Probing High Energy Physics Through Gravitational Waves

by

Tim Whittaker

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Physics

Waterloo, Ontario, Canada, 2021

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Tim Whittaker was the sole author for Chapters 1, and 2 which were written under the supervision of Dr. William East and were not written for publication. This thesis consists in part of a manuscript written for publication. Exceptions to sole authorship of material are as follows:

Chapter 3 This research was conducted at the Perimeter Institute for Theoretical Physics under the supervision of Dr. William East, Dr. Luis Lehner, Dr. Stephen Green and Dr. Huan Yang. The data was produced and analysed by Tim Whittaker. Each author provided intellectual input on manuscript drafts.

Abstract

Over the last few years, gravitational wave detections have become ubiquitous, giving the physics community vast information about fundamental physics. As some of the universe's highest energy events, neutron mergers allow us to explore extreme matter states through the gravitational waves they emit. The gravitational waves from binary neutron star mergers allow us, among other things, to probe the physics of the densest matter, reveal the equation of state of neutron stars, learn about the mechanism behind gamma ray bursts, and test general relativity itself. At the same time, black holes also allow us to test general relativity and probe the fields in their surroundings. In particular, black holes could shine a light on massive boson fields proposed by extensions of the Standard Model. Massive boson under the right circumstances will form bound states around black holes. Under the right conditions, bosons will extract energy and angular momentum from spinning black holes through superradiance. This energy extraction mechanism, along with the bound bosons, causes a boson cloud to grow around the black hole which dissipates its energy through gravitational waves. Detecting these gravitational waves can then help us learn about the bosons bound to the black hole, thereby making the black hole akin to a particle detector. In this thesis we further the understanding of the dynamics of massive boson superradiance instability by extending previous studies to include the self-interactions of the bosons. We then propose a phenomenological model for binary neutron star post-merger waveforms. The proposed model is based on a machine learning technique that requires large amounts of data. We attempt to estimate how much data would be required to have a functional model and discuss the issues that arise when validating the model.

Acknowledgements

I would like to thank my research supervisor Professor William East for his complete support and skillful guidance throughout the project.

I would also like to thank my graduate committee members, Professor Will Percival, Professor Luis Lehner and Professor Huan Yang for critically examining my research performance and providing helpful feedback.

Thank you as well to Stephen Green who has thought me a lot about machine learning.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

With ground-based gravitational wave detectors such as LIGO and VIRGO having detected over 50 gravitational waves from black hole (BH) mergers, [3, 2, 7, 6, 8, 48, 10, 15, 14, 12], and several ones involving neutron stars [5, 13, 11, 4], gravitational wave astronomy is well on its way to giving the scientific community an abundance of data to learn about fundamental physics. We can only expect to see more detection with increased quality with next generation-detectors [1, 16, 23]. Gravitational waves by allowing us to tap into large sums of energy through compact objects open up the possibility of learning about fundamental physics in regimes not previously available. The gravitational signals from these compact objects allow us, among other things, to (1) test our theories of gravity [9, 87], (2) probe particle physics through BH instabilities [37, 22], (3) understand the behaviour of matter in extreme conditions, which we could not recreate in experiments here on earth [5, 13, 11, 4].

A potentially exciting avenue of gravitational wave physics is to access the large sums of energy of BHs and use them as particle detectors. Penrose first showed that one could extract energy from BHs [86] through wave scattering; then, soon after, as we will see in Chapter 2, the Zel'Dovich group demonstrated a similar energy extraction process for bosonic waves through superradiance [103]. Since low-frequency massive bosons form bound states around BHs [88], superradiance leads to an instability where the BH will dump its energy and angular momentum into the bound bosons, leading to gravitational radiation. This means that we can use BHs to detect massive bosons that only interact weakly with gravity, such as axions proposed in [21, 22] or other massive bosons proposed in many extensions of the Standard Model [46, 65, 97]. Detecting the gravitational waves from these sources from the stochastic background is a prospect already being explored in [39, 40]. There have also been studies exploring the dynamics of boson superradiant, which are

weakly coupled to gravity that establish the existence of the instability and explore some of the phenomenology along. These studies considered bosons that have no self-interactions but it is reasonable to expect self-interaction of the bosons at considerable energies. Work still needs to be done to understand the nonlinear dynamics of self-interacting boson fields around BHs and the potential gravitational wave signatures.

Neutron stars, being the densest detected objects after BHs, are of interest in understanding matter's behavior in extreme conditions. They are filled with mysteries such as their equation of state, and the mechanism behind gamma ray bursts [94, 90, 80, 84, 93, 51]. The detection of gravitational waves emitted by neutron star mergers can offer hints to help understand these mysteries. Specifically, the detection of the post-merger gravitational waves from neutron stars which avoid prompt collapse to a black hole, would allow us to constrain the equation of state of neutron stars, test deviation from General Relativity, learn about the behaviour of hot matter at supra-nuclear densities, and help understand the inner engine of gamma ray bursts. So far, detecting a post-merger has not occurred and is prohibitively difficult due to theoretical uncertainties.

In Chapter. 2 of this thesis, we study the superradiance instability of massive bosons modeled by Proca fields with self-interactions around spinning black holes through numerical simulations. This work extends previous studies of the superradiance instability where nonlinear studies of Proca fields with no self-interaction were done. Understanding how the nonlinear potential from the self-interaction affects the instability is crucial since it is possible that the self-interactions may lead to more dynamical processes not present in the purely massive case. In Chapter. 3, we propose a phenomenological model for binary neutron star post-merger gravitational waves and attempt to establish an estimate of data required for the model to be of use. We discuss subtleties that arise when trying to evaluate the performance of the model. This model, if given enough data to train it, will help the community build a waveform bank that can be used to detect the post-merger gravitational waves from binary neutron stars.

Chapter 2

Superradiant Instability and Backreaction of Self-Interacting Massive Vector Fields around Spinning Black Holes

Measurements of stellar and supermassive BHs find their masses to be $\gtrsim 5M_\odot$ for stellar BHs and $\lesssim 10^{10}M_\odot$ for supermassive BHs. That is, a large amount of energy is stored in astrophysical BHs. Under the right conditions, fields around BHs can extract some of the mass and angular momentum of the BHs through a process called superradiance. If the particles are bosonic with small masses ($\lesssim 10^{-10}$ eV), then the bound states of these particles are susceptible to exponential amplification by astrophysically relevant BHs. These bound states then accrue large occupation numbers that lead to many potential detection opportunities. The particle cloud may dissipate its energy through gravitational waves, through bosonic radiation which could be detected here on earth. These signals can then be used to constrain particle models. Thus, BHs can be used as particle detectors. In the case of earth based detectors such as LIGO, there would be an opportunity to probe ultralight bosons of masses $10^{-10} - 10^{-14}$ eV using stellar BHs. The space based detectors such as LISA on the other hand could use supermassive BHs to probe bosons with masses ranging from 10^{-16} to 10^{-20} eV.

Many beyond Standard Model particles are ultralight bosons. Some examples include the QCD axion which is a massive boson conjectured to solve the CP problem in QCD, and the massive dark photon proposed as a dark matter candidate. The ultralight bosons

generically have self-interactions at high energies. Understanding the instability when particles are interacting is crucial to understand the potential signals from these systems. While there have been many studies of the superradiant instability of bosons around spinning BHs, there are no studies following the evolution of the instability for vector bosons with self-interaction. This work attempts to fill this gap.

In this chapter, we study the superradiance instability of vector bosons with self interactions. We model the vector bosons with a Proca field and use a quartic potential as the self-interaction term. Due to the complexity of the equations for the system, we numerically evolved Einstein's field equations coupled to a Proca field with quartic self-interaction in axissymetry. We use a small seed field as the initial data of the Proca field. We find that the attractive self-interaction leads to less energy and angular momentum is extracted at saturation when compared to the case where the Proca field has no self-interaction. When the self-interaction is repulsive, then the Proca field extracts more energy and angular momentum at saturation. This is consistent with the frequency shift observed at saturation, when the self-interaction is attractive, the frequency of the cloud at saturation is greater than in the non interactive case. Similarly, when it is repulsive, the frequency is less than the non interactive case. Comparing this frequency shift to perturbation results, we find that our numerical results deviate quickly when as the self-interaction increases indicating our studied scenarios are beyond the nonrelativistic limit. Finally, we study the possibility of the cloud collapsing in the bosenova scenario. The bosenova scenario states that if the self-interaction strength is large, then the cloud will become unstable as it extracts energy and collapses in itself. We find no evidence of the Proca cloud collapsing in any of our simulations. To test if we were close to the scenario, we study the to profile of the clouds with different self-interactions. We find the self-interaction did not significantly affect the profile of the cloud indicating that we are not close to the bosenova scenario.

2.1 Background

The equivalence principle imposes that all matter must interact with gravitational fields the same way. This means that matter such as dark matter, which interacts weakly with matter in the Standard Model, will be affected by strong gravitational fields such as the gravitational fields of BHs. Additionally, in certain scenarios, the particles can extract energy from the BH with processes such as the Penrose process [86] or the superradiant process [103]. In the case of bosons, at low-frequency radiation stay bound to the BH [88], forming a boson cloud around the BH, which leads to an instability – the superradiance instability. The cloud's gained energy can then be dissipated through gravitational radiation.

The BH superradiant instability opens up the possibility to observe massive boson particles which weakly interact with gravity through the gravitational waves emitted by the bound cloud.

To help understand how superradiance extracts energy and angular momentum from BHs, we will go through the derivation of the instability with a Kerr BH. Superradiance around a Kerr BH can be understood through BH thermodynamics [30, 38]. We begin by considering a Kerr-(A)dS BH with mass M and spin a. In Boyer-Lindquist coordinates, the metric reads as,

$$ds^2 = -\frac{\Delta_r}{\rho^2}\left(dt - \frac{a}{\Sigma}\sin^2\vartheta d\varphi\right) + \frac{\rho^2}{\Delta_r}dr^2 + \frac{\rho^2}{\Delta_\vartheta}d\vartheta^2 + \frac{\Delta_\vartheta}{\rho^2}\left(adt - \frac{r^2+a^2}{\Sigma}d\varphi\right)^2 \tag{2.1}$$

where,

$$\Delta_r = (r^2 a^2)\left(1 - \frac{\Lambda}{3}r^2\right) - 2Mr \qquad \Sigma = 1 + \frac{\Lambda}{3}a^2$$
$$\Delta_\vartheta = 1 + \frac{\Lambda}{3}a^2\cos^2\vartheta \qquad \rho^2 = r^2 + a^2\cos^2\vartheta.$$

We can now define the area (A_H), temperature (T_H), and angular velocity of the event horizon (Ω_H),

$$A_H = \frac{4\pi(r_+^2+a^2)}{\Sigma}, \quad T_H = \frac{r_+\left(1 - \frac{\Lambda}{3}a^2 - \Lambda r_+^2 - a^2 r_+^{-2}\right)}{4\pi(r_+^2+a^2)}, \quad \Omega_H = \frac{a}{r_+^2+a^2}\left(1 + \frac{\Lambda}{3}a^2\right).$$

With $r_+ = M + \sqrt{M^2 - a^2}$ being the location of the event horizon. We now consider the first law which relates changes in the BH's mass M, area A_H, and angular momentum J,

$$\delta M = \frac{k}{8\pi}\delta A_H + \Omega_H \delta J. \tag{2.2}$$

Where $k = 2\pi T_H$. If the matter around the BH obeys the weak energy conditions, then we get the second law which states the BH area can only increase, that is,

$$\delta A \geq 0. \tag{2.3}$$

Now consider the energy and angular momentum flux of a wave with frequency ω and azimuthal number m

$$\delta J/\delta M = m/\omega. \tag{2.4}$$

Now using this in eq. 2.2,

$$\delta M = \frac{\omega k}{8\pi} \frac{\delta A_H}{\omega - m\Omega_H}. \tag{2.5}$$

Now using the second law, $\delta A_H \geq 0$ we get the condition for superradiance,

$$\omega \leq m\Omega_H. \tag{2.6}$$

Where $\Omega_{\rm BH}$ is the BH horizon frequency and m is the azimuthal number of the wave.

This energy and angular momentum extraction process is susceptible to becoming unstable if particles do not escape from the BH, i.e. form a bound state around the BH; this is the superradiance instability. In principle, the extraction will happen until eq. 2.6 becomes an equality. It was shown that low-frequency boson radiation with mass could form bound states around BHs [88]. Massive bosons have a natural mechanism to stay bound to BHs, making them subject to superradiance instability.

Here we will be specifically interested in massive bosons, which have small masses, the ultralight bosons. To motivate the study of ultralight bosons under the superradiant instability, we consider the bosons from the Standard Model, such as the W, Z or Higgs boson; we notice that they are either massless or have large masses. Meanwhile, ultralight bosons have only been proposed as extensions to the Standard Model and have yet to be detected. Since the superradiance instability with a BH of mass M and bosons of mass μ requires $\mu M \ll 1$ [38], to be pertinent, the instability would be sensitive to ultralight bosons. Considering stellar mass BH, then would allow to probe bosons with $\mu \leq 10^{-10}$ eV, which might occur in scenarios such as the string axisverse [21], the QCD axion [20], and dark photon [83]. The superradiance instability then allows us to explore an energy regime not previously available.

Using the superradiance instability as a probe of ultralight bosons was first proposed in [21]. Since then, the superradiance instability has been studied for spin-0 [22, 54, 20, 62], spin-1 [56, 55, 92, 25] and spin-2 fields [36, 41]. These studies have looked at the instability with boson fields with no self-interactions, apart from a few spin-0 studies [101, 24, 62]. The numerical studies of the spin-1 case in axisymmetry find that the boson cloud can extract up to 9% of the BH's mass and up to 38% of the BH's angular momentum. The instability terminates once the cloud's frequency matches the BH's event horizon frequency [56]. In the case of the full general relativistic study with no symmetry assumption, it was found that the superradiant instability efficiently extracts a significant amount of the BH's mass to produce gravitational waves. It was found that the studied scenario would produce gravitational waves with frequency $f \approx (400\text{Hz})(60\text{M}_\odot/\text{M}_0)$ with a characteristic timescale of $t \sim 30$ sec$(M_0/60M_\odot)$ [55], where M_0 is the initial BH mass.

Many constraints on the masses from ultralight bosons have been proposed over the years. Experiments, as well as astrophysical observations, have set these constraints [82]. The superradiance instability will allow constraining the masses with gravitational wave detectors to $10^{-14} < \mu < 10^{-10}$ eV [19, 39, 20].

Here, we focus on studying quartic self-interaction since symmetric self-interacting potentials for ultralight bosons will generically contain a dominant quartic term after the massive term. For instance, the spin-0 self-interaction potential of QCD axions are of the form [52],

$$V(\phi) = \frac{\mu^2}{2}\left(1 - \cos(\phi)\right) \sim \frac{\mu^2}{4}\phi^2 + \lambda\phi^4$$

Where μ is the mass of the cloud and λ is the coupling strength. For small λ, the nonlinear interactions do not affect the superradiance instability but the interactions may become relevant for large λ or large clouds. One could expect the term to have an impact once $\lambda \sim \frac{\mu^2}{\phi^2}$. We will see that the spin-1 superradiance instability does deviate from the purely massive case when $\lambda \sim \frac{\mu^2}{X_a X^a}$ where X_a is the boson vector field.

Recently, work on the superradiant instability with self-interacting spin-0 fields in the non-relativistic regime has been done [24]. While we focus on the spin-1 case here, this work offers hints as to what to look for from the self-interaction. It is shown that having self-interactions lead to new processes for the energy and angular momentum to be transferred. While in the massive case, the BH is the source of energy and angular momentum for the cloud and gravitational radiation dissipates energy and angular momentum to infinity, the self-interaction introduces new ways for the energy to be transferred. Namely, it introduces the emission of relativistic axion waves to infinity, emission of non-relativistic axion waves to infinity, and excitation of forced oscillations. This work also discusses the possibility of the "bosenova scenario." In the bosenova scenario, the self-interaction, when strong enough, as we will see in section, can cause the radius of the cloud to become unstable, leading to collapse. The analysis performed in this work shows that the scenario is ruled out except for a small region small region in the parameter space.

It is worth noting that there is also numerical evidence for the bosenova scenario in the spin-0 case of self-interacting superradiance from massive bosons when the self-interaction is large enough [100, 101]. The numerical studies have found evidence for this occurring by starting the simulation with a field strong enough for the self-interaction to be relevant, which may have introduced the scenario artificially. It is not obvious that the cloud could reach such amplitude with self-interaction. Our study follows the cloud throughout its evolution from a seed field.

In this part of the thesis, we study, through numerical simulations, the superradiance of

a massive vector field with quartic self-interaction in axisymmetry and discuss extending the work to full 3D. We choose to study the vector case as the timescale of the instability is shorter making, the simulations more tractable. In Section. 2.2 we discuss the specifics of the system we are studying and the method for evolving the system, Section. 2.3 discusses the main findings of the axisymmetric simulations, and in Section. 2.4 we summarize.

2.2 Methods

2.2.1 Equations of Motion

To model the massive bosons, we consider a complex Proca field, $X_a = X_a^R + iX_a^I$, with mass μ, and with the non-linear potential, $V(X_a\bar{X}^a)$, to model the self-interaction. We study the complex Proca field around a BH with initial mass M_0 and dimensionless spin $a = 0.99$. The action for the Einstein-Proca system is as follows,

$$S = \int d^4x \left(\frac{R}{4} - \frac{1}{4} F_{ab}\bar{F}^{ab} + V(X_a\bar{X}^a) \right). \tag{2.7}$$

Where R is the Ricci scalar, and $F_{ab} = \nabla_a X_b - \nabla_b X_a$. As discussed in the previous section, we study the quartic potential $V(X_a\bar{X}^a) = (\frac{1}{2}\mu^2 + \lambda X_c\bar{X}^c)X_c\bar{X}^c$, to capture the dynamics of symmetric self-interacting potentials, where λ is the strength of the nonlinear interaction. Varying the action then leads to the following equations of motion for the Proca field with self-interaction,

$$\nabla_a F^{ab} = \left(\mu^2 + 4\lambda X_c\bar{X}^c\right) X^b. \tag{2.8}$$

Where ∇_a is the covariant derivative. We allow λ to take on positive and negative values to represent attractive and repulsive potentials. Note that, unlike the non-interacting Proca case, the Lorenz condition is not implied from eq. 2.8. Also, the real and imaginary parts of the field are coupled, unlike the non-interacting Proca case where the real and imaginary field are essentially independent.

The associated stress-energy tensor for the complex Proca field with self-interaction is given by,

$$\begin{aligned} T_{ab} =& \frac{1}{2}\left(F_{ac}\bar{F}_{bd} + \bar{F}_{ac}F_{bd}\right) g^{cd} - \frac{1}{4} g_{ab} F_{cd}\bar{F}^{cd} \\ &+ \frac{1}{2}\mu^2 \left(X_a\bar{X}_b + \bar{X}_a X_b - g_{ab} X_c\bar{X}^c\right) \\ &+ \frac{1}{2}\lambda X_c\bar{X}^c \left[8(X_a\bar{X}_b + \bar{X}_a X_b) - 4 g_{ab} X_c\bar{X}^c\right]. \end{aligned}$$

The EoM for the spacetime,

$$R_{ab} - \frac{1}{2} g_{ab} R = T_{ab}. \tag{2.9}$$

Where R_{ab} is the Ricci tensor, and g_{ab} is the metric. So the system of equations we want to solve is,

$$\nabla_a F^{ab} = \left(\mu^2 + 4\lambda X_c \bar{X}^c\right) X^b \tag{2.10}$$

$$R_{ab} - \frac{1}{2} g_{ab} R = T_{ab} \tag{2.11}$$

2.2.2 (3+1) Decomposition

To solve the equations of motion numerically, we first perform a (3+1) decomposition. We briefly define the (3+1) decomposition tools here following [18, 17, 61].

Consider a spacetime (M, g) which can be foliated by a one-parameter family of hypersurfaces Σ_t, $t \in \mathbb{R}$, the manifold's topology can then expressed as $M = \Sigma_t \times t$. We want to project our given fields and operators onto and orthogonally to the surface Σ_t. We begin by defining the vector field n_a, which is normal to Σ_t and is normalized as follows,

$$n_a n^a = -1.$$

We can now define a projector, γ_{ab}, onto Σ_t,

$$\gamma^a_b = \delta^a_b + n^a n_b.$$

Using γ_{ab} and n_a, we can write the full 4-dimensional metric as,

$$ds^2 = g_{ab} dx^a dx^b = -\left(\alpha^2 - \beta_a \beta^a\right) dt^2 + 2\gamma_{ab} \beta^a dt dx^b + \gamma_{ab} dx^a dx^b. \tag{2.12}$$

Where α is the lapse function and β is the shift vector. These objects reflect the coordinate degrees of freedom in general relativity. We can now define another useful object, the extrinsic curvature of Σ_t,

$$K_{ab} = -\frac{1}{2} \mathcal{L}_n \gamma_{ab} = -\frac{1}{2\alpha} \left(\partial_t - \mathcal{L}_\beta\right) \gamma_{ab}.$$

Where $\mathcal{L}$ is the Lie derivative. Its trace is,

$$K = -\nabla_c n^c.$$

We now have the tools to decompose the Proca EoMs. We begin by decomposing the Proca vector field as,

$$X_a = \chi_a + n_a \chi \qquad \chi_a := \gamma_a^b \chi_b \qquad \chi := -n^a X_a$$

and introduce the E_a and B_a fields,

$$E_a := F_{ab} n^b \qquad B^i := {}^{(3)}\epsilon^{ijk} D_j \chi_k.$$

Where D_i the covariant derivative associated with the metric on Σ_t, and ${}^{(3)}\epsilon^{abc}$ is the total antisymmetric tensor on Σ_t. We decompose the EoM by projecting eq. 2.8 and

$$\nabla_b \left[\left(\mu^2 + 4\lambda X_c \bar{X}^c \right) X^b \right] = 0$$

onto γ_{ab} and n_a. Additionally, we introduce the auxilary scalar field Z to dampen violations of the Proca field constraints as in [56, 104] by some rate σ. Doing so, we get the following set of equations for the real fields,

$$\alpha^{-1} \left(\partial_t - \mathcal{L}_\beta \right) Z = D_i E^i + (\mu^2 + 4\lambda X_c \bar{X}^c) \chi - \sigma Z \tag{2.13}$$

$$\begin{aligned} \alpha^{-1} \left(\partial_t - \mathcal{L}_\beta \right) E^i = {} & (\mu^2 + 4\lambda X_c \bar{X}^c) \chi^i + {}^{(3)}\epsilon^{ijk} \partial_j B_k \\ & + E^i K + D^i Z - {}^{(3)}\epsilon^{ijk} B_j \partial_k \log \alpha \end{aligned} \tag{2.14}$$

$$\alpha^{-1} \left(\partial_t - \mathcal{L}_\beta \right) \chi_i = -D_i \chi - E_i - \chi \log \alpha \tag{2.15}$$

$$\alpha^{-1} \left(\partial_t - \mathcal{L}_\beta \right) \chi = \Gamma_{(R)} \left(\lambda_{0(R)} - \tilde{\lambda}_{(R)} + \lambda_{m(R)} \right) \tag{2.16}$$

where,

$$\begin{aligned} \Gamma_{(R)} &= \left(1 - \Delta \chi_R^2 - (\Delta \chi_R \chi_I)(1 - \Delta \chi_I^2)^{-1} \right) \\ \lambda_{0(R)} &= -D_i \chi_R^i + \chi_R K - \chi_R^i \partial_i \log \alpha \\ \tilde{\lambda}_{(R)} &= 8\lambda (\chi_I^b [\chi_c^I D_b(\chi_I^c) - \chi_I D_b(\chi_I)] \\ & \quad \chi_I^b [\chi_c^R D_b(\chi_R^c) - \chi_R D_b(\chi_R)] + \\ & \quad \chi_I [\chi_c^I \mathcal{L}_n(\chi_I^c) + \chi_I^c \chi_c^I K] + \\ & \quad \chi_I [\chi_c^I \mathcal{L}_n(\chi_R^c) + \chi_R^c \chi_c^R K] + \\ \lambda_{m(R)} &= \Delta \chi_R \chi_I (1 - \Delta \chi_I^2)^{-1} (\lambda_{0I} - \tilde{\lambda}_I) \\ \Delta &= \frac{8\lambda}{(\mu^2 + 4\lambda X_c \bar{X}^c)}. \end{aligned}$$

The imaginary part of the equations are found by replacing $R \to I$. This is the set of equations we will be solving numerically for the axisymetric case. For the 3D case, we only evolve the real fields. The equations we solve are then,

$$\alpha^{-1}\left(\partial_t - \mathcal{L}_\beta\right) Z = D_i E^i + (\mu^2 + 4\lambda X_c X^c)\chi - \sigma Z \tag{2.17}$$

$$\begin{aligned}\alpha^{-1}\left(\partial_t - \mathcal{L}_\beta\right) E^i = {} & (\mu^2 + 4\lambda X_c X^c)\chi^i + {}^{(3)}\epsilon^{ijk}\partial_j B_k \\ & + E^i K + D^i Z - {}^{(3)}\epsilon^{ijk} B_j \partial_k \log\alpha\end{aligned} \tag{2.18}$$

$$\alpha^{-1}\left(\partial_t - \mathcal{L}_\beta\right)\chi_i = -D_i\chi - E_i - \chi\log\alpha \tag{2.19}$$

$$\alpha^{-1}\left(\partial_t - \mathcal{L}_\beta\right)\chi = \Gamma\lambda_0 \tag{2.20}$$

where,

$$\begin{aligned}\Gamma &= \left(1 - \Delta\chi^2\right) \\ \lambda_0 &= -D_i\chi^i + \chi K - \chi^i\partial_i\log\alpha \\ \Delta &= \frac{8\lambda}{(\mu^2 + 4\lambda X_c X^c)}.\end{aligned}$$

Decomposing the Einstein equations we get the two evolution equations and two constraint equations for the metric,

$$\frac{1}{2\alpha}\left(\partial_t - \mathcal{L}_\beta\right)\gamma_{ab} = -K_{ab} \tag{2.21}$$

$$\left(\partial_t - \mathcal{L}_\beta\right) K_{ab} = -D_a D_b\alpha + \alpha\left(R_{ab} - 2K_{ac}K^c_b + KK_{ab}\right) + 2\alpha S_{ab} \tag{2.22}$$

$$\mathcal{H} = {}^3R + K^2 - K_{ab}K^{ab} - 16\pi\mathcal{E} = 0 \tag{2.23}$$

$$\mathcal{M}_a = D_b K^{ab} - D^a K - 8\pi J^a = 0. \tag{2.24}$$

Where we have R_{ab} the Ricci tensor, ${}^3R$ the Ricci curvature on Σ_t and,

$$\begin{aligned}J^a &= -h^c_a n^d T_{dc} \\ \mathcal{E} &= n^a n^b T_{ab} \\ S_{ab} &= h^c_a h^d_b T_{cd}.\end{aligned}$$

2.2.3 Numerical Methods

Due to the complexity of solving the equations of the Proca-Einstein system in full 3D, we start by solving the equations in axisymmetry with a complex field. We then use

the axisymmetric solution as the initial data for the 3D case. Solving the axisymmetric case allows us to resolve the equations in (2+1) dimensions significantly reducing the computational cost.

