THE UNIVERSITY OF CHICAGO

# HEARING AND SEEING THE UNIVERSE: RESULTS FROM GRAVITATIONAL-WAVE AND OPTICAL STUDIES OF MERGING NEUTRON STARS AND BLACK HOLES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
ZOHEYR DOCTOR

CHICAGO, ILLINOIS
AUGUST 2019

Dedicated to Dr. Colin Day, who showed me the beauty of calculus.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

Nature has been very kind to me during the course of my PhD. Just after beginning my work in gravitational waves, we detected gravitational waves for the first time ever and just two years later we detected a gravitational wave with accompanying electromagnetic signals. These incredible events have made for an exciting few years, both in learning from the experts and in contributing to the field myself. Of course, I wouldn't have been able to be part of this scientific journey without the mentorship of my advisor, Prof. Daniel Holz. Working with Daniel for the last four years has been spectacular. He consistently finds interesting approaches to problems, pinpoints the most relevant questions around a given topic, and makes simple, intuitive, and robust arguments. Daniel was also kind enough to help me secure multiple fellowships and he supported me in traveling to a number of conferences where I made connections with other researchers. I look forward to further collaboration with him.

I have also had the incredible opportunity to work with a number of other researchers in LIGO, Virgo, and the Dark Energy Survey. Principally, I worked closely with Dr. Richard Kessler, and am sincerely indebted to him for the guidance and mentorship he gave me on two of my major projects as a graduate student. Rick was always available to talk to me and help me with my scientific writing. I would not be the writer I am today without him.

Others I am grateful to have worked with include (and are definitely not limited to): Maya Fishbach, Reed Essick, Phil Landry, Ben Farr, Hsin-Yu Chen, Marcelle Soares-Santos, Richard O'Shaughnessy, and Michael Coughlin. I must also acknowledge my thesis committee who I had a number of interesting conversations with coming up to my thesis defense: Robert Wald, David Miller, and Craig Hogan.

Next, I would like to thank all my friends and acknowledge a few here specifically. My girlfriend Darshana Nair has given me consistent support over the last few months as I completed the thesis, and she continues to amaze me with her excitement and inquisitiveness about astrophysics. Janosz Dewberry, my best friend, has been my partner in crime ever

since second grade and to this day we continue to follow parallel paths, enabling us to always work our way up together. I've learned a lot from him both as a friend and as a peer. I would also like to acknowledge David Dewberry and Jannie Korchinski for being my second parents and always having an extra spot for me at their dinner table and house. I also want to acknowledge my friends in Chicago who have been here for me during the course of my PhD, including Akash Dixit, Andrew Ludwig, Sam Whiteley, Aziza Suleymanzade, Ryan McGeehan, Jon Trisnadi, Rhys Povey, Torben Noto, Annie Nelson, Eric Wilson, Valerie Frank, and Gourav Khullar.

Finally, I want to thank my family. My brother Daniel is truly the best brother in the world. He's grown into an amazing man, and no longer do I feel like the older brother. I look up to him for his thoughtfulness, his attention to detail, his care, and his great sense of humor. And of course, I couldn't have made it anywhere without my wonderful parents Munira and Taizoon Doctor. They have consistently fought for me to have all the things I wanted out of life, and have supported me unconditionally all the way. They made sure I went to the best schools and universities possible, gave me a strong moral compass, and also showed me the importance of caring for and loving others.

# ABSTRACT

Since the first direct detection of a gravitational wave by the Laser Interferometer Gravitational-Wave Observatory in 2015, gravitational waves have become an indispensable tool for studying extreme astrophysical phenomena. To date, eleven of these "ripples in spacetime" have been detected – ten from merging pairs of black holes and one from two colliding neutron stars. In principle, electromagnetic and neutrino signals can be combined with gravitational-wave data to form a more complete picture of these compact object mergers. In this thesis, we present four studies of gravitational-wave and optical signatures from black-hole and neutron-star mergers. First we describe a technique for emulating expensive simulations of gravitational waveforms using Gaussian process regression, a method for non-parametrically interpolating functions and their uncertainties. The second topic we address is that of the mass ejected from the collision of the two neutron stars that produced the LIGO-Virgo gravitational wave GW170817. We find that the mass dynamically ejected from the merger should enrich its surroundings with heavy r-process elements, suggesting that neutron-star mergers could have played a significant role in the production of heavy elements we see in our solar system. Next we present the analysis and upper limits from the Dark Energy Survey of an independent optical search for kilonovae, the bright optical transients associated with neutron-star mergers. Finally, we describe Dark Energy Camera optical follow-up of black-hole merger GW170814 and the results of that search.

# CHAPTER 1

# INTRODUCTION

The study of astronomy is humankind's greatest attempt to find our place in the universe and put together a coherent story of how the cosmos came to be the way it is on large scales. We are able to study distant objects and phenomena in our universe because they produce signals which we can capture here on Earth and then compare with theoretical models. For almost all of human history, our knowledge about these objects and phenomena came from the light they emit, i.e. their radiation in the electromagnetic spectrum. However, the 20th Century ushered in new ways of collecting astronomical information: space travel and astro-particle detectors enabled us to gather material from space and detect particles emitted from a wide array of astrophysical processes. Then, in 2015, the landscape of astronomy and astrophysics changed once again: The Laser Interferometer Gravitational-Wave Observatory detected *gravitational waves*, "ripples in spacetime", from the collision of two black holes, opening yet another window on the universe through which we can detect high-energy astrophysical events (Abbott et al., 2016c). While gravitational waves carry considerable information about their sources, combining information from gravitational-wave detections with potential accompanying electromagnetic and astro-particle signals gives us an even richer understanding of the most extreme phenomena in the universe.

To date eleven gravitational waves have been confidently detected and thoroughly characterized, starting with GW150914, the first gravitational wave source directly observed. Ten of these detections came from pairs of colliding black holes hundreds of millions to billions of light years away (The LIGO Scientific Collaboration & the Virgo Collaboration, 2018). The eleventh signal, GW170817, came from a neutron-star merger, and it was followed by emission across the electromagnetic spectrum, allowing the astronomical community to for the first time combine electromagnetic and gravitational-wave data.

This thesis is concerned with *multi-messenger gravitational-wave astrophysics*, the modeling, detection, and characterization of gravitational waves, their accompanying electro-

magnetic or astro-particle signals, and the astrophysical sources of these signals. Some key questions addressed herein include:

- How do features of gravitational-wave sources imprint on the gravitational waves measured here on Earth?
- What are the implications of gravitational-wave signals on their accompanying electromagnetic signals and on the production of heavy elements?
- How often do gravitational-wave events occur in the universe?
- Which gravitational-wave sources also emit light?

In this chapter I outline the central concepts of multi-messenger gravitational-wave astrophysics and give a brief overview of the state of the field. The subsequent chapters present published works in which I am a principal author and contributor.

## 1.1 Gravitational Waves

Gravitational waves (GWs) are a natural consequence of Einstein's theory of general relativity, which has been immensely successful over the last century in predicting and explaining a range of phenomena in astrophysics and cosmology. The theory is encapsulated in the Einstein field equations, 10 coupled, non-linear differential equations:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{1.1}$$

The left-hand side of this equation depends on the spacetime metric $g_{\mu\nu}$, which encodes the spacetime intervals between events and the causal structure of spacetime. The right-hand side features the term $T_{\mu\nu}$, the stress-energy tensor, which records the distribution of matter and energy in the spacetime. In short, the Einstein field equations stipulate that there is an intimate interplay between the curvature of spacetime and the matter and energy in spacetime.

In general, the Einstein field equations are difficult to solve analytically, but in some cases they are tractable without resorting to the use of a computer. For example, if one linearizes general relativity and considers only small perturbations $h_{\mu\nu}$ to a background spacetime metric $\eta_{\mu\nu}$, the perturbations can be described by a wave equation with a source term after making an appropriate gauge choice:

$$\partial^a \partial_a \bar{h}^{\mu\nu} = \left( -\frac{\partial^2}{\partial t^2} + \nabla^2 \right) \bar{h}^{\mu\nu} = -\frac{16\pi G}{c^4} T^{\mu\nu}. \tag{1.2}$$

Here, $\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h^{\alpha}_{\alpha}$, but when the source term is zero, residual gauge freedom allows one to impose $\bar{h}_{\mu\nu} = h_{\mu\nu}$ . Wave solutions to this equation are aptly named gravitational waves, and they are sourced by the matter and energy which are coded into the right-hand side. If we consider gravitational radiation far away from a slowly varying source, Equation 1.2 can be simplified further to depend only on the second time derivative of the traceless quadrupole moment, $Q_{ij}$, and the (luminosity) distance to the source from the observer, $d_L$:

$$h^{ij} = \frac{2}{d_L} \frac{G}{c^4} \frac{d^2}{dt^2} Q^{ij} \tag{1.3}$$

Equation 1.3 implies that gravitational waves are sourced by any spherically-asymmetric accelerating matter, but in almost all cases the magnitude of these waves are negligibly small, because the ratio $G/c^4$ is minuscule ($\sim 10^{-20}$ Mpc/($M_\odot c^2$)). Therefore, only systems with a large second time derivative of their quadrupole moment are good candidates for sources of detectable gravitational waves. In 1974, Hulse and Taylor reported the first binary-neutron-star system containing a pulsar, PSR 1913 + 16, and it was subsequently shown that the orbital period of these two neutron stars was shrinking in accordance with general relativity's prediction of gravitational-wave emission from an orbiting binary system. The orbital decay of PSR 1913 + 16 was strong evidence that gravitational waves exist and that they can be sourced by binary star systems. However, the direct detection of gravitational waves remained elusive for another four decades until the detection of gravitational wave

GW150914 by the Laser Interferometer Gravitational-Wave Observatory.

## 1.2 Direct Detection of Gravitational Waves

At the time of this writing, gravitational waves have only been directly measured with Earth-based interferometers. The Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors in Hanford, WA, and Livingston, LA, and the European Virgo detector are *Michelson interferometers*, which each consist of laser beams traveling back and forth down two kilometer-scale perpendicular arms to measure fluctuations in the relative light travel time down the arms and hence the passage of gravitational waves. These interferometers sensitive to the projection of $h_{\mu\nu}$ onto the interferometer arms which yields a "strain" time series $h(t) = (\delta L_x - \delta L_y)/L$, where $L = L_x = L_y$ are the lengths of the perpendicular arms in the presence of no gravitational wave. Since the recent upgrades of these detectors to their "advanced" configurations, they are capable of reaching noise amplitude spectral densities of $\sim 10^{-22}$ Hz$^{-1/2}$ in the audio band (tens to thousands of Hz).

Although these advanced interferometers are some of the most sensitive instruments ever built, they are not free from noise. For example, they exhibit noise from seismic activity, Poisson fluctuations in the laser amplitude, and radiation pressure of the laser on the mirrors which reflect the laser light back down the interferometer arms. Extracting gravitational-wave signals from the noise background is therefore non-trivial and requires a suite of techniques that prevent misidentification of noise as a gravitational-wave signal.

Two overall search strategies are employed to find gravitational waves buried in the noise. The first strategy, known as a "matched filter" search, compares interferometer time-series strain data $h(t)$ to specific predicted gravitational waveforms. The second strategy does not assume a specific model for the gravitational waveforms and instead looks for simultaneous excess power in multiple interferometers Abbott et al. (2016). When a signal above some significance threshold is discovered by a search pipeline, Bayesian inference techniques are used to find the posterior probability $p(\boldsymbol{\theta}|h(t), S_n(f))$ that parameters $\boldsymbol{\theta}$ of some signal

model $s_{\boldsymbol{\theta}}(t)$ are the correct description of the signal given the observed data $h(t)$ and the assumption of additive Gaussian noise with power spectral density $S_n(f)$. The likelihood of getting $h(t)$ is given by

$$p(h(t)|s_{\boldsymbol{\theta}}(t), S_n(f)) \propto \exp\left(-\frac{1}{2}\langle h(t) - s_{\boldsymbol{\theta}}(t)|h(t) - s_{\boldsymbol{\theta}}(t)\rangle\right) \tag{1.4}$$

where

$$\langle a(t)|b(t)\rangle = 4\Re\int_0^\infty \frac{\tilde{a}(f)\tilde{b}*(f)}{S_n(f)}df \tag{1.5}$$

and $\tilde{q}(f)$ denotes the Fourier transform of a time series $q(t)$. With this likelihood, Bayes' Rule and a choice of priors on the model parameters $p(\boldsymbol{\theta})$ are used to calculate the posterior probabilities of parameters $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta}|h(t), S_n(f)) \propto p(h(t)|s_{\boldsymbol{\theta}}(t), S_n(f))p(\boldsymbol{\theta}) \tag{1.6}$$

Using this scheme, one can infer the physical parameters of the system that produced an observed gravitational wave.

## 1.3 Black-Hole and Neutron-Star Binaries as GW Sources

*Black holes* are regions of spacetime where gravity is so strong that not even light can escape. Although they are just curved yet empty spacetime, they are often referred to as if they were spherical objects, and the remainder of this thesis will treat them as such. Black holes are expected to form from the collapse of massive stars which have undergone a supernova, and up until the detection of GW150914, the existence of these stellar-mass black holes was primarily known indirectly through measurements of x-ray binaries.

Like black holes, *neutron stars* are remnants of massive stars that have exhausted their fuel and gone supernova, leaving nothing but a dense ball of mostly neutrons held up against the force of gravity from nuclear forces and neutron degeneracy pressure. Neutron stars

provide a prime testing ground for physical theories, because all four fundamental forces (electromagnetic, weak, strong, and gravitational) are strongly at play inside and near these objects. Notably, measurements of the masses, radii, and/or tidal deformabilities of neutron stars have direct implications for the behavior of the strong force in dense matter. Observations of gravitational waves enable us to measure the masses, spins, and structures of neutron stars and black holes and test fundamental theories of physics.

The decaying orbit ("inspiral") and eventual collisions of pairs of compact objects (neutron stars and/or black holes) are ideal candidates for gravitational-wave sources that are detectable with ground-based interferometers because black holes and neutron stars are massive objects yet small in spatial extent. This enables them to orbit close to one another before colliding and reach orbital frequencies in the audio band and generate a large time-varying quadrupole moment. Additionally, the inspiral of these binary systems is simply described. Assuming both objects in the binary are massive point particles, Equation 1.3 yields a relation between the frequency $f$ of the observed gravitational wave and the time derivative of the frequency $\dot{f}$ which depends only on a combination of the masses of the two stars $m_1$ and $m_2$ called the chirp mass $\mathcal{M}$ (Abbott *et al.*, 2016c):

$$\mathcal{M} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} = \frac{c^3}{G}\left[\frac{5}{96}\pi^{-8/3} f^{-11/3} \dot{f}\right]^{3/5} \tag{1.7}$$

Various corrections to Equation 1.7 are required to describe the gravitational wave's phase evolution to higher order, but overall the chirp mass drives the observed signal and is therefore relatively well measured compared to other parameters of the system such as the mass ratio $q = m_1/m_2$. Our ability to model the gravitational wave signals from these binaries as a function of the physical parameters of the two compact objects enables us to utilize Equations 1.4 and 1.6 to infer the properties of the merging black holes or neutron stars.

## 1.4 Electromagnetic Counterparts to Compact Binary Mergers

The multi-messenger discovery of GW170817 and its accompanying electromagnetic signals confirmed that neutron-star mergers can source both gravitational and electromagnetic waves. In the case of GW170817, a gamma-ray burst was recorded within seconds of the gravitational wave, optical and infrared signals were detected less than 12 hours after that, and eventually afterglow signals from the gamma ray burst were seen across the electromagnetic spectrum. Whether the binary-black-hole mergers or neutron-star-black-hole mergers that occur in nature are capable of creating electromagnetic counterparts is currently unknown, but future measurements could address this possibility.

There is a wide range of science cases for trying to detect the potential electromagnetic counterparts of gravitational waves from compact binary mergers. (Herein, I refer to these searches to gravitational-wave electromagnetic counterparts as "EM follow-up"). Firstly, any electromagnetic signals from these gravitational-wave-emitting systems imply the presence of matter (or some hitherto unknown physical process). The specific morphology of the electromagnetic counterparts tells us about the configuration and composition of the matter involved in binary mergers, which have implications for the formation of the heavy elements and the structure of the colliding bodies. A counterpart that is well-localized on the sky also enables one to fix the sky position when performing parameter estimation on the gravitational-wave data and also to identify the galaxy hosting the gravitational-wave event. With knowledge of the host galaxy, one can measure the current expansion rate of the universe and further test the formation mechanisms of these binary systems. In sum, finding electromagnetic counterparts to gravitational waves has immense scientific value.

## 1.5 Outline of This Work

This thesis is a compilation of selected published works in multi-messenger gravitational-wave astrophysics that I have led or majorly contributed to with support from independent

collaborators, the LIGO-Virgo Collaboration, and the Dark Energy Survey. Each chapter comprises one of these publications, listed below, and includes a short summary at the beginning which puts the chapter in context. The following chapters are:

- **Chapter 2**: *Statistical Gravitational Waveforms: What to Simulate Next?*
- **Chapter 3**: *Estimating the Contribution of Dyamical Ejecta in the Kilonova Associated with GW17017*
- **Chapter 4**: *A Search for Kilonovae in the Dark Energy Survey*
- **Chapter 5**: *A Search for Optical Emission from Binary-Black-Hole Merger GW170814 with the Dark Energy Camera*

# CHAPTER 2

# STATISTICAL GRAVITATIONAL WAVEFORM MODELS: WHAT TO SIMULATE NEXT?

## 2.1 Summary

To characterize the physical properties of systems we have detected through gravitational waves, we forward model the gravitational-wave signal for different source configurations and apply Equation 1.6 to sample posterior distributions on the source properties. The gravitational-wave signal models are in principle uniquely determined by general relativity and the magnetohydrodynamics of extreme matter, but computing these models for many configurations is prohibitively computationally expensive. To circumvent this, faster approximate models are constructed for use in parameter estimation pipelines. These "approximants" stitch together different methods to build a model for the full inspiral, merger, and ringdown of a gravitational wave from a compact object merger. In this chapter, which is based on *Statistical Gravitational Waveforms: What to Simulate Next?*, my collaborators and I propose a method for directly interpolating expensive simulated gravitational waveforms.

We utilize an interpolation method called *Gaussian process regression* which places a posterior distribution on values of a function $f(x)$ given a few measured values of $f$ at known $x$'s. This enables one to interpolate $f$ and characterize the uncertainty in the interpolation. Consequently, these uncertainties can be propagated through parameter estimation pipelines to guard against biases in the parameter posterior distributions due to imperfect interpolation. Additionally, the uncertainties from the interpolation can guide choices of the parameters of new expensive gravitational waveform simulations in order to minimize the Gaussian process regression model's uncertainty.

This chapter presents the overall method for constructing Gaussian process regression models of gravitational waves. We evaluate the efficacy of these types of models by apply-

ing them to a limited portion of the black-hole-binary parameter space. We find that this technique is capable of generating waveforms that are fast yet accurate to the level of interest for the advanced interferometer era. Our formalism also sets up a bridge between the construction of waveforms and the choices made for initial conditions of expensive numerical relativity simulations.

## 2.2 Introduction

The advent of gravitational wave (GW) detections has reinforced the need for accurate GW waveform models. The discovery and parameter estimates of the first LIGO detections were based on matched-filtering techniques which compared the data with predicted GW waveforms (Abbott et al., 2016c,a; Abbott et al., 2017a). Parameter estimation (PE) of these events has already revealed key astrophysical insights (Abbott *et al.*, 2016c,d). These inferences depend crucially on robust signal detection and parameter estimation, which in turn depend on accurate signal modeling.

Ideally, PE studies with matched-filtering analyses would directly use solutions to Einstein's equations for comparison to observed strain data. At present the most accurate solutions to Einstein's equations for compact binary coalescences (CBCs) come from numerical relativity (NR). Some of these NR simulations have been compared directly to GW150914 and GW151226, with excellent agreement (Lovelace *et al.*, 2016; Abbott et al., 2016a; Abbott *et al.*, 2016). While numerical relativity solutions are accurate, they can take weeks to months to compute (Lovelace *et al.*, 2016), and thus are prohibitively computationally expensive for use in PE.

For binary black holes, PE requires waveform models to cover a 7-d parameter space. For circularized binary black holes, there are 15 source parameters which need to be estimated, eight of which are intrinsic to the binary (each black hole has a 3-d spin vector and a mass). Circularized binary waveform models need only seven intrinsic parameters: the binary's mass ratio and the two spin vectors. This is because the system's total mass sets the frequency

scale for the GW; twice the total mass implies half the gravitational-wave frequency. Since each NR simulation can take many weeks to run, densely filling the 7-d mass-ratio + spin parameter space with NR waveforms for matched-filter analyses is inconceivable with current resources.

To circumvent the computational expense of NR simulations, approximate waveforms ("approximants", e.g. (Husa *et al.*, 2016; Khan *et al.*, 2016a; Taracchini *et al.*, 2014; Babak *et al.*, 2017; Hannam *et al.*, 2014; Bohé *et al.*, 2017)) are instead used for PE (Abbott *et al.*, 2016c). These approximants vary in accuracy, computational expense, source parameter domain, and approximation method, which can limit the potential breadth of PE searches or result in biased parameter estimates (Abbott *et al.*, 2017b). The SEOBNRv3 approximant, for example, is the most complete in that it models a 7-d intrinsic parameter space for CBCs in the effective one-body framework, but it has only been calibrated to a few non-precessing NR simulations and requires significant computational resources to generate waveforms (Pan *et al.*, 2014).

Data-driven models exist as well, such as NR surrogates, which fit to existing NR simulations to facilitate interpolation of waveforms at new parameter points (Blackman *et al.*, 2015, 2017). As computational resources for generating NR waveforms increase, data-driven models can more readily be used for PE as well as calibration and testing of approximate models (e.g. (O'Shaughnessy *et al.*, 2017)). Although recent work has shown that PE studies can be done directly with NR simulations (Lange *et al.*, 2017), robust surrogate and ROM models will still be highly valuable for an array of tasks beyond PE, such as calculations of coalescence rates and tests of General Relativity.

Here we propose a method to obtain reduced-order-model (ROM) waveforms and "interpolation" uncertainties using only a training set of simulations at a small number of points in parameter space. The impetus for using ROMs, which are concise representations of waveforms, is that our method interpolates between waveforms "observed" in simulations and hence benefits computationally from reduced waveform dimensionality. That is, if a time-

series waveform with $N$ points can instead be represented by a list of $m < N$ features, only $m$, rather than $N$, evaluations are required to produce a gravitational waveform with source parameters $\vec{\lambda}$.

For interpolation, we employ a technique known as Gaussian process regression (GPR). The advantage of a Gaussian process (GP) method is that it is *statistical* and fast. We describe our basic method in §2.3, which largely replicates the ROM work of (Pürrer, 2014), but instead of using spline interpolation to provide point predictions we use GPR to provide a statistical interpolation with uncertainties. The primary application of our method would be to use NR simulations to train a flexible, NR-driven GPR waveform model with quantified uncertainties. Although the end goal is to use NR simulations for training, in this paper we only present a proof of concept. Rather than NR waveforms, we use approximant waveforms from IMRPhenomD (Khan *et al.*, 2016b) to train and cross-validate, since they can be generated quickly and for a wide range of parameter values. §5.5 details the results of cross-validating GPR models trained on a small number of IMRPhenomD waveforms with the IMRPhenomD waveforms themselves. We find that the GPR model, trained on just a small subsample of IMRPhenomD waveforms, is able to reproduce IMRPhenomD waveforms to excellent accuracy. Future work will implement an NR training set in the GPR model.

One major advantage of the use of GPR methods is that they naturally provide estimates of the errors in the resulting waveforms. Current waveform approximants used in LIGO PE are implicitly assumed to be perfect, although multiple approximants are used to assess systematic errors. A more refined analysis, leading to improved PE results, would incorporate errors in the waveform approximants, especially as a function of location in parameter space. The GPR methods naturally provide a statistically consistent estimate of these errors. (Moore and Gair, 2014) and (Moore *et al.*, 2016) proposed an improvement to PE in which GPR is used to infer the systematic error on an approximant over the parameter space. They train their GP on the systematic errors between approximants and simulations for parameters where simulations exist. The systematic errors inferred from GPR are then

propagated to the likelihood function used in PE. A proof of concept of this systematic error interpolation method has only been done in 1-d, and it requires an existing approximant off of which to interpolate the errors. In contrast, our method requires no other approximants. Additionally, we consider both 1- and 2-d in this work, though extending to more dimensions is straightforward in principle.

A further advantage of GPR methods and the associated waveform error estimates is that these can be used to optimize the placement of new simulations which can be added to the training set. In §2.5 we describe a method for estimating where in parameter space the GPR model has high error, based only on the GPR uncertainties. This estimate is a natural metric for "greedily" deciding where new simulations are needed to minimize GPR model errors. In effect, our method will tell you the "optimal" place in parameter space to run the next simulation. An iterative procedure of GPR building and NR simulation leads to an efficient training set and GPR interpolation based on fewer training points than a naive regular grid. We outline that iterative procedure here:

1. Use existing (NR) simulations to build a GPR model with uncertainties.

2. Use GPR uncertainties to estimate where in parameter space errors in the GPR model are highest.

3. Generate new simulations at parameter values with high estimated GPR error and rebuild a GPR model with the augmented training set that includes the new simulations.

4. Repeat.

In §5.6 we discuss alternate choices that could be made with respect to GPR modeling of GWs, as the ideas presented herein serve more as a framework for GPR GW models than as an immutable, specific method. We also consider the evaluation time of GPR models, which vary in speed depending on the number of training waveforms and ROM coefficients. We conclude in §5.7.

## 2.3 Method

### *2.3.1 Outline for building GPR models*

To minimize the computational expense of waveform interpolation at source parameters $\vec{\lambda}$, we represent the waveforms with a ROM. We utilize the method described in Pürrer (2014) (hereafter P14), but other choices for building ROMs could also be made (e.g. (Field *et al.*, 2014)). The steps we take to create a GPR waveform interpolation model based on simulations are:

1. Simulate $n_{\rm train}$ frequency-domain waveforms at $\vec{\lambda} = \{\vec{\lambda}_j\}_{j=1}^{n_{\rm train}}$ to cover the parameter space of interest for use as the training set.

2. Project waveforms into a reduced-order basis so that a waveform with parameters $\vec{\lambda}$ is described by a list of coefficients $\{c_i(\vec{\lambda})\}_{i=1}^{m}$, where $m = \dim(\text{basis})$.

3. Regularize each $c_i$ function by subtracting a linear fit of the training values $c_i(\{\vec{\lambda}_j\})$ and normalizing. Define $\tilde{c}_i \equiv \text{Regularize}(c_i)$.

4. Assume each $\tilde{c}_i(\vec{\lambda})$ is a realization of a Gaussian process, and use $\tilde{c}_i(\{\vec{\lambda}_j\})$ as the training set to regress $\tilde{c}_i(\vec{\lambda}) \approx \tilde{c}_{i,\rm GP}(\vec{\lambda})$.

5. Transform $\{\tilde{c}_i(\vec{\lambda})\}$ and their uncertainties to the domain of interest (e.g. frequency domain for LIGO PE) by applying the reverse of the regularizations in step 3 and then projecting the coefficients out of the ROM basis back to the time or frequency domain. The specific operations to go from the ROM to the time or frequency domain depend on which ROM is used.

We discuss all of these steps in detail below.

### *2.3.2 Waveform Generation and Representation*

We generate fixed-chirp-mass, frequency-domain IMRPhenomD waveforms (with a starting frequency of 20 Hz, maximum frequency of 1520 Hz, and frequency resolution of 0.01 Hz) at various source parameter values, and interpolate the amplitudes and phases separately onto sparse frequency grids as in §5.1 of P14[1]. For simplicity, only the $h_+$ components of the waveforms are considered here, but the methods presented herein can be identically applied to $h_\times$, or $h_+$ can be used to calculate $h_\times$ as in equations 6.14 and 6.15 in P14. For the sparse frequency grid, we choose an amplitude grid spacing at frequency $f$ of $\Delta_A = 0.1f$ and a phase grid spacing of $\Delta_\Phi = 0.3f^{4/3}$; P14 found that these grid spacings keep a constant spline interpolation error at all frequencies. The interpolated amplitudes and phases are then packed into the columns of matrices $\mathcal{T}_A$, $\mathcal{T}_\Phi$, and a SVD is performed on each matrix (see §6 of P14):

$$\begin{aligned} \mathcal{T}_A &= V_A \Sigma_A U_A^\top, \\ \mathcal{T}_\Phi &= V_\Phi \Sigma_\Phi U_\Phi^\top. \end{aligned} \tag{2.1}$$

P14 truncates the $V$ matrices to reduce the dimensionality of the ROM, but for simplicity we instead directly compute projection coefficients $c(\tau)$ for each input amplitude or phase $\tau$ (amplitude and phase subscripts dropped for ease of notation):

$$c(\tau) = V^\top \tau. \tag{2.2}$$

Each element of $c$ is a function of $\tau$ and hence a function of the input parameters $\vec{\lambda}$. The following sections describe how the elements of $c$ are interpolated across the parameter space to extract new waveforms with uncertainties.

1. We interpolate fixed-chirp-mass waveforms, whereas P14 interpolated fixed-total-mass waveforms. This results in interpolations which are over the same frequency range rather than interpolations over the same geometric frequency range.

### *2.3.3 Gaussian Process Regression*

GPR is a tool for emulating (i.e., statistically inferring) the behavior of functions of continuous variables. It is commonly used to predict the output of simulations which are too expensive to run for many parameter values (e.g. (Jones *et al.*, 1998; Higdon *et al.*, 2004)). In the case of GWs, we wish to infer new strain waveforms from existing NR simulations. The basic assumption in GPR is that any finite subset of values of a process $f(\vec{x})$ have a joint Gaussian distribution and thus can be described by a mean function $\vec{\mu}$ and covariance function $\mathbf{k}(\vec{x}, \vec{x}')$:

$$f(\vec{x}) \sim \mathcal{GP}(\vec{\mu}, \mathbf{k}(\vec{x}, \vec{x}')). \tag{2.3}$$

Given a list of known values of a function $\mathbf{f} = \{f_1(\vec{x}_1), f_2(\vec{x}_2), ...\}$ at $n_{\rm train}$ training points $X = \{\vec{x}_1, \vec{x}_2, ...\}$, we can calculate the probability distribution for $\mathbf{f}_* = \{f_{1*}(\vec{x}_{1*}), f_{2*}(\vec{x}_{2*}), ...\}$ at new points $X_* = \{\vec{x}_{1*}, \vec{x}_{2*}, ...\}$, using the definition of a Gaussian process: the prediction and the known values have a joint Gaussian distribution. If the mean of the GP prior in Equation 2.3 is zero[2] then:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right), \tag{2.4}$$

where $\mathcal{N}(\mathbf{m}, \mathbf{K})$ is a multivariate normal distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{K}$. $K(X,X)$, $K(X,X_*)$, $K(X_*,X)$, and $K(X_*,X_*)$ are the matrices of covariances between pairs of training and prediction points, the elements of which are calculated using the covariance function $\mathbf{k}(\vec{x}, \vec{x}')$. The conditional probability distribution of $\mathbf{f}_*$ given $\mathbf{f}$ is itself

2. We assume the prior mean is zero, as we only interpolate functions which are de-meaned, i.e. which have their mean subtracted off.

Gaussian (see e.g. (Rasmussen and Williams, 2005)):

$$p(\mathbf{f}_*|\mathbf{f}) = \mathcal{N}\Big( K(X_*, X)K(X, X)^{-1}\mathbf{f}, \\ K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*)\Big). \tag{2.5}$$

With covariances and a training set, $\mathbf{f}$, one can calculate the conditional mean and covariance of $\mathbf{f}_*$. In GPR, the entries of the covariance matrices are computed using a covariance function, or kernel, which is specified by the user and depends on the application. The kernels are symmetric and hence only depend on the distance between points $r = |x - x'|$. Here we primarily consider the squared-exponential kernel, but we discuss the choice of covariance function in §5.6. A squared-exponential covariance constrains the process to be infinitely mean-square differentiable[3] and takes the form of a Gaussian:

$$\mathbf{k}_{SE}(\vec{x}_1, \vec{x}_2) = \mathbf{k}_{SE}(r = |\vec{x}_1 - \vec{x}_2|) = \sigma^2 \exp\left(-\frac{1}{2}r^2/l^2\right). \tag{2.6}$$

where $\sigma$ and $l$ are *hyperparameters* of the process and parameterize the signal variance and length scale, respectively. The hyperparameters of this kernel can be fixed a priori, but are typically chosen to maximize the *hyperlikelihood*, the likelihood of the training data under the GP prior (Equation 2.3). Here we instead maximize the *hyperposterior* which incorporates prior distributions on the hyperparameters:

$$\text{hyperposterior} \propto \text{hyperlikelihood} \times \text{hyperprior}. \tag{2.7}$$

The priors on the hyperparameters, or *hyperpriors*, are discussed further in §5.5.

In addition to selection of hyperparameters, one must choose small values called "nuggets" (Andrianakis and Challenor, 2012) to add to the diagonal of the training points covariance matrix $K(X, X)$ to ensure numerical stability when computing the conditional covariance.

3. See §4.1.1 in Rasmussen and Williams 2005 for the definition of mean-square differentiability.

Computation of the conditional covariance can result in a non-positive-semi-definite matrix due to precision errors when the prior covariance matrix has large values off-diagonal. These nuggets are described further in §5.5.

### *2.3.4 Implementation of GPR-based models*

We now return our focus to the training set projection coefficients $c_i(\vec{\lambda}_j)$ at points $\vec{\lambda}_j$ in the input parameter space. To reconstruct waveforms, P14 interpolates the training projection coefficients over the parameter space using a tensor spline interpolation. We instead use GPR in order to account for uncertainties in the interpolation. Rather than directly interpolate $c_i$, we interpolate *regularized* coefficients $\tilde{c}_i$ to avoid secular changes of the coefficient values over the parameter space and allow modeling with a zero-mean GP. The training values $c_i(\{\vec{\lambda}_j\}_{j=1}^{n_{\rm train}})$ are regularized by first subtracting a linear fit over the parameter space, then normalizing the residual variance to 1 and removing the mean:

$$\begin{aligned} \Delta c_i(\{\vec{\lambda}_j\}) &\equiv c_i(\{\vec{\lambda}_j\}) - \text{linfit}\left(c_i(\{\vec{\lambda}_j\})\right) \\ \tilde{c}_i(\{\vec{\lambda}_j\}) &\equiv \frac{\Delta c_i(\{\vec{\lambda}_j\}) - \text{mean}\left(\Delta c_i(\{\vec{\lambda}_j\})\right)}{\text{std}\left(\Delta c_i(\{\vec{\lambda}_j\})\right)}. \end{aligned} \tag{2.8}$$

We then assume that each $\tilde{c}_i$ is a Gaussian process, and use $\tilde{c}_i(\{\vec{\lambda}_j\})$ as a training set to regress $\tilde{c}_i(\{\vec{\lambda}\})$. Thus for any point in parameter space $\vec{\lambda}$ we can predict $\tilde{c}_i(\vec{\lambda})$ and marginal uncertainties $\delta\tilde{c}_i(\vec{\lambda})$ which can then be transformed back to amplitude/phase by applying the reverse of the transformations used to generate $\tilde{c}_i(\{\vec{\lambda}_j\})$. Assuming $c_i$ is uncorrelated with $c_j$ unless $i = j$, the mean amplitude $A(F_k, \vec{\lambda})$ in the $k$-th frequency bin and covariance

matrix $\Sigma_{kl}(\vec{\lambda})$ are given by[4]:

$$\begin{aligned} A(F_k, \vec{\lambda}) &= \sum_i V^A_{ki} c^A_{i,\mathrm{GP}}(\vec{\lambda}), \\ \Sigma^A_{kl}(\vec{\lambda}) &= \sum_i V^A_{ki} \delta c^A_{i,\mathrm{GP}}(\vec{\lambda}) V^{A\top}_{il}. \end{aligned} \tag{2.9}$$

The phase means and covariances, $\Phi(F_k, \vec{\lambda})$ and $\Sigma^{\Phi}_{kl}(\vec{\lambda})$, can be reconstructed similarly with the corresponding phase projection coefficients.

To assess the accuracy of our GPR model, we compare reconstructed frequency waveform means from GPR $h_{\mathrm{GPR}}(F_k, \vec{\lambda})$ with the waveforms from IMRPhenomD $h(F_k, \vec{\lambda})$ using the *mismatch* function. The mismatch between two frequency domain waveforms $h_1$ and $h_2$ is defined as:

$$\begin{aligned} &\mathrm{mismatch}(h_1, h_2) = \\ &1 - \frac{4}{||h_1||||h_2||} \\ &\times \Re \left( \int_{f_{\min}}^{f_{\max}} \frac{h_1(f) h_2^*(f)}{S_f} \mathrm{d}f \right), \end{aligned} \tag{2.10}$$

where

$$||h|| = 4\Re \left( \int_{f_{\min}}^{f_{\max}} \frac{h(f) h(f)^*}{S_f} \mathrm{d}f \right). \tag{2.11}$$

$f_{\min}$ and $f_{\max}$ are the minimum and maximum frequencies at which the two waveforms are compared, respectively, and $S_f$ is the noise power spectral density of a GW detector. In this paper, we use the aLIGO O1 noise curve `2015-10-01_H1_O1_Sensitivity_strain_asd.txt`[5].

## 2.4 Results

We now implement GPR-based models for three sets of parameter spaces using the `GaussianProcessRegressor` module from `scikit-learn`(Pedregosa *et al.*, 2011):

4. The correlations between $c_i$'s are outside the scope of this paper, and will be considered in future work.

5. This noise curve can be found at https://dcc.ligo.org/LIGO-G1501223/public.

1. constant zero spin

   mass ratio $q \in [1, 6]$

2. equal-and-aligned spin $\chi_1 = \chi_2 \in [-1, 1]$

   constant $q = 1$

3. equal-and-aligned spin $\chi_1 = \chi_2 \in [-0.5, 0.5]$

   $q \in [1, 3]$

All parameter spaces use waveforms with a constant chirp mass $M_c = 20M_\odot$, distance $D = 1$ Mpc, inclination angle $i = 0$, and starting frequency of 20 Hz.

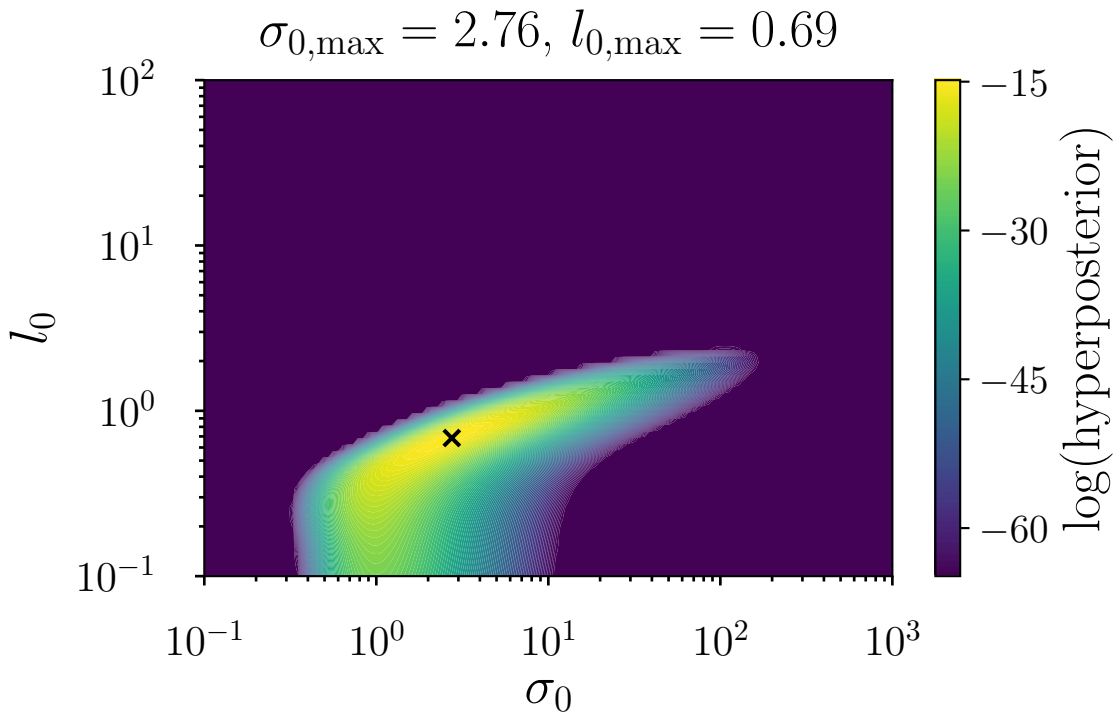


Figure 2.1: Hyperposterior for the first regularized amplitude coefficient $\tilde{c}_0^A$ as a function of kernel covariance scale $\sigma_0$ and length scale $l_0$. The hyperparameter values with largest hyperposterior, $\sigma_{0,\max}$ and $l_{0,\max}$, are marked with an X on the plot and printed at the top.

