Inferring the neutron star equation of state using Machine Learning methods

Jessica Irwin

2025

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

School of Physics & Astronomy

College of Science and Engineering

University of Glasgow

Abstract

The first decade of gravitational wave (GW) detection using the global ground-based GW detector network has facilitated a new era of neutron star observation. From the GW signal produced when two neutron stars (NSs) inspiral and merge, one can directly measure the masses of the two NSs and, importantly, their tidal deformability, a direct measure of the behaviour of matter in the system. This parameter is unique to GW astronomy and therefore offers an independent method to infer the neutron star (NS) equation of state. Due to lack of precision in measurement of NS macroscopic parameters, the equation of state – the relationship between the pressure and density within the ultra-dense neutron-rich matter of a neutron star – is still widely unknown. Though there are various inference schemes to infer the NS equation of state given electromagnetic (EM) and GW observation, these are often computationally and temporally expensive processes.

Recently, the introduction of machine learning (ML) tools in astronomical data analysis have facilitated the handling of large amounts of data and the processing of this data efficiently, to find broad trends or features. These tools will become necessary when considering future GW detection, where we expect increased sensitivity of detectors as well as orders of magnitude more detections, including those of binary neutron star (BNS) mergers. In this thesis, we apply ML methods, notably a type of generative ML model called a Normalising Flow, in developing tools through which we can infer the NS equation of state in current and future observation of gravitational waves (GWs) from BNS mergers.

We firstly introduce a Normalising Flow trained to perform the mapping of equation of state data conditioned on BNS event parameters. Once trained, the Flow can be conditionally sampled to return an equation of state posterior given posterior samples from a single GW event in less than 1 second. Simulation studies demonstrate the validity of the Flow result, alongside the equation of state posterior for the GW event GW170817, which is in agreement with the existing accepted result. The tool facilitates rapid follow-up of GWs from BNS mergers for improved communication with EM astronomers.

In setting the scene for hierarchical inference of the NS equation of state given multiple observations of GWs from BNS mergers, we discuss the performance of Normalising Flows in mapping complex high-dimensionality data sets. The introduction of a new equation of state training data set makes use of an autoencoder for data compression, which achieves root-mean-squared (RMS) error on the equation of state reconstruction on the scale of 10^{-3} for normalised mean-subtracted equations of state. We demonstrate abnormalities in the Normalising Flow's performance in mapping regions of the equation of state space, which manifests as severe spikes and troughs of probability. We highlight the dangers of inconsiderate application of Normalising Flows to mapping any high-dimensionality data set. We finally introduce the regeneration Flow, built to learn the mapping of the joint data and conditional spaces at once, such that it can be sampled repeatedly during training for unlimited training data generation. We demonstrate how this improves Normalising Flow training and reduces the fluctuations in magnitude of probability over the surface of the learned data space, promoting generic learning.

We apply the improved Normalising Flow to hierarchical analysis of the neutron star equation of state, firstly in inferring the combined equation of state given the first two BNS merger observations. We make use of a full ML parameter estimation (PE) pipeline to perform a simulation study of inferring the true equation of state given multiple simulated BNS events associated to three known equations of state. We demonstrate that as we increase the number of events, the quality of sampling the equation of state posterior decreases, suggesting a highly multi-modal space and/or

inaccurate model. We introduce an alternative method for hierarchical inference which is more robust by using the Normalising Flow instead to sample. With the new method, the result of combining information from up to 16 BNS events associated to two out of three simulated equations of state produce a constrained equation of state posterior which agrees with the truth. We highlight the computational expense of the workflow; inference of up to 16 events with the new method takes less than 1 hour. This validates the use of Normalising Flows for hierarchical inference of the NS equation of state in future observing runs, when the number of events are expected to be in the 10s. We suggest substantial future work to improve the sampling quality and, beyond this, tests for the next generation of GW detection to validate the use of Normalising Flows in understanding neutron star matter.

Contents

Declaration

All work presented in this thesis was carried out by the author under the supervision of Prof. Ik Siong Heng and Dr. Chris Messenger in the Institute for Gravitational Research at the University of Glasgow. It has not been submitted for any other degree at the University of Glasgow, or at any other institution. The author is a member of the LIGO-Virgo-KAGRA (LVK) collaboration and therefore their work has benefitted from discussion with other researchers. All work in this thesis was carried out by the author unless otherwise stated.

Chapters 1, 2 and 3 contain introductory material and a literature review regarding neutron star physics, gravitational waves and Machine Learning. These chapters were written by the author with advice from Prof. Ik Siong Heng and Dr. Chris Messenger.

Chapter 4 presents work that builds on that of [1]. The initial concept and Normalising Flow model was developed by co-authors of the publication prior to the beginning the author's PhD. In late 2021/early 2022, the author adopted the project and began development work on the Normalising Flow model. Section 4.1 covers the theory behind the work which was developed by co-authors. The equation of state and macroscopic parameter data sets introduced in section 4.2 were provided by Dr. Arunava Mukherjee. Section 4.2 introduces the work which was undertaken by co-authors with additional investigation done by the author for the purposes of this thesis. The author completed the work discussed in sections 4.3 and the optimisation and the upgrade to the flow discussed in 4.4 under supervision of Prof. Ik Siong Heng and Dr. Chris Messenger. Results from this point onwards were produced by the author. Section 4.5.1 and 4.5.2.1 includes results published in [1].

Chapter 5 presents work undertaken by the author under the supervision of Prof. Ik Siong Heng and Dr. Chris Messenger. The unified equation of state data introduced in 5.1 produced was provided by Dr. Lami Suleiman. Details of the data provided are presented in [2]. The work of these chapters beyond the introduction of data were written by the author of this thesis with the advice of Prof. Ik Siong Heng and Dr. Chris Messenger.

Chapter 6 presents work undertaken by the author alongside work which has been published in [3]. The equation of state data used for the simulation study introduced in section 6.3 and for the publication [3] (also section 6.3.1) was, akin to chapter 5, provided by Dr. Lami Suleiman. The data which is described in section 6.3 and 6.3.1 was provided by Dr. Qian Hu, making use of the pipeline which is the subject of the paper. Results presented in 6.3.1.1 were produced by the author with input from the co-authors of the publication and are those presented in [3]. All results from section 6.3 onwards were produced by the author under advice from Prof. Ik Siong Heng and Dr. Chris Messenger.

Chapter 7 summarises the work presented in all previous chapters and was written by the author.

Acknowledgements

First and foremost, I'd like to thank my supervisors Ik Siong Heng and Chris Messenger. Siong, thank you for reminding me to have thick skin and take any opportunities that come my way. Chris, thank you for helping me out of holes/voids/troughs, and for spending so much time helping me understanding how I got in them. I have appreciated both your honest and kind perspectives throughout my PhD.

I feel very lucky to have had the opportunity to do a PhD, and in particular to complete one amongst world-class researchers. I have been inspired throughout my PhD by my colleagues in the IGR, and in particular owe thanks to Daniel Williams, Michael Williams and Federico Stachurski for their advice and wisdom during our overlapping time working together. More widely, I have appreciated the environment of the Kelvin Building and the company of the people who work in it, who made even the most difficult days bearable. I'd finally like to thank my closest colleagues, those in office 253, who made every day enjoyable and were always there to lend an ear or a biscuit. I feel most lucky to have landed amongst you all.

I'd also like to thank those at LIGO Hanford, in particular Camilla, Rick & Kiyomi, Dripta and Tony for being so welcoming during my time there as a Fellow, as well as Jennie and Brina for their friendship away from home. Thank you to Prof. Jocelyn Read, Dr. Lami Suleiman and others at California State University, Fullerton for hosting me and your endless knowledge on all things neutron stars. Thank you also to the friends I've made along the way; Melissa and Stefano especially, I have appreciated your friendship and advice so much. I'd also like to thank my new colleagues at Utrecht University for being so welcoming and Dr. Sarah Caudill for taking me on as

a post-doc.

I owe many thanks to Dr. Leigh Smith, a genuine inspiration and the best friend a girl could wish to find. You deserve all the awards (and nominations). I'd like to thank Ross for being an unexpected source of happiness during my PhD; for many adventures and endless chat, as well as keeping me fed and watered during the final stages of writing. In particular to Heather, Roslyn and Kathryn; the original women in science, who I could not have made it out of undergrad, never mind finish this thesis, without. Thank you for many years of unlimited comfort in the form of cosy dinners, musicals, Buff Club and sea swims, to name a few. No matter where, I cherish your love and support so much.

Finally, I'd like to thank my family: firstly my grandparents for inspiring me in so many ways, but also for encouraging me to stay interested and enthusiastic. Sam, yes there is now (almost) a doctor on board. And finally to my parents, firstly to prove that this thesis does now actually exist but mostly for my writing retreat and support all throughout my many, many years in education, encouraging me to keep going and do what feels right. I said I'd acknowledge you in my first book in 2008, who knew it would be this.

List of Figures

List of Tables

Chapter 1

Introduction to gravitational waves and neutron stars

We begin introducing NSs by deriving their equations of structure in section 1.1 within the context of General Relativity before outlining our current best understanding of NS structure in section 1.1.1. Primarily, we focus on what we can learn from GW observations of NS mergers; to do so, we introduce GWs in section 1.2 alongside detection methods (section 1.2.2) and data modelling (section 1.2.3). We focus on the parameters of interest with respect to observations of GWs from binary NS mergers in section 1.3 to set the scene for how we can use gravitational wave data to better understand the NS equation of state. We introduce section 1.4 with discussion of the first measurement of GWs from a BNS merger, GW170817 (section 1.4.1). We then summarise observations of NSs made over the decades since their initial discovery (section 1.4.2) and discuss how well we can currently determine NS parameters. Finally, we briefly introduce Bayes inference in section 1.5 and discuss how this is applied to GW PE (section 1.5.1), laying the groundwork for techniques and theory introduced in later chapters.

1.1 An introduction to neutron stars

Our scientific understanding of *neutrons*, let alone NSs, is relatively modern, with the announcement of the discovery of neutrons being made by Chadwick less than 100 years ago in 1932 [5]. Soon after, Baade and Zwicky introduced the concept of *supernovae* [6], a proposed end state of stellar evolution where a dying star collapses. A supernova can result in the complete collapse of material into a singularity or a partial collapse, leaving material behind. In an attempt to account for the energy release in the process, Baade and Zwicky 'tentatively' suggested that a supernova can represent the transition of an ordinary star into a NS. While their discussion of what constitutes a NS is brief, Landau likewise anticipated the existence of dense stars which look like giant nuclei [7] in a similar period. We now know that these stars do indeed exist, and are ultra-dense stellar remnants, with masses on the order of 1 $M_\odot$ and radii of $\sim$ 12 km [8]. NSs are born from the compressed iron core of a type-II supernova, where the material from a massive star collapses and rebounds from the core, leaving behind a (potentially) rapidly spinning, highly magnetised NS in the expelled material. Much of the early discussion surrounding the proposed existence of NSs sought to understand late stages of stellar evolution and how stars fuel themselves [9] - early intuition proved to be correct in postulating what happens when they can no longer do so.

Later in the decade, the discussion of stellar energy sources continued and gained more clarity [10]. Meanwhile, Einstein's Theory of General Relativity [11] was gaining traction as an acceptable theory of the description of gravity. It has since proved to be both locally in agreement with Newtonian dynamics and a valid descriptor of the wider universe [12] through decades of experiment. We refer the reader to [13, 14] for a deeper theoretical background into General Relativity (GR). NSs are interesting laboratories in terms of tests of GR; the highly compacted nuclear matter within a NS curves spacetime dramatically [15] and it is unclear how matter behaves under strong forces like those of stellar collapse, which cannot be replicated on Earth (for reference, the highest pressures that can be achieved terrestrially are those from a diamond anvil cell [16] or shock wave experiments [17], at order 10^{12} dyn cm^{-2} and

at order 10^{15} dyn cm^{-2} respectively). In 1939, Oppenheimer and Volkoff [18], and independently Tolman [19], applied Einstein's theory to understanding the equations of structure of the NS. These equations of structure are the Tolman-Oppenheimer-Volkoff (TOV) equations of hydrostatic equilibrium, which aim to determine a NS's mass and radius given a description of the NS's internal structure assuming GR holds.

We begin our derivation of the TOV equations by introducing Einstein's field equation which describes the generation of curvature in spacetime by mass-energy:

$$\mathbf{G} = 8\pi\mathbf{T}, \tag{1.1}$$

where $\mathbf{G}$ is the Einstein tensor, characterising gravity and is a descriptor of the geometry of spacetime, and $\mathbf{T}$ is the stress-energy tensor, detailing how matter and energy in spacetime affect its curvature [13]. In order to use equation 1.1 to understand the structure of a NS in GR, we need a stress-energy tensor, $\mathbf{T}$, which describes the matter within the NS and how much mass-energy is contained within a unit volume [13]. We first construct the stress-energy tensor and we make some assumptions to simplify the problem: firstly, we assume the NS is static and is non-rotating, spherically symmetric and in its entirety a perfect fluid. A perfect fluid is one which experiences no shear stress or anisotropic pressure. We define its 4-velocity as u^μ, where for a static star $u^\mu = (1, 0, 0, 0)$, and with the density of mass-energy, ρ, and isotropic pressure, P, we define the stress energy tensor

$$T^{\mu\nu} = (\rho + P)u^\mu u^\nu + Pg^{\mu\nu}, \tag{1.2}$$

where similarly $u^\mu u^\nu g_{\mu\nu} = 1$ for the static fluid, given a metric $g_{\mu\nu}$ (discussed later in the derivation). The expression for the four velocity is symmetric and diagonal as $T^{\mu\nu} = 0$ if $\mu \neq \nu$. Local conservation of momentum further requires $T^{\mu\nu}_{;\nu} = 0$.

The field equations above also include the relationship of pressure and density with the star's baryon number density and temperature: $P(n, T)$ and $\rho(n, T)$ respectively. While NSs have temperatures of millions of Kelvin during the majority of their life span [20], they are cold with respect to nuclear standards. We demonstrate why by treating NS matter as an ideal Fermi neutron gas (most baryons in a NS are neutrons,

as opposed to ~50% as in ordinary symmetric matter) and estimating the Fermi temperature. This treatment applies the Pauli exclusion principle which requires that a system of fermions provides an intrinsic pressure, without any thermal contribution. We calculate the Fermi energy, E_F for a NS at densities near nuclear saturation density i.e. $\rho_{\text{sat}} = 2.7 \times 10^{14}$ g cm^{-3}, the density of the nucleus of an atom, via [8]

$$E_F = \frac{\hbar^2}{2m_n}\left(\frac{3\pi^2 N}{V}\right)^{2/3}, \tag{1.3}$$

where m_n is the neutron mass and N/V is the number density of neutrons per unit volume. We expect all stable NSs to reach nuclear saturation density within the outermost 10% of their radius, and then exceed it. The Fermi temperature of pure neutron matter at this density is found to be $T_{\text{Fermi}} \sim 3 \times 10^{11}$ K [13], a temperature greater than the temperature NSs reach within the first few seconds after birth via emission of neutrinos. Thousands of years into their existence, NSs continue to cool further to temperatures of $\sim 10^8$ K [20]. Since the temperature of a NS older than a few seconds has $T \ll T_{\text{F}}$, a NS can be treated as having $T = 0$, or considered 'cold'.

Returning to equation 1.2, we need relationships for the composition of the star, $P(n)$ and $\rho(n)$, and a metric tensor, $g_{\mu\nu}$. The metric tensor, or solely 'metric', serves to describe the local structure of spacetime around the star which we assume to be flat. This is given by

$$ds^2 = -dt^2 + dr^2 + r^2 d\Omega^2, \tag{1.4}$$

where ds is the distance interval in spacetime, t is time, r is the radial distance and

$$d\Omega^2 = d\theta^2 + \sin^2\theta \; d\phi^2, \tag{1.5}$$

where θ is the polar angle and ϕ is the azimuthal angle [13]. This standard metric for flat spacetime is modified to allow for curvature induced by the gravitational influence of a (neutron) star. Therefore, the simple flat metric of equation 1.4 becomes

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 \tag{1.6}$$

where M is the total mass of the star and we are in the regime where $c = G = 1$. This applies for radial co-ordinates *external* to the star, i.e. $r > R$ where R is total stellar

radius. This is the Schwarzschild metric for describing the gravitational field around a spherical star.

The expressions for $P(n)$ and $\rho(n)$ in equation 1.2 can be described by the *equation of state* of NS matter, a relationship between the pressure and density as a function of the baryon number density (and temperature, but we apply $T = 0$). One can describe the equation of state given a full description of the microscopic matter within the star or a phenomenological model, or anywhere in between. A full discussion of possible equation of state models is given in chapter 2 section 2.1. By incorporating all knowledge together as in [13], one arrives at the Tolman-Oppenheimer-Volkoff equations of hydrostatic equilibrium:

$$\frac{dP}{dr} = \frac{-(\rho + P)(m(r) + 4\pi r^3 P)}{r(r - 2m(r))}, \tag{1.7}$$

$$\frac{dm}{dr} = 4\pi r^2 \rho. \tag{1.8}$$

In the above, $m(r)$ is the mass of the NS contained within radius r where $m(r = 0) = 0$. The only requirement in solving these equations is, besides an initial assumption of abiding by GR, the equation of state. The energy density contained within the star is a source of gravity and the pressure of the material prevents the collapse of the star to a black hole (BH). Therefore, the TOV equations encode the structure of a NS in GR.