We assume a specific azimuthal dependence on the Proca field which makes the problem a tractable (2+1) dimensional problem as in [56]. We specifically have,

$$\mathcal{L}_{\hat{\phi}} X_a = imX_a$$

Where $\hat{\phi}$ is the Killing vector in axisymmetry. We restrict our studies to cases where $m = 1$.

To solve the Einstein equations have we need to specify the gauge to deal with the coordinate degree of freedoms. Here, we use the generalized harmonic formulation [89] to evolve the Einstein equations. The idea behind the generalized harmonic formulation is to introduce four source functions, H^a, where,

$$\begin{aligned} H^a &= \Box x^a \\ &= \frac{1}{\sqrt{-g}} \partial_b \left(\sqrt{-g} g^{ab} \right). \end{aligned}$$

Now using these source functions in Einsteins equation and treating H^a as independent variables, we can rewrite the Einsteins equation as,

$$g^{de}\partial_e\partial_d g_{ab} + \partial_b g^{ed}\partial_e g_{ad} + \partial_a g^{de}\partial_e g_{db} + 2H_{(a,b)} - 2H_d\Gamma^d_{ab} + 2\Gamma^e_{db}\Gamma^d_{ea} = -8\pi\left(2T_{ab} - g_{ab}T\right) \quad (2.25)$$

where T is the trace of the stress-energy tensor, Γ^c_{ab} is the Christoffel symbols, and

$$H_{(a,b)} = (\ln\sqrt{-g})_{ab} - \frac{1}{2}\left(\partial_b g^{cd}\partial_c g_{ad} + \partial_a g^{cd}\partial_c g_{bd}\right) - \frac{g^{cd}}{2}\left(\partial_a\partial_c g_{db} + \partial_b\partial_c g_{da}\right).$$

It can be shown that this system of equations is equivalent to the Einstein equation given that the harmonic constraints are satisfied, that is,

$$C^a = H^a - \Box x^a = 0 \quad (2.26)$$

for all times. In this case, we pick the source functions to be the ones from the Kerr BH.

The imaginary and real part of the fields $\{\chi, \chi_i, E^i, Z\}$ are evolved using fourth-order Runge-Kutta time stepping and fourth-order stencils for spatial derivatives. We solve the equations on a grid with 2:1 refinement. The lowest resolution simulation uses 65×33 points on each refinement level. The convergence of the methods for the axisymmetric case can be seen in appendix A.

To solve the system in 3D, we first evolve the system in axisymmetry as above until the instability has begun but still is in the regime where the massive term dominates. We then deform the data according to the method described in appendix B and use this deformed data as initial data. We then only evolve the real part of the EoM, as described in the previous section. To simplify the 3D case further, we fix the spacetime to a Kerr background and only evolve the Proca field.

2.2.4 Measured Quantities

As in [56] we follow the cloud's energy, E, and angular momentum, J throughout the evolution,

$$E = \int \rho_E \sqrt{-g}\ d^3x = -\int \alpha T^t_t \sqrt{-g}\ d^3x \tag{2.27}$$

$$J = \int \rho_J \sqrt{-g}\ d^3x = -\int \alpha T^t_\phi \sqrt{-g}\ d^3x. \tag{2.28}$$

Where we use ρ_E and ρ_J to denote the energy and angular momentum density. We also keep track of some BH quantities. We track the apparent horizon and its area A, and its angular momentum J_{BH}. These quantities are related to the BH's mass M_{BH} by the Christodoulou formula,

$$M_{BH} = \left(M_{irr}^2 + \frac{J_{BH}^2}{4M_{irr}^2}\right)^{1/2}. \tag{2.29}$$

Where $M_{irr} = (A/16\pi)^{1/2}$.

The Newman-Penrose Weyl scalar, Ψ_4, and the scalar, ϕ_2 are also followed through the evolution. These quantities allow us to follow gravitational and Proca radiation in the far field [32, 81]. The quantities are defined using the Newman-Penrose formalism. We first define a set of independent vectors $\mathbf{e}_a$, where, $\mathbf{e}_a\mathbf{e}_b = \eta_{ab}$, this is the tetrad basis. The tetrad basis is related to the coordinate basis through the vierbein, $e_a{}^b$,

$$\mathbf{e}_a = e_a{}^b \partial_b.$$

We then rewrite the metric using tetrads,

$$g_{ab} = \mathbf{e}_a \mathbf{e}_b dx^a dx^b. \tag{2.30}$$

We can now define the null vectors in terms of the tetrads,

$$\ell^a = \frac{1}{\sqrt{2}}\left(e^a_{(0)} + e^a_{(1)}\right) \qquad k^a = \frac{1}{\sqrt{2}}\left(e^a_{(0)} - e^a_{(1)}\right)$$
$$m^a = \frac{1}{\sqrt{2}}\left(e^a_{(2)} + ie^a_{(3)}\right) \qquad \bar{m}^a = \frac{1}{\sqrt{2}}\left(e^a_{(2)} - ie^a_{(3)}\right).$$

The Weyl scalar, Ψ_4, and the scalar, ϕ_2 are then defined as,

$$\Psi_4 = C_{abcd}k^a\bar{m}^b k^c\bar{m}^d. \tag{2.31}$$

Where C_{abcd} is the Weyl tensor. Taking the asymptotic limit of Ψ_4 in the TT gauge shows us how we can extract gravitational waves from it,

$$\Psi_{4(r\to\infty)} = \partial_t^2\left(h_+ + ih_\times\right).$$

With h_+ and $h_\times$ being the two polarization of the metric perturbation. The ϕ_2 scalar is defined as,

$$\phi_2 = F_{ab}\bar{m}^a n^b. \tag{2.32}$$

In the asymptotic limit we have that ϕ_2 gives the radiation from the Proca field.

2.3 Results in axisymmetry

We observe how the self-interaction term affects the properties of the final Proca cloud and the BH through numerical simulations. We impose axisymmetry and focus on $m = 1$ as discussed in the previous section. The instability is triggered by a small seed field close to the BH with dimensionless spin $a = 0.99$ and initial mass M_0. We experiment with $-0.1 < \tilde{\lambda} = \frac{\lambda}{\mu^2} < 1.5$ and $\tilde{\mu} = M_0\mu = \{0.25, 0.30, 0.40, 0.50\}$. We find that the instability saturates at a higher frequency than the massive case leading to less energy being extracted from the BH. Similarly, when λ is negative, we find that the instability saturates at a lower frequency than the massive case, leading to more energy being extracted. We look into the possibility of the "bosenova" collapse and find no evidence of it occurring in our parameter space. We compare the numerical results to a perturbation computation of the frequency shift due to the strength of the self-interaction. We also find no evidence of gravitational radiation far from the BH. Finally, we compare the frequency shift from the non-linear potential obtained through the numerical simulations to the frequency shift expected in the non-relativistic limit. We find that the numerical results quickly deviate from the expected results. This is not too surprising as our tested scenarios are relativistic.

2.3.1 Proca Cloud Energy and Angular Momentum

We follow the energy of angular momentum of the Proca cloud as defined in eq. 2.27,2.28 throughout the evolution of the cloud. Similarly, we keep track of the spin and mass of the BH. Fig. 2.1 shows the energy of the Proca cloud throughout the instability for various values of $\tilde{\lambda}$ and $\tilde{\mu}$. The self-interacting case of the instability follows the growth of the massive instability closely up until saturation, where the self-interacting instability saturates at a lower energy for positive $\tilde{\lambda}$, which is associated with an attractive self-interaction. When the self-interaction takes on the opposite sign, which is associated with a repulsive self-interaction, the instability saturates at a larger energy than the massive case. Fig. 2.2 shows the change of mass of the BH throughout the instability for various $\tilde{\lambda}$ and $\tilde{\mu}$ over the energy of the Proca cloud. We see that the energy of the BH is efficiently being transferred to the Proca cloud. We find that the growth rate of the cloud closely follows the massive case at low energies but quickly becomes sub-exponential at larger energies. This can be seen in Fig. 2.3 where we show $\log(E/M_0)$ through time and see that for large values of $\tilde{\lambda}$, the growth of the cloud deviates from the massive case. This deviation starts the nonlinear term starts to dominate in the potential, i.e., when $\tilde{\lambda} \sim \frac{1}{X_a \bar{X}^a}$. We stop the study at $\tilde{\lambda} = 1.5$ since we found that higher values required higher resolution, which becomes computationally expensive. We can nonetheless see from Fig. 2.7 that the saturation energy follows a linear trend with respect to $\tilde{\lambda}$ and so we could expect similar behaviour for stronger self-interactions.

The angular momentum of the cloud and the BH follow similar trends to the energy of the cloud and BH. Fig. 2.4 shows how the angular momentum of the cloud at saturation is lower when there are self-interactions; this is consistent with the BH losing less angular momentum, as can be seen in Fig. 2.5. Fig. 2.6 shows the growth of the cloud's angular momentum following the massive case until $\tilde{\lambda} \sim \frac{1}{X_a \bar{X}^a}$ where the growth slows down.

We only looked at small deviations from the massive case when looking at negative values of λ. The reason being that when λ gets too negative, the energy then also becomes negative. This might be fixed by looking at a more complete potential than simply a quartic one that is unbounded from below.

2.3.2 Cloud Size & Bosenova

It is conjectured that scalar Proca fields subject to the superradiance instability will undergo "bosenova" collapse under the right circumstance. The idea is that, given a strong enough self-interaction, i.e. λ large enough, the cloud's radius will shrink due to the self-interaction

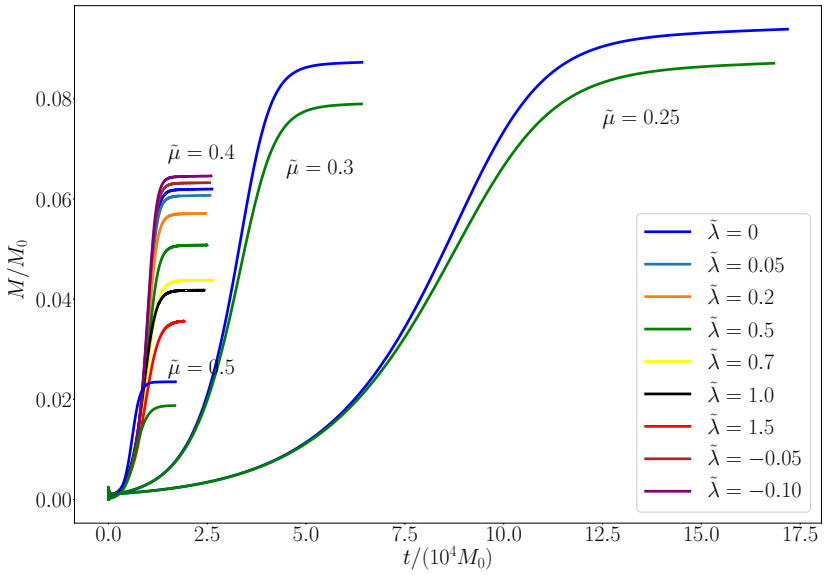

Figure 2.1: Energy of the Proca cloud through time during the superradiance instability for various $\tilde{\mu}$ and $\tilde{\lambda}$. The self-interaction affects the saturation energy of the cloud. Stronger self-interaction causes the cloud to saturate at a lower energy than the purely massive case. Conversly, negative self-interaction leads to higher energy at saturation.

while gaining energy from the BH until the cloud's radius becomes unstable and the cloud collapses. There is some numerical evidence for the scalar case [100, 102] and some evidence in a non-relativistic limit analysis [24] for small regions of the parameter space.

The numerical studies which have investigated this scenario for spin-0 bosons [100, 102] do not evolve the cloud with its nonlinearities up to the point where the bosenova occurs but instead, start with a boson cloud large enough where the non-linear interaction is significant. This may artificially lead to bosenova scenarios. Since we evolve vector bosons in this study, which have a shorter instability timescale than the scalar case, we can follow the whole non-linear evolution of the cloud.

The simulations performed showed no evidence of "bosenova" collapse occurring in the parameter space we have explored. We can nonetheless look at the radius of the Proca cloud at saturation for various λ to gain some insight into the possibility of a collapse. In [24], the Bosenova scenario in the non-relativistic limit is explored by considering the stable minima of the effective potential of the cloud's radius. In the scalar case, the potential of

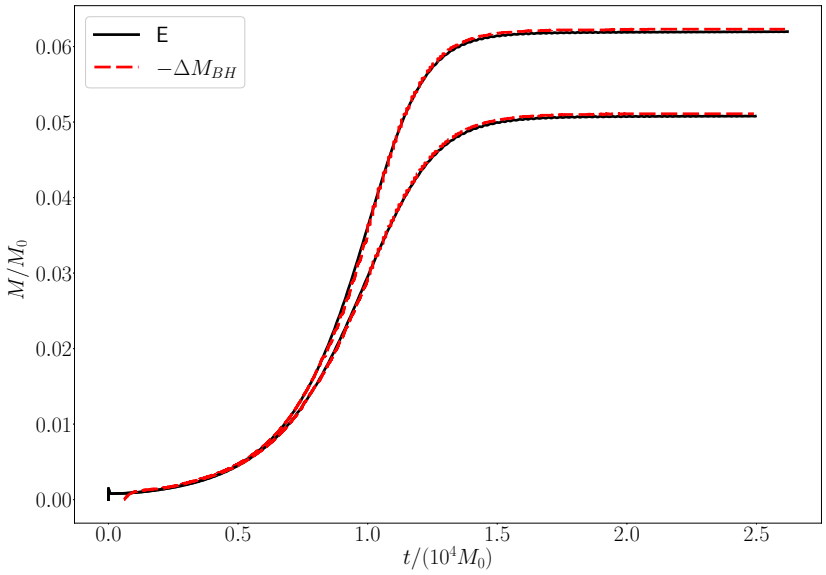

Figure 2.2: Energy of the Proca cloud and the negative change in mass of the BH through the evolution for $\tilde{\lambda} = \{0, 0.5\}$, $\tilde{\mu} = 0.4$, $a = 0.99$ and $M_0 = 1.0$. We can see that all the energy dissipated by the BH is absorbed into the cloud.

the Proca field's radius with self-interaction, ignoring self-gravity goes like,

$$V(\tilde{R}) = \frac{\alpha^4 \varepsilon}{G\mu}\left(\frac{1}{8\tilde{R}^2} - \frac{1}{4\tilde{R}} - \frac{3\alpha^3\varepsilon}{16384\pi G\tilde{R}^3 f^2}\right). \tag{2.33}$$

Where we use notation from [24]. We have the energy scale $f = \mu^2/\lambda$, $\alpha = GM\mu$, and the normalized occupation number $\varepsilon = \frac{N}{GM^2}$. The potential has then has extrema at

$$\tilde{R}^{\pm}_{\text{extrema}} = \frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{9\alpha^3\varepsilon}{4096\pi G f^2}}.$$

Now there is a point as we decrease f where the extrema vanishes and the potential then no longer has a stable minimum which then causes a "bosenova". We can then see how close to the bosenova our cases our by measuring the deviation from the Bohr radius, $a = 1/(\tilde{\mu}\mu)$. In the massive case with $\tilde{\mu} = 0.40$, we have the Bohr radius, $a = 6.25$. This will be the Bohr radius in the non-relativistic limit.

Fig. 2.8 shows the energy density of the clouds with different self-interaction strengths overlayed with the contours of the energy density normalized by the total energy. These

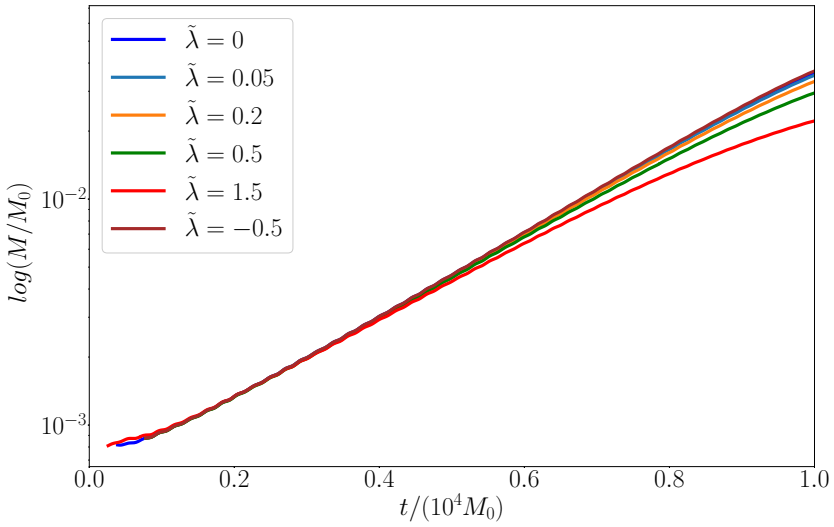

Figure 2.3: Growth rate of the cloud's energy for various self-interaction stregnths. We can see at which times to nonlinear effects of the potential become relevant. Once the nonlinear term become relevent the growth rate diverges downward from the massive only case.

plots illustrate how the energy density distribution of the cloud is not affected by the non-linear potential, and it only affects the total energy. Fig. 2.9 shows the energy density profile along the equatorial plane for $\tilde{\mu} = 0.40$, $\tilde{\lambda} = \{0, 0.05, 0.2, 0.5\}$. We can see again that the self-interaction does not appear to affect the profile of the cloud. We compare these profiles to the one expected from a cloud with a Bohr radius $a = 1/(\tilde{\mu}\mu)$. We find that the clouds all follow this profile closely. This could indicate that the strength of the self-interactions we tested is not strong enough for the bosenova scenario to happen but considering that we go up to $\tilde{\mu} = 0.5$, which is a strong self-interaction, with no deviation, we can know the bosenova scenario, if it occurs, only occurs in extreme cases. Further work with higher λ values is required to test the bosenova scenario properly.

2.3.3 Comparison to non-relativistic limit

The saturation point of the instability can be understood through the condition discussed in chapter 2, $\omega \leq m\Omega_{BH}$, when the cloud's frequency matches the BHs' event horizon frequency. This behaviour is observed in numerical studies for the massive case [56]. The instability with the self-interaction also saturates when $\omega = m\Omega_{BH}$, except that the nonlinear potential

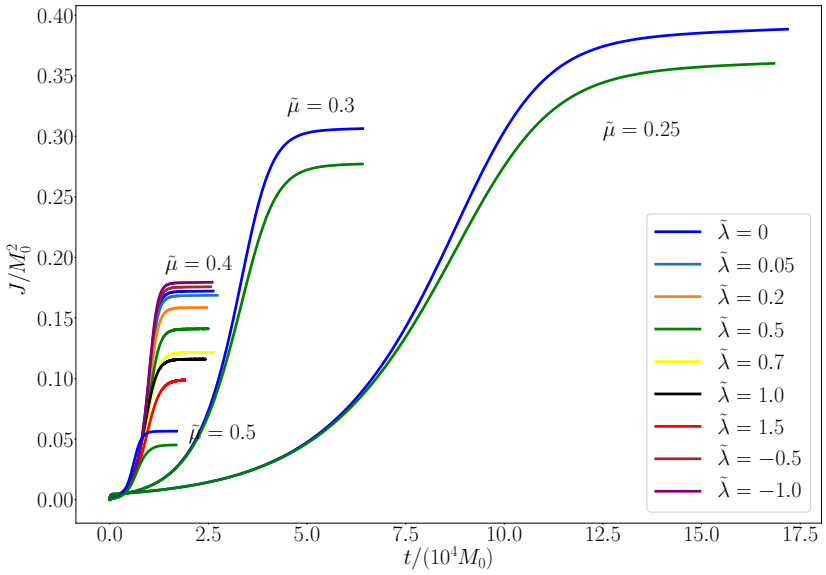

Figure 2.4: Angular momentum of the Proca cloud through the superradiance instability for various $\tilde{\mu}$ and $\tilde{\lambda}$. Positive self-interaction leads to smaller angular momentum at saturation while negative self-interaction leads to larger sums of angular momentum extracted.

causes ω to shift. This section, estimates the frequency shift caused by the nonlinear potential and compares it to the numerical results.

In the spin-0, the frequency shift arising from the nonlinear potential was estimated in ref [24] in the nonrelativistic limit. We follow a similar derivation of the frequency shift for the spin-1 case to compare the shift in frequency according to perturbation theory to the numerical result. We begin with the EoM in the non-relativistic limit,

$$D_\mu F^{\mu\nu} = \mu^2 X^\nu + \lambda X_\mu X^\mu X^\nu. \tag{2.34}$$

Where $F^{\mu\nu} = D_\mu X_\nu - D_\nu X_\mu$. Expanding we get,

$$\left(D^2 - \mu^2\right) X^\nu = \lambda X_\mu X^\mu X^\nu - D^\nu D_\mu X^\mu. \tag{2.35}$$

Now, unlike the massive case, in the self-interacting case we have that,

$$D_\mu X^\mu = -\frac{\lambda}{\mu} D_\mu \left((X_\nu X^\nu) X^\mu\right). \tag{2.36}$$

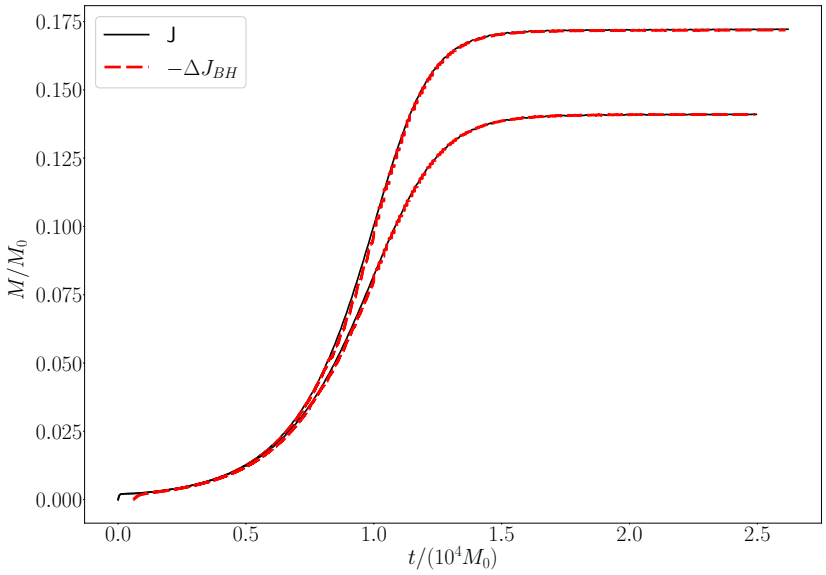

Figure 2.5: Angular momentum of the Proca cloud and the negative change in angular momentum of the BH through the evolution for $\tilde{\lambda} = \{0, 0.5\}$, $\tilde{\mu} = 0.4$, $a = 0.99$ and $M_0 = 1.0$. We can see that all the angular momentum dissipated by the BH is absorbed into the cloud.

