Verbiest, L. Lentati, G. Hobbs, R. van Haasteren, P. B. Demorest, et al. The International Pulsar Timing Array: First data release. *MNRAS*, 458(2):1267–1288, May 2016. doi: 10.1093/mnras/stw347.

[60] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake Vand erPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: https://doi.org/10.1038/s41592-019-0686-2.

[61] A. Weltman, P. Bull, S. Camera, K. Kelley, H. Padmanabhan, J. Pritchard, A. Raccanelli, S. Riemer-Sørensen, L. Shao, S. Andrianomena, E. Athanassoula, D. Bacon, R. Barkana, G. Bertone, C. Bœhm, C. Bonvin, A. Bosma, M. Brüggen, et al. Fundamental physics with the Square Kilometre Array. *PASA*, 37:e002, January 2020. doi: 10.1017/pasa.2019.42.

[62] WFIRST Astrometry Working Group, Robyn E. Sanderson, Andrea Bellini, Stefano Casertano, Jessica R. Lu, Peter Melchior, Mattia Libralato, David Bennett, Michael Shao, Jason Rhodes, Sangmo Tony Sohn, Sangeeta Malhotra, Scott Gaudi, S. Michael Fall, Ed Nelan, Puragra Guhathakurta, Jay Anderson, and Shirley Ho. Astrometry with the Wide-Field Infrared Space Telescope. *Journal of Astronomical Telescopes, Instruments, and Systems*, 5:044005, October 2019. doi: 10.1117/1.JATIS.5.4.044005.

Chapter 5

CONSTRAINING THE STOCHASTIC GRAVITATION-WAVE BACKGROUND WITH PHOTOMETRIC SURVEYS

Wang, Y., Pardo, K., Doré, O., and Chang, T.-C. (2022). "Constraining the Stochastic Gravitation-wave Background with Photometric Surveys". In: *Phys.Rev.D*, 106(8):084006. DOI:10.1103/PhysRevD.106.084006.

5.1 Introduction

The successful detections of compact binary coalescence (CBC) events by the LIGO and Virgo collaborations have opened the possibility of gravitational wave (GW) astronomy [see, e.g., 5–7, 63]. Sensitive to GWs in the 10 $\sim$ 1000 Hz range, the aLIGO network and other proposed ground-based next-generation GW detectors are at the prime frequency range to detect individual mergers between stellar-mass compact objects [see e.g., 8, 56, 59]. To observe GW signals from higher-mass systems requires detector coverage in lower frequency ranges. In the decihertz to millihertz range, a series of space-based interferometer-type GW detectors, such as LISA [12], DeciGO [37] and TianQin [43], will target signals from binaries with masses between 100 $\sim$ 10^7 $M_\odot$ [see, e.g., 12, 45, 70]. The nanohertz range is covered by the Pulsar Timing Array (PTA) method, championed by several collaborations including the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) [55], the European Pulsar Timing Array (EPTA) [26], and the Parkes Pulsar Timing Array (PPTA) [38].

In addition to individual CBCs, an important detection candidate is the stochastic gravitational wave background (SGWB). A prominent SGWB source is the superposition of GWs from supermassive black hole binaries (SMBHBs) throughout cosmic history. While individual GWs can be too weak to be detectable, the combination can be sufficiently loud; indeed, it is predicted that PTAs will detect the SGWB prior to individual GW detections (see, e.g., [58]). Other contributing sources to the SGWB have also been proposed, such as from cosmic strings and phase transitions in the early universe [20, 61].

Recently, the NANOGrav 12.5-yr data analysis detected a common red noise among the observed pulsars with an amplitude of 1.92×10^{-15} at the reference frequency

$1\ \mathrm{yr}^{-1}$ (3.2×10^{-8} Hz) with high confidence [14]. Shortly afterwards, analysis of the 24-yr EPTA dataset revealed a common red noise with an amplitude of 2.95×10^{-15} at the same reference frequency, in rough agreement with the NANOGrav result [21]. Similar detection results have since been published by PPTA [30] and IPTA [13]. While these studies show insufficient evidence of the signal angular correlation, famously known as the Hellings-Downs curve [32], to positively identify the signal as the SGWB, they suggest the prospect of an imminent detection with further data collection.

While PTAs are exceptionally suited for detection in the nanohertz range, there are currently no ongoing or planned observatories to cover the SGWB in the microhertz band. In this range, we expect that the SGWB will be produced by lighter binaries ($10^5 \sim 10^9\ M_\odot$) than those seen by PTAs [see, e.g., 60]. Observing the SGWB in this regime would complement and cross-check the PTA observations, as the SGWB at different frequencies should eventually be consistent with the same population model. Recently, several potential detection methods have been proposed that target this uncovered frequency band gap. For example, GW could be detected via its modifications to asteroid accelerations [25]. It also modulates the phase of continuous GWs from galactic sources [36], from which the SGWB can be inferred. Spaced-based interferometer-type detectors are also proposed [60].

In this paper, we focus on the detection method using relative astrometric measurements [39, 48, 68]. Analogous to the periodic pulse arrival time delay in PTA, the sky position of measured objects oscillates with a passing GW, from which we infer the source properties [16, 53]. In particular, the astrometry method has the advantage of being flexible in its sensitive frequency range with no additional instrument cost to the photometric surveys [68]. By specifically referring to photometric surveys, we wish to emphasize that not only surveys dedicated to astrometry can achieve this purpose; indeed, a promising candidate, the Galactic Bulge Time Domain (GBTD) survey of the Nancy Grace Roman Space Telescope, has the primary objective to observe microlensing signatures in the galactic bulge [1].

Several authors have investigated the potential of this method, both from a theoretical perspective [16, 29, 46, 47, 54] and in close connection with specific surveys [39, 48, 68]. In Ref. [68], we investigate the detection prospect of individual monochromatic GWs in the case of the Nancy Grace Roman Space Telescope (Roman) [4] and *Gaia* [52], both offering high-precision and high-cadence astrometric measurements.

In this work, we elaborate on the effect of survey features on the sensitivity to

SGWB. Explicitly, we compare surveys depending on whether they are sensitive to a largely uniform displacement of all stars within their field of view (FoV). We clarify that, although the detection method itself requires relative astrometry only, the design for absolute astrometry telescopes contributes to higher sensitivity. We also discuss detection implications due to the size of the FoV; we apply an angular power spectrum binning and compute the recoverable signal power for given FoV size.

This paper is organized as follows. In Section 5.2, we review the fundamentals of GW-induced astrometric deflections. We then develop the expected power spectra in both the frequency domain and the spatial domain. In Section 5.3, we examine the respective survey sensitivities under various survey features. We conclude in Section 5.4.

The code used for this analysis and the figures is available at `https://github.com/kpardo/estoiles-p`

5.2 Theory

In this section, we review the basics of GW-induced astrometric deflections. We then discuss two different ways to analyze the deflections produced by the SGWB: the frequency power spectrum and the angular power spectrum. We then show the corresponding instrument noise power spectrum. We emphasize, however, that these signal and noise prescriptions alone are insufficient to gauge the GW detection power of a photometric survey; as we will show in Section 5.3, a more accurate description is contingent on additional survey features.

GW Detection with Astrometry

Upon the passing of a GW, the propagation path of photons are perturbed such that the measured star positions are deflected from their "true" positions. The deflection vector $d\vec{n}$ of the position of a star located at $\vec{n}$ on the celestial sphere is given as [16, 53]

$$dn^i = \frac{n^i + p^i}{2(1 + p_j n^j)} h_{jk} n^j n^k - \frac{1}{2} h^{ij} n_j \ , \tag{5.1}$$

where $\vec{p}$ is the propagation direction of the GW, h_{ij} is the GW strain tensor evaluated at the observer, and the Latin index ranges over the three spatial dimensions. We adopt the Einstein notation, in which repeated upper and lower indices imply summation. In Eqn. (5.1), we adopt the distant source limit, where the distance to the GW source and to the observed stars are much larger than the GW wavelength.

Consequently, the GW perturbations at different star locations are uncorrelated, and can be treated as a source of random noise [16, 53]. In this work, we are interested in GWs with frequencies ranging from $10^{-8} \sim 10^{-5}$ Hz where this distant source limit is almost always valid. Given the observable $d\vec{n}$, it is then possible to infer the source property, i.e., h_{ij}.

The astrometry GW detection method has several merits. Firstly, it is highly versatile in its sensitive frequency range, given as

$$f_{\rm min} \sim \frac{1}{T_{\rm obs}}, \; f_{\rm max} \sim \frac{1}{\Delta t} \; , \tag{5.2}$$

where $T_{\rm obs}$ is the survey lifetime, and Δt is the observational cadence. Therefore, surveys with the right cadence can potentially bridge the microhertz GW frequency spectrum.

In addition, the detection sensitivity is boosted by the number of observed stars and the number of exposures for each star. Given suitable surveys, these factors are typically quite large; for example, in the Galactic Bulge Time Domain Survey (GBTD) of Roman, each of the $\sim 10^8$ observed targets has up to 4.1×10^4 exposures [69].

Furthermore, this detection method requires astrometric measurements only, without additional equipment and observing time; in this way, the GW scientific output is serendipitous given existing surveys. Finally, we note that Eqn. (5.1) suggests that only relative astrometric measurements (i.e., away from *some* fixed reference location) are needed; it is not necessary that we know the absolute astrometric coordinates. We discuss this point in detail in Section 5.3. In the following sections, unless explicitly stated, the astrometric measurements are to be interpreted as relative ones.

Frequency Power Spectrum

One way to analyze the astrometric data is to compute the signal magnitude correlation across time, which can be represented by its Fourier transform, i.e., the SGWB frequency-domain power spectrum [see, e.g., 10, 51].

The SGWB follows the assumption that the time-domain strain amplitude follows a stationary zero-mean Gaussian distribution, and its Fourier transform is a sum of Gaussian modes with frequency-dependent variance, i.e.,

$$\begin{aligned} \langle h(t)h(t')\rangle &= \delta^D(t-t')\sigma_t^2 \\ \langle \tilde{h}(f)\tilde{h}^*(f')\rangle &= \delta^D(f-f')S_h(f) \; , \end{aligned} \tag{5.3}$$

where δ^D is the Dirac delta and σ_t is the constant standard deviation of the strain. The double-sided signal power spectrum, $S_h(f)$, can be expressed as the GW energy density per logarithmic frequency interval, $\Omega_{\rm gw}(f)$, or the characteristic strain amplitude, $h_c(f)$, with the relationship given as [51]:

$$\rho_c c^2 \Omega_{\rm gw}(f) = \frac{\pi c^2}{4G} f^2 h_c^2(f) = 2\frac{\pi c^2}{4G} f^3 S_h(f) \ , \tag{5.4}$$

and the critical density today is given by

$$\rho_c = \frac{3H_0^2}{8\pi G} \ , \tag{5.5}$$

where H_0 is the Hubble constant today.

The above formulae are independent from the specific form of the characteristic strain amplitude (or equivalent quantities). A generic phenomenological model for a source-agnostic SGWB is a power law,

$$h_c(f) \sim A \left(\frac{f}{f_{\rm ref}}\right)^{\alpha} \ , \tag{5.6}$$

completely specified by the spectral slope α and the spectrum amplitude, A, at a reference frequency $f_{\rm ref}$. The slope, in particular, is determined by the nature of the SGWB source. For example, the SGWB from inspiralling black hole binaries has a spectral slope of $-2/3$ [22, 51]; for primordial background (e.g. inflation) and cosmic strings, the spectral slopes are -1 and $-7/6$, respectively [23, 31, 41].

For a survey taking N_m exposures of N_s stars, the *effective* single-exposure single-star noise variance is given by an ensemble average as

$$\langle n_{IJ} n_{I'J'} \rangle = \frac{\sigma_{ss}^2 \delta^K_{II',JJ'}}{N_s N_m} \ , \tag{5.7}$$

where I, J are indices for the exposure and the star and σ_{ss} is the single-star single exposure astrometric noise standard deviation. The symbol δ_K is the Kronecker delta. We assume the measurement noise is both spatially and temporally uncorrelated. In analogy to the GW power spectrum, we define the noise power spectrum as

$$\langle n_{IJ} n_{I'J'} \rangle \equiv \delta^D(t_I - t_{I'}) P_n \delta^K_{JJ'} \ . \tag{5.8}$$

For a finite observation time, the discrete version of the Dirac delta is given by

$$\delta^D(t_I - t_{I'}) \to \frac{1}{\Delta t} \delta^K_{II'} \ , \tag{5.9}$$

where Δt is the (constant) time interval between exposures. It follows that

$$P_n = \Delta t \frac{\sigma_{ss}^2}{N_s N_m} \ . \tag{5.10}$$

Since the deflection signal $|d\vec{n}|$ and the GW amplitude $|h(t)|$ differ only by an $O(1)$ geometric factor, we make the approximation that $|d\vec{n}| \sim |h(t)|$, and Eqn. (5.10) is approximately the strain noise power spectrum. The corresponding characteristic amplitude is given by

$$h_n(f) = \sqrt{2fP_n} = \sigma_{ss}\sqrt{\frac{2f\Delta t}{N_s N_m}} \ , \tag{5.11}$$

where we insert a factor of 2 to convert from a double-sided power spectrum to single-sided. For a survey with fixed cadence, $h_n(f)$ scales as $1/\sqrt{T_{\rm obs}}$, consistent with Ref. [64].

Angular Power Spectrum

Aside from looking at time stream data, we may take any exposure and concentrate on the angular correlations between deflection signals at different sky locations at a given time. This method has been very broadly applied; for example, in the PTA search for SGWB, this spatial correlation is expressed in the form of the Hellings and Downs curve [32]. A similar strategy has been applied to studies of the Cosmic Microwave Background (CMB), galaxy distributions and dark matter distributions [see, e.g., 34, 40, 49, 50, 72]. We note that the most commonly analyzed signals are spin-0 scalar signals, (e.g., CMB temperature map, galaxy counts) and spin-2 tensor signals, (e.g., CMB polarization, weak lensing distortion map). The GW-induced astrometric deflection is a spin-1 vector signal, and will be decomposed with vector spherical harmonics as [29, 34, 46, 50, 54]

$$(dn)_a = \sum_{\ell m} \left[E_{\ell m} Y^E_{(\ell m),a}(\vec{n}) + B_{\ell m} Y^B_{(\ell m),a}(\vec{n}) \right] \ , \tag{5.12}$$

where a indicates two orthogonal unit vectors tangential to the celestial sphere. The E,B basis is chosen such that the coefficients $E_{\ell m}(B_{\ell m})$ transform as scalars (pseudo-scalars) under a local rotation [71].

We assume that the complex coefficients $E_{\ell m}$, $B_{\ell m}$ are drawn from zero-mean normal distributions, with the variance given by

$$\begin{aligned} \langle E_{\ell m} E^*_{\ell' m'} \rangle &= C^E_\ell \delta_{\ell\ell',mm'} \\ \langle B_{\ell m} B^*_{\ell' m'} \rangle &= C^B_\ell \delta_{\ell\ell',mm'} \ . \end{aligned} \tag{5.13}$$

Just as the characteristic strain amplitude, $h_c(f)$, indicates the SGWB source, $C_\ell^{E,B}$ depends on the nature of GW radiation itself, such as its propagation speed and polarization content [46, 47, 54]. In this work, we assume the GW travels at the speed of light and contains only tensor modes. The angular power spectrum is then given as [54]

$$\begin{aligned} C_\ell^E = C_\ell^B &= \frac{12H_0^2 N_\ell^{-2}}{\pi\ell(\ell+1)} \int df\, \frac{\Omega_{\rm gw}(f)}{f^3} |W(f)|^2 \\ &= \frac{8\pi N_\ell^{-2}}{\ell(\ell+1)} \int d\ln f\; h_c^2(f) \\ N_\ell &\equiv \sqrt{\frac{(\ell+2)!}{2(\ell-2)!}}\,. \end{aligned} \tag{5.14}$$

The window function $W(f)$ accounts for the phase difference between two exposures. This factor arises, as the model in Ref. [54] assumes two exposures only, and the "deflection" must be calculated using one of the two exposures as the baseline. In our model, we assume that a true baseline is established by averaging the measurements over the entire observational period. In this way, $W(f)$ is not necessary. We observe that the angular power spectrum is sharply peaked at small ℓ and rapidly drops off as ℓ^{-6} at large ℓ.

From an observational perspective, the angular power spectra $C_\ell^{E,B}(t_I)$ at each time slice t_I can be extracted via the inverse of Eqn. (5.12). Since this angular power spectrum is stationary, the final estimated angular power spectrum can be averaged over the exposures,

$$C_\ell^{E,B} = \frac{1}{N_m} \sum_I C_\ell^{E,B}(t_I)\,. \tag{5.15}$$

This formula can also be understood from the definition of $C_\ell^{E,B}$ as the ensemble average of $E, B_{\ell m}$; for a single exposure the average runs over the m modes, and for multiple exposures it also runs over the (independent) realizations of the angular power spectrum.

The spatial measurement noise is modeled as vectors on the two-sphere with random orientations, with magnitude drawn from a normal distribution as in Eqn. (5.7). Applying a harmonic transform, we obtain the noise power spectrum as

$$C_n^{E,B} = \frac{2\pi\sigma_{ss}^2}{N_m N_s}\,, \tag{5.16}$$

where the factor of 2π reflects the angular normalization and that the power is split evenly between the E, B modes [54].

5.3 Analysis

While Section 5.2 provides the theoretical signal power spectra, they are not necessarily representative of what can be recovered from observational data. In this section, we clarify necessary modifications to the power spectra such that they are applicable for specific types of photometric surveys. Firstly, we describe our reference surveys, the Roman GBTD survey and the *Gaia* astrometric survey. We then discuss the role of absolute and relative astrometry in GW detection. We discuss in depth the effect of subtracting the FoV mean signal and the resulting survey performance. Lastly we introduce angular power binning and explain the implication of limited FoV on GW detection.

Survey Summary

The Nancy Grace Roman Space Telescope[1] is NASA's next flagship observatory after the James Webb Space Telescope. Among other science goals, it aims to probe the evolution of dark energy and large-scale structure by observing billions of galaxies and thousands of supernovae [9]. In terms of probing GWs, the GBTD survey is particularly relevant, where it visits a ~ 1.97 deg^2 patch of sky towards the galactic center. This pointing direction implies large stellar density, and the near infrared sensitive wavelength also leads to less extinction; thanks to these factors, Roman is expected to observe 10^8 stars ($W145_{AB} < 23$) [28] with the GBTD survey, with a single-exposure single-star astrometric uncertainty of 1.1 mas (estimated for $H_{\rm AB} = 21.6$ stars) [69]. The survey comprises of 6 observing seasons, each 72 days long. During each season, each star is visited every 15 minutes, giving a total of 41,000 exposures per star.

Another promising mission as a potential GW probe is the all-sky astrometric survey with *Gaia*[2] [52]. *Gaia* observes on the order of 10^9 stars [17], with a single-exposure single-star astrometric uncertainty of 0.7 mas for $G \sim 20$ stars in *Gaia* Data Release 2 [44]. In the recent *Gaia* Early Data Release 3, the typical uncertainty for six-parameter astrometry (position) is 0.4 mas at $G = 20$ [27]. *Gaia* Data Release 4, which will be based on data during the entire nominal mission lifetime and part of the extended mission, expects a parallax uncertainty of 0.46 mas at $G = 20$

[1] https://roman.gsfc.nasa.gov/
[2] https://sci.esa.int/web/gaia

[3]. *Gaia* also expects to achieve a ten-year total lifetime in the extended mission [18]; the ultimate Data Release 5 expects to have a parallax uncertainty of 0.33 mas for objects of the same magnitude [3] (the astrometric position uncertainty generally has a small difference in value with parallax uncertainty [see, e.g., 27]). Since the difference is within an order of magnitude, we do not update the estimates accordingly, and the general conclusion remains the same. On average, each star is visited 70 times throughout the 5-year observing time [52]. For simplicity, we assume these exposures to be evenly spaced.

Absolute/Relative Astrometry and Mean Subtraction

As is evident from Eqn. (5.3), the SGWB inference relies solely on the deflection vectors and not on the true position vector; indeed, if we observe cycles of a single continuous wave, or a significant period of the stochastic process, the true position can be immediately computed. Therefore, it is not required to have absolute astrometric measurements (as *Gaia* provides), consisting of the absolute star coordinates in, e.g., the extragalactic International Celestial Reference System [42].

What is required, on the other hand, is that the entire deflection vectors should be recovered from the time stream data. This is particularly challenging; since the GW deflection pattern is a large scale signal (see Eqn. (5.14)), deflections within a small FoV appear almost uniform. We shall refer to this almost uniform motion as the FoV mean signal.

For telescopes with a single viewing direction, the mean signal is recoverable under two possibilities. Evidently, a telescope that is in free fall during data collection qualifies, since an inertial frame cannot "absorb" periodic motion. For a "point-and-stare" type telescope like Roman, its reaction wheels are constantly engaged through the fine guidance system for pointing self-calibration [69]. As the mean signal mimics instrument noises such as pointing error and jitter, it may then be corrected *in situ* or get fitted out during the astrometric solution process. In this case, it is in principle possible to reconstruct the mean signal (or part of it) if the pointing system actions are recorded.

In the worst case scenario, we ignore the mean signal and use only the differential deflections across each exposures, which we refer to as the mean-subtracted case. In Ref. [68], we considered this case for individual continuous wave detection and showed that the differential deflections are roughly two orders of magnitude smaller than the mean signal. The sensitivity loss can be approximated by scaling down the

signal characteristic strain amplitude proportionally.

For telescopes with two almost orthogonal viewing directions, such as *Gaia*, the GW signature is typically towards two distinctive directions in the two viewing directions. In this case, the signal is less likely to be absorbed as pointing error. Although the *Gaia* viewing directions were not specifically designed for this purpose, it incidentally satisfies the favoring conditions.

In the following, we consider the signal-to-noise ratio (SNR) defined as (see, e.g., [48])

$$\rho^2 = \int d\ln f \left[\frac{\lambda h_c(f)}{h_n(f)}\right]^2 , \tag{5.17}$$

where h_c is the SGWB characteristic strain amplitude and h_n is the characteristic noise amplitude in Eqn. (5.11), and the sensitivity threshold is fixed to have $\rho = 1$. The scaling factor λ accounts for the signal loss due to mean subtraction. For *Gaia*, we assume $\lambda = 1$, i.e. lossless. As the Roman telescope and system design is still in planning, we do not quote a specific value; instead, we show the range $\lambda = 0.01 \sim 1$, which corresponds to the mean-subtracted case and the full signal case, respectively [68].