The solutions to the TOV equations are found in the limit where $P(R) = 0$, where R is the total radius of the star as $P \to 0$. At total radius R, we can determine the total enclosed mass $M = m(R)$. Solutions over the range of pressure $[P(R), P(R = 0)]$ determine the $M(R)$ relationship of the NS, a direct analogue to the equation of state [21]. Therefore, through the TOV equations, one can relate an equation of state of high density nuclear matter to observable macroscopic properties of NSs and vice versa; from observation of NS properties one can place constraints on viable equations of state of NS matter. The TOV equations are the route from observation to theory in the case of NSs in GR.

Despite assumptions made to simplify the derivation and therefore calculation of

the TOV equations, solving the TOV equations to obtain the macroscopic parameters of total mass M and radius R of the NS is still difficult. In chapter 2, we discuss methods which use the TOV equations to obtain the macroscopic parameters of a NS and the inverse: inferring the underlying structure of an observed NS by integrating equations 1.7 and 1.8 or otherwise.

By deriving equations 1.7 and 1.8, Oppenheimer and Volkoff stated that a NS could not have a stable mass greater than 0.7 $M_\odot$ and assumed that it was 'unlikely that static neutron cores can play any great part in stellar evolution' [18]. We are now aware, however, that these two statements do not hold. Firstly, Tolman, Oppenheimer and Volkoff neglected to account for the effects of nuclear matter at such high densities [22]; this is the primary discrepancy between their stable mass limit and the observations of NSs that have been made. The influence of nuclear theory and experiment on understanding the composition of NS matter is discussed more in chapter 2. Secondly, not only have observations of NSs with $M \geq 2$ $M_\odot$ been made [23], but we also now know that NSs are the remnant of a type-II supernova, a late stage of stellar evolution [24], and have made multiple observations of NSs within a supernova remnant (see [25] for a recent observation). We next summarise our current understanding of NS structure but discuss modern EM and GW observations further in section 1.4.

1.1.1 The structure of a neutron star

Our current best understanding of a NS is presented in figure 1.1.1: we describe an ultra-dense stellar body with thin outer layers composed of a stiff outer and inner crust, where nuclear matter gradually increases in density as we move inwards into the star before the high density supranuclear core, which has density greater than nuclear saturation density ρ_{sat}. We discuss the composition of the NS layer-wise.

The density of the outer crust of a NS spans the range of 10^4 g cm^{-3} to 10^{11} g cm^{-3} [26] and consists of a lattice of neutron-rich ions surrounded by an electron gas; this enforces charge neutrality. Nucleons cluster to form lattice structures of ^{56}Fe in the outermost and lowest density regions, and as density increases moving deeper into the crust, even heavier nuclei become the most prominent and favourable. As density

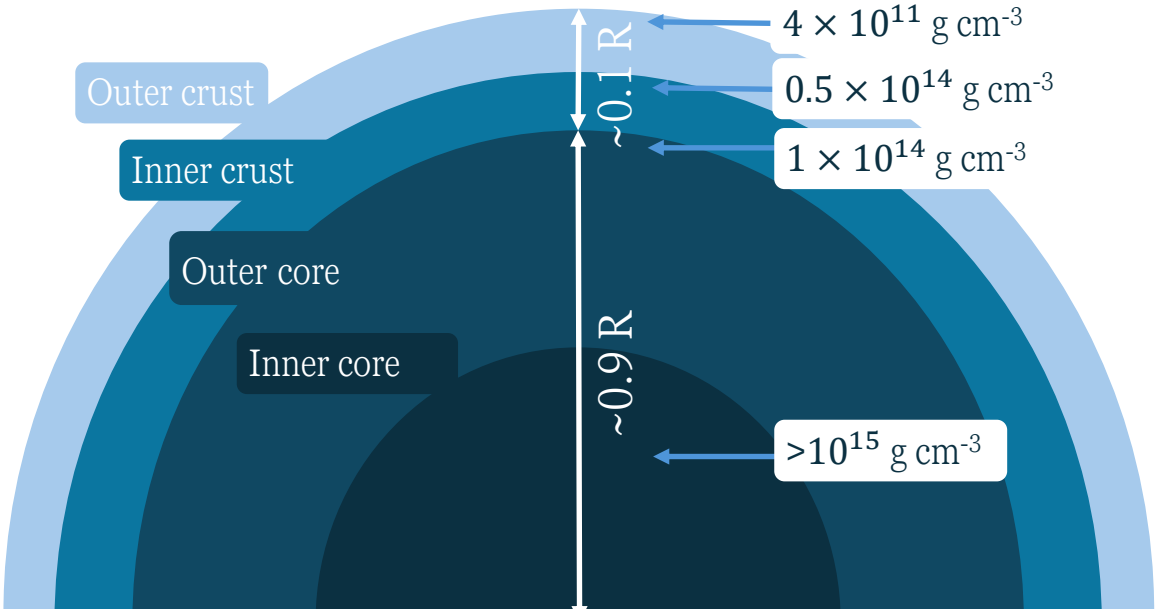


Figure 1.1.1: Illustration of NS shell-like structure. From light to dark blue, in layers of increasing density, are the outer crust, inner crust, outer core and inner core. The two crust layers occupy around 10% of the total radius, R, of the NS. The outer crust has densities less than nuclear saturation density, which is surpassed near the crust-core transition density as density increases further still towards the inner core.

increases, the nuclei also become more neutron-rich [27].

The inner crust follows on as a distinct region beginning at the neutron-drip density of around 10^{11} g cm^{-3}, which signifies the point at which nuclei can no longer retain neutrons and they 'drip' from heavy nuclei [28]. The inner crust therefore consists of a Coulomb lattice but now surrounded by a fluid of degenerate neutrons and electrons. At the most dense region of the inner crust, closest to the core of the NS, the separation between individual nuclei becomes comparable to the radii of the nuclei themselves [29]. Nuclei are thought to find it energetically favourable to rearrange themselves from the stable spherical structure we are familiar with into 'nuclear pasta'; structures which emulate that of pasta shapes (e.g. cylindrical resembling spaghetti or flat, broad sheets resembling lasagne sheets) [30, 31]. This pasta layer may have a thickness of ~ 100 m but due to its density may be responsible for up to half of the total mass of the NS crust [29].

The entire crust (inner plus outer) constitutes around 10% of the total NS radius,

but contributes to only around 1% of the total NS mass [32]. The large contribution of the core of the NS to the overall mass and radius, observation and measurement of the parameters of NSs gives us insight into the behaviour of matter at the extreme densities within the core of a NS. The involvement of the crust in determining the macroscopic properties of the NS is subdominant in comparison to the core [33]. Despite this, the crust of a NS is interesting observationally with respect to phenomena which occur on the surface of the star, for example thermal x-ray emission from the crust [34], starquakes [35] and millimetre-high mountains on the surface which may be the source of gravitational radiation [36]. The relative influence of the crust also demonstrates just how dense the core of a NS can become.

The crust-core transition density is around nuclear saturation density, $\rho_{\text{sat}} = 2.7 \times 10^{14}$ g cm^{-3} [34]. At this density, the uniform nature of matter is restored but instead as a fluid of free neutrons, protons and electrons where the vast majority of nucleons are of course neutrons. These densities cannot be replicated through terrestrial experiment (see chapter 2 for discussion of terrestrial high density nuclear matter experiments) and so astrophysical observation is the primary source of information.

Above approximately twice the nuclear saturation density, we reach the inner core of the NS, where the composition is currently unknown. Particles beyond that of baryonic matter, with more energetically favourable arrangements of quarks at such high densities, for example hyperons or pion condensates - may appear, however the nature of this matter is uncertain. The observation of NSs with mass > 2 M$_\odot$ has prompted further discussion into the composition of NS cores; to support such a high mass NS, where the majority of the mass is contributed by the ultra-dense core of the star, what composition of matter can support itself against the intense gravitational force?

As previously introduced, the equation of state is the relationship that describes the evolution of pressure with density inside the NS. Through observation of macroscopic parameters of NSs, via methods which will be introduced in section 1.4, we can employ the TOV equations to relate macroscopic observables like mass and radius to the

equation of state (this method is introduced in chapter 2 section 2.2.1). Through measurement of macroscopic parameters, we can then *infer* the equation of state of NS matter, and inform our current understanding of high density nuclear matter, and therefore our understanding of the structure of the NS. As such, observation is essential to further our knowledge of NS structure and the equation of state, and our primary method of observation discussed in this work are GWs.

1.2 Gravitational waves

The field of GW research has grown considerably since the first detection of GWs from a binary black hole (BBH) merger in 2015 [37]. Since then hundreds of GWs from binary BH mergers have been made, including both those presented in the third and fourth LIGO Scientific, Virgo and KAGRA (LVK) Collaboration Gravitational-Wave Transient Catalogs (GWTC-3 and GWTC-4) [38, 39]. The network of ground-based detectors currently consists of two LIGO detectors in North America, Virgo in Italy and KAGRA in Japan. More details of these detectors are discussed in 1.2.2. All observations thus far have been that of a compact binary coalescence (CBC); an observation of the merger of two compact objects that produces a GW signal which we have a well-understood theoretical model to compare to. These include the observation of GWs from a BBH mergers but also BNS mergers, with substantially lower masses and matter in the system to now consider, as opposed to a singularity with no matter present. The mass in a BNS system alters the inspiral of the two components in comparison to a BBH system of the same mass, and so has to be treated differently. GWTC-3 and GWTC-4 discuss the range of official GW observations made to-date, only two of which have been confident BNS mergers. We discuss the specific details of BNS mergers and the measurement of the parameters of the system in section 1.3. To date, no detection has been made of GWs from a non-CBC origin; these would include modelled and long duration signals (for example, continuous gravitational waves (CWs) from spinning NSs with crust asymmetries), unmodelled transient signals (for example, 'burst' GWs from supernovae) or unmodelled and long duration signals (for example, a background of GW radiation, akin to the cosmic microwave background radiation

(CMBR)). It is hoped that as current detectors gradually improve in sensitivity and the next-generation of observatories gain traction, a more diverse catalogue of observations will be available. These observations would move towards further confirming or presenting deviations from GR or, more prevalent to this thesis, offering more observations of NSs, whether via CBCs, CWs or something else entirely.

1.2.1 General Relativity and gravitational waves

We have previously introduced Einstein's GR and how it allows us to determine equations of structure of ultra dense spherically symmetric stars. We now demonstrate how GR predicts that the asymmetric acceleration of massive objects in spacetime cause perturbations through the fabric of spacetime, propagating as GWs. We begin again from Einstein's field equation (equation 1.1) which, as previously introduced, details how spacetime guides the movement of matter and how mass dictates the curvature of spacetime. The LHS of equation 1.1 defines the Einstein tensor, $\mathbf{G}$, given equivalently by

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}, \tag{1.9}$$

where $R_{\mu\nu}$ is the Ricci tensor, describing the curvature of spacetime and where R is the Ricci scalar, a further contracted version of $R_{\mu\nu}$ via $R = g_{\mu\nu}R^{\mu\nu}$. The Ricci tensor is a function of the metric and it's derivatives. We refer the reader to [13] for details. The metric tensor $g_{\mu\nu}$ can be expressed as a flat Minkowski spacetime metric $\eta_{\mu\nu} = (-1, 1, 1, 1)$ with a small perturbation:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1.10}$$

where 'small' implies that the perturbation $|h_{\mu\nu}| \ll 1$. By asserting the fact we are in an empty spacetime with no mass or energy, we let $T_{\mu\nu} = 0$ and through solutions following [13], we arrive at the simplified form of the linearised Einstein field equations,

$$\left(\frac{\partial^2}{\partial t^2} + \nabla^2\right)\bar{h}_{\mu\nu} = 0, \tag{1.11}$$

where ∇^2 is the Laplace operator [40] and $\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h$ is the trace-reversed metric perturbation with $h = \eta^{\mu\nu}h_{\mu\nu}$ being the trace. The solution to this equation

takes the form of a plane wave:

$$\bar{h}_{\mu\nu} = A_{\mu\nu} \exp(ik_\alpha x^\alpha). \tag{1.12}$$

This result informs us that GWs are transverse waves which propagate through space with polarisation $A_{\mu\nu}$ and wave tensor k_α at the speed of light. There exist two fundamental polarisations of GWs; plus, h_+, and cross, $h_\times$ which are functions of the two possible degrees of freedom of the polarisation tensor, A_{xx} or A_{xy}, in the transverse-traceless gauge where $\bar{h}_{\mu\nu} = h_{\mu\nu}$. Each of these polarisations, for GWs propagating in the $z-$direction, alter the orientation of free particles in the $x - y$ plane similarly, but at an angle to one another. A GW propagating in the $z-$direction will generally deform free particles in the $x - y$ plane as a linear combination of both polarisations. The magnitude of disruption caused by a GW, h, is the *strain*, describing the relative change in proper distance between two points in spacetime as a GW passes through.

We noted at the beginning of this section that GWs were the result of the asymmetric acceleration of massive objects in spacetime. When we have a source at large distance r, we find the quadrupole formula:

$$h_{ij} = \frac{2}{r}\ddot{Q}_{ij}(t - r) \tag{1.13}$$

where $t = x^0$ and $Q_{ij} = \int \rho(x_i x_j - \frac{1}{3}r^2\delta_{ij})d^3x$ is the reduced quadrupole moment [41]. This tells us that GWs are produced by a 'time-varying mass quadrupole moment' [42]; we require an asymmetry in the mass distribution to produce GWs. GR prohibits both monopole and dipole gravitational radiation due to mass and momentum conservation, respectively. Quadrupole radiation is the first detectable moment. We therefore need non-spherical mass distributions and non-axisymmetric motion to produce gravitational radiation. Astrophysically, this means we can produce GWs through the inspiral and eventual merger of compact objects, such as black holes (BHs) or NSs. We will cover methods of producing GWs from NSs in particular in section 1.3.1.

1.2.2 Ground-based gravitational wave detectors

All currently operational ground-based GW detectors are laser interferometers [43, 44, 45]. While the individual details of these detectors vary, their designs are all based on that of the Michelson interferometer. A Michelson laser interferometer is an instrument composed of two perpendicular arms, down which light travels and reflects at the end to then return and be recombined. The interference pattern of the recombined light informs the observer of the path the light has followed in each of the arms. In a laser interferometer for GW detection, each of the perpendicular arms are optical cavities [46]. A laser input is split in two by a beamsplitter [47] such that laser light travels down each of the perpendicular arms and is circulated within the cavities so to build power [48]. The mirrors which reflect the light at either end of the cavity are highly reflective [49] and act as the gravitational test masses [47]. Upon the output, the laser light is recombined, either constructively or destructively. When a GW passes through the interferometer, the relative lengths of each of the arms changes on a minute scale. It is the relative change in arm length which determines whether the output signal constructively or destructively interferes and therefore whether a GW has passed through or not.

If we suppose a GW signal with amplitude h is arriving at Earth to change the relative arm lengths, L_x and L_y, of the 'x' and 'y' arms of an interferometer where $L_x = L_y \equiv L$ then by [50, 37], the change in arm length ΔL is approximated by

$$\Delta L \approx hL. \tag{1.14}$$

Therefore, increasing the arm length increases the measurable difference in arm length for a given strain signal. The magnitude of the strain signal for a BBH signal of two 30 $M_\odot$ BHs, one of our aforementioned CBC signals, merging at a distance of 400 Mpc away is of the order of 10^{-21} m (comparable to the first BBH merger, GW15091 [37]). It is this reason that typical arm lengths of operational GW detectors are on the order of kilometres: if GW strain h is on the scale of 10^{-21} m by the time it reaches observers on Earth and the interferometer arm length is 4 km for the Laser Interferometer Gravitational-wave Observatory (LIGO)-like detectors, then the measurable difference in arm length given equation 1.2.2 is $\Delta L \simeq 10^{-18}$, the limit

of sensitivity of the advanced detector network [51]. The arm length is also chosen with respect to the frequency response of the detectors; the length of the detector arm is a defined fraction of the wavelength of expected GWs arriving at the detector for maximal distortion measure [52] of signals from CBC sources. This limits the scope of current ground-based interferometric detectors to observation of a number of astrophysical sources.

There are currently 4 ground-based interferometric GW detectors which have been in active observing mode in the current (fourth) observing run. The two LIGO detectors in North America (Livingston, Louisiana and Hanford, Washington [43]) both have 4 km arms [47] and made their first observation of a GW signal in 2015: two merging BHs of masses 36^{+5}_{-4} $M_{\odot}$ and 29^{+4}_{-4} $M_{\odot}$ at a distance of approximately 400 Mpc away, named GW150914 [37]. During the second observing run, the Virgo detector [44] near Cascina, Italy joined the network just over two weeks prior to the first observation of GWs from the merger of two NSs [53], titled GW170817 [54]. More recently, the KAGRA detector, based at the Kamioka observatory in Japan [45], joined the advanced interferometric detector network for the beginning of the fourth observing run [55]. Like Virgo, KAGRA also has a smaller size than both Advanced LIGO detectors with arms of 3 km but has been built into a mountain to reduce the effect of seismic noise sources [45].

There are planned additions to the current advanced detector network [56, 57] in the coming years, prior to the planned next generation of detectors [58, 59] (briefly discussed in chapter 2 section 2.3). Expanding the network geographically increases the fraction of the total sky which is observed by a GW interferometer at a given time, but also increases confidence in and the SNR of a detection; a signal is more likely to be accepted as real if it is present in the measured strain of more than one detector.