Then if we expand D^2 to first order in r_g/r, where $r_g = GM$, we get,

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 + \mu^2\right) X^\mu - \frac{2\tilde{\mu}}{r}\left(\mu + \hat{K}\right) X^\mu = \lambda (X_\rho X^\rho) X^\mu \tag{2.37}$$

where,

$$\hat{K} = \frac{1}{r}\frac{\partial}{\partial r} - 2\left(\mu^2 + \frac{\partial^2}{\partial t^2}\right) - \frac{\hat{L}^2}{r^2} \tag{2.38}$$

is small compared to μ in the non-relativistic case and so can be omitted. Here $\hat{L}^2$ is the total momentum operator. We are then concerned with the following equation,

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 + \mu^2\right) X^\mu - \frac{2\tilde{\mu}}{r}\mu X^\mu = \lambda (X_\rho X^\rho) X^\mu. \tag{2.39}$$

We now expand X^μ in terms of λ to solve perturbatively to first order,

$$X_\mu = X_\mu^{(0)} + \lambda X_\mu^{(1)} + \dots \tag{2.40}$$

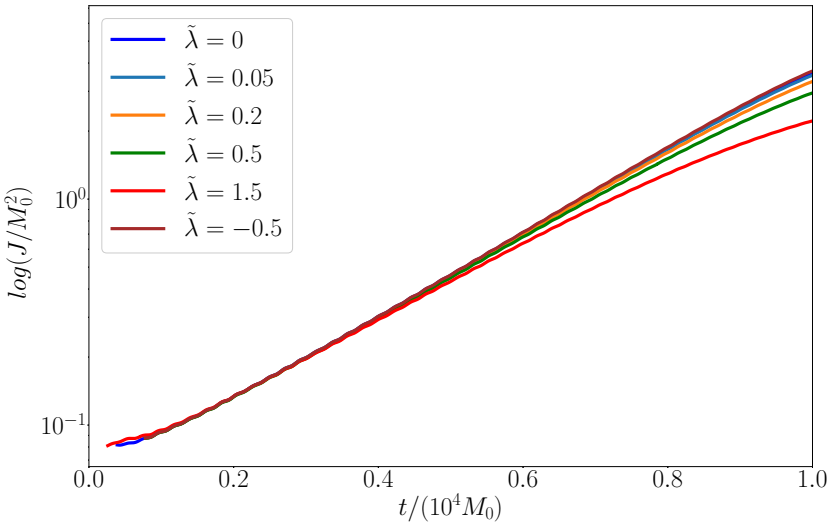

Figure 2.6: Growth rate of the cloud's energy for various self-interaction strengths. We can see at which times to nonlinear effects of the potential become relevant. Once the nonlinear term become relevant the growth rate diverges downward from the massive only case.

At zeroth order we have

$$\left(\frac{\partial^2}{\partial t^2} - \nabla^2 + \mu^2\right) X_\nu^{(0)} - \frac{2\tilde{\mu}\mu}{r} X_\nu^{(0)} = 0. \tag{2.41}$$

As in [25] this admits the bound state,

$$X^\mu = A\left(\Psi^\mu e^{-i\omega t} + c.c.\right) \tag{2.42}$$

where we can write Ψ_μ as a radial and orbital part,

$$\Psi_\mu = R^{n\ell}(r) Y_\mu^{\ell,jm}(\theta, \phi) \tag{2.43}$$

R being the hydrogenic radial wavefunction and $Y_\mu^{\ell,jm}$ being the vector spherical harmonics. We focus on the $\ell = 0$, $m = j = 1$ state as it is the fastest growing bound state [25]. The real part of this state is of the form,

$$X_i^{(0)} = A e^{-\tilde{\mu}\mu r} \begin{pmatrix} \cos(\omega t) \\ -\sin(\omega t) \\ 0 \end{pmatrix} \tag{2.44}$$

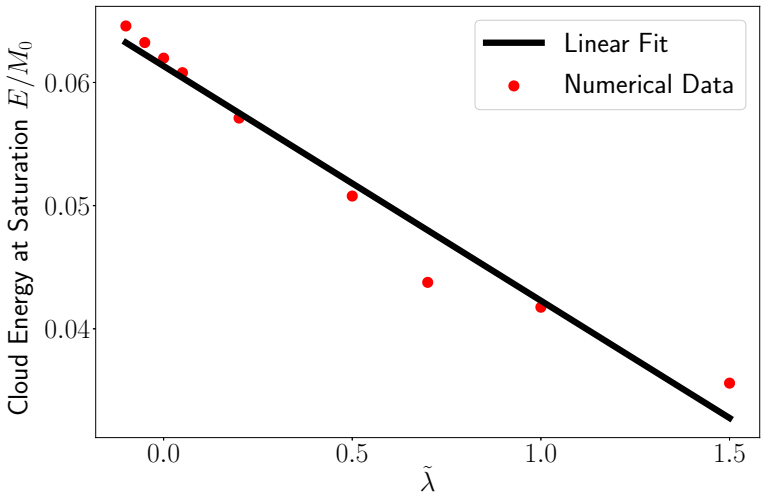

Figure 2.7: Energy of the Proca cloud at saturation as a function of $\tilde{\lambda}$. A linear fit, $y = -0.019\tilde{\lambda} + 0.061$ is overlayed on the numerical data.

$$X_0^{(0)} = A\tilde{\mu}e^{-\tilde{\mu}\mu r}\sin(\theta)\cos(\omega t - \phi) \tag{2.45}$$

The first order energy correction from the nonlinear potential is then,

$$\Delta E^{(1)} = \lambda \int_0^\infty 4\pi r^2 |X^{(0)}|^4 dr. \tag{2.46}$$

The energy of the cloud is related to the frequency through the occupation number, N,

$$E = N\omega$$

and to first order N goes as $E^{(0)}/\mu$. So the frequency shift will be,

$$\Delta\omega^{(1)} = \frac{\lambda\mu\Delta E^{(1)}}{E^{(0)}} \tag{2.47}$$

Using the zeroth order solution we get,

$$\int_0^\infty 4\pi r^2 |X|^4 dr \sim A^4 \int_0^\infty 4\pi r^2 e^{-4\tilde{\mu}\mu r} dr = 4\pi A^4 \left(\frac{2}{(-4\tilde{\mu}\mu)^3}\right)$$

where we take that $X_0^{(0)} \ll X_i^{(0)}$ in the non-relativistic limit [25]. So,

$$\Delta\omega^{(1)} \sim \frac{4\pi\lambda\mu A^4}{E^{(0)}} \left(\frac{2}{(-4\tilde{\mu}\mu)^3}\right) \tag{2.48}$$

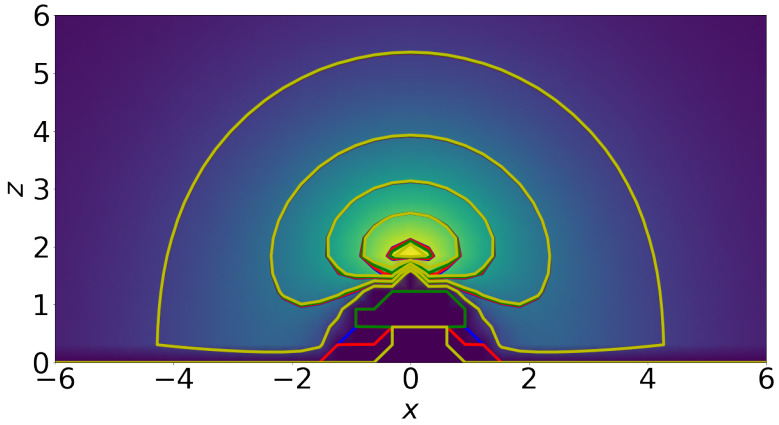

Figure 2.8: Contours of energy density with $\tilde{\lambda} = 0$ (red curves), $\tilde{\lambda} = 0.05$ (green curves), $\tilde{\lambda} = 0.2$ (yellow curves) , and $\tilde{\lambda} = 0.5$ (blue curves) overlapped with the energy density of the $\lambda = 0$ in the background normalized by the total energy of the cloud. We can see that the non linear potential does not affect the profile of the cloud but only the total energy.

Now we can rewrite A^4 in terms of the cloud's energy, $E^{(0)}$.

$$E^{(0)} = \int_0^\infty \rho_W dV \sim \int_0^\infty 4\pi r^2 \frac{\mu^2}{2}((X_0^{(0)})^2 + X_i^{(0)} X^{i(0)})dr$$

from which we get,

$$A^2 \sim -\left(\frac{2E^{(0)}\tilde{\mu}^3\mu}{\pi}\right)$$

Then,

$$\Delta\omega^{(1)} = \frac{4\pi\lambda\mu}{E^{(0)}}\left(\frac{2}{(4\tilde{\mu}\mu)^3}\right)\left(\frac{2E^{(0)}\tilde{\mu}^3\mu}{\pi}\right)^2 = \frac{\lambda}{2\pi}E^{(0)}\tilde{\mu}^3 \tag{2.49}$$

Figures 2.10, 2.11, 2.12 and 2.13 show the predicted frequency shift as a function of $\tilde{\lambda}$ with the black curve and the estimated frequency shift from the numerical simulations for $\tilde{\mu} = 0.25, 0.30, 0.40, 0.50$ as the red dots. To get an idea of goodness of fit from the perturbation calculation, we also fit the $\tilde{\mu} = 0.40$ numerical data with a line of best fit as shown by the blue curve in the figures. The fit and perturbation appear to predict $\Delta\omega$ well for weak self-interactions in most cases apart for $\tilde{\mu} = 0.25$. Since $\Delta\omega \propto E^{(0)}\tilde{\mu}^3$, we can take the fit from the $\tilde{\mu} = 0.40$ and rescale it to the other cases. This rescaled fit is show in

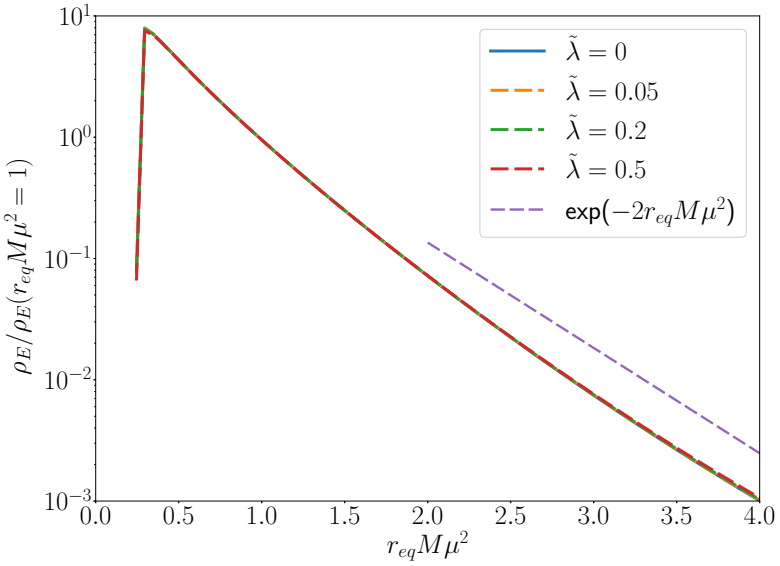

Figure 2.9: Energy density along the equator for various strengths of self- interactions. The length scale of the radius is set by multiplying it by $M\mu^2$. The self-interaction does not seem to affect the profile of the cloud, it only lowers the amplitude. We added the curve $e^{-2r_{eq}M\mu^2}$ to show that the curves follow the profile of a cloud with Bohr radius $a = 1/(\tilde{\mu}\mu)$.

figures 2.11, 2.12 and 2.13. We see that the perturbation computation differs from the fit quickly as $\tilde{\lambda}$ increases. This is to be expected since we are in the relativistic regime where we don't necessarily expect the non-relativistic results to hold.

2.4 Discussion

Combining energy extraction processes such as the superradiance with a confinement mechanism to bound matter near a BH causes an instability which allows for efficient energy extraction from spinning BHs. Bosons have a natural confinement mechanism around BHs; when the Compton wavelength is comparable to, or larger than, the horizon radius of the BH, the bosons will form a bound state. We should then expect the superradiance instability to be triggered when bosons are found around spinning BHs. Furthermore, the mass of astrophysical BHs and the mass of theoretically proposed particles such as axions makes for relevant instability timescales. This scenario opens up the possibility of using

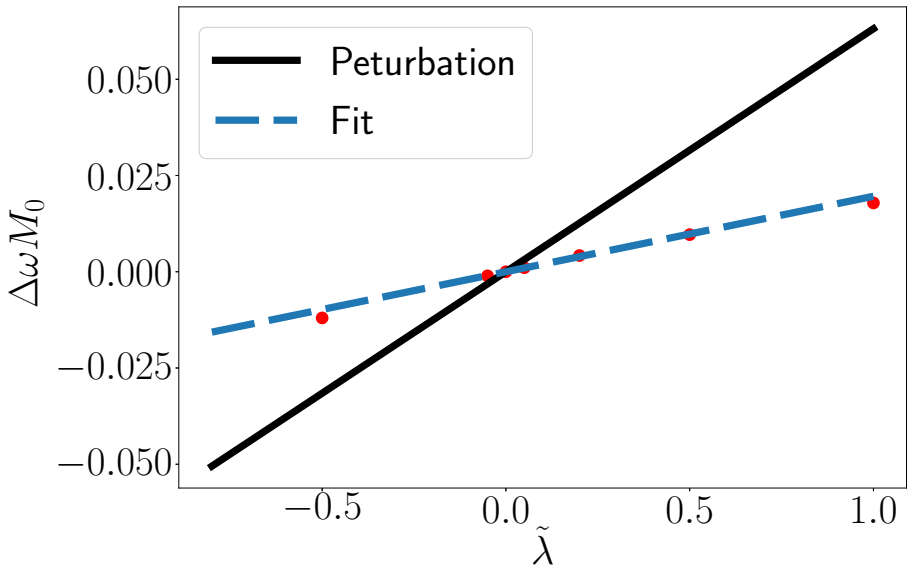

Figure 2.10: Frequency change of the cloud due to the self-interactions as a function of the self-interaction strength. We compare the numerical results (red dots) with the perturbation results (black line) for $\tilde{\mu} = 0.40$. We also include the line of best fit (dashed blue line).

BHs as particle detectors through the gravitation waves emitted by the cloud extracting the BH's energy. Allowing us to constrain a whole new regime of particles. In this chapter, we extended previous works of massive vector superradiance instability with axisymmetry by asking what happens to the instability when the bosons self-interact with a quartic potential.

Considering that we expect the proposed ultralight bosons to have self-interactions, understanding the effects of these self-interactions on the instability is crucial if we hope to detect and understand signals from the superradiance scenario. This work establishes some results in the dynamics of the axisymmetric case. Through numerical simulations of a complex Proca field with quartic self-interaction around a spinning BH in axisymmetry, we find the superradiance instability suppressed when $\lambda > 0$ and amplified when $\lambda < 0$. The instability saturates with a cloud with lower energies and angular momentum when the self-interaction is positive. The Proca field with self-interaction qualitatively behaved in a similar way to the purely massive case, it followed a similar growth until the self-interaction term became large, at which point the growth slowed down and showed no evidence of

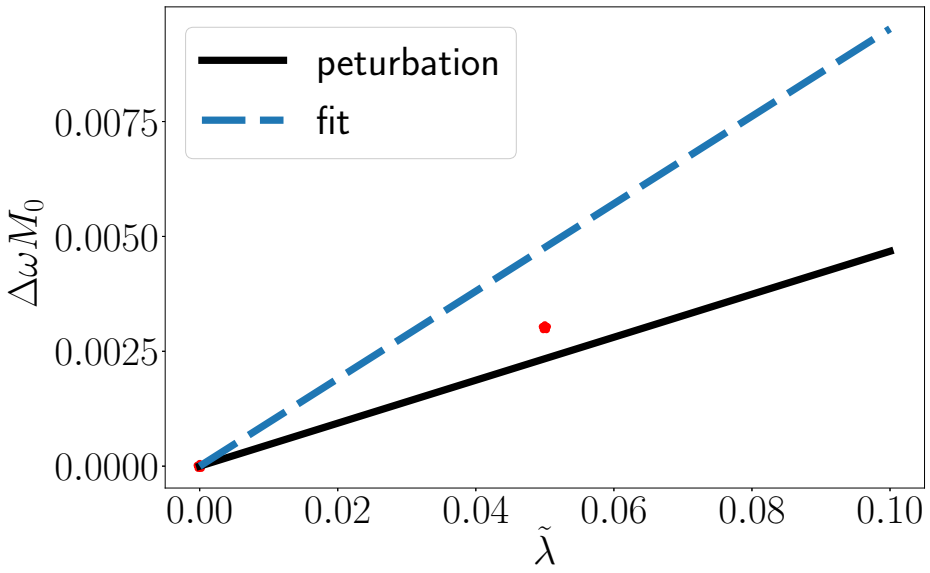

Figure 2.11: Frequency change of the cloud due to the self-interactions. We compare the numerical results (red dots) with the perturbation results (solid black line) for $\tilde{\mu} = 0.50$. We include the rescaled line of best fit from the $\tilde{\mu} = 0.40$ case as the dashed blue line.

radiation in the far field. We looked into the possibility of the bosenova collapse scenario and found no evidence of it occurring in our parameter space. We found that the cloud's profile had not been changed by the self-interaction and followed the expected falloff rate for that given mass. These results are significant as they show that having self-interacting bosons in axisymmetry does not cause the superradiance instability to be terminated by scenarios such as the bosenova collapse. It might be the case that the λ we have tested are not strong enough for the bosenova to occur. More work needs to be done with larger values of λ to properly test the bosenova scenario but considering that we tried cases where $\tilde{\lambda}$ is comparable to $\tilde{\mu}$, we have strong evidence that it will not occur. Finally, we compared the frequency shift $\Delta\omega$ at saturation from the numerical simulations and a perturbation calculation. We find that the perturbation calculation agrees for small values of $\tilde{\lambda}$ but the perturbation computation fails at larger $\tilde{\lambda}$. This is to be expected since we are not in the non-relativistic regime.

We have described the methods to study the 3D case in a fixed background but the simulations have not been completed yet. One interesting aspect that will be worth looking

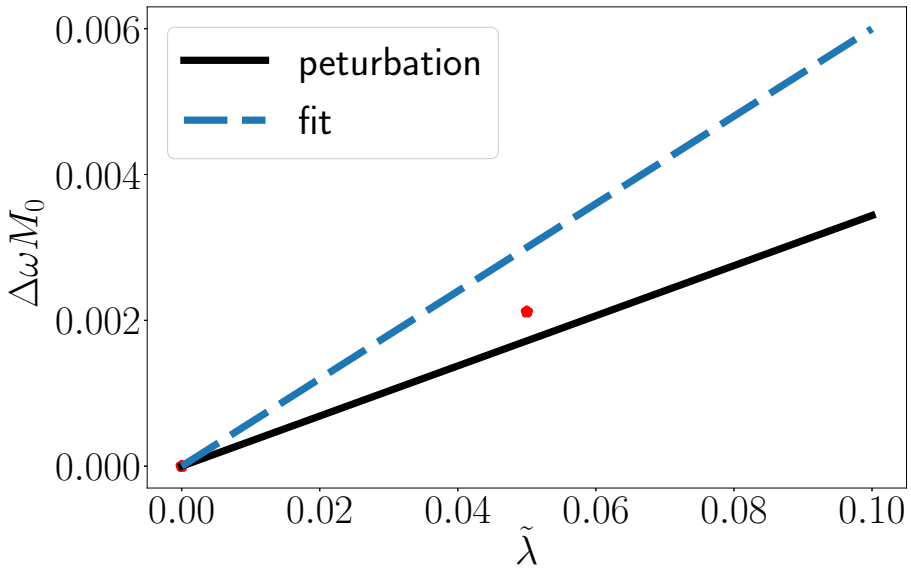

Figure 2.12: Change of the frequency of the cloud due to the self-interactions. We compare the numerical results (red dots) with the perturbation results (solid black line) for $\tilde{\mu} = 0.30$. We include the rescaled line of best fit from the $\tilde{\mu} = 0.40$ case as the dashed blue line.

into is mode mixing. While in the axisymmetric case, we restrict the Proca field to have $m = 1$, in the full 3D case, we do not; if the vector case behaves similarly to the scalar case, then we can expect this mode mixing to lead to radiation from the cloud as described in [24]. We did not observe this radiation in the axisymmetric case but will potentially see it in the 3D case. Finally, we have not observed evidence of the "bosenova' scenario in axisymmetry, but the possibility of the "bosenova" might differ with the relaxed symmetry assumptions and hence worth looking into.

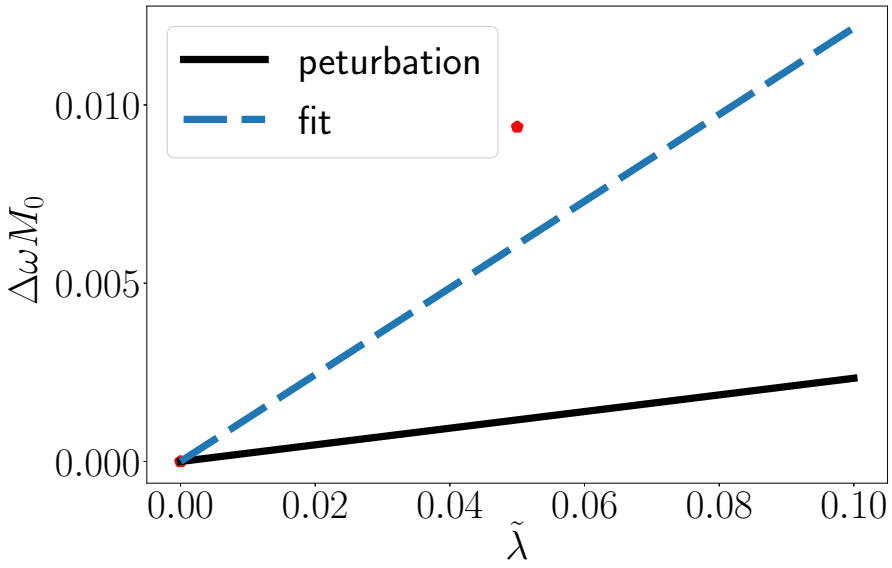

Figure 2.13: Change of the frequency of the cloud due to the self-interactions. We compare the numerical results (red dots) with the perturbation results (solid black line) for $\tilde{\mu} = 0.25$. We include the rescaled line of best fit from the $\tilde{\mu} = 0.40$ case as the dashed blue line.

Chapter 3

Semi-supervised deep learning for postmerger binary neutron star waveforms

When massive iron core stars collapse, they trigger the supernova explosion from which neutron stars are born. These neutron stars are of extreme density with masses around $\sim 1-2M_{\odot}$ and radii around $\sim 10-15$km [79] – matter densities unlike anywhere else in the universe. Neutron stars can then form binary systems which radiate gravitational energy as they inspiral until they merge. Merger events of binary neutron stars are rich in information as they involve, gravitational and electromagnetic radiation. To this date, LIGO/Virgo is believed to have detected up to 5 binary neutron star inspiral gravitational waves. In some cases, after the neutron stars have merged, the remnant does not collapse immediately into a black hole. This remnant emits gravitational waves known as the post-merger signal. This post-merger signal is rich in information about the nature of the remnant object, and the high density and out-of-equilibrium physics of the post-merger processes, which would complement that coming from the accompanying electromagnetic signals. The detection of the post-merger will offer a wealth of information but, at the same time this means that it will be difficult to detect. The detection is difficult since the frequency of the post merger waveforms are in the kHz, where the detectors are less sensitive and the detection of the signal requires a catalog of waveforms spanning the physical parameters and the theoretical uncertainties. Generating this catalog requires performing numerical simulations of these mergers which are expensive. A method to generate quick post-merger waveforms is then sought.

In this chapter, we study the possibility of leveraging the power of machine learning to

build a post-merger waveform model under theoretical uncertainty using the conditional variation autoencoder framework (CVAE) which we implement as a neural network (NN). The idea is to create a model $P(h|\Theta)$ where h is the waveform and Θ are the known parameters. Now since the theoretical models for post-merger waveforms have theoretical uncertainty, we need to marginalize out the uncertainties, that is $P(h|\Theta) = \int dz P(h|\Theta, z)P(z)$ where z is the theoretical uncertainty. It turns out this integral is intractable. As we will see, to get around this we can approximate this distribution by training a CVAE. The CVAE does so by compressing the waveforms h into latent variables z and then decoding the latent variables back into some output h'. The CVAE is trained such that the error between the input and output is minimized. To evaluate the performance of the CVAE, we measure the match between the waveforms from the data and the ones generated by the CVAE. The match is a metric which measures how well a signal projects onto another. This metric is crucial to realize since gravitational wave detections are made by scoring high matches with the detector data. Once trained, the CVAE could have the possibility of generating waveforms seemingly instantaneously.