We show the sensitivity threshold in Figure 5.1, where we do not restrict to the SMBHB SGWB, but rather assume a source-agnostic search [64]. For reference, we also plot the sensitivity curves of IPTA [62, 66] and LISA [57] in gray using phenomenological models. The solid red line shows the best estimate of the common process measured from IPTA DR2 (over $10^{-9} - 4\times10^{-8}$ Hz, consistent with Figure 1 of Ref. [13]),

$$h_{c,\mathrm{IPTA}} \sim 2.8\times10^{-15}\left(\frac{f}{1\ \mathrm{yr}^{-1}}\right)^{-2/3} . \tag{5.18}$$

We note that the population details of potential SMBHBs emitting GWs from 10^{-7} Hz to 10^{-4} Hz are highly uncertain and poorly constrained by observation, if at all. As a heuristic example, we simply extrapolate the SGWB from the nanohertz range (red dashed line). While this treatment is valid until some of the more massive GW sources go near coalescence, it suffices in this work as an example to possible SGWB signals across the wide frequency ranges.

We reiterate that the frequency band difference between Roman and *Gaia* results from the observational cadence (see Eqn. (5.2)). The sensitivity, in addition to the cadence, is affected by the astrometric accuracy and the number of observed stars. Specifically, we observe that within shared frequency ranges, Roman offers better

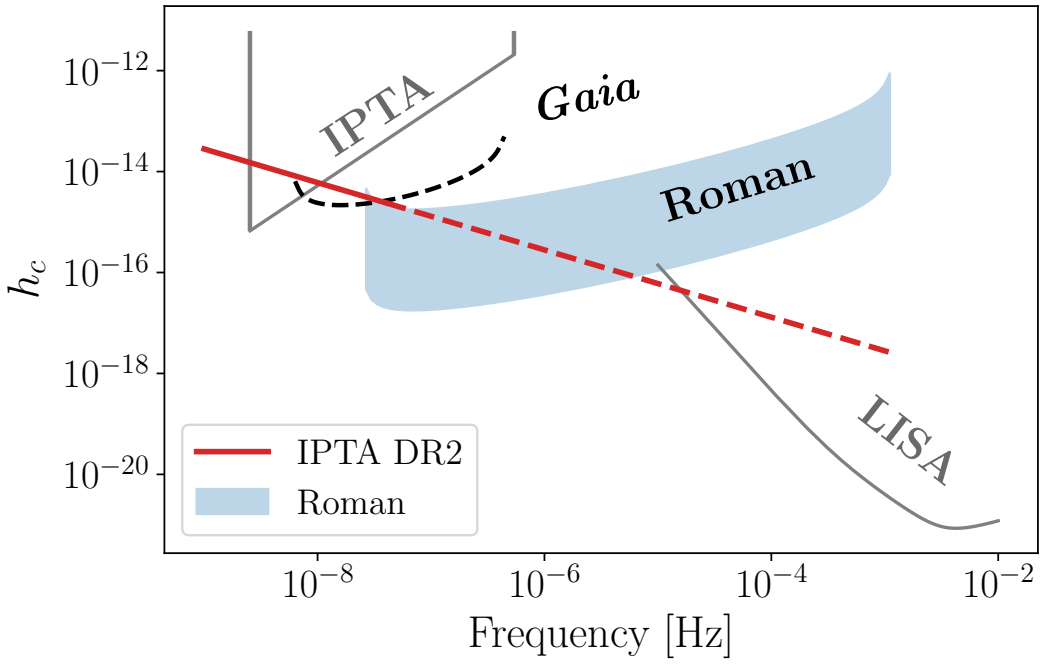

Figure 5.1: Survey sensitivity for a source-agnostic SGWB search. The survey sensitivity of IPTA and LISA are shown in gray curves. The solid black line shows sensitivity of *Gaia*. The range of sensitivity of Roman GBTD survey is shown with the blue solid block, with $\lambda = 0.01 \sim 1$. The solid red line shows the best amplitude estimate from IPTA DR2 with similar frequency range. The dashed red line shows the extrapolated SGWB level in the Roman frequency range.

sensitivity than *Gaia*, even in the mean-subtracted scenario. This is primarily due to its high cadence and larger span in its sensitive frequency. We note that if the SMBHB SGWB is indeed at a similar level as the current best estimate from IPTA, both *Gaia* and Roman can detect its power excess.

Given a fixed redshift distribution, the SGWB amplitude directly implies the local SMBHB remnant density, $\Phi_{\mathrm{BHB},0}$, defined as the number of SMBHB remnant per comoving volume in the local universe. Assuming the same SMBHB population model as in Ref. [13, 19], we can also express the Roman and *Gaia* sensitivity in terms of an upper limit on the local SMBHB number density (in the case of null detection), since $A \sim \sqrt{\Phi_{\mathrm{BHB},0}}$ in Eqn. (5.6). For Roman, the upper limit is $1.0 \times 10^{-5} \sim 1.0 \times 10^{-9}$ Mpc^{-3} for $\lambda = 0.01 \sim 1$. For *Gaia*, the upper limit is 2.3×10^{-6} Mpc^{-3}.

Angular Power Binning

In this section, we discuss the effect of telescope FoV size on the angular power spectrum sensitivity. Firstly, we note that the currently planned telescopes with sufficient astrometric precision for GW measurement are all space-based telescopes with FoV less than 1 deg^2; e.g., Roman (0.26 deg^2) [69], *Gaia* (0.5 deg^2) [2], James Webb Space Telescope ($\sim$ 8 arcmin^2)[3],[4], Hubble Space Telescope (WFC3 NIR, $\sim$ 4 arcmin^2)[5]. Therefore, it is not likely that we can capture GW deflections with significant spatial variation within one exposure. We further discuss in turn the "point-and-stare" type surveys (e.g., Roman) and all-sky scanning surveys (e.g., *Gaia*).

In the case of Roman GBTD survey, which only visits fields close to the galactic center, the measurements are insensitive to angular powers with a scale larger than the FoV size[6]. Moreover, different (ℓ, m) modes of the spherical harmonic decomposition coefficients can be highly coupled. While the exact coupling depends on the FoV geometry, modes with larger difference in value than $\pi/\theta_{\rm fov}$ have much smaller coupling, where $\theta_{\rm fov}$ is the (angular) sidelength of the FoV.

Although *Gaia* offers complete sky coverage after each full-sky scan, it is unclear if it will not suffer the same angular power loss. In the case of the SGWB, the spatial deflection patterns at different times are, by assumption, independent realizations of $C_\ell^{E,B}$. The image of the FoV are also independent, as a consequence. In this way, measurements from different exposures cannot be consistently combined to produce a full-sky map, and *Gaia* suffers the same large-scale power mean-signal loss as Roman.

In the case of individual continuous GW signals, the signal coherence allows, in principle, the construction of a temporal-spatial template, which depends on the GW source property and the attitude history of the telescope [42]. A thorough investigation of such a possibility is beyond the scope of this work. In the following analysis, we shall assume that the large-scale powers cannot be recovered from *Gaia* measurements.

[3] https://www.jwst.nasa.gov/

[4] https://svs.gsfc.nasa.gov/13583

[5] https://hubblesite.org/

[6] In principle, measurements from the several fields Roman visits may be combined on particular time scales (e.g., much shorter than the typical SGWB period of interest). Such reconstruction would require careful modeling of the telescope motion during field-switching; for simplicity, we restrict to scales smaller than the FoV.

To account for the large-scale power loss and mode mixing, we bin the theoretical angular power spectrum and impose a minimum ℓ. The maximum ℓ roughly corresponds to pixel scale, which is much larger than any ℓ-modes contributing significantly to the SGWB, i.e., $\ell \sim 2$. The binned, predicted angular power spectrum can be obtained via the exact mode-coupling matrix of the FoV (i.e., the window) [11, 33, 67]; however, to simplify the calculation and to keep the estimate applicable to generic FoV shapes, we approximate the binning process by directly averaging the theoretical $C_\ell^{E,B}$ within each bin [see, e.g., 33]. Given sufficiently wide bins, different bins have negligible coupling.

We consider a square FoV, where the bins are defined as

$$\begin{aligned} C_q^{E,B} &= \frac{1}{\Delta q} \sum_{\ell_i \in \{\ell_q\}} C_{\ell_i}^{E,B}, \ \Delta q = \left[\frac{\pi}{\theta_{\rm fov}}\right], \\ \{\ell_q\} &= \{\ell_i, \left[q - \frac{1}{2}\right] \Delta q \leq i < \left[q + \frac{1}{2}\right] \Delta q\} \ , \end{aligned} \tag{5.19}$$

where $[\cdot]$ denotes taking the nearest integer and $\theta_{\rm fov}$ is the angular sidelength of the square FoV.

Since the survey is not sensitive to modes larger than the FoV, we discard the first bin when calculating the SNR. With these two steps, the log likelihood is summed over the binned modes as (see, e.g., [34])

$$\begin{aligned} \rho_{E,B}^2 &= \sum_q \left(\frac{C_q^{E,B}}{\Delta C_q^{E,B}}\right)^2 \\ \rho^2 &= \rho_E^2 + \rho_B^2 \ . \end{aligned} \tag{5.20}$$

In the weak signal limit, the diagonal elements of the covariance matrix, $1/\Delta C_q^{E,B}$, are dominated by the noise, instead of cosmic variance. The inverse of the covariance for each bin is given by [24]

$$\Delta C_q^{E,B} = \frac{1}{\Delta_q} \sqrt{\frac{2}{(2\ell_q + 1) f_{\rm sky}}} C_n^{E,B} \ , \tag{5.21}$$

where ℓ_q is the mode in the center of each bin. The factor $f_{\rm sky}$ if the fraction of sky covered by the FoV, and it accounts for the loss of mode power due to partial sky coverage.

It is clear that the SNR loss due to limited FoV is also sensitive to the signal spectral slope; signal concentrated at large scales suffers a higher loss of SNR compared with

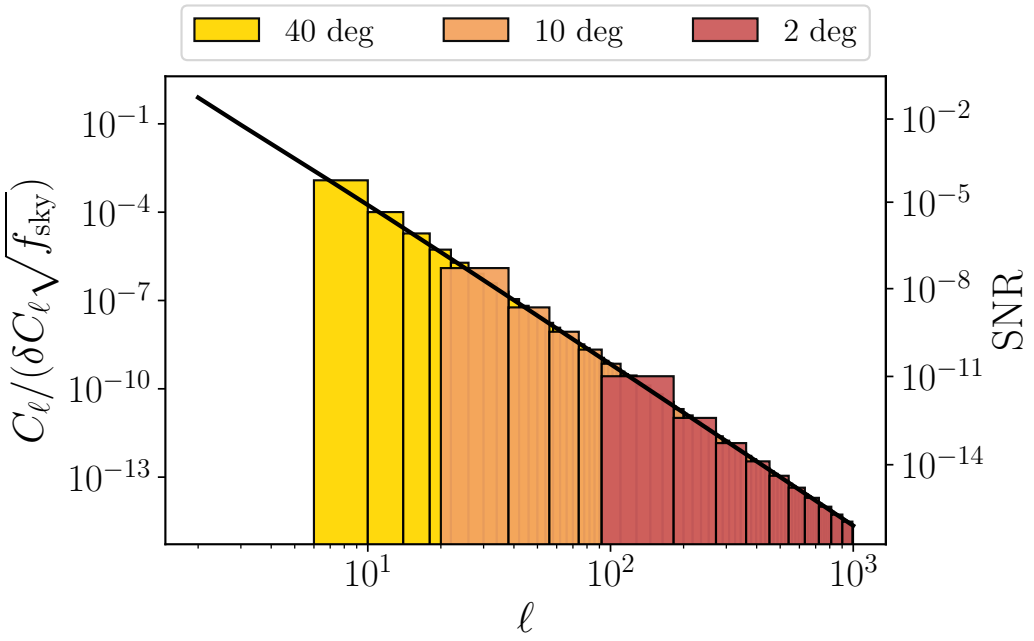

Figure 5.2: The fractional SNR for various FoV size compared with the full sky case. The bar plot corresponds to the left axis and shows the SNR contribution from each bin, normalized by $f_{\rm sky}$. The black solid line shows the full-sky scenario. Using the right axis, the same line shows the total SNR with various threshold $\ell = \pi/\theta_{\rm fov}$ as a fraction of the full sky scenario.

signals that have a flatter spectral profile. As Eqn. (5.14) shows, the angular power spectrum of SGWB decays roughly as ℓ^{-6}, which strongly penalizes the detecting power of limited-FoV surveys.

Figure 5.2 shows the SNR, normalized to the full-sky case, given various FoV sizes. The bar plot corresponds to the left axis, showing the SNR contribution from each bin (see Eqn. (5.19); for ease of presentation, we have normalized all three cases by dividing out the $f_{\rm sky}$ factor). The black solid line shows the full-sky scenario. The right axis and the same black solid line gives the accumulated SNR given various threshold ℓ. We observe that the SNR drops sharply with limited FoV; in the case of Roman with $\theta_{\rm fov} = 0.53$ deg, the fractional SNR is smaller than 10^{-14} compared to the SNR of the full-sky case.

Therefore, only full-sky or nearly full-sky surveys are likely to be sensitive to the angular power spectrum of SGWB. If the SNR is sufficiently large, this method can in principle become a probe for modified gravity theories, as is discussed in Refs. [29, 46, 47, 54]. Just as this method complements PTAs in the frequency domain, it

would also cross-check the non-GR effect with the timing residual measurements. To accomplish this goal, it requires a survey with a nearly all-sky FoV, in additional to sufficient cadence and astrometric accuracy.

Lastly, we comment on the detection prospect difference in frequency domain and angular domain. Since $\tilde{h}(f)$ and $C_\ell^{E,B}$ are decomposition of the same signal, the total signal power, i.e., $\int d\ln f\ h_c^2(f)$, $\sum_\ell C_\ell^E + C_\ell^B$, matches (see Eqn. (5.14)). However, the achievable SNRs are limited differently for these two domains; the frequency domain is limited by observational cadence and the angular sensitivity is limited by sky coverage. In particular, the poor angular domain SNR is attributed to the mismatch between the large-scale nature of GW signal and the limited FoV scale.

5.4 Conclusion

In this work, we examine the SGWB detection prospects using astrometric measurements, with discussion of the effects of telescope features on the sensitivity. We consider two especially promising and representative surveys, the Roman GBTD survey and the *Gaia* all-sky astrometric survey, for in-depth discussions and concrete performance forecasts.

We highlight that whether a uniform deflection signal can be extracted from observational data has a high impact on detector sensitivity. Although absolute astrometry is not required for GW detection, the dedicated telescope design (i.e., two viewing directions) is beneficial for keeping the mean GW signal. For single viewing angle telescopes, it is more likely that we can only use differential deflection data. Thanks to its high-cadence of observations, we found that Roman GBTD survey is sensitive to the SGWB with an SNR of $1.2 - 120$ for frequency power spectrum analysis, depending on whether the mean signal can be captured, if the background is indeed at the level estimated by current PTA efforts. We found that *Gaia* is able to detect the same signal with an SNR of 2.5.

We explain the signal loss due to FoV size by binning the power spectrum and disregarding inaccessible bin powers. As the GW signal is intrinsically a large-scale (quadrupole, in particular) signal, and telescopes with the required astrometric accuracy typically have small FoVs, it is unlikely that this detection method is sensitive to the GW angular patterns. To probe the angular power spectrum and potentially observe non-GR signatures, the future candidates should have significant sky coverage.

As the next generation of photometric telescopes start to go online or go into detailed planning, the possibility of using astrometric measurements to complement existing GW detection strategies is ever more promising. As is shown, Roman and *Gaia* can offer supporting evidence for PTA measurements of the SGWB and probe a complementary SMBHB population in a previously inaccessible frequency range.

Software: astropy [15], matplotlib [35], numpy [65]

References

[1] Galactic Bulge Time Domain Survey. URL https://roman.gsfc.nasa.gov/galactic_bulge_time_domain_survey.html.

[2] Gaia-Taking the Galactic Census, 2006. URL https://www.astro.umd.edu/~olling/Papers/GAIA3_IN_all_info_sheets.pdf.

[3] Expected Science Performance for the Nominal and Extended Mission, 2020. URL https://www.cosmos.esa.int/web/gaia/science-performance.

[4] Roman's Key Science Components, 2020. URL https://www.stsci.edu/roman/about/science-themes.

[5] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, LIGO Scientific Collaboration, and Virgo Collaboration. GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. *Phys. Rev. D*, 93(12):122003, June 2016. doi: 10.1103/PhysRevD.93.122003.

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

[7] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, others, LIGO Scientific Collaboration, and Virgo Collaboration. GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo during the First Half of the Third Observing Run. *Physical Review X*, 11(2):021053, April 2021. doi: 10.1103/PhysRevX.11.021053.

[8] R. X. Adhikari, K. Arai, A. F. Brooks, C. Wipf, O. Aguiar, P. Altin, B. Barr, L. Barsotti, R. Bassiri, A. Bell, et al. A cryogenic silicon interferometer for gravitational-wave detection. *Classical and Quantum Gravity*, 37(16):165003, August 2020. doi: 10.1088/1361-6382/ab9143.

[9] Aniket Agrawal, Teppei Okumura, and Toshifumi Futamase. Constraining neutrino mass and dark energy with peculiar velocities and lensing dispersions

of Type Ia supernovae. *Phys. Rev. D*, 100(6):063534, September 2019. doi: 10.1103/PhysRevD.100.063534.

[10] Bruce Allen and Joseph D. Romano. Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities. *Phys. Rev. D*, 59(10):102001, May 1999. doi: 10.1103/PhysRevD.59.102001.

[11] David Alonso, Javier Sanchez, Anže Slosar, and LSST Dark Energy Science Collaboration. A unified pseudo-C_ℓ framework. *MNRAS*, 484(3):4127–4151, April 2019. doi: 10.1093/mnras/stz093.

[12] Pau Amaro-Seoane, Heather Audley, Stanislav Babak, John Baker, Enrico Barausse, et al. Laser Interferometer Space Antenna. *arXiv e-prints*, art. arXiv:1702.00786, February 2017.

[13] J. Antoniadis, Z. Arzoumanian, S. Babak, M. Bailes, A. S. Bak Nielsen, P. T. Baker, C. G. Bassa, B. Bécsy, et al. The International Pulsar Timing Array second data release: Search for an isotropic gravitational wave background. *MNRAS*, 510(4):4873–4887, March 2022. doi: 10.1093/mnras/stab3418.

[14] Zaven Arzoumanian, Paul T. Baker, Harsha Blumer, Bence Bécsy, Adam Brazier, Paul R. Brook, Sarah Burke-Spolaor, others, and Nanograv Collaboration. The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background. *ApJ*, 905(2):L34, December 2020. doi: 10.3847/2041-8213/abd401.

[15] Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, A. Ginsburg, J. T. Vand erPlas, L. D. Bradley, D. Pérez-Suárez, M. de Val-Borro, T. L. Aldcroft, K. L. Cruz, T. P. Robitaille, E. J. Tollerud, C. Ardelean, T. Babej, others, and Astropy Contributors. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *AJ*, 156(3):123, September 2018. doi: 10.3847/1538-3881/aabc4f.

[16] Laura G. Book and Éanna É. Flanagan. Astrometric effects of a stochastic gravitational wave background. *Phys. Rev. D*, 83(2):024024, January 2011. doi: 10.1103/PhysRevD.83.024024.

[17] A. G. A. Brown et al. Gaia Data Release 2. Summary of the contents and survey properties. *A&A*, 616:A1, August 2018. doi: 10.1051/0004-6361/201833051.

[18] Anthony G. A. Brown. Microarcsecond Astrometry: Science Highlights from Gaia. *ARA&A*, 59, September 2021. doi: 10.1146/annurev-astro-112320-035628.

[19] J. Andrew Casey-Clyde, Chiara M. F. Mingarelli, Jenny E. Greene, Kris Pardo, Morgan Nañez, and Andy D. Goulding. A Quasar-based Supermassive Black Hole Binary Population Model: Implications for the Gravitational Wave Background. *ApJ*, 924(2):93, January 2022. doi: 10.3847/1538-4357/ac32de.

[20] Chia-Feng Chang and Yanou Cui. Stochastic gravitational wave background from global cosmic strings. *Physics of the Dark Universe*, 29:100604, September 2020. doi: 10.1016/j.dark.2020.100604.

[21] S. Chen, R. N. Caballero, Y. J. Guo, A. Chalumeau, K. Liu, G. Shaifullah, K. J. Lee, et al. Common-red-signal analysis with 24-yr high-precision timing of the European Pulsar Timing Array: inferences in the stochastic gravitational-wave background search. *MNRAS*, 508(4):4970–4993, December 2021. doi: 10.1093/mnras/stab2833.

[22] Curt Cutler and Eanna 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.

[23] Thibault Damour and Alexander Vilenkin. Gravitational radiation from cosmic (super)strings: Bursts, stochastic background, and observational windows. *Phys. Rev. D*, 71:063510, 2005. doi: 10.1103/PhysRevD.71.063510.

[24] Scott Dodelson. *Modern cosmology*. 2003.

[25] Michael A. Fedderke, Peter W. Graham, and Surjeet Rajendran. Asteroids for microhertz gravitational-wave detection. *arXiv e-prints*, art. arXiv:2112.11431, December 2021.

[26] R. D. Ferdman, R. van Haasteren, C. G. Bassa, M. Burgay, I. Cognard, A. Corongiu, et al. The European Pulsar Timing Array: current efforts and a LEAP toward the future. *Classical and Quantum Gravity*, 27(8):084014, April 2010. doi: 10.1088/0264-9381/27/8/084014.

[27] Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, M. Biermann, O. L. Creevey, et al. Gaia Early Data Release 3. Summary of the contents and survey properties. *A&A*, 649:A1, May 2021. doi: 10.1051/0004-6361/202039657.

[28] B. Scott Gaudi, Rachel Akeson, Jay Anderson, Etienne Bachelet, David P. Bennett, Aparna Bhattacharya, Valerio Bozza, et al. 'Auxiliary' Science with the WFIRST Microlensing Survey. *BAAS*, 51(3):211, May 2019.