### *2.4.1 1-d GPR in Mass Ratio*

Beginning with Parameter Space 1, we generate $n_{\rm train} = 15$ equally spaced IMRPhenomD waveforms from $q = 1$ to $q = 6$ as a proxy for an NR training set and compute their amplitude and phase projection coefficients as described in §2.3.2. The $i$-th coefficient is then de-trended by removing a linear fit of the 15 training coefficient values $c_i(\{\vec{\lambda}_j\})$. The de-trended $c_i(\{\vec{\lambda}_j\})$ are then normalized by their standard deviation and their mean is removed. We refer to the de-trended, de-meaned, normalized coefficients as $\tilde{c}_i$ and treat each one as a GP:

$$\tilde{c}_i \sim \mathcal{GP}(\vec{0}, \mathbf{k}_i(q, q')). \tag{2.12}$$

The $\tilde{c}_i$ can be trivially transformed back to $c_i$ by reapplying the mean, standard deviation, and linear fit. We take the covariance function for the $i$-th coefficient, $\mathbf{k}_i(q, q')$, to be a squared exponential with hyperparameters $\sigma_i$ and $l_i$.

To ensure numerical stability in the GPR conditional covariance matrix calculation, we add a nugget to the input covariances for each training value. We assume a constant relative error on the training waveform amplitude of $10^{-4}$ at each frequency and transform these errors to errors in the amplitude coefficients, which are used as the kernel nugget. For the training phases, the error at each frequency is kept below the nominal LIGO phase measurement uncertainty of $\sim 0.1$ radians by assuming a constant error of 0.01 radians at each phase value on the sparse frequency grid. These errors are projected to the coefficient errors analogously to the amplitude error case. The nugget levels here are chosen for numerical stability and adequate accuracy, but also roughly correspond to the resolution errors found in NR simulation studies (Lovelace *et al.*, 2016). In principle, the NR errors as a function of source parameters could be incorporated into the nugget values to fully account for the resolutions of different simulations.

To optimize the kernel hyperparameters for each coefficient, we use the `scipy.optimize` implementation (Jones *et al.*, 01 ) of the Broyden-Fletcher-Goldfarb-Shanno algorithm

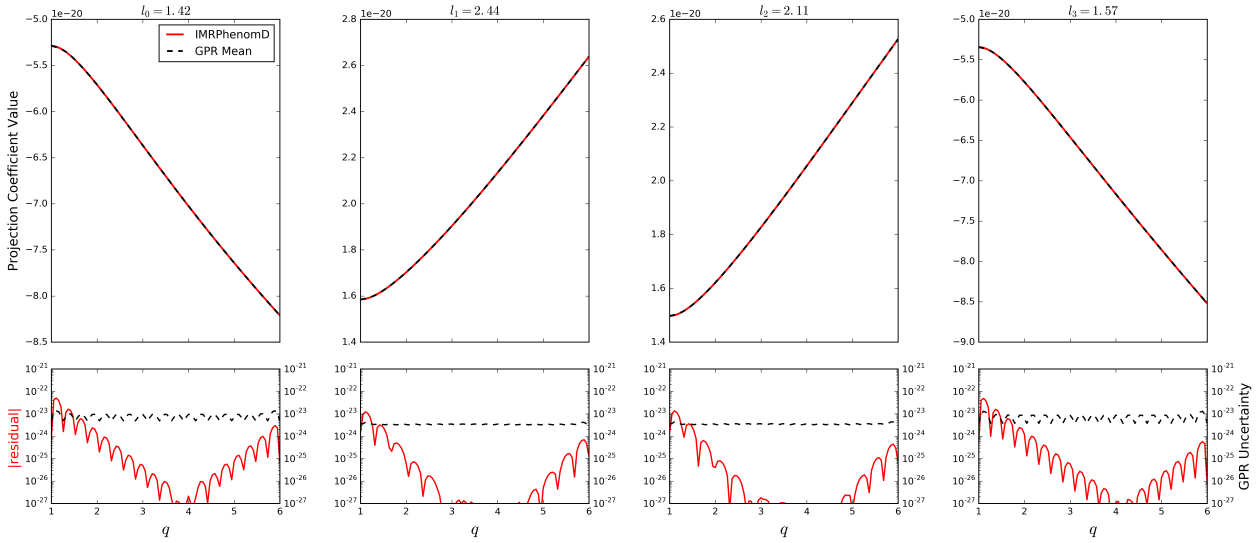


Figure 2.2: First two amplitude coefficients for the IMRPhenomD waveforms and GPR-based waveforms with their residuals. *Top panels*: IMRPhenomD coefficient values (red) and GPR mean coefficient values (black, dashed) for the first two amplitude coefficients. The optimized length scales are shown above the top panels. Although the remaining coefficients are not shown here, they have similar morphologies. *Bottom panels*: the fractional residuals $(|c_i^A - c_{i,\mathrm{GP}}^A|)/c_i^A$ (solid) and the GPR fractional $1\sigma$ uncertainties $\delta c_{i,\mathrm{GP}}^A/c_i^A$ (dashed).

`fmin_l_bfgs_b` (Zhu *et al.*, 1997) to maximize the hyperposterior. We apply log-normal hyperpriors on $\sigma_i$ and $l_i$:

$$\begin{aligned} \log_{10}(\sigma_i) &\sim \mathcal{N}(0, 0.5) \\ \log_{10}(l_i) &\sim \mathcal{N}\left( \log_{10}\left(\frac{1}{2}\mathrm{width}(\{q_j\}_1^{n_{\mathrm{train}}})\right), 1\right). \end{aligned} \tag{2.13}$$

Since the regularized coefficient functions being interpolated have been normalized by their standard deviation, we expect that $\sigma_i \sim 1$, motivating the hyperprior on $\sigma_i$ above. We have less information about the length scales a priori, but we know that they should not be much shorter than the distance between the closest training points, nor should they be much larger than the width of the parameter space. As such, the normal distribution on $\log_{10}(l_i)$ is chosen to have a standard deviation of 1 (i.e. 1 order of magnitude) and peak at half the width of the parameter space spanned by the training set. Figure 2.1 shows the hyperposterior surface for the first amplitude coefficient as a function of the hyperparameters

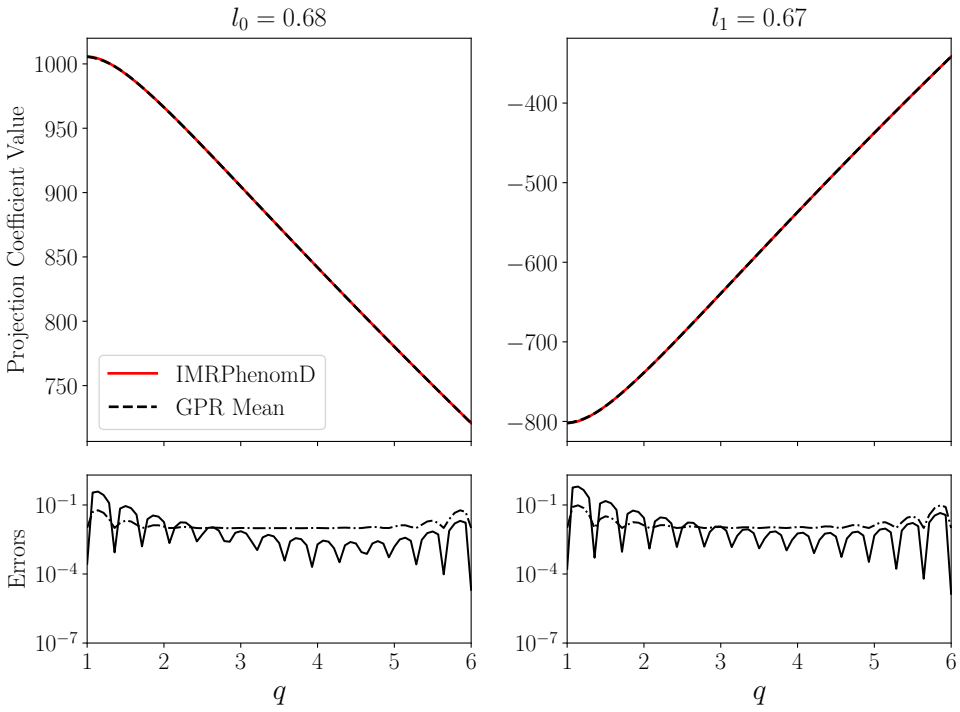


Figure 2.3: First two phase coefficients for the IMRPhenomD waveforms and GPR-based waveforms with their residuals. *Top panels*: IMRPhenomD coefficient values (red) and GPR mean coefficient values (black, dashed) for the first two phase coefficients. The optimized length scales are shown above the top panels. Although the remaining coefficients are not shown here, they have similar morphologies. *Bottom panels*: the residuals $|c_i^\Phi - c_{i,\mathrm{GP}}^\Phi|$ (solid) and the GPR $1\sigma$ uncertainties $\delta c_{i,\mathrm{GP}}^\Phi$ (dashed).

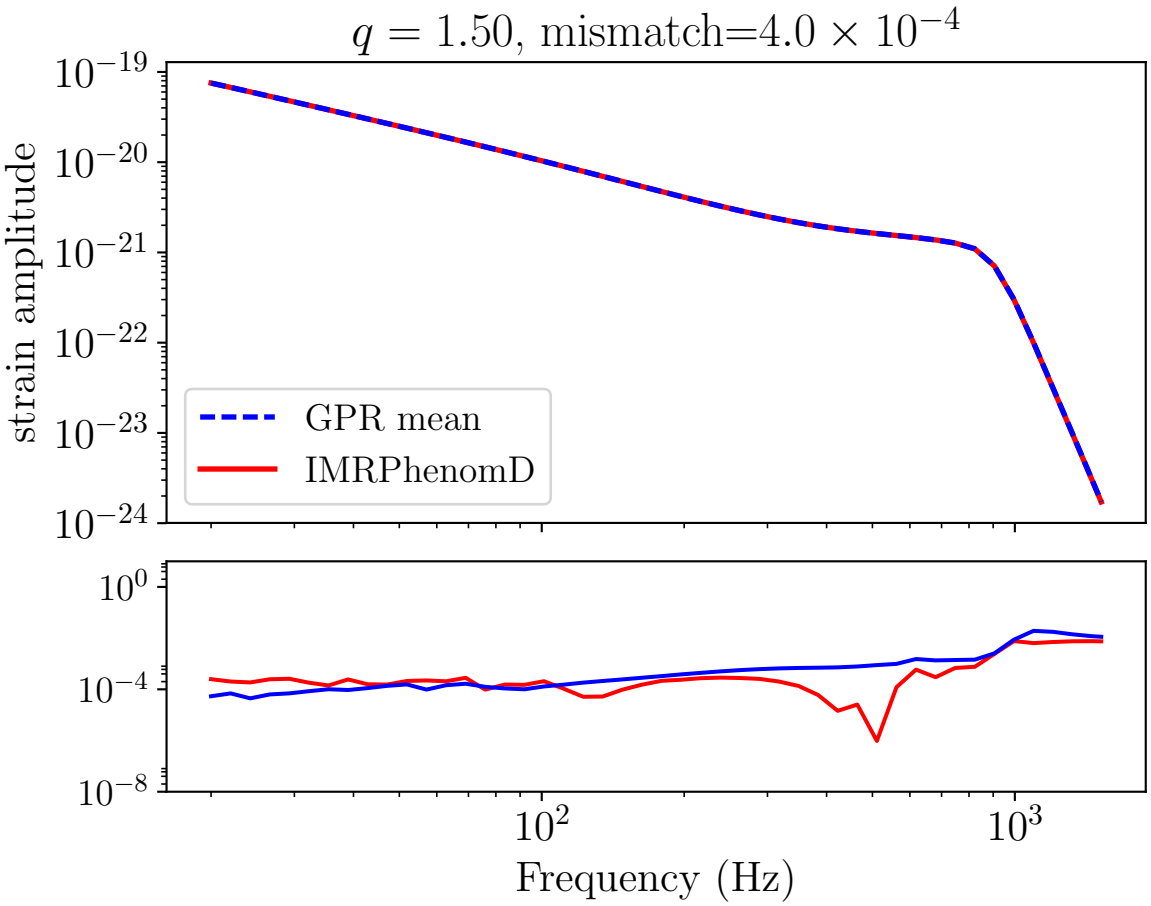


Figure 2.4: Example reconstructed GPR amplitude function. *Top:* The interpolated GPR mean amplitude vs. frequency is shown in blue and the IMRPhenomD amplitude is overlaid in red. *Bottom:* The IMR-GPR-mean residual amplitude is shown in red, and the GPR $1\sigma$ uncertainty is shown in blue as a function of frequency. Both errors are normalized to the IMRPhenomD amplitude.

using the squared-exponential kernel. The hyperposteriors for other coefficients are similar in morphology to those shown here.

With optimized hyperparameters, the GP is used to interpolate each regularized coefficient on a grid five times finer than the training grid. With the GPR model of the $\tilde{c}_i$'s, we can calculate each $c_i$ and transform them back to amplitudes and phases on the sparse frequency grid. Figures 2.2 and 2.3 show the values of the first two amplitude and phase $c_i$'s, respectively, as a function of $q$ from IMRPhenomD and from the GPR-based model. Also shown are the fit residuals, the GPR $1\sigma$ uncertainties, and the optimized length scale

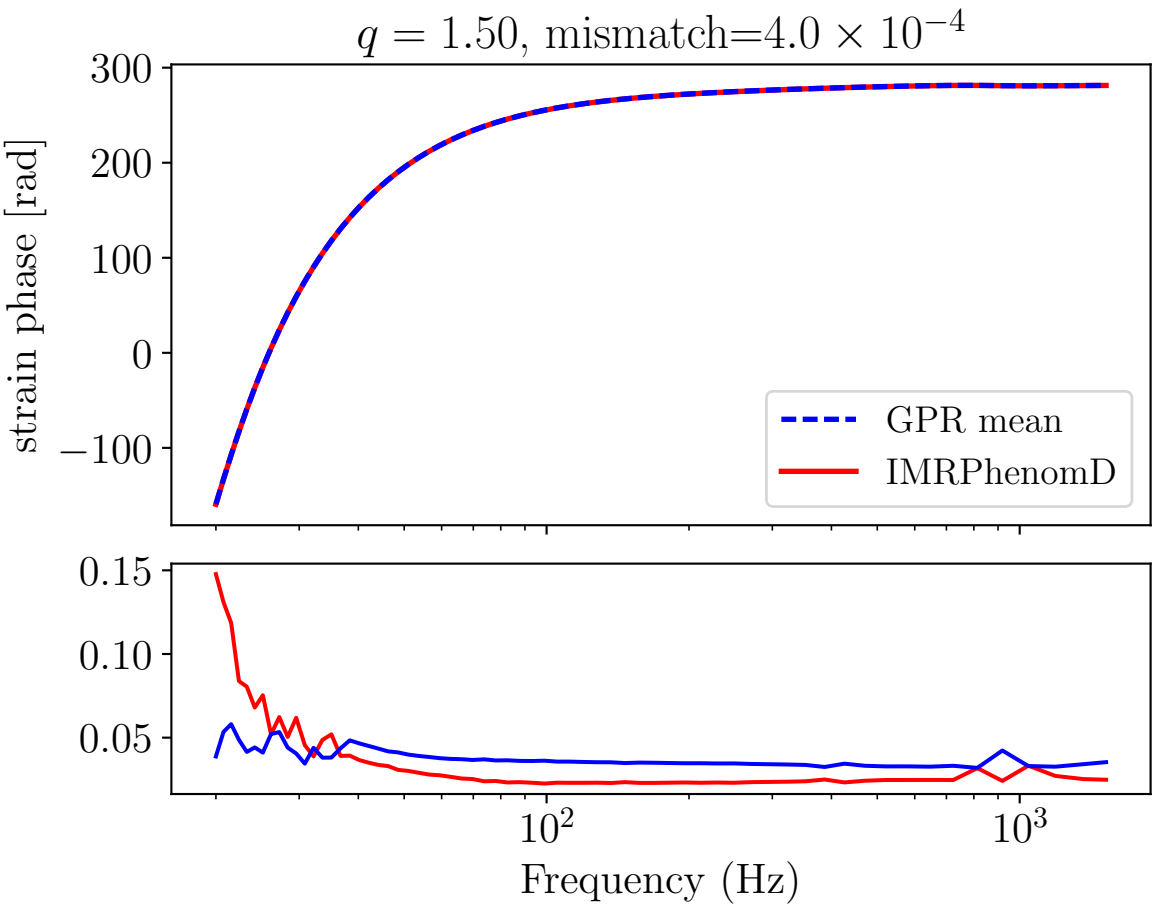


Figure 2.5: Example reconstructed GPR phase function. *Top:* The interpolated GPR mean phase vs. frequency is shown in blue and the IMRPhenomD phase is overlaid in red. *Bottom:* The absolute value of the IMR-GPR residual phase is shown in red, and the GPR $1\sigma$ uncertainty is shown in blue as a function of frequency.

hyperparameter for each coefficient function's GP. Although the GPR errors do not perfectly match the residuals across the parameter space, they are indicative of the maximum error level and of the fact that the errors are largest on the edges of the space. The interpolated coefficients and their uncertainties are then propagated back to amplitudes and phases on the sparse frequency grid.

One such example of a GPR-interpolated waveform is shown in Figures 2.4 and 2.5, which show the amplitude and phase functions, respectively. Notably, the fractional amplitude error between the GPR model and IMRPhenomD waveform is largest at high frequencies, because small errors in the amplitude coefficients combine when projected back to the sparse frequency grid domain. Nonetheless, the GPR mean agrees well with the IMRPhenomD model, and the GPR uncertainties give a reasonable indication of the true error levels at different frequency bins. Since PE is done in the frequency domain, these errors do not need to be propagated to the time domain, although frequency-domain waveforms can be sampled to create a distribution of time-domain waveforms if required.

Figure 2.6 shows the mismatch between the GPR mean waveform and the IMR waveform at various values of $q$. The mismatch increases near the boundaries of the space where there are fewer nearby training points (training points shown in dashed black vertical lines). Indeed, the mismatches and residuals suggest that more training points (i.e. simulations) are needed towards the edge of the space of interest. Increasing the number of training points typically lowers the mismatches, although it depends on the nugget and hyperparameters used in the GP. The relative placement of the simulations is also of interest, which we discuss in §2.5.

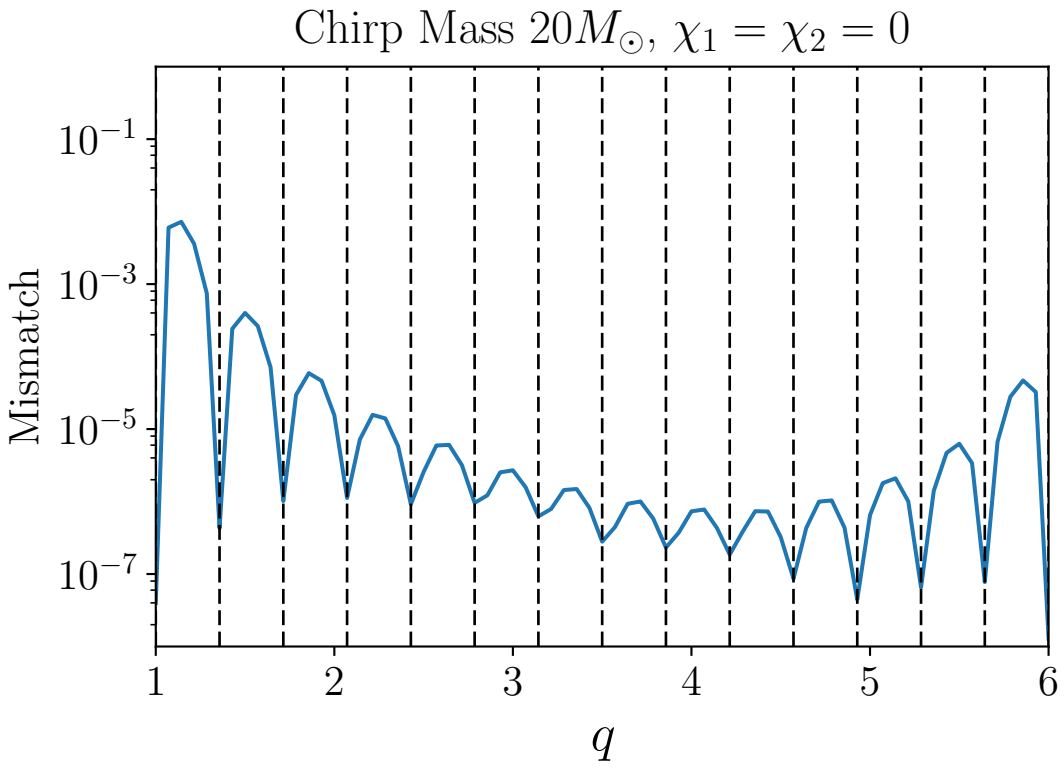


Figure 2.6: Mismatch between the IMRPhenomD waveform and the GPR mean waveform for different mass ratios assuming a constant chirp mass and zero spin. The dashed, vertical lines show the IMRPhenomD training waveform locations used to build the GPR model.

### 2.4.2 1-d GPR in equal-and-aligned spin

Following an analogous method to that presented in §2.4.1, we generate 12 IMRPhenomD waveforms spanning equally-spaced equal-and-aligned spin[6] values from $\chi = -1$ to $\chi = 1$. The chirp mass of $M_c = 20M_\odot$ and mass ratio $q = 1$ are held fixed. A GPR-based waveform model is built using these waveforms as a training set, again using a squared-exponential covariance function to model each $\tilde{c}_i$. The hyperpriors on the length scale and covariance scale are chosen in the same way as in §2.4.1.

The GPR model is evaluated on a grid five times finer than the training grid and is then compared to the waveforms predicted in IMRPhenomD via the mismatch function. Figure 2.7 shows the mismatch between the GPR mean waveform and IMRPhenomD for the 1-d space of equal-and-aligned spin $\chi$. Similarly to the case of mass-ratio space, the mismatch between the GPR model and IMRPhenomD is largest at the boundaries of the training set. Nevertheless, we are able to achieve low mismatches with just a few training points in this case.

### 2.4.3 2-d GPR in equal aligned spin and mass ratio

We again build a training set using IMRPhenomD waveforms as a proxy for NR simulations in order to train a GPR-based model, except here we vary two source parameters: The mass ratio and the value of the equal-and-aligned spins. The training waveforms are generated on a regular grid with 15 points in $q$ and 8 points in $\chi$ from $q \in [1, 3]$ and $\chi \in [-0.5, 0.5]$. The kernel used here is simply an overall covariance $\sigma$ times the product of two squared-exponential kernels, one for the $q$ dimension and one for the $\chi$ dimension. As such, there are now two length scale parameters, $l_q$ and $l_\chi$, as well as the overall covariance $\sigma$ which need to be fit for each coefficient function. Hyperpriors for the hyperparameters are chosen as in previous sections, and the hyperparameters are optimized via the hyperposterior. For

6. The spins of the inspiraling objects are assumed to be equal and aligned with the orbital angular momentum.

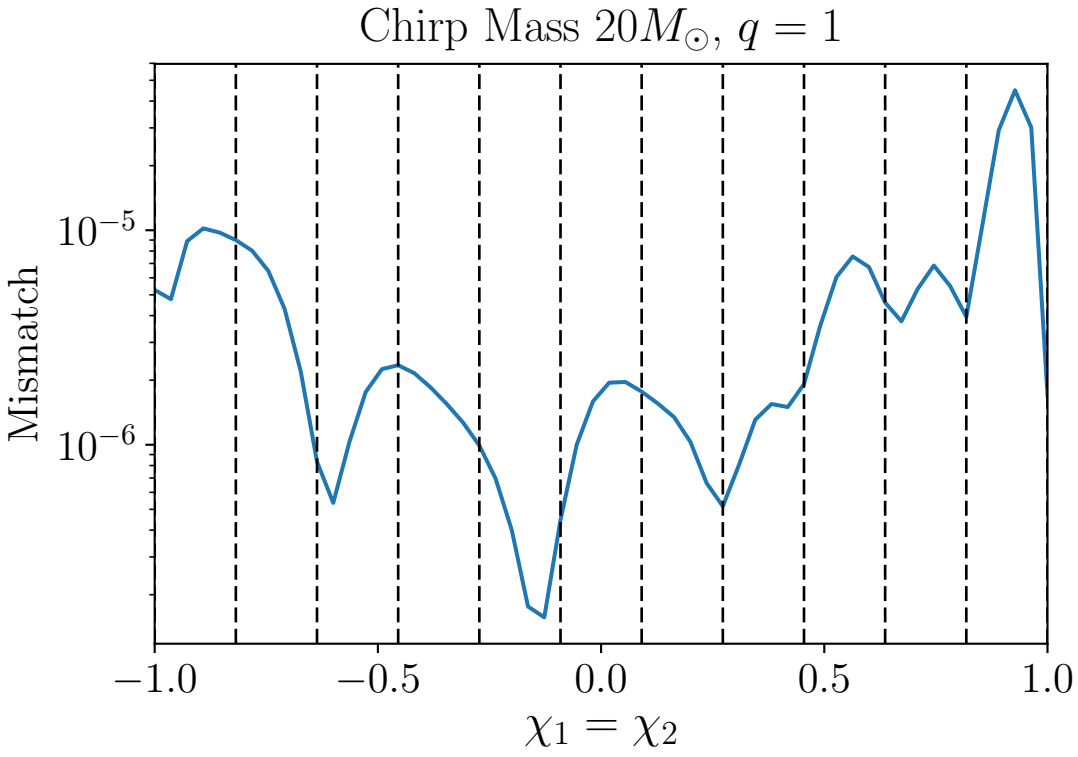


Figure 2.7: Mismatch between the IMRPhenomD waveform and the GPR mean waveform for different equal-and-aligned spin values assuming a constant chirp mass and equal masses ($q = 1$). The dashed, vertical lines show the IMRPhenomD training waveform locations used to build the GPR model.

computational savings, smaller windows in $q$ and $\chi$ are considered here than in the previous sections. Extending to the full parameter space would simply require more training points [7]. Another option would be to decompose the extended domain into smaller overlapping patches and build a GPR model for each patch (e.g. Figure 2 of (Pürrer, 2016)). Although a regular, equal-spacing grid leads to reasonable mismatches in this case, there is no a priori reason to use such a grid, and in practice the existing simulations will have non-regular placements throughout the parameter space.

Figure 2.8 shows the "true" values, interpolations, and residuals of the first amplitude SVD coefficient $c_0^A$ as a function of $q$ and $\chi$. The top left panel is a color map of the coefficient values from IMRPhenomD. The top middle and right panels show $c_{0,\mathrm{GP}}^A$ and $c_{0,\mathrm{spline}}^A$, which are the interpolations of $c_0^A$ with GPR and B-splines, respectively. The bottom panels show the fractional residuals of the interpolants and the estimated error from GPR. The residuals between IMRPhenomD and the GPR model (bottom left panel) for this amplitude coefficient are below the 0.1% level and are comparable to the predicted GPR uncertainties, providing evidence for the accuracy and precision of the GPR model. Comparing the bottom left and bottom right panels of Figure 2.8, the GPR mean is roughly as accurate at the B-spline interpolation for most coefficients, but the spline does not give any information about interpolation errors. The bottom middle panel shows the GPR-estimated fractional errors which give an estimate of the maximum true GPR-IMR residual.

Figure 2.9 is analogous to the mismatch plots shown in §2.4.1 and §2.4.2, but now in two dimensions. The black circles denote the IMRPhenomD training points, and the color map shows the mismatch between the IMRPhenomD waveform and the GPR mean waveform at each point. The mismatches calculated here are all at or below $4.3 \times 10^{-5}$.

7. The computational time to build a GPR given hyperparameters and $N$ training waveforms goes as $\sim O(N^3)$. Additional cost is incurred by the hyperposterior maximization.

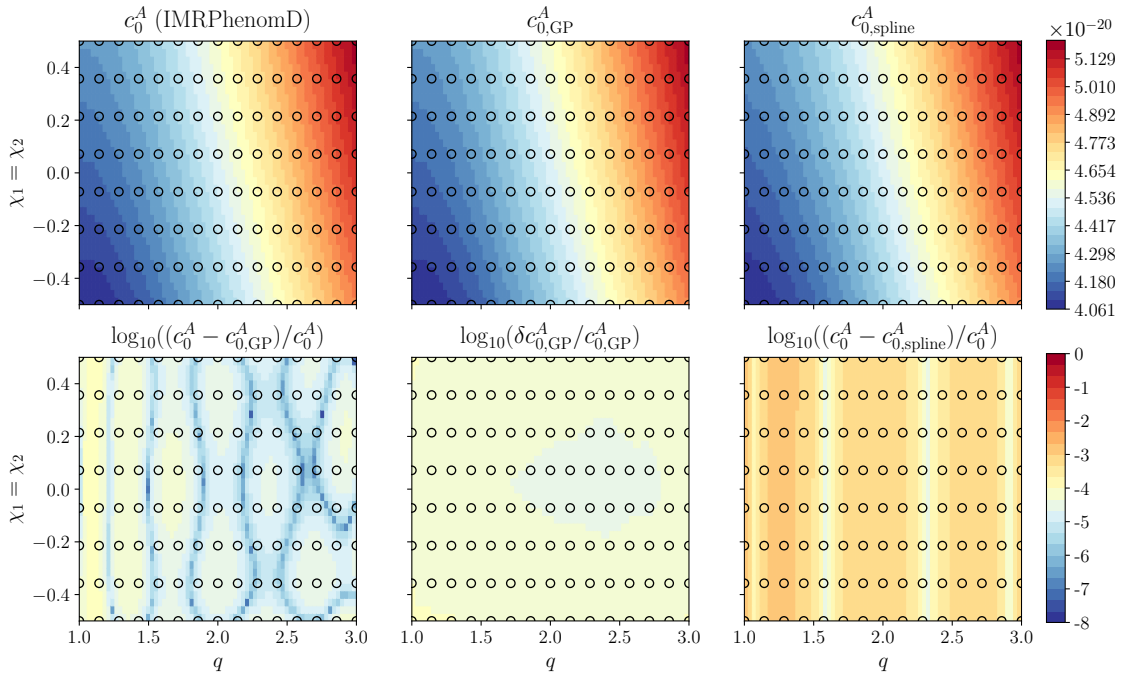


Figure 2.8: First amplitude coefficient $c_0^A$ as a function of $q$ and equal-and-aligned spin $\chi_1 = \chi_2$. Black circles show the training point locations. *Top left*: $c_0^A$ from the accurate model IMRPhenomD from which the training points are generated. *Top middle*: The GPR mean interpolation of $c_0^A$. *Top right*: The B-spline interpolation of $c_0^A$. *Bottom left*: The log of the fractional residual of $c_0^A$ between IMRPhenomD and the GPR mean. *Bottom middle*: the log of the fractional $1\sigma$ uncertainty on $c_0^A$ from the GPR. *Bottom right*: The log of the fractional residual of $c_0^A$ between IMRPhenomD and the B-spline.

## 2.5 Where should we run the next numerical relativity simulation?

Two natural questions arise from considering a GP model: 1) what is the error level of the GPR model for different GW source parameters, and 2) given a pre-existing set of training waveforms, where is the "optimal" placement of an additional training waveform? Thus far, we have evaluated the accuracy of our GPR models by computing the mismatch between the GPR model and the IMRPhenomD, but in practice such comparisons will not be possible, since the "true waveform" (i.e. NR simulation results) will be unknown everywhere other than at the existing simulation points. As such, the mismatch between the GPR mean and the true waveform cannot be used to determine the GPR error level nor can it be used as a

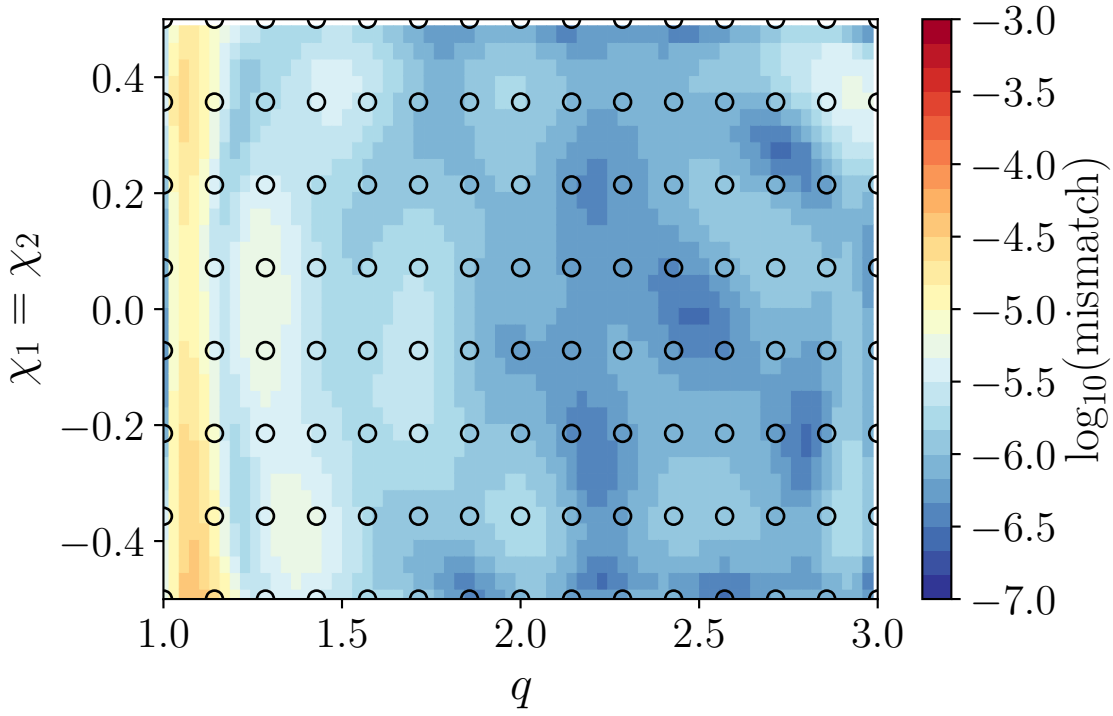


Figure 2.9: Mismatch between IMRPhenomD waveforms and GPR mean waveforms with a regularly-gridded training set. The black circles show the locations of training waveforms from IMRPhenomD used to train the GPR. There are $15 \times 8 = 120$ training points on this grid, and the maximum mismatch in the region is $4.3 \times 10^{-3}$.

parameter of the optimization function which selects new NR simulation parameters. Instead, we propose using the GPR posterior uncertainties to guide the overall error estimation and training set optimization. Here we present a simple metric for estimating GPR waveform errors and for choosing where in parameter space to add a new training waveform. The basic idea is to estimate the mismatch between a GPR and NR waveform with the same parameters based on the spread of GPR samples. Specifically, we estimate the GPR-IMR mismatch by computing the largest mismatch between $M$ GPR samples and the GPR mean. New accurate waveforms can be added where this estimated mismatch is large, analogously to the greedy algorithm 8.1 in (Rasmussen and Williams, 2005). We summarize this training point placement strategy in Algorithm 1.

**Algorithm 1** greedy training point placement

$\{\vec{\lambda}_j\} \leftarrow n_{\rm train}$ initial parameter values, $j \in [1, n_{\rm train}]$

$\{\vec{\lambda}^*_k\} \leftarrow$ fine interpolation grid, $k \in [1, n_{\rm interp}]$

**loop**

Calculate regularized coefficients $\tilde{c}_i(\{\vec{\lambda}_j\})$

$\tilde{c}_i(\vec{\lambda}^*_k) \sim {\rm GPR}(\tilde{c}_i(\{\vec{\lambda}_j\}))$

**for** $k \in [1, n_{\rm interp}]$ **do**

$m \leftarrow 0$, $O_k \leftarrow 0$

**for** $m \in [1, M]$ **do**

$O \leftarrow \texttt{mismatch}(h^{\rm mean}_{\rm GPR}(\vec{\lambda}^*_k), h^{\rm sample}_{\rm GPR}(\vec{\lambda}^*_k))$

$O_k \leftarrow \max(O \cup O_k)$

**end for**

**end for**

$\vec{\lambda}_{n_{\rm train}+1} \leftarrow \vec{\lambda}^*[{\rm argmax}_k(O_k)]$

$\{\vec{\lambda}_j\} \leftarrow \{\vec{\lambda}_j\} \cup \{\vec{\lambda}_{n_{\rm train}+1}\}$

$n_{\rm train} \leftarrow n_{\rm train} + 1$

**end loop**

First, a few training waveforms, preferably on the boundaries of the parameter space $P$, are used to seed a GPR model of the waveforms in $P$. GPR waveforms are interpolated on a fine grid in $P$, and at each grid point the maximum mismatch between the GPR waveform mean and $M$ GPR waveform samples is recorded (hereafter called $O_k$ for the $k$-th interpolation point). $O_k$ at each interpolation grid point is used as a proxy for the true mismatch between the GPR mean and NR in order to determine where to generate a new simulation. By adding a new simulation to the training set at the point with largest $O_k$, the greedy algorithm attempts to minimize error in locations in parameter space with the largest estimated error.

As a proof of concept, we apply a computationally simplified variant of the greedy algorithm to GPR interpolations in the same $q$–$\chi$ space as in §2.4.3. Rather than determining $O_k$ at every point on a dense interpolation grid, we instead partition the space into 100 equally-sized, rectangular domains and determine $O_k$ at a random point in each domain. These 100 $O_k$ values are then used to determine training point placement. Additionally, at each iteration we add a training waveform at the ten points with highest $O_k$ of the 100 computed, rather than just adding one at a time. In the example we show here, we seed the GPR model with 12 initial waveforms: one on each corner of the parameter space and two equally-spaced training waveforms on each edge. We perform 11 iterations (i.e. 122 total training points) of the greedy algorithm, and compute the mismatch between the IMR model and the GPR mean just as in §2.4.3. Figure 2.10 shows these mismatch values over the parameter space. Comparing Figure 2.10 to Figure 2.9, which have 122 and 120 training points, respectively, we see that the iterative method results in lower mismatches than the regular grid across the parameter space. Additionally, the maximum mismatch over the parameter space in the iterative case is $9.3 \times 10^{-5}$, which is an order of magnitude lower than the maximum mismatch over the regular grid of $1.4 \times 10^{-3}$.