1.2.3 Modelling gravitational wave data

We can express the output signal from a GW detector, $d(t)$, in terms of the GW strain signal $h(t)$ plus detector noise, $n(t)$:

$$d(t) = h(t) + n(t). \tag{1.15}$$

Once we have observed data $d(t)$, we wish to determine the confidence of the presence of a signal $h(t)$ within noise $n(t)$, and infer the parameters of the source which produced the signal. GW signal strain, h_t, can be expressed in either the time or frequency domain and is a function of the parameters θ of the source, such that we can express the strain instead as $h(t, \theta)$. We split parameters θ into two categories: intrinsic and extrinsic. Extrinsic parameters are those which change how the detector perceives the signal and describe the positioning and orientation of the source relative to the interferometer [60]. Examples of these parameters are sky location, distance to the source and time of coalescence. Intrinsic parameters, however, are parameters which are inherent to the source itself. These parameters influence how the waveform evolves, e.g. masses and spins. We can use our theoretical knowledge of what a GW signal should look like for a given set of parameters θ using a waveform approximant, built using computationally intensive numerical relativity simulations [61, 62] or instead using phenomenological methods which model the amplitude and the phase of the signal independently in the frequency domain [63]. See [64] for a review of current waveform building methods. Different waveform approximants express the GW waveform produced by a given source using different levels of complexity and including different features, for example precession of the orbital plane of the binary (e.g. [65]). In terms of binary NS waveforms, we are interested in waveform models which allow for a measurable tidal disruption of each of the components in the binary close to merger. We discuss the specifics of parameters measured from GWs from a BNS merger in section 1.3.1.

Given a bank of example waveforms spanning the parameter space of potentially observable binary systems, one can perform *matched filtering* (see [66] and references therein), the method of matching a waveform approximant to observed strain data to find the best-fit waveform via search over waveform parameters. All published GW

events to date have been identified via matched filtering as a search method [64] (with the exception of complementary analyses to existing matched filter results), where the data is constantly scanned for potential template-data matches. Other search methods do exist (for example [67]), including those using machine learning methods (see chapter 3 section 3.4 for discussion). With the bank of template signals of a variety of CBC events, we measure the signal-to-noise ratio (SNR) of the template waveforms given the signal within the noise in the detector as a function of frequency. The template waveform which returns the largest matched-filter SNR beyond a given threshold - defined so to mitigate the influence of noise artefacts - is the optimal template. Given the threshold has been surpassed, one can claim a detection has been made of a CBC event with parameters of that of the template waveform. Once we can confidently claim a detection has been made, one can then begin to determine the underlying parameters of the signal more precisely.

1.3 Gravitational waves from neutron stars

We made an assumption that a NS is static and spherically symmetric in section 1.1 in order to derive the TOV equations of NS structure. In this regime, the exterior spacetime is given by the Schwarzschild metric, which describes the spacetime around a spherically symmetric stellar body [68]. However, if the NS is in a perturbing tidal field induced by a companion NS, the star will have its shape distorted. The deformed star's external spacetime and gravitational field is thus affected, leaving the star in a now spatially inhomogeneous external field, with both stars tidally distorting one another. The existing quadrupole moment due to binary motion is modified, accelerating the production of gravitational radiation.

The tidal deformability of a NS, λ, is a measure of the degree of deformation of the star when it is in an external tidal field: an external perturbing gravitational field which induces a tidal force within the NS [69]. The tidal deformability measures the ratio of the induced quadrupole moment, Q_{ij}, to the perturbing tidal field, ε_{ij}:

$$\lambda \equiv -\frac{Q_{ij}}{\varepsilon_{ij}}. \tag{1.16}$$

The perturbing tidal field ε_{ij} has units 1/length^2 and Q_{ij} has units length^3 [68] and so λ can be considered as a ratio with units of length^5. This suggests that λ is a sensitive function of the NS radius, R, where $\lambda \propto \kappa R^5$. The constant κ relates λ and R via the tidal Love number k_2 for quadrupolar deformation [70, 71] through

$$\lambda = \frac{2}{3G} k_2 R^5. \tag{1.17}$$

The tidal Love number is a dimensionless constant which depends on the equation of state of the NS with values ranging between $0.2 - 0.3$ [68]. Alternatively, we can express the tidal deformability as a dimensionless quantity instead, where we introduce the compactness, $C = M/R$, a measure of how much mass is contained within a unit of radius:

$$\Lambda \equiv \frac{\lambda}{M^5} = \frac{2}{3} k_2 C^{-5}. \tag{1.18}$$

We will primarily use and infer the *dimensionless* tidal deformability in future discussion but will use the terms interchangeably.

Observation of GWs from NSs allows us to measure this tidal parameter as it impacts the phase of the GW. Currently, the most prevalent (and only) source of GWs from NSs are those from binary NS or neutron star-black hole (NSBH) mergers, however there are other theorised sources of GWs from NSs. These include, but are not limited to:

- *Continuous gravitational waves:* Emission of GWs from isolated NSs is predicted to be produced from the rotation of stars with non-axisymmetric matter composition, which results in a time-varying mass-quadrupole [72], a requirement for GW production (as introduced in section 1.2). The non-spherical matter includes NSs which have features on the surface, known as 'mountains' which are millimetres high [73]. While this variety of GWs are yet to be observed, such a signal is expected to be a continuous sinusoid, slowly decaying over time as the pulsar spins down. There are a few particular targets in the search for CWs, for example the Crab and Vela pulsars [74, 75, 76]. Through observation of their EM emission, one can estimate the frequency of GWs they would emit.

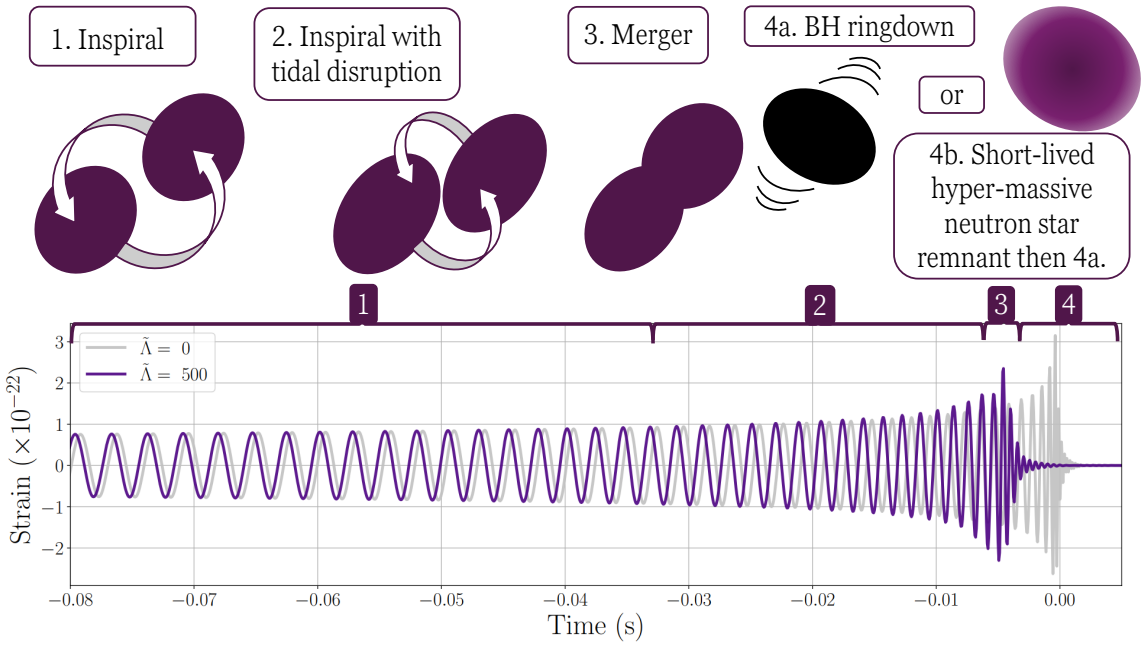


Figure 1.3.1: The inspiral waveform of two 1.4 $M_\odot$ NSs with $\Lambda_1 = \Lambda_2 = 500$ in purple and $\Lambda_1 = \Lambda_2 = 0$ in grey are plotted in the time domain. The waveform was made using the `IMRPhenomD_NRTidalv2` waveform approximant. Four stages of the evolving inspiral waveform are illustrated above, with two potential outcomes of the post-merger phase included.

- *GWs from supernovae:* The creation of a NS via the collapse of a massive star onto its core during a supernova, as introduced in section 1.1, releases significant amounts of energy. While most of this energy is released in the form of neutrinos [77], and our current understanding of supernovae stems from our observation of EM radiation from NSs, we additionally expect to observe the emission of a 'burst' of GWs from such an event [78]. Observation of GWs would allow us to explore the behaviour of ultra-dense matter at the core of the supernova, during the course of the compression and release, before the ultimate production of a NS.

1.3.1 Gravitational waves from binary neutron star mergers

In a NS binary system, each component of the merger is subject to the gravitational field induced by its companion star. The induced quadrupole moment, a result of being in the perturbing gravitational field and having its shape distorted, as introduced above, affects the binding energy of the system. The life of a NS binary system is

detailed in figure 1.3.1: we present the time-domain gravitational waveform of a binary NS system of two 1.4 $M_\odot$ NSs with a combined tidal deformability of $\Lambda_1 = \Lambda_2 = 500$. Four stages of the merger are highlighted and illustrated: point-particle-like inspiral, late-stage inspiral with tidal disruption, merger and ringdown. Initially, the two stars are slowly inspiralling and emitting GWs; a result of the intrinsic quadrupole moment from being in a binary system. At this stage, the measured tidal interaction is negligible; the system could equivalently consist of two BHs (no matter present in the binary system, therefore $\Lambda = 0$). As the inspiral progresses, binding energy is lost from the system in order to instead deform the NSs and the stars fall closer together [79], with a non-negligible and measurable tidal interaction in the GW phase. This is evident in figure 1.3.1 by comparing the GW evolution of two NSs with $\Lambda_1 = \Lambda_2 = 500$ to the waveform of an equivalent system (i.e. same masses of components) instead composed of BHs with $\Lambda_1 = \Lambda_2 = 0$. In the case of two BHs, there is no matter in the system to be tidally perturbed and so there can be no tidal deformability parameter. As a result, the waveform proceeds towards merger due to the emission of gravitational radiation due to the quadrupole moment induced solely by being in a binary system. The result of the eventual merger of the two NSs - whether a NS or BH - depends on the masses of the two components but in the immediate aftermath, the remnant, if a BH, emits gravitational radiation in the ringdown phase, until settling (noted as 4a in figure 1.3.1). Alternatively, the two NSs may coalesce to form a hyper-massive NS, which exists for $\ll 1$ second (noted as 4b in figure 1.3.1) before collapsing to a BH, and similarly ringing down (returning to 4a). The GWs emitted by the hyper-massive NS depend on the mass and equation of state of the short-lived star.

The tidal effects directly impact the phase of the GW signal in the late stages of inspiral. In the frequency domain, the GW phase is [68, 80, 81] expressed as a post-Newtonian (PN) expansion of $x = \left(\frac{\pi G M f}{c^3}\right)^{2/3}$ via

$$\Psi(f) = 2\pi f t_c + \phi_c - \frac{\pi}{4} + \frac{3}{128\eta x^5}\left[1 + \psi_{PP-PN}(x,\eta) + \psi_{\text{tidal}}(x,\eta,\Lambda_1,\Lambda_2)\right] \tag{1.19}$$

where f is the frequency of the GW, t_c is the time of coalescence and ϕ_c is the phase of the coalescence. The symmetric mass ratio is given by $\eta = \frac{m_1 m_2}{M^2}$, where $M = m_1 + m_2$ is the total mass of the binary and m_1 and m_2 are the masses of the

two component NSs (where convention states that $m_1 > m_2$). The second term in the square brackets $\psi_{PP-PN}(x, \eta)$ represents the terms that express the gravitational phase of a point-particle waveform i.e. that used to express the waveform of a binary BH merger. Waveforms of BNS systems are often expressed as that of a point-particle with some correction to account for tidal effects. The tidal correction to the phase ψ_{tidal} is given as [80]

$$\psi_{\text{tidal}}(x, \eta, \Lambda_1, \Lambda_2) = -\frac{39}{2}\tilde{\Lambda}x^5 + \left(-\frac{3115}{64}\tilde{\Lambda} + \frac{6595}{364}\sqrt{1-4\eta}\delta\tilde{\Lambda}\right)x^6. \tag{1.20}$$

Terms with x^N in this expansion are referred to as the N^{th} PN order; the first PN order - the leading term proportional to x which is contained within $\psi_{PP-PN}(x, \eta)$ - is the most influential on the measured phase of the gravitational radiation. The dimensionless combined tidal deformability, $\tilde{\Lambda}$ arrives at the 5^{th} PN order, much less prominent than the effect of the intrinsic time-varying quadrupole:

$$\tilde{\Lambda} = \frac{8}{13}\left[(1 + 7\eta - 31\eta^2)(\Lambda_1 + \Lambda_2) + \sqrt{1-4\eta}\ (1 + 9\eta - 11\eta^2)(\Lambda_1 - \Lambda_2)\right], \tag{1.21}$$

where $\Lambda_1 \equiv \lambda_1/m_1^5$ and $\Lambda_2 \equiv \lambda_2/m_2^5$ are the component dimensionless tidal deformabilities. The 6^{th} PN order includes $\delta\tilde{\Lambda}$, a parameter which quantifies the correction to the combined tidal deformability, given as:

$$\delta\tilde{\Lambda} = \frac{1}{2}\left[\sqrt{1-4\eta}\left(1 - \frac{13272}{1319}\eta + \frac{8944}{1319}\eta^2\right)(\Lambda_1 + \Lambda_2) + \left(1 - \frac{15910}{1319}\eta + \frac{32850}{1319}\eta^2 + \frac{3380}{1319}\eta^3\right)(\Lambda_1 - \Lambda_2)\right]. \tag{1.22}$$

The correction to the dimensionless combined tidal deformability enters at too high a PN order to affect the phase enough to be measured well [82] with the sensitivity of current ground-based GW detectors. Therefore, with only a measurement of $\tilde{\Lambda}$ it is not possible to detach the tidal deformability of the components $\Lambda_{1,2}$ from a single measurement. A measurement of $\delta\tilde{\Lambda}$ would break the degeneracy.

Figure 1.3.2 presents 4 different gravitational waveforms in the time domain and how their phase accumulates differently when the magnitude of combined dimensionless tidal deformability $\tilde{\Lambda}$ varies. In the early stages of inspiral (> 0.1 seconds before merger), all waveforms appear nearly identical. In later stages (< 0.1 seconds before

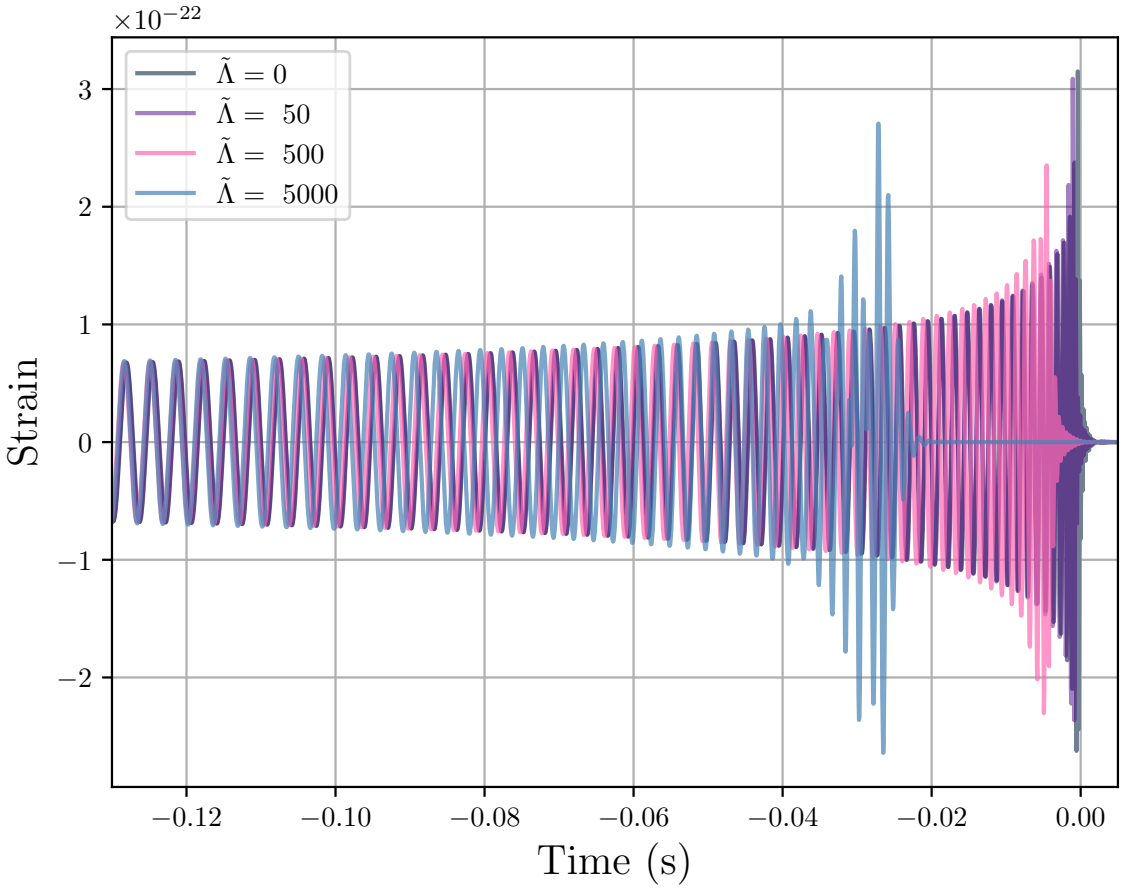


Figure 1.3.2: Four waveforms of a BNS system which consists of two 1.4 $M_\odot$ NSs with $\tilde{\Lambda} = 0, 50, 500, 5000$ made using the `IMRPhenomD_NRTidalv2` waveform approximant. The grey waveform representing the system with $\tilde{\Lambda} = 0$ is for comparison purposes: this would be the equivalent waveform for a system of two 1.4 $M_\odot$ BHs, which mergers at the canonical $t_c = 0$.