By training the CVAE with numerical simulation data, we find that the number of elements in the numerical simulation dataset is not sufficient to train the NN to an acceptable level. The problem being that the CVAE can recreate waveforms it has seen before but fails to recreate ones that it has not seen, that is, it is overfitting. This is reasonable considering the small amount of data we have. We then turn our attention to estimate how much data would be required to train the CVAE. To do so, we build a fiducial model which is a rough model of the numerical simulation data. The fiducial model is a stochastic model which depends on the mean mass and compactness of the remnant. This fiducial model is then used to generate mock waveform data. We find that using the fiducial data, the CVAE can be trained with no overfitting occur when we have $\sim 10,000$ elements. We find that the distribution of the data is well recreated by the CVAE despite it having a poor match. As we will discuss, this poor match appears to be a result of the fiducial model being stochastic. Finally, we find evidence of the latent space learns about the features from the fiducial model which are not exposed during the training. We can see this by the fact that some latent variables are correlated with the compactness.

3.1 Introduction

With over 50 detections of gravitational waves from merging compact binaries since 2015 [3, 2, 7, 6, 8, 48, 10, 15, 14, 12], including up to five involving neutron stars [5, 13, 11, 4], a key goal of gravitational-wave astronomy is to maximize the scientific pay off of current

and future detections. Neutron star mergers are especially interesting since they have the potential to elucidate many physical phenomena from gravitational to nuclear physics.

There have been several proposals to combine observations from multiple events to increase the signal-to-noise ratio (SNR) of specific information (e.g., [76, 98, 99, 35]). For instance, by analyzing different stages of a non-vacuum binary merger, one can extract the signatures of tidal effects, the rate of relaxation to equilibrium of different multipolar perturbations, the connection of a binary's intrinsic parameters to the object formed after the merger, and potential new measurements of the Hubble constant using only gravitational waves (e.g. [77, 42, 78]). Such efforts would help reveal the equation of state (EOS) of neutron stars, probe possible deviations from General Relativity, allow for stringent tests of the *final state conjecture*[1] [85], explore the behavior of hot matter at supra-nuclear densities, and help elucidate the inner engine of gamma ray bursts [94, 90, 80, 84, 93, 51].

In this work, we focus on the post-merger stage of binary neutron star (BNS) coalescence. In particular, we seek a method to extract key properties of the hyper/massive neutron star that results if a prompt collapse to a black hole is avoided. This occurs if the total mass M of the system is not too high with respect to the maximum allowed mass $M_{\rm max}$ of a nonrotating star with the same EOS [$M_T \simeq (1.2-1.7)M_{\rm max}$], see e.g. Refs. [95, 66, 26]. The gravitational radiation from this stage, sourced by the oscillating, differentially-rotating object produced by the merger, is rich in information about the hot EOS of the system, which can be exploited, in particular, to measure the Hubble constant, guide the understanding of the central engine of intense electromagnetic outbursts, and determine the maximum mass of neutron stars.

Achieving this goal will require both knowledge of the waveform produced during this stage, as well as improved sensitivity of detectors at the high frequencies (1–4 kHz) that characterize such waveforms [78, 75]. A theoretical understanding of the expected waveform not only aids in detection, but is crucial for the physical interpretation of such signals. Naturally, the degree of completeness of such knowledge, together with the sensitivity of the detector, impacts the depth of the analysis that can be carried out. For instance, LIGO/Virgo reported no such signal following GW170817 [10], consistent with theoretical expectations that it would be have been too weak to detect. Indeed, the study [10] searched for extra power in spectrograms, but even with detailed models a sensitivity three times greater than aLIGO design would be required for the detection of the leading mode in the aftermerger gravitational-wave signal [96].

Building models for BNS postmerger signals is challenging for several reasons. First, a BNS merger is a highly nonlinear event, which can only be treated numerically. BNS

[1] Sometimes also referred to as the "generalized Israel conjecture."

merger simulations are expensive compared to those of binary black holes, since in addition to solving the Einstein equation, they must incorporate magnetohydrodynamics and microphysical modeling. Typically, this means that BNS simulations converge at lower order in discretization length compared to vacuum simulations. The BNS merger also gives rise to small-scale features, such as turbulence, which are difficult to resolve numerically, but which can significantly affect the gravitational-wave phase and ultimate fate of the remnant. Finally, simulations must be sufficiently long and accurate to smoothly match to perturbation theory in the early inspiral as well and capture the after-merger behavior. Consequently, the number of numerical simulations available at present from which to build a model is limited, and these simulations involve uncertainties not present in binary black hole simulations.

Any model must cover the full parameter space for BNS systems, which in principle includes all parameters for binary black holes, plus the EOS, which for a cold neutron star gives the pressure as a function of energy density. During the inspiral the EOS manifests through the tidal deformabilities of the individual neutron stars, which can be measured through their impact on the phase evolution [58]. Postmerger gravitational waves probe a hot, dense, high-mass regime complementary to the inspiral. Using numerical simulations the dominant postmerger frequency $f_{\rm peak}$ has been connected to the radius $R_{1.6}$ of a nonrotating star of mass 1.6 $\mathrm{M}_\odot$ [27]. However this relation is not exact, but rather depends weakly on the average mass of the neutron stars [91] and mass ratio [73]. The signal is further complicated by the presence of secondary modes, which could contain additional EOS information. Thus, in addition to the challenges of limited numbers of simulations, and uncertainties inherent in simulations, a third challenge is to parametrize the EOS in a manner that would facilitate inference.

In this work, we address these challenges using a deep-learning technique called a CVAE [72] to build a distributional latent-variable model for postmerger signals h. The basic idea is to partition the parameters characterizing the signal into two sets. The first set consists of those parameters θ that we have direct access to from simulations and that we know *a priori* should form part of the characterization of the system. For a postmerger signal, θ could include, e.g., the total mass and spin of the system. The second set of parameters—the latent variables z—includes the EOS as well as any other physics-modeling or numerical differences between simulations. It is *a priori* unclear how best to parametrize these properties of the system and simulations, so this task is left to the CVAE. During training we do not provide any information about latent parameters, rather the CVAE learns to use z to efficiently represent differences in training waveforms $h^{(i)}$ that are not accounted for in the parameters $\theta^{(i)}$.

Through training, the CVAE learns a model $p(h|z,\theta)$ for the waveform h given z and θ.

The CVAE also learns an "encoder" model $q(z|h,\theta)$ for the latent variables z, conditioned on h and θ. Using the encoder to identify latent variables with numerical simulations, we show that z encodes information about the EOS in a useful way. Given a signal, one could then perform inference jointly over θ and z using $p(h|z,\theta)$ to determine the EOS and distinguish modeling uncertainties in simulations.

Existing postmerger models have been designed mostly to facilitate the extraction of ringdown spectral information, guided by the empirical $R_{1.6}$–$f_{\rm peak}$ relation. Past works have included the use of principal component analysis [47], phenomenological waveform modeling [33], agnostic modeling such as BayesWave [49], and searches for specific principal frequencies tied to the physical parameters (e.g. [28, 63, 73, 74, 57, 29, 31].) A major advantage of the CVAE is that it learns automatically to connect features of the waveform to aspects of the EOS, rather than depending on fortuitous discoveries of empirical relations. In this way, it can be applied more generally, and, with sufficient training data, has the potential to discover new relations between the EOS and the waveform.

There have been several recent applications of machine learning to waveform modeling [44, 69, 45] (see also [50]). These approaches use neural networks to interpolate a set of training waveforms across parameter space. Since waveform generation requires forward neural-network passes, these models are fast, and they furthermore allow for rapid computation of derivatives with respect to model parameters, facilitating derivative-based inference algorithms. Our CVAE-based method generalizes these approaches to include a latent space for representing properties of the training waveforms for which a convenient parametrization may not be known in advance. We also note that CVAEs have in the past been applied to gravitational waves [59], but to the problem of parameter estimation rather than waveform modeling.

This paper is organized as follows. In Section. 3.2, we outline our basic approach and introduce the CVAE. Then in Section. 3.3, we build a simplified CVAE model $p(h|z,M)$, depending only on the total mass and the latent variables, for postmerger waveforms using numerical data from simulations. We find that the number of available simulations is at present too small to successfully train the model, so in Section. 3.4, we fit instead to artificial waveforms. By examining the latent space, we find evidence that the neutron star compactness is encoded in the latent space. We discuss the results of this study in Section. 3.5.

3.2 CVAE framework

Suppose we have a set of pairs $(\theta^{(i)}, h^{(i)})$ arising from BNS simulations, where h denotes the signal waveform, and θ denotes the binary system parameters, in particular the constituent masses (and potentially the spins, although we will not consider them here). However, the simulations have additional underlying parameters not included in θ, including the EOS, discretization errors, and additional modeling choices for the matter physics, which would confound an attempt to build a distributional model $p(h|\theta)$. To capture the dependence on these additional parameters, we suppose that they can be represented by a set of latent variables z, and we augment our model by conditioning on these as well, i.e., $p(h|\theta, z)$.

We treat the latent variables as *a priori* unknown, and we not not provide them when building the model. Rather, our aim is to learn a useful latent representation (of the EOS, etc.) based on patterns in the training waveforms. To do so, we fix a suitably restrictive parameterized form for $p(h|\theta, z)$, typically a Gaussian distribution, with mean and covariance specified as outputs of a neural network (with input $(theta, z)$). We also fix a prior $p(z)$, typically standard normal. Only by encoding information in z can the marginalized distribution, now a Gaussian mixture,

$$p(h|\theta) = \int \mathrm{d}z \, p(h|\theta, z)p(z), \tag{3.1}$$

be sufficiently general to represent the training data.

Given parameterized forms for $p(h|\theta, z)$ and $p(z)$, we would like to tune the neural-network parameters to maximize the likelihood that the training outputs $\{h^{(i)}\}_{i=1}^N$ came from the inputs $\{\theta^{(i)}\}_{i=1}^N$ under the model $p(h|\theta)$, i.e., we would like to minimize the *loss* function,

$$L_{\mathrm{MLL}} \equiv \mathbb{E}_{p_{\mathrm{data}}(\theta)} \mathbb{E}_{p_{\mathrm{data}}(h|\theta)} \left[-\log p(h|\theta) \right]. \tag{3.2}$$

However, the integral over z in Eq. (3.1) is intractable, so this cannot be evaluated.

To obtain a tractable loss, we use the variational autoencoder framework [72]. Intuitively, the integral in Eq. (3.1) could be made tractable if we knew which z contributes most for given h and θ, i.e., if we had access to $p(z|h, \theta)$. However, this density is also intractable. As an approximation, one therefore introduces an *encoder* distribution $q(z|h, \theta)$. Then

$$\begin{aligned} \log p(h|\theta) &= \mathbb{E}_{q(z|h,\theta)} \log p(h|\theta) \\ &= \mathbb{E}_{q(z|h,\theta)} \log p(h|z, \theta) - D_{\mathrm{KL}} \left(q(z|h, \theta) \| p(z) \right) \\ &\quad + D_{\mathrm{KL}} \left(q(z|h, \theta) \| p(z|h, \theta) \right). \end{aligned} \tag{3.3}$$

This expression involves the Kullback-Leibler (KL) divergence,

$$D_{\mathrm{KL}}(q(x)\|p(x)) \equiv \mathbb{E}_{q(x)} \log \frac{q(x)}{p(x)}, \tag{3.4}$$

which is nonnegative and vanishes if $q = p$. We therefore take the loss function to be

$$\begin{aligned} L \equiv \mathbb{E}_{p_{\mathrm{data}}(h,\theta)} \big[& -\mathbb{E}_{q(z|h,\theta)} \log p(h|z,\theta) \\ & + D_{\mathrm{KL}} \left(q(z|h,\theta) \| p(z) \right) \big] \qquad (3.5) \\ = L_{\mathrm{MLL}} + \mathbb{E}_{p_{\mathrm{data}}(h,\theta)} & D_{\mathrm{KL}} \left(q(z|h,\theta) \| p(z|h,\theta) \right). \qquad (3.6) \end{aligned}$$

The expression in Eq. (3.5) is now tractable. By Eq. (3.6), if $q(z|h,\theta)$ is identical to the posterior $p(z|h,\theta)$ then the CVAE loss function is equal to the maximum-log-likelihood loss.

The loss (3.5) consists of two terms. By minimizing the first term (the reconstruction loss) the CVAE attempts to reconstruct h as well as possible after first being encoded by $q(z|h,\theta)$ into a latent representation, and then decoded by $p(h|z,\theta)$ into a new waveform. In this sense, the CVAE is similar to a vanilla autoencoder (see, e.g., [60]). The second term (the KL loss) pushes $q(z|h,\theta)$ to match the prior $p(z)$. In this way, it regularizes the model, discouraging $q(z|h,\theta)$ from memorizing the training data. See Fig. 3.1 for an illustration of the overall structure.

Following standard practice, we take all distributions comprising our CVAE to be multivariate normal with diagonal covariance,

$$q(z|h,\theta) = \mathcal{N}(\vec{\mu}_q(h,\theta), \vec{\sigma}_q(h,\theta)), \tag{3.7}$$

$$p(h|z,\theta) = \mathcal{N}(\vec{\mu}_p(z,\theta), \vec{1}), \tag{3.8}$$

$$p(z) = \mathcal{N}(0,1)^l. \tag{3.9}$$

The functions $(\vec{\mu}_q(h,\theta), \vec{\sigma}_q(h,\theta))$ defining the mean and variance of the encoder are given as outputs of an encoder neural network, which takes as input (h,θ). Likewise, the decoder mean $\vec{\mu}_p(z,\theta)$ is given as the output of a decoder network, which takes as input (z,θ). The covariance of the decoder is fixed to the identity, and, following standard practice, the prior over z is taken to be a standard multivariate normal distribution.

When written in the form (3.5), the loss L may be evaluated, and the neural networks trained. Since $q(z|h,\theta)$ and $p(z)$ are both multivariate normal, the KL divergence can be evaluated analytically for any (h,θ). The reconstruction loss is evaluated using a single-sample Monte Carlo approximation as follows:

1. Use the encoder network to evaluate $(\vec{\mu}_q(h,\theta), \vec{\sigma}_q(h,\theta))$.

2. Sample $z \sim q(z|h,\theta)$ by first sampling $\vec{\epsilon} \sim \mathcal{N}(0,1)^l$ and then setting $z = \vec{\mu}_q(h,\theta) + \vec{\epsilon} \odot \vec{\sigma}_q(h,\theta)$. This is known as the *reparameterization trick* [72].

3. Use the decoder network to evaluate $\vec{\mu}_p(z,\theta)$.

4. Evaluate the loss $-\log p(h|z,\theta)$.

The loss L and its derivatives with respect to neural network parameters may then be evaluated on minibatches and minimized using a standard gradient-based stochastic optimizer (we use Adam [70]). The reparameterization trick was used to separate out the stochastic aspect of sampling from $q(z|h,\theta)$ and enable backpropagation for calculating the derivatives.

We now comment on the relation of this procedure to standard fitting techniques for model building. If $p(h|z,\theta)$ were independent of z, then the reconstruction loss would no longer depend on $q(z|h,\theta)$. For the multivariate normal $p(h|\theta)$ given in Eq. (3.8) (now omitting z) this reduces (up to normalization) to the mean squared difference $\frac{1}{N}\sum_{i=1}^{N} \|\vec{h}^{(i)} - \vec{\mu}_p(\theta^{(i)})\|^2$, i.e., $\vec{\mu}_p(\theta)$ becomes a model for the waveform given by a standard least-squares fit to training data. (One could additionally modify the covariance of $p(h|\theta)$, to specialize the fit, e.g., by including information about detector noise properties, or to allow it to be fit as well during training.)

By allowing for dependence on the latent variables z, the marginalized distribution $p(h|\theta)$ can have a more complicated (non-Gaussian) structure. This is needed to build a model from simulations that depend on parameters not included in θ, as these could give rise to different h for the same θ. In the following sections, using a toy model for binary neutron star postmerger waveforms, we provide evidence that the latent variables can learn about the hidden variables of the training data, in this case the EOS. Given a gravitational-wave detection, by using $p(h|z,\theta)$ to perform Bayesian parameter estimation jointly over θ and z, one could, therefore, learn useful information about which of these parameters are preferred; see Fig. 3.2.

The fact that one optimizes L rather than $L_{\rm MLL}$ when training a CVAE gives rise to a possible pitfall. Notice that if the last term in Eq. (3.6) vanishes (i.e., if $q(z|h,\theta)$ and $p(z|h,\theta)$ are identical) then the two losses coincide. Typically, though, the form (3.7) of the encoder $q(z|h,\theta)$ is too restrictive to properly represent the posterior $p(z|h,\theta)$, so $L > L_{\rm MLL}$. Another way equality can be achieved, however, is by ignoring the latent space entirely, i.e., by setting $p(h|z,\theta) = p(h|\theta)$ and $q(z|h,\theta) = p(z|h,\theta) = p(z)$. This is known as posterior collapse. In other words, optimizing L rather than $L_{\rm MLL}$ means that the use of the latent space incurs a cost. If this cost exceeds the gain in reconstruction performance achieved

by using the latent variables, then the CVAE can fail to autoencode [43]. To mitigate this problem, one approach is to multiply the KL loss term by a factor $\beta < 1$ [64], and possibly anneal this term to one during training [34]. In this work, we determine experimentally a fixed value for β that gives good performance. Other approaches include generalizing the form of the encoder distribution so that it more easily approximates the posterior [71]. In Appendix D, we demonstrate these methods on a toy model where the data is a simple sum of sine waves.

3.3 Postmerger waveform model

In this section we apply the CVAE framework to construct a model for postmerger waveforms based on numerical relativity simulations. We build a simplified model $p(h|M, z)$, taking the parameter space to be the total mass of the system, i.e., $\theta = M$, and representing waveforms in terms of several phenomenological fit parameters. We find that the small number of available numerical relativity simulations leads to overfitting, so in the following section we assess the CVAE approach further using synthetic training data.

3.3.1 Training data

We obtain numerical training waveforms from the CoRe database [53]. This comprises 367 BNS waveforms, with varying component masses, spins, and equations of state, as well as different numerical resolutions and starting frequencies. (There are 163 unique choices of masses, spins, and equations of state.) From these inspiral-merger-ringdown waveforms, we extract the postmerger signal by identifying the moment in time of peak gravitational-wave luminosity, and truncating the preceding signal. Note that although the neutron stars come into contact prior to this peak, this definition of the postmerger signal is sufficient for our purposes. We drop all waveforms that exhibit prompt collapse to a black hole, requiring that the post-peak waveform lasts at least 5 ms.

The CoRe simulations have been performed at different resolutions and for varying time durations. To prepare the waveform data for the neural network, we standardize the time resolution and total time by resampling using a quadratic spline and padding with zeros at the end. We represent waveforms using 1000 time samples, with a total duration of 45 ms, which is the duration of the longest postmerger signal. Finally, for simplicity, we only take the real part of the $\ell = m = 2$ component of the complex strain $h = h_+ - ih_\times$, after multiplying by a phase such that the real part vanishes at the merger time, $t = 0$, i.e., $h(t) \equiv \mathrm{Re}(e^{i\theta} h_{22}(t))$, with $h(t = 0) = 0$.

When fitting the model directly to these strain data sets, we find that the zero padding leads to poor waveform reconstruction with overdamping at late times (see App. E). Instead, we find it more effective to use a compressed representation, where waveforms are expressed in terms of parameters of a phenomenological model [33],

$$\begin{aligned} h(t) = \alpha_1 e^{-t/\tau_1} \{ & \sin(2\pi f_1 t) + \sin[2\pi(f_1 - f_{1\epsilon})t] \\ & + \sin[2\pi(f_1 + f_{1\epsilon})t]\} \\ + \alpha_2 e^{-t/\tau_2} & \sin\left(2\pi f_2 t + 2\pi\gamma_2 t^2 \right. \\ & \left. + 2\pi\xi_2 t^3 + \pi\beta_2\right). \end{aligned} \tag{3.10}$$

Here, $f_{1\epsilon} = 50$ Hz, and the parameters $\{\alpha_1, \alpha_2, \tau_1, \tau_2, f_1, f_2, \beta_2, \gamma_2, \xi_2\}$ are determined by least-squares fit. This model is motivated by the fact that BNS postmerger signals tend to consist of two damped sinusoids, with corrections added for improved accuracy [33].

The parameterized form (3.10) is capable of representing the numerical-relativity postmerger signals with $\mathcal{M} > 0.8$, where

$$\mathcal{M}(h, \tilde{h}) = \frac{\langle h|\tilde{h}\rangle}{\sqrt{\langle h|h\rangle\langle\tilde{h}|\tilde{h}\rangle}}, \quad \langle h|\tilde{h}\rangle = \int_0^{t_f} h(t)\tilde{h}^*(t)dt, \tag{3.11}$$

is the match between two waveforms, assuming a flat noise spectrum. We additionally restrict to those waveforms with $\mathcal{M} > 0.9$. This reduces our data set to 123 waveforms. Finally, we use principle component analysis whitening to decorrelate and standardize the waveform parameters (see App. F).

To summarize, our training data consist of 123 (M, h) pairs, where h is a 9-component vector describing a postmerger signal. This discards labels for the mass ratio, spins, EOS, and numerical resolutions, but keeps that of the total mass. This omitted information, however, gives rise to differences between the waveforms, so it (along with any other differences between simulations) should be understood as latent variables. In the following subsection we construct a CVAE to model these waveforms and characterize the latent information.

3.3.2 Experiments

We model $p(h|M)$ with a CVAE with latent dimension $l = 4$. The encoder and decoder are fully-connected networks with rectified linear unit nonlinearities. Exploring various

hyperparameter choices, we find best performance with a 2-hidden-layer encoder with $(100, 50)$ units, a decoder with inverted structure, and $\beta = 0.003$.

We split our dataset into training (60%) and validation (40%) sets, and train for 4000 epochs. The training history is plotted in Fig. 3.3. Since the validation loss is much higher than the training loss, we conclude that the CVAE is overfitting, and that the training set is too small for the network to successfully generalize to previously unseen waveforms. We have also tried methods for mitigating overfitting such as adding dropout layers. There was no significant improvements using that method.

3.4 Extended training set

The number of available numerical simulations is at present too small to properly train a CVAE to model postmerger waveforms. Nevertheless, we can assess the viability of the approach for the future by training on synthetic data. In this section, we first construct a "fiducial" waveform model $p_{\rm fid}(h|\bar{M}, \mathcal{C})$ based on the CoRe waveforms. The fiducial model depends on the average mass $\bar{M} = (M_1 + M_2)/2$ and the compactness $\mathcal{C} = GM/(Rc^2)$ of the remnant neutron star. We then generate a much larger training set by sampling from this model, and we show that with this we can successfully train the CVAE. This exercise also provides an estimate for the number of numerical waveforms needed to train a CVAE.

3.4.1 Fiducial model

We now construct the fiducial model $p_{\rm fid}(h|\bar{M}, \mathcal{C})$ for postmerger waveforms h given $\bar{M}$ and $\mathcal{C}$. We do not have direct access to $\mathcal{C}$ from the numerical-relativity simulations, so instead we extract it based on the dominant postmerger frequency [73, 33, 91]. Indeed, Ref. [33] found that, for a remnant with $\bar{M} = 1.6$ M$_\odot$, the dominant postmerger frequencies approximately satisfy the relations,

$$f_1 = \left(a_5 + a_6\mathcal{C} + a_7\mathcal{C}^2 + a_8\mathcal{C}^3\right)(1.6\ M_\odot/\bar{M}), \tag{3.12}$$

$$f_2 = \left(a_9 + a_{10}\mathcal{C} + a_{11}\mathcal{C}^2\right)(1.6\ M_\odot/\bar{M}), \tag{3.13}$$

with the numerical parameters $a_5 = -35.17$ kHz, $a_6 = 727.99$ kHz, $a_7 = -4858.54$ kHz, $a_8 = 10989.88$ kHz, $a_9 = -3.12$ kHz, $a_{10} = 51.90$ kHz, and $a_{11} = -89.07$ kHz. On dimensionful grounds, we have added an extra factor $1/\bar{M}$ to extend these relations to other BNS masses. We therefore take Eq. (3.13) as given and use it to *define* $\mathcal{C}$ in terms of

f_2 and $\bar{M}$. After solving for $\mathcal{C}$, we plot all the parameters from fitting Eq. (3.10) to the numerical waveforms versus $\mathcal{C}$ in Fig. 3.4. The parameters of f_1 are then fitted to our data. We find $a_5 = 10.65$ kHz, $a_6 = -262.69$ kHz, $a_7 = 2329.10$ kHz, and $a_8 = -6254.30$ kHz. These values are different from the ones found in Ref. [33], but as can be seen in Fig. 3.4, the curve defined by [33] is within one standard deviation of the parameters we find.