[29] Sebastian Golat and Carlo R. Contaldi. All-sky analysis of astrochronometric signals induced by gravitational waves. *arXiv e-prints*, art. arXiv:2201.03903, January 2022.

[30] Boris Goncharov, R. M. Shannon, D. J. Reardon, G. Hobbs, A. Zic, et al. On the Evidence for a Common-spectrum Process in the Search for the Nanohertz Gravitational-wave Background with the Parkes Pulsar Timing Array. *ApJ*, 917(2):L19, August 2021. doi: 10.3847/2041-8213/ac17f4.

[31] Leonid P. Grishchuk. REVIEWS OF TOPICAL PROBLEMS: Relic gravitational waves and cosmology. *Physics Uspekhi*, 48(12):1235–1247, December 2005. doi: 10.1070/PU2005v048n12ABEH005795.

[32] R. W. Hellings and G. S. Downs. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. *ApJ*, 265:L39–L42, February 1983. doi: 10.1086/183954.

[33] Eric Hivon, Krzysztof M. Górski, C. Barth Netterfield, Brendan P. Crill, Simon Prunet, and Frode Hansen. MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets. *ApJ*, 567(1):2–17, March 2002. doi: 10.1086/338126.

[34] Wayne Hu. Weak lensing of the CMB: A harmonic approach. *Phys. Rev. D*, 62(4):043007, August 2000. doi: 10.1103/PhysRevD.62.043007.

[35] John D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science and Engineering*, 9(3):90–95, May 2007. doi: 10.1109/MCSE.2007.55.

[36] María José Bustamante-Rosell, Joel Meyers, Noah Pearson, Cynthia Trendafilova, and Aaron Zimmerman. Gravitational Wave Timing Array. *arXiv e-prints*, art. arXiv:2107.02788, July 2021.

[37] Seiji Kawamura, Takashi Nakamura, Masaki Ando, Naoki Seto, Tomotada Akutsu, et al. Space gravitational-wave antennas DECIGO and B-DECIGO. *International Journal of Modern Physics D*, 28(12):1845001, January 2019. doi: 10.1142/S0218271818450013.

[38] Matthew Kerr, Daniel J. Reardon, George Hobbs, Ryan M. Shannon, Richard N. Manchester, Shi Dai, Christopher J. Russell, Songbo Zhang, et al. The Parkes Pulsar Timing Array project: second data release. *PASA*, 37:e020, June 2020. doi: 10.1017/pasa.2020.11.

[39] Sergei A. Klioner. Gaia-like astrometry and gravitational waves. *Classical and Quantum Gravity*, 35(4):045005, February 2018. doi: 10.1088/1361-6382/aa9f57.

[40] L. Legrand, C. Hernández-Monteagudo, M. Douspis, N. Aghanim, and Raúl E. Angulo. High-resolution tomography for galaxy spectroscopic surveys with angular redshift fluctuations. *A&A*, 646:A109, February 2021. doi: 10.1051/0004-6361/202039049.

[41] L. Lentati, S. R. Taylor, C. M. F. Mingarelli, A. Sesana, S. A. Sanidas, A. Vecchio, et al. European Pulsar Timing Array limits on an isotropic stochastic gravitational-wave background. *MNRAS*, 453(3):2576–2598, November 2015. doi: 10.1093/mnras/stv1538.

[42] L. Lindegren, J. Hernández, A. Bombrun, S. Klioner, U. Bastian, M. Ramos-Lerate, A. de Torres, et al. Gaia Data Release 2. The astrometric solution. *A&A*, 616:A2, August 2018. doi: 10.1051/0004-6361/201832727.

[43] Jun Luo, Li-Sheng Chen, Hui-Zong Duan, Yun-Gui Gong, Shoucun Hu, Jianghui Ji, Qi Liu, et al. TianQin: a space-borne gravitational wave detector. *Classical and Quantum Gravity*, 33(3):035010, February 2016. doi: 10.1088/0264-9381/33/3/035010.

[44] X. Luri, A. G. A. Brown, L. M. Sarro, F. Arenou, C. A. L. Bailer-Jones, A. Castro-Ginard, J. de Bruijne, T. Prusti, C. Babusiaux, and H. E. Delgado. Gaia Data Release 2. Using Gaia parallaxes. *A&A*, 616:A9, August 2018. doi: 10.1051/0004-6361/201832964.

[45] Jianwei Mei, Yan-Zheng Bai, Jiahui Bao, Enrico Barausse, Lin Cai, Enrico Canuto, Bin Cao, Wei-Ming Chen, et al. The TianQin project: Current progress on science and technology. *Progress of Theoretical and Experimental Physics*, 2021(5):05A107, May 2021. doi: 10.1093/ptep/ptaa114.

[46] Deyan P. Mihaylov, Christopher J. Moore, Jonathan R. Gair, Anthony Lasenby, and Gerard Gilmore. Astrometric Effects of Gravitational Wave Backgrounds with non-Einsteinian Polarizations. *Phys. Rev. D*, 97(12):124058, 2018. doi: 10.1103/PhysRevD.97.124058.

[47] Deyan P. Mihaylov, Christopher J. Moore, Jonathan R. Gair, Anthony Lasenby, and Gerard Gilmore. Astrometric effects of gravitational wave backgrounds with nonluminal propagation speeds. *Phys. Rev. D*, 101(2):024038, January 2020. doi: 10.1103/PhysRevD.101.024038.

[48] Christopher J. Moore, Deyan P. Mihaylov, Anthony Lasenby, and Gerard Gilmore. Astrometric Search Method for Individually Resolvable Gravitational Wave Sources with Gaia. *Phys. Rev. Lett.*, 119(26):261102, December 2017. doi: 10.1103/PhysRevLett.119.261102.

[49] Andrina Nicola, David Alonso, Javier Sánchez, Anže Slosar, Humna Awan, Adam Broussard, Jo Dunkley, Eric Gawiser, Zahra Gomes, Rachel Mandelbaum, Hironao Miyatake, Jeffrey A. Newman, Ignacio Sevilla-Noarbe, Sarah Skinner, and Erika L. Wagoner. Tomographic galaxy clustering with the Subaru Hyper Suprime-Cam first year public data release. *J. Cosmology Astropart. Phys.*, 2020(3):044, March 2020. doi: 10.1088/1475-7516/2020/03/044.

[50] Takemi Okamoto and Wayne Hu. Cosmic microwave background lensing reconstruction on the full sky. *Phys. Rev. D*, 67(8):083002, April 2003. doi: 10.1103/PhysRevD.67.083002.

[51] E. S. Phinney. A Practical Theorem on Gravitational Wave Backgrounds. *arXiv e-prints*, art. astro-ph/0108028, August 2001.

[52] T. Prusti et al. The Gaia mission. *A&A*, 595:A1, November 2016. doi: 10.1051/0004-6361/201629272.

[53] Ted Pyne, Carl R. Gwinn, Mark Birkinshaw, T. Marshall Eubanks, and Demetrios N. Matsakis. Gravitational Radiation and Very Long Baseline Interferometry. *ApJ*, 465:566, July 1996. doi: 10.1086/177443.

[54] Wenzer Qin, Kimberly K. Boddy, Marc Kamionkowski, and Liang Dai. Pulsar-timing arrays, astrometry, and gravitational waves. *Phys. Rev. D*, 99(6):063002, March 2019. doi: 10.1103/PhysRevD.99.063002.

[55] Scott Ransom, A. Brazier, S. Chatterjee, T. Cohen, J. M. Cordes, M. E. DeCesar, P. B. Demorest, J. S. Hazboun, M. T. Lam, R. S. Lynch, M. A. McLaughlin, S. M. Ransom, X. Siemens, S. R. Taylor, and S. J. Vigeland. The NANOGrav Program for Gravitational Waves and Fundamental Physics. In *Bulletin of the American Astronomical Society*, volume 51, page 195, September 2019.

[56] David Reitze, Rana X. Adhikari, Stefan Ballmer, Barry Barish, Lisa Barsotti, GariLynn Billingsley, Duncan A. Brown, Yanbei Chen, et al. Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. In *Bulletin of the American Astronomical Society*, volume 51, page 35, September 2019.

[57] Travis Robson, Neil J. Cornish, and Chang Liu. The construction and use of LISA sensitivity curves. *Classical and Quantum Gravity*, 36(10):105011, May 2019. doi: 10.1088/1361-6382/ab1101.

[58] Pablo A. Rosado, Alberto Sesana, and Jonathan Gair. Expected properties of the first gravitational wave signal detected with pulsar timing arrays. *MNRAS*, 451(3):2417–2433, August 2015. doi: 10.1093/mnras/stv1098.

[59] B. Sathyaprakash, M. Abernathy, F. Acernese, P. Amaro-Seoane, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker, N. Beveridge, S. Birindelli, S. Bose, L. Bosi, S. Braccini, C. Bradaschia, T. Bulik, et al. Scientific Potential of Einstein Telescope. *arXiv e-prints*, art. arXiv:1108.1423, August 2011.

[60] Alberto Sesana, Natalia Korsakova, Manuel Arca Sedda, Vishal Baibhav, Enrico Barausse, Simon Barke, Emanuele Berti, et al. Unveiling the gravitational universe at μ-Hz frequencies. *Experimental Astronomy*, 51(3):1333–1383, June 2021. doi: 10.1007/s10686-021-09709-9.

[61] Xavier Siemens, Jeffrey Hazboun, Paul T. Baker, Sarah Burke-Spolaor, Dustin R. Madison, Chiara Mingarelli, Joseph Simon, and Tristan Smith. Physics Beyond the Standard Model With Pulsar Timing Arrays. *BAAS*, 51(3): 437, May 2019.

[62] S. R. Taylor, M. Vallisneri, J. A. Ellis, C. M. F. Mingarelli, T. J. W. Lazio, and R. van Haasteren. Are We There Yet? Time to Detection of Nanohertz Gravitational Waves Based on Pulsar-timing Array Limits. *ApJ*, 819(1):L6, March 2016. doi: 10.3847/2041-8205/819/1/L6.

[63] The LIGO Scientific Collaboration and the Virgo Collaboration. GW190412: Observation of a Binary-Black-Hole Coalescence with Asymmetric Masses. *arXiv e-prints*, art. arXiv:2004.08342, April 2020.

[64] Eric Thrane and Joseph D. Romano. Sensitivity curves for searches for gravitational-wave backgrounds. *Phys. Rev. D*, 88(12):124032, December 2013. doi: 10.1103/PhysRevD.88.124032.

[65] Stéfan van der Walt, S. Chris Colbert, and Gaël Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science and Engineering*, 13(2):22–30, March 2011. doi: 10.1109/MCSE.2011.37.

[66] J. P. W. Verbiest, L. Lentati, G. Hobbs, R. van Haasteren, P. B. Demorest, et al. The International Pulsar Timing Array: First data release. *MNRAS*, 458(2): 1267–1288, May 2016. doi: 10.1093/mnras/stw347.

[67] Benjamin D. Wandelt, Eric Hivon, and Krzysztof M. Górski. Cosmic microwave background anisotropy power spectrum statistics for high precision cosmology. *Phys. Rev. D*, 64(8):083003, October 2001. doi: 10.1103/PhysRevD.64.083003.

[68] Yijun Wang, Kris Pardo, Tzu-Ching Chang, and Olivier Doré. Gravitational wave detection with photometric surveys. *Phys. Rev. D*, 103(8):084007, April 2021. doi: 10.1103/PhysRevD.103.084007.

[69] WFIRST Astrometry Working Group, Robyn E. Sanderson, Andrea Bellini, Stefano Casertano, Jessica R. Lu, Peter Melchior, Mattia Libralato, David Bennett, Michael Shao, Jason Rhodes, Sangmo Tony Sohn, Sangeeta Malhotra, Scott Gaudi, S. Michael Fall, Ed Nelan, Puragra Guhathakurta, Jay Anderson, and Shirley Ho. Astrometry with the Wide-Field Infrared Space Telescope. *Journal of Astronomical Telescopes, Instruments, and Systems*, 5:044005, October 2019. doi: 10.1117/1.JATIS.5.4.044005.

[70] Kent Yagi. Scientific Potential of Decigo Pathfinder and Testing GR with Space-Borne Gravitational Wave Interferometers. *International Journal of Modern Physics D*, 22(1):1341013, January 2013. doi: 10.1142/S0218271813410137.

[71] Matias Zaldarriaga. Nature of the E-B decomposition of CMB polarization. *Phys. Rev. D*, 64(10):103001, November 2001. doi: 10.1103/PhysRevD.64.103001.

[72] Matias Zaldarriaga and Uroš Seljak. All-sky analysis of polarization in the microwave background. *Phys. Rev. D*, 55(4):1830–1840, February 1997. doi: 10.1103/PhysRevD.55.1830.

Chapter 6

TOLERANCE TO ASTROPHYSICAL MODEL UNCERTAINTY IN GRAVITATIONAL-WAVE DARK SIREN HUBBLE MEASUREMENT

6.1 Introduction

A fundamental topic in modern cosmology is the understanding of the growth history of the universe. In particular, the Λ Cold Dark Matter (ΛCDM) model has proven to be very successful with current cosmological and astrophysical observations [23]. The values of the cosmological parameters in this model are constrained with modern observations to increasing level of precision, such as that of the Hubble constant. However, inconsistency of the Hubble constant values estimated with different methods appears, which is famously known as the Hubble tension.

On one hand, the Hubble constant can be constrained from early-universe measurement on the cosmic microwave background (CMB) power spectrum. The Planck survey on CMB yields a Hubble constant of $H_0 = 67.4 \pm 0.5$ km/s/Mpc [44]. An independent late-universe measurement relies on "standard candles" [21, 22, 47] with known luminosity such as Cepheids in Type Ia supernovae hosts. For example, Milky Way Cepheids in the early data release 3 (EDR3) of *Gaia* measure a Hubble constant of 73.2 ± 1.3 km/s/Mpc [48]. Evidently, the two methods yield estimates that are 4.2σ apart, posing a strong challenge if we require consistency within the same model [35, 55].

As gravitational-wave (GW) astronomy advances with successful event detections [1, 2, 8, 52], the possibility to use GW events as standard sirens for studying the Hubble constant draws attention [7, 16, 50]. Specifically, the GW source luminosity distance can be inferred from the waveform; if its redshift information becomes available, the Hubble constant and other cosmological parameters can be constrained.

In the simplest case, the GW event is accompanied by an electromagnetic counterpart, and its redshift is determined with an EM telescope. In this case, the GW standard siren is also called a "bright" siren. The sole example to date is GW170817 [8], a binary neutron star merger. Its gamma-ray burst signal was registered independently by the Fermi Gamma-ray Burst Monitor, and the Anti-Coincidence Shield at the International Gamma-Ray Astrophysics Laboratory (INTEGRAL) spectrometer[3].

This event alone constrained the Hubble constant to 70^{+12}_{-8} km/s/Mpc [4]. The bright sirens do not represent the full potential of standard sirens, as they make up only a small fraction of the total GW events. To start with, GW source population synthesis studies show that binary neutron star (BNS) events are fewer than binary black hole (BBH) events by a factor of $O(10)$ [5, 12]. Among the detectable BNS events, only 0.1% is estimated to have observable EM counterpart [14].

While individual GW events without EM counterpart cannot constrain cosmological parameters due to missing redshift, the total detected population could be informative in statistical inference of cosmological parameters [50]. This method is also called the GW dark siren. In short, the detected catalog can be compared to an "expected" distribution as a function of the cosmology and the astrophysics [57]. Depending on the main focus parameter space, we can further define a sub-category called the "spectral siren", where the focus is primarily on comparing source mass distribution [17, 24–26]. In this work, we focus on the redshift distribution, which is derived from galaxy distribution with the assumption that GW events are biased tracers of galaxies [27, 30, 31].

Such method has been applied to GW catalogs; in the latest Third LIGO-Virgo-KAGRA Gravitational-Wave Transient Catalog (GWTC-3), 47 events are analyzed against the galaxy catalog GLADE+ [20], and together with GW170817 yield $H_0 = 68^{+8}_{-6}$ km/s/Mpc [53]. While the current detector sensitivity limits the application of this method (see, e.g., Fig 7 in Ref. [30]), it is predicted that current detector upgrades and third-generation detectors can produce GW catalog that constrains the Hubble constant towards the percent level [11, 16, 51, 59].

In the dark siren method, the expected distribution is typically a mixture model between a galaxy catalog and theoretical population models [27, 30, 31, 57]. This treatment is necessary, since galaxy catalogs can be incomplete due to telescope magnitude limits. Since the relation between GW events and host galaxy characteristics is still an active area of research [28], the detected events can well come from galaxies that are missing from catalog. In this case, the redshift information is known only to the degree of the smooth overall galaxy mass function. It is evident that, with an incorrect theoretical model, the inferred cosmology is susceptible to bias. In Ref. [7], the effect of varying the event rate redshift evolution is explored; however, for this catalog, the Hubble constant inference is dominated by the bright siren GW170817, which is not affected by astrophysical model assumption (see Figure 7 in Ref. [7]). In Ref. [57], bias incurred by elemental astrophysical model

substructures (modeled as gaussian peaks) is illustrated.

In this work, we seek to quantify the susceptible bias using a Fisher information framework. Compared with Monte Carlo simulations (e.g., [30, 31]), this method is more cost-efficient and produces quick estimates of the bias. Compared with Ref. [57], this paper considers realistic mixture models with simulated galaxy catalogs, thus emphasizing the interplay between galaxy survey precision and GW catalog selection.

This chapter is organized as follows; in Section 6.2, we show the Fisher information framework. In Section 6.3, we highlight how galaxy catalog and GW measurement uncertainties are factored into our simulation and introduce model variations. In Section 6.4 and 6.5, we examine the simulation cases and explain results on galaxy model error tolerance. In Section 6.6, we summarize our findings and outline directions for further studies.

6.2 Fisher for Dark Siren

In this section, we specify the Fisher formalism for calculating statistical and systematic error on the Hubble constant. As discussed in Section 6.1, the statistical dark siren method amounts to comparing the detected GW population to an expected rates model, $r(\hat{D}, \hat{\Theta})$, where $\hat{\Theta}$ is a vector of parameters including, e.g., the sky location, $\hat{\Omega}$, and redshifted mass parameter of the binary, $(\hat{m}_1, \hat{m}_2)$. Throughout this work, we use $\hat{\cdot}$ for detector-frame GW quantities, $\tilde{\cdot}$ for galaxy catalog quantities, and no additional superscripts for the underlying true distribution. We also focus on estimating the Hubble constant alone and assume a standard ΛCDM cosmology.

We discuss the mass contribution by the end of this section, and we focus on the luminosity distance and angular sky location, $\hat{D}, \hat{\Omega}$, for now. The expected rate of GW events is given by

$$r_{\rm det}(\hat{D}, \hat{\Omega}) = f_{\rm gw}(\hat{D}, \hat{\Omega}) r(\hat{D}, \hat{\Omega}) \ , \tag{6.1}$$

where $f_{\rm gw}$ is the completeness fraction of GW detection, i.e., for sources at $\hat{D}$, $\hat{\Omega}$, what fraction of sources exist in the catalog.

The full rate, $r(\hat{D}, \hat{\Omega})$, is a convolution of the underlying GW event population, $r(D, \Omega)$, with the detector sensitivity kernel. We model the luminosity distance direction and angular direction sensitivity kernel as a gaussian distribution with $(\hat{D}, \hat{\sigma}_D)$ and a three dimensional von Mises-Fisher distribution with parameter $\hat{\kappa}$. They are assumed to be independent, and their joint distribution is their product.

The underlying full rate can be written as a mixture model between an empirical part from galaxy catalog and a supplementary theoretical component. The proportion depends on the completeness fraction of the catalog, $f_{\rm g}$. For magnitude-limited surveys, the completeness fraction is computed from distribution of galaxy absolute magnitude and the luminosity distance.

We assume that the galaxy catalog is in the form of $(z_i, \sigma_i, \Omega_i)$ representing the measured redshift, redshift uncertainty and angular position of galaxy i. The uncertainty kernel shares a similar form as the GW kernel, except that the gaussian distribution is given in redshift, and the parameter κ_i is much larger than $\hat{\kappa}$. The rate of GW events in galaxy i is given by r_i^g. The true galaxy mass function is written as $r_{\rm true}(z)$. Performing analytical integrals, we obtain

$$\begin{aligned} r(\hat{D},\hat{\Omega}) &= \int r(D,\Omega) K(D,\hat{D},\Omega,\hat{\Omega}) dD d\Omega \\ &= \int [r_{\rm cat}(D,\Omega) + (1-f_g(D)) r_{\rm true}(D,\Omega)] K(D,\hat{D},\Omega,\hat{\Omega}) dD d\Omega \\ &\equiv r_{\rm cat}(\hat{D},\hat{\Omega}) + r_{\rm theo}(\hat{D}) \ , \end{aligned} \tag{6.2}$$

$$r_{\rm cat}(\hat{D},\hat{\Omega}) = \sum_i \frac{r_i^g}{\sigma_i \hat{\sigma}_D} \frac{\hat{\kappa}}{(2\pi)^2} e^{\hat{\kappa}(\Omega_i\cdot\hat{\Omega}-1)} \int \exp\left[-\frac{(z-z_i)^2}{2\sigma_i^2}\right] \exp\left[-\frac{(D(z)-\hat{D})^2}{2\hat{\sigma}_D^2}\right] dz \ , \tag{6.3}$$

$$r_{\rm theo}(\hat{D}) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_D} \int [1-f_g(D(z))] r_{\rm true}(z) \exp\left[-\frac{(D(z)-\hat{D})^2}{2\hat{\sigma}_D^2}\right] dz \ , \tag{6.4}$$

where we separate into the catalog-piece and theoretical piece of the rate in the last line of Eqn. (6.2).