To address the question of whether we can estimate the true mismatches using $O_k$, we compare the maximum of the 100 $O_k$ values and the maximum GPR mean-IMR mismatch

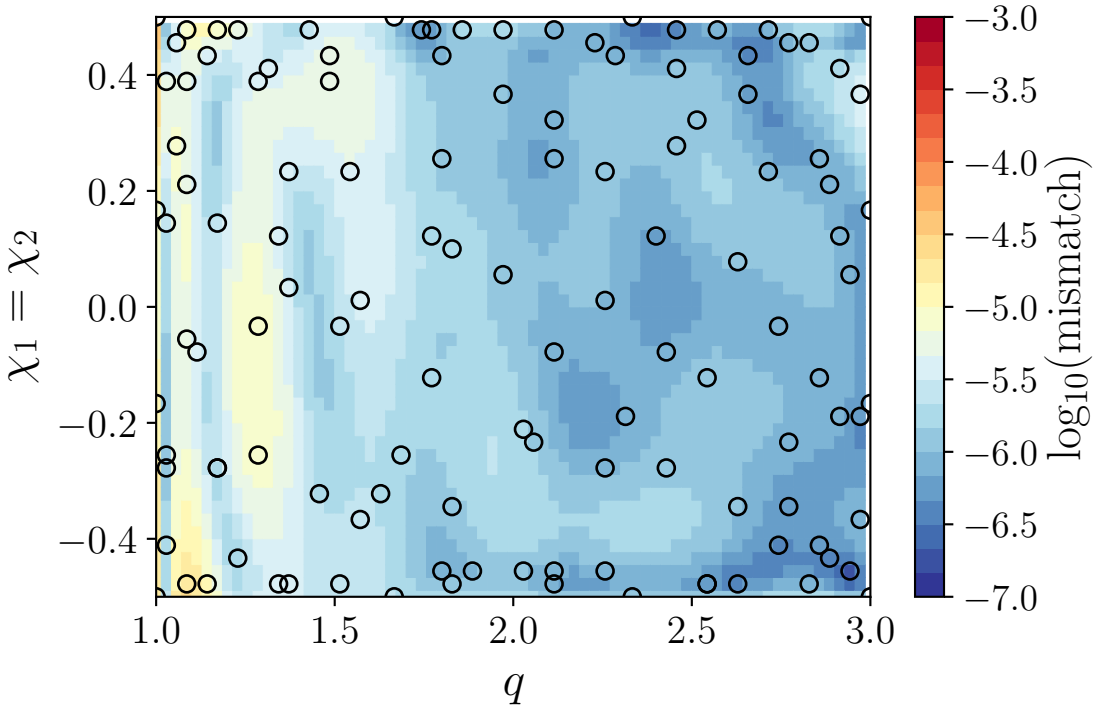


Figure 2.10: Mismatch between IMRPhenomD waveforms and GPR mean waveforms with an iteratively-built training set. The training set shown here was seeded with 12 initial IMRPhenomD waveforms and 10 points were added at each iteration based on 100 samples of $O_k$ across the space. The black circles show the locations of training waveforms from IMRPhenomD used to train the GPR. In this example, 10 iterations were performed, yielding $10 \times 10 + 12 = 122$ training points and a maximum mismatch in the region of $3.4 \times 10^{-5}$.

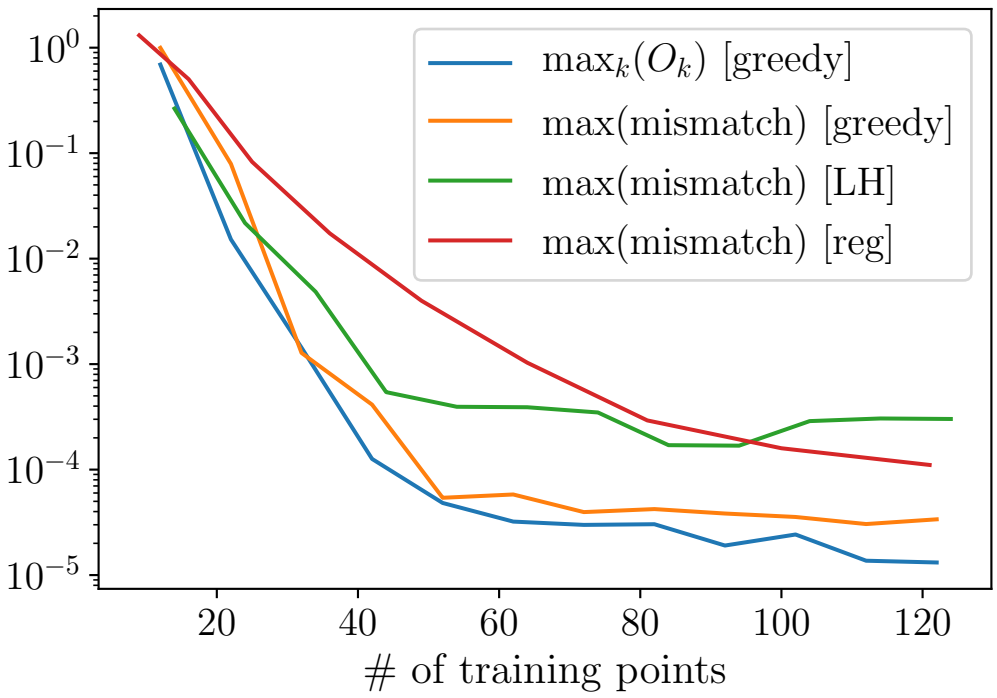


Figure 2.11: Maximum $O_k$ value using the greedy algorithm and maximum mismatch between the GPR mean and IMRPhenomD for different training point schemes and numbers of training points. For the greedy training point placement, twelve training waveforms on the boundaries of the space from IMRPhenomD seed the GPR model at the first iteration. At each subsequent iteration, the ten points with the highest $O_k$ (out of 100 points tested) decide the locations for new training waveforms. The maximum $O_k$ with greedy placement is shown in blue, and maximum mismatches for greedy, Latin-hypercube, and square grids are shown in orange, green, and red, respectively.

over the parameter space at each iteration in Figure 2.11. Using a greedy grid, the maximum $O_k$ value (blue) tracks the maximum GPR mean-IMR mismatch (orange) to within an order of magnitude, though the maximum $O_k$ would likely be higher if more than 100 $O_k$ samples were taken. On the last iteration, we calculate $O_k$ on the fine interpolation grid used in §2.4.3 rather than just sampling 100 $O_k$ values. This yields the $O_k$ map shown in Figure 2.12. By construction, $O_k$ is relatively constant over the parameter space. Additionally, the maximum $O_k$ on this finer grid is $9.3 \times 10^{-5}$, which bounds the maximum GPR-IMR mismatch of $3.4 \times 10^{-5}$, further suggesting that the maximum $O_k$ value can be used to estimate the maximum true error level of the GPR mean. As such, $O_k$ can indicate when a sufficient number of training waveforms have been used.

It is worth noting that the maximum mismatch of $4.3 \times 10^{-5}$ on the 120-point regular grid from §2.4.3 is comparable to the maximum mismatch of $3.4 \times 10^{-5}$ using the 122-point greedy grid. However, this fact does not indicate that regular training grids are as effective as greedy grids: On the regular grid, the number of training points in the $q$-direction (15) and in the $\chi$-direction (8) were tuned to achieve low mismatches. In practice though, such tuning will not be possible since (a) the true GPR-mean error will not be known at points without simulations, and (b) building different regular grids for tuning would use significant simulation resources.

To compare the greedy algorithm to other training point placement schemes, Figure 2.11 also shows the maximum mismatch over the $q$-$\chi$ space between GPR and IMR as a function of the number of training points for a Latin hypercube (LH) grid (green) and a regular, square grid (red). For the LH case, a training point is placed at each corner of the space, and then the space is LH sampled with multiples of 10 additional training points. In other words, for each trial, training points are put on the corners of the space, and a new hypercube with $10 \times n$ partitions per axis ($n \in [1, 12]$) is constructed and randomly populated with training points under the constraint that there is exactly one training point in each row and column of the hypercube. For the square grid case, an equally-spaced $n \times n$ training

grid spanning the space of interest is created for $n \in [3, 11]$. Examining Figure 2.11, the greedy algorithm is able to achieve mismatches considerably below the Latin hypercube or square grid mismatches for the same numbers of training points, suggesting that the greedy algorithm is the best simulation placement strategy when the training grid cannot be tuned with trial and error.

From a theoretical standpoint, it is not surprising that the greedy grid tends to be more accurate than these other "pseudo-uniform" sampling techniques. To see this, consider the conditional covariance $K_{\text{cond}} = K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*)$ from Equation 2.5. Note that as the elements of $K(X_*, X)$ get small, the second term diminishes, making the whole expression approach the prior covariance $K(X_*, X_*)$. Assuming the training set is relatively uniform over the input space, points on the edges of the space have fewer nearby training points and hence result in smaller elements of $K(X_*, X)$. In the center of the space, there are many nearby training points, so $K(X_*, X)$ tends to have larger elements, which decreases the elements of the conditional covariance. In effect, the GPR model has higher uncertainties near the boundaries when the training points are uniformly spread out. Neither LH sampling nor square grids take into account that the GPR uncertainties are highest near the edges. On the other hand, the greedy algorithm accounts for the GPR uncertainties and preferentially puts training points near the boundary, as can be seen in Figures 2.10 and 2.12.

We would like to stress that greedy GPR grids are only optimal when using GPR as the interpolation method. For other interpolation schemes such a grid can be highly sub-optimal or even not applicable at all. It is also worth noting that the choice of GP kernel will affect the greedy grid properties. For example, the squared-exponential kernel tends to pile training points near the edges of the space (akin to e.g. P14 and Field *et al.* (2014)) compared to the Matern 5/2 kernel (which we discuss in §5.6). Additionally, properties of the ROM or surrogate model used for the GPR could also affect the greedy grid properties. Despite these limitations, if a GP model with a given kernel is desired, a greedy grid is an

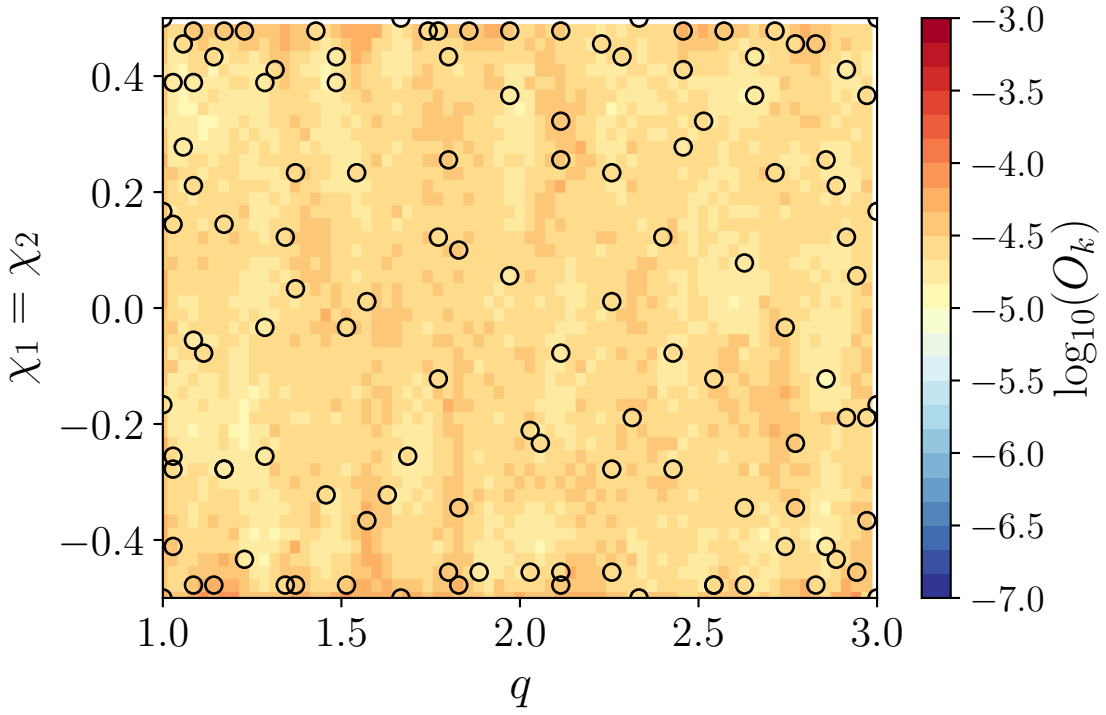


Figure 2.12: Maximum GPR sample-mean mismatch $O_k$ over 20 samples calculated at each point on the fine interpolation grid based on the same iteratively-built training set shown in Figure 2.10. The black circles show the locations of training waveforms from IMRPhenomD used to train the GPR. The maximum $O_k$ in the region is $9.3 \times 10^{-5}$.

effective strategy for minimizing errors.

One counter-argument to using $O_k$ to measure the error is that there could be sharp features in the coefficient functions which are not sufficiently sampled by the training points and hence are poorly interpolated. This could indeed be true in some cases, but it is an issue that applies to any interpolation scheme. A benefit of using GPR is that if sharp features exist, some of which are sampled by the training points, the hyperparameter optimization will select shorter length scales and larger covariances and hence increase the overall coefficient uncertainty across the space. Additionally, the GPR conditional distributions are Gaussian, meaning that large excursions from the mean are not ruled out — they are just less likely.

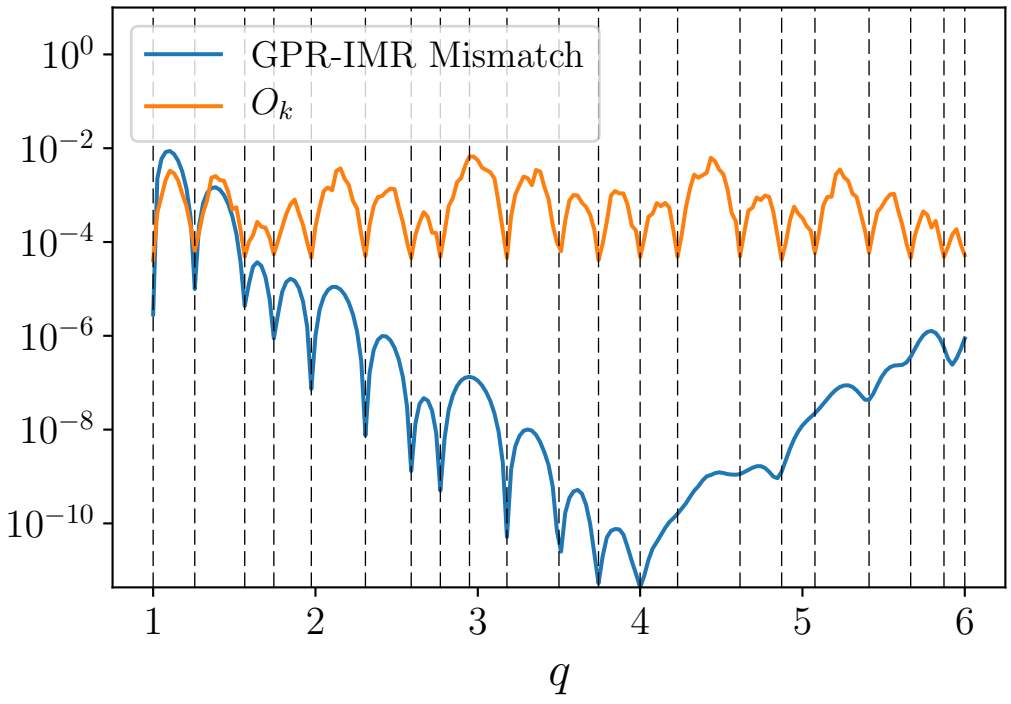


Figure 2.13: Mismatch between the GPR mean using a Matern 5/2 kernel and IMRPhenomD (blue), and the mismatch estimated using $O_k$ based on 20 GPR samples at each interpolation point (orange).

## 2.6 Discussion

In §5.5 we presented three examples of GPR-based ROM models trained on a subset of simulated waveforms. These example models produced accurate mean waveforms and quantified uncertainties across the parameter spaces of interest. §2.5 showed that further improvements in speed and accuracy can be made to GPR-based models through use of the greedy algorithm. Although we have made specific choices in our implementation, it is to be emphasized that our method is completely general. For example, different ROM or surrogate models could be used rather than the SVD-based ROM we employ, hyperparameters could instead be treated as nuisance parameters and marginalized over, the GP kernel could be changed, more sophisticated coefficient regularizations could be applied, and the greedy algorithm could be modified to incorporate other constraints. We discuss a few of these possibilities here. First, let us consider using the greedy algorithm with $O_k$ as the metric for placing

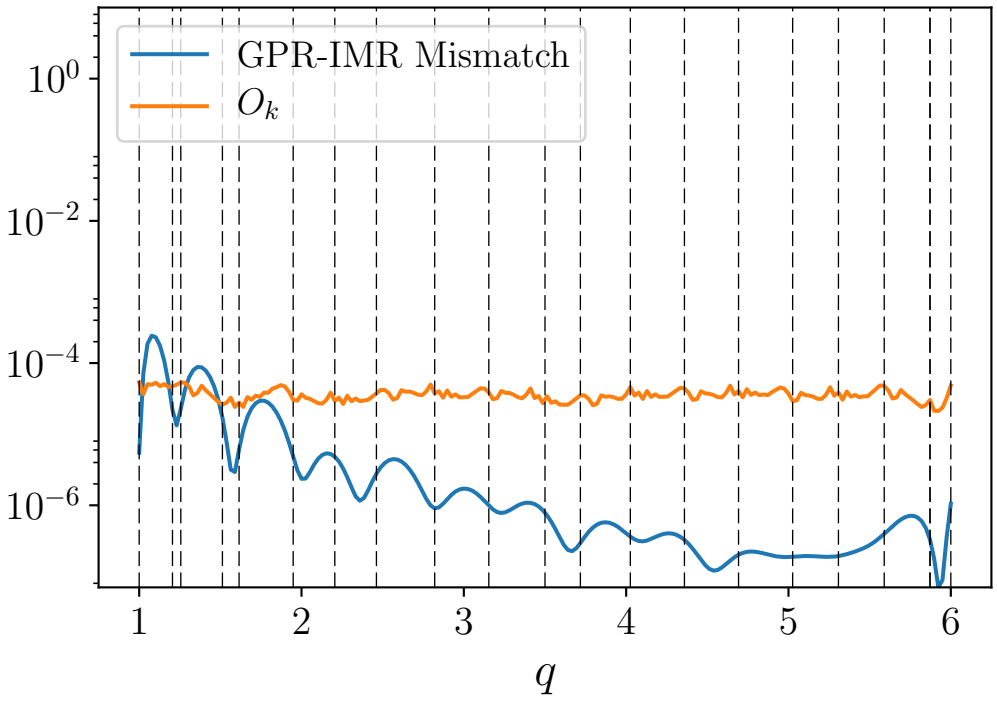


Figure 2.14: Mismatch between the GPR mean using a squared-exponential kernel and IMRPhenomD (blue), and the mismatch estimated using $O_k$ based on 20 GPR samples at each interpolation point (orange).

new simulations. It is worth emphasizing that greedy training point placement with $O_k$ does not strictly result in the smallest possible training set for a desired error level. Rather, the greedy algorithm attempts to flatten the error across the parameter space by adding training points where the error is estimated to be highest. A principal limitation to just using $O_k$ to guide simulations is that it does not encode other constraints or priorities. Two possible modifications to our greedy strategy include weighting $O_k$ by the expected simulation cost at certain parameter values and weighting by a prior on the source population parameters. Future work will investigate these possibilities. Next we discuss the kernel functions. The kernel functional form encodes our assumptions about the smoothness and fluctuations of a GP. Another common kernel choice, other than the squared-exponential kernel adopted

above, is the Matern 5/2 covariance:

$$\mathbf{k}_{\nu=5/2}(r) = \sigma^2 \left(1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}r}{l}\right), \tag{2.14}$$

The squared-exponential covariance constrains the GP to be infinitely mean-square differentiable, imposing a strong smoothness condition on the interpolations. The Matern 5/2 kernel is less restrictive in that it only demands the process be twice mean-square differentiable. To illustrate the effects of different kernel choices, we compare with the squared-exponential covariance. For both the squared-exponential and the Matern 5/2 kernels, we apply Algorithm 1 to build a training set to interpolate waveforms for $q \in [1, 6]$ as in §2.4.1. In each case, we begin with three training waveforms to seed the algorithm: one waveform on each end of the space and one in the center. At each iteration, $O_k$ is calculated on the fine interpolation grid from §2.4.1 and a new point is placed at the point of highest $O_k$.

Figure 2.13 shows the real GPR-IMR mismatch and the mismatch estimated from $O_k$ using the Matern 5/2 kernel, and Figure 2.14 shows the same for a squared-exponential kernel. In both cases, the maximum $O_k$ over the space bounds the real mismatch, except near $q = 1$ where the coefficient functions vary faster than elsewhere in the space (see Figures 2.2 and 2.3). The squared-exponential kernel is able to keep the error relatively constant over the space. Additionally, the maximum mismatch using the squared exponential is about two orders of magnitude lower than when using the Matern 5/2 kernel. On the other hand, use of the Matern kernel results in significantly lower error on the interior of the space. Further study of kernel effects will be required as new portions of parameter space are explored with the GPR model. In particular, higher dimensional GPR models may require more complex kernels, which can be constructed by summing or multiplying pre-existing kernel functions. Additionally, kernels with compact support should be considered, as they can allow faster GP evaluation and enhanced computational stability.

We now shift our focus to the execution time of generating GPR waveforms. The exam-

ples shown here were designed to run in less than one day on one computing node with 16 cores, but future work would make use of more cores, allowing, for example, a larger ROM basis, or interpolation and $O_k$ calculation on a finer grid. To illustrate the scaling of required time and resources with the number of GP training points, we perform a GPR on one coefficient in the $q$-$\chi$ space for different numbers of training values and evaluation points. The numbers of training and evaluation points determine the sizes of the matrices that must be multiplied in Equation 2.5 and hence the execution time. It is worth noting that $K(X,X)^{-1}$ or $K(X,X)^{-1}\mathbf{f}$ with optimized hyperparamters can be pre-calculated, allowing the most computationally intensive step in the GPR-building to be done just once ahead of time. We assume here that the optimization and matrix inversion steps have already been done and simply look at the evaluation time of a GP.

The timing results are shown in Figure 2.15, which plots the conditional GP mean and covariance evaluation time per coefficient per interpolated point at $n_{\text{interp}}$ points as a function of the number of training points. That is, we evaluate the GP mean and covariance at $n_{\text{interp}}$ points and divide the total time by $n_{\text{interp}}$ to show the time per interpolated coefficient value. In principle, the total evaluation time should scale directly with the number of interpolated points, but Figure 2.15 indicates that interpolating more points at once gives an overall speedup. This is due to the overhead in constructing the training-test covariance matrices in the `scikit-learn` implementation of the GPR. This overhead is further evidenced by the fact that the $n_{\text{interp}} = 100$ and $n_{\text{interp}} = 1000$ curves in Figure 2.15 converge. Future work will consider alternate GPR implementations to mitigate such overhead, since typically only single waveform evaluations are required.

Note that to build a full GPR waveform, a GP must be evaluated for each of the $\sim 100$ phase and amplitude coefficients. In the case that the overhead cannot be bypassed ($n_{\text{interp}} = 1$, blue curve), it would take $\sim 1$ minute per waveform evaluation, assuming 100 coefficients and 1000 training points. If the overhead can be entirely removed ($n_{\text{interp}} = 1000$, red curve), each waveform evaluation would instead take $\sim 200$ms. This can be compared to the

evaluation time for the spline-based ROM in P14 of $\sim 1$ ms depending on the system's total mass (see Figure 1 of P14). Further speedups to our model could be achieved by lowering the number of ROM coefficients, decreasing the number of training points, or evaluating the coefficients in parallel. To decrease the number of training points, domain decomposition could be used as suggested in §2.4.3. If domains of $\sim 100$ training points were used, the GP evaluation times would fall by over an order of magnitude (evaluation time goes as $n_{\rm train}^2$). Another possibility is to apply the subset of regressors method, which effectively considers a only a subset of rows of $K(X,X)^{-1}\mathbf{f}$ in Equation 2.5. This scheme reduces the size of the matrices multiplied when calculating the GP conditional mean and covariance (see §8.3.1 in (Rasmussen and Williams, 2005)).

As mentioned before, the most expensive step of building a GPR model is in training, where the hyperparameters must be optimized and $K(X,X)$ must be inverted. The matrix inversion computation time scales as $O(n_{\rm train}^3)$, since $K(X,X)$ is an $n_{\rm train} \times n_{\rm train}$ matrix (see e.g. (Rasmussen and Williams, 2005)). Given the timescale for generating an NR waveform, we would not expect more than $O(10,000)$ NR waveforms in the near future, so the matrix inversion step should not be prohibitively expensive, especially since the kernel for each coefficient can be handled in parallel, and a $O(1000)$ training set can be optimized on a personal computer in minutes and only needs to be done once. Also, as mentioned earlier, smaller domains with fewer training waveforms can be handled with separate GPR models or a subset of regressors can be used, mitigating the need for the inversion of very large covariance matrices.

In this section, we have discussed a few of the practical and computational challenges for GPR-based GW models as well as techniques for overcoming these difficulties. Implementation of GPR models in 3 or more input dimensions will require the use of some of these techniques, hence the consideration of only up to 2 dimensions in this work. Moving to higher dimensions, we envision needing domain decomposition, for example, to keep the number of training waveforms (and thus evaluation times) small. Beyond building and evaluating

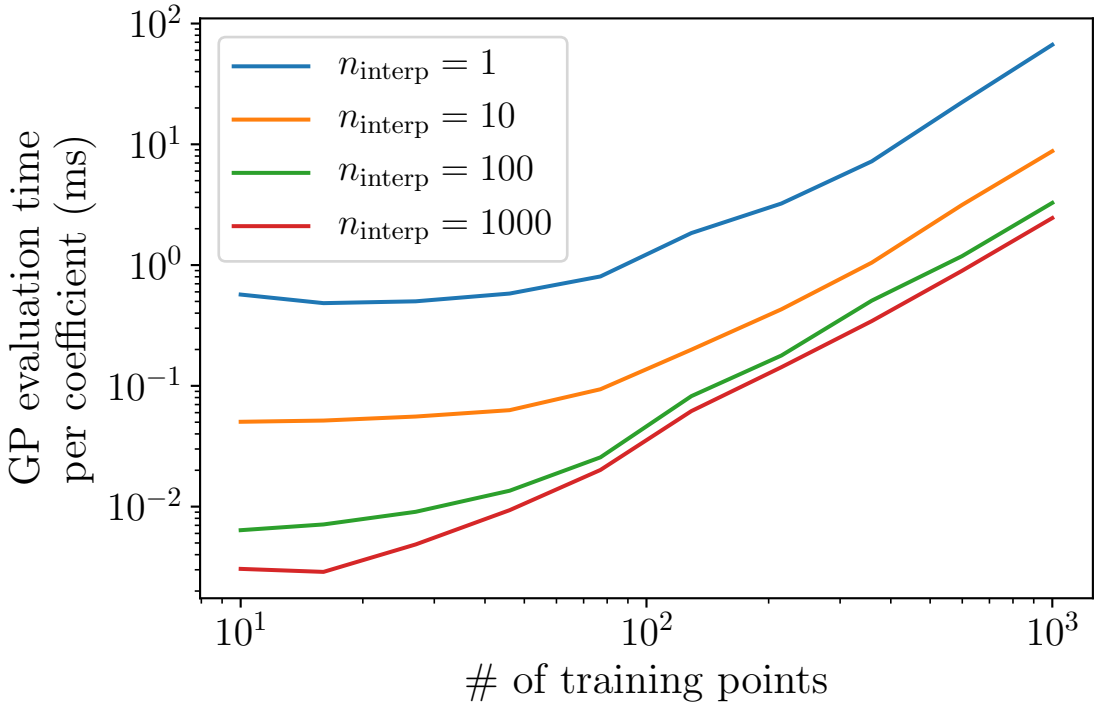


Figure 2.15: Time to evaluate (mean and variance) one coefficient per interpolation point in two dimensions as a function of the number of training points. Times shows are from GP evaluations on a 2.6GHz Intel E5-2670 CPU.

such higher-dimensional models, more strategic cross-validation methods than those used here will be required to ensure the models' robustness: dense, rectangular cross-validation grids of $O(50)$ points per dimension quickly become prohibitive in higher dimensions. Despite these challenges to higher-dimensional GPR models, we believe further work on these models is warranted. Our preliminary studies with 3-d GPR models in the space of mass ratio and un-equal, aligned spins show that reasonable error levels can be achieved with few training points, but further validation beyond the scope of this paper is needed before any concrete conclusions can be drawn. Additionally, we emphasize that the ultimate strength of these GPR-based models is their ability to produce quantified interpolation uncertainties which can be propagated to parameter estimates. Although a number of other models exist (e.g. P14) covering the parameter spaces we have examined here, GPR waveforms are the only ones that come with analytic uncertainties that vary with GW frequency and source parameters. Along with our future work on 3-and-higher dimensional source modeling with GPR, we hope to introduce methods for neatly incorporating these waveforms (and uncertainties) into MCMC parameter estimation techniques.

## 2.7 Conclusions

We have demonstrated that Gaussian process regression can be used to construct reduced-order-model waveforms with uncertainties using only a few existing simulations, and that these uncertainties can guide the choice of future simulations. The overall motivation for such GPR models is that GPR uncertainties can be propagated to the parameter estimation of compact binary coalescences in order to remove bias in estimates due to systematic waveform errors. Figures 2.4 and 2.5 show example amplitude and phase functions with uncertainties from the GPR model.

This work has also shown that GPR can model waveforms accurately over a parameter space of interest. Figures 2.9 and 2.10 show that with a sufficient number of training sim-

ulations, the error level of the GP model can be reduced to levels adequate for parameter estimation with LIGO data, especially if Algorithm 1 is used to construct the training set. Such greedy algorithms will be a particularly useful tool for efficiently choosing the parameters of new simulations in the nominal 7-d parameter space of interest to LIGO. Since the methods presented here are general, they could in principle be applied to other scenarios such as eccentric-orbit Laser Interferometer Space Antenna sources or neutron-star binaries with tidal deformability parameters.

Another finding of this work is that the error level of the GPR model can be estimated from the GPR itself rather than through cross-validation. Figure 2.11, which compares the maximum true GPR error to the maximum estimated error, demonstrates that the GPR uncertainties can alone be used to estimate the maximum error level of the GPR model. This allows one to know when a GPR model has reached a desired error level and does not require further simulations in the parameter region of interest.

Finally, we describe future directions for this work. In the immediate future, GPR models will be applied to three- or higher-dimensional parameter spaces to test the robustness of these models as the complexity grows. In particular, we will see if the ROM coefficient functions can generally be treated as being uncorrelated. Additionally, a wider range of kernel functions will be explored than what has been presented here. In the longer term, GPR training sets will be built directly from NR simulations, rather than utilizing a stand-in approximant. Apart from doing PE studies with an NR-based GPR model, the model could also be used to validate or study families of approximants. In sum, GPR models present an exciting frontier for NR-simulation-driven models of GW waveforms.

# CHAPTER 3

# ESTIMATING THE CONTRIBUTION OF DYNAMICAL EJECTA IN THE KILONOVA ASSOCIATED WITH GW170817

## 3.1 Summary

The gravitational-wave signal from a neutron-star merger is uniquely described by the masses, spins, and structures (i.e. equation of state) of the two neutron stars assuming a circularized binary. Similarly, the ejection of neutron-rich material from the merger also depends only on the initial parameters of the inspiraling neutron stars. Therefore, gravitational-wave measurements of neutron-star mergers can be used to estimate properties of the mass ejected in the collision. The mass that is expelled undergoes r-process nucleosynthesis, heats up, and glows brightly to power an optical / near-infrared transient called a "kilonova". This chapter investigates the implications of the gravitational wave signal GW170817 for binary neutron star merger mass ejection and kilonova morphology.

In an ideal world, the gravitational-wave and electromagnetic data could be simultaneously leveraged to make joint inferences on the parameters of the inspiraling neutron stars and their eventual violent collision. However, the gravitational waves and electromagnetic counterparts are observed by different teams, so making joint use of the two data sets is logistically difficult. As a result, gravitational-wave and electromagnetic analyses have to be done separately (until eventual data releases) and then the results of those analyses have to be compared at a high level through the published literature. This chapter is essentially the gravitational-wave half of a global multi-messenger analysis and was published to show the breadth of constraints that one can make with gravitational-wave data alone.

This chapter is based on the LIGO-Virgo Collaboration publication *Estimating the Contribution of Dynamical Ejecta in the Kilonova Associated with GW170817*, for which I was one of the primary authors. My main contribution to the analysis was in the last section, which describes the implications of GW170817 for heavy element production in neutron star

mergers over the age of the universe. A debate has been raging over the past few decades as to the origin of the heavy r-process elements such as gold that we see here on Earth and in space. It was previously unknown whether core-collapse supernovae or neutron-star mergers were the primary site of r-process-element production, but this analysis of GW170817 (among others) suggested that neutron star mergers could be the dominant r-process site.

## 3.2 Introduction

On August 17, 2017, 12:41:04 UTC, the LIGO – Virgo gravitational-wave (GW) observatory network, composed of LIGO Hanford Observatory, LIGO Livingston Observatory, and Virgo, recorded GWs consistent with a binary neutron star (BNS) inspiral and merger (Abbott et al., 2017c). This signal was subsequently named GW170817.

In addition to the GW signature, the merger of a BNS system is expected to have multiple electromagnetic signatures over different time scales (Nakar, 2007; Metzger & Berger, 2012). The LIGO-Virgo sky localization of GW170817 (Abbott et al., 2017c) spurred an intensive multi-messenger campaign covering the whole electromagnetic spectrum to search for counterparts (see Abbott et al. 2017d for an extended list). Within hours, broadband observations — backed by archival data investigation — revealed an optical transient (Coulter et al., 2017; Soares-Santos et al., 2017; Valenti et al., 2017; Arcavi, Hosseinzadeh, Howell, et al., 2017; Tanvir et al., 2017; Pian, D'Avanzo, et al., 2017; Lipunov et al., 2017), a type of transient called a kilonova (Li & Paczynski, 1998; Metzger, 2017) originating from neutron-rich matter unbound from the system (e.g., Evans et al., 2017; McCully et al., 2017; Smartt et al., 2017; Troja et al., 2017).

Broadly, two types of ejecta are expected to contribute to kilonovae: dynamical ejecta produced at the time of merger (Rosswog et al., 1999; Metzger et al., 2010; Roberts et al., 2011; Rosswog, 2013; Barnes & Kasen, 2013; Tanaka & Hotokezaka, 2013; Hotokezaka et al., 2013; Bauswein et al., 2013; Sekiguchi et al., 2016; Radice et al., 2016; Dietrich et al., 2017b; Dietrich & Ujevic, 2017; Bovard et al., 2017), and post-merger winds produced by

the remnant system, for example from an accretion disk around a black hole or massive neutron star (Dessart et al., 2009; Perego et al., 2014; Martin et al., 2015; Kiuchi et al., 2015; Fernández et al., 2015; Kasen et al., 2015; Foucart et al., 2016; Shibata et al., 2017; Siegel & Metzger, 2017; Ciolfi et al., 2017; Fujibayashi et al., 2017).

Both electromagnetic and GW measurements rely on models to connect the underlying properties and composition of the ejecta to their respective observations. The process of interpreting ejecta based on electromagnetic observations is described in Alexander et al. (2017); Arcavi et al. (2017); Chornock et al. (2017); Covino et al. (2017); Cowperthwaite et al. (2017); Diaz et al. (2017); Drout et al. (2017); Evans et al. (2017); Kasen et al. (2017); McCully et al. (2017); Nicholl et al. (2017); Pian et al. (2017); Smartt et al. (2017); Tanaka et al. (2017); Troja et al. (2017); Abbott et al. (2017d). We use phenomenological calculations that estimate the dynamical ejecta mass from the pre-coalescence binary properties, which GW observations can constrain. This mass is a critical ingredient needed to predict contribution of dynamical ejecta to the EM light curve associated with this kilonova transient. Going forward, this procedure would also assist in the interpretation of future followup observations where a dim counterpart was detected, or none at all.

This Letter shows how dynamical ejecta masses obtained from GW parameter estimates of GW170817 via phenomenological fits to numerical models for the mass and velocity of dynamically ejected matter in BNS systems (Dietrich & Ujevic, 2017) (hereafter DU17) can predict kilonova light curves. Similar numerical work has produced fitting formulae in the case of neutron-star black-hole (NSBH) binaries (Kawaguchi et al., 2016). While the GW detection of GW170817 cannot rule out the presence of a black hole companion, the BNS interpretation is favored (Abbott et al., 2017c). Consequently, we do not include the NSBH scenario in this work, and only employ the fitting formulas for ejecta mass and velocity from BNS simulations (DU17). The GW170817 analysis extracted the BNS source parameters using Bayesian inference (Abbott et al., 2017c), and those results are used here to estimate the mass of the dynamical ejecta. This approach accounts for the dependence of the amount

of ejected matter on the size and stiffness (Kawaguchi et al., 2016) of the components of the binary, characterized by the equation of state (EOS) and its influence on the mass-radius relationship (Lattimer & Prakash, 2001; Özel & Freire, 2016).

Bayesian inference with a GW signal model applied to the strain data provides a posterior distribution of component masses ($m_i$) and dimensionless spins ($\chi_i \equiv c|\mathbf{S}_i|/(Gm_i^2)$, where $\mathbf{S}$ is the angular momentum of the NS) consistent with the observations (Veitch et al., 2015). Assuming neutron stars spins are small ($\chi \leq 0.05$, hereafter "low spin"), we obtain distributions of ejecta between $10^{-3}$ and $10^{-2}$ $M_\odot$. Allowing for larger neutron stars spins ($\chi \leq 0.89$, hereafter "high spin") pushes some ejecta values higher, of the order of $10^{-1}M_\odot$ at its highest. In this Letter, we focus on dynamical sources, so it is important to recall that this analysis may not account for a significant fraction of the ejecta mass; winds could produce comparable or even more ejecta than from dynamical sources. Using the GW-derived dynamical ejecta estimates, the derived light curves vary significantly between the adopted models, in both color evolution and time and magnitude of peak emission; in extreme cases, they can reach beyond 15th magnitude in optical bands.

Like supernovae (Terasawa et al., 2001), neutron star mergers are believed to contribute to the abundance of heavy elements (Lattimer & Schramm, 1974) through r-process nucleosynthesis (Burbidge et al., 1957). If so, the frequency of kilonova events should then be intimately tied to the overall abundance of r-process generated material (Lattimer & Schramm, 1974). Using our GW estimates of dynamical ejecta masses and the merger rates inferred from the BNS discovery ($1540^{+3200}_{-1220}$ $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$) (Abbott et al., 2017c), we estimate a present-day r-process density of $10^{1.7} - 10^{3.2}$ $M_\odot\mathrm{Mpc}^{-3}$ contributed by BNS mergers. Under the assumption that all BNS mergers produce the same amount of dynamical ejecta that we infer for GW170817, this estimate is consistent with the Galactic values and suggests the associated nucleosynthesis is one of the primary contributors to r-process abundances.

## 3.3 Predicted Dynamical Ejecta Mass

The amount of ejecta from binary mergers in general depends on the masses and EOS of the two components, their rotation, and, most importantly for post-merger winds, the neutrino/radiation hydrodynamics and the magnetic fields, e.g. Hotokezaka et al. (2013); Martin et al. (2015); Sekiguchi et al. (2016); Radice et al. (2016); Dietrich et al. (2017b); Siegel & Metzger (2017). Based on detailed numerical studies of merging, irrotational binaries, the phenomenological fits devised by DU17 relate the dynamical ejecta mass $M_{\rm ej}$ to the gravitational mass of the component stars ($m$), their baryonic mass ($m_b$), and their radii $R$ (or equivalently compactnesses $C = Gm/Rc^2$). Contributions due to winds were not included in the simulations used by DU17, and thus are not part of the fits for $M_{\rm ej}$, even though they may lead to comparable ejecta masses.

Because the EOS in neutron stars is poorly constrained, two approaches are taken to describe the bulk properties of the binary components. In the first approach, we assume an EOS and infer $m_b$ and $C$ from the binary's measured gravitational masses using a zero-temperature non-rotating model (computed using the Oppenheimer-Volkoff equations, Oppenheimer & Volkoff 1939). Different EOS will predict different radii and baryonic masses for the same gravitational masses and, as such, will affect the amount of ejecta and the predicted light curve of the kilonova. The EOS of cold, dense, degenerate matter is poorly constrained (see Oertel et al. 2017 for a recent review), so we evaluate a representative selection of the EOS considered in Özel & Freire (2016). The tidal deformabilities allowed by GW170817 (Abbott et al., 2017c) do disfavor stiffer EOS; however, many remain compatible with our measurements. Due to observational constraints, we restrict ourselves to EOS that have a maximum mass above 1.97 $M_\odot$ (Demorest et al., 2010; Antoniadis et al., 2013). Specifically, we consider EOS calculations from Glendenning (1985), GNH3; Müther et al. (1987), MPA1; Wiringa et al. (1988), WFF1-2; Engvik et al. (1996), ENG; Müller & Serot (1996), MS1, MS1b; Akmal et al. (1998), APR3-4; Douchin & Haensel (2001), SLy; and Lackey et al. (2006), H4.