We fit a generalized linear model for the remaining waveform parameters in terms of the compactness, i.e.,

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \log(\tau_1) \\ \log(\tau_2) \\ \frac{\bar{M}}{1.6M_\odot} f_1 \\ \frac{\bar{M}}{1.6M_\odot} f_2 \\ \beta_2 \\ \frac{\bar{M}^2}{1.6M_\odot} \gamma_2 \\ \frac{\bar{M}^3}{1.6M_\odot} \xi_2 \end{pmatrix} \sim \mathcal{N}(\vec{B} + \vec{A}\mathcal{C}, \Sigma), \tag{3.14}$$

where

$$\vec{A} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ (a_5/\mathcal{C} + a_6 + a_7\mathcal{C} + a_8\mathcal{C}^2) \\ (a_9/\mathcal{C} + a_{10} + a_{11}\mathcal{C}) \\ a_{12} \\ a_{13} \\ a_{14} \end{pmatrix}, \tag{3.15}$$

$$\vec{B} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \\ b_9 \end{pmatrix}, \tag{3.16}$$

and where Σ is a generic covariance matrix, and a_i and b_i are constants which are obtained by fitting the numerical waveforms except for $\{a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}\}$, where we use the values discussed previously. We determine Σ by first subtracting the mean of $\{\alpha_1, \alpha_2, \log(\tau_1), \log(\tau_2), \beta_2, \gamma_2, \xi_2\}$, and subtracting Eq. (3.12) and Eq. (3.13) for the frequencies. This method of computing Σ is conservative in the sense that it overestimates the noise in the model. The covariance matrix can also be estimated by first removing $\vec{A}\cdot\mathcal{C} + \vec{B}$ from the data, or by simultaneously fitting for the linear dependence and covariance. However, we have checked that doing the former only reduces the magnitude of the covariance $\sim 2\%$, and does not significantly affect the results of the estimate.

Note that we took logarithms of τ_1 and τ_2 so as to normalize their distributions. The values of the parameters are recorded in appendix G. We add a factor of $1/\bar{M}$ in the frequencies to roughly model mass dependence of the frequency on the mass, inspired by Ref. [73]. The estimated variance on γ_2 and ξ_2 is large which leads to issues with the waveforms at late times. On dimensional grounds, we also add a factor of $1/\bar{M}^2$ to γ_2 and $1/\bar{M}^3$ to ξ_2.

3.4.2 Experiments

With the fiducial model, we are now able to generate artificial postmerger signals, which we take as training data for the CVAE. We first generate a dataset with 1,010,000 samples $h \sim p_{\rm fid}(h|\bar{M}, \mathcal{C})$ by randomly choosing the remnant mass, $1.2\ M_\odot < \bar{M} < 1.7\ M_\odot$ and $\mathcal{C} \in (0.1, 0.2)$ from uniform distributions. We pick the compactness in this range in accordance to the range given in [33]. The dataset is then whitened and normalized. We then train CVAEs (with $\ell = 4$) using 500, 1000, 5000, 8000, 9000, 10^4, 1.5×10^4, 10^5, and 10^6 elements. We set aside 10^4 elements for the validation set.

While the CVAE showed no signs of overfitting with a dataset with 10,000 elements or more, it failed to get a good match, with $\mathcal{M} = 0.27$ for the training set and $\mathcal{M} = 0.24$ for the validation set. Given the stochastic nature of the fiducial model, using the match as the metric to evaluate the performance of the CVAE may not be the best choice of metric. By generating a set of waveforms using the fiducial model with fixed $\bar{M}$ and $\mathcal{C}$ then computing the match between these waveforms, we get an average match of 0.38. We should not expect the CVAE to produce a match greater than this since, if the CVAE is well trained, then the CVAE model could recreate any of the waveforms from the fiducial model giving an average match close to 0.38. What is more important then is that the distribution of parameters reproduced by the CVAE is similar to the ones from the fiducial model. By looking at Fig. 3.5, we see that the recreated parameters from the CVAE (orange dots) are

similar to the input parameters (blue dots); this figure shows that the CVAE can recreate the parameters well. The amplitude part of the waveforms are well recreated by the CVAE, as shown in Fig. 3.6, where we have the amplitude part of the match is computed and plotted versus different training set sizes. Some more work needs to be done to find the correct metric. Details of the CVAE architecture are provided in Appendix G. We train the CVAE over 2000 epochs with a batch size of 150. The training of the largest dataset took approximately three days on a laptop.

We can now address the question of how the CVAE encodes information about the compactness in the latent space. In Fig. 3.7 we plot the latent distribution of encoded waveforms $h \sim p(\bar{M}, \mathcal{C})$ for fixed $\bar{M} = 1.35$ $M_\odot$, but sampling $\mathcal{C} \sim p(\mathcal{C})$. We see that the latent variables clearly encode information, as they trace a path in the latent space rather than sampling the normal distribution. However, there is no obvious connection to the compactness $\mathcal{C}$.

To probe whether correlations between z and $\mathcal{C}$ can be obtained under more idealized conditions, we train the CVAE once again, but this time based on a fiducial model with a reduced noise level. Fig. 3.8 shows the latent space of the CVAE trained with the noise reduced by half, i.e., $\Sigma \to 0.5\Sigma$. We see that there is a clear correlation between z_0 and $\mathcal{C}$. The CVAE has therefore learned about the compactness based on waveforms alone, without ever having been provided compactness information. This is evidence of the NN learning the simple $\mathcal{C} - f_2$ relationship. While this is a simple relationship to uncover, it shows how the CVAE can expose interesting relationships in the latent space.

3.5 Discussion

The detection of a gravitational wave signal from the post-merger oscillations of a BNS would be a major accomplishment with important consequences for our understanding of physics in extreme regimes. Given the weaker amplitude and high frequency of such a signal, a good model will be essential in making and learning from such a detection. However, our ability to build such a model is hampered by a number of theoretical uncertainties, including our lack of knowledge of the true neutron star EOS, uncertainties regarding the behavior of magnetized matter, viscous, cooling, and other microphysical effects, and difficulties adequately resolving turbulence and other small scale effects, while still covering the allowed parameter space. Machine learning offers a way to build partially informed models with relatively few assumptions about the unknown physics. To build the CVAE, we assume that the unknown physics can be encoded in a small latent space, and that there is a way to interpolate between different theoretical models of the unknown physics. By

training a neural network with samples covering not only different unknown physics (e.g. different EOSs), but also different theoretical errors (e.g. numerical resolutions used in the simulations), we can attempt to build a model that interpolates over both, and thus is sufficiently robust to be used to look for real signals. The neural network approach also easily adapts to incorporate new information as it is obtained, and does so with a low associated computational cost at training time. In this approach, the main cost lies in the amount of data that is required to train the model. In this case, training the model with 10^6 elements took approximately three days on a laptop. This training time can be reduced significantly by taking advantage of GPUs.

By building CVAEs for toy model data, we first showed that the CVAE is capable of learning simple waveform time series consisting of a sum of sine functions. We then moved on to training the CVAE on real numerical waveforms using the CoRe database [53], and found that the neural network is overfitting, which is unsurprising given the dearth of data to train on, and the complexity of the waveforms. Instead of training directly on the time series from the numerical waveforms, we fit a nine parameter function [Eq. (3.10)] to waveforms, and used these parameters as the training data. We combined this with some preprocessing to improve the learning outcome, but still found that we were overfitting. Given that we could not avoid overfitting with the data we had, we turned our attention to estimating how much data would be required for the CVAE to work. To perform the estimate, we built a fiducial model of the parameters to generate data to train on. We generated data using the fiducial model, trained a CVAE with four latent variables, and found that we could create a CVAE which would recreate a similar parameter space to the fiducial model without overfitting with $\sim 10,000$ training elements. Unfortunately, the CVAE did not manage to recreate the phases accurately enough to get sufficient matches. Nonetheless, the CVAE model does show evidence of learning how to recreate the waveforms by being able to reproduce a similar parameter distribution to the fiducial model and encoding the compactness in the latent space. A better metric to evaluate the model needs to be found. The particular choice of fitting function we used was chosen mainly for illustration, and other choices, for example using a greater number of free parameters, should also work with these methods, though would presumably increase the required size of the training set.

In investigating the latent space generated by the training process, we found evidence that the latent space was encoding the hidden variables, but we were unable to obtain a clear correspondence between the main hidden variable (the neutron star compactness) and some combination of the latent variables. A readily apparent relationship does show up once the noise of the fiducial model is reduced. This suggests that in the case without reduced noise, the latent space still encodes the compactness, but in a more complicated way. This relation could perhaps be found by training a secondary neural network for this

purpose. We leave further study of the latent space to future work.

In this work, we considered two ways to present the data to the neural network: the full time series, and the parameters of a model which was fit to the time series. However, these are far from the only choices. Another method which we attempted, but found poor results, is to decompose the spectrum of the waveforms with principle component analysis [47], and then train on the coefficients. We found that we needed more than 10 principle components to get a match above 0.90. The combination of large variance and needing many coefficients made the learning of the data more difficult.

It is also important to keep in mind that the estimate given in Sec. 3.4.2 depends heavily on the specific neural network architecture (number of layers, number of neurons per layer, etc.) that we train on. The space of neural network implementations of a CVAE is large, and the one we picked might not be the optimal one despite our search. One could envision trying to adapt tools such as autoML [68], which searches over the space of neural networks, to build a CVAE optimal for the post-merger data.

From our studies, we conclude that more training data is required to train the CVAE to properly represent BNS post-merger waveforms. We found that with $\sim 10^4$ waveforms, the CVAE could recreate similar parameters to the fiducial model but could not get a good fit. Given that the CVAE recreates the parameter space of the fiducial model well and the stochastic nature of the fiducial model, we suspect that the choice of metric to evaluate the CVAE should not be match. More work needs to be done to determine the correct metric.

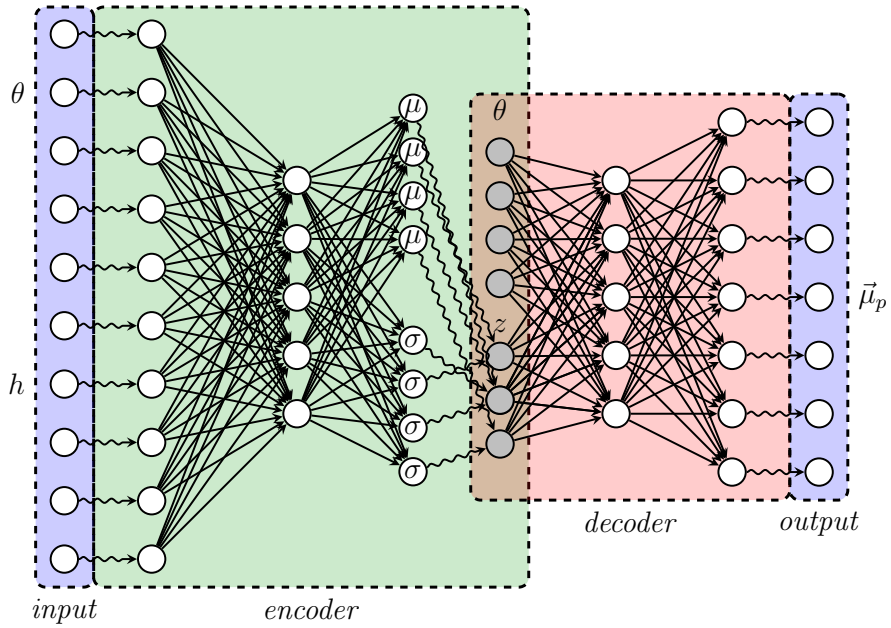

Figure 3.1: Diagram of a CVAE. The white circles of the image represent the neurons of the neural network, which are nonlinear functions, while the connection between the neurons are represented by the black arrows. The CVAE in the diagram is broken up into four pieces: the input, the encoder, the decoder, and the output. The information flows through the CVAE from left to right and is compressed in the latent space. At the intersection of the encoder and decoder, we have the latent space, which consists of the latent variables z and the conditional variables θ where $\dim(z) < \dim(h)$. This latent layer is connected to the previous layer with a different function than the other layers, represented by the squiggly arrows. The last layer of the encoder outputs $\vec{\mu}_q, \vec{\sigma}_q$ which are used to define a normal distribution $\mathcal{N}(\vec{\mu}_q, \vec{\sigma}_q)$. The latent layer then samples from this distribution, $z \sim q(z|h,\theta) = \mathcal{N}(\vec{\mu}_q(h,\theta), \vec{\sigma}_q(h,\theta))$. The last layer of the of the encoder consists the $\vec{\mu}_q(h,\theta)$ and $\vec{\sigma}_q(h,\theta)$ variables which are, respectively, used as the mean and standard deviation of the normal distribution sampled from the latent variables. These variables facilitate interpolation by forcing the latent variables to be normally distributed.

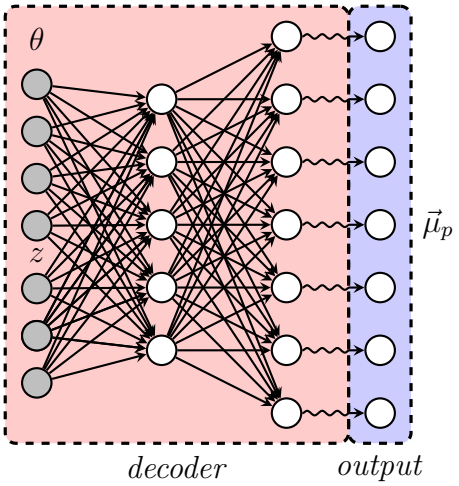

Figure 3.2: After the CVAE is trained, one can disregard the input and the encoder and only consider the layers after the latent space. Doing so results in a model $\vec{\mu}_p = \text{Dec}(\theta, z)$ where Dec is the decoder network.

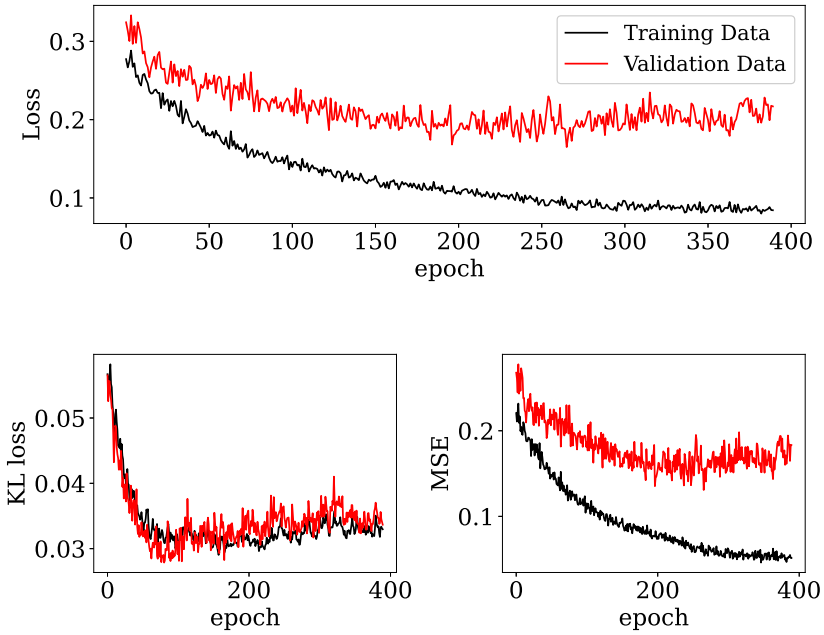

Figure 3.3: Training history for a CVAE trained with numerical waveforms in the representation (3.10). The upper plot shows the total loss (3.5), while the bottom two plots show the KL (left) and reconstruction loss (right) contributions to the total. The red and black curves correspond to the training and validation sets, respectively. The gap between training and validation sets is an indication of overfitting.

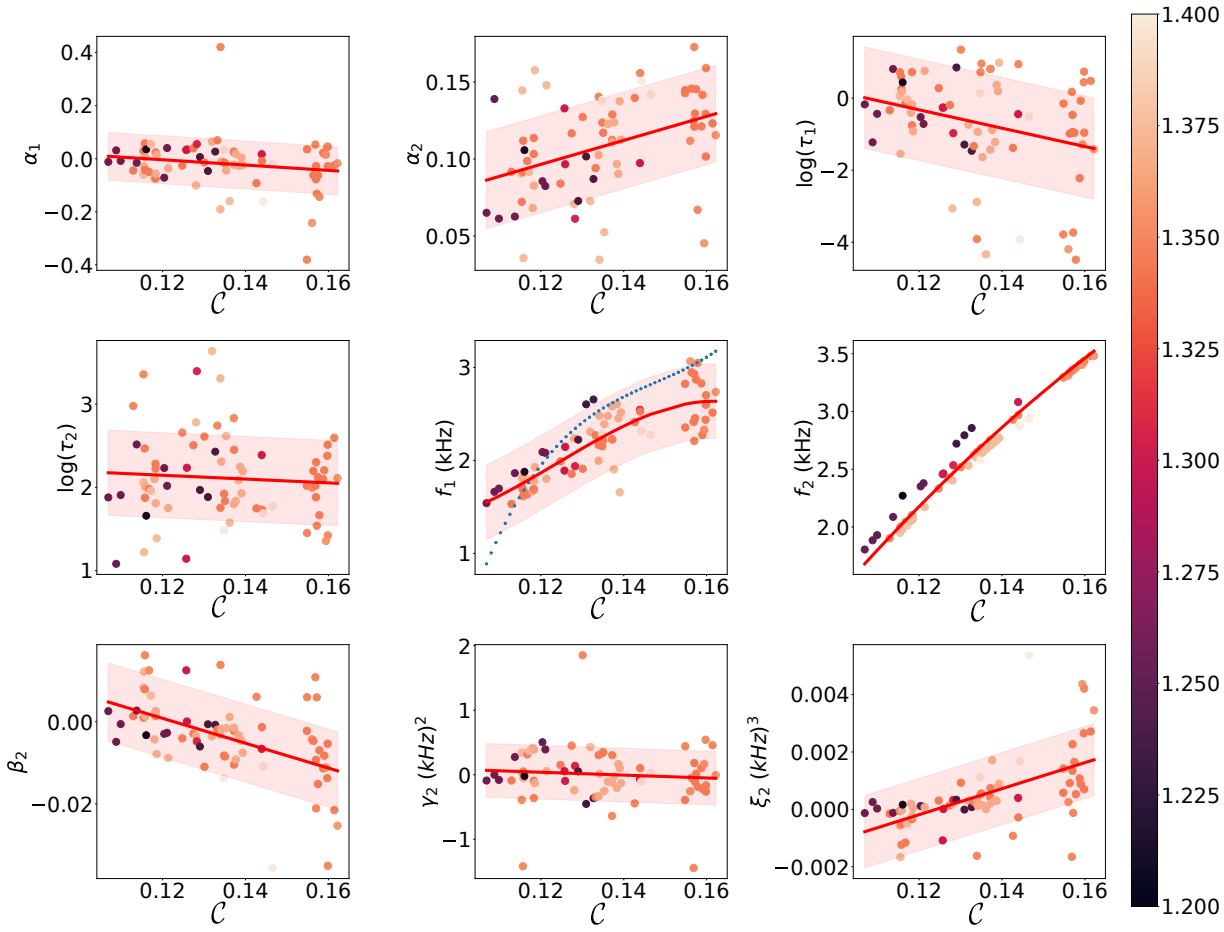

Figure 3.4: The parameters in Eq. (3.10), obtained from fitting the numerical waveforms, versus compactness $\mathcal{C}$, color coded by the average mass ($\bar{M}$) of the two stars in units of solar mass. Since we do not have direct access to $\mathcal{C}$ from the simulations, here we define it be given by $f_2 = (b_0 + b_1\mathcal{C} + b_2\mathcal{C}^2)(1.6\ M_\odot/\bar{M})$ with $b_0 = -3.12$ kHz, $b_1 = 51.90$ kHz, and $b_2 = -89.07$ kHz. We then fit for the parameters of $f_1 = (a_0 + a_1\mathcal{C} + a_2\mathcal{C}^2 + a_3\mathcal{C}^3)(1.6\ M_\odot/\bar{M})$ to our data. The blue curve in the f_1 plot represents the fit from Ref. [33]. Though this curve roughly lies within the range of the data points, we find different values for the best fit parameters for the mean of f_1. The remaining parameters, on the other hand, appear to be roughly linear in $\mathcal{C}$, with some noise. We overlay the fiducial model at a fixed $\bar{M} = 1.35\ M_\odot$ on the numerical data: the solid red line is the mean of the data, i.e. $\vec{A}\cdot\mathcal{C} + \vec{B}$, while the shaded regions are the the one standard deviation ranges. From this, we see that the fiducial model captures most of the numerical data.

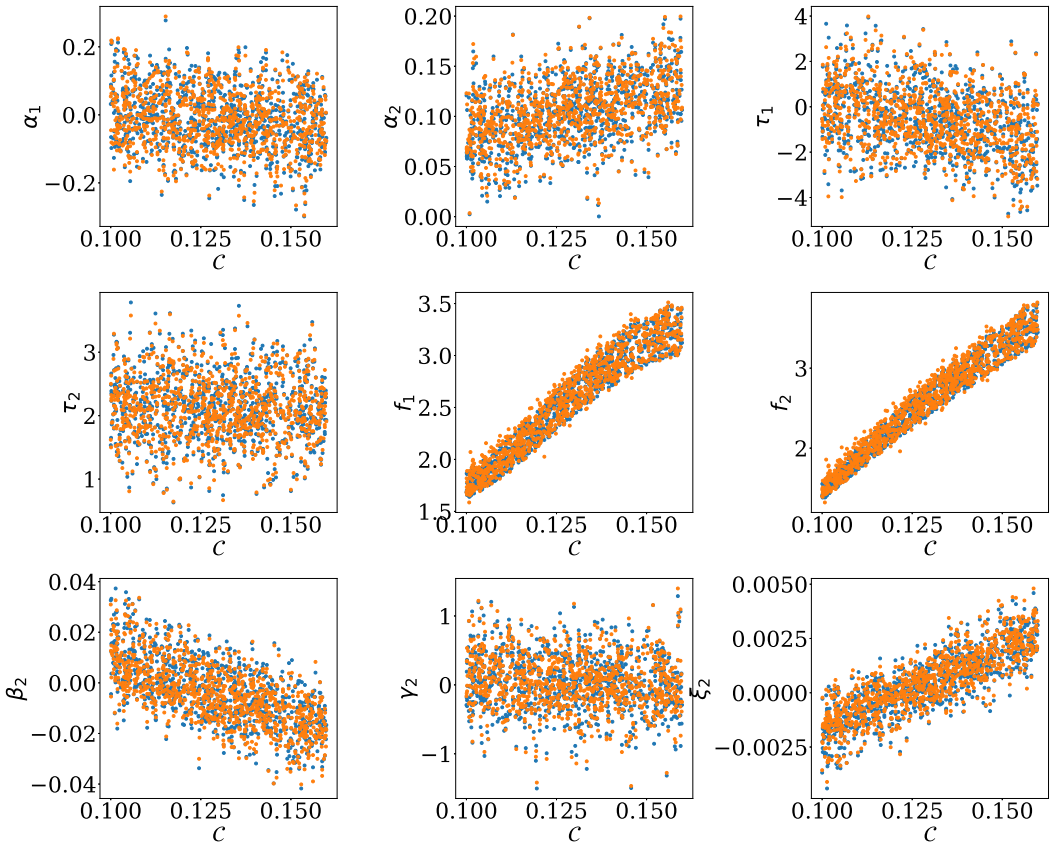

Figure 3.5: Parameters from the fiducial model (blue dots) with recreated parameters from the CVAE (orange dots.) The recreated data is generated by inputting the fiducial model data into the CVAE. The CVAE was trained with 10,000 elements. Here we fix the value of $\beta = 0.0008$, which is different from the β value used in the numerical data experiment.

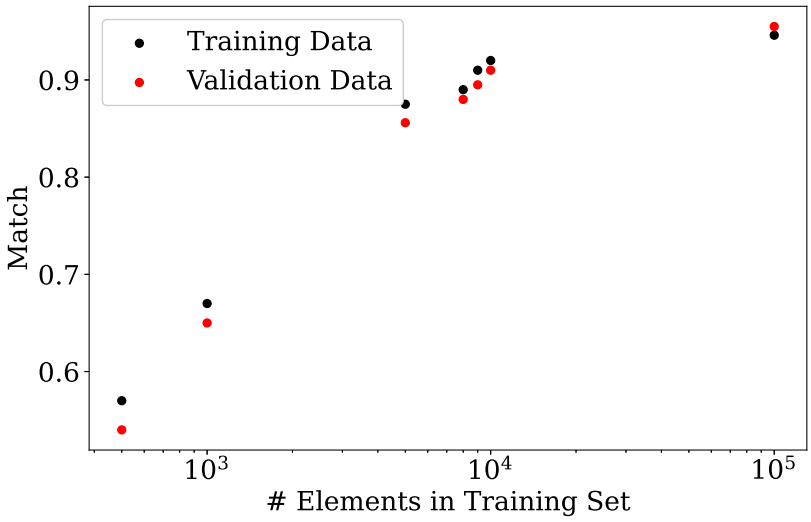

Figure 3.6: Amplitude match of training and testing data with varying sizes of dataset for the fiducial model data. The neural network was trained and tested over five different datasets for each dataset size. Here we again fix the value of $\beta = 0.0008$.