Treating the detected GW events as Poisson samples from this rates model, the Fisher information can be written as [57]

$$\mathcal{I}_{H_0} = \int r_{\rm det}(\hat{D},\hat{\Omega}) \left(\frac{d\log r_{\rm det}}{dH_0}\right)^2 d\hat{D} d\hat{\Omega} \ , \tag{6.5}$$

and the statistical uncertainty is given by

$$\delta H_{0,\rm s} = 1/\sqrt{\mathcal{I}_{H_0}} \ . \tag{6.6}$$

Consider any rate model $q(D)$ derived from a fixed redshift rate $q(z)$ with dz/dD. The Hubble constant derivative is given as

$$\frac{d\log q(D)}{dH_0}|_D = -\frac{1}{H_0}\left(\frac{d\log q(D)}{d\log D}+1\right), \tag{6.7}$$

which can be shown by applying the chain rule. This result can be directly applied to $r_{\rm cat}(D,\Omega)$ and the integrand of $r_{\rm theo}$. Since the completeness fractions are functions of luminosity distance in the detector frame, they do not change under variation of H_0. Accordingly,

$$\frac{d\log r_{\rm det}(\hat{D},\hat{\Omega})}{dH_0} = \frac{d\log r(\hat{D},\hat{\Omega})}{dH_0}\,. \tag{6.8}$$

The Hubble constant derivative of the detected GW rate is then given by

$$\frac{d\log r_{\rm det}(\hat{D},\hat{\Omega})}{dH_0} = \frac{1}{r(\hat{D},\hat{\Omega})}(I_1+I_2) \tag{6.9}$$

$$\begin{aligned} I_1 &= -\frac{1}{H_0}r_{\rm cat}(\hat{D},\hat{\Omega}) - \frac{1}{H_0}J_1 \\ I_2 &= -\frac{1}{H_0}r_{\rm theo}(\hat{D}) - \frac{1}{H_0}J_2 \end{aligned} \tag{6.10}$$

$$J_1 = \sum_i \frac{r_i^g}{\sigma_i\hat{\sigma}_D}\frac{\hat{\kappa}}{(2\pi)^2}e^{\hat{\kappa}(\Omega_i\cdot\hat{\Omega}-1)}\int \exp\left[-\frac{(z-z_i)^2}{2\sigma_i^2}\right]\frac{z_i-z}{\sigma_i^2}\exp\left[-\frac{(D(z)-\hat{D})^2}{2\hat{\sigma}_D^2}\right]D(z)\frac{dz}{dD}dz \tag{6.11}$$

$$J_2 = \frac{1}{\sqrt{2\pi}\hat{\sigma}_D}\int [1-f_g(D(z))]\frac{dr_{\rm true}(z)}{dz}\exp\left[-\frac{(D(z)-\hat{D})^2}{2\hat{\sigma}_D^2}\right]D(z)\frac{dz}{dD}dz\,. \tag{6.12}$$

The derivative of $r_{\rm true}$ can be analytical if the redshift evolution is given by some parametric function of redshift; in this work we use finite differencing to approximate its value.

This Fisher formalism can also be adapted to estimate bias from the true Hubble constant value, if the assumed models deviate from the true model, $\tilde{r}$. For instance, both the theoretical galaxy mass function, $r_{\rm true}(z)$, and the GW event bias, r_i^g, are active areas of research. Semi-analytic models and hydrodynamic simulations on galaxy redshift evolution attempt to include various physical processes such as gas cooling, star formation, feedback, etc., but the model details vary [54]. GW event bias[1] can also be redshift dependent to reflect increase in merger efficiency in low

[1] In this work, the most common usage of "bias" refers to systematic errors in contrast to statistical ones. In a few places, such as the present location, it is inherited from cosmology studies and denotes that GW sources are tracers of galaxies.

metallicity galaxy environment [28]. In the most general form, the deviation can be written as

$$\tilde{r}(\hat{D},\hat{\Omega}) = r(\hat{D},\hat{\Omega}) + \delta r(\hat{D},\hat{\Omega}) \ . \tag{6.13}$$

The resulting bias on the Hubble constant can be shown to be [57]

$$\delta H_{0,\mathrm{b}} = \frac{1}{\mathcal{I}_{H_0}} \int \frac{d\log r(\hat{D},\hat{\Omega})}{dH_0} \delta r d\hat{D} d\hat{\Omega} \ . \tag{6.14}$$

While the statistical uncertainty scales as $1/\sqrt{T}$, where T is the total observation time, the error term δr is also proportional to T and the bias stays constant. Especially in the context of third-generation detectors where we expect large numbers of high precision GW events, the role of the bias will be increasingly important in determining the level of total error budget.

Since GW measurement probes redshifted mass rather than the intrinsic mass, the event rate dependence on redshifted mass is also subject to changes in the Hubble constant. Specifically, the full rate model with explicit mass dependence is given by

$$\begin{aligned}
\frac{\partial N}{\partial \hat{M}\partial \hat{D}} &= \int \frac{\partial N}{\partial M \partial D} P(M|\hat{M}) P(D|\hat{D}) dD dM \\
&= \int \frac{\partial N}{\partial m \partial D} \frac{1}{1+z(D)} P(M|\hat{M}) P(D|\hat{D}) dD (1+z(D)) dm \\
&= \int \frac{dN}{dD} p(m) P(M|\hat{M}) P(D|\hat{D}) dD dm \\
&= \int \left(\int \frac{dN}{dD}(m(1+z(D)), D) P(M = m(1+z(D))|\hat{M}) P(D|\hat{D}) dD \right) p(m) dm \ .
\end{aligned} \tag{6.15}$$

The Hubble constant derivative of this piece with mass information is given as

$$\begin{aligned}
\frac{d}{dH_0}\left(\frac{\partial N}{\partial D \partial M}\right) &= \frac{d}{dH_0}\left(\frac{\partial N}{\partial D \partial m}\frac{1}{1+z(D)}\right) \\
&= \frac{d}{dH_0}\left(\frac{dN}{dD} p(m) \frac{1}{1+z(D)}\right) \\
&= \left(\frac{d}{dH_0}\frac{dN}{dD}\right) p(m) \frac{1}{1+z(D)} + \frac{dN}{dD} p(m) \frac{1}{(1+z(D))^2} \frac{D}{H_0} \frac{dz}{dD} \ ,
\end{aligned} \tag{6.16}$$

where the rate and luminosity distance direction derivatives are given in Eqn. (6.2) and Eqn. (6.9).

6.3 Simulation

In this section, we explain various components in our simulation and enumerate parameters of the model cases. Firstly, we describe the simulation volume. With the desirable GW angular uncertainty in mind, we simulate a cone with radius $\theta_{\rm max} = 4$ deg using the `Healpy` package in Python [29, 60]. We set the resolution parameter `NSIDE` $= 2^9$, which gives 3784 equal-areal pixels within the simulated region. The redshift ranges from 0.01 to 2. The lower bound ($\sim$ 43 Mpc) is sufficiently small to include the bulk expected GW population. The upper bound $z = 2$ is picked in light of realistic galaxy catalog depth; even though third-generation GW detectors may detect significant events at even higher redshift, the host galaxy information can be so incomplete that the dark siren becomes subject to above-tolerance bias. We assume that all-sky data is available for both galaxy catalog and GW catalog, and we further assume that all sky patches are statistically similar, including galaxy and GW event rates, distribution and measurement quality. Since the Fisher information scales linearly with angular coverage, we add a multiplicative factor equal to the fraction of simulated area compared to the whole sky. We note that the potential bias, $\delta H_{0,b}$, is independent from sky coverage. This suggests that the all-sky case is the most optimistic in terms of reducing statistical error and consequently the total error budget.

We create two broad groups of models that treat GW measurement differently. In the first group, which we refer to as the fiducial models, GW uncertainties are modeled using simple scaling functions with redshift. In the second group, GW uncertainties are computed from Fisher matrix with realistic third-generation detector network sensitivity. We refer to the latter as realistic models. In the following sections, we go into details about the setup for each.

Galaxy Simulation

We now discuss simulation parameters related to the galaxy catalog. There are three relevant aspects of galaxy catalog, the measurement quality, depth and overall rates, and we discuss each in turn.

Galaxy Redshift Measurement Precision

In our simulation, we consider two levels of redshift measurement precision with the form $\sigma_i(z) = \sigma_{i,0}(1+z)$ (recall that the index i refer to galaxy index in the catalog). In most fiducial cases and all cases with Fisher uncertainty, we set $\sigma_{i,0} = 0.003$,

corresponding to typical value for spectroscopic surveys. For photometric survey redshift measurement, the target redshift error is $\sigma_{i,0} = 0.01$. We explore the effect of larger redshift uncertainty with the model labeled *photo*.

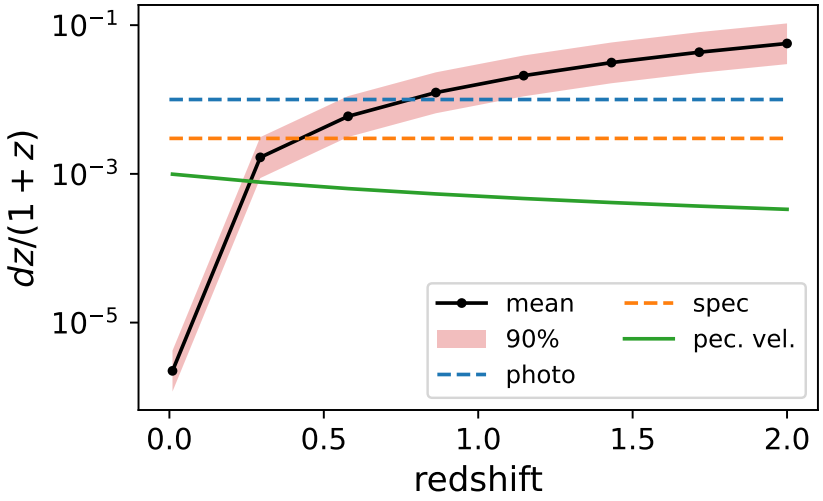

Figure 6.1: Redshift error budget comparison between galaxy measurement, GW measurement and peculiar velocity. The luminosity distance error is computed from a binary with intrinsic mass $(m_1, m_2) = (8.28\ M_\odot, 7.42\ M_\odot)$, and the light red band marks the 90% percentile across sky locations.

Aside from galaxy survey measurement error, the galaxy peculiar velocity also introduce error on the inferred redshift. In Ref. [33, 40–42], the authors studied the effect of considering galaxy-mass-dependent peculiar velocity in GW bright sirens, which are attainable at small redshift. In Ref. [59], the peculiar velocity is included as a constant in galaxy redshift uncertainty. In Figure 6.1, we compare the error from galaxy measurement, GW measurement and peculiar velocity. We show both spectroscopic and photometric redshift uncertainty value. For peculiar velocity, we adopt a typical value of 500 km/s [59]. We use the reference GW source with intrinsic mass $(m_1, m_2) = (8.28\ M_\odot, 7.42\ M_\odot)$ and compute the volume-weighted luminosity distance uncertainty (see below for details). The light red block represents the 90 percentile over sky locations. We make a few observations. Firstly, at all redshifts, the peculiar velocity gives smaller error than the nominal spectroscopic and photometric uncertainties, especially with increasing redshift. This suggests that, given our galaxy survey assumptions, it is reasonable to ignore peculiar velocity and focus on the main source of error. This assumption, however, is expected to break down for very small redshift galaxies, since their magnitude allows for better quality measurement than the typical values quoted here. If a more

realistic or a real galaxy catalog is used, it is necessary to re-evaluate and incorporate peculiar velocity. We also observe that, the redshift budget from GW contribution is in general smaller than galaxy contribution up until $z \sim 1$. For the assumed galaxy magnitude limit, the completeness fraction is typically already quite small at these redshifts (e.g. Figure 6.5 and Figure 6.4). This comparison suggests that in the radial direction, the dominating source of uncertainty comes from galaxy catalogs rather than from GWs with third-generation sensitivity. In the remainder of this work, we will neglect the contribution from peculiar velocity.

Galaxy Survey Completeness Fraction

The depth of a galaxy survey is related to many factors, such as the target galaxy magnitude or presence of special emission lines. For example, one of the targets of the Dark Energy Spectroscopic Instrument (DESI) [38] is Luminous Red Galaxies (LRG). Such galaxies exhibit prominent 4000Å spectral break, and they will be the primary observation targets for DESI in redshift range $0.4 < z < 1.0$ [58]. In the redshift range of $0.6 < z < 1.6$, another group of important targets are the Emission Line Galaxies (ELG), and the spectral signature is the [O II] doublet [45]. Since GW event distribution among different galaxy types is still an ongoing area of research [15], we make the simplifying assumption that the GW bias, i.e., its relative density per galaxy, is uniform. Under this premise, although galaxy surveys that target specific emission lines can reach large redshift, the catalog sample excludes galaxies without these signatures, and is consequently quite incomplete for our purpose.

Instead, we consider a magnitude-limited galaxy catalog such as the Bright Galaxy Survey (BGS) [32] of DESI. The DESI BGS is especially promising due to its large sky coverage and its spectroscopic redshift precision. We note that DESI BGS is expected to have a ~14,000 deg^2 footprint [32], which is not fully consistent with our all-sky catalog assumption. We expect that other surveys focusing on different sky patches can fill each other's gaps and provide a joint catalog (e.g. see GLADE [20]). For example, the Wide Survey of Euclid [37] is expected to cover ~15,000 deg^2 that complements the footprint of DESI. This survey is also expected to offer photometric redshift error with $\sigma_i/(1+z) < 0.05$ and spectroscopic measurement with $\sigma_i/(1+z) < 0.01$ [37]. Consequently, we neglect sky coverage details and focus only on the instrumentation magnitude limit.

The magnitude-limited completeness fraction depends on the absolute magnitude

distribution, which can be given as empirical fits from local surveys in the Schechter form, [6, 30, 31]

$$p(M) \propto 10^{-0.4(\alpha_M+1)(M-M^*)} \exp\left[-10^{-0.4(M-M^*)}\right] . \tag{6.17}$$

For the shape parameters, we adopt the empirical fits in Ref. [13] derived from local galaxy catalog from SDSS near $z = 0.1$. For r-band magnitude, the best-fit parameters are given by $M^* = -20.44 + 5\log_{10} h, \alpha_M = -1.05$, where h is the Hubble parameter. The absolute magnitude limits are given by $(-24.26, -16.11)$. The absolute magnitude is converted to the apparent magnitude m with $m = M + 5\log_{10}(D/10\ \mathrm{pc})$. Integrating up to the magnitude limit of the galaxy survey, we obtain the corresponding completeness fraction.

In realistic surveys, the observation completeness can be further dependent on more instrumentation and scheduling details. During DESI BGS, the Bright objects (r-band magnitude $r < 19.5$) are top-priority objects and are observed with high completeness; after four telescope passes, $> 80\%$ of BGS Bright targets are assigned a fiber for spectroscopic measurement. The majority of the Faint objects (r-band magnitude $r < 20.175$) has a lower priority, and they only achieve a fiber assignment efficiency of 60% [32] with four passes. The fiber assignment efficiency counts towards the galaxy completeness fraction in addition to the magnitude limit. However, this scaling is dependent on specific observation priorities and scheduling strategy, and we neglect it in this present study for the sake of generality.

In our models, we mostly adopt an apparent magnitude limit of $m_g = 20.175$ as an optimistic baseline, corresponding to the Faint targets in DESI BGS [32]. In the fiducial model named `bright`, we set $m_g = 19.5$ to simulate a shallower catalog (mimicking the DESI Brights target). We also set $m_g = 22$ for the fiducial model `deepz` to assess the performance gain with a deeper survey.

Galaxy Number Density

We compute the galaxy number density using galaxy mass function fitted from observation. For the galaxy number distribution, $\Psi(\sigma_v, z)$, where σ_v is the surface velocity dispersion of the galaxy, we first adopt a modified Schechter function fitted from the Sloan Digital Sky Survey (SDSS) catalog [18],

$$\Psi(\sigma_v, 0) = \phi_* \left(\frac{\sigma_v}{\sigma_*}\right)^{\alpha} \exp\left[-\left(\frac{\sigma_v}{\sigma_*}\right)^{\beta}\right] \frac{\beta}{\sigma_v \Gamma\left(\alpha/\beta\right)} , \tag{6.18}$$

where $\phi_* = 8.0 \times 10^{-3} h^3$ Mpc^{-3}, $\sigma_* = 161$ km/s, $\alpha = 2.32$ and $\beta = 2.67$. To account for the redshift dependence, we apply a multiplicative scaling factor,

$$\Psi(\sigma_v, z) = \Psi(\sigma_v, 0) \frac{\Psi_{\rm hyd}(\sigma_v, z)}{\Psi_{\rm hyd}(\sigma_v, 0)} \,, \tag{6.19}$$

where $\Psi_{\rm hyd}(\sigma_v, z)$ is the fitted function from hydrodynamical simulations in Ref. [54]. The integrated surface velocity dispersion ranges from $\sigma_v = 100$ to 300 km/s to include the peak of this distribution. We note that this Schechter function form is fitted from a selection of early-type galaxy samples. Since the details of GWs as galaxy tracers are still topics of active research (see, e.g., Ref. [15]) and may not be limited to massive early-type galaxies, this rate might be an underestimate. Moreover, with improved galaxy survey sensitivity, the small σ_v end could improve from more measurement on faint galaxies. As a sanity check, the resulting galaxy number density at the smallest redshift in our simulation is 0.002 Mpc^{-3} (i.e. average comoving distance separation is roughly 7.9 Mpc.). To gauge the effect of changes in galaxy number density, we set a fiducial model with reduced galaxy density, labeled `sparse`.

Catalog generation

Combining the intrinsic galaxy mass function and the observational completeness, we obtain the expected rates of galaxies in catalog. We then generate Poisson samples and distribute them evenly in the angular direction over our simulated sky patch. We note that we do not further distinguish the generated galaxy samples, e.g., their stellar mass etc. It is possible to assign galaxies different hosting rates, r_i^g, which alters the distribution shape. Since this variation is a straightforward adaptation to the current model, we keep our model general and exclude highly-model-dependent modifications. We also note that the Fisher information from the catalog piece can vary based on the exact catalog realization. However, due to the large number of expected galaxies, the total Fisher information is very stable; for the `optimistic` model parameters (see Table 6.1), we generate 50 catalog realizations and find that the Fisher information standard deviation is 0.053% of the average. For computational efficiency concerns, in the final catalogs reported here, we use only a single galaxy realization for each set of galaxy magnitude limit and overall galaxy density scaling, $(m_g, f_{\rm g,red})$. We list the number of simulated galaxies for each case in Table. 6.3.

Gravitational-wave Simulation

Similar to galaxy simulation, inputs from GWs include measurement quality, completeness fraction and overall event rate. As mentioned, our test models are divided into two broad categories; for realistic models, the test GW event is drawn from an observationally constrained BBH mass function, and the measurement uncertainties are computed from Fisher information with a third-generation detector network. In the following sections, we go through each factor.

GW measurement quality

In the following sections, we discuss measurement error in luminosity distance and angular error. In all cases, the GW redshift kernel is modeled as a gaussian with $(\hat{D}, \hat{\sigma}_D)$. The angular kernel is modeled with the von Mises-Fisher distribution with $p = 3$,

$$p(\Omega, \hat{\Omega}) = C(\kappa) e^{\kappa \cos \Omega \cdot \hat{\Omega}}, \quad C(\kappa) = \frac{\kappa}{4\pi \sinh \kappa}, \tag{6.20}$$

where κ controls the spread of the distribution. The value of κ is set such that $p(1\sigma)^2$ of the probability is contained within a circular cap with area equal to $\sin\theta \Delta\theta \Delta\phi$, where $\Delta\theta, \Delta\phi$ are the polar angle and azimuthal angle uncertainty of the GW event.

In the fiducial models, we consider a simple scaling function for the fractional luminosity distance error, $d \log D = d \log D_0 (1 + \hat{z})$. In most cases, we adopt $d \log D_0 = 0.001$, and we discuss the effect from using smaller SNR events $d \log D_0 = 0.01$ with the catalog named `lum`. We also adopt a scaling function for the angular uncertainty parameter, $\kappa(\hat{z}) = \kappa_0/(1 + \hat{z})$. In the base case, we set $\kappa_0 = 10^4$, which corresponds to a typical scale of $d\theta_0 = 0.64$ deg. From our Fisher information error data hindsight, this is a reasonable value given the detector network we consider. To gauge the effect of worse localization error, in the fiducial model named `ang`, we set $\kappa_0 = 10^3$, giving $d\theta_0 = 2.03$ deg. The choice of these target localization quality is motivated by Ref. [39] (Figure 8) and Ref. [36] (Figure 3), where localization error can be below 1 deg in radius or less with next generation GW detector added to the network. For `ang` only, we increase the simulated region to a radius of 10 deg, to ensure we do not suffer from edge effects.

For realistic models, we compute measurement error using Fisher information formalism. Firstly, we draw GW source intrinsic parameter samples. We consider the `POWER LAW + PEAK` model with a smoothing function at the low mass end.

Distribution of the primary mass, m_1, is given as [9]

$$\pi(m_1) = [(1-\lambda_g)\mathcal{B}(m_1) + \lambda_g G(m_1)]S(m_1) \ , \tag{6.21}$$

where $\mathcal{B}$ is the normalized power-law distribution with a slope of $-\alpha$ and a high-mass cutoff at $m_{\rm max}$, and $G(m_1)$ is a normalized gaussian distribution with (μ_g, σ_g). The sigmoid smoothing function, $S(m_1)$, tapers the distribution at the low mass end with parameter δ_m. The distribution of the secondary mass, m_2, follows a power law with a slope of $-\beta$. In this work, we adopt the parameters $\alpha = 3.78, \beta = 0.81, \mu_g = 32.27\ M_\odot, \sigma_m = 3.88\ M_\odot, \lambda_g = 0.03, \delta_m = 2.5\ M_\odot$, following Ref. [57], which are mostly identical to the inferred parameters in Ref. [53]. We assume the primary black hole mass is between $m_{\rm min} = 6.5\ M_\odot, m_{\rm max} = 112.5\ M_\odot$. Using this distribution, we sequentially generate (m_1, m_2) pairs.

To compute GW signal-to-noise ratio (SNR) and parameter uncertainties, we assume a detector network of CE+LL+ET, a Cosmic Explorer [46] at current LIGO Hanford location, aLIGO Livingston and Einstein Telescope [49]. The noise power spectral density (PSD) is imported from the `psd` module from `PyCBC` [43][2]. The detector antenna patterns, f_p, f_c, are obtained through `PyCBC.Detector.antenna_pattern`[3]. The plus, cross waveforms, h_p, h_c, are obtained via `PyCBC.get_fd_waveform` using approximant `IMRPhenomHM` [34].