In the second case, we take an approach that does not assume a specific EOS to compare against our EOS specific results. The internal structure of the neutron stars in a binary is encoded in the gravitational waveform through the (dimensionless) tidal deformabilities (denoted $\Lambda$) of the neutron stars (Flanagan & Hinderer, 2008; Damour et al., 2012; Del Pozzo et al., 2013; Wade et al., 2014). One can infer $m_b$ and $C$ from the binary's measured gravitational masses and tidal deformabilities by applying fits from Coughlin et al. (2017) and Yagi & Yunes (2017), which give $m_b(m, C)$ and $C(\Lambda)$, respectively. While some error is incurred using these additional fits, it is small compared to the estimated uncertainty of the fits for the dynamical ejecta properties and the intrinsic uncertainty in current numerical relativity simulations. Specifically, for the EOS considered by Yagi & Yunes (2017), the error in the tidal deformability-compactness relation is $< 10\%$ for nuclear EOS, while for the baryonic mass fit, the maximum error found by Coughlin et al. (2017) is $< 3\%$. When applying these fits, we also exclude cases with component masses above 3 $M_\odot$, a standard upper bound on neutron star masses (Kalogera & Baym, 1996), and restrict the compactness to be below the Buchdahl bound (Buchdahl, 1959) of $4/9 \simeq 0.44$, which similarly only affects a few cases.

### *3.3.1 Sources of Uncertainties in Ejecta Mass Estimation*

Many caveats must be considered when assessing the uncertainty in estimates of $M_{\rm ej}$. The amount of ejecta from mergers also depends on various microphysics, such as the particular treatment of thermal effects, neutrino transport, and magnetic fields (Dessart et al., 2009; Perego et al., 2014; Bauswein et al., 2013; Sekiguchi et al., 2016; Radice et al., 2016; Bovard et al., 2017; Ciolfi et al., 2017), which lead to uncertainties about the ejecta's structure, angular distribution, and composition (Kasen et al., 2013; Tanaka & Hotokezaka, 2013; Barnes et al., 2016). These parameters are not included in the $M_{\rm ej}$ fits in DU17. Additionally, the DU17 fits ignore the effects of spin on dynamical ejecta, which can change the amount of ejecta (Kastaun & Galeazzi, 2015; Dietrich et al., 2017a; Kastaun et al., 2017). In particular,

aligned spin can increase torque in the tidal tail and lead to more ejecta, which is most notable for unequal mass configurations. To understand the effect of spin on dynamical ejecta, additional better resolved simulations are needed.

Systematic uncertainties are also of concern. The accuracy of the $M_{\rm ej}$ fit from DU17 relies on the underlying numerical relativity simulations. Simulation choices for input physics (nuclear EOS and microphysics), inclusion of different neutrino transport models, and chosen grid resolution can all result in large systematics. For example, comparison of numerical relativity predictions of $M_{\rm ej}$ differ by a factor of $\sim$4 (Sekiguchi et al., 2016; Lehner et al., 2016; Bovard et al., 2017). Further, the error on ejecta masses from numerical simulations likely has an absolute component, leading to increasing relative errors for low ejecta masses — for additional discussion see Endrizzi et al. (2016); Ciolfi et al. (2017). The error at low masses is not symmetric since $M_{\rm ej}$ cannot be negative, potentially biasing the phenomenological fits of DU17 to overestimate the ejecta mass. Additionally, there are also systematic uncertainties introduced by the specific form of the fit, where all EOS effects are contained in the values of $m_b$ and $C$ for a given $m$. Finally, as discussed in Abbott et al. (2017c) and Sec. 3.3.2, the waveform model used to infer the masses and tidal deformabilities from the gravitational wave signal introduces its own systematic uncertainties, though these are estimated to be smaller than those of the DU17 $M_{\rm ej}$ fit.

All these considerations will contribute to the uncertainty in the $M_{\rm ej}$ fit from DU17; the error is a mixture of systematic errors that need investigation with dedicated future studies and numerical simulations. To model some part of this error, we will treat the average relative error of the fit quoted in DU17 (72%) as a statistical error for any results used here and defer a more robust error analysis to future work. We include an estimate of the error of the $M_{\rm ej}$ fit from DU17 by replacing each ejecta mass sample with a random value consistent with a Gaussian distribution in $\log_{10} M_{\rm ej}$ centered on the value and with standard deviation of $\log_{10} 1.72$, as motivated in Section 3.3.1. This method excludes zero ejecta masses and errors for small ejecta masses $\lesssim 10^{-3}\,M_\odot$ are not well modeled. The ejecta mass fit is based

on simulations with non-zero ejecta mass. The full parameter space likely also contains cases with little or no ejecta mass, for example, systems exhibiting prompt black hole formation. Since we reported in Abbott et al. (2017b) that prompt collapse can only be excluded for extreme EOS such as MS1, and the fit at values below $3 \times 10^{-3}\, M_{\odot}$ strongly overestimates the ejecta mass compared to the NR data points, the fit cannot reliably exclude zero ejecta mass below this value. Figure 3.1 shows that, in the low-spin cases, the number of samples less than $M_{\rm ej} < 3 \times 10^{-3}{\rm M}_{\odot}$ is typically $\sim 10 - 15\%$ of the cumulative total for most. In extreme cases, this fraction is up to 50%, but also arises from EOS which have been disfavored in Abbott et al. (2017c). In the high-spin cases, this number is typically smaller, around 5-10%, but can reach up to 25% in the extreme cases. We also discard the few samples when the fit predicts a negative value.

### *3.3.2 Ejecta Mass Predictions*

We evaluate the $M_{\rm ej}$ fit using the binary parameters derived from the gravitational-wave analysis (Abbott et al., 2017c). These parameters include the gravitational masses, tidal deformabilities, and spins of the component stars, though the spins are not used in evaluating the fit. Bayesian inference provides a distribution of these parameter values as a set of independent samples drawn from the posterior (Veitch et al., 2015; Abbott et al., 2016a). As a quantity that derives from these binary parameters, $M_{\rm ej}$ then is also represented as a statistical sample.

While the estimation of $M_{\rm ej}$ does not include the component spins as an input, they are an important degree of freedom in the waveform models used in the GW analysis. We consider two sets of GW parameter samples, defined by the choices for the prior on the spin magnitude. The two spin priors considered here are $\chi \leq 0.89$ (our "high spin" case with the upper limit dictated by the waveform model used), and $\chi \leq 0.05$ (our "low spin" case, slightly above the largest inferred spin at merger of a neutron star in a binary neutron star system that will merge within a Hubble time (Burgay et al., 2003)). While the waveform models

used only include the effects of the spin components along the orbital angular momentum, the spin priors assume isotropic spin directions. The very highest spins allowed in the high spin posterior set exceed the mass-shedding limit ($\chi \sim 0.7$ for the EOS considered in Lo & Lin 2011), but the small density of posterior samples in this region lies outside the 90% credible intervals. More importantly, the high-spin posterior on the primary mass contains samples with masses well above the maximum mass allowed for a static NS for any of the EOS we consider; we simply exclude from consideration any samples with such unsupported masses for each EOS.

There are also systematic errors introduced by the waveform model used. As discussed in Abbott et al. (2017c), analysis with a different waveform model changes the 90% credible bounds on the masses by $\sim$ 15% in the high-spin case (with no changes in the low-spin case), and the bounds on the tidal deformabilities by $\sim$ 20–30% in both low- and high-spin cases. Since these differences are below the systematic errors of the DU17 fit, we do not attempt to account for them here. The true systematic errors from waveform models may be significantly larger than those estimated in this comparison; making such assessments is the subject of future work.

Figure 3.1 reports cumulative probability distributions for the dynamical ejecta for a selection of the EOS tested. While all cases predict ejecta concentrated between $10^{-3}$ – $10^{-2}\mathrm{M}_{\odot}$, the high spin results allow for larger median ejecta values in general — maximum values can exceed a tenth of a solar mass. Since the DU17 fits for $M_{\mathrm{ej}}$ neglect spin, the differences in ejecta for the cases shown in Figure 3.1 are driven by the imprint of the spin choices inherent in the GW analysis that was input into this analysis.

Figure 3.2 shows the distribution of ejecta masses using the SLy EOS, illustrating how the ejecta mass tends to scale with the component mass distribution. Among the EOS tested, SLy is nearer the lower side of ejecta distributions in both the estimated median and maximum ejecta. The fits themselves imply an ejecta distribution strongly dependent on the mass of the primary ($m_1$) and the difference between the primary and secondary masses.

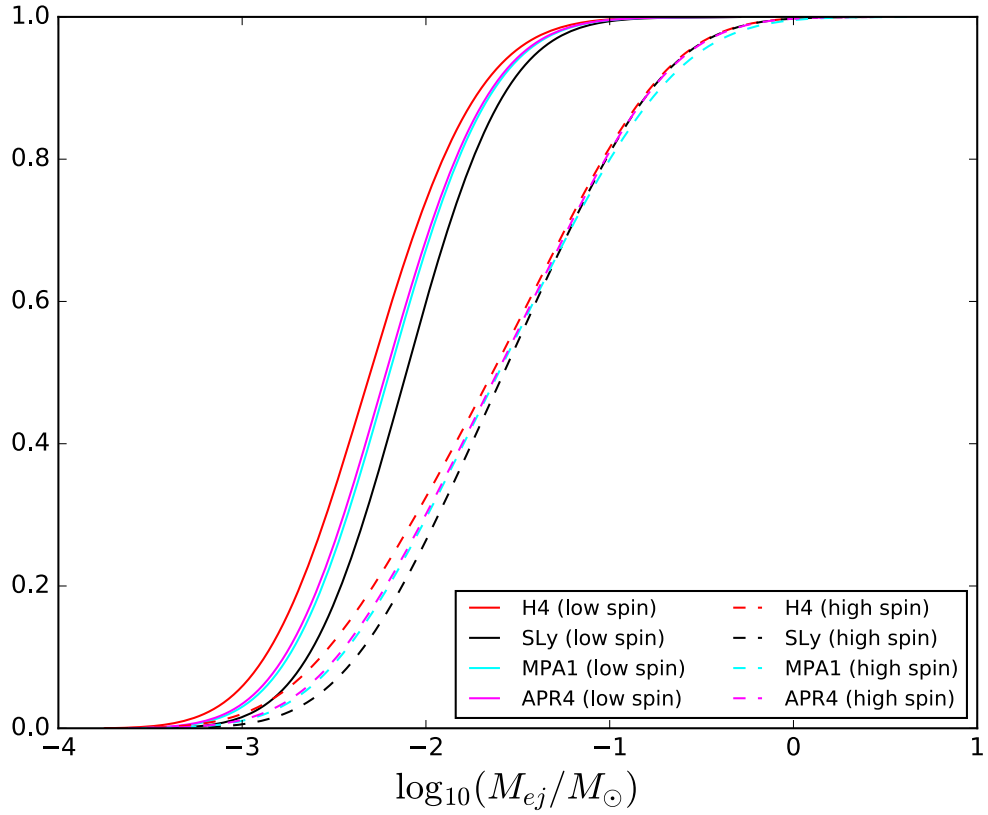


Figure 3.1: The figure above displays the cumulative distribution function of the dynamical ejecta mass predicted for a representative selection of the EOS in the study. The low spin case are traced in solid colors and the high spin case are dashed curves.

However, applying the fit uncertainty smears the ejecta distribution over the difference of the component masses. This effect is most evident in the marginal distributions plotted as histograms on the sides of the Figure 3.2 panels. Since the high spin distribution has more posterior samples away from equal mass systems, as well as larger primary masses overall, more samples give rise to larger ejecta masses. While this only affects the high spin case, those EOS which allow for larger maximum masses also allow for a larger maximum ejecta values, typically $M_{\rm ej} > 10^{-1}$ M$_\odot$ (above the maximum ejecta mass of $6.5 \times 10^{-2}$M$_\odot$ in the simulations to which the $M_{\rm ej}$ fit has been calibrated). This is a natural consequence of larger maximum masses corresponding to larger differences between $m_1$ and $m_2$, as illustrated in Figure 3.2.

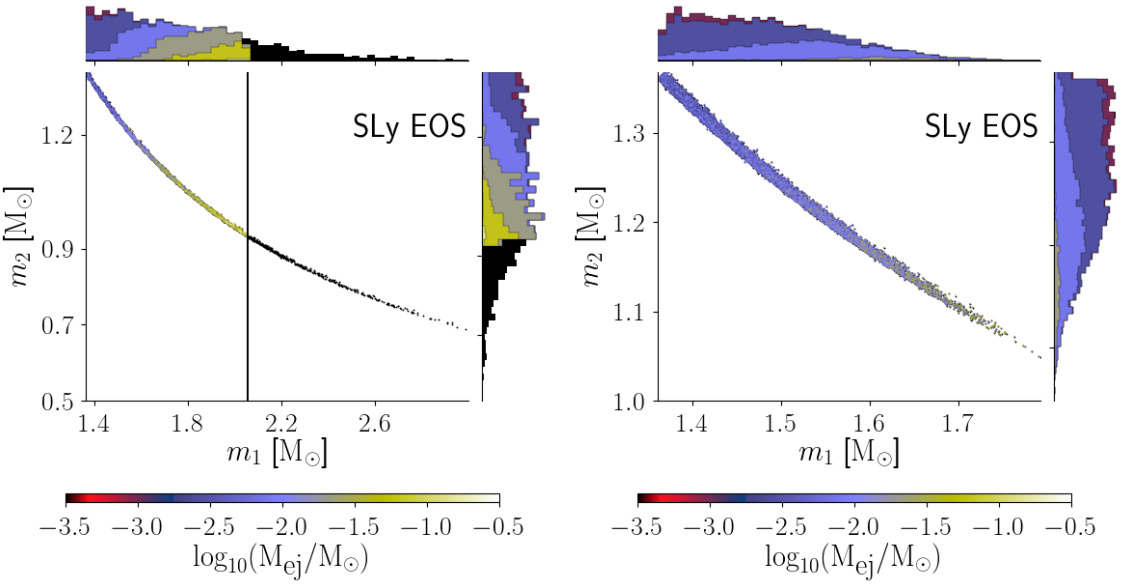


Figure 3.2: The left (high-spin prior) and right (low-spin prior) panels above show the distribution of the primary ($m_1$) and secondary ($m_2$) masses from GW measurements. The color of each point indicates the predicted dynamical ejecta mass for each sample that the SLy EOS allows. In the left-hand plot, black markers correspond to $m_1$ values that are disallowed by the maximum mass of the EOS (marked by a vertical line). The underlying black histograms to the top and right of each plot are the one-dimensional marginalized histograms of the masses. The stacked histograms on top of them in various colors show the binary masses that create ejecta masses above logarithmically spaced thresholds of $1\times10^{-3}$, $3\times10^{-3}$, $8\times10^{-3}$, $2\times10^{-2}$, $6\times10^{-2}$, and $2\times10^{-1}$M$_\odot$ where only the first four are nonzero in the right-hand plot.

## 3.4 Kilonova Light Curve Models

Current kilonova emission models (Li & Paczynski, 1998; Metzger, 2017; Barnes et al., 2016; Tanaka et al., 2017) produce spectral energy distributions between the ultraviolet (UV) and the near-infrared (NIR). Generally, there are two different physical processes that require modeling. First, the conversion of dynamical and wind ejecta material into r-process elements (i.e., the nucleosynthesis) (Kasen et al., 2013, 2015; Barnes et al., 2016; Rosswog et al., 2017; Metzger, 2017), and second, the production of an associated electromagnetic transient (Metzger et al., 2010; Kasen et al., 2013; Barnes et al., 2016; Rosswog et al., 2017). Beyond these considerations, there are still several important nuclear physics ingredients that are unknown, such as opacity and heating rate, and can lead to large uncertainties in

light curve prediction (see, e.g., Rosswog et al. 2017). We do not attempt to model these uncertainties.

We briefly describe here three parameterized models used to generate light curves in this work. Wollaeger et al. (2017) use radiative transfer simulations and provide analytic fits for the peak time, bolometric luminosity, and color corrections as a function of ejecta parameters. The Wollaeger et al. (2017) lightcurves are scaled as a function of ejecta mass and velocity, which changes both the time of peak luminosity as well as peak magnitude. We obtain the velocity from additional fits in DU17, and assume an opacity of 10 $cm^2/g$ thus modeling the presences of lanthanides. Conversely, Metzger (2017) provides a toy model for blue kilonova with opacity 0.1 $cm^2/g$ for lanthanide-free matter. DU17 use the radiative Monte-Carlo (MC) simulations of Tanaka & Hotokezaka (2013) and derive an analytical model for kilonova emission driven by dynamical ejecta from a BNS merger. No wind contribution is included in DU17 although winds can potentially dominate (Kiuchi et al., 2015; Ciolfi et al., 2017; Siegel & Metzger, 2017). The dynamical ejecta models tend to predict redder and more slowly rising NIR than wind-driven models.

Light curves from dynamical ejecta models depend significantly on the thermalization efficiency, the radiation transport simulations used, and other assumptions (Metzger & Fernandez, 2014; Coughlin et al., 2017; Rosswog et al., 2017). In our analysis we do not consider observational error from extinction in the light curve prediction, as it is likely smaller than the systematic error of the models (Kawaguchi et al., 2016).

## 3.5 Predicted Kilonova Light Curves

In conjunction with the mass and tidal estimates for the low spin case, we calculate the mass and velocity of dynamical ejecta as described in Section 3.3. Using the light curve models of DU17; Metzger (2017); Wollaeger et al. (2017), we show the absolute and apparent magnitudes consistent with these estimates of dynamical ejecta in Figure 3.3. Here, we employ the `DZ2` model from Wollaeger et al. (2017), and set 40 Mpc (near the median of

the GW distance posterior (Abbott et al., 2017c,a)) as the fiducial distance to the event for calculating the apparent magnitudes. DU17 exhibits the features of most lanthanide-rich dynamical ejecta models, with a rapid fade in the blue and a late rise in the NIR. Wollaeger et al. (2017), which also considers the contribution from the wind ejecta of $0.005\,M_{\odot}$, is both brighter, has a slower fade in the blue, and a faster fade in the NIR. The model in Metzger (2017) — adopted here only considering dynamical ejecta — is between these two models, originally brighter in the blue and NIR bands (g,r,i,z) than either of these models, but fades more quickly than Wollaeger et al. (2017).

Employing the lower opacity blue-peaked model in Metzger (2017) and GW inferred distance, we can calculate the distribution of peak times and observed peak magnitudes in a given photometric band. Since the source resides at a low redshift, we neglect the cosmological redshift of the source. Figure 3.4 shows the peak i-band magnitudes from those light curves versus the time of peak i-band magnitude when considering the low spin distribution. The samples from the high spin distribution produce the peaks which are brighter by one magnitude on average. This is understood from the ejecta distributions in Figure 3.2 — the low spin distribution tends to produce less ejecta and hence is less luminous. We note again that the light curves in Figure 3.3 are calculated with a distance fixed to the source, while the magnitudes in Figure 3.4 fold in the distance inferred from the GW data. Thus, a wider spread arises from the variance in the GW-only distance posterior distribution. Including the distance values from the GW posteriors better estimates the variation that would arise in a prediction from only GW information as opposed to having constraints from electromagnetic measurements.

The estimates presented here are a proof-of-principle study to illustrate what is possible at present with forward modeling from GW observations. Particularly if available before EM observations begin, or in a situation before a confident counterpart has been identified (e.g., due to poor sky localization), analysis driven by the GW data can inform EM followup observations and interpretation, particularly in cases where (due to geometric effects and

observational delays) the dynamical ejectas effect on the light curve is enhanced. Predictions of peak times in the the emission and the color evolution are useful for comparison with early observations, and provide falsifiable predictions to evaluate models of the source.

## 3.6 Abundance of r-Process Material

The r-process and s-process are the two known mechanisms by which heavy elements can be synthesized (Burbidge et al., 1957). To assess the contribution of the r-process to the observed abundances of heavy elements (Arnould et al., 2007; Sneden et al., 2008), one can identify the abundances expected from the s-process alone, and hence the r-process residual. Type II SNe can produce r-process elements, but they may not produce the observed abundance patterns (e.g., Freiburghaus et al., 1999). BNS mergers could also account for these elements. However, quantifying the contribution of those mergers has remained elusive due to poor constraints on both the rate of mergers as well as the amount of matter ejected in each merger. With GW170817, we are able to constrain both of these quantities significantly from data.

If BNS mergers are to produce most of the observed r-process elements in the Milky Way (MW), the mergers must occur with a sufficiently high rate and eject significant amounts of r-process material. Assuming dynamical ejecta dominate over winds, the mass fraction $X_{\rm rp}$ of r-process nuclei in the MW should be proportional to the merger rate density $\mathcal{R}$ and dynamical ejecta mass $M_{\rm ej}$, with a proportionality constant set by the local galaxy density and the MW age and mass. Following Qian (2000), we estimate that the merger rate and ejecta per event are approximately related by $\mathcal{R} \simeq 600 (f_{\rm rp} M_{\rm ej}/10^{-2}{\rm M}_\odot)^{-1} {\rm Gpc}^{-3}{\rm yr}^{-1}$. In this relationship, $f_{\rm rp} \equiv M_{\rm rp}/M_{\rm ej}$ is the fraction of matter dynamically ejected in NS mergers that is converted to heavy r-process elements rather than lighter products, e.g., $\alpha$ particles. The value of $f_{\rm rp}$ depends on details of the dynamics, geometry, and neutrino illumination of the ejected matter, all of which change the electron fraction ($Y_e$) distribution of ejected matter (see, e.g., Kasen et al., 2015; Goriely et al., 2015). However, various studies have

suggested significant r-processing of ejecta material (e.g., Goriely et al., 2011, 2015; Wanajo et al., 2014; Just et al., 2015; Radice et al., 2016). The red band in the left panel of Figure 3.5 shows this relationship between $\mathcal{R}$ and $M_{\rm ej}$ for $f_{\rm rp} \in [0.5, 1]$ (e.g., Goriely et al., 2015). Also shown in the left panel are the constraints on the local rate density of BNS mergers from GW170817 (gray) and the range of ejecta masses typically considered in the literature (blue). The overlap of these constraints suggests that BNS mergers could account for all of the observed r-process abundance.

A more detailed calculation of r-process enrichment from the dynamical ejecta of BNS mergers can be done using the specific distributions of $M_{\rm ej}$ and $\mathcal{R}$ inferred from GW170817. Under the assumption that all binary mergers have the same ejecta mass as that inferred from GW170817, we calculate the average dynamically ejected local r-process material density according to

$$\rho_{\rm rp} = f_{\rm rp} M_{\rm ej} \mathcal{R} \frac{\int_0^{t_h} \int_0^t \dot{\rho}_*(\tau) p_{\rm delay}(t-\tau) d\tau dt}{\int_0^{t_h} \dot{\rho}_*(\tau) p_{\rm delay}(t_h - \tau) d\tau}, \tag{3.1}$$

where $t_h$ is the Hubble time.[1]

In this expression, $\dot{\rho}_*$ is the cosmological star formation rate, assumed to follow Madau & Dickinson (2014); $p_{\rm delay}$ is the delay time distribution of NS mergers, $p_{\rm delay}(t) \propto t^{-1}$ (see, e.g., O'Shaughnessy et al., 2008; Dominik et al., 2012), with minimum delay time of 10 Myr; and $\mathcal{R}$ is the present-day merger rate density for NS mergers. The denominator is a normalization factor which scales the present-day merger rate density to $\mathcal{R}$.

In the right panel of Figure 3.5, we plot the distribution of $\rho_{\rm rp}/f_{\rm rp}$ for a few representative EOS using our $M_{\rm ej}$ distributions and the rates inferred from GW170817. On the top axis, we also show $X_{\rm rp}/f_{\rm rp} = (\rho_{\rm rp}/f_{\rm rp})/\rho_*$, where $\rho_* = \int_0^{t_h} \dot{\rho}_*(t) dt$. If $f_{\rm rp} = 1$, the range $10^{1.7}{\rm M}_\odot{\rm Mpc}^{-3} - 10^{3.2}{\rm M}_\odot{\rm Mpc}^{-3}$ brackets our 90% credible intervals on $\rho_{\rm rp}$ for all EoS. Both $\rho_{\rm rp}$ and $X_{\rm rp}$ are shown normalized to $f_{\rm rp}$, since $f_{\rm rp}$ depends on unknown details of the merger. The gray band in the right panel of Figure 3.5 shows the MW mass abundance

1. We assume ΛCDM cosmology with TT+lowP+lensing+ext parameters from Ade et al. (2016).

of r-process elements, derived from Arnould et al. (2007). As long as $f_{\rm rp} \gtrsim 10\%$ of the dynamically ejected mass is converted to heavy r-process elements, dynamical ejecta could account for all of the MW r-process abundance. We have not factored in many modeling details such as the relative abundance pattern of r-process elements, the value of $f_{\rm rp}$, the relative contribution of dynamical versus wind ejecta, and uncertainties in star formation history of the Universe.

## 3.7 Conclusions

In this Letter, we derived estimates for the dynamical ejecta mass produced by the BNS merger GW170817, as well as the corresponding kilonovae light curves and r-process nucleosynthesis yields, without additional photometric or EM spectral data. These estimates have the GW data as their foundation and use a fit to a wide variety of simulations to obtain dynamical ejecta masses from these data. Our predictions for light curves include a range of possible magnitudes and time scales of emission. In general, for the blue model in Metzger (2017) in the i-band, we predict peak magnitudes concentrated between $\sim 19$ and $\sim 17$ for a merger consistent with our low spin results, and peak magnitude between $\sim 19$ and $\sim 16$ — typically lasting twice as long — for mergers consistent with high spin results. Such predictions can guide expectation as to whether or not future, perhaps more distant, counterparts would be observable with a given facility. The predictions from the GW inference for the dynamically unbound matter depend strongly on the allowed spin configurations in the GW model, which in turn influence the predicted light curves. The low spin results predict smaller ejecta masses on the whole, and as such, a bright kilonova event (e.g., $> 16$ magnitude) may indicate a faster spinning NS component. We stress that the phenomenological fits used to predict $M_{\rm ej}$ themselves are not corrected for spin effects, so this increased brightness occurs because of degeneracies in the GW parameter estimates between spin and mass ratio.

We have also presented predicted light curves derived from other models in the literature.

Our results show that when large amounts of ejecta mass are allowed, the light curves have brighter peaks and are longer-lived. They differ in color evolution, however (compare DU17 and Wollaeger et al. (2017) for example) and electromagnetic observations combined with these curves could hint towards mixtures of different ejecta material compositions (Metzger, 2017). For example, strong emission observed in both blue and red bands could imply sectors of material containing both high and low electron fractions. However, the Metzger model, as implemented here, neglects post-merger wind effects, and in general, these conclusions only hold under the assumption that dynamical ejecta dominate the mass ejection.

Our results suggest that dynamical ejecta from rare NS mergers could be an important and inhomogeneous source of r-process elements in the galaxy (Ji et al., 2016; Beniamini et al., 2016). If more than $f_{\rm rp} \gtrsim 10\%$ of the mass ejected from mergers is converted to r-process elements, our prediction for average r-process density in the local universe is consistent with the Galactic abundance. Our approach does not address the contribution from winds, which could eject a substantial overall mass but may (Siegel & Metzger, 2017) or may not (Rosswog et al., 2017) have the wide range of $Y_e$ needed to produce all r-process abundances (i.e., the second and third r-process peak). Our approach is also not as detailed as full multi-species chemical enrichment calculations used to interpret observations of individual elements in targeted populations (see, e.g., Côté et al., 2017). As Advanced LIGO and Virgo approach design sensitivity, these observational constraints should rapidly shrink, enabling more precise tests of the BNS r-process nucleosynthesis paradigm. Additionally, present and future electromagnetic observations should provide complementary information to directly constrain those parameters that our analysis cannot.

Finally, if electromagnetic measurements are consistent with a total ejecta mass (dynamical and wind) of $\gtrsim 0.01 M_\odot$, and if we require consistency with low neutron star spins, then one possible conclusion is that winds contribute significantly to the total ejected mass. However, if winds dominate, then the dynamical ejecta mass will be an important but potentially difficult to measure component in the light curve, which our calculations can supply. Addi-

tionally, with so much material ejected per event, to be consistent with our inferred detection rate, we would predict that only a fraction of the ejecta can form r-process elements.

The coincidence of GW170817 and GRB170817A was an exceptionally rare event, allowing for a unique set of measurements to be made about the processes driven by the BNS merger. Future observations should facilitate the refinement of these measurements. The observation of GW170817 suggests that in the upcoming year-long third observing run (Abbott et al., 2016b) with a three instrument GW network, there will likely be more GW observations of BNS. In the coming years, GW measurements will allow for better understanding of populations of kilonova events.

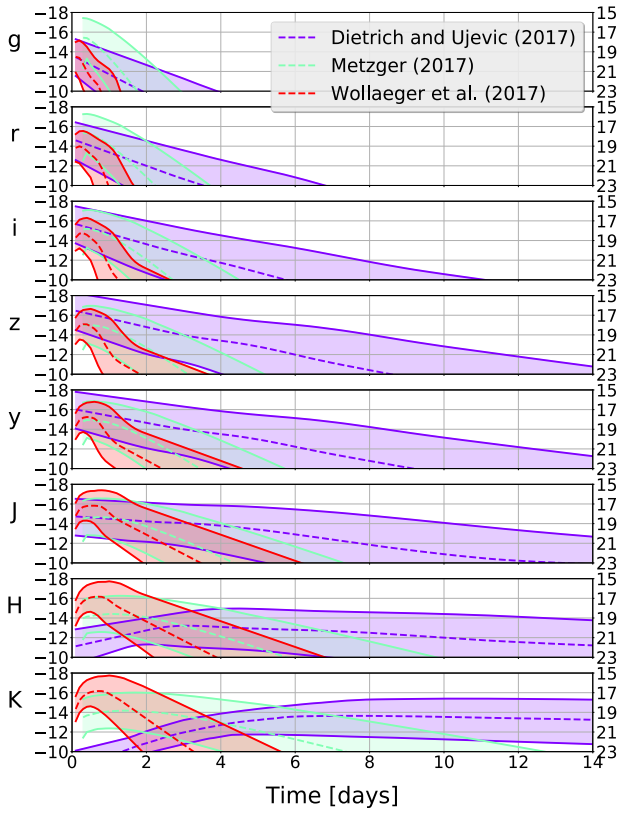


Figure 3.3: Absolute (left vertical axis) and apparent (right vertical axis) magnitudes of light curves consistent with parameter estimation for astrophysical spins for the kilonova models of DU17; Metzger (2017); Wollaeger et al. (2017) in $grizyJHK$ filters. In particular, the `DZ2` model is employed from Wollaeger et al. (2017). The dashed lines show the median light curve, while the shaded intervals show the 90% intervals. In addition to including the average relative error (72%) of the ejecta mass fitting formula, we include 1 mag errors on the intervals to account for errors in the models themselves (Coughlin et al., 2017). The lower percentiles are not conservative as we cannot definitively exclude zero ejecta mass due to unmodeled systematics. The fiducial distance to the event is 40 Mpc.

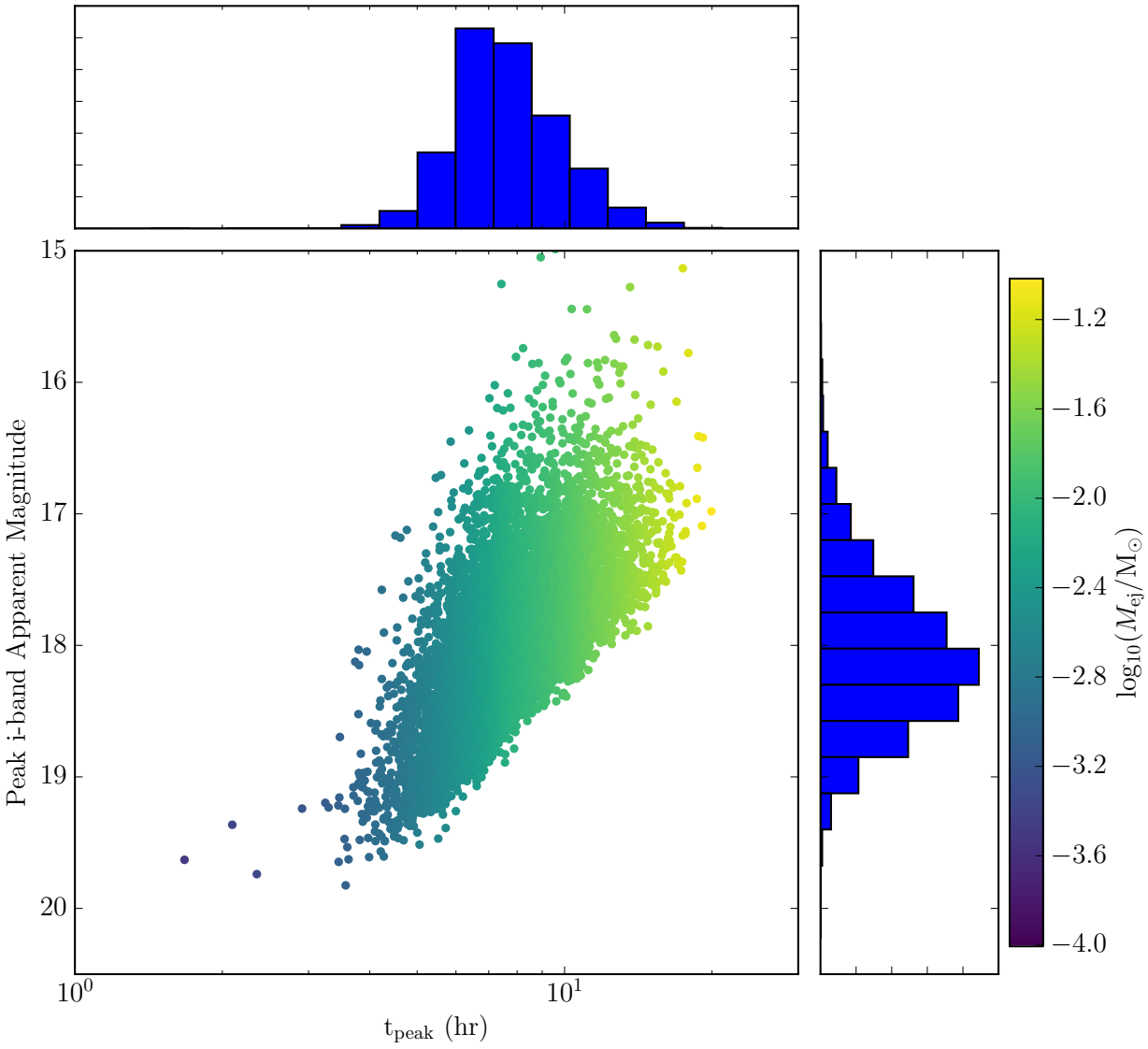


Figure 3.4: Inferred peak i-band apparent magnitude vs. time of peak i-band magnitude with the blue model in Metzger (2017) and low-spin sample distribution (marginal distributions on $M_{\rm ej}$and time of peak shown on top and right). Apparent magnitudes are calculated from the dynamical ejecta only, using the GW inferred distance.

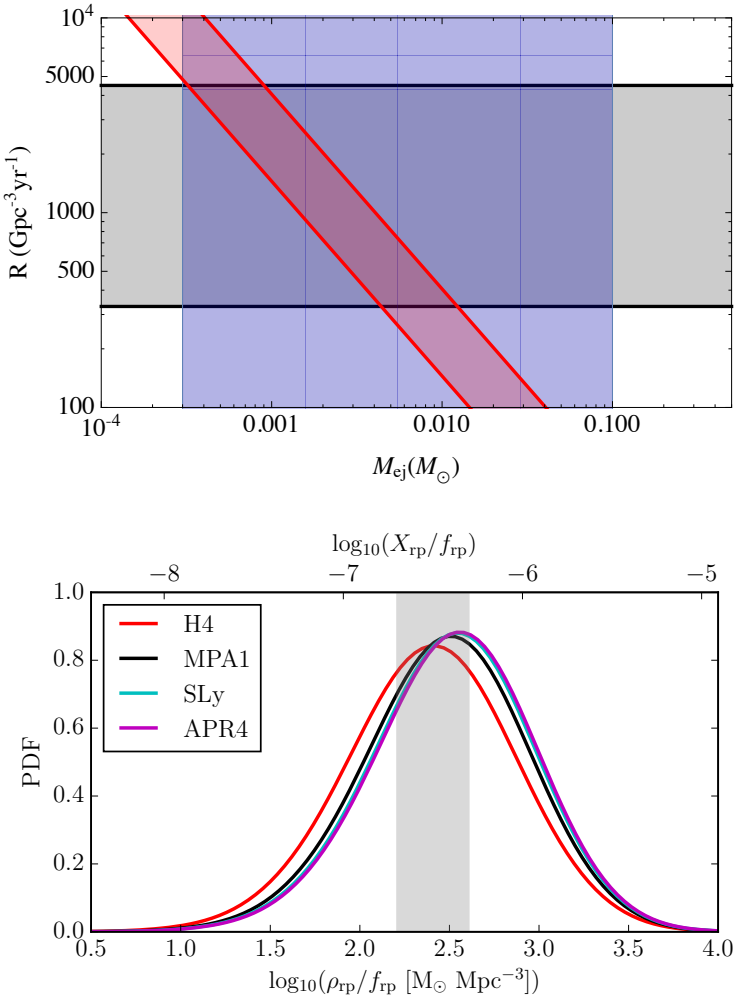


Figure 3.5: *Left panel*: Plot of the present-day BNS merger rate density $\mathcal{R}$ versus dynamical ejecta masses $M_{\rm ej}$. The solid gray band corresponds to the event rate range deduced from GW170817. The solid blue band shows the approximate range of conceivable dynamical ejecta masses, based on the ejecta models used in this work. The red band shows the approximate range of r-process elements per unit volume, based on Galactic observations, an approximate density of MW-like galaxies (0.01 Mpc$^{-3}$), a range of Galactic masses, and r-process formation efficiencies $f_{\rm rp}$ between 0.5 and 1. Configurations in the intersection of all three bands correspond to cases where dynamical ejecta from BNS mergers are solely responsible for r-process element formation. *Right panel*: Probability distributions of r-process material density and abundance (normalized by $f_{\rm rp}$) from dynamical ejecta for different EOS at $z = 0$. The lower (upper) bound on the 90% credible interval for $\rho_{\rm rp}/f_{\rm rp}$ over all EOS is $10^{1.7}$ M$_\odot$Mpc$^{-3}$ ($10^{3.2}$M$_\odot$Mpc$^{-3}$). The vertical gray band shows the Solar r-process abundance (Arnould et al., 2007).

# CHAPTER 4

# A SEARCH FOR KILONOVAE IN THE DARK ENERGY SURVEY

## 4.1 Summary

As discussed in the previous chapter, the detection of GW170817 and its electromagnetic counterparts provided confirmation that neutron-star mergers can power explosive optical/near-infrared transients known as kilonovae. Kilonovae had been well-studied prior to 2017 though, and interestingly, key work related to the r-process and neutron-star mergers had been done in the 20th Century by scientists affiliated with the University of Chicago including Maria Goeppert Mayer, David Schramm, and James Lattimer. With some theoretical guidance on kilonova observables, kilonova search programs were put together, which primarily focused on follow-up of short gamma-ray bursts or potential gravitational waves. However, one could in principle detect a neutron-star merger and its associated kilonova in a similar manner to supernova detection: search for transient point sources in deep optical images of the night sky without the use of external gamma-ray or gravitational-wave triggers. This chapter presents such a search using data from the Dark Energy Survey and is based on the publication *A Search for Kilonovae in the Dark Energy Survey*, which I led.

The Dark Energy Survey regularly observes 10 3-square-degree fields with the Dark Energy Camera in order to detect supernovae and measure cosmological parameters. We used this supernova dataset to search for kilonovae that exhibit features similar to a set of models from Barnes and Kasen (2013). By simulating Dark Energy Camera observations of kilonovae and supernovae, we calculated our search's kilonova detection efficiency and false positive rate. When applied to the Dark Energy Survey first-two-years dataset, our search algorithms yielded no excess of detections over the background number of supernovae expected. Given the non-detection of kilonovae, we set upper limits on the rate of kilonovae as a function of kilonova intrinsic brightness. This was the first time independent optical/near-infrared

limits had been set on kilonova emission.

The upper limits that our search set are consistent with the neutron-star merger rates inferred from the subsequent detection of GW170817 that occurred just months after the publication of this work. Additionally, our analysis of the Dark Energy Survey data highlighted some important considerations for both optical follow-up of gravitational waves and for future independent optical searches with e.g. the Large Synoptic Survey Telescope.