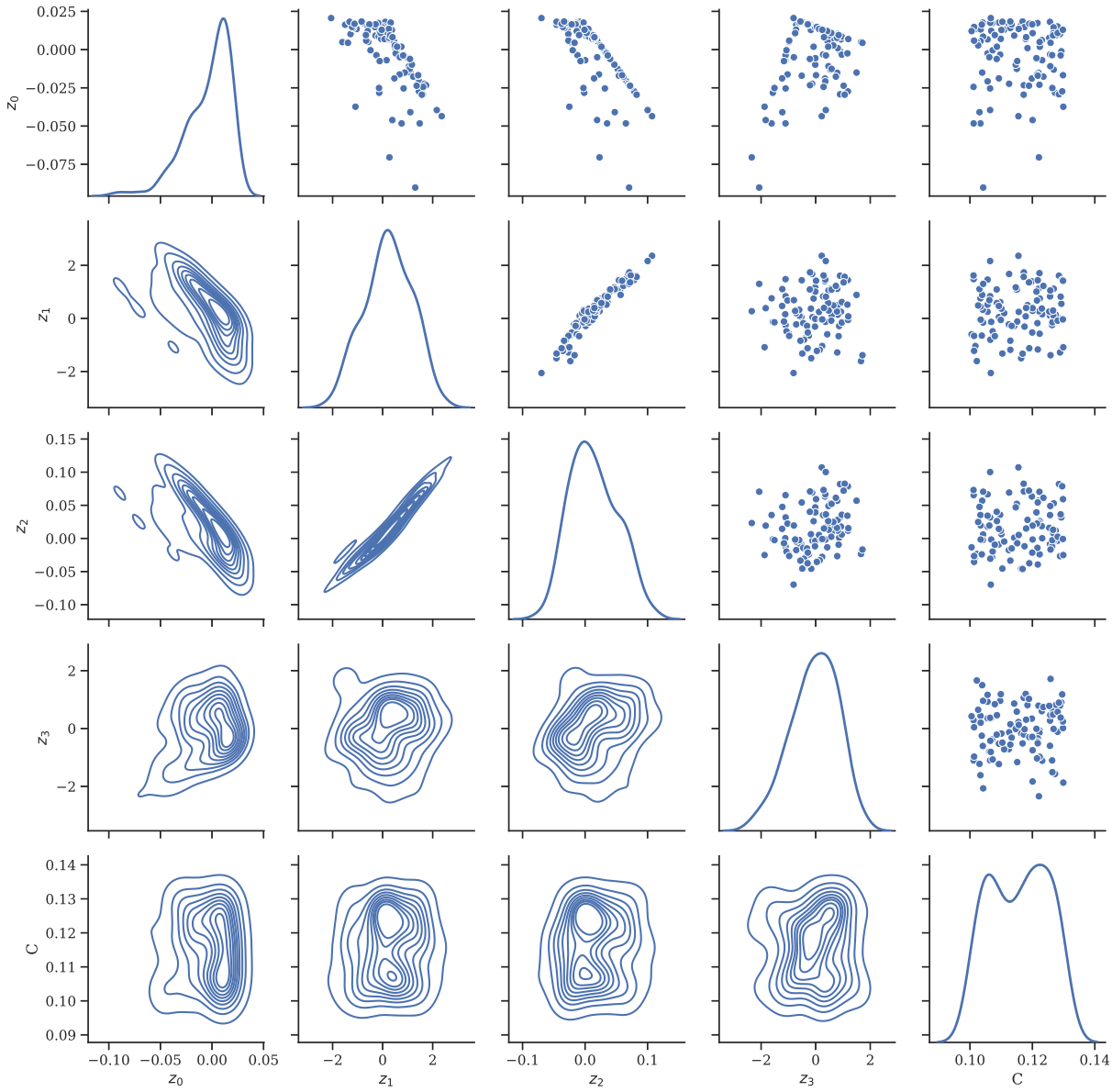

Figure 3.7: Pairplot of latent variables generated with data generated with the fiducial model with fixed masses and random compactness. That is we sample $z \sim p(z|\bar{M} = 1.35M_\odot, h(\bar{M} = 1.35M_\odot, \mathcal{C}))p(\mathcal{C})$. We can see that the data traces out a one dimensional path with noise in the latent space in some of the latent variables. This can be seen clearly with the variables (z_1, z_2). This indicates the CVAE is avoiding posterior collapse. By looking at the compactness column one would hope to see a relationship between the compactness and the latent variables, unfortunately the relationship is not obvious here.

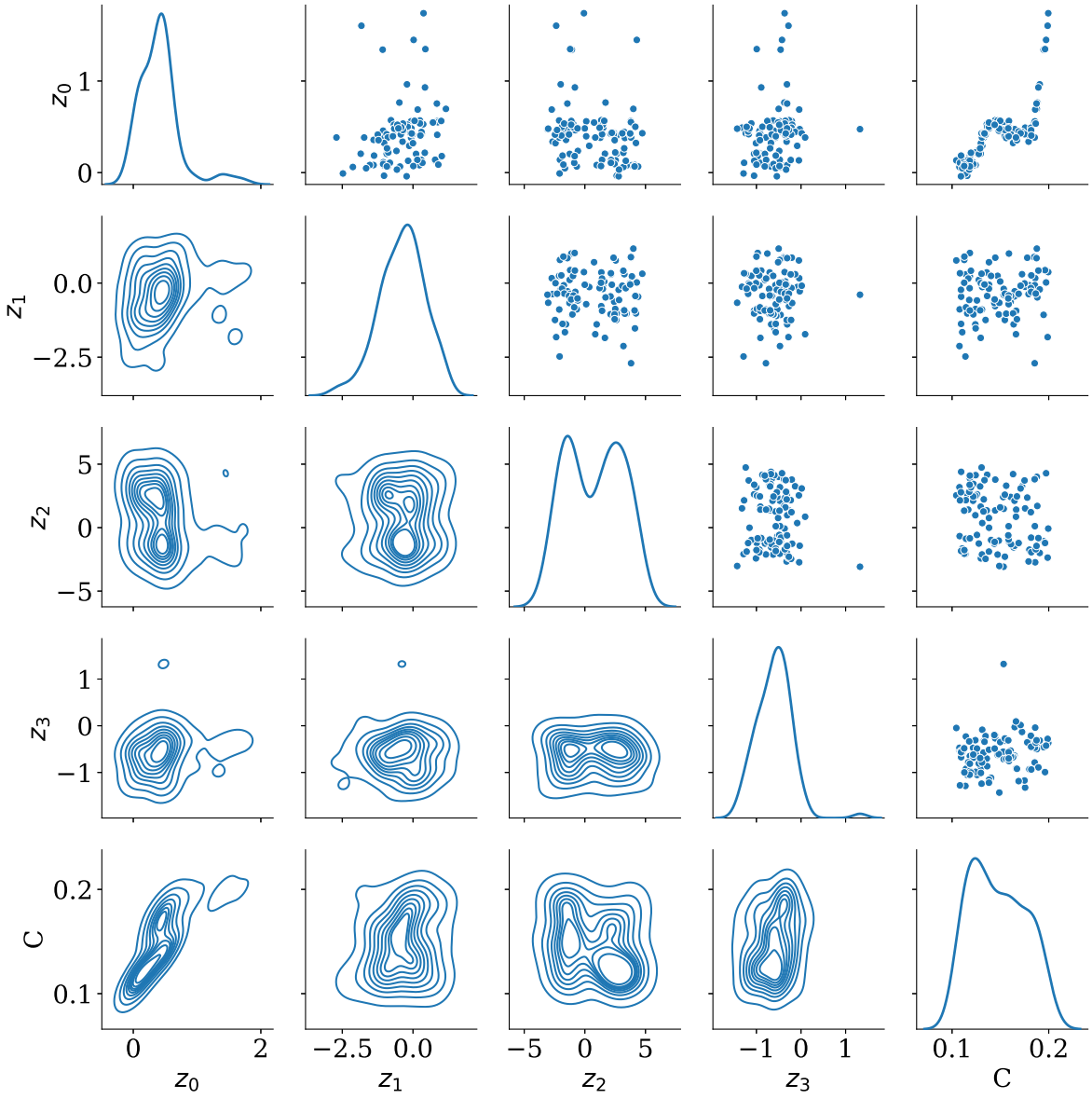

Figure 3.8: Similar to Fig. 3.7, but for a fiducial model with half the variance ($\Sigma \to 0.5\Sigma$), we show a pairplot of latent variables generated with fixed masses and random compactness. That is, we sample $z \sim p(z|\bar{M} = 1.35M_\odot, h(\bar{M} = 1.35M_\odot, \mathcal{C}))p(\mathcal{C})$. In this case we can see that there is a relationship between the first latent variable z_0 and the compactness.

References

[1] J Aasi, B P Abbott, R Abbott, T Abbott, M R Abernathy, K Ackley, C Adams, T Adams, P Addesso, and et al. Advanced ligo. *Classical and Quantum Gravity*, 32(7):074001, Mar 2015.

[2] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, and et al. Gw151226: Observation of gravitational waves from a 22-solar-mass binary black hole coalescence. *Phys. Rev. Lett.*, 116:241103, Jun 2016.

[3] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, and et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016.

[4] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, and et al. Gw190425: Observation of a compact binary coalescence with total mass 3.4 m. *The Astrophysical Journal*, 892(1):L3, Mar 2020.

[5] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, Affeldt, and et al. Gw170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017.

[6] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Afrough, and et al. Gw170814: A three-detector observation of gravitational waves from a binary black hole coalescence. *Phys. Rev. Lett.*, 119:141101, Oct 2017.

[7] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, and et al. Gw170104: Observation of a 50-solar-mass binary black hole coalescence at redshift 0.2. *Phys. Rev. Lett.*, 118:221101, Jun 2017.

[8] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, and et al. Gw170608: Observation of a 19 solar-mass binary black hole coalescence. *The Astrophysical Journal*, 851(2):L35, Dec 2017.

[9] 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. Tests of general relativity with gw150914. *Physical Review Letters*, 116(22), May 2016.

[10] B.P. Abbott, R. Abbott, T.D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R.X. Adhikari, V.B. Adya, C. Affeldt, and et al. Gwtc-1: A gravitational-wave transient catalog of compact binary mergers observed by ligo and virgo during the first and second observing runs. *Physical Review X*, 9(3), Sep 2019.

[11] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, D. Agarwal, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, T. Akutsu, K. M. Aleman, G. Allen, A. Allocca, P. A. Altin, A. Amato, S. Anand, A. Ananyeva, S. B. Anderson, W. G. Anderson, M. Ando, S. V. Angelova, S. Ansoldi, J. M. Antelis, S. Antier, S. Appert, Koya Arai, and et al. Observation of gravitational waves from two neutron star–black hole coalescences. *The Astrophysical Journal Letters*, 915(1):L5, jun 2021.

[12] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen, A. Allocca, P. A. Altin, A. Amato, S. Anand, A. Ananyeva, S. B. Anderson, and et al. Gw190412: Observation of a binary-black-hole coalescence with asymmetric masses. *Phys. Rev. D*, 102:043015, Aug 2020.

[13] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen, A. Allocca, P. A. Altin, A. Amato, S. Anand, A. Ananyeva, S. B. Anderson, and et al. GW190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object. *The Astrophysical Journal*, 896(2):L44, jun 2020.

[14] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, and et al. Gw190521: A binary black hole merger with a total mass of 150 $M_{\odot}$. *Phys. Rev. Lett.*, 125:101102, Sep 2020.

[15] R. Abbott et al. GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. 10 2020.

[16] F Acernese, M Agathos, K Agatsuma, D Aisa, N Allemandou, A Allocca, J Amarni, P Astone, G Balestri, G Ballardin, and et al. Advanced virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, Dec 2014.

[17] Miguel Alcubierre. *Introduction to 3+1 Numerical Relativity*. International Series of Monographs on Physics. Oxford University Press, 2008.

[18] Miguel Alcubierre, Juan Carlos Degollado, and Marcelo Salgado. Einstein-maxwell system in3+1form and initial data for multiple charged black holes. *Physical Review D*, 80(10), Nov 2009.

[19] Asimina Arvanitaki, Masha Baryakhtar, Savas Dimopoulos, Sergei Dubovsky, and Robert Lasenby. Black hole mergers and the qcd axion at advanced ligo. *Physical Review D*, 95(4), Feb 2017.

[20] Asimina Arvanitaki, Masha Baryakhtar, and Xinlu Huang. Discovering the qcd axion with black holes and gravitational waves. *Physical Review D*, 91(8), Apr 2015.

[21] Asimina Arvanitaki, Savas Dimopoulos, Sergei Dubovsky, Nemanja Kaloper, and John March-Russell. String axiverse. *Physical Review D*, 81(12), Jun 2010.

[22] Asimina Arvanitaki and Sergei Dubovsky. Exploring the string axiverse with precision black hole physics. *Physical Review D*, 83(4), Feb 2011.

[23] John Baker, Jillian Bellovary, Peter L. Bender, Emanuele Berti, Robert Caldwell, Jordan Camp, John W. Conklin, Neil Cornish, Curt Cutler, Ryan DeRosa, Michael Eracleous, Elizabeth C. Ferrara, Samuel Francis, Martin Hewitson, Kelly Holley-Bockelmann, Ann Hornschemeier, Craig Hogan, Brittany Kamai, Bernard J. Kelly, Joey Shapiro Key, Shane L. Larson, Jeff Livas, Sridhar Manthripragada, Kirk McKenzie, Sean T. McWilliams, Guido Mueller, Priyamvada Natarajan, Kenji Numata, Norman Rioux, Shannon R. Sankar, Jeremy Schnittman, David Shoemaker, Deirdre

Shoemaker, Jacob Slutsky, Robert Spero, Robin Stebbins, Ira Thorpe, Michele Vallisneri, Brent Ware, Peter Wass, Anthony Yu, and John Ziemer. The laser interferometer space antenna: Unveiling the millihertz gravitational wave sky, 2019.

[24] Masha Baryakhtar, Marios Galanis, Robert Lasenby, and Olivier Simon. Black hole superradiance of self-interacting scalar fields. *Physical Review D*, 103(9), May 2021.

[25] Masha Baryakhtar, Robert Lasenby, and Mae Teo. Black hole superradiance signatures of ultralight vectors. *Physical Review D*, 96(3), Aug 2017.

[26] A. Bauswein, T. W. Baumgarte, and H.-T. Janka. Prompt merger collapse and the maximum mass of neutron stars. *Physical Review Letters*, 111(13), Sep 2013.

[27] A. Bauswein, H.T. Janka, K. Hebeler, and A. Schwenk. Equation-of-state dependence of the gravitational-wave signal from the ring-down phase of neutron-star mergers. *Phys. Rev. D*, 86:063001, 2012.

[28] A. Bauswein and N. Stergioulas. Unified picture of the post-merger dynamics and gravitational wave emission in neutron star mergers. *Phys. Rev.*, D91(12):124056, 2015.

[29] Andreas Bauswein and Nikolaos Stergioulas. Spectral classification of gravitational-wave emission and equation of state constraints in binary neutron star mergers. 2019.

[30] Jacob D. Bekenstein. Extraction of energy and charge from a black hole. *Phys. Rev. D*, 7:949–953, Feb 1973.

[31] Sebastiano Bernuzzi. Neutron Star Merger Remnants. *Gen. Rel. Grav.*, 52(11):108, 2020.

[32] Nigel T. Bishop and Luciano Rezzolla. Extraction of gravitational waves in numerical relativity. *Living Reviews in Relativity*, 19(1), Oct 2016.

[33] Sukanta Bose, Kabir Chakravarti, Luciano Rezzolla, B. S. Sathyaprakash, and Kentaro Takami. Neutron-star radius from a population of binary neutron star mergers. *Physical Review Letters*, 120(3), Jan 2018.

[34] Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space, 2016.

[35] Richard Brito, Alessandra Buonanno, and Vivien Raymond. Black-hole Spectroscopy by Making Full Use of Gravitational-Wave Modeling. *Phys. Rev.*, D98(8):084038, 2018.

[36] Richard Brito, Vitor Cardoso, and Paolo Pani. Massive spin-2 fields on black hole spacetimes: Instability of the schwarzschild and kerr solutions and bounds on the graviton mass. *Physical Review D*, 88(2), Jul 2013.

[37] Richard Brito, Vitor Cardoso, and Paolo Pani. Black holes as particle detectors: evolution of superradiant instabilities. *Classical and Quantum Gravity*, 32(13):134001, Jun 2015.

[38] Richard Brito, Vitor Cardoso, and Paolo Pani. Superradiance. *Lecture Notes in Physics*, 2020.

[39] Richard Brito, Shrobana Ghosh, Enrico Barausse, Emanuele Berti, Vitor Cardoso, Irina Dvorkin, Antoine Klein, and Paolo Pani. Gravitational wave searches for ultralight bosons with ligo and lisa. *Physical Review D*, 96(6), Sep 2017.

[40] Richard Brito, Shrobana Ghosh, Enrico Barausse, Emanuele Berti, Vitor Cardoso, Irina Dvorkin, Antoine Klein, and Paolo Pani. Stochastic and resolvable gravitational waves from ultralight bosons. *Physical Review Letters*, 119(13), Sep 2017.

[41] Richard Brito, Sara Grillo, and Paolo Pani. Black hole superradiant instability from ultralight spin-2 fields. *Phys. Rev. Lett.*, 124:211101, May 2020.

[42] Juan Calderon Bustillo, Tim Dietrich, and Paul D. Lasky. Higher-order gravitational-wave modes will allow for percent-level measurements of hubble's constant with single binary neutron star merger observations, 2020.

[43] Xi Chen, Diederik P. Kingma, Tim Salimans, Yan Duan, Prafulla Dhariwal, John Schulman, Ilya Sutskever, and Pieter Abbeel. Variational lossy autoencoder, 2017.

[44] Alvin J. K. Chua, Chad R. Galley, and Michele Vallisneri. Reduced-order modeling with artificial neurons for gravitational-wave inference. *Phys. Rev. Lett.*, 122(21):211101, 2019.

[45] Alvin J. K. Chua, Michael L. Katz, Niels Warburton, and Scott A. Hughes. Rapid generation of fully relativistic extreme-mass-ratio-inspiral waveform templates for LISA data analysis. *Phys. Rev. Lett.*, 126(5):051102, 2021.

[46] Michele Cicoli, Mark Goodsell, Joerg Jaeckel, and Andreas Ringwald. Testing string vacua in the lab: from a hidden cmb to dark forces in flux compactifications. *Journal of High Energy Physics*, 2011(7), Jul 2011.

[47] James Alexander Clark, Andreas Bauswein, Nikolaos Stergioulas, and Deirdre Shoemaker. Observing Gravitational Waves From The Post-Merger Phase Of Binary Neutron Star Coalescence. *Class. Quant. Grav.*, 33(8):085003, 2016.

[48] The LIGO Scientific Collaboration and the Virgo Collaboration. Gw190412: Observation of a binary-black-hole coalescence with asymmetric masses, 2020.

[49] Neil J Cornish and Tyson B Littenberg. Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches. *Classical and Quantum Gravity*, 32(13):135012, Jun 2015.

[50] Elena Cuoco et al. Enhancing Gravitational-Wave Science with Machine Learning. *Mach. Learn. Sci. Tech.*, 2(1):011002, 2021.

[51] Simone Dall'Osso, Bruno Giacomazzo, Rosalba Perna, and Luigi Stella. Gravitational waves from massive magnetars formed in binary neutron star mergers. *The Astrophysical Journal*, 798(1):25, Dec 2014.

[52] Giovanni Grilli di Cortona, Edward Hardy, Javier Pardo Vega, and Giovanni Villadoro. The qcd axion, precisely. *Journal of High Energy Physics*, 2016(1):34, Jan 2016.

[53] Tim Dietrich, David Radice, Sebastiano Bernuzzi, Francesco Zappa, Albino Perego, Bernd Brueugmann, Swami Vivekanandji Chaurasia, Reetika Dudi, Wolfgang Tichy, and Maximiliano Ujevic. Core database of binary neutron star merger waveforms and its application in waveform development, 2018.

[54] Sam R. Dolan. Instability of the proca field on kerr spacetime. *Physical Review D*, 98(10), Nov 2018.

[55] William E. East. Massive boson superradiant instability of black holes: Nonlinear growth, saturation, and gravitational radiation. *Physical Review Letters*, 121(13), Sep 2018.

[56] William E. East and Frans Pretorius. Superradiant instability and backreaction of massive vector fields around kerr black holes. *Physical Review Letters*, 119(4), Jul 2017.

[57] Paul J. Easter, Paul D. Lasky, Andrew R. Casey, Luciano Rezzolla, and Kentaro Takami. Computing Fast and Reliable Gravitational Waveforms of Binary Neutron Star Merger Remnants. 2018.

[58] Eanna E. Flanagan and Tanja Hinderer. Constraining neutron star tidal Love numbers with gravitational wave detectors. *Phys. Rev. D*, 77:021502, 2008.

[59] Hunter Gabbard, Chris Messenger, Ik Siong Heng, Francesco Tonolini, and Roderick Murray-Smith. Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy. 9 2019.

[60] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[61] Eric Gourgoulhon. 3+1 Formalism and Bases of Numerical Relativity. *arXiv e-prints*, pages gr–qc/0703035, March 2007.

[62] Andrei Gruzinov. Black hole spindown by light bosons, 2016.

[63] Matthias Hanauske, Kentaro Takami, Luke Bovard, Luciano Rezzolla, Jos A. Font, Filippo Galeazzi, and Horst Stocker. Rotational properties of hypermassive neutron stars from binary mergers. *Phys. Rev.*, D96(4):043004, 2017.

[64] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. 2016.

[65] Bob Holdom. Two u(1)'s and ϵ charge shifts. *Physics Letters B*, 166(2):196–198, 1986.

[66] Kenta Hotokezaka, Koutarou Kyutoku, Hirotada Okawa, Masaru Shibata, and Kenta Kiuchi. Binary neutron star mergers: Dependence on the nuclear equation of state. *Physical Review D*, 83(12), Jun 2011.

[67] Harold Jeffreys. *Theory of probability*. Clarendon Press, Oxford, 1961.

[68] Haifeng Jin, Qingquan Song, and Xia Hu. Auto-keras: An efficient neural architecture search system, 2018.

[69] Sebastian Khan and Rhys Green. Gravitational-wave surrogate models powered by artificial neural networks. *Phys. Rev. D*, 103(6):064015, 2021.

[70] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[71] Diederik P. Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improving variational inference with inverse autoregressive flow, 2017.

[72] Diederik P Kingma and Max Welling. Auto-encoding variational bayes, 2013.

[73] Luis Lehner, Steven L. Liebling, Carlos Palenzuela, O. L. Caballero, Evan O'Connor, Matthew Anderson, and David Neilsen. Unequal mass binary neutron star mergers and multimessenger signals. *Class. Quant. Grav.*, 33(18):184002, 2016.

[74] Luis Lehner, Steven L. Liebling, Carlos Palenzuela, and Patrick M. Motl. m=1 instability and gravitational wave signal in binary neutron star mergers. *Phys. Rev.*, D94(4):043003, 2016.

[75] Denis Martynov et al. Exploring the sensitivity of gravitational wave detectors to neutron star physics. *Phys. Rev. D*, 99(10):102004, 2019.

[76] J. Meidam, M. Agathos, C. Van Den Broeck, J. Veitch, and B. S. Sathyaprakash. Testing the no-hair theorem with black hole ringdowns using TIGER. *Phys. Rev.*, D90(6):064009, 2014.

[77] C. Messenger and J. Read. Measuring a cosmological distance-redshift relationship using only gravitational wave observations of binary neutron star coalescences. *Phys. Rev. Lett.*, 108:091101, 2012.

[78] Haixing Miao, Huan Yang, and Denis Martynov. Towards the Design of Gravitational-Wave Detectors for Probing Neutron-Star Physics. *Phys. Rev. D*, 98(4):044044, 2018.

[79] M.C. Miller, F.K. Lamb, A.J. Dittmann, S. Bogdanov, Z. Arzoumanian, K.C. Gendreau, S. Guillot, A.K. Harding, W.C.G. Ho, J.M. Lattimer, R.M. Ludlam, S. Mahmoodifar, S.M. Morsink, P.S. Ray, T.E. Strohmayer, K.S. Wood, T. Enoto, R. Foster, T. Okajima, G. Prigozhin, and Y. Soong. Psr j0030+0451 mass and radius from nicer data and implications for the properties of neutron star matter. , 887(1):L24, December 2019.