We compute uncertainties from Fisher matrix for each mass pair on a grid of parameter space, $(\hat{z}, \hat{\Omega}, \iota)$. The redshift grid is evenly spaced from $\hat{z} = 0.01$ to 2. The ι grid (source orbital inclination) is evenly spaced from $\iota = 0.01$ to $\pi/2 - 0.01$. The angular position $\hat{\Omega}$ is the `healpy`[60] grid pixel with `NSIDE=4` (in total 192 equal-areal pixels).

In principle, these samples can be fed into Monte Carlo integration for Hubble error calculation; we note that in practice such a simulation is time-consuming, since both the catalog piece and the theoretical piece in the Fisher information must be recomputed from convolution between the galaxy catalog and the new uncertainties. In the current scope of this study, we use the data from only one mass pair sample $(m_1, m_2) = (8.28\ M_\odot, 7.42\ M_\odot)$. We have checked that this mass pair is close to the peak probability in the mass distribution, thus may be used as a proxy to the expected population majority.

[2] Specifically, the adopted PSD versions are `CosmicExplorerP1600143`, `aLIGODesignSensitivityP1200087` and `EinsteinTelescopeP1600143` respectively.

[3] Specifically, with detectors `H1`,`L1` and `E1`.

The corresponding parameter uncertainty is a weighted average over sky location and binary inclination angles, $(\hat{\Omega}, \hat{\iota})$. For sky coverage, we take a flat average. The threshold ι depends on selection criterion. In our realistic simulations, we examine using either the SNR or the angular error. We find that at each redshift, the relative change in the criterion compared to when $\iota = 0$ demonstrates a universal pattern, and we define this averaged scaling function as (here we use SNR, ρ, as the example)

$$g(\iota) = \text{interp1d}\left(\iota, \frac{1}{N_z}\sum_{i=0}^{N_z}\frac{\rho(\iota, z_i)}{\rho(0, z_i)}\right) . \tag{6.22}$$

Specifically, we used linear interpolation from `scipy.interp1d` [56]. At each redshift, the threshold ι can be computed from the inverse function, and we take the volume-weighted average as the nominal uncertainty, e.g.,

$$\begin{aligned} \iota_{\text{thresh}}(z) &= g^{-1}\left(\frac{\rho_{\text{thresh}}}{\rho(\iota = 0, z)}\right) \\ \delta\bar{\theta} &= \frac{1}{N_\Omega}\sum_{i=1}^{\text{NPIX}}\frac{\int_{\iota_{\text{thresh}}}^{\pi/2}\delta\theta\sin\iota d\iota}{\int_{\iota_{\text{thresh}}}^{\pi/2}\sin\iota d\iota} . \end{aligned} \tag{6.23}$$

To visualize the resulting rates model, we produce four model cases with $\rho_{\text{thresh}} = 100, 200$ (*rho_sm*, *rho_lg*) and $\theta_{\text{thresh}} = 0.4$ deg, 1 deg (*theta_sm*, *theta_lg*).

GW completeness fraction

As is seen in Eqn. (6.1), GW completeness fraction provides the overall scaling for how many events we can expect to have for the Hubble measurement. In the fiducial models, we adopt a phenomenological approach and use a sigmoid function,

$$f_{\text{gw}} = \frac{1}{1 + \exp((\hat{z} - \mu_{\text{cf}})/\sigma_{\text{cf}})} , \tag{6.24}$$

and we hold $\sigma_{\text{cf}} = 0.15$, which controls the decay rate of the completeness fraction.

In models with simulated measurement uncertainty, the GW completeness fraction can be directly obtained by enforcing selection criterion. Assuming that the orbital axis of the binary is uniformly distributed over the sphere, the completeness fraction is given by $f_{\text{gw}} = 1 - \cos\iota_{\text{thresh}}$ (see Eqn. (6.23)). This setup brings an important difference from the fiducial models. In the fiducial cases, the rates model is only affected by an overall scaling upon completeness fraction changes, since we assume source selection is independent from measurement quality. In the realistic case,

the completeness fraction is closely related to averaged measurement quality; as the selected GW sources decrease in number, the selection is more stringent, and individual event quality improves.

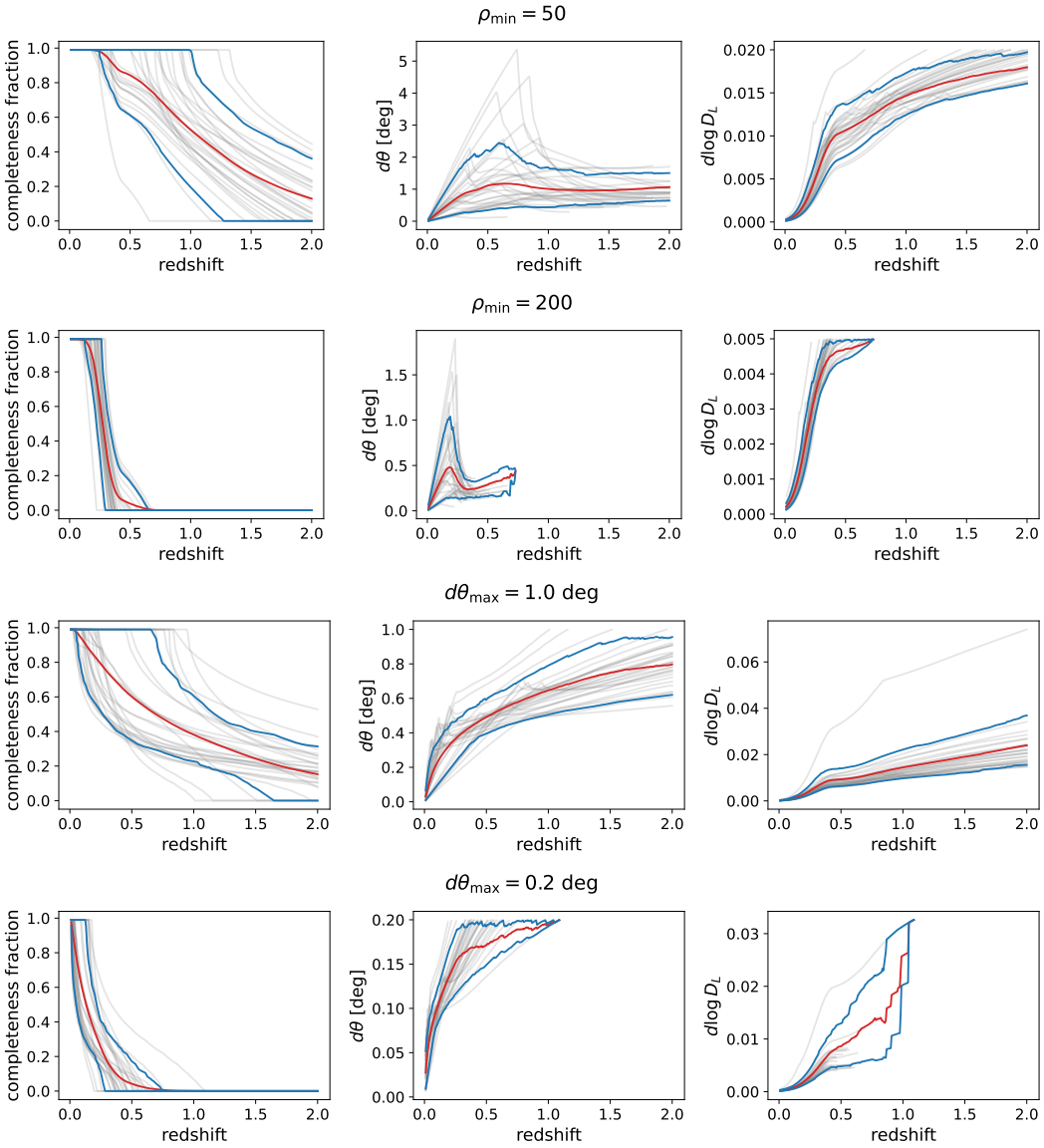

Figure 6.2: Example completeness fraction and GW measurement uncertainties. From *Top* to *Bottom* panels, the selection criteria are for an SNR of 50, 200, and a maximum angular uncertainty (radius) of 1, 0.2 deg. In each panel, the light gray traces show results for randomly selected 30 sky pixels. The red trace shows the pixel-average that is adopted in the final Fisher calculation. The blue bracket shows the 80 percentile. In the uncertainty panels, the traces end when no sky pixels contain above-threshold sources.

In Figure 6.2, we show the weighted completeness fraction, angular uncertainty and

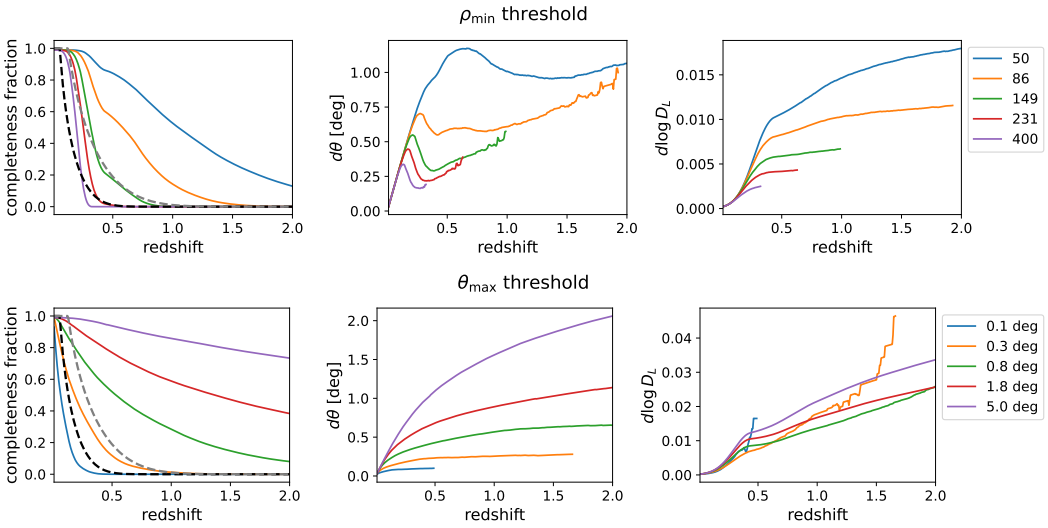

Figure 6.3: Volume-averaged completeness fraction and GW measurement uncertainties. The *Top* panel shows the SNR criterion and the *Bottom* panel shows the angular resolution cap. The galaxy completeness fraction with magnitude limit of 20.175 and 22 are shown with dashed black and gray traces, respectively.

luminosity distance uncertainty given several selection criteria. We show results for a random selection of 30 sky pixels (gray trace), and show the adopted mean value (red trace) and the 80 percentile (blue trace). In Figure 6.3, we show the adopted mean value for a range of thresholds that we scan over in the final simulation. In panels for uncertainties, the traces end when no sky pixels contain qualifying sources. We observe peculiar behavior in, e.g. the angular uncertainty when using SNR as the threshold, shown as a bump in intermediate redshift values. This is due to the difference in uncertainty growth rate for localization and SNR. By excluding the inclination values with sub-threshold SNR, the remaining *selected* GW sources with better measurement quality may have a smaller weighted angular uncertainty. We also show the galaxy completeness fraction with a magnitude limit of 19.5 and 20.175 in dashed black and gray lines for comparison.

GW event number density

The GW number density is constrained from observation; based on events up to GW170104, the local BBH merger rate is constraintted to be 103^{+110}_{-63} Gpc^{-3}yr^{-1} [2]. With GWTC2, the local merger rate for BBH is updated to $23.9^{+14.3}_{-8.6}$ Gpc^{-3}yr^{-1} [10]. We assume that this rate is a constant in redshift and across galaxies, due to the lack of relevant observational constraints. In our fiducial models, we found

that the GW uncertainties can be more conservative compared to actual Fisher matrix results, resulting in a more pessimistic assessment on susceptibility to bias. To better showcase different effects, we artificially raise the GW event rates to 239 $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ for most fiducial models. The realistic case is shown with model named `realistic_bbh_rate`. In the realistic models, we found that using realistic rates already produces simulation results with sufficient clarify.

Based on variations of the above model parameters, we present eight fiducial models and four models with realistic GW uncertainties, summarized in Table. 6.1 and 6.2. All models except `ang` share the same simulated volume. Realistic models have identical galaxy parameters and GW occupation rate as `realistic_bbh_rate`.

6.4 Simulation Result Analysis

In this section, we analyze our simulation results and discuss implication to the GW dark siren Hubble measurement. In Figure 6.6 and Figure 6.7, we show angular cross section of the catalog-piece rates model, $r_{\mathrm{cat}}(\hat{D})$, at different redshifts. In Figure 6.4 and Figure 6.5, we unwrap $r_{\mathrm{cat}}(\hat{D})$ at $\theta = 3.3$ deg to show one slice along the redshift direction. All panels in these four figures have independent color scale.

In Figure 6.4, we analyze the catalog-piece rate model by comparing to the base case, `optimistic`. The models `lum`, `ang` and `photo` show rate changes with poorer measurement sensitivity. In the redshift direction, while GW fractional errors in `lum` increase an order of magnitude and the redshift error in `photo` increases only by a factor of 3, the redshift-direction blurring effect is much more apparent in `photo`. In the angular direction, the GW angular uncertainty in `ang` increases only by 24%, yet the angular direction rates becomes significantly flatter. This contrast suggests that the redshift uncertainty is dominated by galaxy catalog errors, and the angular direction is dominated by GW localization.

Comparing `sparse` to `optimistic`, we observe that the rates model becomes more sharply peaked. In `deepz`, the catalog redshift reach is much deeper, although the higher redshift section rates model becomes less spiky due to larger redshift error and more galaxies blending together.

In Figure 6.8, we plot the differential Fisher information for various fiducial models prior to scaling with GW completeness fraction. For clarify, all traces except the leftmost one have been shifted to the right. As expected, the overall scale of the Fisher information is similar for models with the same galaxy and GW number density, since galaxy completeness and measurement quality only contribute to the

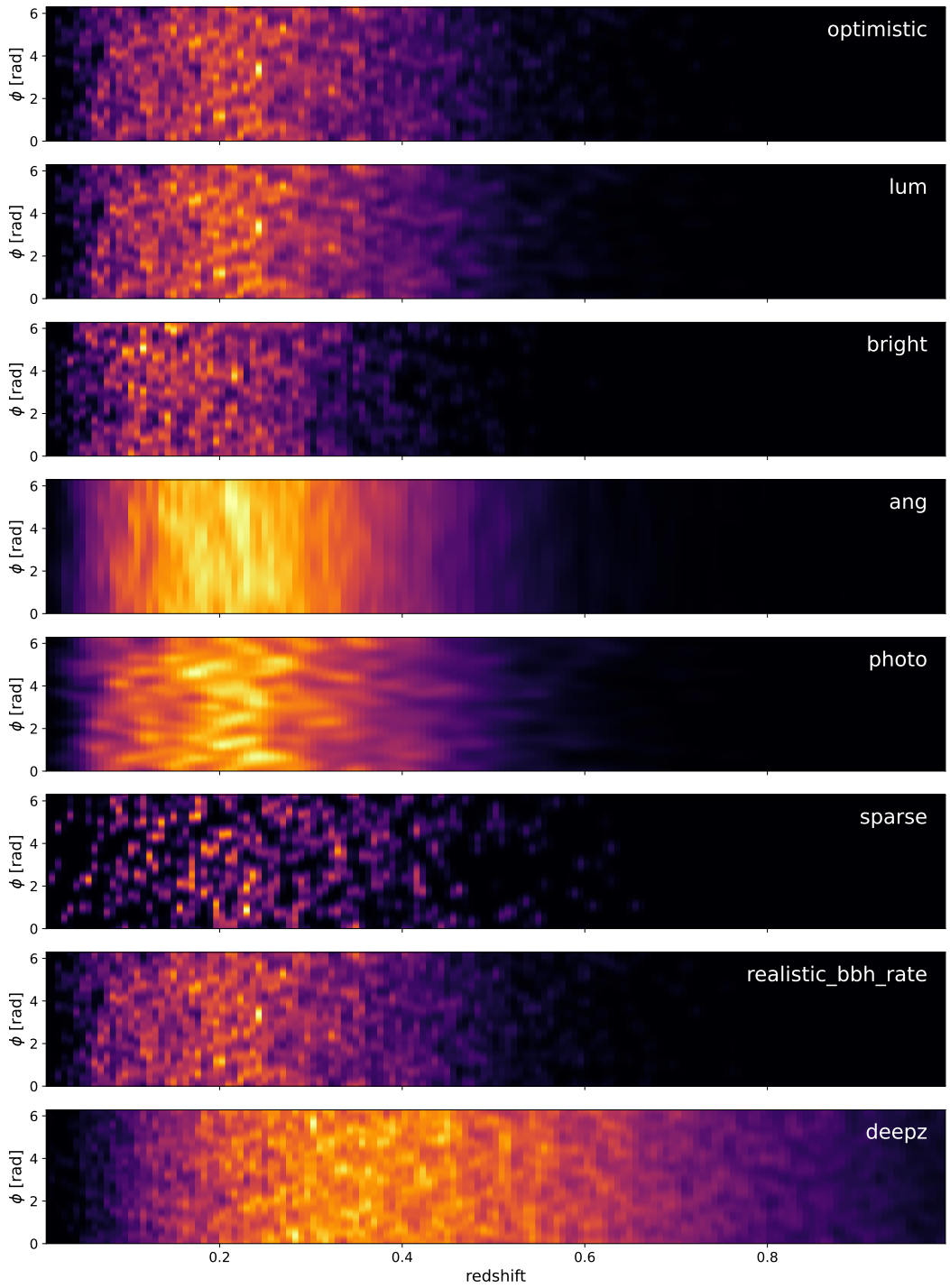

Figure 6.4: Slice of the expected GW event rate due to galaxy catalog in the simulated volume at $\theta = 3.3$ deg. Each panel represents a fiducial catalog with the catalog name shown on the upper right corner. See text and Table 6.1 for model parameters. Note that each panel has an independent color scale.

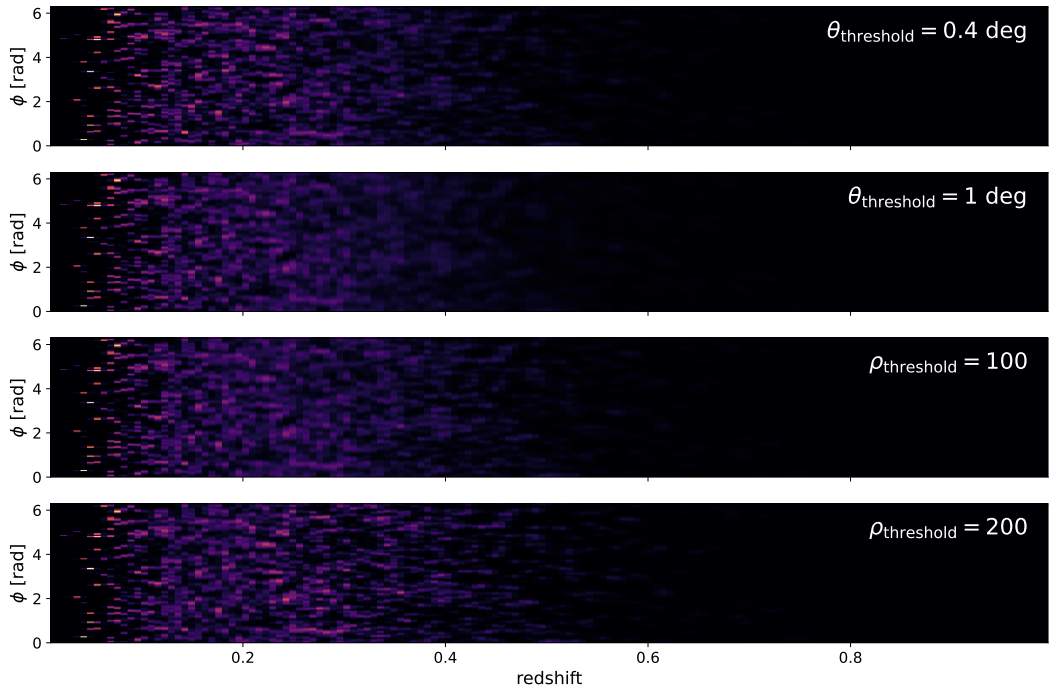

Figure 6.5: Slice of the expected GW event rate due to galaxy catalog in the simulated volume at $\theta = 3.3$ deg. Each panel represents a catalog with the catalog name shown on the upper right corner. See text and Table 6.1 for model parameters. Note that each panel has an independent color scale.

ratio between the catalog piece and the theoretical piece, as well as the smoothness. The smoothing degree and spiky region location are consistent with the event rate angular and redshift direction cross section (e.g. comparing *photo* and *deepz*).

In Figure 6.9, we show the differential Fisher information for realistic models using either SNR or localization as the selection criterion. Since in these models the GW completeness fraction is fixed given a selection criterion, the differential Fisher information shown has already been scaled with $f_{\rm gw}$. In both panels, we observe that more stringent selection criterion produces taller peaks at small redshift, but the redshift depth also drastically reduces. The final Fisher information depends on the competition between these two factors.