## 4.2 Introduction

The recent detections by LIGO of gravitational waves (GW) from binary black hole mergers (Abbott et al., 2016c,a) have motivated searches for electromagnetic (EM) counterparts to gravitational waves (Abbott et al., 2016b; Annis et al., 2016; Soares-Santos et al., 2016; Cowperthwaite et al., 2016). Theoretical and numerical studies suggest that outflows of energetic neutron-rich material during a binary neutron star (BNS) merger enable r-process nucleosynthesis (e.g. Li and Paczyński 1998). The decay of these r-process elements results in isotropic thermal emission and is called a "kilonova" (KN). Metzger and Berger (2012) compared different EM counterparts of GW sources and concluded that the kilonova has promising detectability with current instruments. Observations of KNe could constrain models of neutron star mergers, and an accurate redshift measurement would allow GW measurements to be used as cosmological distance probes (Schutz, 1986; Dalal et al., 2006). While properties of optical KN light curves remain uncertain, a number of models predict KNe to be dim ($M_i \sim -14$ mag), red ($i - z \sim 1$ mag), and short-lived ($\sim$1 week) (Barnes and Kasen, 2013; Tanaka and Hotokezaka, 2013; Metzger and Fernández, 2014; Lippuner and Roberts, 2015).

An interesting KN candidate was reported by Tanvir et al. (2013) and Berger et al. (2013). The candidate was first identified by the gamma ray burst GRB130613B, which triggered both the *Swift* Burst Alert Telescope and Konus-Wind. Following GRB130613B, the teams reported two epochs of Hubble Space Telescope observations in the V and H bands as well

as optical observations with the Inamori Magellan Areal Camera and Spectrograph (IMACS) and the Low Dispersion Survey Spectrograph (LDSS3) on the Magellan Telescopes. At 9.4 days after the GRB, the source was found to have $M_H \sim -15.2$ mag and $M_V \gtrsim -13.3$ mag, indicating a red $V-H$ color $\gtrsim 1.9$ mag. Other r-process events possibly involving a neutron star have been reported in Jin et al. (2016a), Jin et al. (2016b), and Ji et al. (2016).

Here we describe an independent search for KNe from the first two seasons of data from the Dark Energy Survey supernova program (DES-SN) (Bernstein et al., 2012; Diehl et al., 2016) using the Dark Energy Camera (DECam: Flaugher et al. 2015). KN searches triggered by GRBs or GWs will be an integral part of GW multi-messenger astronomy, but here we use existing DES-SN data to search for KNe without an external trigger. With four optical broadband filters, 30-square-degree coverage in the supernova fields, and ∼1 week cadence in each filter with excellent depth per visit, the DES-SN sample is well-suited for a KN search. Light curve simulations of KNe and supernova (SN) backgrounds are used to inform the analysis and selection criteria that we apply to the DES-SN data sample to look for KNe. The KN simulations are based on spectral energy distributions (SED) resulting from radiation transport calculations of KN-merger models from Barnes and Kasen (2013) (hereafter BK13). These calculations fold in r-process element opacities using atomic data for heavy elements rather than approximating their opacities with that of iron. As a result, BK13 predict KN light curves that are dimmer, redder, and longer-lived than those predicted with iron opacities. These models are still highly uncertain, so our analysis reports volumetric rate limits for each KN model and over a wide range of absolute brightnesses.

We make a preliminary estimate of our potential KN sensitivity using two approximate calculations based on previously published results. Both approximations assume a KN peak $i$-band magnitude of $M_i \sim -14$ mag and 100% search efficiency within the DES-SN limiting magnitudes and sky areas described in § 4.3. The first approximation assumes an optimistically large KN rate given by the LIGO 90% confidence upper limit on binary neutron star mergers of $1.26 \times 10^4$ yr$^{-1}$Gpc$^{-3}$ (The LIGO Scientific Collaboration et al., 2016).

With these assumptions, we would expect to find 1-2 KNe in our DES-SN sample. The second calculation is based on the hypothetical correspondence between KNe and short-hard gamma-ray bursts (SGRB, Paczynski 1986, Narayan et al. 1992). Assuming that the true event rate of SGRBs is the lower limit found in Fong et al. (2015) of 90 $\mathrm{yr}^{-1}\mathrm{Gpc}^{-3}$ and that 1/2 of SGRBs are associated with KNe, the probability of seeing a KN in our sample is $\sim 4 \times 10^{-3}$.

We present results from the first untriggered optical search for KNe, which is complementary to the LIGO search for BNS mergers based on predicted GW signals. LIGO directly probes for mergers by looking for characteristic "chirp" GW signals from sources in its detection volume (The LIGO Scientific Collaboration et al., 2016). The DES KN search described here is sensitive to optical emission from such mergers. While we do not discuss optical follow-up to LIGO triggers, the methods and results presented here will inform strategies and contaminant rejection techniques for follow-up (Soares-Santos et al., 2016). The outline of this paper is as follows. In § 4.3 and § 4.4, we describe the DES data sample and the simulations, respectively. § 5.4 details selection requirements and our KN search efficiency for different models of host galaxy noise. The results and discussion of the analysis are presented in § 5.5 and § 5.6, and we conclude in § 5.7.

## 4.3 DES-SN Data Sample

For DES-SN, the CTIO Blanco-4m telescope and DECam were used to make repeated observations of ten 3 $\mathrm{deg}^2$ fields. Each field was observed in *griz* bands with central wavelengths of 4830, 6430, 7830, 9180 Å, respectively. Eight of these fields were "shallow" fields with an average single-visit depth of $\sim 23.5$ mag in each band. The other two were "deep" fields with an average single-visit depth of $\sim 24.5$ mag in each band. The exposure time and number of exposures per visit are given in Table 1 of Kessler et al. (2015) (hereafter K15). The cadence was approximately one visit per week in each band and field. Transients were detected using the difference-imaging pipeline `DiffImg` described in K15 to find events with signal-to-noise

ratio above 5 that also pass automated scanning to reject subtraction artifacts (Goldstein et al., 2015). To select candidates for SN science, DES-SN requires a detection on two separate nights, mainly to reject asteroids[1]. However, *all* candidates with two detections are saved, even if the two detections are on the same night.

For our KN search, we consider events with $i$ and $z$ detections close in time, since models predict that KNe will have higher SNR in the DES $i$ and $z$ bands than in $g$ and $r$. Additionally, KNe are expected to fade on the timescale of repeat SN-field observations ($\sim$ 1 week). In the shallow fields, we require an $i$ and $z$ detection on the same night since all four (*griz*) bands are observed within $\sim$ 20 minutes anyway. In the deep fields, adjacent-night $i$ and $z$ detections are accepted since the four bands are not always observed on the same night. We define a *trigger* to be the first paired $i$ and $z$ detections of a candidate[2]. Longer time separations between $i$ and $z$-band detections are not considered as triggers since KNe are expected to dim significantly between repeated observations of the fields.

Using the first two seasons of DES-SN, these criteria result in 3487 triggers in the shallow fields, and 1236 triggers in the deep fields, most of which are SNe and asteroids. Our data sample includes PSF-fitted measurements of the flux and its uncertainty in all four bands and all epochs, the angular separation between $i$ and $z$-band detections on the trigger night (for asteroid rejection), the matched host galaxy, and its photometric redshift. The absolute photometric calibration of the sample has been validated at the 2% level (K15).

## 4.4 Simulations of DES Light Curves

Monte Carlo simulations of KN and SN light curves, using the `SNANA` software package (Kessler et al., 2009), are employed to tune selection criteria, to determine the KN search efficiency, and to predict backgrounds. For an arbitrary light curve model, the `SNANA` sim-

1. See sections 3.2.4 and 3.3 in K15 for more details about science candidates

2. Pairing requires that the RA and DEC of the $i$ and $z$ detections are within 1", and thus the angular (radial) separation extends out to $\sim \sqrt{2} \times 1$".

ulation uses the observing conditions (PSF, zero point, sky noise) at each DES-SN epoch and passband to generate a redshifted[3] flux and uncertainty that would have been measured by `DiffImg`. The `SNANA` simulation generates light curves at the catalog-level and does not use images. As described in section 7 of K15, `SNANA` also does not fully account for image subtraction artifacts on bright galaxies. To better characterize the KN efficiency, we perform an additional study of fake point sources overlaid on the search images near low-redshift galaxies and processed with `DiffImg`.

**Kilonovae:** We simulate KN light curves using rest-frame SED models from BK13. Their models have evolved since the start of this analysis (Barnes et al., 2016), but here we restrict ourselves to the original light curves of Barnes and Kasen (2013). The implications of the new models are discussed in § 5.7. The BK13 models are parameterized by the velocity $\beta = v/c$ and mass $M$ of the matter ejected from the merger. BK13 generate nine SEDs corresponding to $\beta = [0.1, 0.2, 0.3]$ and $M = [0.001, 0.01, 0.1]\ M_{\odot}$, which we use in our analysis. Calculated *griz* light curves for all nine models are shown in Fig 4.1. These light curves are generically brightest in the $z$ band, with decreasing brightness in bluer filters. At the time of maximum $z$-flux, the ratio of $g$, $r$, and $i$-band flux to $z$-band flux is 0.04, 0.2, and 0.5, respectively, when averaged over all nine models. Notably, the rest-frame brightness of these models spans 4-5 magnitudes, and the distribution over model parameters is unknown, adding to the uncertainty in overall KN detectability. In the `SNANA` simulation, KN light curve start times are randomly generated during the first season (Y1: $56534 < \mathrm{MJD} < 56698$) and second season (Y2: $56877 < \mathrm{MJD} < 57067$), and the redshift distribution is assumed to follow a constant co-moving volumetric rate. Our analysis considers KN light curves with and without host galaxy flux. When including host galaxy flux, the light curves are generated in the simulation such that the KN rate at a particular location is proportional to the background host brightness as described in § 4.5.4. To compute our sensitivity over

3. The `SNANA` simulations use a FLRW cosmology with $H_0 = 70$km/s/Mpc, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$.

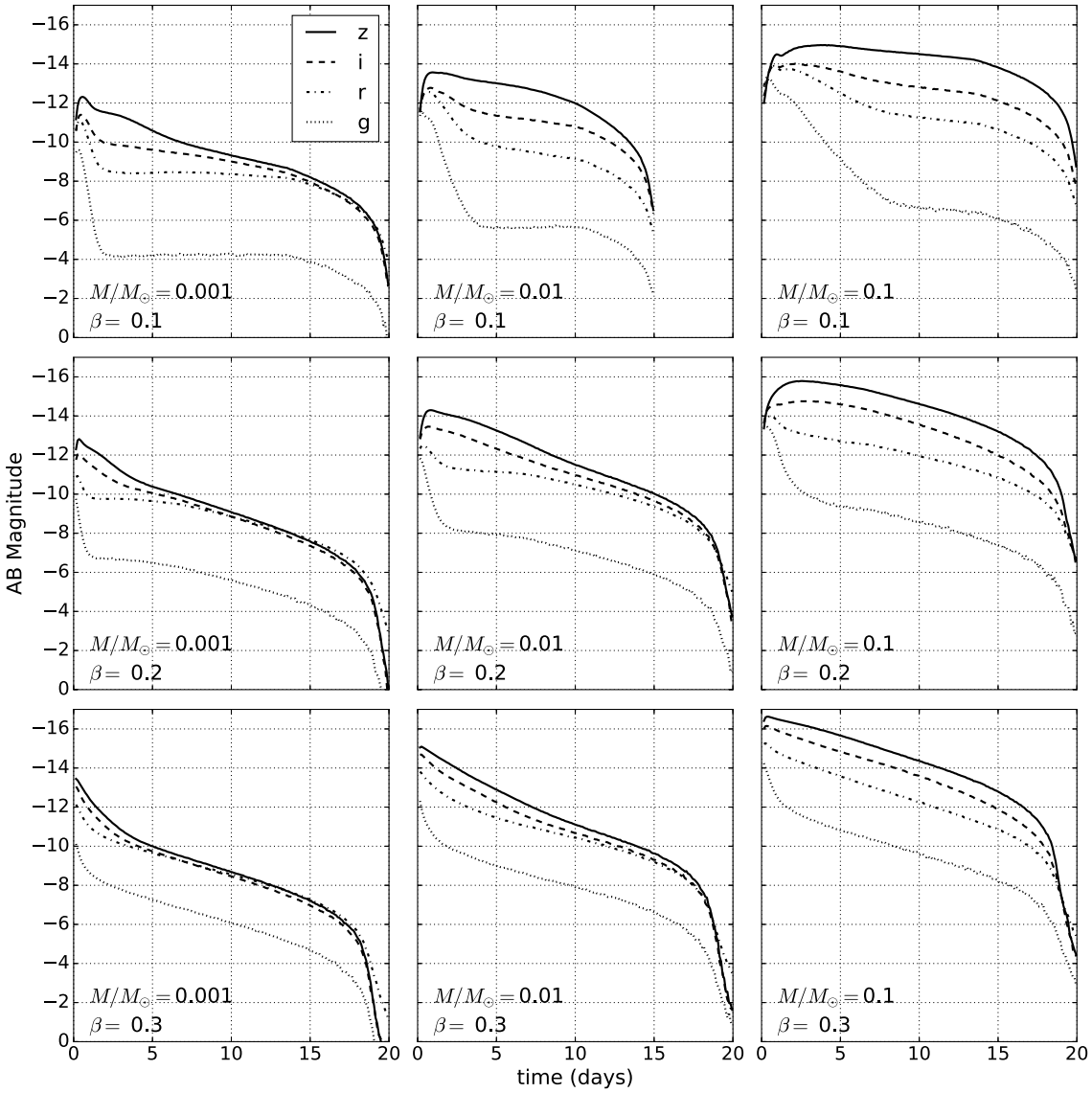


Figure 4.1: Computed *griz* broadband light curves from integrating the nine BK13 spectral energy distributions. The bottom left of each panel shows the BK13 model parameters.

a wide range of KN brightness, absolute magnitude offsets are applied as described in § 5.4 and § 5.5.

**Supernovae:** We perform simulations of Type Ia SNe using the SALT-II light curve model (Guy et al., 2010), the volumetric rate vs. redshift from Dilday et al. (2008), and the 'G10' intrinsic scatter model and the stretch and color populations from Kessler et al. (2013). For core collapse (CC) SNe we use CC templates based on IIbc, II-P, and IIn SNe as described

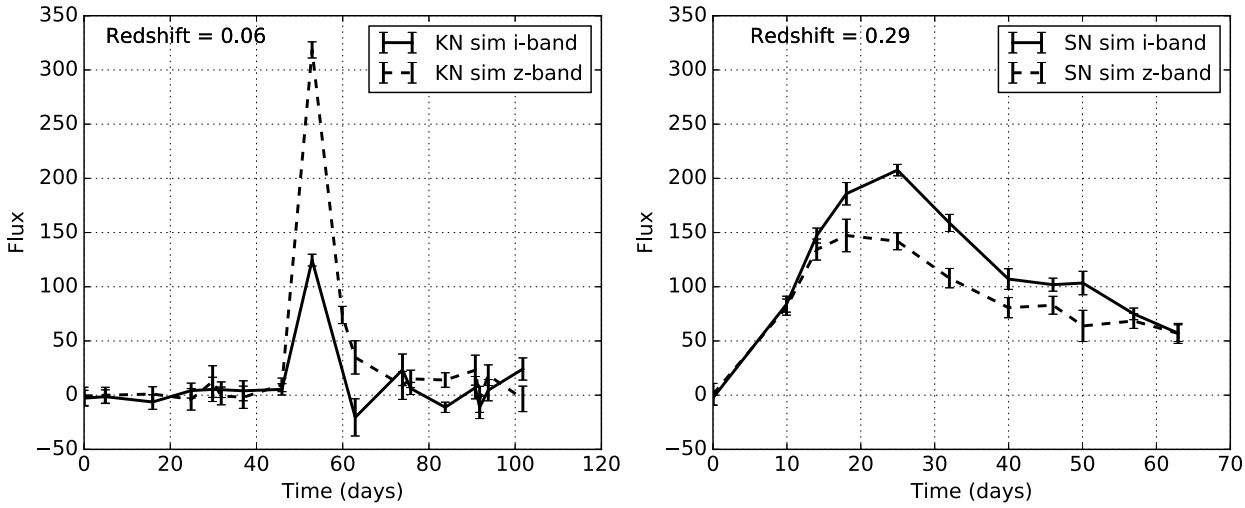


Figure 4.2: Observed KN and SN light curves in the $i$ and $z$ filters, as simulated with `SNANA`. The KN is based on the BK13 model with $\beta = 0.3$, $M = 0.1 M_{\odot}$ and redshift $z = 0.06$. The SNIa is simulated with SALT-II color $c = 0.03$, stretch parameter $x_1 = -0.75$, and redshift $z = 0.29$. Magnitudes are given by $27.5 - 2.5 \log_{10}(\text{Flux})$; e.g., the shallow-field detection limit of mag=23.5 corresponds to Flux=40. The error bars show the simulated flux and uncertainties for each observation; the lines connect these simulated points to guide the eye.

in Kessler et al. (2010) along with the rates from Li et al. (2011). Example $i$ and $z$-band light curves simulated for a typical KN and SNIa are shown in Fig. 4.2. In general, the SNe tend to be bluer and have longer timescales than KNe. SN times of peak flux are simulated in the ranges MJD = 56450 to 56740 for Y1 and MJD = 56840 to 57110 for Y2. Each window extends $\sim 2$ months before the start of the observing season, and $\sim 1$ month after the end. These extended time windows account for long rise and fall times of SN light curves as well as the $(1 + z)$ time dilation for higher redshift SNe. SNe are generated in the redshift range $z < 1.35$. The probability of matching to a host galaxy is taken from Bernstein et al. (2012)[4]. Note that the simulation does not account for SNe matched to incorrect hosts. For each simulated host galaxy, the associated photometric redshift ($z_{\text{phot}}$) is based on the photometric redshift distribution in the Science Verification (SV) catalogue described in Bonnett et al. (2015). To make SN background estimates with high statistical precision, we generate $3.7 \times 10^5$ SNIa and $4.3 \times 10^6$ CC SNe; this corresponds to 40 of our

4. See $m_i < 24$ and $\kappa_{Ia} = 0.5$ column of Table 18.

DES-SN data sets.

**Fakes on nearby galaxies:** Following K15, we place fake point sources ("fakes") of random magnitudes onto CCD images to determine the efficiency of `DiffImg` in identifying sources on low-redshift galaxies. While the `SNANA` simulation accounts for increased noise from bright galaxies, processing fakes with `DiffImg` includes unmodeled inefficiencies that are not characterized by our simulation. As described in K15, there exists a "surface brightness anomaly" which degrades the `DiffImg` search efficiency for events on nearby galaxies. Pan-STARRS1 has also seen a similar anomaly (see Fig. 6 in Rest et al. 2014). To determine the effects of this anomaly on our search, we compute and compare the detection efficiencies from the `SNANA` and `DiffImg`+fakes methods for a range of source magnitudes and background surface brightnesses ($m_{\rm SB}$), where $m_{\rm SB}$ is the magnitude per square arcsecond measured on the template image. These two methods are used to independently quantify the results of our KN search, which we present in § 5.5. Since our analysis relies on $i$ and $z$-band detections, the product of the $i$ and $z$-band efficiencies gives the trigger efficiency. Due to limited resources and a small number of low-redshift galaxies ($\sim 100$) in the SN fields, we only test fake foreground magnitudes from 21st to 25th magnitude, and we do not run the difference imaging pipeline on fake BK13 light curves themselves.

**Not Simulated:** Cowperthwaite and Berger (2015) (hereafter CB15) identify a number of objects other than SNe that could be KNe contaminants. These backgrounds are not simulated due to their low expected observation rate and are mostly removed with an $i-z$ color cut (see Fig. 3 in CB15). CB15 also found background objects with redder $i-z$ colors that could be consistent with the KN color: the Pan-STARRS fast transient PS1-13ess with $i-z=0.62$ mag (Drout et al., 2014) and Type ".Ia" SNe (Shen et al., 2010). Tables 1 and 2 in CB15 show that the observation rate of these contaminants is about two orders of magnitude lower than the SN Ia observation rate, but still enough to contaminate KN signals. Nevertheless, these objects fail our selection requirements as described in § 4.5.2.

## 4.5 Analysis

Our analysis extracts volumetric rest-frame rate limits from the DES-SN data sample using estimates of search time-volume and search efficiency. Following Dilday et al. (2008) (hereafter D08), the average KN rest-frame volumetric rate we infer is given by

$$R = N_{\rm KN}/[\widetilde{\epsilon_{\rm KN}VT}], \tag{4.1}$$

where $\widetilde{\epsilon_{\rm KN}VT}$ is the effective time-volume probed by the survey for a volume $V$, observation time $T$, and efficiency as a function of redshift $\epsilon_{\rm KN}(z)$. $\widetilde{\epsilon_{\rm KN}VT}$ is computed from D08:

$$\widetilde{\epsilon_{\rm KN}VT} = (\Theta T)\int_{z_{min}}^{z_{max}} dz\epsilon(z)u^2(z)\frac{du}{dz}\frac{1}{(1+z)} \tag{4.2}$$

Here, $\Theta$ is the solid angle probed by the survey in all fields and $u$ is the FLRW metric comoving distance. The KN efficiency $\epsilon(z)$ is defined as the fraction of simulated KNe at redshift $z$ passing the selection requirements described in § 4.5.1. Dependence of $\epsilon(z)$ on factors such as cadence and observing conditions are integrated out over the two DES seasons. We do not consider local density perturbations, which could affect volumetric rate estimates at low redshifts.

### *4.5.1 Selection Requirements*

Here we describe the selection requirements (cuts) applied to the data and simulated events. The cuts were designed to exclude SNe while maximizing KN efficiency.

1. A *KN trigger* requires paired $i$- and $z$-band detections, where a detection is described in K15. Both detections pass automated image scanning described in Goldstein et al. (2015). We also require that both detections have a signal-to-noise ratio (SNR) greater than 5. The $i$- and $z$-band detections must occur on the same night for a KN trigger except in the deep fields where adjacent night detections are allowed due to longer

deep-field exposure times. A candidate can only KN trigger once: the first paired $i+z$ detection of a candidate is taken as the KN trigger and subsequent $i+z$ detections of the same candidate are not considered additional triggers.

2. For each $i$- and $z$-band detection in cut 1, we require zero point (ZP) $>$ $\mathrm{ZP}_{\mathrm{median}}-0.2$ mag: $\mathrm{ZP}_{\mathrm{median}}$ is the median of the ZP distribution for KN trigger $i$ and $z$ detections, and it is calculated independently for the shallow and deep fields. This cut ensures reasonable atmospheric transparency during the KN trigger observations.

3. For each $i$- and $z$-band detection in cut 1, we require PSF full width at half maximum $< 2.0$ arcsec.

4. The ratio $g$-flux/$z$-flux $< 0.15$ on the KN trigger night. If the $i$ and $z$ bands are observed on adjacent nights (in the deep fields), the $g$-band flux on the earliest night of the KN trigger is used. Considering all nine BK13 models, the $g$-flux is at most 0.09 times the $z$-flux at time of peak $z$-flux. The cut is well above the maximum $g$-flux/$z$-flux to accept events with reasonably large Poisson fluctuations.

5. $r$-flux/$z$-flux $< 0.4$: Similar to the $g$-flux/$z$-flux cut except using the $r$ band rather than the $g$ band. Considering all nine BK13 models, the $r$-flux is at most 0.37 times the $z$-flux at time of peak $z$-flux.

6. There is at least one $i$ or $z$ band observation (regardless of SNR) 3 to 10 days[5] after the KN trigger: This ensures sufficient data to examine the light curve evolution.

7. There is at least one observation (independent of SNR) 2 to 14 days before the trigger: ensures the ability to identify (and reject) light curves that begin before the trigger.

8. There is at least one observation (independent of SNR) 20 to 100 days after the trigger: ensures the ability to identify (and reject) light curves that continue after any KN would have faded.

5. All temporal cuts are made in the observer frame.

9. There is no SNR $> 4$ observation 2 to 14 days before the trigger: rejects objects that are bright before the KN trigger.

10. There is no SNR $> 4$ observation more than 20 days after the trigger: We define the time difference between the last single-band detection and the KN trigger $\Delta t \equiv t_{\rm last} - t_{\rm trigger}$ and require $\Delta t < 20$ days to reject objects with long timescales.

11. Veto events matched to a host galaxy with $z_{\rm phot} > 0.3$: Since we are sensitive to KNe at low redshift ($z < 0.3$), we remove events associated with a high redshift galaxy. The requirement for matching a source to a host galaxy is a source-host galaxy angular separation $< 2''$ and a directional light radius separation of $d_{LR} < 2$ (Sako et al., 2014; Gupta et al., 2016). This matching requirement is more strict than the K15 requirement of $d_{LR} < 4$, and was adjusted to avoid too many false matches to hosts, which lowers the KN efficiency. If a source is matched to a host, it is vetoed if the host has $z_{\rm phot} > 0.3$. Given the DES-SN limiting magnitudes, transients detected with $z_{\rm phot} > 0.3$ have peak absolute magnitude brighter than $\sim -16.5$ mag in the deep fields and $-17.5$ mag in the shallow. Note that transients with no detected host galaxy will have no $z_{\rm phot}$ and thus this veto will not apply.

12. shape $< 0$. Here we define the 'shape' as the rate of change of the $z$-band flux (normalized to the KN trigger $z$-band flux) between the night of the KN trigger and the next $z$-band observation:

$$\text{shape} = \frac{1}{F_1} \frac{F_2 - F_1}{t_2 - t_1}, \tag{4.3}$$

where $F_1$ and $t_1$ are the $z$-band flux and MJD of the KN trigger, respectively. $F_2$ and $t_2$ are the flux and MJD of the next $z$-band observation after the KN trigger, respectively. This shape cut removes events that do not exhibit a declining light curve after the trigger.

13. *angular separation* ($\delta_{iz}$) between the KN trigger $i$ and $z$ observations of the transient

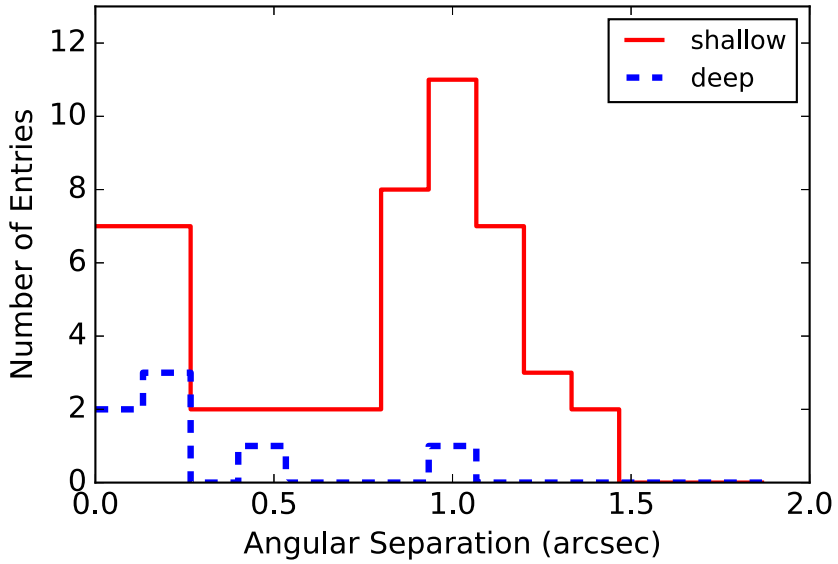


Figure 4.3: Distributions of the angular separation between KN trigger $i$ and $z$ observations $\delta_{iz}$ for the DES-SN data. The first ten cuts are applied. A few objects with outlying $\delta_{iz}$ are not shown on the plot.

$< 0.6$": This cut removes contamination from asteroids. The $\delta_{iz}$ distribution is shown in Fig. 4.3 for the deep and shallow fields after applying the first 10 cuts. The deep field exposure times for the $i$ and $z$ bands are 1800 and 3600 seconds respectively, long enough that moving asteroids fail the PSF-shape requirement. The deep field sample is thus dominated by non-moving transients, and the $\delta_{iz}$ distribution is peaked near 0.1", consistent with the astrometric precision. The shallow field exposure times are much shorter, 200 and 400 seconds for $i$ and $z$ band respectively, and thus slow-moving asteroids are included by the 1" trigger-matching requirement in `DiffImg`. Fig. 4.3 shows that the shallow field $\delta_{iz}$ distribution has two components: 1) a non-moving component at $\delta_{iz} \sim 0.1$ and 2) a moving component with more events as $\delta_{iz}$ increases.

Further evidence for asteroid contamination is the excess of events in the four shallow fields near the ecliptic where most asteroids orbit. After applying the first 10 cuts, there are a total of 42 events in these four fields, while the four shallow fields away

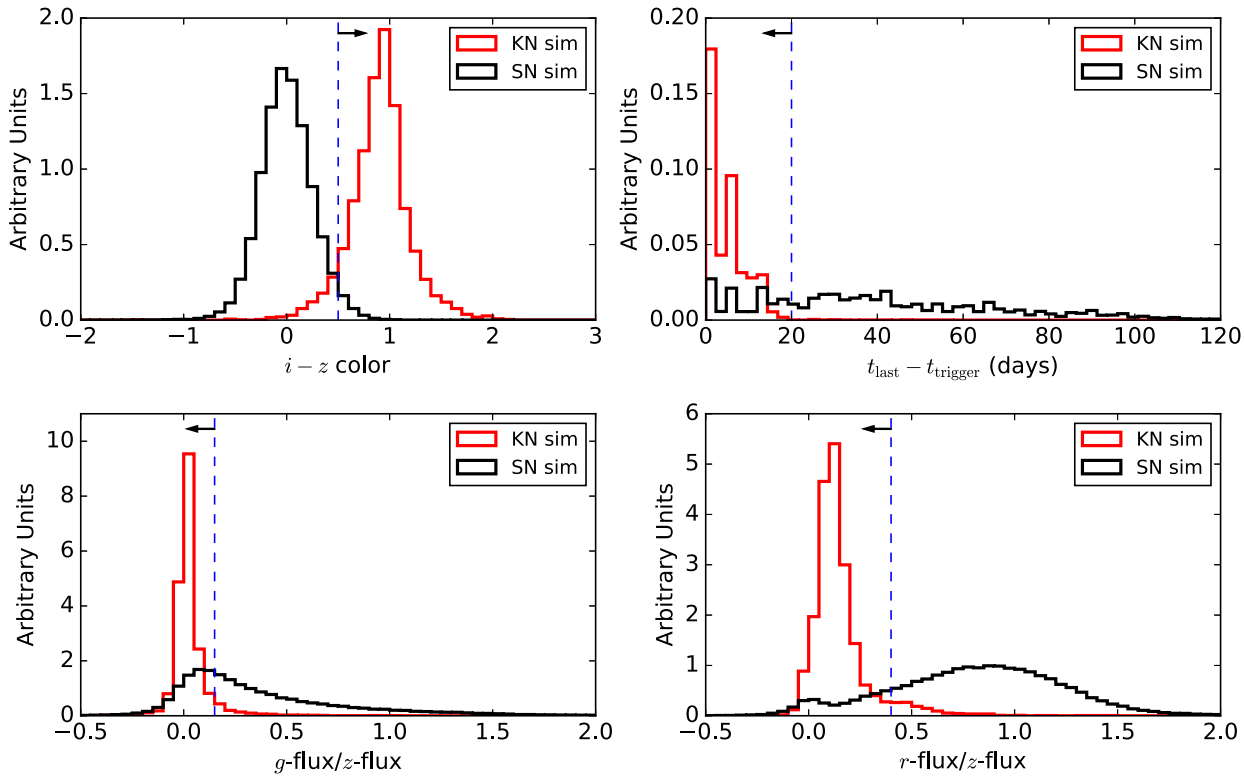


Figure 4.4: Simulated distributions of KN and SN triggers (cut 1). The dashed vertical lines show the values of the cuts, and the arrows show the selected sample. *Top left panel:* $i-z$ colors. *Top right:* Time between first trigger and last single-band detection. *Bottom left:* ratio of $g$-flux to $z$-flux for a trigger. Negative values are allowed, since negative fluxes can occur due to forced photometry. *Bottom right:* ratio of $r$-flux to $z$-flux for a trigger.

from the ecliptic have a total of 11 remaining events. In contrast, the two deep fields (one near the ecliptic and one away from the ecliptic) each contain the same number of events (4) after cut 10, showing that the deep fields are less susceptible to asteroid contamination.

14. KN trigger $i-z$ color $> 0.5$ mag: Motivated by the BK13 predictions that KNe are very red, we require $i-z > 0.5$ for the trigger bands.

We exclude one cut at a time from the analysis to determine the effectiveness of each cut. We find that the four cuts which maximally suppress SN background, in order of effectiveness, are cuts 9, 14, 10, and 11. Though the $g$- and $r$-band cuts (4 and 5) are not among the most effective cuts for SNe, we still require these color cuts to reject other backgrounds,

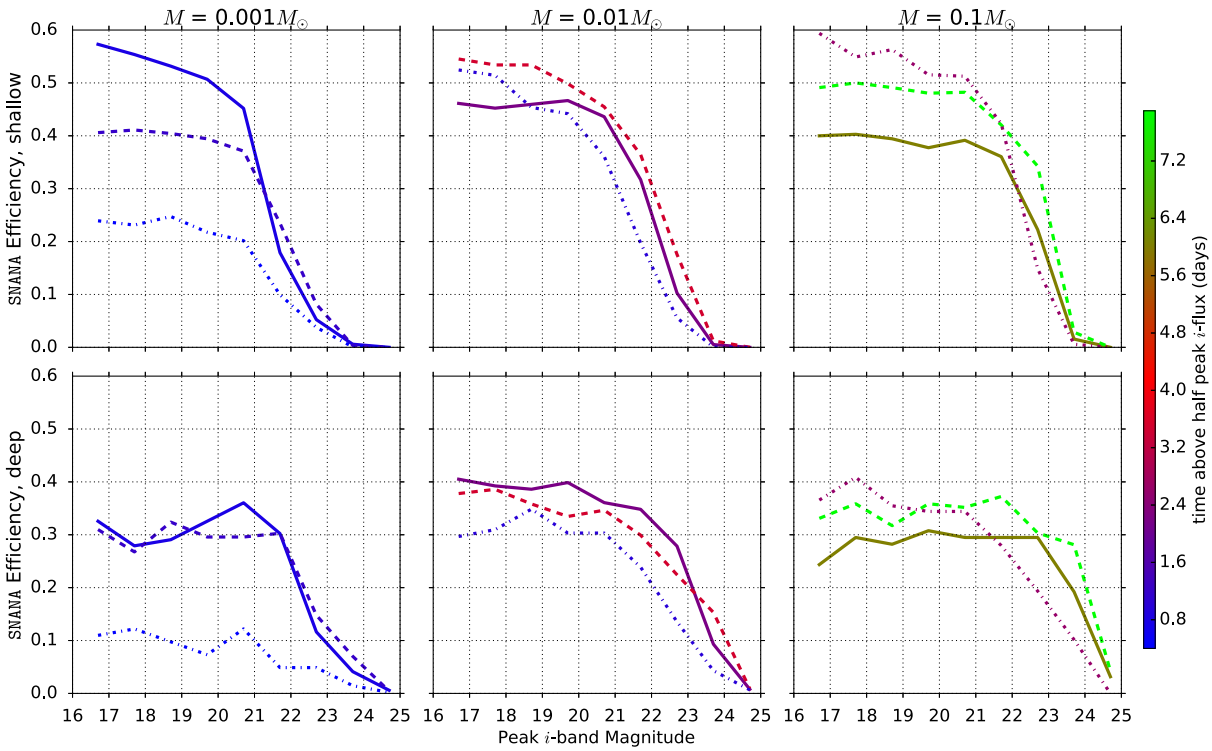


Figure 4.5: KN search efficiency as a function of peak apparent $i$-band magnitude for the BK13 models in the shallow and deep fields at fiducial redshift $z = 0.02$. Efficiencies do not account for host galaxy noise. The solid, dashed, and dashed-dotted lines correspond to BK13 models with $\beta = 0.1, 0.2, 0.3$, respectively. The colors of the lines represent each model's $t_{\rm half}$, the time above half maximum flux in the $i$-band.

as described in § 4.5.2. The simulated distributions of the three colors (cuts 4, 5, and 14) and $\Delta t$ (cut 10) are shown in Fig. 4.4 for KNe and SNe. The KN distributions shown here are based on a random assortment of the nine BK13 models placed uniformly in co-moving volume out to $z = 0.15$ in the shallow fields and $z = 0.2$ in the deep fields.

The number of data and simulated events remaining after each cut is shown in Table 4.1. No events from the DES-SN sample pass all of our cuts, and the SN simulation predicts $1.1 \pm 0.2$ total background events.

### *4.5.2 Efficiency for Transients that are Not Simulated*

The cuts are tuned solely on the SN simulations but are effective at removing other backgrounds as well. As mentioned in §4.4, Type ".Ia" SNe and Pan-STARRS fast transients

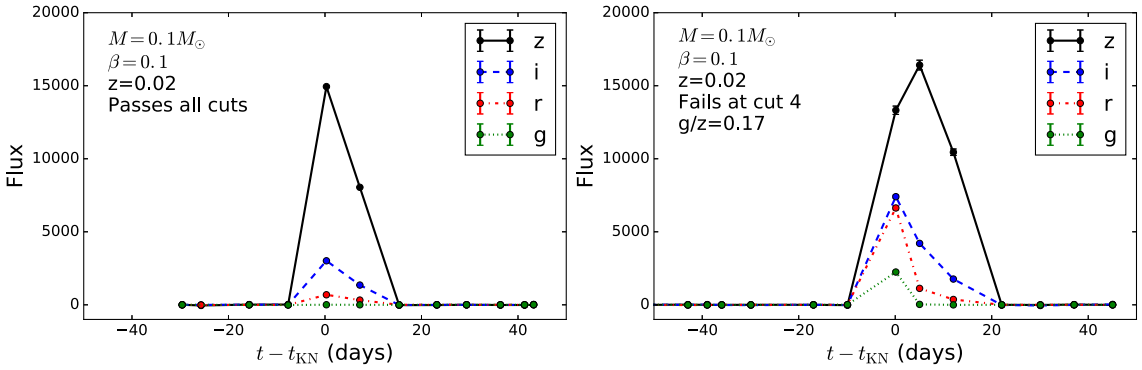


Figure 4.6: Example *griz* $\beta = 0.1$, $M = 0.1 M_\odot$ light curves at magnitude 18 that pass the cuts (left) and fail the cuts at cut 4 (right). The points show the observed fluxes and lines are drawn between them to guide the eye. Fluxes are defined such that magnitudes are given by $27.5 - 2.5 \log_{10}(\text{Flux})$; e.g., mag=22.5 for Flux=100.

like PS1-13ess have $i - z$ colors consistent with KNe. PS1-13ess is bright in the $g$-band relative to the $z$-band (Drout et al., 2014), so objects like it are rejected with a cut on the $g$ brightness (cut 4). ".Ia" models from Shen et al. (2010) produce spectra that are initially a blue continuum, but redden at late times as the model dims. Such light curves are removed with cuts 4, 5, and 9, which remove events with bright $r$ or $g$-band flux before or during the KN trigger.

### *4.5.3 KN efficiency with no host galaxy*

To calculate KN rates using Equations 4.1 and 4.2, the KN efficiency as a function of redshift or apparent magnitude is required. Fig. 4.5 shows the efficiency versus peak apparent magnitude for the nine BK13 models. To generate these curves, we vary the absolute magnitude of BK13 light curves and simulate them at a fiducial redshift of $z = 0.02$ with no host galaxy noise. Here, a fiducial redshift is used so that the dependence of efficiency on observed brightness can be compared between models, while being agnostic to each model's absolute magnitude. Our final rate calculations, however, use analogous efficiency curves and include redshifting based on absolute magnitudes from BK13. At bright apparent magnitudes, the efficiencies do not reach unity due to the DES-SN cadence and selection requirements.

To further illustrate selection effects, Fig. 4.5 shows the timescale of each KN model, parameterized by the time above half maximum flux in the $i$ band $t_{\rm half}$. Near the detection limit (mag 23.5 for shallow fields, mag 24.5 for deep), KNe with longer $t_{\rm half}$ are more detectable, because there are more chances to make a detection than for short-$t_{\rm half}$ KNe. However, this relation between efficiency and $t_{\rm half}$ does not hold at magnitudes much brighter than the detection limit. For a bright KN, there are multiple chances to detect the light curve before it falls below the detection threshold, so the selection requirements rather than $t_{\rm half}$ drive the efficiency for each model.