merger), the tidal effects begin to differentiate each of the waveforms causing them to evolve differently. For larger $\tilde{\Lambda}$, the phase evolves more rapidly and as such the binary inspirals more quickly and the two stars merge more quickly relative to a BNS system with a smaller $\tilde{\Lambda}$. A larger $\tilde{\Lambda}$ is the result of larger $\Lambda_{1,2}$; for a given mass of star, a larger Λ suggests that the star is less compact as per equation 1.18 and so the matter in the star is more extended. As such, the star is more easily disturbed by tidal forces, a quality which accelerates the inspiral. The implications of small or large Λ on the equation of state and composition of NS matter is discussed in chapter 2. In contrast, a small Λ suggests the star is more compact, with more compressed mass per unit radius. The star is less easily deformed and the matter requires more energy to be disturbed. NSs with small Λ are able to get closer together during the inspiral stage,

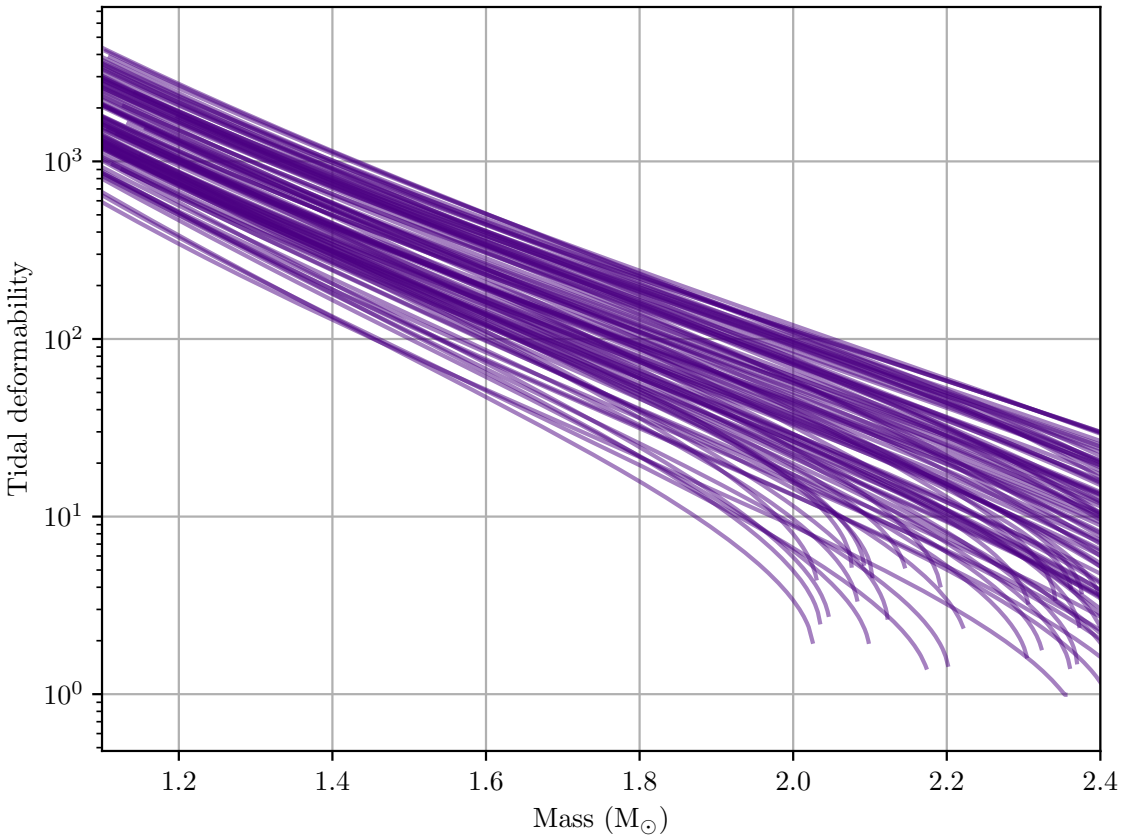


Figure 1.3.3: Relationship between mass and dimensionless tidal deformability for a variety of NS equations of state. 100 piecewise polytropic equations of state are presented in the $m - \Lambda$ plane within the range of known (observed) NS masses.

resulting in a longer waveform and a later merger of the two components.

For realistic NSs, i.e. $m_{1,2} \sim [1, 2]\ M_\odot$, with realistic internal structure (meaning the equation of state), tidal deformabilities are large. Figure 1.3.3 presents the relationship between mass and tidal deformability for a variety of piecewise polytropic NS equations of state, the details of which are covered in chapter 2 section 2.1. This particular model is widely accepted as a realistic phenomenological NS equation of state model in the literature. NSs belonging to each of these polytropic equations of state in $P - \rho$ would live on the equivalent $m - \Lambda$ curve in figure 1.3.3. The plot presents a variety of equations of state within the range of mass at which NSs have been observed, mainly as radio pulsars. For a given NS mass, e.g. 1.6 $M_\odot$, the range of possible tidal deformabilities for a range of equations of state span more than an order of magnitude. Our current best understanding of the NS equation of state means that

for a given mass measurement, the tidal interaction is still broadly unknown and spans a broad region of parameter space. It is key, therefore, to observe NSs and measure their tidal deformability parameters to be able to constrain the equation of state.

1.4 Observations of neutron stars

1.4.1 GW170817

The first observed GW event originating from a binary NS merger, GW170817 [54], provided a completely independent measurement of NS masses [83, 4] which complemented the existing NS binary mass measurements [84] from binary radio pulsars. The detection additionally allowed for the first ever measurement of the NS tidal deformability.

The estimated masses of the components of GW170817 are in the range of $[0.86, 2.26]$ $M_\odot$, in agreement with average masses of NSs [85] and below known masses of BHs [86]. However, the masses of the components alone does not eradicate the possibility that GW170817 is the merger of at least one black hole. We need further evidence of matter in the system to prove that they are NSs and not solar mass BHs. Proof of matter in the system is evident through either emission of EM radiation from the merger or non-zero measurement of tidal effects.

The initial measurement of the tidal parameters of GW170817 in [54] allowed Λ_1 and Λ_2 to vary independently with no prior assumption or constraint. This assumption equivalently allows for $\Lambda_{1,2} = 0$ if either of the components are instead BHs. This analysis resulted in constraint on $\tilde{\Lambda}$ of < 800 at 90% confidence, a result which has since been replicated [87, 83], corroborating the support for smaller $\tilde{\Lambda}$. The authors of [54] present different relationships for a variety of equations of state of NSs, ranging from very compact and high pressure NSs with small Λ to those more extended with larger Λ, alongside the tidal measurement in the $\Lambda_1 - \Lambda_2$ plane. They find their constraint in the $\Lambda_1 - \Lambda_2$ space prefers equations of state which prefer more compact stars, a result in agreement with radius measurements from x-ray measurement (see

section 1.4.2).

More complex inference of $\Lambda_{1,2}$ in [4] makes the assumption that the two components of GW170817 are indeed NSs which have the same equations of state. This assumption places stricter constraints on the properties of each of the components, restricting the prior to eliminate regions of parameter space which would be inconsistent with properties of a BNS merger (as we currently understand it). The implications of a more stringent analysis on the understanding of the nature of NS matter for GW170817 is discussed in chapter 2 section 2.2.1.

The short gamma-ray burst, GRB 170817A, detected 1.7 seconds after the time of coalescence of GW170817, further validated the source of gravitational radiation being a BNS merger [88, 89]. Alongside the near-coherent timing of these two signals, the improved localisation of GW170817 due to information from both LIGO detectors and Virgo meant that the sky map associated to the event was small: the GW event was localised to 28 deg^2. Therefore both GRB 170817A and GW170817 were able to be associated to the same host galaxy, NGC 4993, and the resultant kilonova of the event was found in multiple wavebands of EM radiation [90]. Assuming GRB 170817A was a product of a BNS merger, independent analysis was performed to investigate the properties of the merger [91, 92, 93]. These analyses have their own assumptions and model-dependencies which account for how the EM counterpart varies relative to the mass and composition of the component NSs. Observation of the kilonova and it's light curve, which decays in the days and weeks following the merger, informs us of the material ejected in the merger, the quantity of which is determined again by the properties of the component NSs (i.e. compactness and tidal deformability) [94].

To date, there have been two observations of GWs from BNS mergers: GW170817 (as discussed) and GW190425 [95]. We focus our discussion on GW170817 due to the greater confidence in its origin as a BNS merger, a title further assured by the EM counterpart and non-negligible tidal measurement. Additionally, GW170817 remained in the detector's sensitive band for > 1 minute and had a matched-filter SNR of 32.4 [96]. Comparatively, GW190425 had an SNR of 12.9 in LIGO Livingston and 2.5 in

Virgo (LIGO Hanford was not observing at the time), meaning it was only classified as a confident detection in one detector. This is detrimental to the sky map associated to the event, which extends over a large area of the sky. This makes EM follow up investigations much more difficult. Inference of the source properties yields a higher mass system than that of GW170817, but cannot rule out the possibility of the system consisting of at least one black hole [95]. As can be inferred from figure 1.3.3, a higher mass NS has a lower Λ and therefore, the parameter makes a smaller alteration to the waveform. Combined with the smaller SNR of the event, inference of the parameters of GW190425 is less informative than that of GW170817.

The two BNS merger events observed by the ground-based GW detector network have indirectly inferred the rate of galactic NS mergers to be less than expected, relative to pre-observing expected rates. By considering the binary pulsar population as determined by radio observation [97] and supernovae rates from population synthesis estimates [98], it was expected that BNS mergers would be the most prominent source of GWs. As the observing runs commenced, this rate evolved and was reassessed relative to the new observable volume of events in the universe due to change in sensitivity of the detector, as well as including the prior on the merger rate to include the observation of GW170817. The updated rate was therefore $[110, 3840]$ $\mathrm{Gpc}^{-3}\mathrm{y}^{-1}$ [53]. The addition of GW190524 as an independent event restricts the expected rate of BNS mergers further to 1090^{+1720}_{-800} $\mathrm{Gpc}^{-3}\mathrm{y}^{-1}$ [95]. The current rate remains consistent with estimated rates of the galactic NS population which are in known NS binaries, observed via pulsar measurements [99].

1.4.2 Modern electromagnetic observations of neutron stars

NSs were first discovered by Bell Burnell while a PhD student. Their identification of a regularly repeating radio signal was found to be a rapidly rotating NS [100] with jets of radiation emanating from it's magnetic poles, known as a *pulsar*. Since this first detection, radio observations of pulsars have formed the majority of NS observations to date [101]. After the first observations made of radio emission from NSs following Hewish et al. [100] came the identification of the NS-pulsar binary system PSR

B1913+16 by Hulse and Taylor [102]. The two bodies in this binary are indeed NSs, but only one body is emitting radiation as an observable pulsar. Both bodies, however, are orbiting one another and as a result, are emitting GWs continuously (the reasons why are discussed in section 1.3). Energy is thus lost through the form of gravitational radiation and as a result, the orbital period of the pulsar slowly decays [103]. Hulse and Taylor indeed observed a delay in the orbital phase of PSR B1913+16 and in 1993, won the Nobel Prize in Physics for this discovery [104].

Since these two fundamental discoveries of NS systems, observations have been made of thousands of NSs through a variety of media. These observations have ultimately led to a greater understanding of their internal structure and extreme surroundings. Growing numbers of observations of binary pulsar systems in the decade following the discovery made by Hulse and Taylor found that the precise determination of the pulsar spin frequency as well as understanding of relativistic orbital effects allows for measurement of component masses [105]. Through measurement of the Keplerian parameters of the binary system and accounting for relativistic effects of the radio signal passing through the gravitational potential well induced by the binary [106], the component masses can be uniquely determined via the Keplerian mass function [107]. These masses were found to be broadly consistent with a theorised universal mass of 1.4 $M_\odot$ [105]. This mass was expected to be 'canonical' given the nature of their supernova birth from the iron core of a massive star [84, 23], meaning all NS were expected strictly to have a mass of around 1.4 $M_\odot$.

However, in more recent years, more precise observational methods have allowed for improved mass determination [108, 109].Employing the method of measuring component masses given measurement of NS spin frequency and Keplerian orbital dynamics with a correction for strong field gravity, Martínez et al. [109] find potentially the smallest mass of NS to be measured: 1.174 ± 0.004 $M_\odot$. Identification of very small NS masses gives more information about the potential evolution of the NS and its progenitor [110]. For such a small mass NS to exist, it is possible that its progenitor was likewise smaller in mass due to being accreted by the companion NS, which underwent the supernova explosion prior. Therefore, the supernova collapse

involved less matter collapsing onto the iron core, and therefore less mass compressed into the soon-to-be NS.

The mass range of NSs is now understood to be broad, from the aforementioned lower limit set by Martínez et al. to an unknown but confidently > 2 $M_\odot$ upper limit, supported by recent observations of high-mass radio pulsars [23, 111, 112]. So far, the maximum NS mass is found to be that of 2.35 ± 0.17 $M_\odot$ [113], belonging to a NS accreting material from a low-mass ($< 1M_\odot$) companion. By measuring light curves of different wavelengths and fitting for parameters of the system relative to an underlying theoretical model following [114], the authors can then use the system binary parameters to determine component masses using the method outlined above. There are many factors which are at play in determining the upper limit of masses for NSs, primarily whether the NS is in a system where it is accreting matter from a companion. We will not discuss said systems, but rather acknowledge that these systems can support a stable NS of high mass. High mass NSs are interesting targets as it can allow us to place an upper limit on the maximum pressure supported before collapsing to a BH [85], and therefore place constraints on viable equations of state via the TOV equations.

Measurement of the radius of a NS is less straight-forward and is subject to assumptions which accrue large uncertainties [115]. Most ideal are simultaneous mass and radius measurements [116], first attempted via observation of x-rays from NSs in binary systems. Özel [117] applied the method of Van Paradijs [118] to determine the mass and radius of x-ray bursts from NSs in accreting binaries. The luminosity of such bursts depend on the radius of the explosive photosphere of the NS and its mass. However, the method used is prone to uncertainty which originates from the lack of understanding of the NS atmosphere, and model uncertainties impact the quality of the measurement [119].

Detection of NSs in a low-mass x-ray binary (LMXB) allow for observation of thermal emission from the stellar surface when the system is accreting less material from the companion compared to the systems described above [120], or not at all. The

thermal emission, a remnant of heating from previous accretion, is faintly measurable with x-ray observatories and measurement of their angular size along with knowledge of their distance allows for measurement of the NS radius. For observations of NSs where we have a good estimate of their distance (in a well-known globular cluster, for example), radii measurements are estimated to be $[8, 16]$ km [119]. Again, uncertainties around the NS atmosphere contribute error towards the measurement, along with uncertainties regarding the amount of x-ray absorbing matter in the interstellar medium.

Radius measurements of NSs almost entirely consisted of those of x-rays from NSs in LMXBs, until the NS Interior Composition Explorer (NICER) collaboration made their first observations of the pulsar PSR J0030+0451 [121, 122, 123]. Aboard the International Space Station, the x-ray instrument observes hotspots on the NS surface whose emission, when not on the rotational pole of the rapidly spinning star, is modulated. This is observed as an x-ray pulse which has propagated through the curved gravitational potential of the NS, providing a measurement of both total mass and equatorial radius [124]. Their first observation of PSR J0030+0451 returned constraints of $M = 1.44^{+0.15}_{-0.14}$ $M_\odot$ and $R = 13.02^{+1.24}_{-1.06}$ km [121]. This result agrees with the independent companion analysis of Riley et al. [122] and together their results provide direct constraint of the composition of cold, supranuclear NS matter via combined inference of mass and radius and integration of equations 1.7 and 1.8.

1.5 Bayesian inference

We introduce Bayesian inference as a method to infer the probability of a system's parameters, given an observation of GWs emitted by that system. We wish to use the observation of a given GW event to inform our knowledge of the parameters of the source, while accounting for our current knowledge of those parameters and our understanding of how likely we are to observe such a signal with those parameters. This formulation allows us to assign probability on an event-by-event basis, but also combine theoretical knowledge and our understanding of previous GW events into our current understanding of new events.