In both groups, we observe that the Fisher information peak is misaligned with the galaxy distribution, which peaks around $z = 2$. While at higher redshift the expected events are more numerous, their contribution to constraining the Hubble constant decreases due to their flatter profile. This phenomenon becomes more dramatic in realistic models, where the Fisher information shows an additional peak at very low

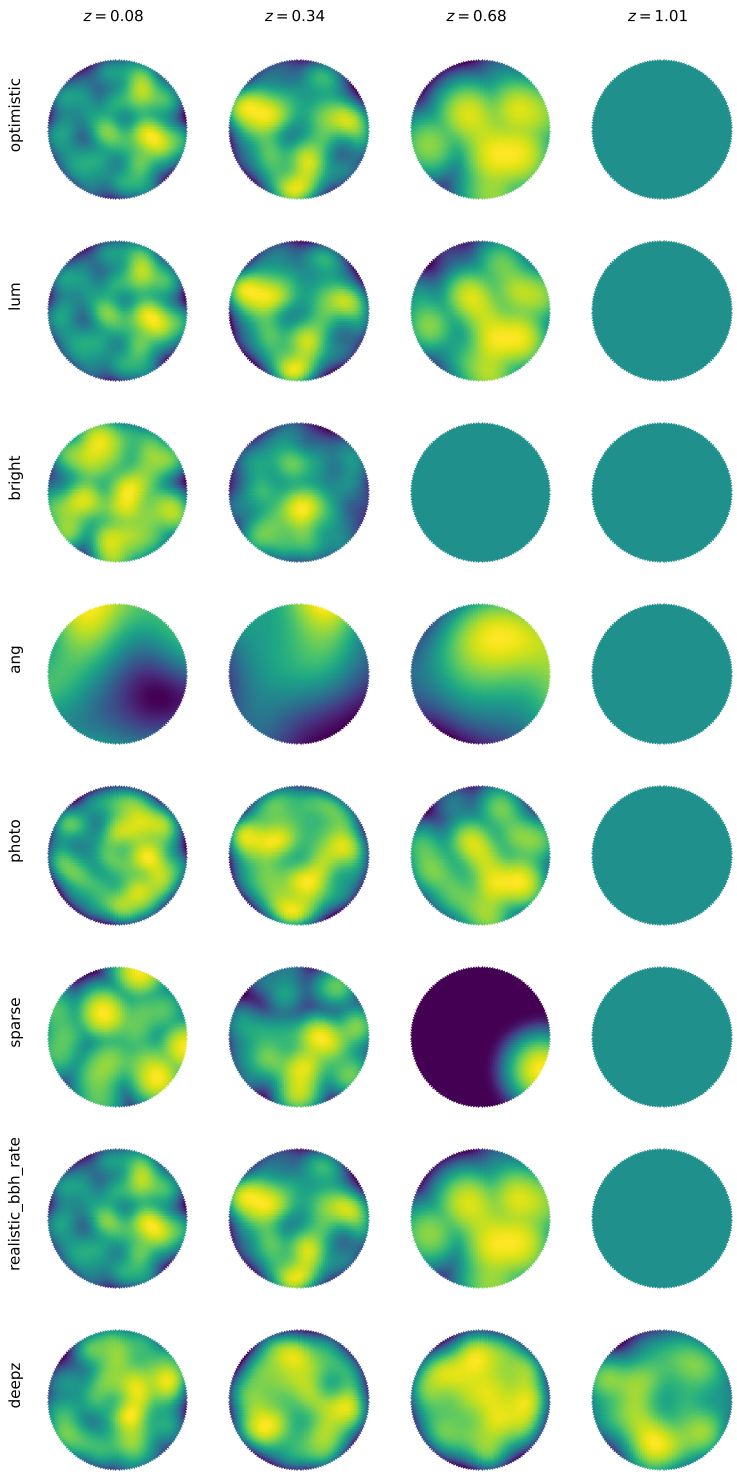

Figure 6.6: Angular cross section of the expected GW event rate due to galaxy catalogs. The galaxy catalog is convolved with fiducial values of GW measurement uncertainty. From *Top* to *Bottom*, catalog names are labeled on the left. From *Left* to *Right*, the redshift is $(0.08, 0.34, 0.68, 1.01)$. Note that each panel has an independent color scale. See text and Table 6.1 for simulation parameters.

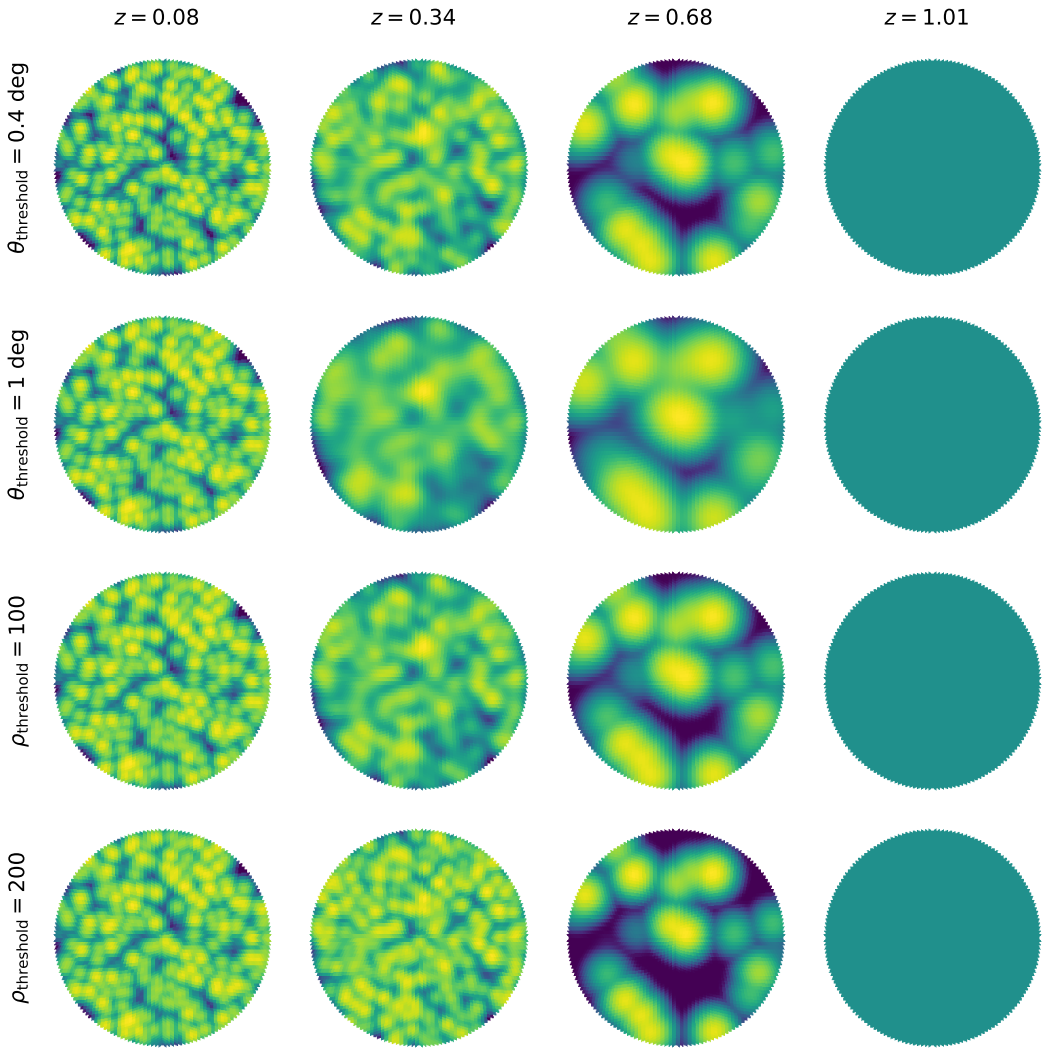

Figure 6.7: Angular cross section for simulated catalogs with realistic GW uncertainties at different redshifts. The GW event selection criterion is noted on the left. From *Left* to *Right*, the redshift is $(0.08, 0.34, 0.68, 1.01)$. See text and Table 6.2 for simulation parameters. Note that each panel has an independent color scale.

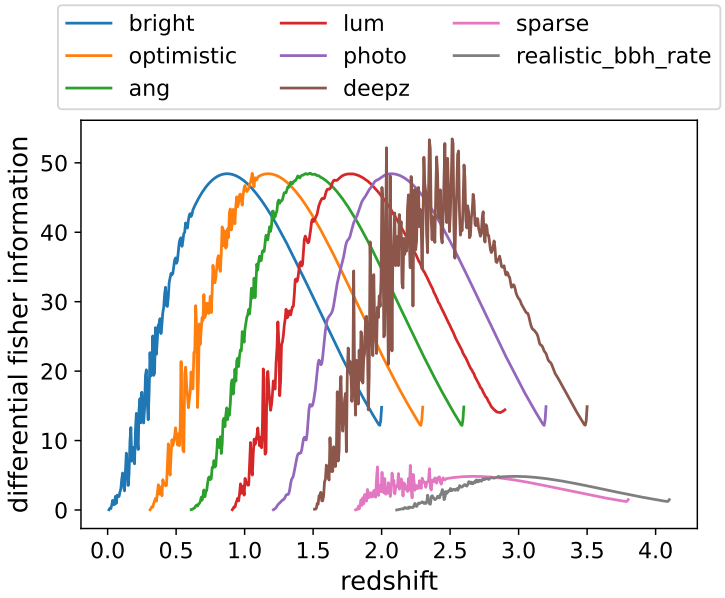

Figure 6.8: Full differential Fisher information for various fiducial models. The GW completeness fraction has not been applied. For presentation clarity, all but the leftmost trace in each panel have been shifted to larger redshift values.

redshift. In combination with Figure 6.5, this demonstrates that low-measurement-error events at small redshift contribute significantly to constraining the Hubble constant albeit the disadvantage in number.

6.5 Galaxy Model Error Tolerance

As is shown in Eqn. (6.14), bias from incorrect galaxy model assumption is sensitive to the form of rates difference. In Ref. [57], the considered galaxy model correction are local overdensities of galaxies, and they are modeled as individual gaussian bumps in the rates model. In Ref. [19], the authors consider galaxy clustering and generate test catalogs given real space correlation function, rather than having uniformly distributed Poisson samples; in this case the model error would be the deviation from these large scale structures. In this work, we consider the case where the assumed galaxy mass function has a different redshift dependence from the true distribution,

$$r = r_{\rm true}(1+\hat{z})^{\beta} \ . \tag{6.25}$$

In this case, the catalog piece $r_{\rm cat}$ remains unchanged while the supplemented theoretical rate is multiplied with $(1+\hat{z})^{\beta}$. The corresponding error rate term is then given by $\delta r = \left[1-(1+\hat{z})^{\beta}\right] r_{\rm true}$. The statistical uncertainty and the bias is computed according to Eqn. (6.6) and Eqn. (6.14).

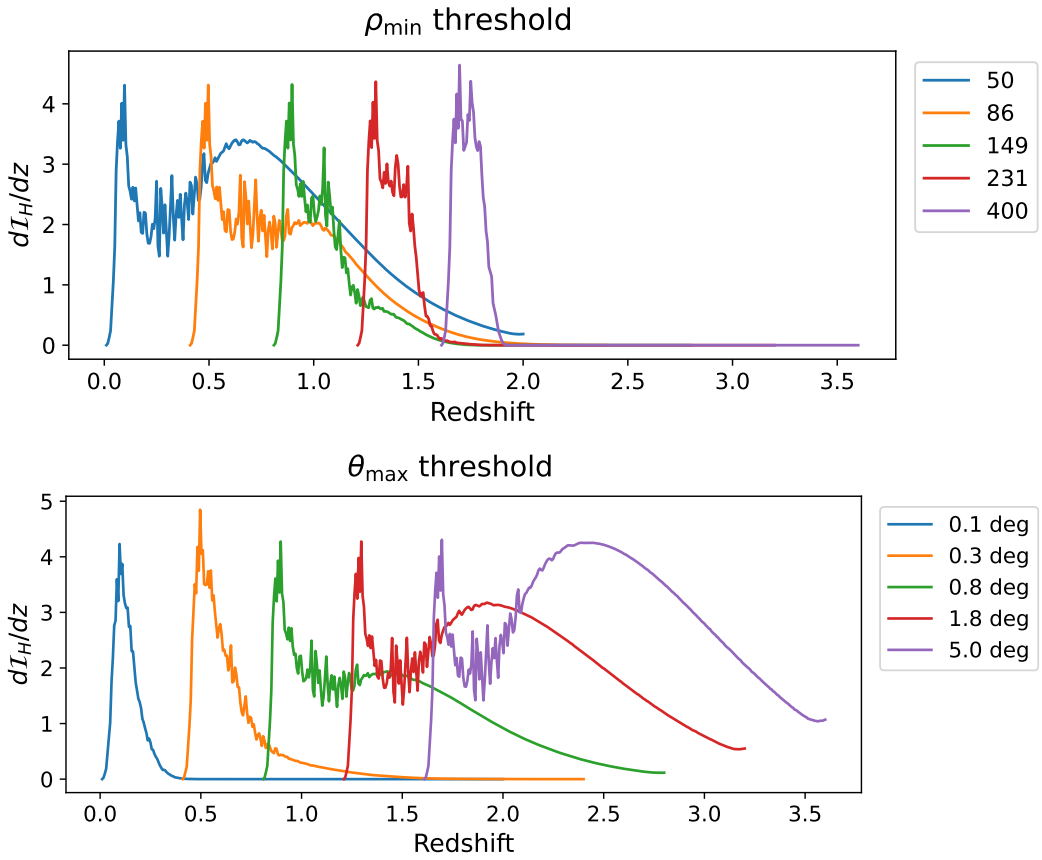

Figure 6.9: Differential fisher information for realistic models given different selection criteria. The GW completeness fraction is applied. In all except the leftmost traces, the redshift coordinate has been artificially shifted to the right for visual clarify.

Intuitively, mitigating potential bias and reducing statistical uncertainties of Hubble measurement place different demands on the GW events used for inference. On one hand, the potential bias does not benefit from more events; it requires that the incorrect rate piece plays less role in the Fisher information. Therefore, reducing bias requires placing emphasis on close-by events where galaxy catalog is highly complete. On the other hand, the statistical uncertainty depends only on the total Fisher information, which benefits greatly from including high-redshift events occupying greater volume. Therefore, considering both pieces is essential for balancing the dark siren precision and accuracy. In the following test, we investigate that, given a fixed total error budget, what the maximum tolerated redshift evolution parameter β is under various GW event selection.

In Figure 6.10, we study fiducial models and vary the GW selection completeness

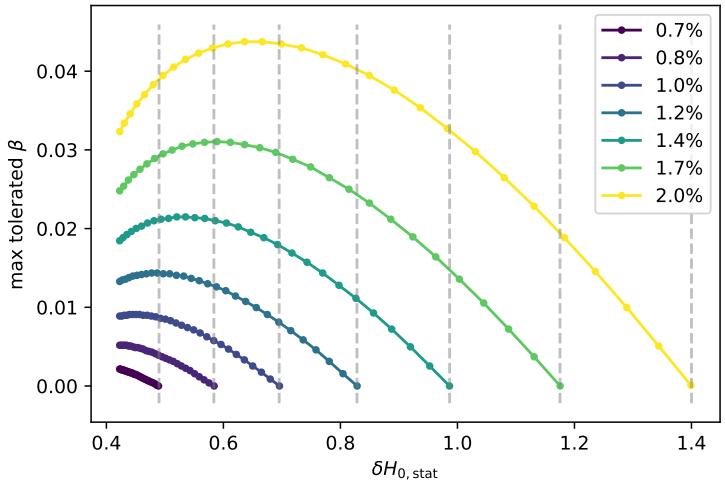

Figure 6.10: Maximum galaxy model error tolerance given various total error budget using the fiducial models `realistic_bbh_rate`. The target total error budget is geometrically distributed between 0.7% and 2%. The horizontal axis has a unit of km/s/Mpc.

fraction, i.e., $\mu_{\rm cf}$ in Eqn. (6.24), and search for the maximum galaxy mass function redshift bias β for a fixed total error budget. This figure shows the result for the model `realistic_bbh_rate`. The searched total error budget values are geometrically distributed from 0.7% to 2%, and we mark them with vertical dashed lines. The right end of the traces marks either when the statistical error saturates the total budget, or when $\mu_{\rm cf} = 0.01$ at the edge of our simulation volume. The left end marks either when the statistical error is 10% of the total budget, or when $\mu_{\rm cf} = 1.8$ close to the maximum simulated redshift. As the total error budget changes, we observe the development of a sweet spot; prior to the sweet spot, (right-hand-side of the figure), the improvement in statistical error is faster than the increasing contribution from $r_{\rm theo}$, which leaves greater room for theoretical model induced bias. However, as we keep adding deeper sources, the growth of Fisher information is suppressed due to the smoothing of the rates model, and is soon outpaced by the bias brought by the large theoretical population. This observation suggests that, given the same error budget, if we would like to prioritize robustness against variations in astrophysical models, it can be desirable to apply additional filtering and disregard distant events.

In Figure 6.11, we show the galaxy model error tolerance given different fiducial model parameter assumptions for a 1% total error in the *Top* panel. In the *Bottom* panel, we show the corresponding GW completeness fraction parameter $\mu_{\rm cf}$. The overlay of the traces demonstrates that given a similar galaxy catalog, small

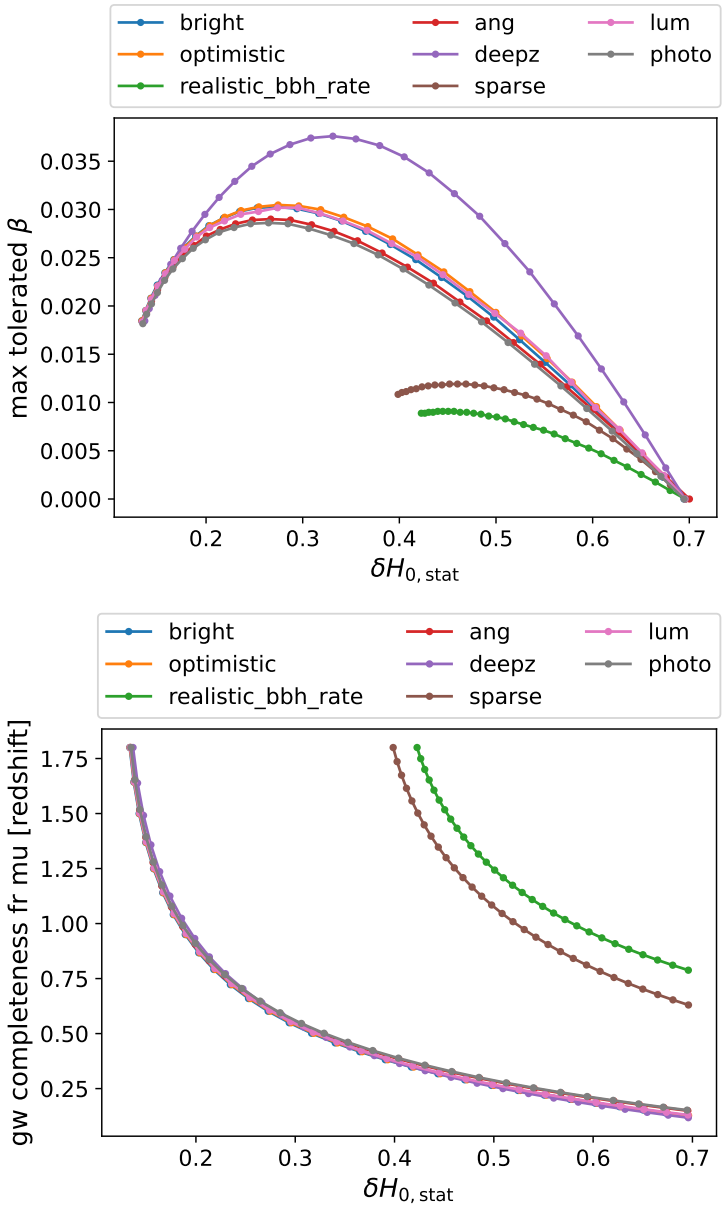

Figure 6.11: Maximum galaxy model error tolerance with 1% total error budget for fiducial simulation models. The *Left* panel shows the maximum β, and the *Right* panel shows the corresponding $\mu_{\rm cf}$ in GW completeness fraction. See text for ending point definition. The horizontal axis has a unit of km/s/Mpc.

variations of measurement uncertainty does not significantly vary the overall Fisher information, and the same statistical uncertainty is achieved with similar GW observational depth. By comparing the error tolerance, however, we observe that `sparse` shows significant improvement over `realistic_bbh_rate`, even though they both represent a density reduction of a factor of 10. This is due to the fact that the Fisher

information is highly sensitive to the *derivative* of the rates model; changing the GW density only applies an overall scaling, but increasing galaxy density leads to more blending between neighboring galaxy probability region and "flattens out" the derivative. In the completeness figure, we also see that `sparse` achieves the same level of $\delta H_{0,\mathrm{stat}}$ at smaller redshift, further supporting the derivative argument.

In Figure 6.12, we show tolerance to β using realistic GW uncertainties as we apply different source selection criteria. In the *Top* panel, the minimum SNR ρ_{min} is geometrically distributed from 50 to 400, and in the *Bottom* panel, the maximum angular localization radius θ_{max} ranges from 0.2 deg to 5 deg following a geometric distribution. The black dashed line corresponds to the right vertical axis and shows the number of expected GW sources under each threshold condition. For the SNR selection, none of our searched value gives a statistical uncertainty below 0.7%, and the corresponding trace is absent.

For stringent total error budget, e.g. 0.7% and 0.8%, the behavior is generally similar to that for fiducial models, where the reduction in statistical error brings advantage to model error tolerance. With larger error budget, however, we observe a qualitative difference; rather than developing selection sweet spots, the error tolerance increases exponentially with increasing statistical uncertainty. In this case, it becomes more advantageous to apply stringent GW source selection and pick out close-by and well-localized sources. Since the galaxy catalog is highly complete at these distances (see e.g., Figure 6.3), the inference is very robust to galaxy mass function assumptions. As an order-of-magnitude estimate, the GW uncertainty comoving volume for a target angular resolution of 0.2 deg and a redshift uncertainty of 0.003 (i.e. spectroscopic) at $z = 0.2$, is ~ 300 Mpc^3. Using the reference galaxy density of 0.002 Mpc^{-3}, this gives < 1 galaxy in this volume, on average. Therefore, the dark sirens have effectively become bright sirens in this case.