For comparison between models, Table 4.2 shows the efficiency of each $M = 0.1M_{\odot}$ model at magnitude 18 ($\sim$ 5 mag brighter than threshold) and $z = 0.02$ in the shallow fields with each cut. The cut on the ratio of g-flux to z-flux (cut 4) significantly reduces the $\beta = 0.1$, $M = 0.1M_{\odot}$ efficiency, showing how the interplay between color and cuts will affect the efficiency of bright sources. Although cut 4 diminishes the $\beta = 0.1$, $M = 0.1M_{\odot}$ efficiency, the cut value of $g/z = 0.15$ was chosen to maximize the signal over all nine BK13 models while removing the background SNe. Example 18th magnitude $\beta = 0.1$, $M = 0.1M_{\odot}$ light curves are shown in Fig. 4.6. The light curve in the left panel of Fig. 4.6 passes the cuts, while that in the right panel fails at cut 4 because of its high g-flux on the detection night. The discrepancy in measured g-flux between these two light curves is due to the timing of the observations with respect to the light curve start times, exemplifying the effect of the cadence on the search efficiency.

### *4.5.4 KN efficiency with underlying host galaxy*

Here we investigate two anomalous effects from difference imaging on bright galaxies that result in efficiency losses: 1) excess missed detections of point sources, and 2) excess flux scatter which affects analysis selection requirement efficiency. For the first effect, poor subtractions can result in KN events that are not detected or that result in detections which look like mis-subtractions and thus fail the automated scanning. To investigate the detection

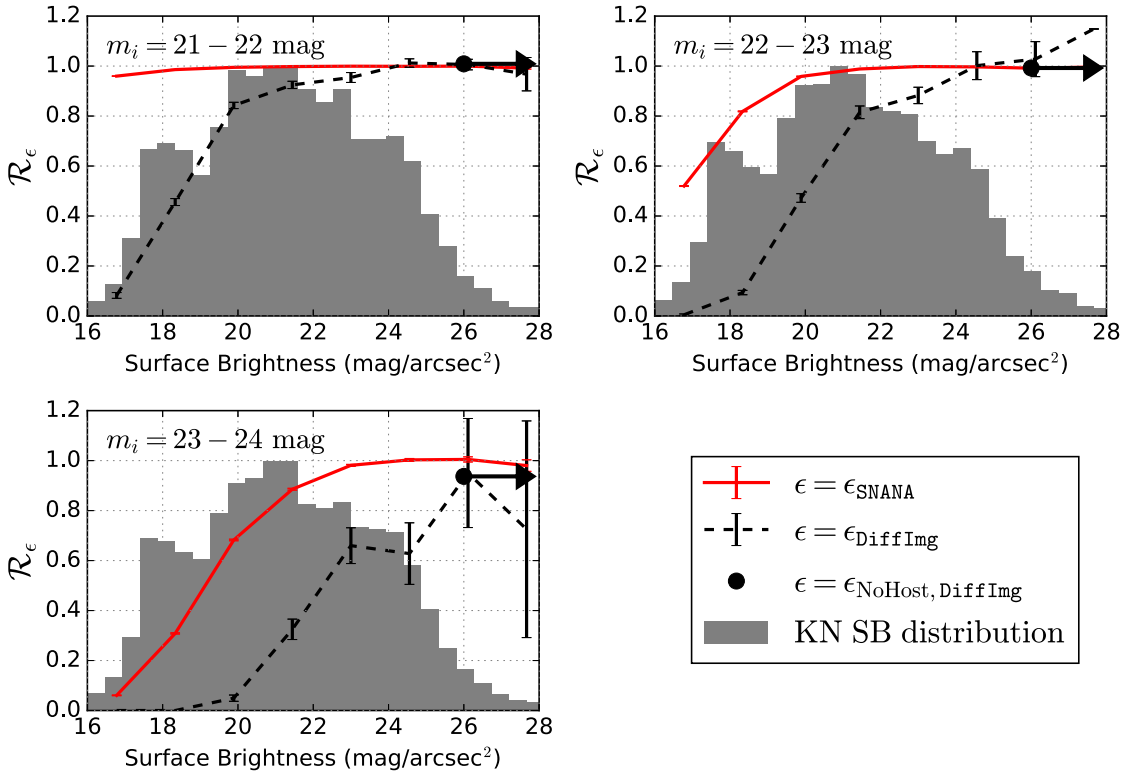


Figure 4.7: $\mathcal{R}_\epsilon$ in the DES-SN shallow fields vs. $m_{\rm SB}$ for KN $i$-band magnitudes (a) $m_i$ = 21-22, (b) 22-23, and (c) 23-24. The solid red line is based on the `SNANA` simulation, and the dashed black line is from fake point sources processed by `DiffImg`. The gray regions show the $i$-band surface brightness distribution in the shallow-field, calculated in `SNANA` from the DES+2MASS catalogue Sersic profiles. The black dot and arrow shows $\mathcal{R}_\epsilon$ for `DiffImg` fakes with $m_{\rm SB} > 26$, i.e. with no host galaxy. The error bars show $1\sigma$ uncertainties.

efficiency losses, we use fakes to measure single-band, single-epoch *detection* efficiencies for `DiffImg` ($\epsilon_{\texttt{DiffImg}}$[6]) as a function of background surface brightness ($m_{\rm SB}$). The single-epoch efficiency in the `SNANA` simulation ($\epsilon_{\texttt{SNANA}}$) accounts for host galaxy Poisson noise as well as excess flux scatter from the SB anomaly, but does not account for detection losses. We therefore use the `DiffImg` detection efficiencies, measured with fake sources on galaxies, to correct $\epsilon_{\texttt{SNANA}}$.

The second image-subtraction effect is from detected KNe with excess flux scatter. We characterize the impact on the analysis efficiency by including this excess scatter in the sim-

6. $\epsilon$ includes the host galaxy noise unless a "noHost" subscript is included

ulation (see Figs. 9-10 in K15). In this analysis, we do not consider the effect of correlations between measured fluxes in different bands and correlations between fluxes and detections. Additionally, our analysis does not account for the effects of image subtraction artifacts on $\delta_{iz}$.

To characterize the detection efficiencies with and without a host galaxy, and to compare `SNANA` and `DiffImg`, we define an efficiency ratio,

$$\mathcal{R}_\epsilon \equiv \epsilon/\epsilon_{\mathrm{NoHost,SNANA}} \tag{4.4}$$

where $\epsilon$ is the detection efficiency of either `SNANA` or `DiffImg` and $\epsilon_{\mathrm{NoHost,SNANA}}$ is the `SNANA` efficiency with no host galaxy. Since $\epsilon_{\mathrm{NoHost,SNANA}}$ is the highest possible efficiency, we expect $\mathcal{R}_\epsilon \leq 1$, except for statistical fluctuations. Fig. 4.7 shows the ratio $\mathcal{R}_\epsilon$ for the $i$-band shallow fields, where $\epsilon$ is 1) $\epsilon_{\mathtt{SNANA}}$, 2) $\epsilon_{\mathrm{NoHost,DiffImg}}$, and 3) $\epsilon_{\mathtt{DiffImg}}$. To compute our KN search sensitivity, we have similar information for the $z$-band and the deep fields.

The simulation of $\epsilon_{\mathtt{SNANA}}$ shown in Fig. 4.7 demonstrates that host Poisson noise has almost no effect on the detection efficiency for source magnitude $m_i = 21-22$ mag. However, near the detection limit ($m_i = 23-24$ mag), $\epsilon_{\mathtt{SNANA}}$ falls to half the no-host efficiency when the surface brightness reaches $m_{\mathrm{SB}} \sim 19$ mag/asec$^2$, which is comparable to sky noise of 19.6 mag/asec$^2$. The efficiency loss from host Poisson noise can not be mitigated, and represents the upper limit on the efficiency from `DiffImg`.

Fig. 4.7 also shows that the `DiffImg` efficiency with no host galaxy noise $\epsilon_{\mathrm{NoHost,DiffImg}}$ agrees with the no-host efficiencies predicted using `SNANA` ($\epsilon_{\mathrm{NoHost,SNANA}}$). For $m_{\mathrm{SB}} \gtrsim 26$, $\epsilon_{\mathtt{SNANA}}$ and $\epsilon_{\mathtt{DiffImg}}$ converge to $\epsilon_{\mathrm{NoHost,SNANA}}$, demonstrating that very faint backgrounds do not degrade efficiency. For brighter backgrounds, the `DiffImg` efficiency ($\epsilon_{\mathtt{DiffImg}}$) deviates from the `SNANA` efficiency ($\epsilon_{\mathtt{SNANA}}$) that simulates Poisson noise. At a given surface brightness, $\epsilon_{\mathtt{DiffImg}}$ is lower than $\epsilon_{\mathtt{SNANA}}$, because of the SB anomaly described in K15[7]. For large

7. Note that K15 characterized excess scatter in the measured flux, not degraded efficiency: since there are many opportunities to detect the light curve of a bright SNIa, SNIa detection efficiency is not affected

background brightnesses, the SB anomaly has a significant effect on the efficiency. For fakes with 21 mag $< m_i <$ 22 mag and 17 mag $< m_{\rm SB} <$ 18 mag/asec$^2$, $\sim 60\%$ are detected on the subtracted image, but only 2% of these pass automated image scanning because of poor quality subtractions. An example search and difference image for an undetected fake on a bright galaxy is shown in Fig. 4.8. This particular fake has an $i$-band magnitude of 21.2 mag and $m_{\rm SB} = 17.6$ mag/asec$^2$. The subtracted image shows a dipole structure and fails the automated scanning.

To estimate the total search sensitivity degradation from host galaxy noise, the efficiency is weighted by the KN distribution of background surface brightness $m_{\rm SB}$[8]. This distribution has not been predicted, so here we generate KNe at surface brightnesses following the Sersic profiles of the "DES+2MASS" and SV galaxy catalogues. This procedure assumes that the KN rate density follows the background surface brightness. The DES+2MASS catalogue includes galaxies from the 2MPZ catalogue within 200 Mpc (Bilicki et al., 2014). We use the following information from the galaxy catalogues: sky coordinates, redshift, $i$- and $z$-band magnitude, absolute magnitude, and shape profile. The gray-shaded region in Fig. 4.7 shows the $i$-band surface brightness distribution for KNe derived from the DES+2MASS+SV galaxy catalogue Sersic fits assuming the KN volumetric rate follows galaxy profiles. While the $m_{\rm SB}$ distribution depends on the choice of galaxy catalog and model assumptions, the efficiency curves are much less sensitive to these choices. Additionally, we have not investigated potential $\delta_{iz}$ artifacts from image subtractions.

We next compute the average $\epsilon_{\texttt{DiffImg}}/\epsilon_{\texttt{SNANA}}$ (ie. excess detection efficiency loss) as a function of KN magnitude weighted by the $m_{\rm SB}$ distribution. For $m_{\rm SB}$ below and above where $\epsilon_{\texttt{DiffImg}}$ is known, the efficiency ratio is set to 0 and $\texttt{max}(\epsilon_{\texttt{DiffImg}}/\epsilon_{\texttt{SNANA}})$, respectively. We then use `SNANA` to simulate analogous efficiency curves to those in Fig. 4.5, this time including host galaxy Poisson noise and SB-anomaly excess flux scatter. These new efficiency

by the SB anomaly.

8. Uncertainties on efficiency are shown in Fig. 4.7 for illustrative purposes, but are not propagated through the analysis since host galaxy model uncertainties dominate.

Figure 4.8: Search (left) and difference (right) images for an undetected fake with source magnitude $m_i = 21.2$ mag and $m_{\rm SB} = 17.6$ mag/asec$^2$. The fake is shown at the center of the green circle. Each image is 2.3'×2.3'.

curves are then down-weighted by the excess detection efficiency loss to give the overall KN efficiency as a function of KN magnitude. Since the brightest fakes in the `DiffImg` test are $m_i = 21$ mag, we linearly extrapolate $\epsilon_{\texttt{DiffImg}}/\epsilon_{\texttt{SNANA}}$ to mag 16 where $\epsilon_{\texttt{DiffImg}}/\epsilon_{\texttt{SNANA}}$ is assumed to be 1. The recomputed efficiencies which account for the SB anomaly are used to calculate KN volumetric rates in § 5.5.

## 4.6 Results

### *4.6.1 Event selection and contamination*

The number of events passing each cut for the data, simulated KNe, and simulated Ia and CC SNe are shown in Table 4.1. For the simulations, efficiencies are shown as well. The KN efficiencies in Table 4.1 are based on a different simulation than the 18th magnitude KN efficiencies in Table 4.2. Table 4.1 shows the efficiencies for BK13 models simulated randomly in co-moving volume out to redshift 0.15 in the shallow fields and 0.2 in the deep fields with no host galaxies. In comparison to Table 4.2, the KN efficiency here is much lower (2%) because the simulated redshift range goes beyond the detectable distance of most BK13 models. Fig. 4.9 shows the efficiency per redshift bin ($\mathrm{d}\epsilon/\mathrm{d}z$) and cumulative efficiency ($\epsilon$) as

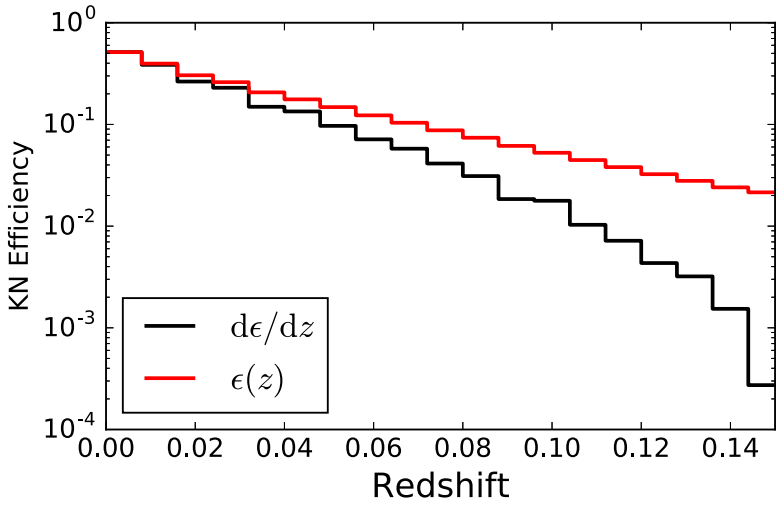


Figure 4.9: Efficiency per redshift bin $d\epsilon/dz$ and cumulative efficiency $\epsilon(z)$ in the shallow fields.

a function of redshift for this KN simulation in the shallow fields. The cumulative efficiency falls off as the redshift range is increased, showing that the redshift range and KN absolute magnitude are driving the overall efficiency shown in Table 4.1. At redshifts near 0, the efficiency plateaus at $\sim 0.35$ rather than 1 because 30% of the low-redshift KNe do not meet the KN trigger requirement, and there is added loss from the other selection requirements. At $z = 0.15$ the cumulative efficiency in Fig. 4.9 matches the KN efficiency shown in the shallow column of Table 4.1.

During the first two years of DES-SN, 3487 and 1236 KN triggers were identified in the shallow and deep fields, respectively. As mentioned earlier, asteroids contaminate the shallow fields, so the SN simulations under-predict the number of data triggers. The deep-field triggers however are well described by the SN simulations. Fig. 4.10 shows the distributions of host galaxy $z_{\rm phot}$ and $i-z$ color for the SN simulations (normalized to 2 seasons) and the data in the deep fields. The simulated distributions agree well with those from the data and show that the SN simulations predict the non-asteroid background levels. After all cuts, no events remain, which is consistent with our simulation prediction of $1.1 \pm 0.2$ SN events: 0.96 in the deep fields and 0.17 in the shallow fields. There is a 33% chance of finding 0 events

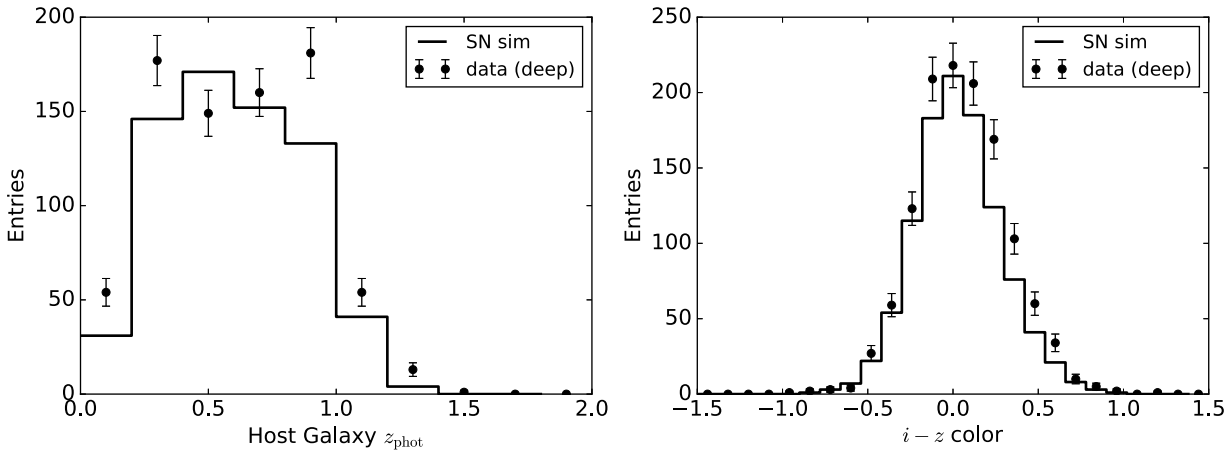


Figure 4.10: Left: for objects matched to a host galaxy, the distribution of deep field KN trigger $z_{\rm phot}$ is shown for the data (solid circles) and the SN simulation (histogram). Right: Distribution of KN trigger i-z color. The simulation has been re-scaled by 1/40 to correspond to two DES seasons.

in the sample given the SN prediction.

While no events pass our cuts in this search, SNe could contaminate future KN searches (e.g. Cowperthwaite et al. 2016). Fig. 4.11 shows example light curves for simulated CC and Ia which pass all the cuts. About 40% of the simulated SN background are CC, suggesting that Ia and CC SNe contribute similarly to the KN-search background. The Ia and CC contaminants have mean redshifts in the simulation of 1.1 and 0.8, respectively, but do not have measured $z_{\rm phot}$ values, and thus can not be vetoed by cut 11. Fig. 4.12 shows the Ia and CC efficiency as a function of simulated redshift. At low redshift, the SNe have low efficiency because they exhibit blue light curves and their light curves pass the detection threshold well past 20 days and thus fail the veto (cut 10). However, sufficiently redshifted SNe can take on the red colors of KN light curves and pass cuts 4, 5, and 14. These simulated contaminants illustrate that high-redshift SNe with no photometric redshift are the primary background for KN searches.

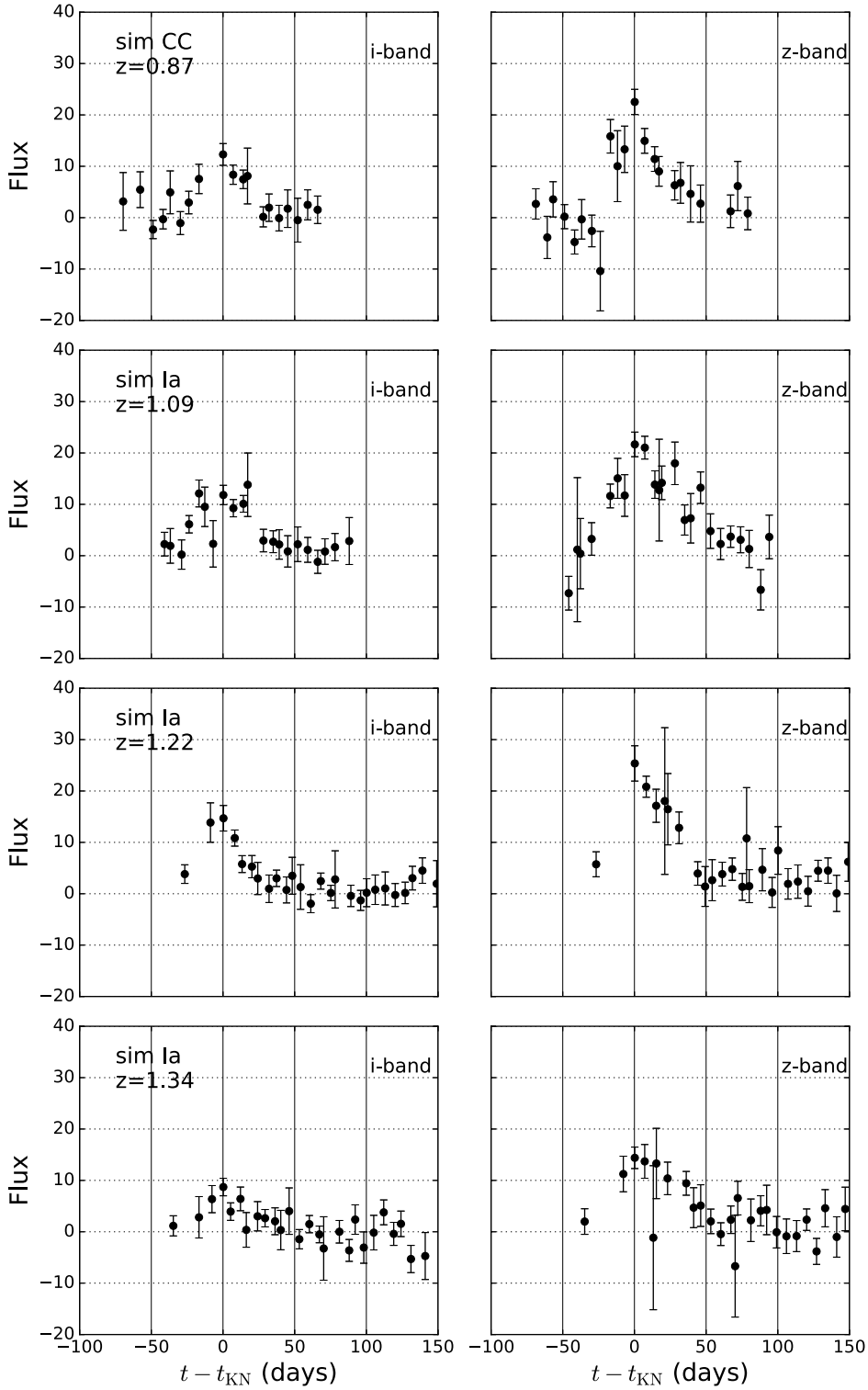


Figure 4.11: Example $i$ and $z$-band light curves for simulated CC and Ia which passed all cuts. Fluxes are defined such that magnitudes are given by $27.5 - 2.5\log_{10}(\text{Flux})$; e.g., mag=25 for Flux=10.

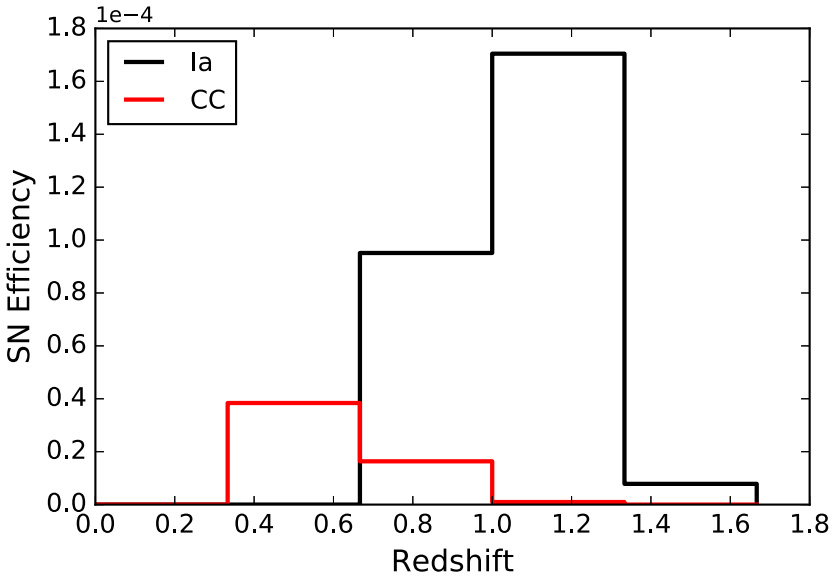


Figure 4.12: Simulated efficiency of Ia and CC SNe as a function of redshift. Note that the y-axis is scaled by $10^{-4}$.

### *4.6.2 Rate limits*

Here we calculate 90% upper rate limits using Equations 4.1 and 4.2, efficiencies from § 5.4, and two different assumptions: no host galaxy noise and host galaxy noise including the detection and selection-requirement efficiency loss from the SB anomaly. For these rate calculations, efficiencies differ from those presented in § 5.4, as redshift effects are included rather than simulating at the fixed redshift $z = 0.02$. The first assumption with no host galaxy corresponds to each BNS system being ejected from its host galaxy or occurring in a faint galaxy. For this assumption, we use $\epsilon_{\mathrm{NoHost,SNANA}}$ (Fig. 4.5) to calculate the rate limits. The red points in Fig. 4.13 show the upper rate limits for the nine BK13 models assuming all BNS systems are "kicked" out of galaxies. For the brightest model, the upper rate limit is $2.4{\times}10^4$ Gpc$^{-3}$yr$^{-1}$, while for the dimmest it is $1.0{\times}10^7$ Gpc$^{-3}$yr$^{-1}$. We also compute rate limits as a function of absolute magnitude for BK13 models, which are shown as black lines in Fig. 4.13.

Next we consider the second assumption where all BNS systems remain in the galaxy and are distributed based on background SB, whose distribution is that described in § 4.5.4. We

use $\epsilon_{\texttt{DiffImg}}(z)$ rather than $\epsilon_{\mathrm{NoHost},\texttt{SNANA}}(z)$ for detections and include the excess flux scatter effects in the selection-requirement efficiency. In this case, the limits are $5.2{\times}10^{4}\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ and $4.6{\times}10^{7}$ $\mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$ for the brightest and dimmest BK13 models, respectively. Fig. 4.14 is analogous to Fig. 4.13, but shows the rate limits for KNe in galaxies and includes detection and selection-requirement efficiency loss from Poisson noise and the SB anomaly. The rate limits on all nine models with and without host galaxy noise are given in Table 4.3. The KN rate upper limits with host galaxy noise are $\sim$ 3 times higher than the upper limits for KNe with no detectable host. The SB anomaly is largely responsible for the decreased sensitivity: the rates calculated with just host Poisson noise are only 10% higher than with no host galaxy, showing the SB anomaly has a significant impact on the KN search.

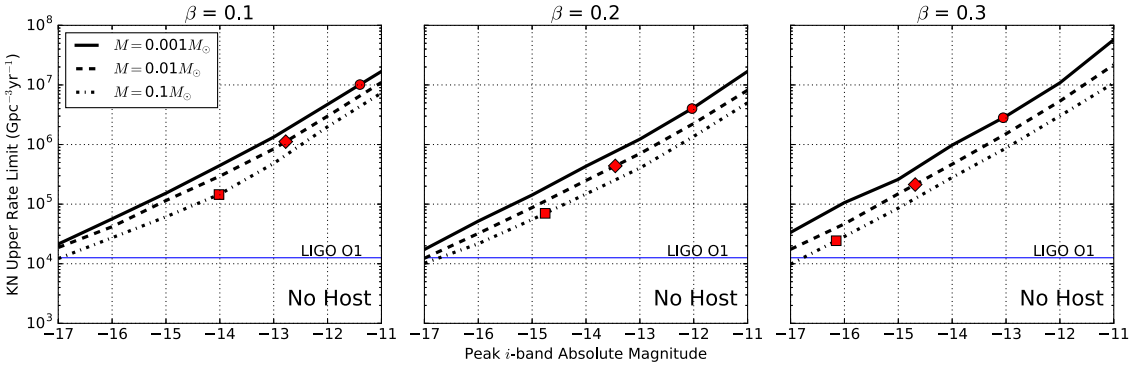


Figure 4.13: The 90% upper rate limit for the nine BK13 models using efficiencies calculated with `SNANA`. Each BK13 model is offset in absolute magnitude and the search efficiency is determined to calculate the rate. The red points show the rate limits for each model with no magnitude offset. The blue line shows the upper limit set in Advanced LIGO O1 (The LIGO Scientific Collaboration et al., 2016).

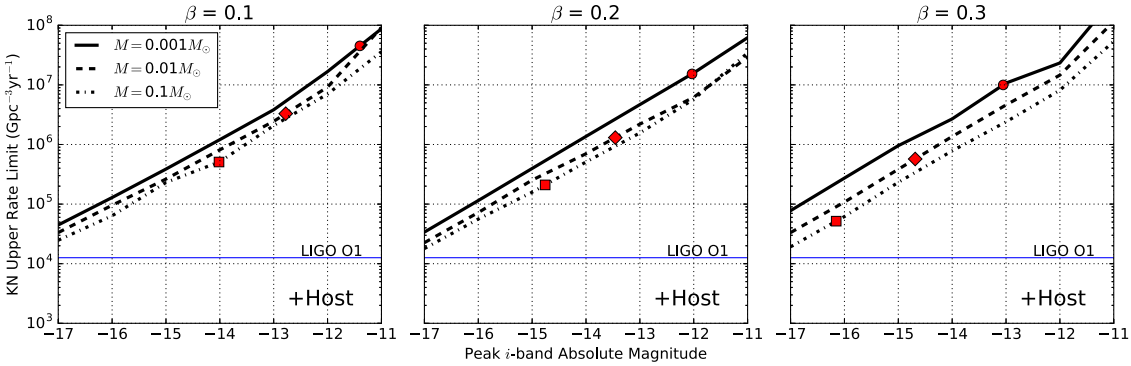


Figure 4.14: The same as Fig. 4.13, except using `DiffImg` efficiencies accounting for the SB anomaly.

Table 4.1: Number of Events and Simulated Efficiencies for each Selection Requirement

| | Shallow Fields | | | | | Deep Fields | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Cuts | Data | sim SNa | $\text{eff}_{\text{CC}}$b | $\text{eff}_{\text{Ia}}$b | $\text{eff}_{\text{KN}}$bc | Data | sim SNa | $\text{eff}_{\text{CC}}$b | $\text{eff}_{\text{Ia}}$b | $\text{eff}_{\text{KN}}$bc |
| 1. i+z trigger | 3487 | 1214 | 0.01 | 0.28 | 4.3 | 1236 | 1062 | 0.045 | 0.89 | 3.9 |
| 2. ZP $<$ $\text{ZP}_{\text{median}}$ | 3300 | 1172 | 0.01 | 0.27 | 4.1 | 1180 | 1010 | 0.043 | 0.85 | 3.7 |
| 3. trigger PSF($i$,$z$) $<$ 2.0 " | 3074 | 1139 | 0.0098 | 0.27 | 3.9 | 1176 | 1004 | 0.042 | 0.84 | 3.6 |
| 4. gflux/zflux $<$ 0.15 | 550 | 237 | 0.0019 | 0.057 | 3.7 | 407 | 562 | 0.019 | 0.53 | 3.4 |
| 5. rflux/zflux $<$ 0.4 | 209 | 34 | 0.0005 | 0.0058 | 3.6 | 213 | 280 | 0.009 | 0.27 | 3.3 |
| 6. observed 3-10 days after trigger | 195 | 28 | 0.0004 | 0.0045 | 3.1 | 169 | 214 | 0.0069 | 0.21 | 2.9 |
| 7. observed 2-14 days before trigger | 172 | 22 | 0.0003 | 0.0033 | 3 | 152 | 189 | 0.0061 | 0.18 | 2.8 |
| 8. observed 20 to 100 days after trigger | 164 | 20 | 0.0003 | 0.0032 | 2.6 | 148 | 175 | 0.0055 | 0.17 | 2.7 |
| 9. no SNR $>$ 4 observed 2 to 14 days before trigger | 61 | 5.1 | $8\times10^{-5}$ | 0.0008 | 2.6 | 24d | 42 | 0.0014 | 0.041 | 2.5 |
| 10. no SNR $>$ 4 observed 20 days after trigger | 53 | 4.0 | $6\times10^{-5}$ | 0.0007 | 2.5 | 8 | 11 | 0.0004 | 0.0096 | 2.5 |
| 11. veto $z_{phot}$ $>$0.3 | 51 | 1.9 | $3\times10^{-5}$ | 0.0004 | 2.3 | 2 | 6.2 | 0.0002 | 0.0058 | 2.2 |
| 12. shape $<$ 0.0 | 47 | 1.4 | $2\times10^{-5}$ | 0.0002 | 2.2 | 2 | 3.9 | 0.0001 | 0.0035 | 2.2 |
| 13. $\delta_{iz}$ $<$ 0.6 arcsec | 12 | 1.4 | $2\times10^{-5}$ | 0.0002 | 2.2 | 1 | 3.9 | 0.0001 | 0.0035 | 2.2 |
| 14. i-z color $>$ 0.5 | 0 | 0.17 | $3\times10^{-6}$ | $2\times10^{-5}$ | 2.2 | 0 | 0.96 | $3\times10^{-5}$ | 0.0009 | 2.1 |

Since 40 SN data sets are simulated for better statistics, the numbers of simulated SN events are divided by 40 to correspond to one DES-SN data set. The statistical uncertainty in each sim SN line is $\sqrt{N_{\text{SN}}/40}$, where $N_{\text{SN}}$ is the number of simulated events on a given line. $\text{eff}_{\text{CC}}$, $\text{eff}_{\text{Ia}}$, and $\text{eff}_{\text{KN}}$ are the simulated efficiencies of CC SNe, SNIa, and KNe, respectively. They are quoted in percent. The KN efficiencies are based on a random assortment of BK13 light curves generated randomly in co-moving volume out to redshift 0.15 in shallow fields and 0.2 in deep fields assuming no host galaxy noise. The KN efficiencies shown here are small due to the large redshift range of the simulation. The 2.8$\sigma$ difference between the SN simulation prediction and data in Cut 9 (deep fields) is not unexpected: We estimate a $\sim$ 30% chance of there being such a deviation in the table by performing a Monte Carlo simulation of data values in the table given the numbers from the SN simulation.

Table 4.2: KN efficiency for peak $m_i = 18$. BK13 $M = 0.1M_\odot$ models simulated in shallow fields with `SNANA`, where each model is scaled to have peak $i$ band magnitude $m_i = 18$.

| | KN efficiency for | | |
|---|---|---|---|
| Cut | $\beta = 0.1$ | $\beta = 0.2$ | $\beta = 0.3$ |
| 1. i+z trigger | 0.99 | 0.99 | 0.99 |
| 2. ZP $<$ $\text{ZP}_{\text{median}}$ | 0.88 | 0.90 | 0.88 |
| 3. trigger PSF($i$,$z$) $<$ 2.0 arcsec | 0.77 | 0.77 | 0.78 |
| 4. gflux/zflux $<$ 0.15 | 0.64 | 0.72 | 0.78 |
| 5. rflux/zflux $<$ 0.4 | 0.59 | 0.72 | 0.78 |
| 6. observed 3 to 10 days after trigger | 0.55 | 0.69 | 0.73 |
| 7. observed 2 to 14 days before trigger | 0.53 | 0.66 | 0.71 |
| 8. observed 20 to 100 days after trigger | 0.47 | 0.60 | 0.64 |
| 9. no SNR $>$ 4 observed 2 to 14 days before trigger | 0.45 | 0.58 | 0.63 |
| 10. no SNR $>$ 4 observed $>$ 20 days after trigger | 0.44 | 0.58 | 0.62 |
| 11. veto $z_{phot}$ $>$0.3 | 0.41 | 0.53 | 0.55 |
| 12. shape $<$ 0.0 | 0.40 | 0.50 | 0.55 |
| 13. $\delta_{iz}$$<$ 0.6 arcsec | 0.40 | 0.50 | 0.55 |
| 14. i-z color $>$ 0.5 mag | 0.40 | 0.50 | 0.55 |

Table 4.3: The 90% upper rate limits for the BK13 models

| 2*$M/M_\odot$ | 2*$\beta$ | 2*peak $M_i$ | upper rate limit ($\text{Gpc}^{-3}\text{yr}^{-1}$) | |
|---|---|---|---|---|
| | | | no host | + host |
| 0.001 | 0.1 | $-11.4$ | $1.0 \times 10^7$ | $4.6 \times 10^7$ |
| 0.001 | 0.2 | $-12.0$ | $4.0 \times 10^6$ | $1.5 \times 10^7$ |
| 0.001 | 0.3 | $-13.1$ | $2.8 \times 10^6$ | $1.0 \times 10^7$ |
| 0.01 | 0.1 | $-12.8$ | $1.1 \times 10^6$ | $3.3 \times 10^6$ |
| 0.01 | 0.2 | $-13.5$ | $4.4 \times 10^5$ | $1.3 \times 10^6$ |
| 0.01 | 0.3 | $-14.7$ | $2.1 \times 10^5$ | $5.7 \times 10^5$ |
| 0.1 | 0.1 | $-14.0$ | $1.5 \times 10^5$ | $5.1 \times 10^5$ |
| 0.1 | 0.2 | $-14.8$ | $7.0 \times 10^4$ | $2.1 \times 10^5$ |
| 0.1 | 0.3 | $-16.2$ | $2.4 \times 10^4$ | $5.2 \times 10^4$ |

## 4.7 Discussion

The upper limits on the KN rate set in our analysis are a few orders of magnitude above other estimates. The LIGO Collaboration set an upper limit on the BNS merger rate of $1.26 \times 10^4$ $\mathrm{yr}^{-1}\mathrm{Gpc}^{-3}$ (The LIGO Scientific Collaboration et al., 2016), and Fong et al. (2015) estimate a true SGRB rate of $270^{+1580}_{-180}$ $\mathrm{yr}^{-1}\mathrm{Gpc}^{-3}$. Although the limits presented in this work are not the most stringent, our analysis is independent of and consistent with other experiments and it directly sets limits on KN-like optical emission.

A key finding in this work is that the SB subtraction anomaly and SN background signals are major limitations in using DECam to search for KNe. Based on our results, the efficiency loss from underlying host galaxies reduces the KN search sensitivity by a factor of 3 relative to KNe with an undetected host. Further study is also required to determine the sensitivity reduction within the LIGO 80-170 Mpc range (Abbott et al., 2016d). A small part of this reduced sensitivity is from Poisson noise, and cannot be reduced. The major source of degradation, however, is from the difference imaging software. This software was developed and optimized several years ago to discover high-redshift SNe, with little motivation to find faint sources on bright galaxies. With new motivation to find EM counterparts to LIGO-triggered GW events, the software optimizations need to be revisited with the goal of improving the KN efficiency.

The cuts needed to adequately remove SNe and other backgrounds remove about half of the KN triggers, as shown in Tables 4.1 and 4.2. Our analysis reduced the predicted SN background to $1.1 \pm 0.2$ events in our sample, but this background will increase with higher search sensitivity or with looser cuts. While the MJD range of SN explosions will be reduced in follow-up to LIGO triggers, analysis cuts would be relaxed for two reasons. First, to maximize area coverage, only $i$ and $z$ band were used in Soares-Santos et al. (2016) (no $g, r$), which removes veto cuts 4 and 5 from § 4.5.1. Second, there is no pre-trigger veto unless there are serendipitous observations such as in Cowperthwaite et al. (2016). They detected a red, rapidly declining source in the follow-up observations of GW151216 (Abbott

et al., 2016a) with DEcam, but the source was vetoed since the Pan-STARRS Survey for Transients had identified it 94 days before the GW event. These relaxed cuts are likely to increase backgrounds.

Our analysis also establishes the need for robust asteroid rejection. The $\sim$ 5-minute separation between $i$ and $z$ exposures in our shallow-field sample (75 minutes for deep fields) is adequate for removing asteroids. To maximize the area of future GW follow-up observations, shorter $i$ and $z$ exposure times may be used, and therefore consecutive $i+z$ exposures would be separated by less than the 5 minutes in our DES-SN sample. To reduce asteroid contamination, a minimum time between $i$ and $z$ exposures should be considered, such as observing several $i$-band pointings before repeating in the $z$-band. The additional telescope slews may cost more overhead than the extra filter changes, and thus EM programs would benefit from a more rigorous analysis of observation strategies.

For future surveys searching for KNe, our analysis has a few important implications. One implication is an estimate of the KN rate limits that could be set with other surveys. The Large Synoptic Survey Telescope (LSST), for example, will image $\sim$ 1000 times more sky area than DES-SN, suggesting that LSST could further constrain the rate limits presented here if the LSST cadence is comparable to that of DES-SN (LSST Science Collaboration et al., 2009). Our analysis also shows that multi-band observations will be essential for robust background rejection. Lastly, further optimizations to the survey cadence and temporal spacing of exposures could be made to maximize the gains from a KN search. In particular, a faster cadence would maximize the chance for detection of short-lived KNe, and a minimum time between the $i$ and $z$ exposures is needed to limit asteroid contamination.