We begin with Bayes' theorem:

$$p(\vec{\theta}|h, M) = \frac{p(h|\vec{\theta}, M)p(\vec{\theta}|M)}{p(h|M)}, \tag{1.23}$$

where $p(\vec{\theta}|h, M)$ is the *posterior* probability of parameters $\vec{\theta}$ given observed GW strain h and a chosen model, M. Representing our initial understanding of the source parameters given our chosen model but regardless of the observed data is the *prior*, $p(\theta|M)$. The probability of observing the GW strain h given our model M and its parameters $\vec{\theta}$ is the *likelihood*, $p(h|\vec{\theta}, M)$. The *evidence* is given as $p(h|M)$, which acts to normalise the posterior probability distribution, such that

$$\int p(\theta|h, M)\ d\theta = 1, \tag{1.24}$$

however, we focus our discussion on the prior, likelihood and posterior. The posterior probability density $p(\vec{\theta}|h, M)$ of parameters $\vec{\theta}$ of a GW source is a joint probability density distribution on all N parameters of the source, both extrinsic and intrinsic. For analysis of a BBH merger, the posterior density distribution consists of $N = 15$ dimensions [125] describing the nature of the signal. In order to find the marginalised posterior probability density for a single parameter, for example the mass of the primary component in the merger m_1, one can marginalise over all other parameters to obtain the marginal 1-dimensional posterior probability density:

$$p(\theta_i|h, M) = \int p(\theta_i|h, M) \prod_{k \neq i} d\theta_k. \tag{1.25}$$

In practice, we express the posterior probability density of all N= 17 parameters of a BNS system together in PE in order to construct credible intervals and regions of confidence. The total number of parameters for a BNS system accounts for tidal parameters (discussed in section 1.3) and neglects parameters which accounts for the eccentricity of the system, assuming the orbits of the two NSs around one another have been circularised by the time we come to observe GW emission from the inspiral. By marginalising, we are finding our estimate of the posterior probability density $p(\vec{\theta}|h, M)$ accounting for uncertainty in all other $N - 1$ parameters defined by the joint posterior probability density.

1.5.1 Parameter estimation

GW PE applies the method of Bayesian inference to find the posterior probability density of parameters of a GW source given detection of a noisy signal, $d(t)$, introduced in equation 1.15. This can equivalently be expressed in the frequency domain, such that

$$d^*(f) = h^*(f) + n^*(f) \tag{1.26}$$

given that the Fourier transform is a linear transformation. Once we have observed data where we believe there to be a GW signal present, we find the joint posterior probability density, $p(\vec{\theta}|d(f))$, on a set of unknown parameters $\vec{\theta}$ given observation of a signal over a range of frequency $d(f)$. This requires some set up, primarily in construction of the priors on each of the parameters $p(\vec{\theta})$ and the choice of likelihood $p(d(f)|\vec{\theta})$. The likelihood is a description of the measurement made [126] and in GW data analysis it is common to use a Gaussian likelihood. The noise in a GW detector is characterised by the power spectral density (PSD), $S_n(f)$, likewise a function of frequency. We choose to perform the analysis in the frequency domain as the prominence of a given source of noise changes as a function of frequency. The likelihood is therefore modelled as the product of independent probabilities across frequency bins. We give the likelihood as

$$\mathcal{L}(d^*(f)|\vec{\theta}) = \sum_{i=0}^{N-1} \frac{1}{2\pi S_n(f_i)} \exp\left(\frac{-|d^*(f_i) - h^*(\vec{\theta}, f_i)|^2}{2S_n(f_i)^2}\right), \tag{1.27}$$

where $h^*(\vec{\theta}, f_i)$ is a theoretical model of the GW signal expected given parameters $\vec{\theta}$ i.e. a template waveform [127].

We must also choose our prior probability distributions to perform Bayesian inference. The choice of prior for a given parameter θ expresses our knowledge of this parameter *prior* to any measurement or observation being made. Once we have made an observation and have inferred the properties of this parameter given some method, we can then update our prior knowledge of the parameter via equation 1.23. Often, we have no prior knowledge of the parameters we are measuring, and in such a scenario we choose to have a minimally informative prior. A uniform prior gives every region of the parameter space equal probability, but is not invariant under

parameter transformation. The least informative prior, therefore, is the Jeffrey's prior [128]. Alternatively, we can have the prior distribution express our existing knowledge of a given parameter. For example, in GW inference of NS parameters, we can use mass priors which are informed by existing observations of NSs through EM observation or nuclear experiment (see chapter 2 section 2.2.1). See [129, 130] for related work on using physics-informed priors in inference of properties of NSs with GWs. Given the choice of likelihood and priors, we wish to evaluate the posterior probability of the parameters θ. Due to the high dimensionality of the full posterior probability distribution, we need efficient sampling methods to evaluate the function of the space without resulting to stepping along a grid which is inefficient in such a high dimensionality space.

1.5.1.1 Markov-chain Monte-Carlo

One example of a stochastic sampling method is MCMC which provides an efficient method to sample from the posterior probability distribution, our target distribution $p(x)$, the most generic implementation being through evaluation of the Metropolis-Hastings algorithm [131, 132]. We define the current state of the sampler, x_t, as a Markov chain, which is a sequence of random variables where the future state x_{t+1} depends only on the current state, and is independent of any previous steps. To determine the transition from step x_t to x_{t+1}, we must propose a potential step to be taken. To do so, we define a proposal distribution, which is based on the current state x_t of the sampler. Often, this proposal is a Gaussian distribution centred around x_t. Our proposed step is a sample drawn from proposal distribution $x' \sim q(x'|x_t)$ which depends on the current step of the algorithm. We evaluate the relative probabilities of the current state and the proposed state via [133]

$$\alpha = \frac{p(x')q(x_t|x')}{p(x_t)q(x'|x_t)}. \tag{1.28}$$

If $\alpha \geq 1$, the step is accepted and $x_{t+1} = x'$. Otherwise, the state is assigned probability α and is accepted or rejected relative to $u \sim U[0, 1]$, where we accept the proposed sample $x_{t+1} = x'$ if $\alpha > u$, otherwise it is rejected and $x_{t+1} = x_t$. This random walk is performed indefinitely, unless the user defines a condition at which to stop or a

finite run time, with the result being samples from the target distribution, $p(x)$, the posterior probability distribution.

In order to make this process computationally efficient, there are many variants of the standard MCMC set-up. Ensemble MCMC is the process of performing the Metropolis-Hastings algorithm in parallel with multiple 'walkers', where each walker explores the target probability distribution while in communication with the other walkers. This is a computationally efficient extension for problems in high dimensions; more walkers can explore the probability distribution more effectively. Additionally, when problems are multi-modal, an ensemble of walkers prohibits getting stuck on local maxima or minima. A very common example of ensemble MCMC is given by the python package `emcee` [134], the use of which will be discussed in chapter 6.

By sampling the 'chains' at the end of the MCMC sampling routine, i.e. the path of steps taken by each of the walkers in the ensemble, one can accrue posterior samples. It's standard to first allow the walkers time to settle in the parameter space, and then sample late stages of each of the walker's paths, leaving an initial 'burn-in' stage [126]. To determine sampling quality, one can measure the autocorrelation length, a measure of the number of steps required to produce statistically independent samples of the target distribution, or posterior probability density [134]. An alternative measure is the number of unique steps taken by the walkers. This measures the acceptance rate of proposed samples and can flag potential issues with the multi-modality of the target distribution, or the choice of proposal distribution.

1.6 Conclusion

In this chapter, we have firstly introduced the concept and theory behind NS physics, and how we can relate their equations of state to their macroscopic parameters through General Relativity and the TOV equations. We covered our current best understanding of the structure of a NS. More relevant to the current work, we also introduced GWs through a common reference point in General Relativity, and discussed how we can detect, model and infer source parameters of GWs. To do so,

we additionally introduced Bayes' theorem, which will become particularly relevant for statistical and computational methods employed in this thesis.

Next, we covered the specifics of how GWs observed from the merger of BNSs provides a unique method through which to understand the composition of NS matter. The measurement of the tidal deformability of a NS is unique to GW observation and is complementary to our current understanding from EM observations. The observation of the GW event GW170817 in particular provided constraints on the tidal deformability of a NS for the first time, and alongside the EM emission from the event, confirmed its status as the merger of (at least one) NS.

Our current understanding of NS structure is by no means comprehensive. Since being hypothesised and later observed for the first time as a radio pulsar, extensive observations have been made using a variety of methods. Radio observations of NSs continue to dominate our catalogue of understanding, but recent observations of binary NS mergers with GWs have offered a new method through which we can observe and measure NS parameters, but also to understand better the composition of NS matter.

As ground based GW detectors improve and as the next generation of detectors begins, we hope to add to our current collection of GW observations of NS mergers, and perhaps observe GWs from NSs from a different origin (for example, continuous GWs or those from the post-merger phase). Our current understanding of NSs may evolve, and particularly mysterious properties of NSs may no longer be so. In the next chapter, we learn more about the NS composition and their equations of state, and how we can infer the composition of a NS from observation of their macroscopic parameters.

Chapter 2

The dense matter equation of state

Our previous introduction of NSs - how they come to be, our current best understanding of their composition and their local environment - classified them as some of the most extreme bodies in our universe, with the most dense matter we know of. It is uncovering the nature of this matter and its true composition which is the goal of many of the studies listed in this chapter.

In chapter 1 we introduced how, through employing General Relativity and making assumptions about the state of the matter in a NS, we can arrive at the TOV equations. These equations encapsulate the structure of a NS and describe a relationship between macroscopic observables of a NS and its internal structure. The internal structure is given by the *equation of state* of neutron-rich matter, the only input required (apart from boundary conditions) to solve the TOV equations [22]. We introduce the equation of state as the foundational relationship between the pressure, P, temperature T and density ρ (or alternatively energy density, ϵ) of a fluid. The ideal gas law, $P = nk_BT$, for example, is an equation of state of an ideal gas. However, NSs are far from ideal; the temperature of a NS, while high by stellar standards [20], is very low relative to the Fermi temperature [22] and so can be approximated as having zero-temperature (discussed in chapter 1). With this treatment, we therefore consider the equation of state of a NS to be a relationship between P and ρ within the star.

The equation of state is bounded at the low density crust to have $P = 0$ and

increasing pressure as we approach the core. As such, an equation of state of NS matter is a monotonically increasing function of pressure with density. This satisfies *thermodynamic stability*; the pressure only increases as density increases as we travel further into the star. Another physical requirement for a valid equation of state is *causality*: the speed of sound within a NS should never be greater than the speed of light. The speed of sound in a NS describes the stiffness of the equation of state [135] and is calculated as it's gradient

$$c_s^2 \equiv \frac{dP}{d\epsilon}. \tag{2.1}$$

where ϵ is the energy density. For $c = 1$, causality requires that $c_s^2 \leq 1$ as an absolute limit and thermodynamic stability requires that $c_s^2 > 0$ [136].

Stiff equations of state are those where the pressure increases rapidly with density and therefore supports more repulsive force to combat the gravitational force pushing down on the star. An illustration of a stiff and a soft equation of state is given in figure 2.0.1. A stiff star can support a larger radius for a given mass, and is therefore more extended. Therefore the tidal deformability λ (introduced in chapter 1 section 1.3), is also large. Conversely, soft equations of state are those where pressure does not increase as steeply with density, with NSs of smaller mass and smaller volume (radius), which are more difficult ton tidally deform with small λ.

The true, universal equation of state of NS matter is currently unknown; the region of uncertainty, particularly at the highest densities within the inner core - is wide and so could exhibit some exotic phase of matter, or even phase transitions, the effect of which are illustrated in figure 2.0.2. Here, the phase transition appears as a sharp change in density at a given pressure, signifying a transition between two phases of matter within the star. Figure 2.0.2 presents one example of a first order phase transition, expected to occur within the core of a NS where the behaviour of matter at such high densities is widely unknown, however similar behaviour may also occur at the crust-core boundary. The presence of phase transitions may also be probed by observation of macroscopic observables of the NSs through features in the relations between those observables, for example NS mass, m, and radius, R [137].

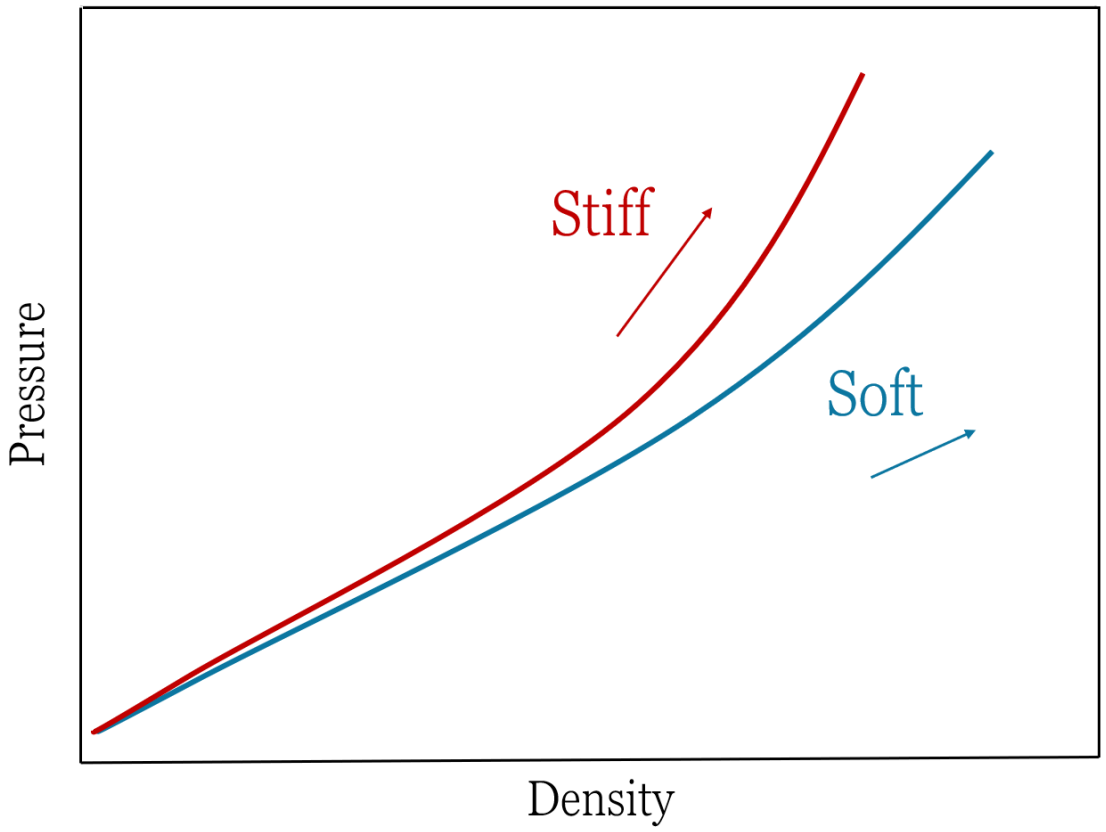


Figure 2.0.1: Illustration of two example equations of state in pressure vs. density demonstrating the difference between soft and stiff behaviour.

The equation of state of low density NS matter - the crust of the NS - describes constituents of matter which we are familiar with terrestrially, i.e. protons, neutrons and electrons [34] albeit in an extreme astrophysical environment [2]. The behaviour of such matter, up to and around nuclear saturation density, can be probed by physical experiment (discussed in section 2.2). Analyses often can comfortably make the choice of using a fixed crust equation of state without this assumption having influence on the overall result [138, 139]. Theoretical models attempt to model the NS equation of state based on our best knowledge of high density nucleonic matter extrapolated to densities beyond which we can reach with terrestrial experiments. However, to truly constrain the equation of state, we use information gleaned from astrophysical observations of NSs; by measuring the macroscopic parameters of NSs - mass m, radius R and more recently tidal deformability λ - we can infer the composition of matter required to support such a star with these properties.

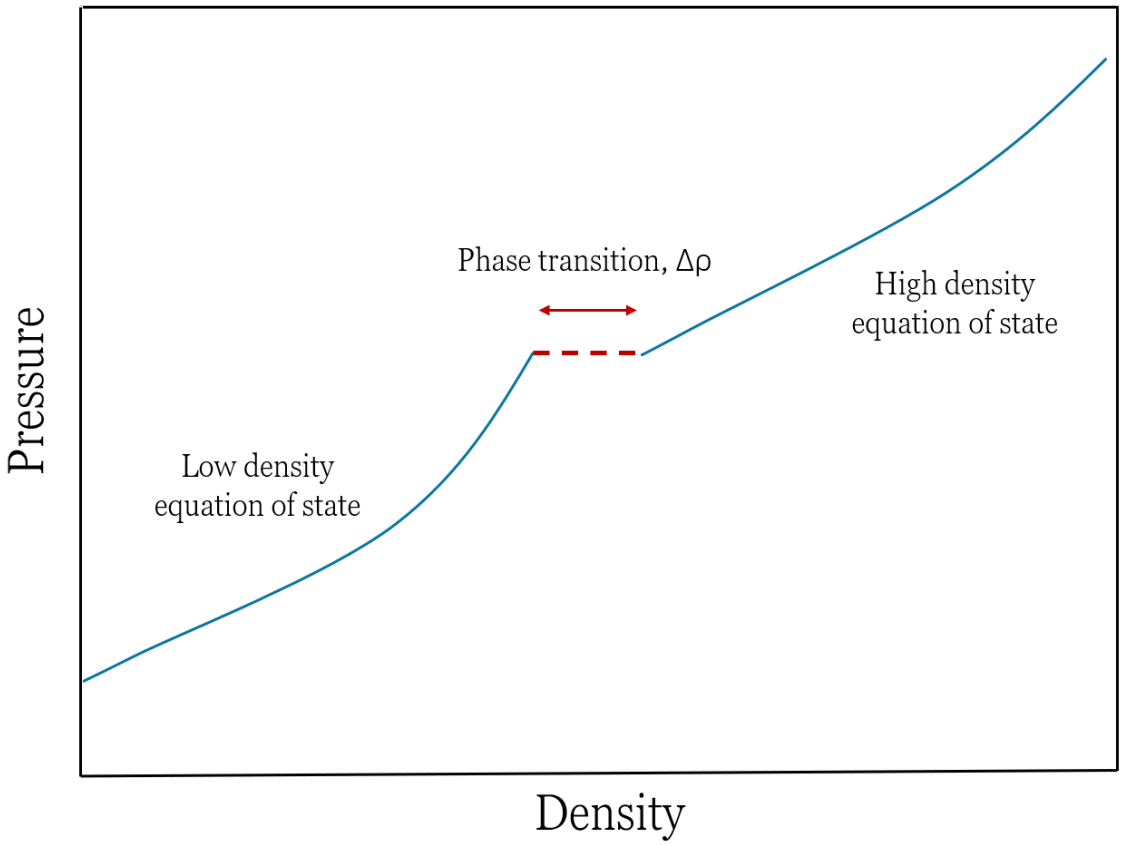


Figure 2.0.2: Illustration of an example equation of state in pressure vs. density with a first order phase transition.