We now discuss the expected GW event number. From the two panels in Figure 6.12, roughly 50 GW events can constrain the Hubble constant value to below 2%. The Hubble constant inference using these effective bright sirens can be robust to astrophysical model error up to $\sim$ 60% change in galaxy mass function. We further observe that the number of included sources increase exponentially to achieve smaller statistical error. This again reflects the fact that deep redshift events offer poor constraining power due to the lack of supporting galaxy catalog information. In the actual analysis, the Bayesian inference framework considers the posterior of

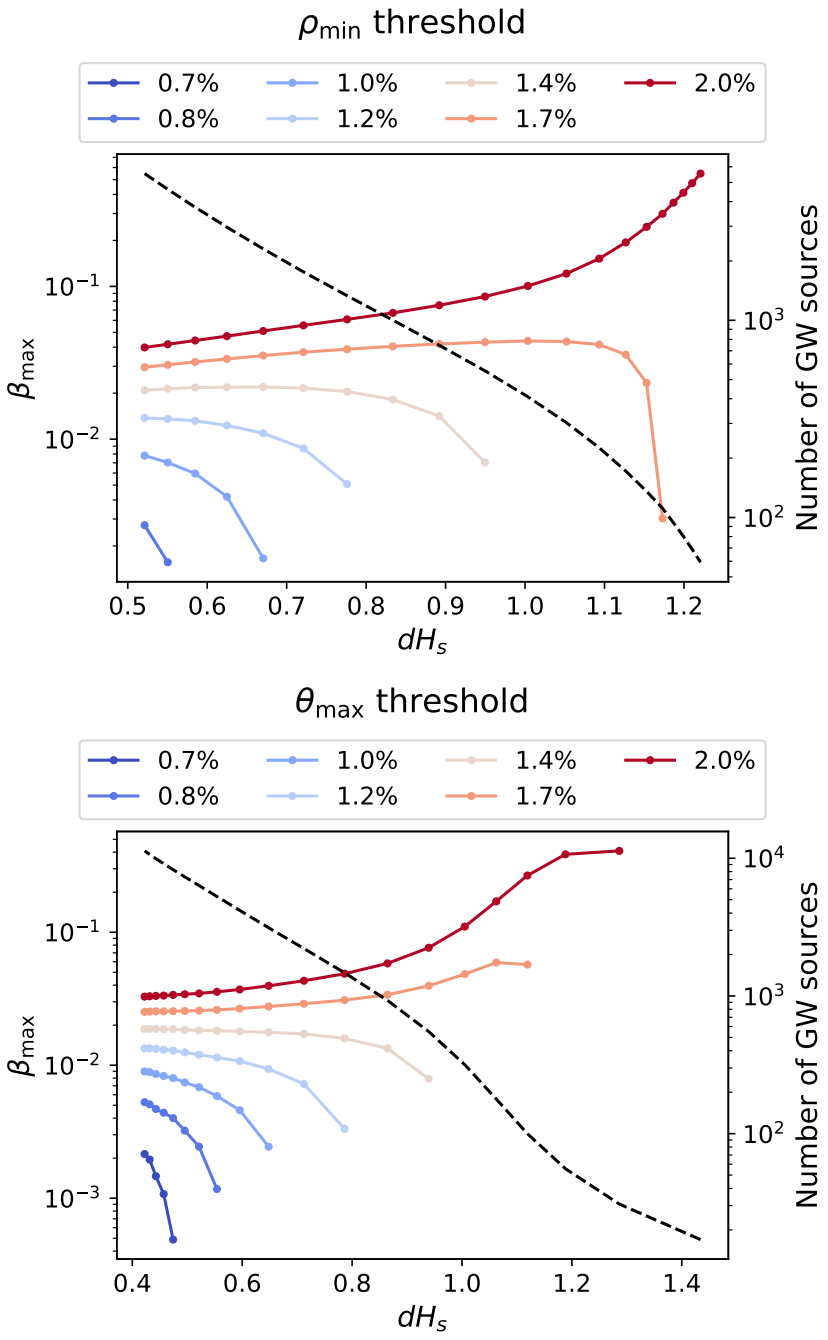

Figure 6.12: Maximum galaxy model error tolerance given various total error budget. The *Top* panel shows SNR selection and the *Bottom* shows the angular resolution criterion. The target total error budget is geometrically distributed between 0.7% and 2%. The right vertical axis shows the expected number of GW sources for each threshold. In the *Top* panel, no selection criterion gives statistical uncertainty at 0.7% H_0, and the corresponding trace is absent. The horizontal axis has a unit of km/s/Mpc.

each GW event instead of using Gaussian error approximation, and the analysis on this exponentially growing GW catalog would pose heavy computational strain with moderate payoff.

Finally, we comment on the overall magnitude of galaxy theoretical model error tolerance. The fiducial models explore effects from different error sources, and the realistic models highlight the transition between dark siren and effective bright siren. We discuss each aspect in turn.

The fiducial models show that, in general, tolerance to β is on the order of $O(0.01)$ if the dark siren method were to achieve a Hubble measurement with a percent-level total error budget. This is expected to be applicable for realistic models as well, if localization errors are larger than our assumed scenario, i.e., CE+LL+ET. Comparing `sparse` and `realistic_bbh_rate`, we see that increasing galaxy spacing or reducing density improves the rate model sharpness, which translates into more robustness against model bias by $\sim 20\%$. While galaxy density is in reality fixed, more complex host galaxy model may achieve similar effect. We discuss this scenario in the next section. Although in the `deepz` simulation we also use high GW event occupation rate, the max β tolerance is higher than the baseline `optimistic` by 23%. This suggests that increasing high-precision galaxy catalog depth is very important for managing potential bias. This result thus highlights the challenge in controlling the bias in the GW dark siren Hubble measurement. Even for low-redshift universe ($z < 0.1$), the Schechter function on the distribution of galaxy luminosity function or surface velocity dispersion function is only constrained with a parameter uncertainty on the order of 5% with the SDSS dataset [18]. Redshift evolution model uncertainties are also limited by the lack of deep redshift observations, and a percent-level knowledge can be challenging.

We now discuss the implication from realistic model tests. Most notably, the error tolerance increases exponentially in the regime when the close-by effective bright sirens already provide sufficient statistical power to constrain the Hubble constant. For our particular condition, this shift occurs when the total error budget is between $1.4\% - 2.0\%$. This result shows that, if GW localization produces sufficient effective bright sirens, the gain for including higher redshift faint sources is limited; while the statistical error is suppressed, susceptibility to bias quickly grows, and the total error budget does not improve. We note that this behavior is not present in fiducial model simulations. This is mostly because the fiducial model localization assumption is more conservative and we have yet to reach the effectively bright siren scenario.

6.6 Conclusion and Future Work

In this work, we use the Fisher information formalism to quantify potential bias in the GW dark siren measurement of the Hubble constant in the context of third-generation GW detector network. We use a mixture model between galaxy catalog and theoretical galaxy number density and simulate the full-sky Fisher information. For both galaxy survey and GW measurement parameters, we create variation models to explore the range of behavior. We assume a power law error in the galaxy mass function and find the maximum error tolerance given a range of total error budgets.

We found that the galaxy redshift error, galaxy completeness fraction (i.e. magnitude limit) and the GW angular localization error are significant factors that contribute to model error tolerance. In the case where GW dark sirens are not effectively localized and close-by sources do not provide the desirable Hubble measurement precision, the error tolerance trend can demonstrate sweet spots, beyond which including further and fainter GW events in the inference starts to impose harsher requirement on our knowledge of the galaxy mass function. Over the range of considered parameters, we notice that to achieve a total Hubble measurement budget of 1%, the maximum β is on the order of $O(0.01)$. In contrast, when well-localized sources (i.e. small measurement uncertainty) can already satisfy the target error budget, it is preferred to only use these effective bright sirens; this dramatically improves tolerance to astrophysical model uncertainties $\beta_{\max} \sim O(1)$. Our study thus shows the importance of jointly considering galaxy survey and GW measurement, and the simulation results can contribute to the decision for GW source selection strategy for dark siren inference of cosmological parameters.

We now discuss several directions for further investigation on this project. Firstly, we can improve on the galaxy catalog model with galaxy clustering. On one hand, galaxies within the same group are closer together, and stronger overlaying of rate volume from each galaxy could lead to a more featureless rates model. On the other hand, the large scale structure "troughs" can be wider than the flat galaxy distribution separation, thus adding more large scale features. Which factor becomes dominate should be clarified with an improved model. We can also add non-uniform galaxy hosting rates for GW depending on galaxy type (e.g. Ref. [15]). Different from an overall redshift evolution in GW occupation rate, this creates more local rates difference, and the smoothing from overlaying rate volume may be mitigated. Both improvement directions focus on modifying the "spikiness" of the Fisher information, and can produce similar results shown with the model

sparse.

Secondly, integration over different GW mass pairs and source location can be done explicitly. In this current approach, we select a random mass pair and treat it as an average case. Overall, the expected effect is a slight addition to the Fisher information, since the detector-frame mass distribution also changes with Hubble constant. However, if we assume that GW sources of all masses (within the detectable range of ground-based detectors) are evenly distributed in galaxies, the rates profile for each GW species can be quite similar. In our simulation, we have shown that sharp features provides the most information, and it may not be expected that smooth mass profile variation adds a lot to our estimate.

Moreover, the simulation results can be refined by considering larger simulation volume with high resolution. We note that the completeness fraction for more relaxed selection thresholds extends well beyond $z = 2$. This creates artifacts in, e.g. the right edge of Fisher information in Figure 6.8 and Figure 6.9 and potentially brings inaccuracies in the β tolerance results. In our searched parameter space, we attempted to avoid thresholds that are too relaxed and include major contribution from higher redshift. In addition, the Fisher information typically drops well before this limit due to the smoothing of the rates model (see, e.g., Figure 6.8), indicating that our current calculation still captures the major effect. These regions are relatively far away from the most interesting parameter space of this work, and should not alter the overall trend and the main conclusions.

References

[1] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116 (6):061102, February 2016. doi: 10.1103/PhysRevLett.116.061102.

[2] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, B. Agarwal, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, others, LIGO Scientific, and Virgo Collaboration. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Phys. Rev. Lett.*, 118(22):221101, June 2017. doi: 10.1103/PhysRevLett.118.221101.

[3] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough,

B. Agarwal, M. Agathos, (INTEGRAL, et al. Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A. *ApJ*, 848(2):L13, October 2017. doi: 10.3847/2041-8213/aa920c.

[4] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, B. Agarwal, M. Agathos, et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551(7678):85–88, November 2017. doi: 10.1038/nature24471.

[5] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, Ligo Scientific Collaboration Kagra Collaboration, VIRGO Collaboration, et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, 21(1):3, April 2018. doi: 10.1007/s41114-018-0012-9.

[6] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, LIGO Scientific Collaboration, and Virgo Collaboration. A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo. *ApJ*, 909(2):218, March 2021. doi: 10.3847/1538-4357/abdcb7.

[7] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, LIGO Scientific Collaboration, Virgo Collaboration, et al. A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo. *ApJ*, 909(2):218, March 2021. doi: 10.3847/1538-4357/abdcb7.

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

[9] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, LIGO Scientific Collaboration, Virgo Collaboration, et al. Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. *ApJ*, 913(1):L7, May 2021. doi: 10.3847/2041-8213/abe949.

[10] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, et al. Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. *ApJ*, 913(1):L7, May 2021. doi: 10.3847/2041-8213/abe949.

[11] Viviane Alfradique, Miguel Quartin, Luca Amendola, Tiago Castro, and Alexandre Toubiana. The lure of sirens: joint distance and velocity measurements with third-generation detectors. *MNRAS*, 517(4):5449–5462, December 2022. doi: 10.1093/mnras/stac2920.

[12] Marek Biesiada, Xuheng Ding, Aleksandra Piórkowska, and Zong-Hong Zhu. Strong gravitational lensing of gravitational waves from double compact binaries-perspectives for the Einstein Telescope. *J. Cosmology Astropart. Phys.*, 2014(10):080–080, October 2014. doi: 10.1088/1475-7516/2014/10/080.

[13] Michael R. Blanton, David W. Hogg, Neta A. Bahcall, J. Brinkmann, Malcolm Britton, Andrew J. Connolly, István Csabai, Masataka Fukugita, Jon Loveday, Avery Meiksin, Jeffrey A. Munn, R. C. Nichol, Sadanori Okamura, Thomas Quinn, Donald P. Schneider, Kazuhiro Shimasaku, Michael A. Strauss, Max Tegmark, Michael S. Vogeley, and David H. Weinberg. The Galaxy Luminosity Function and Luminosity Density at Redshift z = 0.1. *ApJ*, 592(2):819–838, August 2003. doi: 10.1086/375776.

[14] Matteo Califano, Ivan de Martino, Daniele Vernieri, and Salvatore Capozziello. Constraining ΛCDM cosmological parameters with Einstein Telescope mock data. *MNRAS*, 518(3):3372–3385, January 2023. doi: 10.1093/mnras/stac3230.

[15] Liang Cao, Youjun Lu, and Yuetong Zhao. Host galaxy properties of mergers of stellar binary black holes and their implications for advanced LIGO gravitational wave sources. *MNRAS*, 474(4):4997–5007, March 2018. doi: 10.1093/mnras/stx3087.

[16] Hsin-Yu Chen, Maya Fishbach, and Daniel E. Holz. A two per cent Hubble constant measurement from standard sirens within five years. *Nature*, 562 (7728):545–547, October 2018. doi: 10.1038/s41586-018-0606-0.

[17] David F. Chernoff and Lee S. Finn. Gravitational Radiation, Inspiraling Binaries, and Cosmology. *ApJ*, 411:L5, July 1993. doi: 10.1086/186898.

[18] Yun-Young Choi, Changbom Park, and Michael S. Vogeley. Internal and Collective Properties of Galaxies in the Sloan Digital Sky Survey. *ApJ*, 658 (2):884–897, April 2007. doi: 10.1086/511060.

[19] Charles Dalang and Tessa Baker. The clustering of dark sirens' invisible host galaxies. *J. Cosmology Astropart. Phys.*, 2024(2):024, February 2024. doi: 10.1088/1475-7516/2024/02/024.

[20] G. Dálya, R. Díaz, F. R. Bouchet, Z. Frei, J. Jasche, G. Lavaux, R. Macas, S. Mukherjee, M. Pálfi, R. S. de Souza, B. D. Wandelt, M. Bilicki, and P. Raffai. GLADE+ : an extended galaxy catalogue for multimessenger searches with advanced gravitational-wave detectors. *MNRAS*, 514(1):1403–1411, July 2022. doi: 10.1093/mnras/stac1443.

[21] T. de Jaeger, L. Galbany, S. González-Gaitán, R. Kessler, A. V. Filippenko, F. Förster, M. Hamuy, P. J. Brown, DES Collaboration, et al. Studying Type II supernovae as cosmological standard candles using the Dark Energy Survey. *MNRAS*, 495(4):4860–4892, July 2020. doi: 10.1093/mnras/staa1402.

[22] Suhail Dhawan, Saurabh W. Jha, and Bruno Leibundgut. Measuring the Hubble constant with Type Ia supernovae as near-infrared standard candles. *A&A*, 609: A72, January 2018. doi: 10.1051/0004-6361/201731501.

[23] Scott Dodelson. *Modern Cosmology*. 2003.

[24] Jose María Ezquiaga and Daniel E. Holz. Spectral Sirens: Cosmology from the Full Mass Distribution of Compact Binaries. *Phys. Rev. Lett.*, 129(6):061102, August 2022. doi: 10.1103/PhysRevLett.129.061102.

[25] Amanda M. Farah, Thomas A. Callister, Jose María Ezquiaga, Michael Zevin, and Daniel E. Holz. No need to know: astrophysics-free gravitational-wave cosmology. *arXiv e-prints*, art. arXiv:2404.02210, April 2024. doi: 10.48550/arXiv.2404.02210.

[26] Will M. Farr, Maya Fishbach, Jiani Ye, and Daniel E. Holz. A Future Percent-level Measurement of the Hubble Expansion at Redshift 0.8 with Advanced LIGO. *ApJ*, 883(2):L42, October 2019. doi: 10.3847/2041-8213/ab4284.

[27] Andreas Finke, Stefano Foffa, Francesco Iacovelli, Michele Maggiore, and Michele Mancarella. Cosmology with LIGO/Virgo dark sirens: Hubble parameter and modified gravitational wave propagation. *J. Cosmology Astropart. Phys.*, 2021(8):026, August 2021. doi: 10.1088/1475-7516/2021/08/026.

[28] Maya Fishbach, Daniel E. Holz, and Will M. Farr. Does the Black Hole Merger Rate Evolve with Redshift? *ApJ*, 863(2):L41, August 2018. doi: 10.3847/2041-8213/aad800.

[29] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *ApJ*, 622 (2):759–771, April 2005. doi: 10.1086/427976.

[30] R. Gray, C. Messenger, and J. Veitch. A pixelated approach to galaxy catalogue incompleteness: improving the dark siren measurement of the Hubble constant. *MNRAS*, 512(1):1127–1140, May 2022. doi: 10.1093/mnras/stac366.

[31] Rachel Gray, Ignacio Magaña Hernandez, Hong Qi, Ankan Sur, Patrick R. Brady, Hsin-Yu Chen, Will M. Farr, Maya Fishbach, Jonathan R. Gair, Archisman Ghosh, Daniel E. Holz, Simone Mastrogiovanni, Christopher Messenger, Danièle A. Steer, and John Veitch. Cosmological inference using gravitational wave standard sirens: A mock data analysis. *Phys. Rev. D*, 101(12):122001, June 2020. doi: 10.1103/PhysRevD.101.122001.

[32] ChangHoon Hahn, Michael J. Wilson, Omar Ruiz-Macias, Shaun Cole, David H. Weinberg, John Moustakas, Anthony Kremin, Jeremy L. Tinker, et al. DESI Bright Galaxy Survey: Final Target Selection, Design, and Validation. *arXiv e-prints*, art. arXiv:2208.08512, August 2022. doi: 10.48550/arXiv.2208.08512.

[33] Cullan Howlett and Tamara M. Davis. Standard siren speeds: improving velocities in gravitational-wave measurements of H_0. *MNRAS*, 492(3):3803–3815, March 2020. doi: 10.1093/mnras/staa049.

[34] Chinmay Kalaghatgi, Mark Hannam, and Vivien Raymond. Parameter estimation with a spinning multimode waveform model. *Phys. Rev. D*, 101(10): 103004, May 2020. doi: 10.1103/PhysRevD.101.103004.

[35] Marc Kamionkowski and Adam G. Riess. The Hubble Tension and Early Dark Energy. *arXiv e-prints*, art. arXiv:2211.04492, November 2022. doi: 10.48550/arXiv.2211.04492.

[36] Kevin A. Kuns, Hang Yu, Yanbei Chen, and Rana X. Adhikari. Astrophysics and cosmology with a decihertz gravitational-wave detector: TianGO. *Phys. Rev. D*, 102(4):043001, August 2020. doi: 10.1103/PhysRevD.102.043001.

[37] R. Laureijs, J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, et al. Euclid Definition Study Report. *arXiv e-prints*, art. arXiv:1110.3193, October 2011. doi: 10.48550/arXiv.1110.3193.

[38] Michael Levi, Lori E. Allen, Anand Raichoor, Charles Baltay, Segev BenZvi, Florian Beutler, Adam Bolton, Francisco J. Castander, et al. The Dark Energy Spectroscopic Instrument (DESI). In *Bulletin of the American Astronomical Society*, volume 51, page 57, September 2019. doi: 10.48550/arXiv.1907.10688.

[39] Michele Maggiore, Chris Van Den Broeck, Nicola Bartolo, Enis Belgacem, Daniele Bertacca, Marie Anne Bizouard, Marica Branchesi, Sebastien Clesse, Stefano Foffa, Juan García-Bellido, Stefan Grimm, Jan Harms, Tanja Hinderer, Sabino Matarrese, Cristiano Palomba, Marco Peloso, Angelo Ricciardone, and Mairi Sakellariadou. Science case for the Einstein telescope. *J. Cosmology Astropart. Phys.*, 2020(3):050, March 2020. doi: 10.1088/1475-7516/2020/03/050.

[40] Suvodip Mukherjee, Archisman Ghosh, Matthew J. Graham, Christos Karathanasis, Mansi M. Kasliwal, Ignacio Magaña Hernandez, Samaya M. Nissanke, Alessandra Silvestri, and Benjamin D. Wandelt. First measurement of the Hubble parameter from bright binary black hole GW190521. *arXiv e-prints*, art. arXiv:2009.14199, September 2020. doi: 10.48550/arXiv.2009.14199.

[41] Constantina Nicolaou, Ofer Lahav, Pablo Lemos, William Hartley, and Jonathan Braden. The impact of peculiar velocities on the estimation of the Hubble constant from gravitational wave standard sirens. *MNRAS*, 495(1): 90–97, June 2020. doi: 10.1093/mnras/staa1120.

[42] Harshank Nimonkar and Suvodip Mukherjee. Dependence of peculiar velocity on the host properties of the gravitational wave sources and its impact on the measurement of Hubble constant. *MNRAS*, 527(2):2152–2164, January 2024. doi: 10.1093/mnras/stad3256.

[43] Alex Nitz, Ian Harry, Duncan Brown, Christopher M. Biwer, Josh Willis, Tito Dal Canton, Collin Capano, Thomas Dent, Larne Pekowsky, Soumi De, Miriam Cabero, Gareth S Cabourn Davies, Andrew R. Williamson, Duncan Macleod, Bernd Machenschalk, Francesco Pannarale, Prayush Kumar, Steven Reyes, dfinstad, Sumit Kumar, Márton Tápai, Shichao Wu, Leo Singer, veronica villa, Sebastian Khan, Stephen Fairhurst, Koustav Chandra, Alex Nielsen, Shashwat Singh, and Thomas Massinger. gwastro/pycbc: v2.1.1 release of pycbc, March 2023. URL https://doi.org/10.5281/zenodo.7746324.

[44] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, et al. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, September 2020. doi: 10.1051/0004-6361/201833910.

[45] A. Raichoor, J. Moustakas, Jeffrey A. Newman, T. Karim, S. Ahlen, Shadab Alam, S. Bailey, et al. Target Selection and Validation of DESI Emission Line Galaxies. *AJ*, 165(3):126, March 2023. doi: 10.3847/1538-3881/acb213.

[46] David Reitze, Rana X. Adhikari, Stefan Ballmer, Barry Barish, Lisa Barsotti, GariLynn Billingsley, Duncan A. Brown, Yanbei Chen, et al. Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. In *Bulletin of the American Astronomical Society*, volume 51, page 35, September 2019.

[47] Adam G. Riess, William H. Press, and Robert P. Kirshner. A Precise Distance Indicator: Type IA Supernova Multicolor Light-Curve Shapes. *ApJ*, 473:88, December 1996. doi: 10.1086/178129.

[48] Adam G. Riess, Stefano Casertano, Wenlong Yuan, J. Bradley Bowers, Lucas Macri, Joel C. Zinn, and Dan Scolnic. Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM. *ApJ*, 908(1):L6, February 2021. doi: 10.3847/2041-8213/abdbaf.