## 4.8 Conclusion

We have presented a search for KNe in the Dark Energy Survey supernova fields. Simulations of kilonovae and supernovae were performed to tune cuts, determine search efficiency, and assess background levels. In our analysis of the first two DES-SN seasons, we find no KN

candidates and set the first untriggered optical search limits on the rate of BK13-like KN events. The most serious issue for our search is a factor of $\sim 3$ loss in sensitivity due to difference-imaging for faint sources on bright galaxies. For the brightest KNe considered in this analysis, our limits are comparable to the limits set by the LIGO collaboration from gravitational wave observations (The LIGO Scientific Collaboration et al., 2016). During the course of our analysis, Barnes et al. (2016) updated the BK13 KN light curve models to account for the efficiency with which radioactive decay products thermalize the ejecta. The light curves for these updated models (BK16) are still characteristically red, but are fainter compared to BK13, especially at late times. At peak, the BK16 fiducial model bolometric luminosity is roughly half of that predicted without accounting for thermalization efficiency, which sets BK16 rate limits from our search $\gtrsim 3$ times higher than the BK13 limits. Since the BK16 models dim faster than those of BK13, a BK16 search in our data would suffer additional sensitivity loss from missed detections. Like BK13, the BK16 models are brightest in the infrared, so the basic methods and cuts presented herein are still applicable to a BK16 model search, though re-optimization of the cut values could be performed. This work sets the stage for further KN searches with DECam and other large-field-of-view telescopes such as Pan-STARRS1 or Large Synoptic Survey Telescope.

## 4.9 Acknowledgments

We would like to thank Dan Kasen for providing us with the BK13 light curves. We also gratefully acknowledge support from the Kavli Institute for Cosmological Physics at the University of Chicago. Z.D. is supported by the NSF Graduate Research Fellowship Program, grant DGE-1144082. R.J.F. is supported in part by NSF grant AST-1518052 and from a fellowship from the Alfred P. Sloan Foundation.

Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education

Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, Texas A&M University, and the OzDES Membership Consortium.

The DES data management system is supported by the National Science Foundation under Grant Number AST-1138766. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2012-39559, ESP2013-48274, FPA2013-47986, and Centro de Excelencia Severo Ochoa SEV-2012-0234. Research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) including ERC grant agreements 240672, 291329, and 306478.

This work was supported in part by the Kavli Institute for Cosmological Physics at the University of Chicago through grant NSF PHY-1125897 and an endowment from the Kavli Foundation and its founder Fred Kavli.

# CHAPTER 5

# A SEARCH FOR OPTICAL EMISSION FROM BINARY-BLACK-HOLE MERGER GW170814 WITH THE DARK ENERGY CAMERA

## 5.1 Summary

Do mergers of two stellar-mass black holes emit optical light? In this chapter, we set out to test this hypothesis by optically following up gravitational wave GW170814 with the Dark Energy Camera. This chapter is based on a publication by the Dark Energy Survey and others, A Search for Optical Emission from Binary-Black-Hole Merger GW170814 with the Dark Energy Camera, for which I was the primary author.

GW170814 was the first detection of a gravitational wave with a non-negligible signal in three detectors (LIGO Livingston, LIGO Hanford, and Virgo). Since a third interferometer was online, the direction of incidence of the gravitational wave was much better triangulated than what was possible with only two detectors. Additionally, the small region of the sky from which GW170814 likely came was in the middle of the region surveyed by the Dark Energy Survey. This had the added benefit that there were existing images of the area to which we could compare our follow-up images.

A day after the GW170814 detection, we began taking images of a large portion of GW170814's likely sky origin, corresponding to an almost 90% chance of pointing at the gravitational-wave source. This probability surpassed what any previous optical binary-black-hole follow-up programs had achieved. In order to maximize the sky area we covered with our images, we only took images in one filter and sacrificed the distinguishing power that comes with having measurements in multiple filters. The search images were compared with older Dark Energy Survey images of the region to find objects that were changing in brightness.

We applied a set of criteria to each transient point-source in our images to pare down the list of candidates in our search to only the most interesting. In order to do this in an unbiased manner, we applied our criteria to a small control sample of data before finally unblinding the entire dataset and performing the full analysis. Of the $\sim$ 60k candidates in our search images, our analysis identified two candidates that were interesting as potential gravitational-wave counterparts. Further investigation of these objects showed that neither could be related to GW170814: one candidate was a variable star and the other was either a variable star or a very high-redshift galaxy beyond the GW170814 distance. The non-detection of optical emission from GW170814 in our search disfavors models of binary-black-hole emission that fades in a few days after the merger.

## 5.2 Introduction

Since the first binary black hole (BBH) merger detection in September, 2015 (Abbott et al., 2016), mergers of two black holes have become a mainstay of gravitational-wave (GW) astrophysics. The first five observed BBHs, found only by the Hanford and Livingston Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors, offered significant astrophysical insight into the BBH mass distribution and event rates (Abbott et al., 2016; Abbott et al., 2016a; Abbott et al., 2016a; Abbott et al., 2016c, 2017a,b). For electromagnetic (EM) follow-up, however, the two LIGO detectors alone place poor constraints on the sky position, typically a few hundred deg$^2$.

To date, no compelling optical counterparts to BBH mergers have been identified. However, Stalder et al. (2017) found optical candidate ATLAS17aeu in their follow-up of GW170104 and hypothesize a chance coincidence. Additionally, a weak gamma-ray burst in coincidence with GW150914 was reported in Connaughton et al. (2016), but its association with GW150914 is still under dispute. There are three (not mutually exclusive) reasons for non-detections: (1) the probable sky regions of previous BBH detections were not searched comprehensively, (2) the BBH emission could not be identified or distinguished from background

transients, and/or (3) optical emission from BBH mergers is non-existent or below the detectable threshold at the times of the existing observations. Theoretical models have been proposed which could produce EM signals (e.g. de Mink & King, 2017; Loeb, 2016; Perna et al., 2016; Stone et al., 2017; McKernan et al., 2018), but these models are highly speculative. With little theoretical guidance, there is a need for more complete searches for BBH EM emission while also controlling false-positive event rates. Detection of BBH EM counterparts would be of immense scientific value (e.g. Phinney, 2009), as it could constrain the formation environments of BBHs, the behavior of matter in strong field gravity, and cosmological parameters such as the Hubble constant[1].

Thus far, a number of optical follow-up campaigns have been conducted to search for BBH counterparts (e.g. Smith et al., 2018; Stalder et al., 2017; Smartt et al., 2016b; Cowperthwaite et al., 2016; Soares-Santos et al., 2016; Yoshida et al., 2017; Morokuma et al., 2016; Smartt et al., 2016a; Lipunov et al., 2017). However, the large probable sky areas of the "double-coincident" LIGO detections (Hanford and Livingston detectors only) curtailed searches for EM counterparts from BBH mergers. For example, Soares-Santos et al. (2016) observed 102 $\mathrm{deg}^2$ of the GW150914 high-probability sky region with the optical imager Dark Energy Camera, (DECam: Flaugher et al., 2015), corresponding to 38% of the initial LIGO sky map probability. After accounting for the lack of existing images (templates) for difference-imaging, a shift in the sky map in a reanalysis of LIGO data, and other efficiency losses, only 3% of the probable GW150914 sky area was searched and analyzed. Similarly, the DECam follow-up campaign of GW151226 reported in Cowperthwaite et al. (2016) covered 29 $\mathrm{deg}^2$, just $\sim 2\%$ of the final GW151226 high probability region. In contrast, with the three-detector network including the Virgo interferometer, the smaller 28-$\mathrm{deg}^2$ 90% localization region of neutron-star merger GW170817 enabled 81% DECam coverage of the final LIGO-Virgo sky map and identification of the EM counterpart (Abbott et al., 2017; Soares-Santos et al.,

1. Even without an optical counterpart to a BBH, it is possible to measure the Hubble constant with a BBH GW sky map and galaxy catalog as in e.g. Schutz (1986); Chen et al. (2018); The DES Collaboration et al. (2019).

2017). These searches were all performed in the $i$ and $z$ bands, requiring two tilings of the search area. We note that these DECam searches attempted to tile maximal sky map probability, but for the nearby events such as GW170817, targeting based on galaxy catalogs can be successful (e.g. Coulter et al., 2017; Valenti et al., 2017; Arcavi et al., 2017).

On August 14, 2017, the LIGO-Virgo Collaboration (LVC), with the addition of the Virgo detector, made the first "triple-coincident" detection of GWs from a BBH event, GW170814, and provided a much tighter constraint on the sky position of the source than those of previous BBH detections (Abbott et al., 2017c; The LIGO Scientific Collaboration & the Virgo Collaboration, 2018). The detection of GW170814, with its 87-deg$^2$ 90%-localization region, enabled our team to perform a comprehensive search of the sky area for BBH merger optical counterparts and significantly improve our sensitivity to BBH merger EM emission models.

We report on our search for optical counterparts to GW170814 using the Dark Energy Camera. In §5.3, we describe the parameters and cadence of our follow-up observations, which extended to 12 days after the GW170814 trigger and covered 225 deg$^2$. §5.4 describes the analysis. Finally, §5.5 presents the results of the analysis, which we then comment on in §5.6 and §5.7.

## 5.3 Search and Light Curves

On August 14, 2017 at 10:30:43 UTC, the LVC reported a signal consistent with the inspiral and merger of two black holes of masses $30.5^{+5.7}_{-3.0} M_\odot$ and $25.3^{+2.8}_{-4.2} M_\odot$ at a luminosity distance of $540^{+130}_{-210}$ Mpc and redshift[2] $z = 0.12^{+.03}_{-.04}$ (Abbott et al., 2017c). LIGO and Virgo sent out a `Bayestar` sky map 2 hours after the trigger (LIGO-Virgo Collaboration, 2017a; Singer & Price, 2016) and we captured our first DECam image of the probability region at 06:00 UTC on August 15, 19.5 hours after the GW detection. DECam is an optical imager, installed on the Blanco 4-m telescope at the Cerro Tololo Inter-American Observatory. It has a 3-deg$^2$

2. Assuming cosmology of Planck Collaboration (2016)

field of view and is equipped with several broadband optical/NIR filters ($u,g,r,i,z,Y,VR$), making it well-suited to search for faint transients over large sky areas (Flaugher et al., 2015).

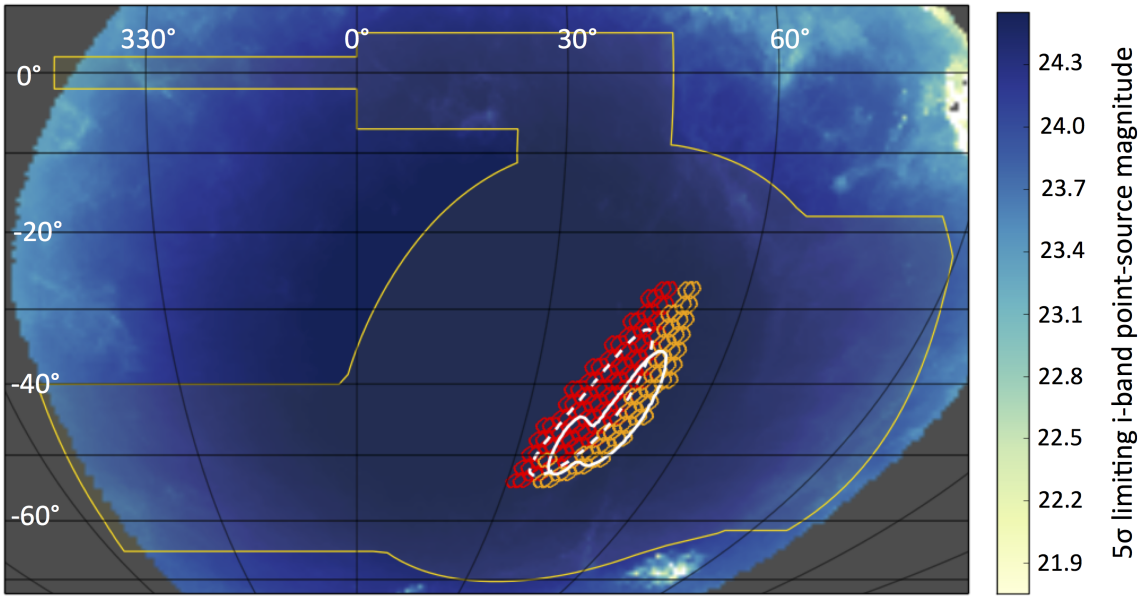


Figure 5.1: Dithered tiling performed for GW170814 overlaid on the GW170814 90%-confidence sky area contours. The red hexes show the individual pointings that were performed in our search on the first night of observations. The orange hexes represent the tiles that were not observed until the second night or later due to the sky map change. The white dotted contour shows the initial `Bayestar` map and the solid white contour represents the final sky map from the LIGO-Virgo O1-O2 GW catalog. The region enclosed by the yellow contour corresponds to the DES footprint, and the background color shows the estimated $5\sigma$ point-source limiting magnitude for a 90-second exposure which accounts for air mass and dust extinction (see Neilsen et al., 2016).

We imaged the high-probability area of the `Bayestar` sky map in the $i$-band with 90-sec exposures, corresponding to a $5\sigma$ point-source depth of $\approx$ 23 mag. Our strategy of imaging the most probable regions was similar to that used in Soares-Santos et al. (2016) and Cowperthwaite et al. (2016) (which surveyed in $i$ and $z$ bands), but in order to maximize the sky area coverage we only surveyed in the $i$-band. Our search covered 225 deg$^2$, corresponding to 90% of the initial `Bayestar` map, 90% of the `LALInference` sky map that was released with the original GW170814 LIGO-Virgo publication (Abbott et al., 2017c), and 86% of the

final map from the O1-O2 GW catalog (The LIGO Scientific Collaboration & the Virgo Collaboration, 2018). These estimates account for chip gaps on the camera but not masking of bright stars. A preliminary `LALInference` map accounting for calibration uncertainties was sent after our first night of observations, causing a shift in the search region on the second night of observations and onward (LIGO-Virgo Collaboration, 2017b). Figure 5.1 shows our tiling over the LIGO-Virgo sky maps. Observations of the region of interest were taken in epochs which began roughly 0.8, 1.8, 2.8, 5.8, 8.8, 9.8, 10.8, and 11.8 days after the GW event, and each epoch's tiling spanned about 4 hours. The $5\sigma$ limiting point-source magnitude in the $i$-band was approximately 24 over all the tiles on the first night of observations. The first DECam image was taken on August 15, 2017 at 06:05:31 UTC. Our cadence was chosen to have a dense sampling in time, but observing conditions and follow-up of GW170817 introduced larger gaps between the third and eighth nights.

We processed the images from our search using the Dark Energy Survey's (DES) transient detection pipeline as in Soares-Santos et al. (2016) and Herner et al. (in preparation). The pipeline consists of a single-epoch processing stage (Morganson et al., 2018; Bernstein et al., 2012) followed by a stage which takes the difference of search images and template images to identify sources with fluctuating brightness (`DiffImg`, Kessler et al., 2015). Template images were available from existing DES data since the LIGO-Virgo sky maps were contained in the DES footprint. The sources detected in the pipeline are used to generate candidate light curves: A candidate requires at least two detections by the pipeline, and for each candidate a light curve is constructed from a PSF-fitted flux at each observation. The pipeline also removes persistent point sources in the DES Y1 catalog that are brighter than 20.5 mag in any band.

We split the data into two samples because of a shift in the GW sky map after the first night of observations. This shift prompted a change in the patch of sky we targeted, creating inhomogeneity in the data sample since the cadence of observations was not uniform over the full area we imaged. The red and orange "hexes" in Figure 5.1 show which fields

were observed the first night versus only on later nights, respectively. The first dataset, D1, includes the $N_{D1} = 42368$ candidates which were first observed $\sim 0.8$ days after the GW trigger when we were targeting the `Bayestar` sky map. D2 contains the $N_{D2} = 17192$ candidates observed for the first time after acquiring the preliminary `LALInference` sky map. Over the full GW170814 follow-up campaign, the median number of observations per candidate is 8 and 5 for D1 and D2, respectively.

## 5.4 Analysis

To identify candidates of interest, we apply selection requirements (or "cuts") to the full set of candidates produced by `DiffImg`. We present these criteria in §5.4.2 and have chosen them to (a) minimize contamination from both astrophysical transients such as supernovae and asteroids as well as artifacts in the data and (b) identify "fast transients" which quickly decline after the merger. `SNANA` simulations (Kessler et al., 2009) of Type Ia and core-collapse SN light curves (using the SALT-II Ia light curve model of Guy et al. (2010) and Ibc, IIp, IIn core-collapse templates from Kessler et al. (2010)) provide guidance on cuts to remove supernovae. A full optimization and exploration of the cuts is not explored here and is left for future analyses. We choose these cuts using a *control sample* of candidates which are away from the highest probability regions of the `LALInference` sky map, as described in §5.4.1. The number of candidates remaining in the control sample after applying cuts is used to infer the number of candidates expected in the full sample. This inference is detailed in §5.4.3.

Our analysis also makes use of two auxiliary tools: 1) A machine-learning (ML) algorithm, `autoScan`, trained on DES difference images that produces a score for each difference-imaging detection between 0 and 1 with high scores corresponding to high-confidence point-source-like sources (Goldstein et al., 2015), and 2) object-fitting algorithms which classify persistent DES sources as galaxies or stars depending on the sources' spatial extent (Drlica-Wagner et al., 2018; Sevilla-Noarbe et al., 2018). We use these tools to identify high-confidence point-source

| | | | |
|---|---|---|---|
| $N_{D1}$ | 42368 | $N_{D2}$ | 17192 |
| $N_{C1}$ | 12381 | $N_{C2}$ | 3867 |
| $N_{B1}$ | 29987 | $N_{B2}$ | 13325 |

Table 5.1: The number of candidates in the two subsets of full (D), control (C), and blinded (B) samples.

detections in our search and to match these detections with stars and galaxies.

### *5.4.1 Control Sample*

To reduce potential bias in tuning the analysis cuts to reject all events, the cuts are optimized on a control sample. The control sample comprises a random third of all `DiffImg` candidates, and candidates within 4.5 deg of the maximum a posteriori point of the Abbott et al. (2017c) `LALInference` sky map are excluded. There is an $\sim 8\%$ and $\sim 10\%$ chance that the true location of GW170814 is in the control region based on the Abbott et al. (2017c) and The LIGO Scientific Collaboration & the Virgo Collaboration (2018) sky maps, respectively.

As with the full data set, we split our control sample into two subsamples. The first subsample C1 (with $N_{C1} = 12381$ candidates) comprises the control candidates in D1. The second subsample C2 contains the control candidates ($N_{C2} = 3867$ candidates) in D2. We apply the cuts in §5.4.2 to the two control subsamples and record the sets of candidates c1 and c2 (with $N_{c1}$ and $N_{c2}$ candidates respectively) passing cuts out of the totals.

The remaining data (which we call the *blinded sample*) is similarly split into two subsamples B1 and B2 for events first observed when targeting the `Bayestar` map and `LALInference` map, respectively. In total, Subsample B1 contains $N_{B1} = 29987$ candidates and B2 contains $N_{B2} = 13325$ candidates. Since B1, B2, C1, and C2 are mutually exclusive, we have $N_{D1} = N_{B1} + N_{C1}$ and $N_{D2} = N_{B2} + N_{C2}$. Table 5.1 summarizes the numbers of candidates in each subsample.

### *5.4.2 Selection Requirements*

Below we list the cuts applied to the candidates:

1. Raw Sample: All candidates produced by `DiffImg`.

2. 1st Epoch ML>0.7: Using the `autoScan` machine-learning score ($0 < \mathrm{ML} < 1$) that was trained with DES data (Goldstein et al., 2015) to remove non-point-source-like detections, we require $\mathrm{ML} > 0.7$ for the first observation. This cut eliminates image artifacts that arise in the difference imaging. For reference, the DES Supernova program requires $\mathrm{ML} > 0.5$, but for *two* separate detections of a candidate rather than just one detection. Our requirement is more stringent since we are looking for rapidly fading sources and therefore only cut on the first-epoch ML. Our stricter $\mathrm{ML} > 0.7$ requirement lowers the numbers of single-epoch false positives by a factor of $\sim 2$ compared with $\mathrm{ML} > 0.5$, while lowering the efficiency by only a few percent at signal-to-noise ratio 10 (Goldstein et al., 2015).

3. Host Galaxy $z < 0.3$: Using galaxies from the DES Y3 Gold catalog, a candidate is matched to a host galaxy if it is within four times the *directional light radius* of the galaxy (Gupta et al., 2016). The directional light radius is the radius of a potential host galaxy in the direction of the candidate transient and is dependent on the survey. Each galaxy is also fit with a Directional Neighborhood Fitting (DNF) photometric redshift $z_{\mathrm{DNF}}$ with uncertainty $\Delta z_{\mathrm{DNF}}$ (De Vicente et al., 2016). If the candidate is matched to a galaxy and the best match galaxy satisfies $z_{\mathrm{DNF}} - \Delta z_{\mathrm{DNF}} > 0.3$, the candidate is removed from the sample. This cut removes events that are clearly associated with galaxies beyond the estimated GW redshift of $z = 0.12^{+.03}_{-.04}$.

4. 2nd Observation S/N $\geq 2$: The candidate must have a measured signal-to-noise ratio (S/N) of at least 2 on the second observation. Measurements within one hour of each other are not considered separate observations for this cut. This cut rejects asteroids and difference imaging artifacts.

5. Greater than $2\sigma$ decline: There must be a $> 2\sigma$ decline in the flux between the first and second epochs that a candidate was observed. A similar cut was implemented in Soares-Santos et al. (2016) and Cowperthwaite et al. (2016). $\sigma$ is the quadrature sum of the flux errors on the two epochs. If multiple measurements of a candidate were taken in the same epoch (i.e. in the same night), we use the first measurement of the epoch. If we did not observe the candidate on the second epoch, it is removed from the sample. We note that the effect of this cut depends sensitively on the observational choices of the follow-up campaign, not just the astrophysics of the potential EM source.

6. $N_{\rm obs} \geq 4$: To ensure that we can examine each candidate's light curve over a broad portion of the follow-up campaign, the candidate must have been observed at least $N_{\rm obs} = 4$ times, regardless of S/N.

7. Late-time S/N $< 6$: After one week from the GW event, the S/N of all observations of a candidate must be less than 6. This requirement removes objects that are bright at late times such as supernovae and variable stars.

8. No Late-time Brightening: To isolate fading transients, we require that after 48 hours from the GW event, there is no increase in flux of the candidate greater than $3\sigma$, where $\sigma$ is the quadrature sum of uncertainties on adjacent flux measurements.

9. Visual Inspection: Subtracted image stamps identified as artifacts (e.g. cosmic rays) are removed from the sample.

After applying these cuts to the control sample, $N_{c1} = 1$ and $N_{c2} = 0$ candidates remain.

### *5.4.3 Expectation of Number of Candidates in Full Sample*

Given $N_{c1}$ and $N_{c2}$ out of $N_{C1}$ and $N_{C2}$ candidates passing in the control fields, respectively, we expect $\langle N_{b1} + N_{b2} \rangle = N_{c1}N_{B1}/N_{C1} + N_{c2}N_{B2}/N_{C2} = 2.4$ events in B1+B2, which we interpret as the mean of a Poisson distribution. This interpretation does not account for

small differences in Milky Way reddening and stellar density over the search region. In §5.5, we analyze the blinded sample and compare our expectations to the number of candidates passing the cuts.

## 5.5 Results

Table 5.2 shows the effect of the cuts on the full sample, which includes the control sample. It also shows the initial $i$-band magnitudes and sky positions for the events passing all cuts. After analyzing the blinded sample, one more candidate is found, leaving a total of two candidates passing cuts in the control and blinded samples, with ID numbers 1 and 2, respectively. Finding one candidate passing cuts in the blinded sample is consistent with the 2.4 expected background events derived from the control sample presented in §5.4.3. The light curves for both events and their sky positions are shown in Figure 5.3.

Upon visual inspection of the two candidates, neither is an obvious subtraction artifact or cosmic ray. Here we do not show examples of subtraction artifacts and cosmic rays that would be cut by visual inspection since visual inspection did not end up removing any candidates in this analysis. However, the template images for both candidates contain a bright source at the position of the candidates. The template, search and difference images from the first epoch of observations of each candidate are shown in Figure 5.2.

A deeper search through the DES high-quality object catalog ("Y3 Gold") reveals that Candidate 1 is associated with an object that is classified as either a galaxy at $z \sim 0.9$, or a star, depending on the classifier used. A multi-epoch, multi-object fitting algorithm classifies the object's PSF as a candidate star, whereas the single-object fit categorizes the object as a galaxy (Drlica-Wagner et al., 2018; Sevilla-Noarbe et al., 2018). Notably, the object is too faint to meet the brightness cutoff for inclusion in our star veto catalog and it is not vetoed by our host galaxy redshift cut (cut 3) because we only include the highest confidence galaxies in the host-galaxy matching. Fitting each band to a constant flux for all archival observations of the object (DES Years 1-4) results in a $\chi^2$/DOF of $48.6/17 = 2.9$

| Cuts | $N_{\rm seq}$ | $N_{\rm only}$ | $N_{\rm LO}$ |
|---|---|---|---|
| 1. Raw Sample | 59560 | – | – |
| 2. 1st Epoch ML $> 0.7$ | 1206 | 1206 | 258 |
| 3. Unmatched or Host $z < 0.30$ | 730 | 31119 | 8 |
| 4. 2nd Obs S/N $\geq 2.0$ | 663 | 44181 | 4 |
| 5. $> 2.0$ sigma decline | 45 | 5570 | 65 |
| 6. $N_{\rm obs} \geq 4$ | 31 | 50029 | 2 |
| 7. Late-time S/N $< 6$ | 4 | 27571 | 21 |
| 8. No Late-time Brightening | 2 | 36499 | 4 |
| 9. Visual Inspection | 2 | – | 2 |

| Candidate # | RA | DEC | $m_i$ |
|---|---|---|---|
| 1 | 42.35047° | -40.32632° | 22.5 |
| 2 | 47.63365° | -36.36045° | 21.9 |

Table 5.2: *Top*: Candidates remaining in the full data sample after applying cuts. $N_{\rm seq}$ is the number of candidates remaining after applying each cut sequentially, $N_{\rm only}$ is the number of candidates after applying an individual cut, and $N_{\rm LO}$ is the number of candidates if a cut is "left out" but all the rest are applied. *Bottom*: sky coordinates and initial $i$-band magnitude $m_i$ of the two candidates passing all cuts.

and $p$-value $p(\chi^2 \geq 48.6|{\rm DOF} = 17) = 7 \times 10^{-5}$, indicating previous variability of the source. These archival fluxes are shown in Figure 5.3. Spectroscopic observations of this source could clarify if the object is a star or galaxy.

Candidate 2 is also associated with a DES Y3-Gold object and is classified as a star by both classifiers and constant-flux fits to archival observations yield a $\chi^2/{\rm DOF}$ of $25.7/14 = 1.8$ and $p$-value of $p(\chi^2 \geq 25.7|{\rm DOF} = 14) = 0.03$ (see Figure 5.3). However, the star is also too faint (by 0.16 mag) to meet the brightness cutoff for the star veto catalog of our pipeline and hence was not removed by the 20.5 mag persistent-point-source cut in §5.3.

## 5.6 Discussion

Although our search identified two interesting candidates, it is unlikely that either candidate is associated with GW170814. Neither candidate is located in the 90% confidence region of the `LALInference` sky map, and both are associated with existing objects in DES catalogs

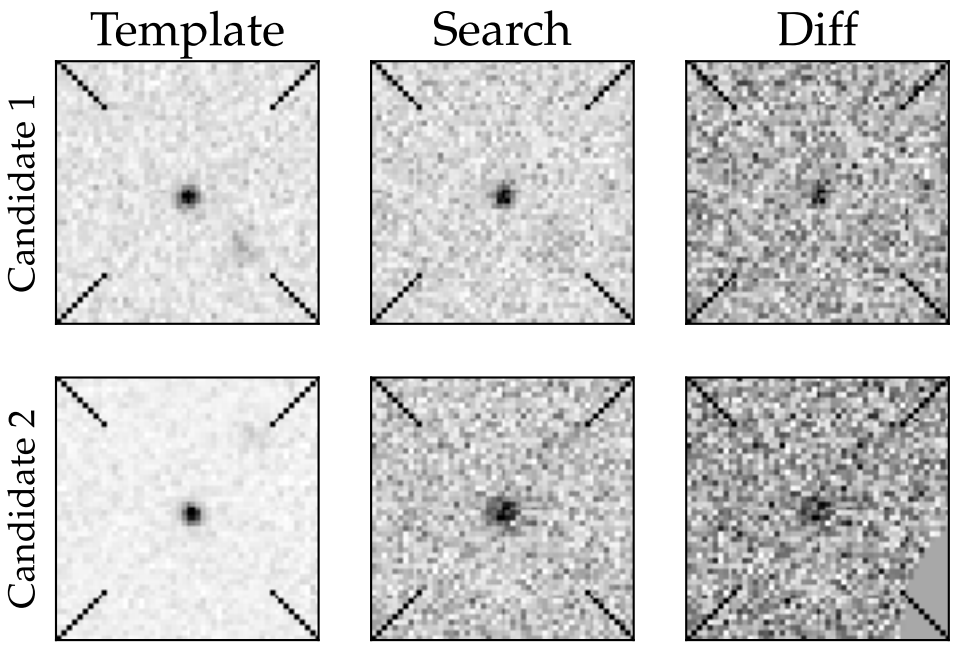


Figure 5.2: Template, search, and difference image stamps for candidates passing cuts. The top row shows the $i$-band images for Candidate 1, and the bottom for Candidate 2. The search and difference images are from the the first epoch of observations of the candidate. Each stamp is 13.2" x 13.2".

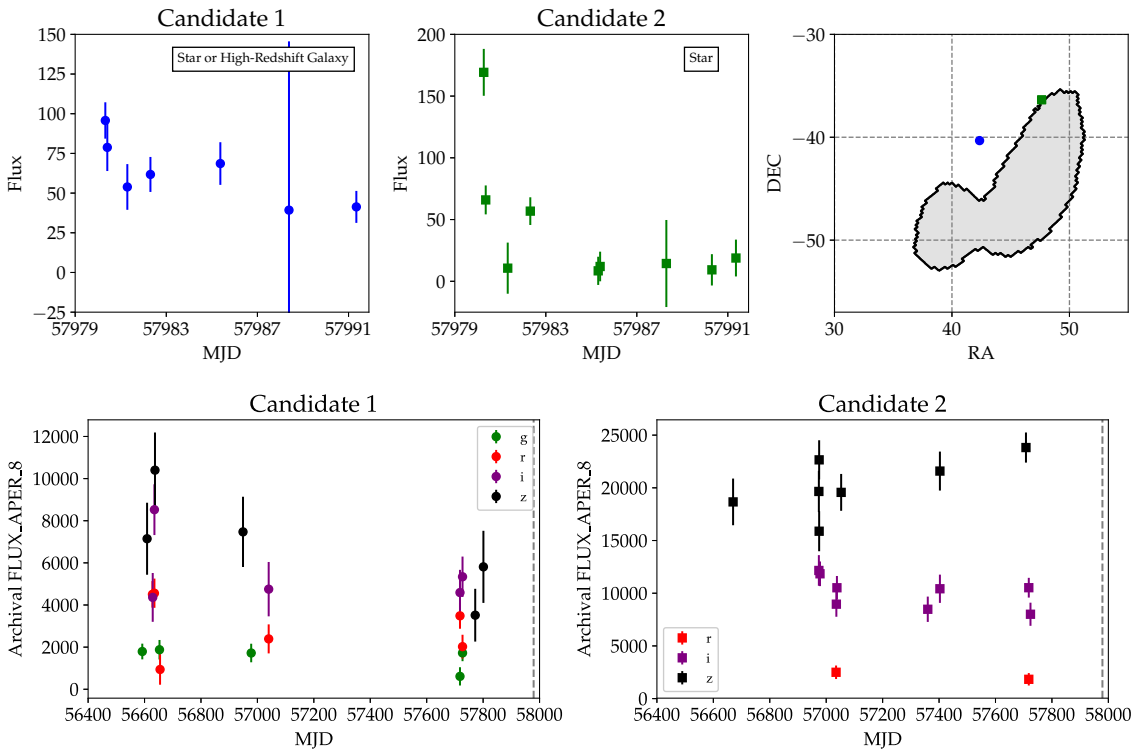


Figure 5.3: Light curves, archival fluxes, and sky positions for the two candidates passing all cuts. *Top:* The left panel shows the $i$-band light curve for Candidate 1 (associated with Y3 Gold star or high-redshift galaxy), and the middle panel shows the same for Candidate 2 (associated with Y3 Gold star). The flux is defined in relation to AB magnitude as $m_{\mathrm{AB}} = -2.5\log_{10}(\mathrm{Flux}) + 27.5$. The right panel overlays the sky positions of the two candidates on the 90% credible region of the `lalinference` sky map (gray). *Bottom:* Available archival flux measurements in $g,r,i,z$ bands at the locations of Candidates 1 (left) and 2 (right). These `FLUX_APER_8` fluxes are taken with 22.22-pixel apertures and are not from difference imaging and therefore cannot be directly compared to those in the top panels. The vertical, gray, dashed line on the far right of the two plots indicates the GW170814 merger time.

that are inconsistent with our expectations of the GW source. Candidate 2 is likely the transient behavior of a variable star and is consistent with the number of background candidates expected in the blinded sample. Candidate 1 could also be stellar variability, or it could be a signal associated with a distant galaxy. Assuming it is a galaxy, the DNF photometric redshift of the object is $z = 0.95 \pm 0.12$, far beyond the possible redshift of GW170814 at that sky position: The 99% upper limit on the GW distance along the Candidate 1 line of sight is 615 Mpc, whereas the galaxy distance is $6380^{+1010}_{-980}$ Mpc assuming the LCDM cosmology parameters of Planck Collaboration (2016). We note however that photometric redshifts can occasionally have catastrophic failures.

An alternative explanation for the persistent emission from the two candidates is that one or both of these candidates is associated with a quasar. If either is a quasar, it is unlikely to be at the low redshifts of interest for GW170814 (Pâris et al., 2017). Spectroscopic follow-up of the persistent sources associated with Candidates 1 and 2 could resolve whether we have mis-categorized them.

We conclude that these two candidates are not associated with GW170814, and thus we find no EM counterpart associated with the BBH merger over the 225 deg$^2$ region that we surveyed with 86% sky map coverage. We have not yet computed the efficiency, which is needed to set rate limits on BBH merger emission, but this rate-limit analysis is underway using `SNANA` simulations similar to those used in Soares-Santos et al. (2016). Our rate-limit analysis will also re-evaluate the cuts to maximize possible BBH model efficiency while minimizing supernova background events. Qualitatively though, the analysis presented here covers 86% of the GW sky map and searches for events with rapidly declining light curves. The non-detection of an EM counterpart in our sample results in stringent limits on fast-declining optical models brighter than $i \sim 23$ mag 1-2 days after the BBH coalescence. This search is not sensitive to models that fade faster than the time between the first two observations due to Cut 4. Assuming a flat-in-frequency optical spectrum from 4000 Å to 7000 Å and the GW170814 median distance, this $i \sim 23$ mag limit corresponds to a

luminosity limit of $L_{\rm optical} \sim 5 \times 10^{41}$ erg/s.

Our results constrain the space of models put forth in e.g. Stone et al. (2017), de Mink & King (2017). For example, Stone et al. (2017) posits that BBH mergers occurring in the gaseous environments of AGN disks could be accompanied by gas accretion onto the final merged black hole that powers luminosities of order $L \sim 10^{40}$ erg/s lasting a few years, but highly super-Eddington accretion might result in a brighter and shorter-lived transient that our analysis is sensitive to. Our search also narrows the feasibility of models from de Mink & King (2017), which predict emission with luminosities of approximately $L \sim 10^{42}$ erg/s occurring on fast timescales. The search performed here is tailored to remove longer-lived transients, and therefore it does not constrain long-lived BBH counterparts, such as the supernova association suggested in Loeb (2016) and later discussed in Woosley 2016 and D'Orazio & Loeb 2018.

Aside from identifying interesting candidates, our search for counterparts to GW170814 is a test-bed for future BBH follow-up analyses where the sky map credible areas will be small enough to be completely tiled in less than one night using DECam. For a real-time search for future counterparts, we consider resources to spectroscopically follow $\sim 10$ candidates, which we would want to identify within roughly two days of the GW trigger. In this scenario, we only apply the first five cuts in Table 5.2, since the remaining cuts depend on observations beyond two days. Through cut 5 ($> 2\sigma$ decline), our search finds 45 candidates. Of these, we find that four candidates (including Candidates 1 and 2) are associated with DES-catalog objects that are either galaxies beyond our redshift cut (cut 3), or stars, and are thus uninteresting as black-hole-merger counterparts. Excluding these four candidates, our real time search would find 41 candidates over 225 deg$^2$ or $\sim 16$ candidates per 87 deg$^2$ (the 90% credible area of the final GW170814 sky map). For comparison, Cowperthwaite & Berger (2015) predicts $\sim 19$ Type Ia SNe detected at $z < 0.25$ over a 7-day, 87-deg$^2$ search. This suggests that the first five cuts are adequate to find interesting spectroscopic targets over a region the size of the GW170814 sky map.

Future work will incorporate simulations of BBH and SN light curves to assess the efficiency and false alarm rate of our search. If several BBH events are followed up with no EM counterpart found, a combined analysis will be needed to set limits on BBH EM emission.

## 5.7 Conclusion

Although our search identified two interesting candidates, it is unlikely that either candidate is associated with GW170814. Neither candidate is located in the 90% confidence region of the `LALInference` sky map, and both are associated with existing objects in DES catalogs that are inconsistent with our expectations of the GW source. Candidate 2 is likely the transient behavior of a variable star and is consistent with the number of background candidates expected in the blinded sample. Candidate 1 could also be stellar variability, or it could be a signal associated with a distant galaxy. Assuming it is a galaxy, the DNF photometric redshift of the object is $z = 0.95 \pm 0.12$, far beyond the possible redshift of GW170814 at that sky position: The 99% upper limit on the GW distance along the Candidate 1 line of sight is 615 Mpc, whereas the galaxy distance is $6380^{+1010}_{-980}$ Mpc assuming the LCDM cosmology parameters of Planck Collaboration (2016). We note however that photometric redshifts can occasionally have catastrophic failures.

An alternative explanation for the persistent emission from the two candidates is that one or both of these candidates is associated with a quasar. If either is a quasar, it is unlikely to be at the low redshifts of interest for GW170814 (Pâris et al., 2017). Spectroscopic follow-up of the persistent sources associated with Candidates 1 and 2 could resolve whether we have mis-categorized them.

We conclude that these two candidates are not associated with GW170814, and thus we find no EM counterpart associated with the BBH merger over the 225 deg$^2$ region that we surveyed with 86% sky map coverage. We have not yet computed the efficiency, which is needed to set rate limits on BBH merger emission, but this rate-limit analysis is underway using `SNANA` simulations similar to those used in Soares-Santos et al. (2016). Our rate-

limit analysis will also re-evaluate the cuts to maximize possible BBH model efficiency while minimizing supernova background events. Qualitatively though, the analysis presented here covers 86% of the GW sky map and searches for events with rapidly declining light curves. The non-detection of an EM counterpart in our sample results in stringent limits on fast-declining optical models brighter than $i \sim 23$ mag 1-2 days after the BBH coalescence. This search is not sensitive to models that fade faster than the time between the first two observations due to Cut 4. Assuming a flat-in-frequency optical spectrum from 4000 Å to 7000 Å and the GW170814 median distance, this $i \sim 23$ mag limit corresponds to a luminosity limit of $L_{\rm optical} \sim 5 \times 10^{41}$ erg/s.

Our results constrain the space of models put forth in e.g. Stone et al. (2017), de Mink & King (2017). For example, Stone et al. (2017) posits that BBH mergers occurring in the gaseous environments of AGN disks could be accompanied by gas accretion onto the final merged black hole that powers luminosities of order $L \sim 10^{40}$ erg/s lasting a few years, but highly super-Eddington accretion might result in a brighter and shorter-lived transient that our analysis is sensitive to. Our search also narrows the feasibility of models from de Mink & King (2017), which predict emission with luminosities of approximately $L \sim 10^{42}$ erg/s occurring on fast timescales. The search performed here is tailored to remove longer-lived transients, and therefore it does not constrain long-lived BBH counterparts, such as the supernova association suggested in Loeb (2016) and later discussed in Woosley 2016 and D'Orazio & Loeb 2018.