In this chapter, we discuss different equation of state modelling methods, before discussing inference methods which constrain the NS equation of state at different densities. We then focus on equation of state inference using data from GW observations, briefly reviewing the literature on current inference methods. We finally introduce equation of state inference for the next-generation of ground-based GW detectors and discuss the upcoming challenges and requirements for future tools.

2.1 Modelling the neutron star equation of state

A variety of equation of state models exist which aim to describe the $P-\rho$ relationship within the star with a given physically-informed composition of matter. These include quark models [140], nucleonic models [141] and hybrid models [142], each of which are built to describe the matter within a NS on a microscopic scale. An example of such an equation of state is SLy4 [143], a member of a family of SLy equations of

state [144] which describe high density pure neutron matter [145], and as such are common in the analysis of NS composition. SLy4 is a hadronic unified equation of state, meaning that from crust to core, the composition of the NS matter is described by the same equation of state. Equations of state like SLy4 are built to describe the behaviour of this matter under the extreme conditions present in a NS, and as such obey the physical laws that define matter at this density, namely thermodynamic stability and causality. There are many alternative models which are built to describe the nature of matter within a NS with different underlying physical phenomena in precise detail - those cited above are just a few. However physically informed models, including SLy4, are complex and require many parameters to describe them. Also, they each describe the NS interior using one physical description of the composition of matter.

Alternatively, we can describe the NS equation of state with phenomenological models; these are relations which instead aim to describe the behaviour of the equation of state relationship in pressure and density, rather than being defined by nuclear theory. These are common in inference methods, where we use observation to infer the parameters of a given model in order to describe the overall equation of state. This is less computationally expensive than solving a physically motivated equation of state or integrating the TOV equations. One equation of state model whose use is prevalent in the literature is the piecewise polytropic equation of state [146] of Read et al., where the $P - \rho$ relationship is represented as a series of adjoining polytropes up to a maximum density. A polytropic equation of state is described by:

$$P(\rho) = K_i \rho^{\Gamma_i} \tag{2.2}$$

where the pressure P of the ith polytrope is a function of rest-mass density ρ and adiabatic index Γ_i with coefficient K_i. Beyond a minimum density, ρ_0, polytropes are built for a series of density intervals, described by their own K_i and Γ_i, each bounded such that $\rho_{i-1} \leq \rho \leq \rho_i$. Read et al. demonstrate in [146] that 6 parameters, ρ_0, Γ_1, ρ_1, Γ_2, ρ_2 and Γ_3, are necessary to construct a 3-piece polytropic model, given that

$$K_1 = \frac{P(\rho_0)}{\rho_0^{\Gamma_1}}. \tag{2.3}$$

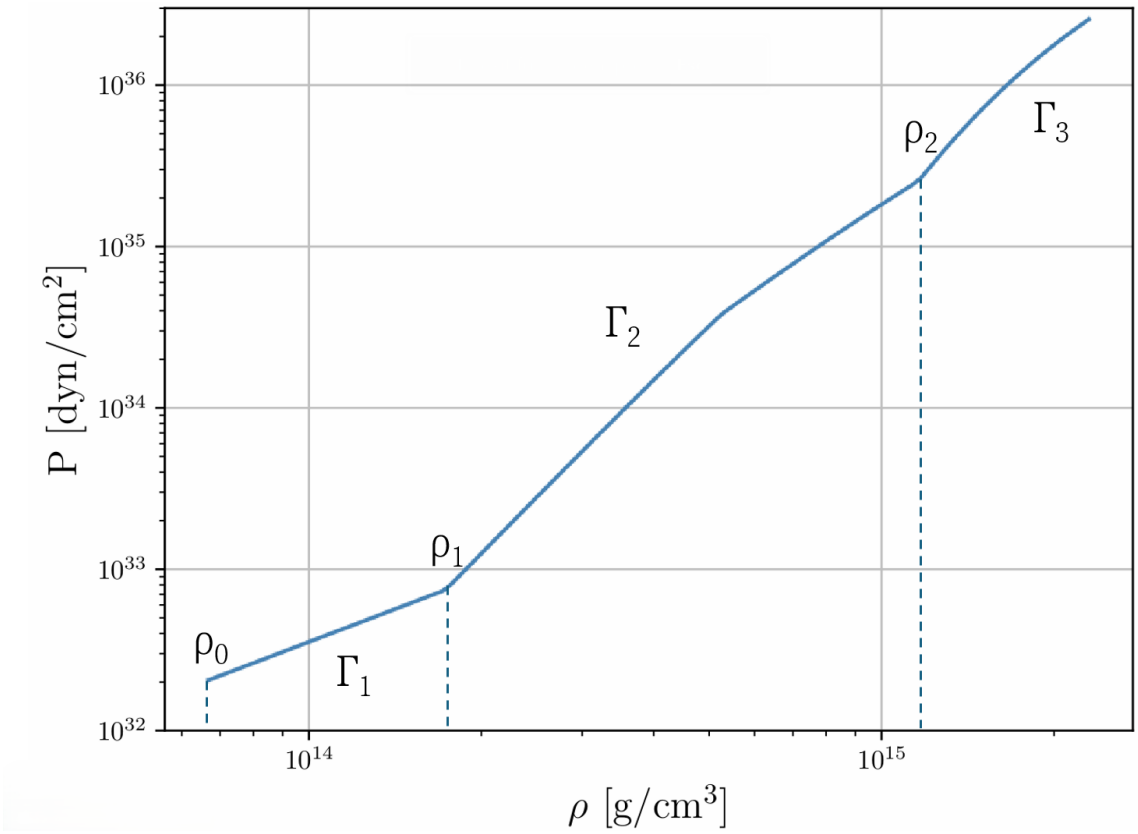


Figure 2.1.1: Plot of pressure vs. density of an example piecewise polytropic equation of state with 3 pieces, parameterised by adiabatic indices $\Gamma_{1,2,3}$ and with joining densities $\rho_{0,1,2}$.

However, if densities between adjoining polytropes are fixed, only 4 parameters are therefore necessary: ρ_0, Γ_1, Γ_2, Γ_3. This allows the user to define the pressure at ρ_0, where $P_0 = P(\rho_0)$ and the rest of the pieces by their respective Γ_i between the fixed densities. Read et al find that using fixed adjoining densities of $\rho_1 = 10^{14.7}$ g cm^{-3} and $\rho_2 = 10^{15.0}$ g cm^{-3} reduces the residual when fitting a given set of theoretically motivated equations of state to a minimum.

Carney, Wade and Irwin (not the author) [147] compare the piecewise polytropic model to an alternative parameterisation: Lindblom's spectral representation [148]. Lindblom states that this alternative parameterisation can reconstruct a set of realistic equations of state with a smaller number of parameters than Read et al. within a comparable margin of error. Additionally, they state that the parameterisation is suitable for complex (with phase transitions) at a comparable accuracy to piecewise polytropic equations of state, as well as simple (smooth) equations of state. Lindblom

suggests that the spectral method can be improved further to tackle phase transitions in the equation of state by using more spectral parameters, and modelling the equation of state above and below the phase transition pressure with different fits, however this increases the complexity of the fit. When tasked with reconstructing the simulated parameters of a GW signal of a BNS merger associated to a known equation of state, Carney, Wade and Irwin [147] find that both the polytropic and spectral models recover near identical tidal information from the signal, demonstrating that each equation of state model interprets the same information relative to the true underlying equation of state, and any inconsistency in information learned is not due to the equation of state model choice. However, in reconstructing the true simulated equation of state, the polytropic model introduces additional statistical error at the joining densities of each of the polytropic segments, an error which is entirely avoided by using the spectral model which builds a smooth $P - \rho$ relationship. Both models are, however, flexible in the construction; for example, the polytropic model can be represented by less parameters than is presented in [147], mitigating error induced at connecting pieces.

A unified approach to modelling the high density equation of state is a solution somewhere between full nuclear descriptions of the equation of state and phenomenological models. Davis et al. [2] propose an efficient construction of NS equations of state by extracting nuclear parameters from the high density equation of state described phenomenologically (for example, piecewise polytropic) and using these parameters to calculate the low density equation of state. As a result, the low density equation of state is described by a meta-model [149], a 'model of models' which constructs an equation of state given nuclear parameters which have existing constraints, for example given nuclear laboratory experiments or observation. The crust equation of state is stitched to the core equation of state at a matching density point, which occurs somewhere near saturation density by detaching the nuclear parameters from the given equation of state at this point and carrying these to create the low density equation of state. This equation of state construction method is such that each equation of state is consistently and uniquely described by nuclear parameters from high to low density, with a minimal number of parameters. The benefits of this will be described in the

section 2.3, where we consider equation of state constraints with future GW detectors.

2.2 Inferring the neutron star equation of state with observations

While we intend to build models which match the observed physical phenomena or describe the underlying physics, we also wish to use our observations to inform our understanding of the equation of state. Different observations of physical phenomena provide us with information on different density regions of the nuclear equation of state. Figure 2.2.1 from [130] presents example equations of state over a broad pressure and density range and the different methods by which we can constrain the equation of state in different density regimes. Up to twice the nuclear saturation density, terrestrial nuclear experiments can probe the equation of state of high density neutron-rich matter through experiments which are discussed more in depth in the next paragraph. From $2\rho_{\rm sat}$ to around $4\rho_{\rm sat}$, GW observations of BNS mergers, in particular the inspiral phase where we measure the tidal interaction of the two components (as introduced in chapter 1 section 1.3) provide observation of the behaviour of neutron-rich matter at these densities. Beyond this, we require observation of EM emission from NSs and also observation of GWs from the post-merger phase of two NSs, where we might support high mass NSs, hosting ultra-dense matter for a very short period before eventual collapse.

Understanding the equation of state of high density nuclear matter through terrestrial experiments [150, 151, 152, 153] is a difficult task, given the densities that are required to be replicated are around and above the nuclear saturation density, $\rho_{\rm sat} = 2.7 \times 10^{14}$ g cm^{-3}. At this density inside a NS, we expect the composition of matter to be highly asymmetric, where the number of neutrons is much greater than the number of protons. To understand the behaviour of matter under these conditions, terrestrial experiments aim to measure the *nuclear symmetry energy*, $\boldsymbol{J}$, a parameter which quantifies the difference between symmetric nuclear matter (equal quantities of protons and neutrons) and pure neutron matter [154]. In the high density regime, this

parameter describes the energy required to go from symmetric to pure neutron matter [155] and is key to understanding the overall behaviour of the system. The slope of the symmetry energy, L, evaluated at saturation density, determines the pressure of symmetric nuclear matter (at ρ_{sat}) [154]. By providing the dominant source of pressure at this density, L is therefore known to be positively correlated with the NS radius [156]; a larger L at saturation density infers a more rapid increase in energy required to maintain the system with density and therefore a stiffer equation of state and larger radius for a given mass [157, 154].

The parity radius experiment (P-REX) aims to measure the thickness of the 'neutron skin' in Pb^{208} [152], mimicking a NS crust. In neutron-rich nuclei, the radius of the volume that neutrons occupy become larger than that of the protons due to

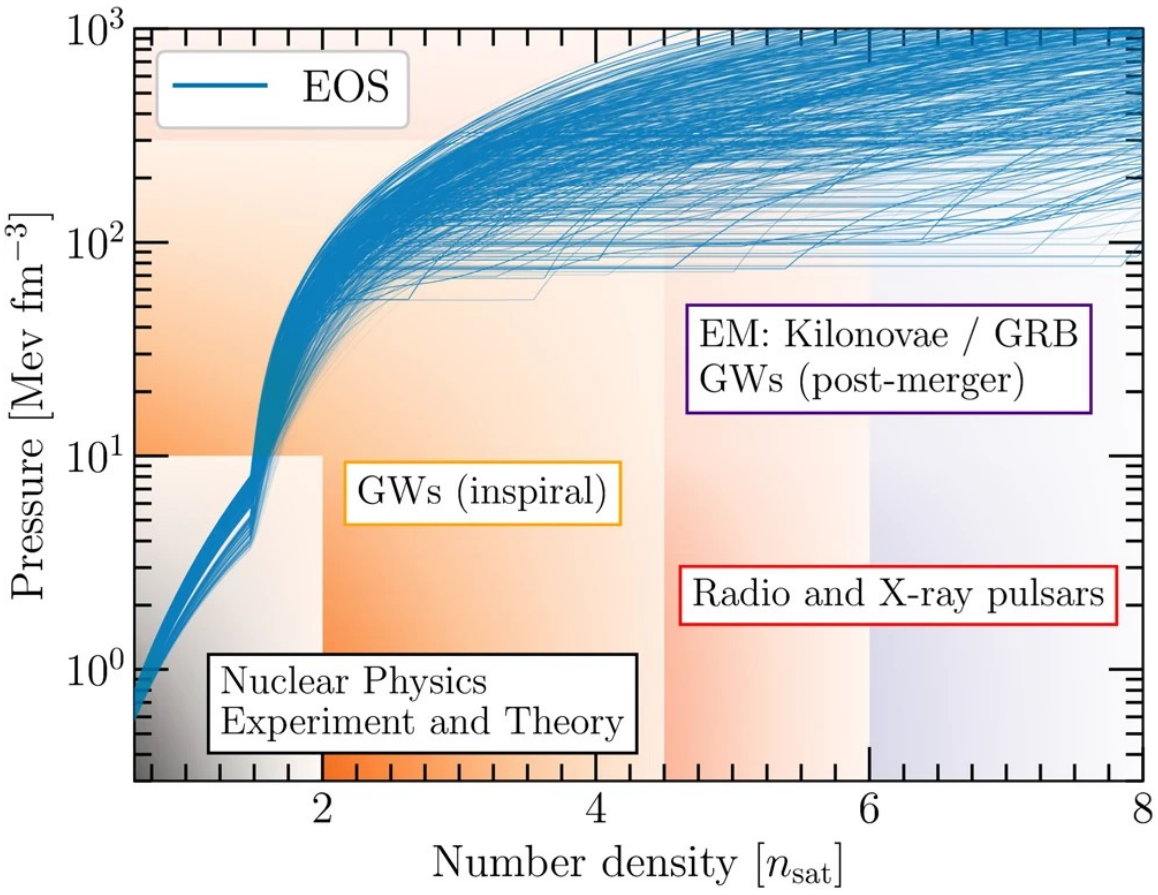


Figure 2.2.1: A plot of number density in units of nuclear saturation density (n_{sat}) vs. pressure for example high density nuclear equations of state. Different regions are shaded in different colours, representing the regions of density or pressure which are informed by different experimental or observational constraints. From [130].

differences in binding energy (a direct impact of the symmetry energy of asymmetric matter) [158]. Assuming that the same physical phenomena are operating on the neutrons within the lead isotope as are inside a NS, measurement of the thickness of the neutron skin allows for direct inference of the saturation density and its slope, and provides complementary information for inference of the NS radius [155].

In the higher density regime, above $2\rho_{\text{sat}}$, we rely on astrophysical observations. Given that NS densities can reach up to $\sim 10\rho_{\text{sat}}$ within the core [96], observation of astrophysical phenomena provides a means of measurement of the properties of high density neutron rich matter. Measurement of NS masses and radii through a variety of electromagnetic means has thus far placed constraints on the NS equation of state, and in particular constrains the equation of state through the direct mapping from $M - R$ to $P - \rho$ relationships [21]. The details on observation of macroscopic parameters of NSs through a variety of electromagnetic methods are discussed in chapter 1 section 1.4.2.

Additionally, the maximum mass of the NS is defined by the pressure at the highest densities within the star; Özel and Psaltis [160] found that varying the pressure at densities of around $\sim 7\rho_{\text{sat}}$ varies the $m - R$ mostly in the determination of the maximum mass allowed by the equation of state. Inference of the maximum allowed mass via observation is difficult, and is defined currently by a few select observations of heavy NSs [161, 162]. From these observations, lower limits have been placed on the maximum allowed mass, which enforces that all valid equations of state must allow for such massive NSs to be stable.