[80] Elias R. Most, L. Jens Papenfort, Veronica Dexheimer, Matthias Hanauske, Stefan Schramm, Horst Stcker, and Luciano Rezzolla. Signatures of quark-hadron phase transitions in general-relativistic neutron-star mergers. *Physical Review Letters*, 122(6), Feb 2019.

[81] Ezra Newman and Roger Penrose. An approach to gravitational radiation by a method of spin coefficients. *Journal of Mathematical Physics*, 3(3):566–578, 1962.

[82] Ken K.Y. Ng, Salvatore Vitale, Otto A. Hannuksela, and Tjonnie G.F. Li. Constraints on ultralight scalar bosons within black hole spin measurements from the ligo-virgo gwtc-2. *Physical Review Letters*, 126(15), Apr 2021.

[83] Paolo Pani, Vitor Cardoso, Leonardo Gualtieri, Emanuele Berti, and Akihiro Ishibashi. Perturbations of slowly rotating black holes: Massive vector fields in the kerr metric. *Phys. Rev. D*, 86:104017, Nov 2012.

[84] Vasileios Paschalidis. General relativistic simulations of compact binary mergers as engines for short gamma-ray bursts. *Classical and Quantum Gravity*, 34(8):084002, Mar 2017.

[85] R. Penrose. Gravitational Collapse: the Role of General Relativity. *Nuovo Cimento Rivista Serie*, 1, 1969.

[86] R. PENROSE and R. M. FLOYD. Extraction of rotational energy from a black hole. *Nature Physical Science*, 229(6):177–179, Feb 1971.

[87] Scott E. Perkins, Remya Nair, Hector O. Silva, and Nicolas Yunes. Improved gravitational-wave constraints on higher-order curvature theories of gravity, 2021.

[88] WILLIAM H. PRESS and SAUL A. TEUKOLSKY. Floating orbits, superradiant scattering and the black-hole bomb. *Nature*, 238(5361):211–212, Jul 1972.

[89] Frans Pretorius. Numerical relativity using a generalized harmonic decomposition. *Classical and Quantum Gravity*, 22(2):425–451, jan 2005.

[90] Davdi Radice, Sebastiano Bernuzzi, Walter Del Pozzo, Luke F. Roberts, and Christian D. Ott. Probing extreme-density matter with gravitational-wave observations of binary neutron star merger remnants. *The Astrophysical Journal*, 842(2):L10, Jun 2017.

[91] Luciano Rezzolla and Kentaro Takami. Gravitational-wave signal from binary neutron stars: A systematic analysis of the spectral properties. *Physical Review D*, 93(12), Jun 2016.

[92] Joo G. Rosa and Sam R. Dolan. Massive vector fields on the schwarzschild spacetime: Quasinormal modes and bound states. *Physical Review D*, 85(4), Feb 2012.

[93] Milton Ruiz, Ryan N. Lang, Vasileios Paschalidis, and Stuart L. Shapiro. Binary neutron star mergers: A jet engine for short gamma-ray bursts. *The Astrophysical Journal*, 824(1):L6, Jun 2016.

[94] Yuichiro Sekiguchi, Kenta Kiuchi, Koutarou Kyutoku, and Masaru Shibata. Effects of hyperons in binary neutron star mergers. *Physical Review Letters*, 107(21), Nov 2011.

[95] Masaru Shibata and Kenta Hotokezaka. Merger and mass ejection of neutron star binaries. *Annual Review of Nuclear and Particle Science*, 69(1):4164, Oct 2019.

[96] Andoni Torres-Rivas, Katerina Chatziioannou, Andreas Bauswein, and James Alexander Clark. Observing the post-merger signal of GW170817-like events with improved gravitational-wave detectors. *Phys. Rev.*, D99(4):044014, 2019.

[97] Steven Weinberg. A new light boson? *Phys. Rev. Lett.*, 40:223–226, Jan 1978.

[98] Huan Yang, Vasileios Paschalidis, Kent Yagi, Luis Lehner, Frans Pretorius, and Nicols Yunes. Gravitational wave spectroscopy of binary neutron star merger remnants with mode stacking. *Phys. Rev.*, D97(2):024049, 2018.

[99] Huan Yang, Kent Yagi, Jonathan Blackman, Luis Lehner, Vasileios Paschalidis, Frans Pretorius, and Nicolas Yunes. Black hole spectroscopy with coherent mode stacking. *Phys. Rev. Lett.*, 118(16):161101, 2017.

[100] H. Yoshino and H. Kodama. Bosenova collapse of axion cloud around a rotating black hole. *Progress of Theoretical Physics*, 128(1):153190, Jul 2012.

[101] Hirotaka Yoshino and Hideo Kodama. Gravitational radiation from an axion cloud around a black hole: Superradiant phase. *Progress of Theoretical and Experimental Physics*, 2014(4), Apr 2014.

[102] Hirotaka Yoshino and Hideo Kodama. The bosenova and axiverse. *Classical and Quantum Gravity*, 32(21):214001, Oct 2015.

[103] Ya. B. Zel'Dovich. Generation of Waves by a Rotating Body. *Soviet Journal of Experimental and Theoretical Physics Letters*, 14:180, August 1971.

[104] Miguel Zilho, Helvi Witek, and Vitor Cardoso. Nonlinear interactions between black holes and proca fields. *Classical and Quantum Gravity*, 32(23):234003, Nov 2015.

APPENDICES

Appendix A

Convergence Axisymmetry

To verify that the integration converges, we run the simulation at three different resolutions. We verify that the measured proca and black hole quantities converge to the same solution at different resolutions. We also take a look at the constraints at the different resolutions.

Figures. A.4,A.2 shows the energy and angular momentum quantities at three different resolutions. The medium resolution being twice the resolution of the low one, and the high one being twice the resolution of the medium one. We can see that the three resolutions converge the same solution. Similarly, we look at the BHs quantities at different resolutions. Figures. A.5,A.3, show the angular momentum and mass of the cloud at two different resolutions. Again, the higher resolution is twice the resolution of the lower one. We see again that the quantities converge to the same solution at both resolutions. Since the numerical scheme is fourth order, we expect that when doubling the resolution of a simulation, the norm should decrease by a factor of 16. Figure. A.1 shows the convergence of the norm of the constraint violation. We see from this figure that we do not get this factor. This is due to the norm being violated initially by the perturbation introduced with the Proca field.

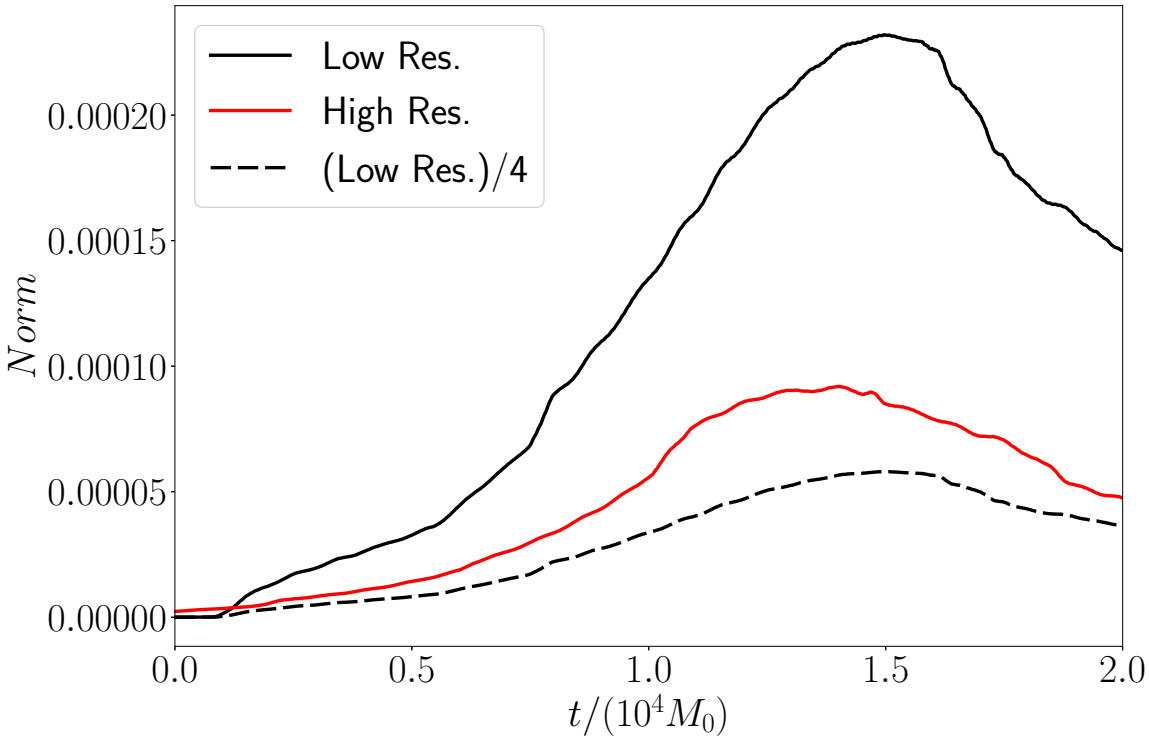

Figure A.1: Norm of the constraints violation at low and high resolution through time. The high resolution simulation is double the resolution of the low resolution one. We expect the numerical scheme to converge to fourth order, and so we expect the norm to decrease by a factor of 16 when the resolution is doubled. We do not observe this here since the constraints are initially violated by the Proca field perturbation.

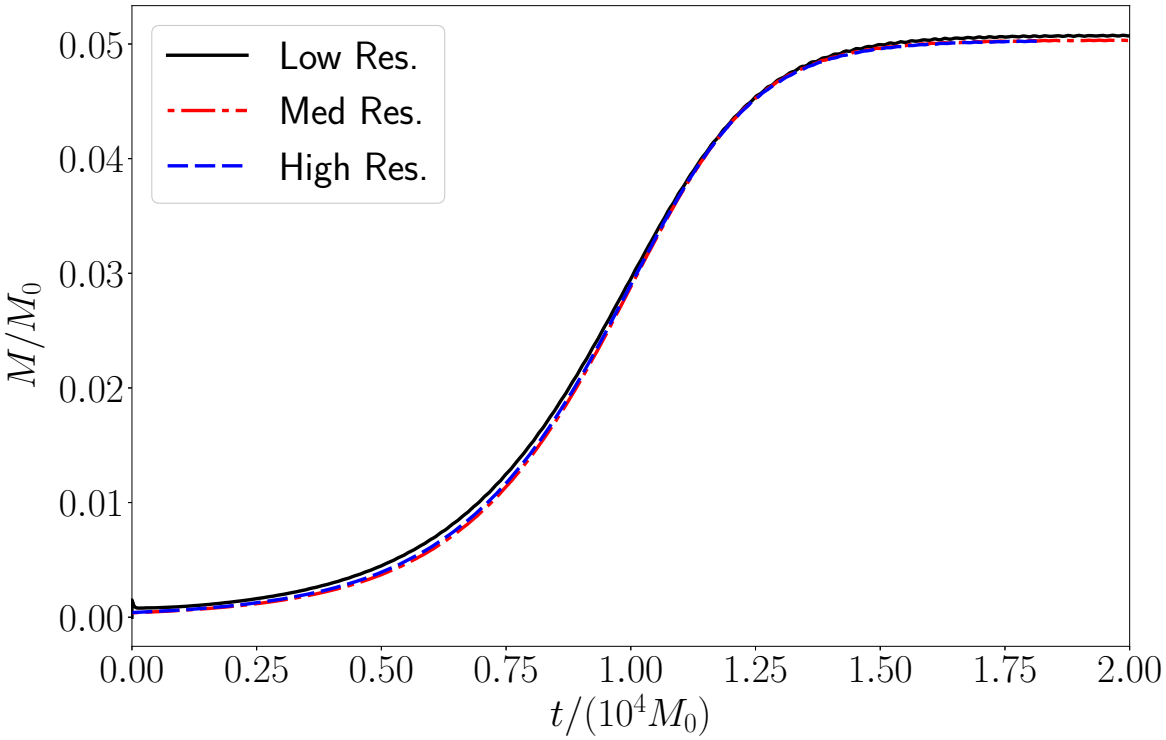

Figure A.2: Cloud energy as a function of time for $\tilde{\mu} = 0.5$, $\lambda = 0.5$, $a = 0.99$, at three different resolutions. The finer resolution simulations are double the resolution of the coarser ones and started with half the amplitude of the Proca cloud. They are then time shifted. We see that three simulations match, indicating convergence.

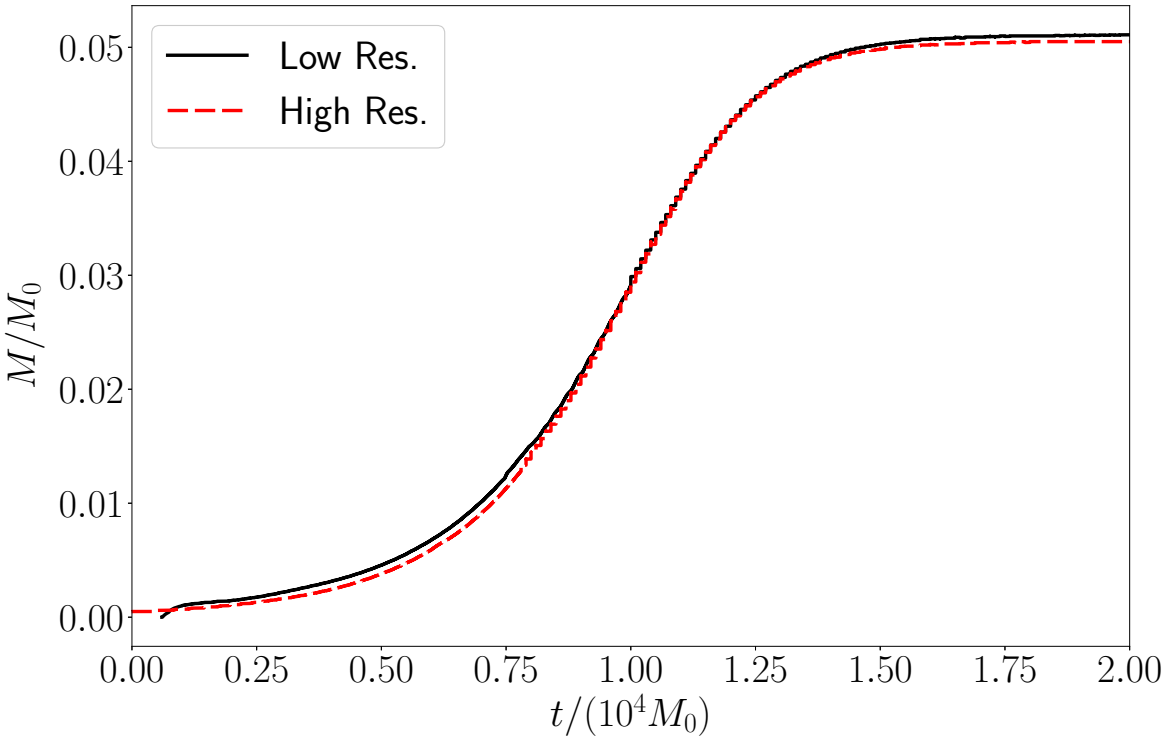

Figure A.3: BHs mass as a function of time for $\tilde{\mu} = 0.5$, $\lambda = 0.5$, $a = 0.99$, at two different resolutions. The finer resolution simulations are double the resolution of the coarser ones. We see that two simulations match, indicating convergence.

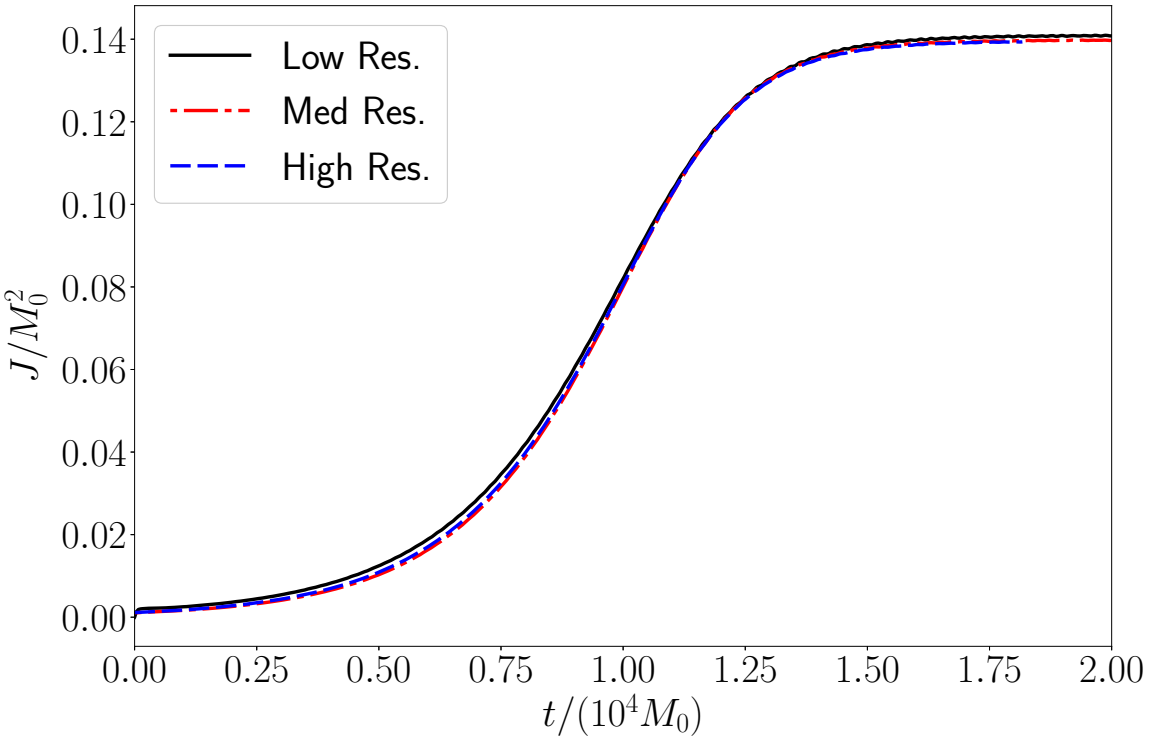

Figure A.4: Cloud angular momentum as a function of time for $\tilde{\mu} = 0.5$, $\lambda = 0.5$, $a = 0.99$, at three different resolutions. The finer resolution simulations are double the resolution of the coarser ones and started with half the amplitude of the Proca cloud. They are then time shifted. We see that three simulations match, indicating convergence.

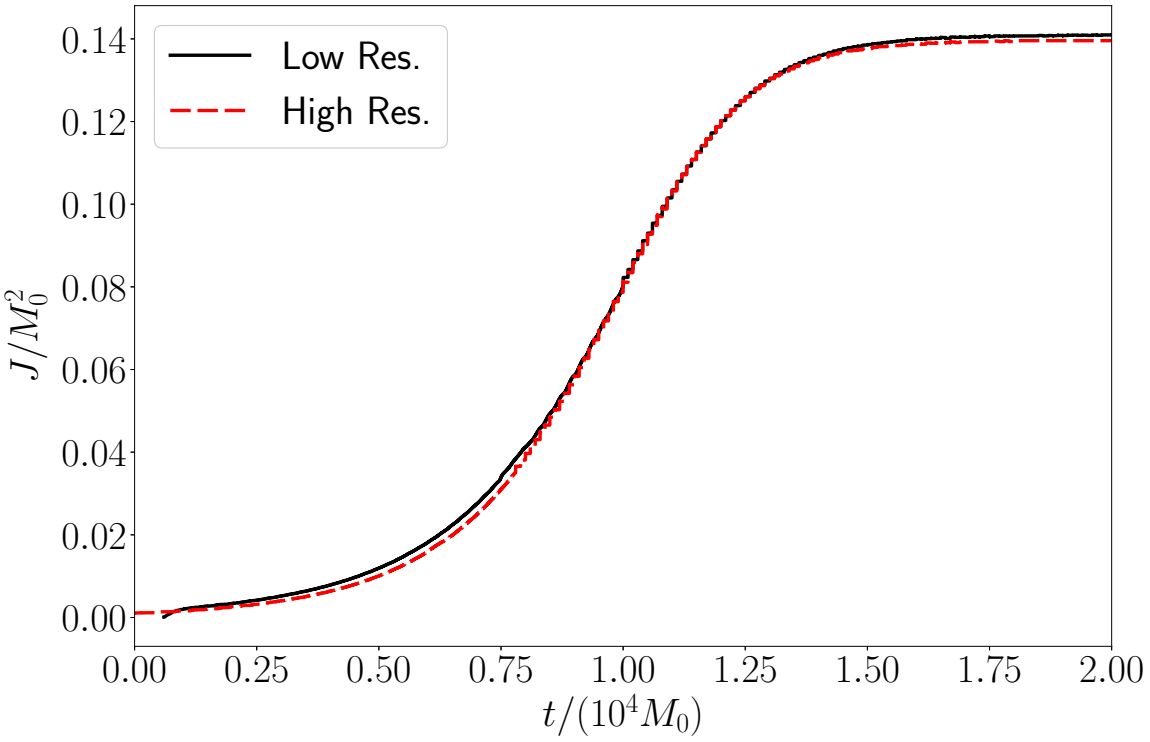

Figure A.5: BHs angular momentum as a function of time for $\tilde{\mu} = 0.5$, $\lambda = 0.5$, $a = 0.99$, at two different resolutions. The finer resolution simulations are double the resolution of the coarser ones. We see that two simulations match, indicating convergence.

Appendix B

Deforming Axisymmetric Data

Suppose we have initial data (E^i, χ^i, χ) which satisfy the axisymmetric massive Proca constraints, we want to find a new set of data $(\tilde{E}^i, \tilde{\chi}^i, \tilde{\chi})$ which satisfy the full nonlinear Proca constraints.

Let $\tilde{\chi} = \chi + g$ and $\tilde{\chi}_i = \chi_i$, then we have that,

$$D_i E^i = -\mu^2 \chi \tag{B.1}$$

Now suppose λ is small, we want to find g such that,

$$\tilde{D}_i \tilde{E}^i = -(\mu^2 + 4\lambda \tilde{X}_i \tilde{X}^i)\tilde{\chi} \tag{B.2}$$

Expanding eq. B.2 with $\tilde{\chi} \to \chi + g$ we get,

$$D_i E^i = -(\mu^2(\chi + g) + 4\lambda(\chi_i \chi^i - g^2 + \chi^2)(\chi + g)) \tag{B.3}$$

Now since, (E^i, χ^i, χ) satisfy the massive Proca equations,

$$0 = -(\mu^2 + 4\lambda(\tilde{\chi}_i \tilde{\chi}^i - g^2 + \chi^2))(\tilde{\chi} - g) \tag{B.4}$$

$$0 = 4\lambda g^3 + \chi g^2 - \left[\mu^2 + 4\lambda X_i X^i\right] g - 4\lambda X_i X^i \chi \tag{B.5}$$

We can solve for g by solving for the cubic,

$$ag^3 + bg^2 + cg + d = 0$$

Where,

$$\begin{aligned} a &= 4\lambda \\ b &= \chi \\ c &= -\mu^2 - 4\lambda X_i X^i \\ d &= -4\lambda X_i X^i \chi \end{aligned}$$

We can analytically solve for g when $\lambda \neq 0$ with,

$$g = \left(q + \sqrt{q^2 + (\ell - p^2)^3}\right)^{1/3} + \left(q - \sqrt{q^2 + (\ell - p^2)^3}\right)^{1/3} + p$$

$$\begin{aligned} p &= -\frac{b}{3a} \\ q &= p^3 + \frac{bc - 3ad}{6a^2} \\ \ell &= \frac{c}{3a} \end{aligned}$$

This equation will works for cubics with a single real root, that is, in the case where

$$\Delta = 18abcd - 4b^3 d + b^2 c^2 - 4ac^3 - 27a^2 d^2 < 0$$

We can approximate the solution with no symmetry assumption by first evolving the fields in axissymmetry until the instability sets in and then deforming the fields as proposed above and evolving the new fields in full 3D.

Appendix C

Waveform model with partial information

Thanks to numerous developments in solving the equations of general relativity, the binary black hole waveform is largely known (for comparable mass ratios). The corresponding waveform models, whether they are constructed phenomenologically by matching numerical waveforms with analytical approximations, or through other methods such as the effective-one-body approach, are all *deterministic*. On the other hand, to construct the waveform model for post-merger neutron stars, we are facing a different problem where our current knowledge is insufficient to predict an accurate waveform given the orbital parameters. This problem is likely to persist until there are significant improvements in our understanding of the neutron star EOS (which likely will come from the detections themselves) and in our modelling of the merger/post-merger process with a sufficiently complete physical prescription.