[49] B. Sathyaprakash, M. Abernathy, F. Acernese, P. Amaro-Seoane, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker, N. Beveridge,

S. Birindelli, S. Bose, L. Bosi, S. Braccini, C. Bradaschia, T. Bulik, et al. Scientific Potential of Einstein Telescope. *arXiv e-prints*, art. arXiv:1108.1423, August 2011.

[50] B. F. Schutz. Determining the Hubble constant from gravitational wave observations. *Nature*, 323(6086):310–311, September 1986. doi: 10.1038/323310a0.

[51] Arman Shafieloo, Ryan E. Keeley, and Eric V. Linder. Will cosmic gravitational wave sirens determine the Hubble constant? *J. Cosmology Astropart. Phys.*, 2020(3):019, March 2020. doi: 10.1088/1475-7516/2020/03/019.

[52] The LIGO Scientific Collaboration, the Virgo Collaboration, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen, et al. Properties and astrophysical implications of the 150 Msun binary black hole merger GW190521. *arXiv e-prints*, art. arXiv:2009.01190, September 2020.

[53] The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, R. Abbott, H. Abe, F. Acernese, K. Ackley, N. Adhikari, R. X. Adhikari, V. K. Adkins, et al. Constraints on the cosmic expansion history from GWTC-3. *arXiv e-prints*, art. arXiv:2111.03604, November 2021. doi: 10.48550/arXiv.2111.03604.

[54] Paul Torrey, Sarah Wellons, Francisco Machado, Brendan Griffen, Dylan Nelson, Vicente Rodriguez-Gomez, Ryan McKinnon, Annalisa Pillepich, Chung-Pei Ma, Mark Vogelsberger, Volker Springel, and Lars Hernquist. An analysis of the evolving comoving number density of galaxies in hydrodynamical simulations. *MNRAS*, 454(3):2770–2786, December 2015. doi: 10.1093/mnras/stv1986.

[55] Licia Verde, Tommaso Treu, and Adam G. Riess. Tensions between the early and late Universe. *Nature Astronomy*, 3:891–895, September 2019. doi: 10.1038/s41550-019-0902-0.

[56] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272, February 2020. doi: 10.1038/s41592-019-0686-2.

[57] Hang Yu, Brian Seymour, Yijun Wang, and Yanbei Chen. Uncertainty and Bias of Cosmology and Astrophysical Population Model from Statistical Dark Sirens. *ApJ*, 941(2):174, December 2022. doi: 10.3847/1538-4357/ac9da0.

[58] Rongpu Zhou, Biprateep Dey, Jeffrey A. Newman, Daniel J. Eisenstein, K. Dawson, S. Bailey, et al. Target Selection and Validation of DESI Luminous Red Galaxies. *AJ*, 165(2):58, February 2023. doi: 10.3847/1538-3881/aca5fb.

[59] Liang-Gui Zhu, Yi-Ming Hu, Hai-Tian Wang, Jian-dong Zhang, Xiao-Dong Li, Martin Hendry, and Jianwei Mei. Constraining the cosmological parameters using gravitational wave observations of massive black hole binaries and statistical redshift information. *Physical Review Research*, 4(1):013247, March 2022. doi: 10.1103/PhysRevResearch.4.013247.

[60] Andrea Zonca, Leo Singer, Daniel Lenz, Martin Reinecke, Cyrille Rosset, Eric Hivon, and Krzysztof Gorski. healpy: Python wrapper for HEALPix. Astrophysics Source Code Library, record ascl:2008.022, August 2020.

catalog name	σ_g	m_g	$d \log \hat{D}$	$\hat{\kappa}$	equiv. θ_0 [deg]	$f_{\mathrm{g,red}}$	$r_{\mathrm{obs,GW}}$ [$\mathrm{Gpc^{-3}yr^{-1}}$]
optimistic	0.003	20.175	0.001	10^4	0.8	1.	239
realistic_bbh_rate	0.003	20.175	0.001	10^4	0.8	1.	23.9
bright	0.003	19.5	0.001	10^4	0.8	1.	239
photo	0.01	20.175	0.001	10^4	0.8	1.	239
ang	0.003	20.175	0.001	10^3	2.3	1.	239
lum	0.003	20.175	0.01	10^4	0.8	1.	239
sparse	0.003	20.175	0.001	10^4	0.8	0.1	239
deepz	0.003	22	0.001	10^4	0.8	1.	239

Table 6.1: Table of fiducial model catalogs parameters. From left to right, the columns represent catalog name, galaxy redshift uncertainty σ_g, galaxy survey magnitude limit m_g, fractional uncertainty in GW event luminosity distance $d \log \hat{D}$, GW angular uncertainty parameter coefficient $\hat{\kappa}$, the equivalent angular uncertainty at $z = 0$, θ_0, galaxy number density scalar $f_{\mathrm{g,red}}$ and GW number density $r_{\mathrm{obs,GW}}$. See text for detailed explanation of parameters.

catalog name	threshold type	threshold value
theta_sm	angular uncertainty	0.4 deg
theta_lg	angular uncertainty	1. deg
rho_sm	SNR	100
rho_lg	SNR	200

Table 6.2: Table of model catalog parameters using realistic GW measurement uncertainty from Fisher information calculations. From left to right, the columns represent catalog name, source selection threshold type and the corresponding threshold value. The galaxy catalog parameters and number density scaling is the same as the fiducial model *realistic_bbh_rate*.

catalog	*optimistic*	*bright*	*sparse*	*deepz*
$N_{\rm gal}$	7047	3392	684	45251

Table 6.3: Number of simulated galaxies in each model. Note that models with the same galaxy completeness fraction and number density share the same galaxy catalog.

Appendix A

KEPLER AS ASTROMETRY PROBE OF GRAVITATIONAL-WAVE

Due to the small expected signal, the astrometry detection method benefits greatly from a large number of observed targets and high cadence. For this reason, the Roman Space Telescope and *Gaia* are considered especially promising candidates. In this appendix, we discuss our ongoing work using archival data from the *Kepler* mission for GW constraints.

Launched in 2009, the *Kepler* Mission is a space-borne observatory designed to monitor ~170,000 stars to look for exoplanets in the Milky Way [3, 4]. Although measuring GWs astrometrically was not among the original goals for *Kepler*, its survey design has characteristics that are suitable for probing GWs. Throughout its mission, the spacecraft points towards the same patch of sky. Given the 30-minute cadence and a total observation time of 3.5 years, the number of exposures are around 61,000 [8]. The single-star single-exposure astrometric error is magnitude dependent, but in general a level of 1 mas is achievable (Figure 1 in Ref. [8]). These factors suggest that the strain sensitivity of *Kepler* is roughly 4.8×10^{-14} rad (see Eqn. (4.10)). For a supermassive binary black hole with a chirp mass of $10^9\ M_\odot$ at a frequency of 5×10^{-7} Hz, the detectable range is within 40 Mpc. Although such a source is optimistic, this demonstrates that *Kepler* could potentially be sensitive to GWs from neighboring physical sources.

Another advantage of *Kepler* is its 105 sq degrees field-of-view (FoV) [3], which is about 400 times larger than that of Roman. In Chapter 4, we highlight the loss of the mean signal due to the large-scale nature of GWs. For Roman FoV. the mean-free signal magnitude is about 100 times smaller than the full signal. With a larger FoV, the mean-free signal takes up a larger fraction of the full signal; even in the case where we can only rely on relative astrometry, the resulting signal is more optimistic in *Kepler* than in Roman.

Our work is mostly focused on understanding the systematics and building a pipeline to clean the *Kepler* data. Firstly, we apply quality filtering to stars in the *Kepler* catalog. We first remove stars with measurable rotations flagged in Ref. [7, 9, 10], since the centroid measurements can change superficially due to stellar surface brightness

changes during rotation. We then download the long cadence lightcurve data from the Mikulski Archive for Space Telescopes (MAST)[1]. The centroid measurement quality can be affected by noise of known physical origin, such as coronal mass ejection and cosmic ray activity [11]. Telescope motion, such as when the spacecraft gradually settles back to the target FoV after downlinking data to Earth every 30 days, can also contaminate the data [11]. We offer the option to filter exposures with known issues.

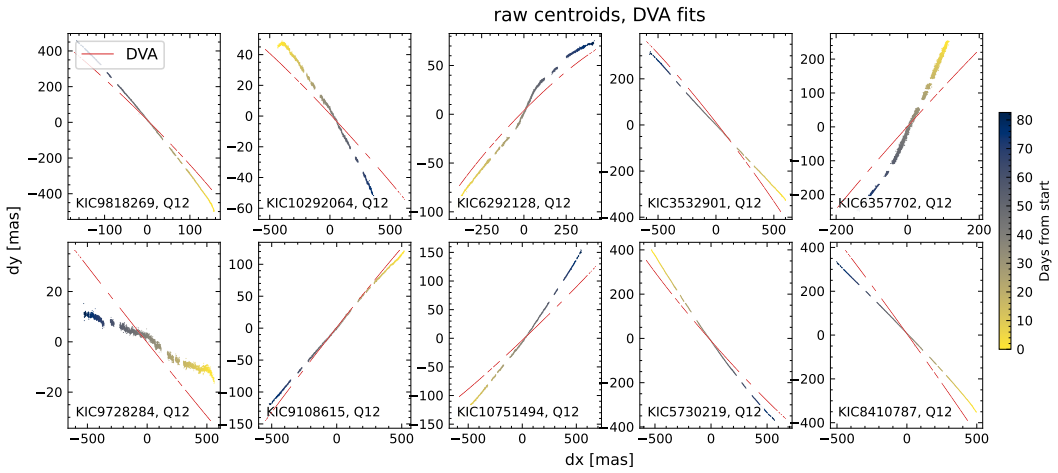

Figure A.1: Raw centroid data with differential velocity aberration fits for a selection of *Kepler* targets in Quarter 12. The horizontal and vertical axes represent the row and column in each output. The centroid scatter color scale shows the number of days since the quarter beginning.

We find that the largest motion in the centroids is due to differential velocity aberration (DVA) as the telescope is moving with respect to stars. We use the Python wrapper package `RaDec2Pix`[2] to compute the expected DVA for each star at each exposure. The DVA can then be removed from raw centroids, reducing the single-quarter centroid movements from several hundreds of mas to $O(10-100)$ mas. However, accounting for the DVA fully requires knowing the telescope pointing and velocity and the true sky location of the stars. In the `RaDec2Pix` package, the telescope pointing is loaded from a fixed input array, and the velocity is taken from the `spiceypy` package [1] that provides nominal spacecraft velocity. We also use the nominal ra and dec which ultimately come from legacy surveys in the *Kepler* input catalog. These steps can introduce errors, and in practice, we found only part of the DVA is removed, and the residual is mostly still dominated by DVA-like

[1] `https://archive.stsci.edu/missions-and-data/kepler`
[2] `https://github.com/stevepur/Kepler-RaDex2Pix`

motion. In Figure A.1, we show the raw centroid data (`MOM_CENTR`) for a random selection of ten stars out of a curated sample. For this sample, we discard stars with noticeable rotation and have measured magnitude beyond the saturation limit (`kic_kepmag`>11.4, [5]). We then select the brightest 100, and 89 have valid data in Quarter 12. The breaks in data represent a duration of roughly three days after each data downlink, when the centroids show spurious movement due to telescope motion. The *Kepler Input Catalog* ID for each star is shown in the lower left corner in each panel. The DVA is shown as red scatters.

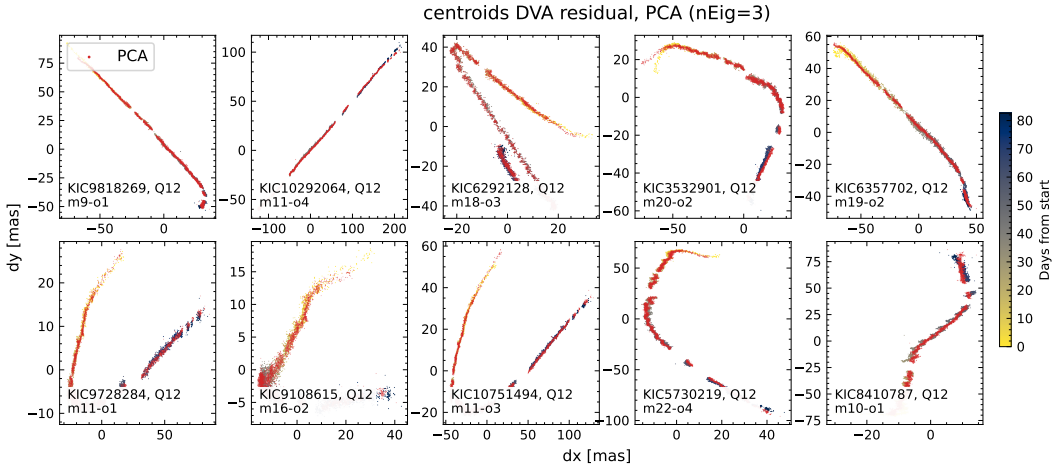

Figure A.2: Centroid residual after DVA correction. The best-fit model using the leading three PCA eigenvectors is shown in red. The horizontal and vertical axis represent the Cartesian coordinate tangential to the celestial sphere along the ra and declination direction, respectively.

We also build the functionality to perform principal component analysis (PCA) on the DVA residual. This is motivated for removing FoV-scale aberrations due to, e.g. thermal effects. Although ancillary data on primary mirror temperatures are available [11], we do not attempt to forward-model the resulting motion due to instrument complexity and the temperature data resolution (four sensors exist on the primary mirror). Removing leading eigenvectors, the residual is reduced to $O(1-10)$ mas, which is consistent with the quoted astrometric error. In Figure A.2, we show the residual after accounting for DVA using the same set of stars. We also show the 3-eigenvector fit with red scatters. In this example, the PCA is performed on the 89 stars. The detector module and output number for each star is noted in the bottom left in each panel. Different from Figure A.1 that shows the row-column coordinate, this figure shows the projected astrometry location of stars on the sky, and the horizontal and vertical axes represent unit vector directions along ra and

dec. In Figure A.3, we show the centroid residual after removing the leading three eigenvectors in PCA decomposition.

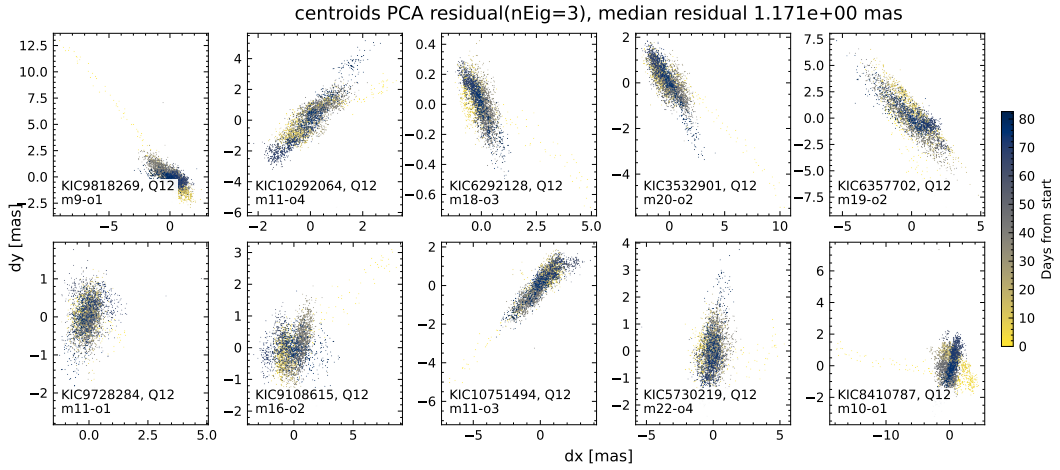

Figure A.3: Centroid residual after removing three leading PCA eigenvectors. The horizontal and vertical axis represent the Cartesian coordinate tangential to the celestial sphere along the ra and declination direction, respectively.

We now discuss several challenges and future directions to this project. The *Kepler* dataset is extremely rich, and it is important to understand which data to query for our purpose. For example, the stellar centroids can be computed from flux data in several ways. In the simpler case, it is possible to define an integration area and compute the first moment of flux (see, e.g., Ref. [2]). In the *Kepler* lightcurve file, this quantity is labeled `MOM_CENTR` and we have mostly used this in the testing process. It is also possible to use the point spread function in each detector output to fit for the most likely incident chip location. Only the Photometric Performance Assessment (PPA) stars offer this measurement (labeled as `PSF_CENTR`) [6], and the total number is about 10 times smaller than the whole catalog. Understanding how these quantities are constructed and their recommended usage is extremely important for removing systematics as much as possible and retaining a faithful representation of the stellar centroids. Instead of computing DVA, it is also possible to use another data column named `POS_CORR`. This column is fitted from motion polynomials, and supposedly captures effects such as DVA, telescope pointing and thermal effects [2]. In our current tests using the PPA stars among the brightest non-saturated 100 stars in quarter 12, we do not observe a substantial qualitative difference between using these different data products.

It is also important to investigate how to combine quarters. *Kepler* FoV rotates

90 degrees every three months, which means the same stars will fall to a different part of chip [4]. Consequently, the systematics can be quite different. In fact, in Ref. [2], only quarters with the same telescope rotation orientation are analyzed together for more stable results. Currently, our analysis has been limited to a single quarter. Understanding how to best combine different quarters will be paramount for exploiting the full statistical power of the survey measurement and revealing the potential of *Kepler* as a GW probe.

In conclusion, we have set the base for using archival *Kepler* data to probe GWs. We built the pipeline to clean data and to remove several systematic noise factors. In the current test, the cleaned centroids are consistent with the expected error budget. We also outline modeling and technical future challenges for improving this pipeline and finally applying to the *Kepler* dataset.

References

[1] A. Annex. SpiceyPy, a Python Wrapper for SPICE. In LPI Editorial Board, editor, *Third Planetary Data Workshop and The Planetary Geologic Mappers Annual Meeting*, volume 1986 of *LPI Contributions*, page 7081, June 2017.

[2] G. Fritz Benedict, Angelle M. Tanner, Phillip A. Cargile, and David R. Ciardi. A Technique to Derive Improved Proper Motions for Kepler Objects of Interest. *AJ*, 148(6):108, December 2014. doi: 10.1088/0004-6256/148/6/108.

[3] William J. Borucki. KEPLER Mission: development and overview. *Reports on Progress in Physics*, 79(3):036901, March 2016. doi: 10.1088/0034-4885/79/3/036901.

[4] William J. Borucki, David G. Koch, Jack J. Lissauer, Gibor B. Basri, John F. Caldwell, William D. Cochran, Edward W. Dunham, John C. Geary, David W. Latham, Ronald L. Gilliland, Douglas A. Caldwell, Jon M. Jenkins, and Yoji Kondo. The Kepler mission: a wide-field-of-view photometer designed to determine the frequency of Earth-size planets around solar-like stars. In J. Chris Blades and Oswald H. W. Siegmund, editors, *Future EUV/UV and Visible Space Astrophysics Missions and Instrumentation.*, volume 4854 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 129–140, February 2003. doi: 10.1117/12.460266.

[5] Ronald L. Gilliland, Jon M. Jenkins, William J. Borucki, Stephen T. Bryson, Douglas A. Caldwell, Bruce D. Clarke, Jessie L. Dotson, Michael R. Haas, Jennifer Hall, Todd Klaus, David Koch, Sean McCauliff, Elisa V. Quintana, Joseph D. Twicken, and Jeffrey E. van Cleve. Initial Characteristics of Kepler Short Cadence Data. *ApJ*, 713(2):L160–L163, April 2010. doi: 10.1088/2041-8205/713/2/L160.

[6] Jie Li, Christopher Allen, Stephen T. Bryson, Douglas A. Caldwell, Hema Chandrasekaran, Bruce D. Clarke, Jay P. Gunter, Jon M. Jenkins, Todd C. Klaus, Elisa V. Quintana, Peter Tenenbaum, Joseph D. Twicken, Bill Wohler, and Hayley Wu. Photometer performance assessment in Kepler science data processing. In Nicole M. Radziwill and Alan Bridger, editors, *Software and Cyberinfrastructure for Astronomy*, volume 7740 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 77401T, July 2010. doi: 10.1117/12.856788.

[7] A. McQuillan, T. Mazeh, and S. Aigrain. Rotation Periods of 34,030 Kepler Main-sequence Stars: The Full Autocorrelation Sample. *ApJS*, 211(2):24, April 2014. doi: 10.1088/0067-0049/211/2/24.

[8] David G. Monet, Jon M. Jenkins, Edward W. Dunham, Stephen T. Bryson, Ronald L. Gilliland, David W. Latham, William J. Borucki, and David G. Koch. Preliminary Astrometric Results from Kepler. *arXiv e-prints*, art. arXiv:1001.0305, January 2010. doi: 10.48550/arXiv.1001.0305.

[9] A. R. G. Santos, R. A. García, S. Mathur, L. Bugnet, J. L. van Saders, T. S. Metcalfe, G. V. A. Simonian, and M. H. Pinsonneault. Surface Rotation and Photometric Activity for Kepler Targets. I. M and K Main-sequence Stars. *ApJS*, 244(1):21, September 2019. doi: 10.3847/1538-4365/ab3b56.

[10] A. R. G. Santos, S. N. Breton, S. Mathur, and R. A. García. Surface Rotation and Photometric Activity for Kepler Targets. II. G and F Main-sequence Stars and Cool Subgiant Stars. *ApJS*, 255(1):17, July 2021. doi: 10.3847/1538-4365/ac033f.

[11] Jeffrey E. Van Cleve, Jessie L. Christiansen, Jon M. Jenkins, Douglas A. Caldwell, Thomas Barclay, Stephen T. Bryson, Christopher J. Burke, Jennifer Campbell, Joseph Catanzarite, Bruce D. Clarke, Jeffrey L. Coughlin, F. Girouard, Michael R. Haas, Todd C. Klaus, Jeffrey J. Kolodziejczak, Jie Li, Sean D. McCauliff, Robert L. Morris, Fergal Mullally, Elisa V. Quintana, Jason Rowe, Anima Sabale, Shawn Seader, Jeffrey C. Smith, Martin D. Still, Peter G. Tenenbaum, Susan E. Thompson, Joesph D. Twicken, Akm Kamal Uddin, and Khadeejah Zamudio. Kepler Data Characteristics Handbook. Kepler Science Document KSCI-19040-005, id. 2. Edited by Doug Caldwell, Jon M. Jenkins, Michael R. Haas and Natalie Batalha, December 2016.