The past few decades of NS observations and nuclear experiment have thus resulted in various constraints being placed on our current knowledge of the NS equation of state. The result of these constraints are expressed in figure 2.2.2 from an analysis performed by Ng et al [159]. The figure presents the combined equation of state constraint from multiple methods including radio observation of heavy pulsars and x-ray observation of millisecond pulsars (primarily by the NICER experiment [123, 163, 122, 121]), both introduced in chapter 1 section 1.4.2, alongside constraints from GW observations of

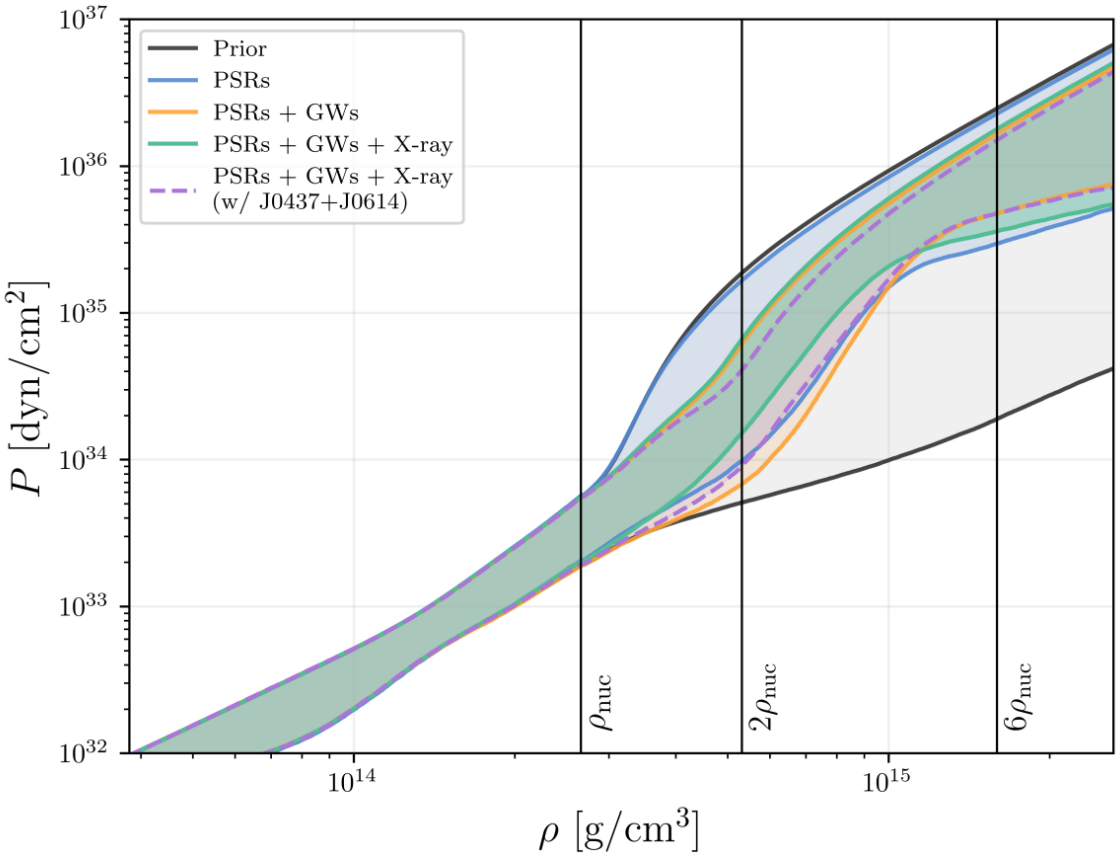


Figure 2.2.2: Equation of state constraints from radio observation of massive pulsars (blue), GW observations of BNS mergers (orange), and observation of x-ray emission from millisecond pulsars (green and purple) from an analysis performed in [159] which uses nuclear-informed priors (grey).

BNS mergers. Their analysis includes constraints from nuclear theory and experiment through use of a nuclear-physics-informed prior developed by [2]. More details of this data set will be discussed in 5 section 5.1 when we incorporate it into our analysis. It is clear from figure 2.2.2 (and complementary analysis which is in broad agreement; for example, [164, 165, 166]) that the current observations, theory and experiment have constrained the equation of state space so to rule out particularly *soft* equations of state from the nuclear-informed prior, and each of the electromagnetic constraints broadly agree on a general trend. However, from all existing observations, including those not discussed here (for example [167, 94]), which aim to observe the behaviour of NS matter as it undergoes further compression, none are particularly strong to provide a very stringent constraint on the equation of state currently.

2.2.1 Inferring the neutron star equation of state with GWs

Observations of GWs from merging NSs allow us to measure the dimensionless parameters $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$ from the gravitational waveform, which are the combined tidal deformability of the system and the difference in tidal deformability of the components, respectively. As discussed in chapter 1, these parameters appear in the gravitational waveform of a BNS merger at the 5^{th} post-Newtonian order [168]. As stated in chapter 1 equations 1.21 and 1.22, these are by definition reparameterisations of the individual dimensionless tidal deformability of each component, Λ_1 and Λ_2, and the symmetric mass ratio, η [169]. It is not trivial to recover the tidal deformability of each of the components; $\delta\tilde{\Lambda}$ is not as well recovered from the waveform as $\tilde{\Lambda}$ with current GW detectors [168], as discussed fully in chapter 1 section 1.3.1.

2.2.1.1 Universal relations

In seeking a method to resolve the individual component tidal parameters, Yagi and Yunes [82] discuss *universal relations*: relationships that have been found to exist between macroscopic parameters of NSs that are approximately independent of equation of state. The reason of the universality of these relations is currently unknown, but it is thought to be at least in part due to the similarity of NSs in the region which we probe with the parameters that we measure [170]. However, the matter composition in these regions of interest is by no means well-defined and so this alone cannot explain the universality. In practice, these relationships allow us to automatically recover parameters of a BNS given measurements of others, assisting the broader task of parameter estimation [82]. The primary universal relation which is relevant to this work is the *Binary Love relation*. This is a quasi-universal relation between the symmetric $\Lambda_s \equiv (\Lambda_1 + \Lambda_2)/2$ and anti-symmetric $\Lambda_a \equiv (\Lambda_2 - \Lambda_1)/2$ tidal deformability of two NSs in a BNS merger. While the relationship between Λ_1 and Λ_2 for a given mass of NS depends heavily on the equation of state, it has been found that the relationship between Λ_a and Λ_s does not. Therefore, by solving simultaneously, one can resolve the component tides more easily. In PE of a GW signal from a BNS merger, measurement of Λ_a and Λ_s therefore allows for accurate recovery of both individual tidal deformability parameters of each star [82]. However, universal

relations assume a simplified BNS system, where both stars are cold [171], slowly and uniformly rotating in GR [172], and do not account for phase transitions within the stars [173]. Continued independent measurement of these macroscopic parameters can therefore (in)validate universal relations moving forward.

2.2.1.2 Brief literature review

Currently, we are limited by the number of joint observations of both mass and radius/tidal deformability of the same source. Therefore, we rely on Bayesian inference methods to use individual measurements to infer probabilities of given equations of state. This allows us to incorporate prior knowledge, e.g. from nuclear theory, into the analysis. We refer the reader to chapter 1 section 1.5 to recap the theory of Bayesian inference if required. We discuss Bayesian inference within the context of GW observations of NSs only.

Prior to any detection of GWs from NS mergers, or of GWs from any CBC for that matter, Read et al. [174] used numerical relativity to simulate BNS merger waveforms as deviations from that of BBH waveforms (often modelled as point particles) in order to estimate the deviation induced by the presence of matter in the system. By simulating the late stages of inspiral and merger of two 1.35 $M_\odot$ NSs, each associated to a underlying realistic piecewise polytropic equation of state [146], they expected that Advanced LIGO would be able to distinguish between a BNS waveform produced and that of point particles at an effective distance of 100 Mpc with a single Advanced-LIGO-configuration GW detector. Read et al. projected that one observation of a BNS merger at a distance of 100 Mpc could constrain the NS radius to within $\sim 10\%$, or approximately 1 km. This can be equivalently considered as an error in pressure in the $P - \rho$ plane; at a density of 5×10^{14} g cm^{-3} one can determine the pressure to within $\sim 10^{33}$ dym cm^{-2}.

Del Pozzo *et al.* [79] performed a similar but fully Bayesian investigation to determine how many detections of BNS signals with second generation GW detectors (Advanced LIGO era) would be required to achieve considerable constraints on the NS equation of state, given the BNS waveform is discriminable from that of point

particles. They suggest that $\mathcal{O}(10)$ observations of BNS mergers would be enough to constrain the equation of state in the $m-\Lambda$ plane, providing distinguishable preference for either stiff or soft equations of state. However, their analysis considers only 3 distinct equations of state defined in the literature, rather than a continuous array of parametric equations of state and a spectrum of stiffness vs. softness. Another Bayesian analysis was completed by Chatziioannou et al. [175] in order to determine whether GW observations of NS mergers could distinguish between different internal compositions of NSs. They found that an observation of a GW signal with an SNR of 20 could provide indication of possible non-nucleonic equations of state (potentially non-NS in nature), while an SNR of 30 could either confirm or deny entirely. To differentiate between different flavours of non-nucleonic equations of state, higher SNR observations are required, but the authors state that second generation GW detectors (advanced LIGO) will not be able to differentiate between equations of state that differ only at high central densities (i.e. NSs with exotic inner core content).

The announcement of the GW observation of the BNS merger GW170817 came in October 2017 [54], the highest SNR GW event at the time [38]. This observation was the first of it's kind: direct observation of properties of ultra-dense matter undergoing tidally induced stress and strain. The LIGO-Virgo collaboration soon published their own equation of state inference from the event [4] in which they present constraints on the NS parameters and composition with two methods, incorporating information from the EM follow-up campaign post-announcement [176, 177, 90, 178]. Their method samples uniformly in symmetric tidal deformability, $\Lambda_{\rm s} \in [0, 5000]$, among all other parameters of interest in PE (method discussed in chapter 1 section 1.2). The authors then apply the Binary Love universal relation [82] to return the asymmetric tidal deformability, $\Lambda_{\rm a}$ and thus the component tides, Λ_1 and Λ_2. The second method samples directly the parameters of Lindblom's spectral model [148], γ_i, where the rest-mass density is a function of these parameters and pressure, $\rho(p;\gamma_i)$,which is inverted to give the equation of state in $P(\rho)$. Sampling the equation of state parameters directly allows for physical constraints on the equation of state to be directly enforced in the analysis, and for all physically viable equations of state to be valid within the prior bounds set on the equation of state parameters. The TOV equations are

integrated to recover Λ_1 and Λ_2, and as such both methods can be compared in determining the tidal deformability of both components. The LHS of figure 2.2.3 presents the joint $\Lambda_1 - \Lambda_2$ posterior of the component NSs of GW170817 from [4]. The shaded green region presents the posterior given from the method which employs the binary love universal relation. The green, blue and orange solid lines present the 90% credible intervals for posteriors found using the universal relations, the parameterised spectral equation of state and a result from an alternative analysis performed by the collaboration [83], where component tidal deformability parameters are sampled independently and the TOV equations are solved to find the possible equations of state, which is not assumed to be common between the two NSs. The shaded region represents the region which violates convention ($\Lambda_1 < \Lambda_2$) and so is unphysical. The solid black lines present the $\Lambda_1 - \Lambda_2$ relationship for given equation of state models from the literature. Both of the analyses they perform prefer softer equations of state where tidal deformability is smaller, inferring more compact stars. The RHS of figure 2.2.3 also presents the equation of state posterior as declared in [4]. Relative to their equation of state prior bounds, their result pushes to a softer equation of state posterior to both 50% and 90% confidence. This is in agreement with the LHS figure; a softer equation of state makes for a slower increase in pressure with density into the core of the NS. This means there is less pressure support throughout the star against the intense gravitational force, and a smaller radius of NS. A smaller, compacted NS is less easily tidally disturbed as a result.

The collaboration's method of returning the equation of state posterior – through building equations of state given the posterior samples in equation of state parameters γ, and integrating the TOV equations along with the mass samples to return tidal deformability – adds significant computational time at each stage. Adding extra dimensions to sample over in PE is non-trivial, alongwith the added constraint of integrating the TOV equations for each of the 1000s of equations of state. In chapter 4, we discuss a method to return a non-parametric equation of state posterior without need for any of these additional steps beyond standard PE after a BNS merger event.

After the observation and initial equation of state inference published by the

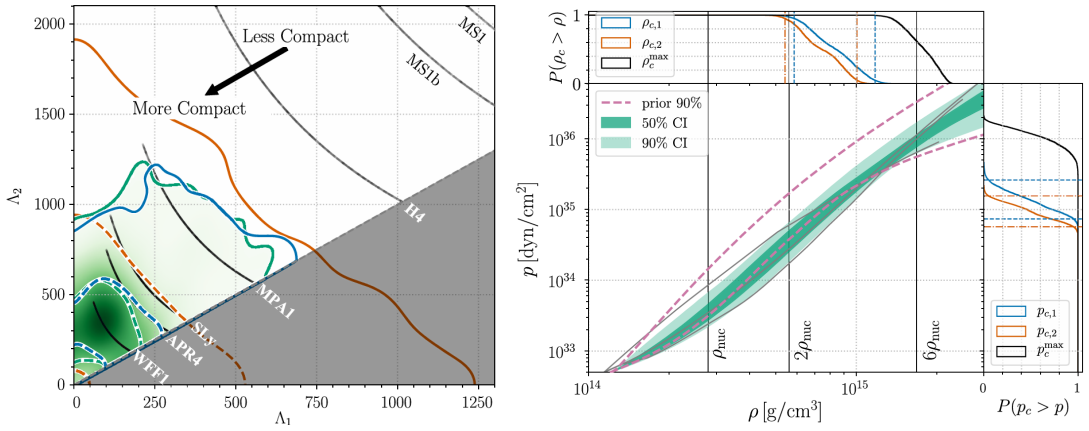


Figure 2.2.3: LHS: joint $\Lambda_1 - \Lambda_2$ posterior from GW170817, where the shaded green region is the posterior to 90% confidence, compared to example equations of state from the literature in solid grey lines. Results from other analysis methods performed in [4] are presented in blue and orange solid and dashed lines to 90% and 50% confidence, respectively. The shaded lower grey corner represents the out-of-bounds $\Lambda_1 \geq \Lambda_2$ region. RHS: plot of pressure vs. density for the equation of state posterior to 50% and 90% confidence in dark and light green bands, respectively, from [4]. Their training prior bounds are presented in dashed pink lines. The upper and RHS plots present 1-dimensional cumulative probability densities for ρ_1, ρ_2, and $\rho_{\rm max}$ as well as P_1, P_2, and $P_{\rm max}$.

collaboration [4], more sophisticated analyses and discussion of how to improve constraints by combining information from multiple BNS mergers were the next topic of discussion. One example is that of Landry and Essick [179], who introduce a non-parametric method of inferring the NS equation of state given observation of BNS mergers using Gaussian process regression [180], also known as Gaussian processes. As discussed above, parametrising the NS equation of state restricts the equation of state to a small number of parameters which describe a pre-defined morphology. This is an efficient alternative to describing existing equation of state models with fewer parameters, however in inference, this can severely limit the extent to which the equation of state can be described, especially if the true, universal equation of state exhibits strong features characteristic of phase transitions which the chosen parameterisation cannot emulate to high precision without an increase in number of parameters [148, 147]. Alternatively, Gaussian processes allow for a 'statistical map' to exist between mass and tidal deformability parameters of an event, $[m_{1,2}, \Lambda_{1,2}]$, and a diverse set of synthetic non-parametric, thermodynamically stable and causal equations of state which align with current observational and experimental constraints.

The Gaussian process models the underlying functional relationship between energy density μ and pressure P as a multivariate Gaussian. Each realisation from the function that the Gaussian process describes is a selection of correlated random variables. The Gaussian process is a construction of a conditional distribution for a set of arbitrary points in the function space f_i, where $f_i = f(x_i)$, conditioned on known values f_i^*; in this case, some known equations of state from nuclear theory:

$$p(f_i|x_i, \{f_j^*, x_j^*\}) = \frac{p(f_i|f_j^*, \{x_i, x_j^*\})}{p(f_j^*, \{x_j^*\})}. \tag{2.4}$$

Once conditioned on candidate equations of state informed by nuclear theory, individual samples from the Gaussian process yield physically motivated samples from the equation of state prior. The corresponding $m - \Lambda$ relationship associated to each synthetically generated equation of state by integrating the TOV equations and the GW likelihood (modelled by a kernel density estimate (KDE)) can be calculated via Monte Carlo integration.

The authors applied their method in [181] to inferring the NS equation of state with data from GW170817 and find that soft equations of state are preferred, consistent with other prominent analyses, even with their most diffuse prior. Landry, Essick and Chatziioannou [182] demonstrate the extension of the workflow to combine information from multiple astrophysical sources, notably from the seminal constraints placed on the mass and radius of the first pulsar observed by NICER [163].

2.2.2 Combining information from multiple GW events

Since the observation of GW170817, many tools have been developed to infer the equation of state of individual BNS merger events and constraints have been placed on the composition of high density nuclear matter. These constraints, however, are limited by the uncertainty on the measurement of macroscopic parameters, and the limited number of detections made. Future ground-based detectors will have improved sensitivity and bandwidth and thus will measure parameters of interest with greater precision [58, 59]. Additionally, the number of BNS mergers is expected to increase [183] with many more opportunities for observation and further constraint. However,

we now need adapted tools to be able to combine the information from multiple observations for improved constraint. We introduce hierarchical Bayesian inference before introducing existing methods.