Nevertheless, although current numerical simulations do not provide the full answer, they do converge on certain features (such as the dominant peak in the post-merger spectrum). It is natural to ask how one could construct a waveform model utilizing this *partial* information, while taking into account the inherent uncertainties. To answer this question, we present a general framework for waveforms with partial information.

A *non-deterministic* waveform model can be written as $h = h(C; \mathcal{I})$. The control variables C are the physical quantities that determine the waveform assuming perfect knowledge. In the binary case, they are the orbital parameters, which can be determined from the inspiral waveform measurement (which in general will be much louder than the post-merger signal) to a certain accuracy. The latent variables $\mathcal{I}$ encapsulate all the

theoretical uncertainties and modelling errors. In the limit that a waveform model becomes deterministic, $\dim(\mathcal{I}) = 0$.

C.0.1 Detection

For the purpose of detection, the control variables are already constrained by the inspiral measurement. As a result, the posterior distribution of the control variables $P_{\rm in}(C)$ can be used as a prior distribution for post-merger detection. One way to quantify the statistical significance of a post-merger event detection is using the hypothesis test framework. Using the post-merger gravitational wave data stream s, we compare the following two hypotheses:

$$\begin{aligned} &\mathcal{H}_1 : s = h + n\,, \\ &\mathcal{H}_2 : s = n \end{aligned} \tag{C.1}$$

where n represents noise.

The significance of detection can be characterized by the Bayes factor, which is defined as

$$\mathcal{B} := \frac{P(s|\mathcal{H}_1)}{P(s|\mathcal{H}_2)}\,, \tag{C.2}$$

and the evidence function $P(s|\mathcal{H}_i)$ can be computed by

$$P(s|\mathcal{H}_i) = \int dC d\mathcal{I} \mathcal{L}_i(C,\mathcal{I}) P_{\rm in}(C) P_{\rm in}(\mathcal{I})\,, \tag{C.3}$$

where $P_{\rm in}(\mathcal{I})$ is the prior weight function for the latent variables. The likelihood function $\mathcal{L}_i$ is given by

$$\begin{aligned} \mathcal{L}_1 &= \frac{1}{Z_n} \prod_{f>0} \exp\left(-\frac{2|s-h|^2}{S_n}\right)\,, \\ \mathcal{L}_2 &= \frac{1}{Z_n} \prod_{f>0} \exp\left(-\frac{2|s|^2}{S_n}\right)\,, \end{aligned} \tag{C.4}$$

where S_n is the single-side spectral density of the detector, and Z_n is a common normalization constant. The larger the Bayes factor is, the more significant the detection.

According to the Jeffreys scale of interpretation of Bayes Factor [67], the evidence is "very strong" if $\mathcal{B} > 10$.

Even without prior knowledge of the control variables, we can still define the SNR of an event as

$$\text{SNR} := 2\text{Max}_{C,\mathcal{I}} \left(\int df \frac{h(C,\mathcal{I})s^* + h.c.}{S_n} \right) . \tag{C.5}$$

Given a threshold in SNR, it is straightforward to evaluate the average rate of detecting a false signal due to the detector noise background, which is often referred to as the false-alarm rate. This part is similar to binary black hole detection. Generally speaking, for a given false alarm rate, the threshold SNR is higher if the latent space is larger.

C.0.2 Parameter estimation

While we will not focus on using the model discussed in this paper for parameter estimation, here we briefly outline how such a non-deterministic waveform model might be used for parameter estimation. After we compute the posterior distribution of C and $\mathcal{I}$ based on the gravitational-wave data s, the posterior distribution of the control variables can be obtained by marginalizing the latent variables:

$$P(C|s) = \int d\mathcal{I} P(C,\mathcal{I}|s) . \tag{C.6}$$

In reality, the computational cost for constructing the posterior distribution of C and $\mathcal{I}$ increases with the size of the latent space. Therefore, an efficient non-deterministic waveform model should minimize the size of the latent space while still being able to fit the true waveforms.

The latent space encodes information about the unknown physical parameters and so it is also possible to use the latent space to infer the unknown physical parameters. This can be accomplished if a map between the unknown physical parameters and latent space is built. In general this is not an easy task as the mapping can be quite complicated, but we have seen that this is possible in Fig. 3.8 with our fiducial model.

Appendix D

Toy Model

As a simple demonstration of a CVAE being used as a generative model, we train a CVAE on toy data which consists of a sum of sine functions. This is motivated by the observation that the post-merger waveform seems to be dominated by a small number of main frequencies which are believed to not vary significantly in time [47].

We generate $\mathcal{O}(10^3)$ waveforms of the form

$$w = \sum_{i=1}^{3} \sin(\omega_i t + \phi_i) \tag{D.1}$$

where $\phi_i \in [0, 2\pi]$ and $\omega_i \in [1, 4]$, $t \in [0, 15]$ with a sample frequency of $1000/15$. The phase and frequencies are picked randomly from a uniform distribution. The data set is then split into a training and a testing set. We randomly pick 60% of the total set for training, and the other 40% as the testing set. Once the network is trained, we measure the match using eq. 3.11.

We train the neural network using the frequencies as the conditional variables and let the latent space have three variables. The neural network has three encoding layers with $\{100, 50, 10\}$ neurons in the layers. The decoder has the same structure but inverted. The encoder and decoder have rectified linear unit functions as the activation functions, and the latent and output have linear functions as activation functions. Figure D.1 is an example output of the waveform recreated by the CVAE. We found that the CVAE was easily able to recreate the waveform given sufficient amounts of data (~ 600 waveforms).

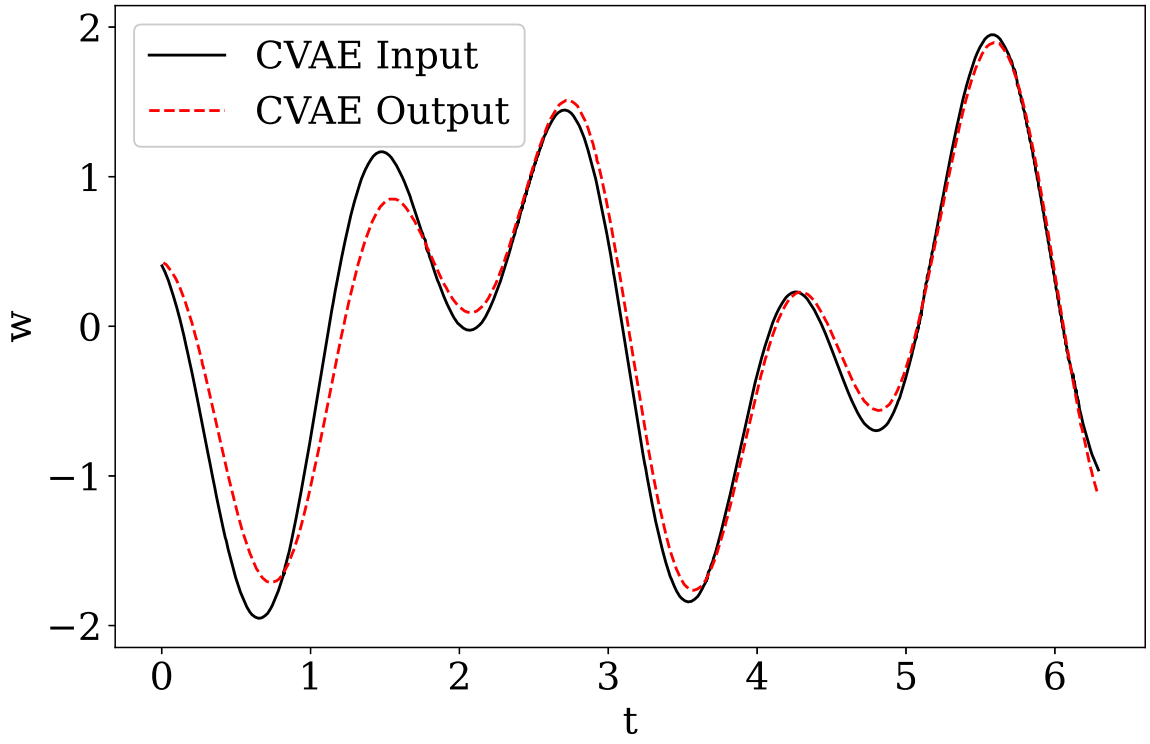

Figure D.1: A comparison of the waveform generated by Eq. (D.1) (black curve) to the one recreated by the CVAE (red curve). The waveform is produced by sampling the latent space and picking the latent variables which produce the highest match. The match is 0.95.

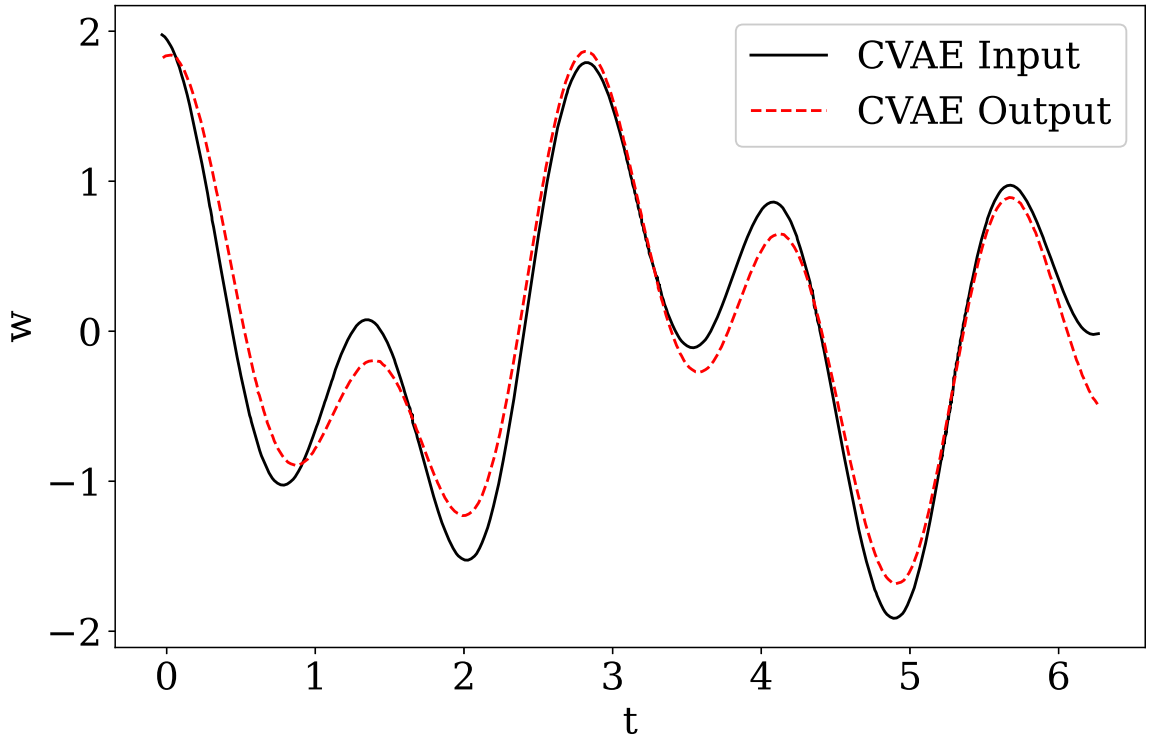

Figure D.2: To see that the latent space learned information about the parameters ϕ_i we generate a new wave using Eq. (D.1) with the same ω_i as Fig D.1 but different ϕ_i. We search over our latent space to see if we can also recreate that waveform. Here we see the waveform generated by Eq. (D.1) (black curve) could be recreated by the CVAE (red curve).

Appendix E

Fitting in time domain

In this section we discuss the issues of training the CVAE directly on the numerical waveforms. We show that training directly on the numerical waveforms with the amount of data we have leads to overfitting.

We constructed a CVAE similar to the one in section D.1 where the input layer took in the time series of the numerical waveform. The training set consisted of the numerical waveforms and we augmented the dataset by adding random phases to the waveforms. Using this method we increased our dataset to 500 elements.

After training, we compute the match of the recreated waveforms, finding that the neural network was in fact capable of recreating the waveforms it trained on, but could not interpolate between different waveforms. That is, the neural network was overfitting. The overfitting can be seen in Fig. E.1 where we have the history of the MSE throughout the training. We see that the MSE of the training and testing sets is diverging, indicating that the CVAE is learning the training set well, but is not able to interpolate. This is expected given the small amount of training data. In the following section, we will discuss methods to alleviate this problem, and provide an estimate of the amount of data required for this model to function.

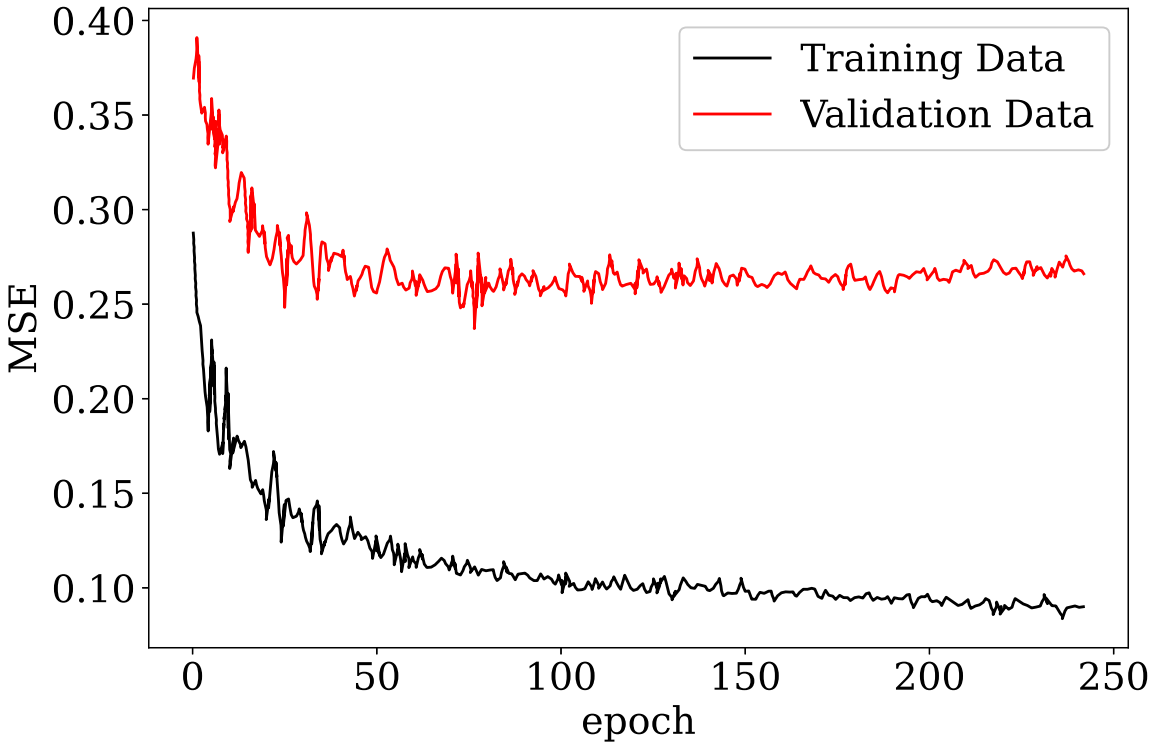

Figure E.1: Training history of the CVAE with time domain numerical waveform data. The plot depicts the MSE contribution of the loss function, that is how well can it recreate the waveform. The difference between the training data and validation data is a symptom of overfitting.

Appendix F

Data preprocessing

In this section we discuss the details of the data processing done to the data in the main text. The technique used here to preprocess the data is called data whitening, the basic idea is to center the dataset, normalize it, and remove the correlations.

Certain subsets of the parameters in the model given by Eq. (3.10) turn out to be strongly correlated. This is illustrated by the correlation matrix (Fig. F.1) of the data. Intuitively, you would suspect that this would make it easier for the CVAE to reproduce the data. This turns out to not be completely true since it creates an incentive to only learn the highly correlated parameters and disregard the uncorrelated ones. This occurs since the correlated terms end up dominating the mean square term of the cost function. We also found that our data tends toward posterior collapse when the correlations are left as is.

To deal with these issues, we can use some basic data processing methods such as the whitening transform. This transformation consists of three steps: (i) centering the dataset, (ii) decorrelating the dataset, and (iii) rescaling the dataset. To center the dataset, we simply find the mean of every feature and subtract it from the dataset. This guarantees that the mean of the new dataset is 0. To decorrelate the data we find the eigenvalues λ_i and eigenvectors v_i of the covariance matrix and define $V = [v_1, v_2, ..., v_n]$ where the columns are the eigenvectors. The decorrelated data X_{dec} is found with respect to the original data X by projecting X on V:

$$X_{\text{dec}} = X \cdot V \ . \tag{F.1}$$

Finally, to rescale the data, we divide through by the eigenvalues by defining the diagonal matrix $P = \text{diag}(1/\sqrt{\lambda_1 + \epsilon}, 1/\sqrt{\lambda_2 + \epsilon}, ..., 1/\sqrt{\lambda_n + \epsilon})$,

$$X_{\text{whiten}} = X_{\text{dec}} \cdot P \tag{F.2}$$

where X_{whiten} is the whitened data, and ϵ is a small number to deal with the cases where the eigenvalue is zero (though this does not arise for the cases we consider). Once the data is whitened, we can also tune the parameter β in the cost function such that no posterior collapse occurs. The tuning of the parameter is done by varying the parameter over many training sessions to find the optimal value [64]. After performing these steps, the correlation matrix of the transformed data will just be the identity matrix.

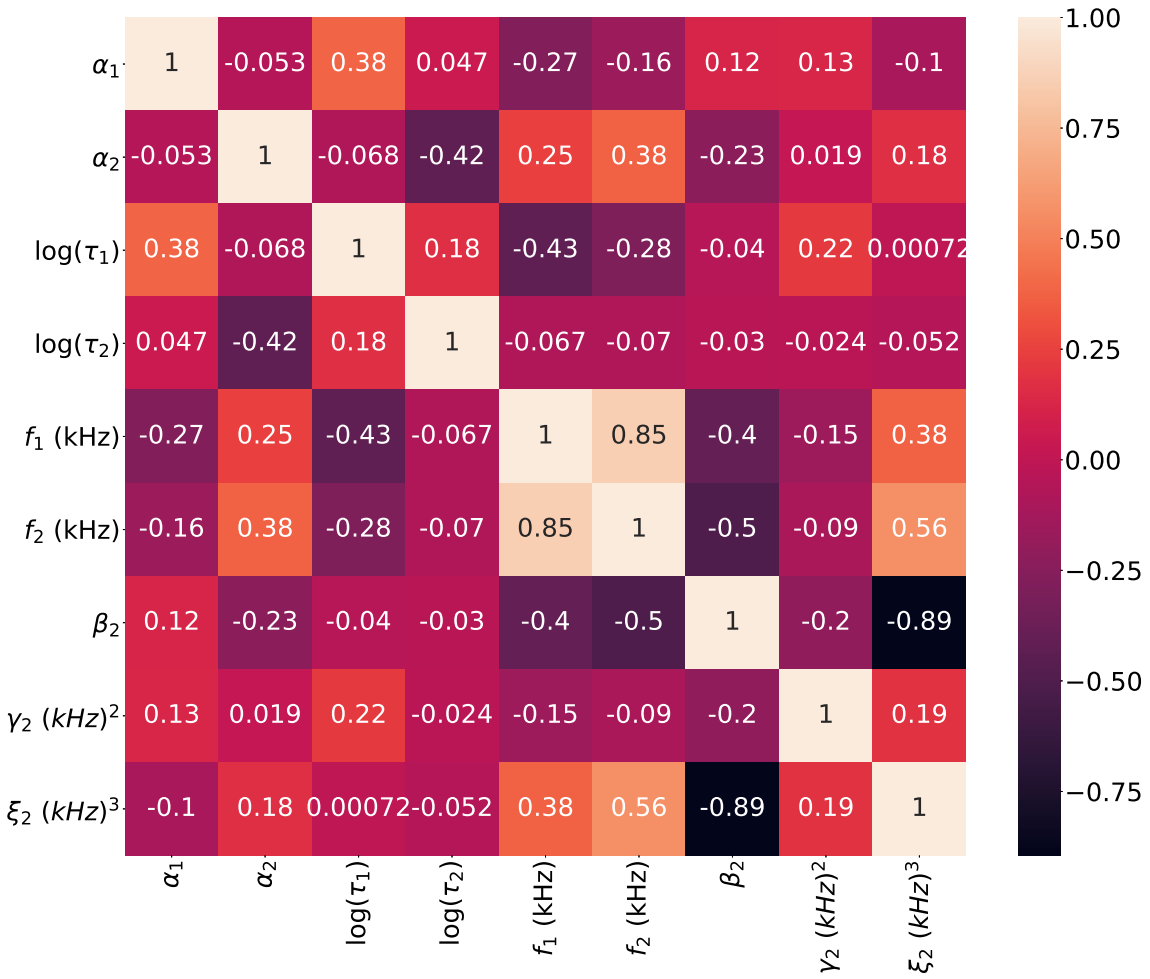

Figure F.1: Normalized covariance matrix of the parameters defined in Eq. (3.10), when fitted to the numerical relativity waveforms.

Appendix G

Fiducial Model and CVAE Parameters

Tables G.1 and G.2 give the parameters for Eq. (3.14) presented in Sec. 3.4.1 by fitting the numerical waveforms from the CoRe database [53]. The fits are found by first finding the compactness using the f_2 - $\mathcal{C}$ relation. Then we fit f_1 with the compactness to Eq. (3.12), and perform a linear fit to the other parameters. This gives us the vectors $\vec{A}$ and $\vec{B}$, with the specific values shown in Tables G.1 and G.2. We then compute the covariance matrix Σ, finding the values shown in Table G.3.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}
-0.99	0.78	2.73	2.41	-6254.30	2329.10	-262.69	10.65	-3.12	51.90	89.07	-0.03	-2.22	0.04

Table G.1: Parameters for $\vec{A}$ in the fiducial model

b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_8
0.12	0.00	2.73	2.42	0.00	0.00	0.04	0.31	0.00

Table G.2: Parameters for $\vec{B}$ in the fiducial model

8.23×10^{-3}	-1.54×10^{-4}	4.95×10^{-2}	2.19×10^{-3}	-4.39×10^{-3}	4.28×10^{-3}	1.08×10^{-4}	4.85×10^{-3}	-1.20×10^{-5}
-1.54×10^{-4}	1.00×10^{-3}	-3.04×10^{-3}	-6.87×10^{-3}	-8.23×10^{-4}	-3.41×10^{-3}	-7.05×10^{-5}	2.55×10^{-4}	7.06×10^{-6}
4.95×10^{-2}	-3.04×10^{-3}	2.01	1.28×10^{-1}	-8.74×10^{-2}	1.12×10^{-1}	-5.41×10^{-4}	1.29×10^{-1}	1.30×10^{-6}
2.19×10^{-3}	-6.87×10^{-3}	1.28×10^{-1}	2.64×10^{-1}	-3.34×10^{-3}	8.79×10^{-3}	-1.48×10^{-4}	-5.18×10^{-3}	-3.41×10^{-5}
-4.39×10^{-3}	-8.23×10^{-4}	-8.74×10^{-2}	-3.34×10^{-3}	2.92×10^{-2}	-2.66×10^{-15}	5.44×10^{-5}	-1.05×10^{-2}	-3.34×10^{-5}
4.28×10^{-3}	-3.41×10^{-3}	1.12×10^{-1}	8.79×10^{-3}	-2.66×10^{-15}	7.24×10^{-2}	1.32×10^{-3}	9.34×10^{-3}	-1.96×10^{-4}
1.08×10^{-4}	-7.05×10^{-5}	-5.41×10^{-4}	-1.48×10^{-4}	5.44×10^{-5}	1.32×10^{-3}	9.19×10^{-5}	-7.99×10^{-4}	-1.09×10^{-5}
4.85×10^{-3}	2.55×10^{-4}	1.29×10^{-1}	-5.18×10^{-3}	-1.05×10^{-2}	9.34×10^{-3}	-7.99×10^{-4}	1.75×10^{-1}	1.01×10^{-4}
-1.20×10^{-5}	7.06×10^{-6}	1.30×10^{-6}	-3.41×10^{-5}	-3.34×10^{-5}	-1.96×10^{-4}	-1.09×10^{-5}	1.01×10^{-4}	1.61×10^{-6}

Table G.3: Elements of the covariance matrix Σ in the fiducial model

Finally, the details of the CVAE used in Sec. 3.4.2 are given in Table G.4.

	Input	Layer 1	Layer 2	Latent Layer	Layer 3	Layer 4	Output
Number of Neurons	9	345	335	4	335	345	9
Activation Function	–	ReLu	ReLu	Linear	ReLu	ReLu	tanh
Type		Dense	Dense	Lambda	Dense	Dense	

Table G.4: Parameters for the CVAE used in Sec. 3.4.2. All the layers in the neural network consist of dense layers, except for the latent space where we have a lambda layer which takes the output of layer 2 as the mean and variance of a multivariate normal and outputs a sample from the multivariate normal.