Aside from identifying interesting candidates, our search for counterparts to GW170814 is a test-bed for future BBH follow-up analyses where the sky map credible areas will be small enough to be completely tiled in less than one night using DECam. For a real-time search for future counterparts, we consider resources to spectroscopically follow $\sim 10$ candidates, which we would want to identify within roughly two days of the GW trigger. In this scenario, we only apply the first five cuts in Table 5.2, since the remaining cuts depend on observations beyond two days. Through cut 5 ($> 2\sigma$ decline), our search finds 45 candidates. Of these,

we find that four candidates (including Candidates 1 and 2) are associated with DES-catalog objects that are either galaxies beyond our redshift cut (cut 3), or stars, and are thus uninteresting as black-hole-merger counterparts. Excluding these four candidates, our real time search would find 41 candidates over 225 deg$^2$ or $\sim 16$ candidates per 87 deg$^2$ (the 90% credible area of the final GW170814 sky map). For comparison, Cowperthwaite & Berger (2015) predicts $\sim 19$ Type Ia SNe detected at $z < 0.25$ over a 7-day, 87-deg$^2$ search. This suggests that the first five cuts are adequate to find interesting spectroscopic targets over a region the size of the GW170814 sky map.

Future work will incorporate simulations of BBH and SN light curves to assess the efficiency and false alarm rate of our search. If several BBH events are followed up with no EM counterpart found, a combined analysis will be needed to set limits on BBH EM emission.

# REFERENCES


Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, http://dx.doi.org/10.1103/PhysRevLett.116.241102, Phys. Rev. Lett., 116, 241102

—. 2016b, http://dx.doi.org/10.1007/lrr-2016-1, LRR, 19, 1

—. 2017a, http://dx.doi.org/http://doi.org/10.1038/nature24471, Nature, in press

—. 2017b, http://dx.doi.org/10.3847/2041-8213/aa920c, ApJL, in press

—. 2017c, http://dx.doi.org/10.1103/PhysRevLett.119.161101, Physics Review Letters, 119, 61101

—. 2017d, http://dx.doi.org/10.3847/2041-8213/aa91c9, ApJL, in press

Ade, P. A. R., Aghanim, N., Arnaud, M., et al. 2016, http://dx.doi.org/10.1051/0004-6361/201525830, Astronomy and Astrophysics, 594, A13

Akmal, A., Pandharipande, V. R., & Ravenhall, D. G. 1998, http://dx.doi.org/10.1103/PhysRevC.58.1804, Phys. Rev. C, 58, 1804

Alexander, K. D., et al. 2017, http://dx.doi.org/10.3847/2041-8213/aa905d, ApJL, in press

Antoniadis, J., Freire, P. C. C., Wex, N., et al. 2013, http://dx.doi.org/10.1126/science.1233232, Science, 340, 6131

Arcavi, I., Hosseinzadeh, G., Howell, D. A., et al. 2017, http://dx.doi.org/10.1038/nature24291, Nature, in press

Arnould, M., Goriely, S., & Takahashi, K. 2007, http://dx.doi.org/10.1016/j.physrep.2007.06.002, Physics Reports, 450, 97

Barnes, J., & Kasen, D. 2013, http://dx.doi.org/10.1088/0004-637X/775/1/18, Astrophys. J., 775, 18

Barnes, J., Kasen, D., Wu, M.-R., & Martínez-Pinedo, G. 2016, http://dx.doi.org/10.3847/0004-637X/829/2/110, ApJ, 829, 110

Bauswein, A., Goriely, S., & Janka, H.-T. 2013, http://dx.doi.org/10.1088/0004-637X/773/1/78, ApJ, 773, 78

Beniamini, P., Hotokezaka, K., & Piran, T. 2016, http://dx.doi.org/10.3847/0004-637X/832/2/149, ApJ, 832, 149

Bovard, L., Martin, D., Guercilena, F., et al. 2017, ArXiv e-prints, http://arxiv.org/abs/1709.09630arXiv:1709.09630 [gr-qc]

Buchdahl, H. A. 1959, http://dx.doi.org/10.1103/PhysRev.116.1027, Phys. Rev., 116, 1027

Burbidge, E. M., Burbidge, G. R., Fowler, W. A., & Hoyle, F. 1957, http://dx.doi.org/10.1103/RevModPhys.29.547, Rev. Mod. Phys., 29, 547

Burgay, M., D'Amico, N., Possenti, A., et al. 2003, http://dx.doi.org/10.1038/nature02124, Nature, 426, 531

Chornock, R., et al. 2017, http://dx.doi.org/10.3847/2041-8213/aa905c, ApJL, in press

Ciolfi, R., Kastaun, W., Giacomazzo, B., et al. 2017, http://dx.doi.org/10.1103/PhysRevD.95.063016, Physical Review D, 95, 063016

Côté, B., Belczynski, K., Fryer, C. L., et al. 2017, http://dx.doi.org/10.3847/1538-4357/aa5c8d, ApJ, 836, 230

Coughlin, M., Dietrich, T., Kawaguchi, K., et al. 2017, ArXiv e-prints, http://arxiv.org/abs/1708.07714, arXiv:1708.07714 [astro-ph.HE]

Coulter, D. A., et al. 2017, http://dx.doi.org/10.1126/science.aap9811, Science, in press

Covino, S., Wiersema, K., Fan, Y. Z., et al. 2017, http://dx.doi.org/10.1038/s41550-017-0285-z, Nature Astronomy, in press

Cowperthwaite, P., et al. 2017, http://dx.doi.org/10.3847/2041-8213/aa8fc7, ApJL, in press

Damour, T., Nagar, A., & Villain, L. 2012, http://dx.doi.org/10.1103/PhysRevD.85.123007, Physical Review D, 85, 123007

Del Pozzo, W., Li, T. G. F., Agathos, M., Van Den Broeck, C., & Vitale, S. 2013, http://dx.doi.org/10.1103/PhysRevLett.111.071101, Phys. Rev. Lett., 111, 071101

Demorest, P., Pennucci, T., Ransom, S., Roberts, M., & Hessels, J. 2010, http://dx.doi.org/10.1038/nature09466, Nature, 467, 1081

Dessart, L., Ott, C., Burrows, A., Rosswog, S., & Livne, E. 2009, http://dx.doi.org/10.1088/0004-637X/690/2/1681, ApJ, 690, 1681

Dessart, L., Ott, C. D., Burrows, A., Rosswog, S., & Livne, E. 2009, http://dx.doi.org/10.1088/0004-637X/690/2/1681, ApJ, 690, 1681

Diaz, M. C., Macri, L. M., et al. 2017, http://dx.doi.org/10.3847/2041-8213/aa9060, ApJL, in press

Dietrich, T., Bernuzzi, S., Ujevic, M., & Tichy, W. 2017a, http://dx.doi.org/10.1103/PhysRevD.95.044045, Phys. Rev. D, 95, 044045

Dietrich, T., & Ujevic, M. 2017, http://dx.doi.org/10.1088/1361-6382/aa6bb0, Class. Quant. Grav., 34, 105014

Dietrich, T., Ujevic, M., Tichy, W., Bernuzzi, S., & Brügmann, B. 2017b, http://dx.doi.org/10.1103/PhysRevD.95.024029, Physical Review D, 95, 024029

Dominik, M., Belczynski, K., Fryer, C., et al. 2012, http://dx.doi.org/10.1088/0004-637X/759/1/52, ApJ, 759, 52

Douchin, F., & Haensel, P. 2001, http://dx.doi.org/10.1051/0004-6361:20011402, Astronomy and Astrophysics, 380, 151

Drout, M. R., et al. 2017, http://dx.doi.org/10.1126/science.aaq0049, Science, in press

Endrizzi, A., Ciolfi, R., Giacomazzo, B., Kastaun, W., & Kawamura, T. 2016, http://dx.doi.org/10.1088/0264-9381/33/16/164001, Class. Quant. Grav., 33, 164001

Engvik, L., Bao, G., Hjorth-Jensen, M., Osnes, E., & Ostgaard, E. 1996, http://dx.doi.org/10.1086/177827, ApJ, 469, 794

Evans, P. A., et al. 2017, http://dx.doi.org/10.1126/science.aap9580, Science, in press

Fernández, R., Kasen, D., Metzger, B. D., & Quataert, E. 2015, http://dx.doi.org/10.1093/mnras/stu2112, , 446, 750

Flanagan, É. É., & Hinderer, T. 2008, http://dx.doi.org/10.1103/PhysRevD.77.021502, Phys. Rev. D, 77, 021502

Foucart, F., O'Connor, E., Roberts, L., et al. 2016, http://dx.doi.org/10.1103/PhysRevD.94.123016, Physical Review D, 94, 123016

Freiburghaus, C., Rembges, J.-F., Rauscher, T., et al. 1999, http://stacks.iop.org/0004-637X/516/i=1/a=381, ApJ, 516, 381

Fujibayashi, S., Sekiguchi, Y., Kiuchi, K., & Shibata, M. 2017, ArXiv e-prints, http://arxiv.org/abs/1703.10191arXiv:1703.10191 [astro-ph.HE], ApJ, accepted

Glendenning, N. K. 1985, http://dx.doi.org/10.1086/163253, ApJ, 293, 470

Goriely, S., Bauswein, A., & Janka, H.-T. 2011, http://dx.doi.org/10.1088/2041-8205/738/2/L32, ApJL, 738, L32

Goriely, S., Bauswein, A., Just, O., Pllumbi, E., & Janka, H.-T. 2015, http://dx.doi.org/10.1093/mnras/stv1526, , 452, 3894

Hotokezaka, K., Kiuchi, K., Kyutoku, K., et al. 2013, http://dx.doi.org/10.1103/PhysRevD.87.024001, Physical Review D, 87, 024001

Ji, A. P., Frebel, A., Chiti, A., & Simon, J. D. 2016, http://dx.doi.org/10.1038/nature17425, , 531, 610

Just, O., Bauswein, A., Pulpillo, R. A., Goriely, S., & Janka, H.-T. 2015, http://dx.doi.org/10.1093/mnras/stv009, , 448, 541

Kalogera, V., & Baym, G. 1996, http://dx.doi.org/10.1086/310296, ApJL, 470, L61

Kasen, D., Badnell, N. R., & Barnes, J. 2013, http://dx.doi.org/10.1088/0004-637X/774/1/25, ApJ, 774, 25

Kasen, D., Fernandez, R., & Metzger, B. 2015, http://dx.doi.org/10.1093/mnras/stv721, , 450, 1777

Kasen, D., et al. 2017, http://dx.doi.org/10.1038/nature24453, Nature, in press

Kastaun, W., Ciolfi, R., Endrizzi, A., & Giacomazzo, B. 2017, http://dx.doi.org/10.1103/PhysRevD.96.043019, Physical Review D, 96, 043019

Kastaun, W., & Galeazzi, F. 2015, http://dx.doi.org/10.1103/PhysRevD.91.064027, Physical Review D, 91, 064027

Kawaguchi, K., Kyutoku, K., Shibata, M., & Tanaka, M. 2016, http://dx.doi.org/10.3847/0004-637X/825/1/52, ApJ, 825, 52

Kiuchi, K., Sekiguchi, Y., Kyutoku, K., et al. 2015, http://dx.doi.org/10.1103/PhysRevD.92.064034, Physical Review D, 92, 064034

Lackey, B. D., Nayyar, M., & Owen, B. J. 2006, http://dx.doi.org/10.1103/PhysRevD.73.024021, Phys. Rev. D, 73, 024021

Lattimer, J. M., & Prakash, M. 2001, http://dx.doi.org/10.1086/319702, ApJ, 550, 426

Lattimer, J. M., & Schramm, D. N. 1974, http://dx.doi.org/10.1086/181612, ApJL, 192, L145

Lehner, L., Liebling, S. L., Palenzuela, C., et al. 2016, http://dx.doi.org/10.1088/0264-9381/33/18/184002, Class. Quant. Grav., 33, 184002

Li, L.-X., & Paczynski, B. 1998, http://stacks.iop.org/1538-4357/507/i=1/a=L59, ApJL, 507, L59

Lipunov, V. N., et al. 2017, http://dx.doi.org/10.3847/2041-8213/aa92c0, ApJL, in press

Lo, K.-W., & Lin, L.-M. 2011, http://dx.doi.org/10.1088/0004-637X/728/1/12, ApJ, 728, 12

Madau, P., & Dickinson, M. 2014, http://dx.doi.org/10.1146/annurev-astro-081811-125615, , 52, 415

Martin, D., Perego, A., Arcones, A., et al. 2015, http://dx.doi.org/10.1088/0004-637X/813/1/2, ApJ, 813, 2

McCully, C., Hiramatsu, D., Howell, D. A., et al. 2017, http://dx.doi.org/10.3847/2041-8213/aa9111, ApJL, in press

Metzger, B. D. 2017, http://dx.doi.org/10.1007/s41114-017-0006-z, LRR, 20, 3

Metzger, B. D., & Berger, E. 2012, http://dx.doi.org/10.1088/0004-637X/746/1/48, ApJ, 746, 48

Metzger, B. D., & Fernandez, R. 2014, http://dx.doi.org/10.1093/mnras/stu802, , 441, 3444

Metzger, B. D., Martínez-Pinedo, G., Darbha, S., et al. 2010, http://dx.doi.org/10.1111/j.1365-2966.2010.16864.x, , 406, 2650

Müller, H., & Serot, B. D. 1996, http://dx.doi.org/10.1016/0375-9474(96)00187-X, Nucl. Phys., A606, 508

Müther, H., Prakash, M., & Ainsworth, T. L. 1987, http://dx.doi.org/10.1016/0370-2693(87)91611-X, Phys. Lett., B199, 469

Nakar, E. 2007, http://dx.doi.org/10.1016/j.physrep.2007.02.005, , 442, 166

Nicholl, M., et al. 2017, http://dx.doi.org/10.3847/2041-8213/aa9029, ApJL, in press

Oertel, M., Hempel, M., Klähn, T., & Typel, S. 2017, http://dx.doi.org/10.1103/RevModPhys.89.015007, Rev. Mod. Phys., 89, 015007

Oppenheimer, J. R., & Volkoff, G. M. 1939, http://dx.doi.org/10.1103/PhysRev.55.374, Phys. Rev., 55, 374

O'Shaughnessy, R., Belczynski, K., & Kalogera, V. 2008, http://dx.doi.org/10.1086/526334, ApJ, 675, 566

Özel, F., & Freire, P. 2016, http://dx.doi.org/10.1146/annurev-astro-081915-023322, , 54, 401

Perego, A., Rosswog, S., Cabezón, R. M., et al. 2014, http://dx.doi.org/10.1093/mnras/stu1352, , 443, 3134

Pian, E., D'Avanzo, P., et al. 2017, http://dx.doi.org/10.1038/nature24298, Nature, in press

Qian, Y.-Z. 2000, http://stacks.iop.org/1538-4357/534/i=1/a=L67, ApJL, 534, L67

Radice, D., Galeazzi, F., Lippuner, J., et al. 2016, http://dx.doi.org/10.1093/mnras/stw1227, , 460, 3255

Roberts, L. F., Kasen, D., Lee, W. H., & Ramirez-Ruiz, E. 2011, http://dx.doi.org/10.1088/2041-8205/736/1/L21, ApJL, 736, L21

Rosswog, S. 2013, http://dx.doi.org/10.1098/rsta.2012.0272, Phil. Trans. R. Soc. A, 371, 20120272

Rosswog, S., Feindt, U., Korobkin, O., et al. 2017, http://dx.doi.org/10.1088/1361-6382/aa68a9, Class. Quant. Grav., 34, 104001

Rosswog, S., Liebendörfer, M., Thielemann, F.-K., et al. 1999, Astronomy and Astrophysics, 341, 499

Sekiguchi, Y., Kiuchi, K., Kyutoku, K., Shibata, M., & Taniguchi, K. 2016, http://dx.doi.org/10.1103/PhysRevD.93.124046, Physical Review D, 93, 124046

Shibata, M., Kiuchi, K., & Sekiguchi, Y.-i. 2017, http://dx.doi.org/10.1103/PhysRevD.95.083005, Physical Review D, 95, 083005

Siegel, D. M., & Metzger, B. D. 2017, ArXiv e-prints, http://arxiv.org/abs/1705.05473arXiv:1705.05473 [astro-ph.HE]

Smartt, S. J., et al. 2017, http://dx.doi.org/10.1038/nature24303, Nature, in press

Sneden, C., Cowan, J. J., & Gallino, R. 2008, http://dx.doi.org/10.1146/annurev.astro.46.060407.145207, , 46, 241

Soares-Santos, M., et al. 2017, http://dx.doi.org/10.3847/2041-8213/aa9059, ApJL, in press

Tanaka, M., & Hotokezaka, K. 2013, http://dx.doi.org/10.1088/0004-637X/775/2/113, ApJ, 775, 113

Tanaka, M., Utsumi, Y., Mazzali, P. A., et al. 2017, http://dx.doi.org/10.1093/pasj/psx121, PASJ, in press

Tanaka, M., Kato, D., Gaigalas, G., et al. 2017, ArXiv e-prints, http://arxiv.org/abs/1708.09101arXiv:1708.09101 [astro-ph.HE]

Tanvir, N. R., et al. 2017, http://dx.doi.org/10.3847/2041-8213/aa90b6, ApJL, in press

Terasawa, M., Sumiyoshi, K., Kajino, T., Mathews, G. J., & Tanihata, I. 2001, http://stacks.iop.org/0004-637X/562/i=1/a=470, ApJ, 562, 470

Troja, E., Piro, L., van Eerten, H. J., et al. 2017, http://dx.doi.org/10.1038/nature24290, Nature, in press

Valenti, S., et al. 2017, http://dx.doi.org/10.3847/2041-8213/aa8edf, ApJL, in press

Veitch, J., Raymond, V., Farr, B., et al. 2015, http://dx.doi.org/10.1103/PhysRevD.91.042003, Phys. Rev. D, 91, 042003

Wade, L., Creighton, J. D. E., Ochsner, E., et al. 2014, http://dx.doi.org/10.1103/PhysRevD.89.103012, Phys. Rev. D, 89, 103012

Wanajo, S., Sekiguchi, Y., Nishimura, N., et al. 2014, http://dx.doi.org/10.1088/2041-8205/789/2/L39, ApJL, 789, L39

Wiringa, R. B., Fiks, V., & Fabrocini, A. 1988, http://dx.doi.org/10.1103/PhysRevC.38.1010, Phys. Rev. C, 38, 1010

Wollaeger, R. T., Korobkin, O., Fontes, C. J., et al. 2017, ArXiv e-prints, http://arxiv.org/abs/1705.07084arXiv:1705.07084 [astro-ph.HE]

Yagi, K., & Yunes, N. 2017, http://dx.doi.org/10.1016/j.physrep.2017.03.002, , 681, 1

B. P. Abbott, R. Abbott, T. D. Abbott, *et al.*, \doibase10.1103/PhysRevLett.116.061102 Physical Review Letters **116**, 061102 (2016a), http://arxiv.org/abs/1602.03837 arXiv:1602.03837 [gr-qc] .

B. P. Abbott, R. Abbott, T. D. Abbott, *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), \doibase10.1103/PhysRevLett.116.241103 Phys. Rev. Lett. **116**, 241103 (2016b).

B. P. Abbott, R. Abbott, T. D. Abbott, *et al.* (LIGO Scientific and Virgo Collaboration), \doibase10.1103/PhysRevLett.118.221101 Phys. Rev. Lett. **118**, 221101 (2017a).

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., \doibase10.1103/PhysRevLett.116.241102 Physical Review Letters **116**, 241102 (2016c), http://arxiv.org/abs/1602.03840 arXiv:1602.03840 [gr-qc] .

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., \doibase10.3847/2041-8205/818/2/L22 Astrophysical Journal **818**, L22 (2016d), http://arxiv.org/abs/1602.03846 arXiv:1602.03846 [astro-ph.HE] .

G. Lovelace, C. O. Lousto, J. Healy, M. A. Scheel, A. Garcia, R. O'Shaughnessy, M. Boyle, M. Campanelli, D. A. Hemberger, L. E. Kidder, H. P. Pfeiffer, B. Szilágyi, S. A. Teukolsky, and Y. Zlochower, \doibase10.1088/0264-9381/33/24/244002 Classical and Quantum Gravity **33**, 244002 (2016), http://arxiv.org/abs/1607.05377 arXiv:1607.05377 [gr-qc] .

B. P. Abbott *et al.* (Virgo, LIGO Scientific), \doibase10.1103/PhysRevD.94.064035 Phys. Rev. **D94**, 064035 (2016), http://arxiv.org/abs/1606.01262 arXiv:1606.01262 [gr-qc] .

S. Husa, S. Khan, M. Hannam, M. Prrer, F. Ohme, X. Jimnez Forteza, and A. Boh, \doibase10.1103/PhysRevD.93.044006 Phys. Rev. **D93**, 044006 (2016), http://arxiv.org/abs/1508.07250 arXiv:1508.07250 [gr-qc] .

S. Khan, S. Husa, M. Hannam, F. Ohme, M. Prrer, X. Jimnez Forteza, and A. Boh, \doibase10.1103/PhysRevD.93.044007 Phys. Rev. **D93**, 044007 (2016a), http://arxiv.org/abs/1508.07253 arXiv:1508.07253 [gr-qc] .

A. Taracchini *et al.*, \doibase10.1103/PhysRevD.89.061502 Phys. Rev. **D89**, 061502 (2014), http://arxiv.org/abs/1311.2544 arXiv:1311.2544 [gr-qc] .

S. Babak, A. Taracchini, and A. Buonanno, \doibase10.1103/PhysRevD.95.024010 Phys. Rev. **D95**, 024010 (2017), http://arxiv.org/abs/1607.05661 arXiv:1607.05661 [gr-qc] .

M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, \doibase10.1103/PhysRevLett.113.151101 Physical Review Letters **113**, 151101 (2014), http://arxiv.org/abs/1308.3271 arXiv:1308.3271 [gr-qc] .

A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Pürrer, V. Raymond, T. Chu, H. Fong, P. Kumar, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, M. A. Scheel, and B. Szilágyi, \doibase10.1103/PhysRevD.95.044028 Physical Review D **95**, 044028 (2017), http://arxiv.org/abs/1611.03703 arXiv:1611.03703 [gr-qc] .

B. P. Abbott *et al.* (Virgo, LIGO Scientific), \doibase10.1088/1361-6382/aa6854 Class. Quant. Grav. **34**, 104002 (2017b), http://arxiv.org/abs/1611.07531 arXiv:1611.07531 [gr-qc] .

Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, \doibase10.1103/PhysRevD.89.084006 Physical Review D **89**, 084006 (2014), http://arxiv.org/abs/1307.6232 arXiv:1307.6232 [gr-qc] .

J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, \doibase10.1103/PhysRevLett.115.121102 Physical Review Letters **115**, 121102 (2015), http://arxiv.org/abs/1502.07758 arXiv:1502.07758 [gr-qc] .

J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, @noop ArXiv e-prints 1701.00550 (2017), http://arxiv.org/abs/1701.00550 arXiv:1701.00550 [gr-qc] .

R. O'Shaughnessy, J. Blackman, and S. E. Field, @noop ArXiv e-prints (2017), http://arxiv.org/abs/1701.01137 arXiv:1701.01137 [gr-qc] .

J. Lange, R. O'Shaughnessy, M. Boyle, J. Calderón Bustillo, M. Campanelli, T. Chu, J. A. Clark, N. Demos, H. Fong, J. Healy, D. Hemberger, I. Hinder, K. Jani, B. Khamesra, L. E. Kidder, P. Kumar, P. Laguna, C. O. Lousto, G. Lovelace, S. Ossokine, H. Pfeiffer, M. A. Scheel, D. Shoemaker, B. Szilagyi, S. Teukolsky, and Y. Zlochower, @noop ArXiv e-prints (2017), http://arxiv.org/abs/1705.09833 arXiv:1705.09833 [gr-qc] .

M. Pürrer, \doibase10.1088/0264-9381/31/19/195010 Classical and Quantum Gravity **31**, 195010 (2014), http://arxiv.org/abs/1402.4146 arXiv:1402.4146 [gr-qc] .

S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, \doibase10.1103/PhysRevD.93.044007 Phys. Rev. D **93**, 044007 (2016b).

C. J. Moore and J. R. Gair, \doibase10.1103/PhysRevLett.113.251101 Physical Review Letters **113**, 251101 (2014), http://arxiv.org/abs/1412.3657 arXiv:1412.3657 [gr-qc] .

C. J. Moore, C. P. L. Berry, A. J. K. Chua, and J. R. Gair, \doibase10.1103/PhysRevD.

93.064001 Physical Review D **93**, 064001 (2016), http://arxiv.org/abs/1509.04066 arXiv:1509.04066 [gr-qc] .

S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, \doibase10.1103/PhysRevX.4.031006 Physical Review X **4**, 031006 (2014), http://arxiv.org/abs/1308.3565 arXiv:1308.3565 [gr-qc] .

We interpolate fixed-chirp-mass waveforms, whereas P14 interpolated fixed-total-mass waveforms. This results in interpolations which are over the same frequency range rather than interpolations over the same geometric frequency range.

D. R. Jones, M. Schonlau, and W. J. Welch, \doibase10.1023/A:1008306431147 J. of Global Optimization **13**, 455 (1998).

D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne, \doibase10.1137/S1064827503426693 SIAM Journal on Scientific Computing **26**, 448 (2004), http://arxiv.org/abs/https://doi.org/10.1137/S1064827503426693 https://doi.org/10.1137/S1064827503426693 .

We assume the prior mean is zero, as we only interpolate functions which are de-meaned, i.e. which have their mean subtracted off.

C. E. Rasmussen and C. K. I. Williams, @noop *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (The MIT Press, 2005).

See §4.1.1 in Rasmussen and Williams 2005 for the definition of mean-square differentiability.

I. Andrianakis and P. G. Challenor, \doibasehttp://dx.doi.org/10.1016/j.csda.2012.04.020 Computational Statistics & Data Analysis **56**, 4215 (2012).

The correlations between $c_i$'s are outside the scope of this paper, and will be considered in future work.

This noise curve can be found at https://dcc.ligo.org/LIGO-G1501223/public.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, @noop Journal of Machine Learning Research **12**, 2825 (2011).

E. Jones, T. Oliphant, P. Peterson, *et al.*, `http://www.scipy.org/` SciPy: Open source scientific tools for Python, (2001–).

C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, `\doibase10.1145/279232.279236` ACM Trans. Math. Softw. **23**, 550 (1997).

The spins of the inspiraling objects are assumed to be equal and aligned with the orbital angular momentum.

The computational time to build a @ plus3@ minus4@ @ plus2@ @ plus2@ minus2@ GPR given hyperparameters and $N$ training waveforms goes as $\sim O(N^3)$. Additional cost is incurred by the hyperposterior maximization.

M. Pürrer, `\doibase10.1103/PhysRevD.93.064041` Physical Review D **93**, 064041 (2016), http://arxiv.org/abs/1512.02248 arXiv:1512.02248 [gr-qc]

.

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Phys. Rev. Lett., 116, 061102. `https://link.aps.org/doi/10.1103/PhysRevLett.116.061102`

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, ApJl, 818, L22

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, Phys. Rev. Lett., 119, 161101. `https://link.aps.org/doi/10.1103/PhysRevLett.119.161101`

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016b, Physical Review Letters, 116, 241103

—. 2016c, Physical Review X, 6, 041015

—. 2017a, Physical Review Letters, 118, 221101

—. 2017b, ApJl, 851, L35

—. 2017c, Physical Review Letters, 119, 141101

Arcavi, I., McCully, C., Hosseinzadeh, G., et al. 2017, ApJl, 848, L33

Bernstein, G. M., Abbott, T. M. C., Desai, S., et al. 2017, , 129, 114502

Chen, H.-Y., Fishbach, M., & Holz, D. E. 2018, , 562, 545

Connaughton, V., Burns, E., Goldstein, A., et al. 2016, ApJl, 826, L6

Coulter, D. A., Foley, R. J., Kilpatrick, C. D., et al. 2017, Science, 358, 1556

Cowperthwaite, P. S., & Berger, E. 2015, ApJ, 814, 25

Cowperthwaite, P. S., Berger, E., Soares-Santos, M., et al. 2016, ApJl, 826, L29

de Mink, S. E., & King, A. 2017, ApJl, 839, L7

De Vicente, J., Sánchez, E., & Sevilla-Noarbe, I. 2016, , 459, 3078

D'Orazio, D. J., & Loeb, A. 2018, Physical Review D, 97, 083008

Drlica-Wagner, A., Sevilla-Noarbe, I., Rykoff, E. S., et al. 2018, ApJs, 235, 33

Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, , 150, 150

Goldstein, D. A., D'Andrea, C. B., Fischer, J. A., et al. 2015, , 150, 82

Gupta, R. R., Kuhlmann, S., Kovacs, E., et al. 2016, , 152, 154

Guy, J., Sullivan, M., Conley, A., et al. 2010, Astronomy and Astrophysics, 523, A7

Herner, K., et al. in preparation

Kessler, R., Bassett, B., Belov, P., et al. 2010, , 122, 1415

Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, , 121, 1028

Kessler, R., Marriner, J., Childress, M., et al. 2015, , 150, 172

LIGO-Virgo Collaboration. 2017a, LIGO/Virgo G297595: Identification of a GW Binary Merger Candidate, ,

—. 2017b, LIGO/Virgo G297595: Updated localization from LIGO and Virgo data, ,

Lipunov, V. M., Kornilov, V., Gorbovskoy, E., et al. 2017, , 465, 3656

Loeb, A. 2016, ApJl, 819, L21

McKernan, B., Ford, K. E. S., Bellovary, J., et al. 2018, ApJ, 866, 66

Morganson, E., Gruendl, R. A., Menanteau, F., et al. 2018, ArXiv e-prints, arXiv:1801.03177

Morokuma, T., Tanaka, M., Asakura, Y., et al. 2016, Publications of the Astronomical Society of Japan, 68, L9

Neilsen, Jr., E. H., Bernstein, G., Gruendl, R., & Kent, S. 2016, doi:10.2172/1250877

Pâris, I., Petitjean, P., Ross, N. P., et al. 2017, Astronomy and Astrophysics, 597, A79

Perna, R., Lazzati, D., & Giacomazzo, B. 2016, ApJ, 821, L18

Phinney, E. S. 2009, in Astronomy, Vol. 2010, astro2010: The Astronomy and Astrophysics Decadal Survey

Planck Collaboration. 2016, Astronomy and Astrophysics, 594, A13

Schutz, B. F. 1986, , 323, 310

Sevilla-Noarbe, I., Hoyle, B., Marchã, M. J., et al. 2018, ArXiv e-prints, arXiv:1805.02427

Singer, L. P., & Price, L. R. 2016, Physical Review D, 93, 024013

Smartt, S. J., Chambers, K. C., Smith, K. W., et al. 2016a, , 462, 4094

—. 2016b, ApJl, 827, L40

Smith, G. P., Bianconi, M., Jauzac, M., et al. 2018, arXiv e-prints, arXiv:1805.07370

Soares-Santos, M., Holz, D. E., Annis, J., et al. 2017, ApJl, 848, L16

Soares-Santos, M., Kessler, R., Berger, E., Annis, J., et al. 2016, ApJl, 823, L33

Stalder, B., Tonry, J., Smartt, S. J., et al. 2017, ApJ, 850, 149

Stone, N. C., Metzger, B. D., & Haiman, Z. 2017, , 464, 946

The DES Collaboration, the LIGO Scientific Collaboration, the Virgo Collaboration, et al. 2019, arXiv e-prints, arXiv:1901.01540

The LIGO Scientific Collaboration, & the Virgo Collaboration. 2018, arXiv e-prints, arXiv:1811.12907

Valenti, S., David, Sand, J., et al. 2017, ApJl, 848, L24

Woosley, S. E. 2016, ApJl, 824, L10

Yoshida, M., Utsumi, Y., Tominaga, N., et al. 2017, Publications of the Astronomical Society of Japan, 69, 9

Abbott, B. P., Abbott, R., Abbott, T. D., et al. (2016a). Gw151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence. *Phys. Rev. Lett.*, 116:241103.

Abbott, B. P., Abbott, R., Abbott, T. D., et al. (2016b). Localization and broadband follow-up of the gravitational-wave transient gw150914. *The Astrophysical Journal Letters*, 826(1):L13.

Abbott, B. P., Abbott, R., Abbott, T. D., et al. (2016c). Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, 116(6):061102.

Abbott, B. P., Abbott, R., Abbott, T. D., et al. (2016d). Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo. *Living Reviews in Relativity*, 19.

Annis, J., Soares-Santos, M., Berger, E., et al. (2016). A Dark Energy Camera Search for Missing Supergiants in the LMC After the Advanced LIGO Gravitational Wave Event GW150914. *ArXiv e-prints*.

Barnes, J. and Kasen, D. (2013). Effect of a High Opacity on the Light Curves of Radioactively Powered Transients from Compact Object Mergers. *ApJ*, 775:18.

Barnes, J., Kasen, D., Wu, M.-R., and Mart'inez-Pinedo, G. (2016). Radioactivity and thermalization in the ejecta of compact object mergers and their impact on kilonova light curves. *ArXiv e-prints*.

Berger, E., Fong, W., and Chornock, R. (2013). An r-process Kilonova Associated with the Short-hard GRB 130603B. *Astrophys. J.*, 774:L23.

Bernstein, J. P., Kessler, R., Kuhlmann, S., et al. (2012). Supernova Simulations and Strategies for the Dark Energy Survey. *ApJ*, 753:152.

Bilicki, M., Jarrett, T. H., Peacock, J. A., Cluver, M. E., and Steward, L. (2014). Two Micron All Sky Survey Photometric Redshift Catalog: A Comprehensive Three-dimensional Census of the Whole Sky. *ApJs*, 210:9.

Bonnett, C., Troxel, M. A., Hartley, W., et al. (2015). Redshift distributions of galaxies in the DES Science Verification shear catalogue and implications for weak lensing. *ArXiv e-prints*.

Cowperthwaite, P. S. and Berger, E. (2015). A Comprehensive Study of Detectability and Contamination in Deep Rapid Optical Searches for Gravitational Wave Counterparts. *ApJ*, 814:25.

Cowperthwaite, P. S., Berger, E., Soares-Santos, M., et al. (2016). A DECam Search for an Optical Counterpart to the LIGO Gravitational Wave Event GW151226. *ArXiv e-prints*.

Dalal, N., Holz, D. E., Hughes, S. A., and Jain, B. (2006). Short GRB and binary black hole standard sirens as a probe of dark energy. *Physical Review D*, 74(6):063006.

Diehl, H. T., Neilsen, E., Gruendl, R., et al. (2016). The dark energy survey and operations: years 1 to 3.

Dilday, B., Kessler, R., Frieman, J. A., et al. (2008). A Measurement of the Rate of Type Ia Supernovae at Redshift z  0.1 from the First Season of the SDSS-II Supernova Survey. *ApJ*, 682:262–282.

Drout, M. R., Chornock, R., Soderberg, A. M., et al. (2014). Rapidly Evolving and Luminous Transients from Pan-STARRS1. *ApJ*, 794:23.

Flaugher, B., Diehl, H. T., Honscheid, K., et al. (2015). The Dark Energy Camera. , 150:150.

Fong, W., Berger, E., Margutti, R., and Zauderer, B. A. (2015). A Decade of Short-duration Gamma-Ray Burst Broadband Afterglows: Energetics, Circumburst Densities, and Jet Opening Angles. *ApJ*, 815:102.

Goldstein, D. A., DAndrea, C. B., Fischer, J. A., et al. (2015). Automated Transient Identification in the Dark Energy Survey. , 150:82.

Gupta, R. R., Kuhlmann, S., Kovacs, E., et al. (2016). Host Galaxy Identification for Supernova Surveys. *ArXiv e-prints*.

Guy, J., Sullivan, M., Conley, A., et al. (2010). The Supernova Legacy Survey 3-year sample: Type Ia supernovae photometric distances and cosmological constraints. *Astronomy and Astrophysics*, 523:A7.

Ji, A., Frebel, A., Chiti, A., and Simon, J. (2016). R-process enrichment from a single event in an ancient dwarf galaxy. *Nature*, 531(7596):610–613.

Jin, Z.-P., Fan, Y.-Z., and Wei, D.-M. (2016a). An r-process macronova/kilonova in GRB 060614: evidence for the merger of a neutron star-black hole binary. In *European Physical Journal Web of Conferences*, volume 109 of *European Physical Journal Web of Conferences*, page 08002.

Jin, Z.-P., Hotokezaka, K., Li, X., Tanaka, M., D'Avanzo, P., Fan, Y.-Z., Covino, S., Wei, D.-M., and Piran, T. (2016b). The Macronova in GRB 050709 and the GRB-macronova connection. *Nature Communications*, 7:12898.

Kessler, R., Bassett, B., Belov, P., et al. (2010). Results from the Supernova Photometric Classification Challenge. , 122:1415–1431.

Kessler, R., Bernstein, J. P., Cinabro, D., et al. (2009). SNANA: A Public Software Package for Supernova Analysis. , 121:1028–1035.

Kessler, R., Guy, J., Marriner, J., et al. (2013). Testing Models of Intrinsic Brightness Variations in Type Ia Supernovae and Their Impact on Measuring Cosmological Parameters. *ApJ*, 764:48.

Kessler, R., Marriner, J., Childress, M., and The DES Collaboration (2015). The Difference Imaging Pipeline for the Transient Search in the Dark Energy Survey. , 150:172.

Li, L.-X. and Paczyński, B. (1998). Transient Events from Neutron Star Mergers. *ApJl*, 507:L59–L62.

Li, W., Chornock, R., Leaman, J., et al. (2011). Nearby supernova rates from the lick observatory supernova search. iii. the rate-size relation, and the rates as a function of galaxy hubble type and colour. *Monthly Notices of the Royal Astronomical Society*, 412(3):1473–1507.

Lippuner, J. and Roberts, L. F. (2015). r-process Lanthanide Production and Heating Rates in Kilonovae. *ApJ*, 815:82.

LSST Science Collaboration, Abell, P. A., Allison, J., et al. (2009). LSST Science Book, Version 2.0. *ArXiv e-prints*.

Metzger, B. D. and Berger, E. (2012). What is the Most Promising Electromagnetic Counterpart of a Neutron Star Binary Merger? *Astrophys. J.*, 746:48.

Metzger, B. D. and Fernández, R. (2014). Red or blue? A potential kilonova imprint of the delay until black hole formation following a neutron star merger. , 441:3444–3453.

Narayan, R., Paczynski, B., and Piran, T. (1992). Gamma-ray bursts as the death throes of massive binary stars. *ApJl*, 395:L83–L86.

Paczynski, B. (1986). Gamma-ray bursters at cosmological distances. *ApJl*, 308:L43–L46.

Rest, A., Scolnic, D., Foley, R. J., et al. (2014). Cosmological Constraints from Measurements of Type Ia Supernovae Discovered during the First 1.5 yr of the Pan-STARRS1 Survey. *ApJ*, 795:44.

Sako, M., Bassett, B., Becker, A. C., et al. (2014). The Data Release of the Sloan Digital Sky Survey-II Supernova Survey. *ArXiv e-prints*.

Schutz, B. (1986). Determining the Hubble constant from gravitational wave observations. *Nature*, 323:310–311.

Shen, K. J., Kasen, D., Weinberg, N. N., Bildsten, L., and Scannapieco, E. (2010). Thermonuclear .ia supernovae from helium shell detonations: Explosion models and observables. *The Astrophysical Journal*, 715(2):767.

Soares-Santos, M., Kessler, R., Berger, E., et al. (2016). A Dark Energy Camera Search for an Optical Counterpart to the First Advanced LIGO Gravitational Wave Event GW150914. *ArXiv e-prints*.

Tanaka, M. and Hotokezaka, K. (2013). Radiative transfer simulations of neutron star merger ejecta. *The Astrophysical Journal*, 775(2):113.

Tanvir, N. R., Levan, A. J., Fruchter, A. S., et al. (2013). A 'kilonova' associated with the short-duration $\gamma$-ray burst GRB 130603B. , 500:547–549.

The LIGO Scientific Collaboration, the Virgo Collaboration, Abbott, B. P., Abbott, R., Abbott, T. D., et al. (2016). Upper limits on the rates of binary neutron star and neutron-star–black-hole mergers from Advanced LIGO's first observing run. *ArXiv e-prints*.