2.2.2.1 Hierarchical Bayesian inference

As covered in chapter 1 section 1.5, Bayes' theorem expresses the posterior probability density function (PDF) of parameters $\vec{\Omega}$ given observed data h as

$$p(\vec{\Omega}|h, M) = \frac{p(\vec{\Omega}, M)p(h|\vec{\Omega}, M)}{p(h, M)}, \tag{2.5}$$

where $p(\Omega, M)$ is the prior probability of parameters $\vec{\Omega}$, $p(h|\vec{\Omega}, M)$ the probability of observing data h given parameters $\vec{\Omega}$, or the likelihood of parameters $\vec{\Omega}$, and $p(h, M)$ the evidence supporting the observed data. In the case of GW observations of CBC signals, if we have N BNS merger events, we assume that each measurement of GW strain, h is statistically independent. We wish to determine the combined probability of the parameters $\vec{\Omega}$, which each describe a property of the NSs, given all N events. For this we combine the individual likelihoods of observing each of the N events given the parameters $\vec{\Omega}$ [184]:

$$p(\vec{\Omega}|h, M) \ \propto \ p(\vec{\Omega}, M) \ \prod_{i=1}^{N} \ p(h_i|\vec{\Omega}, M). \tag{2.6}$$

This does not require the determination of the individual event posteriors, but rather is a computationally effective way to determine an overall posterior from an arbitrary number of observed events. Hierarchical Bayesian inference methods concern themselves with instead having an efficient likelihood calculation, $p(h_i|\vec{\Omega}, M)$ which can be combined. Information from each BNS merger event is given through the measured parameters of the system, $\vec{\theta}$, which in particular includes m_1, m_2, Λ_1 and Λ_2, among other observable parameters, for example the spins of the two NSs. Analyses which investigate low and high spin scenarios have upper limits on the prior of 0.05 and 0.89 respectively. The overall equation of state posterior given N BNS merger events is given by:

$$p(\vec{\Omega}|h, M) \ \propto \ p(\vec{\Omega}, M) \ \prod_{i=1}^{N} \ \int \ p(h_i|\vec{\Omega}, \theta_i, M) \ p(\theta_i|\vec{\Omega}, M) \ d\theta_i. \tag{2.7}$$

where we have marginalised over the GW observable parameters $\vec{\theta}$. Our overall goal is to then infer the set of parameters $\mathbf{\Omega}$ which is common to all events. The mathematical explanation of hierarchical Bayesian inference in introduced again and in more depth in chapters 5 and 6 when in use.

2.2.2.2 Current hierarchical inference methods

The analysis discussed previously of Del Pozzo et al. [79] use a Bayesian method to combine information from multiple BNS sources. By assuming a universal equation of state, they find the combined posterior on the tidal deformability by constructing the likelihood, $\mathcal{L}(d_i|\vec{\theta}, \lambda_{1,2})$ of observed data for the i^{th} event, d_i, where $\lambda_{1,2}$ are coefficients of a linear expression of the tidal deformability, λ:

$$\lambda(m) = \lambda_0 + \lambda_1(m - 1.4M_\odot)/M_\odot \tag{2.8}$$

and $\vec{\theta}$ are other observed parameters of the system. The parameter space is explored by evaluation of the likelihood given the prior through nested sampling in the manner of [185]. In simulated signals which are built with parameters that are associated to a true underlying equation of state, they find, after a few tens of sources are incorporated into the analysis, the true tidal deformability is recovered and is clearly distinguishable from the others, each of which are respectively associated to a soft, moderate and stiff equation of state. However, to re-cap the literature review in section 2.2.1.2, the method employed by the authors uses only 3 strict equations of state from the literature, rather than a continuous set of parameterised equations of state that may be soft or stiff to different degrees, and so despite their clear result, the analysis differs in comparison to those which employ purely parameterised equations of state. The analysis of Del Pozzo et al. further investigates with how many BNS events it would be apparent that macroscopic parameters of the BNS system (standard data products of parameter estimation) prefer softer/stiffer equations of state. The simplified parameterisation of λ here restricts the way in which we express the $\lambda(m)$ relationship for the three discrete equations of state used in the analysis. The authors find the gradient of their model of $\lambda(m)$ as a straight line difficult to measure but the intercept, or measurement of λ at a mass of 1.4 $M_\odot$, is easier to determine.

Lackey and Wade [80] instead infer the parameters of an equation of state model directly given a 'realistic' population of N BNS events. Their realistic population corresponds to a year of advanced LIGO observation with an expected event rate of 40 mergers of two 1.4 $M_\odot$ NSs detected per year with SNR > 8 [186]. They wish to find the posterior PDF $p(\vec{\Omega}, \vec{\theta}|h)$ of equation of state parameters $\vec{\Omega}$ and parameters of the system $\vec{\theta}$ given observed GW data h. The inferred equation of state parameters $\vec{\Omega}$ are that of the piecewise polytropic model of Read et al. [146] while the GW parameters $\vec{\theta}$ they consider are a reduced set of parameters which assume zero spins, meaning their overall combined equation of state and GW parameter space is 11-dimensional. By marginalising over the non-equation of state sensitive parameters (distance, right ascension, declination, phase, inclination, time of coalescence, phase at coalescence, and difference in dimensionless tidal deformability, $\delta\tilde{\Lambda}$, assuming this parameter is not measurable at advanced LIGO sensitivity), they construct quasi-likelihoods for the equation of state-sensitive parameters $\vec{\theta}_{\rm in} = \{m_1, m_2, \tilde{\Lambda}, \log(p_1), \Gamma_1, \Gamma_2, \Gamma_3\}$ along with equation of state parameters $\vec{\Omega}$ for each of the N events. Marginalising over extrinsic parameters reduces the dimensionality of the problem and retains only the equation of state-sensitive information, given that the equation of state-insensitive parameters that we have marginalised over and the piecewise polytropic parameters are independent. Additional assumptions are made so that the problem simplifies to calculating the quasi-likelihood of each of the N events via additional marginalisation over masses $\{m_1, m_2\}$ and taking the product via equation 2.6. The posterior PDF is sampled using Markov Chain Monte Carlo (MCMC) and each posterior sample of $\mathbf{\Omega}$ corresponds to an equation of state in the $P - \rho$ plane. Their result is consistent with the true injected equation of state but with sharp features which are spikes in uncertainty associated to the transition densities between polytropic pieces, an error which is a direct consequence of their choice of equation of state parameterisation. With ≥ 5 events, they state that the analysis is dominated by the loudest 5 events, and more than 5 events does not significantly improve the constraint of the equation of state or measurability of intrinsic parameters. They discuss how more sophisticated equation of state parameterisations are necessary to ensure that the chosen parameterisation does not negatively impact the recovery of the equation of state, especially when

considering more complex equations of state.

Golomb and Talbot [187] use a GMM, a linear combination of multi-dimensional Gaussians, to model the multi-dimensional likelihood of each event in their analysis. Constructing this GMM acts as an analytical model for the likelihood $\mathcal{L}(h|\vec{\theta})$ of observed data h modelled by parameters $\vec{\theta}$. They likewise use equation 2.6 to explore the parameter space and evaluate the likelihood of given equations of state $\mathbf{\Omega}$ and the prior $p(\mathbf{\Omega})$, using nested sampling to do so. The computational time taken to perform their analysis scales with the number of events. They quote that each likelihood evaluation took $\leq$ 50 ms in a combined analysis of 37 events - this is repeatedly calculated during nested sampling and so dictates the overall time taken, which depends on the sampler settings and performance. They state that the time taken to complete one likelihood evaluation is comparable to that of hierarchical analyses of the mass distribution of a BBH population in the same detector set up [188], meaning the increased complexity of NS parameter and equation of state inference does not inhibit their analysis.

In comparison, the 'rapid' analysis of Ray et al. [189] instead uses a kernel density estimate (KDE) to mitigate the cost accrued with increasing number of events by re-using posterior samples in mass and tidal deformability $[m_1, m_2, \Lambda_1, \Lambda_2]$ from single event PE analyses, standard data products of a GW detection. The likelihood of a 4-parameter spectral equation of state represented by parameters $\vec{\gamma}$ given a reparameterised set of parameters of the NS component masses and tides (chirp mass, $\mathcal{M}$, mass ratio q, and tidal parameters $\tilde{\Lambda}$, assuming the posterior distribution of BNS observable parameters is largely independent of $\delta\tilde{\Lambda}$) is given as:

$$\mathcal{L}(h|\vec{\gamma}) \sim \int p(q, \tilde{\Lambda}(\mathcal{M}, q, \vec{\gamma}|h))dq \tag{2.9}$$

which is modelled by a KDE through

$$\mathcal{L}(h|\vec{\gamma}) \sim \int_0^1 K_i(q, \tilde{\Lambda}(\mathcal{M}_i, q, \vec{\gamma}|h))dq \tag{2.10}$$

By substituting the above into equation 2.6, they can evaluate the approximate posterior of equation of state parameters, $\vec{\gamma}$ which models the overall $\mathbf{\Omega}$, given the

prior $p(\mathbf{\Omega})$ and sample the posterior distribution using MCMC. They quote that for 10 BNS events, their analyses takes on the order of ~ 1 day to complete.

Current hierarchical analyses focus on simplifying and accelerating the likelihood calculation of equation 2.6 so that the combinatorial process is not limited by costly likelihood evaluations. Expressing the likelihood using either a KDE or a GMM is an effective solution to quick individual likelihood evaluation but the analyses still scale with number of events and/or posterior samples per event and are limited in their flexibility while remaining computationally inexpensive. The analyses summarised above operated primarily with regards to high density equation of state inference; the methods broadly choose to parameterise the equation of state using phenomenological models, which limit the ability to express complex equations of state. These assumptions are valid within the bounds of the advanced ground-based GW detector network, where numbers of BNS mergers remain low and where measurement accuracy is not expected to be particularly constraining. In next generation detectors like ET [58] or Cosmic Explorer (CE) [59], BNS mergers are expected to be detected at a rate of approximately 10^5 per year [58] and as such we require analysis which can combine information from the 10^5 events within a sensible timescale which is also computationally inexpensive. Additionally it will be beneficial to have tools which tolerate more precise parameter measurement and consistently map macroscopic parameters to the equation of state without losing information due to modelling inaccuracies or generality.

2.3 Future gravitational wave detectors as probes of the neutron star equation of state

Next generation ground based GW detectors ET and CE have proposed sensitivities at least an order of magnitude greater than that of the design sensitivity of the current ground based detectors [58, 59]. The improvement in sensitivity refers to the noise curve, a measure of the magnitude of disruption in spacetime that the detector is sensitive to as a function of the frequency of the GW emission. As a result, it is

expected that the NS equation of state will be constrained to such a level that the NS radius can be determined, through integration of the TOV equations, to within hundreds of metres of uncertainty (on the percent level of the total radius) [190]. Both ET and CE are more sensitive relative to current ground-based detectors but also across a wider band of frequency [59], however the ET/CE sensitivity improvement is not uniform over their frequency bands [191]. In the case of detection of BNS mergers, this also means that more of the inspiral signal is detectable as well as the post merger signal [192]. Figure 2.3.1, from [58] but from analysis completed in [192], presents the noise curves of previous, current and future ground-based GW observatories in strain per $\sqrt{\text{Hz}}$ (the square root of the power spectral density) as a function of frequency. The grey lines associated to different ground-based GW detectors represent the limit of sensitivity of each of these instruments, below which a signal with a given stain per $\sqrt{\text{Hz}}$ will not be observable. A mock-BNS signal at a distance of 100 Mpc is plotted in blue, to demonstrate that, with the current advanced LIGO set-up, we can in principle observe the NS inspiral from the early stages of inspiral (up to 500 Hz) to just before the merger ($\sim$ 2000 Hz). However, with ET, the merger itself and also the post-merger signal will also be detectable.

In the next generation of ground-based GW detection, we expect on the order of 10^5 observations of BNS mergers per year [58] and approximately 10^3 of these mergers to have an EM counterpart [193]. Pipelines for fast and efficient detection [194, 195, 196] and parameter estimation [3, 197, 198, 199] of these GW signals are increasing in their numbers. Efficient identification and inference of the vast number of signals facilitates rapid communication with observers, while also minimises the amount of compute power spent on analysing the many, many signals [200].

Once PE of these observations is performed, rapid equation of state inference would provide EM astronomers with complementary information which could further inform whether an EM counterpart will be observable. Özel and Freire [120] state that the fate of the merger remnant depends on the equation of state, therefore determining whether we expect subsequent emission of EM or GW radiation as the remnant NS settles or complete expulsion of material, leaving behind a remnant BH. The mass

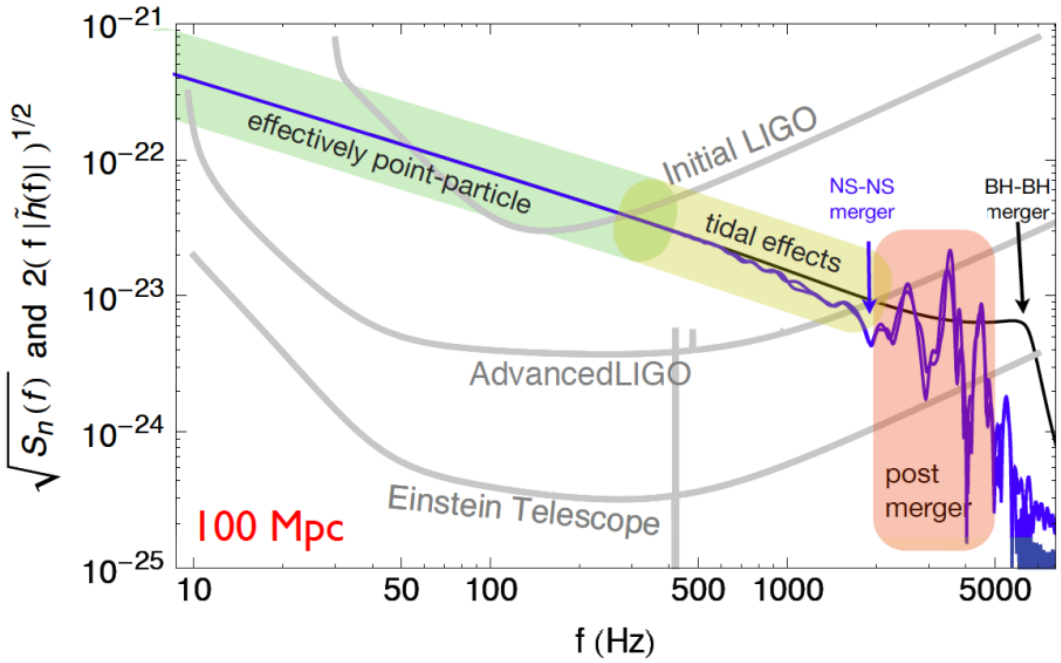


Figure 2.3.1: Plot of detector strain per $\sqrt{\text{Hz}}$ as a function of frequency of GW signal. Noise curves for different ground based observatories are plotted in grey alongside a mock BNS waveform. The point-particle-like stage of the BNS waveform is highlighted in green and the region at which the tidal effect become prominent is highlighted in yellow. The post-merger signal is highlighted in red. The merger and post-merger signal will be observable with ET. From [58].

of the material ejected from the kilonova resulting from a BNS merger also depends on the equation of state [201]; conversely, Zhao et al. [202] explored how kilonova lightcurves can be used to probe the NS equation of state. Qiumu et al. [203] find that softer equations of state have more ejecta material as well as brighter kilonova emission. Therefore, rapid knowledge of the equation of state post-BNS merger could inform astronomers if EM emission will be visible, minimising wasted observation time.

To better understand how next-generation detectors will compare when observing BNS mergers, Chatziioannou [191] found that a GW170817-like event observed with next-generation detectors would have an SNR of 1000. Further, they found that observing 100 sources each with an SNR of 200 would give an uncertainty on (dimensionless) $\tilde{\Lambda}$ of 100, error of $\sim 10\%$, for a 1.6 $M_\odot$ NS when considering the $m - \Lambda$ equivalent representation of the equation of state (see chapter 1 figure 1.3.3 for an

example and typical scales of Λ possible with current equation of state constraint). As measurement precision improves with more sensitive detectors and the number of observations increase drastically, we need to be careful in how we infer the equation of state using measured parameters. With current detectors, quasi-universal relations have been proven to be valid and be an effective method to infer one macroscopic parameter after measurement of another (see section 2.2.1.1). However, in current GW detectors, measurement precision of macroscopic parameters is still very broad, where discrepancy can be hidden. Suleiman and Read [204] find that using quasi-universal relations to extract parameters with next-generation detectors adds additional uncertainty, broader than the uncertainty present with direct inference. As a result, it will be necessary to infer the equation of state directly rather than use equation of state-insensitive relations.

Finally, the uncertainties associated with uninformed crust-core matching for non-unified equations of state become, in the future with next generation detectors, comparable to measurement uncertainty [205]. The assumption of a unique crust equation of state is no longer valid and the uncertainty associated to the choice of crust model may now express itself as an incorrect determination of macroscopic parameters [2]. Therefore unified equations of state, like those introduced in [2], are the necessary choice for equation of state inference with next generation detectors.

2.4 Conclusions

This chapter has introduced the concept of the equation of state: a fundamental relationship between the pressure and density of a NS. We have introduced current methods to model the equation of state phenomenologically, including using the piecewise polytropic parameterisation which will be particularly important when introducing the equation of state data used in chapters 4 and 5. We also introduced unified equations of state, which are particularly important when considering equation of state inference with next generation GW detectors, where the assumption of a fixed NS crust equation of state may no longer be valid.

Not only can we model the $P - \rho$ relationship with phenomenological and theory-informed models, we can use both observation and experiment to inform our knowledge. We have introduced how nuclear experiments may tell us about physical phenomena in the NS crust, near nuclear saturation density, and the nature of the microphysics. Astrophysical observation, including observation of GWs from BNS mergers, is currently our only method of observing the result of interactions of neutron-rich matter at ultra-high densities, several times nuclear saturation densities. The underlying physics of neutron-rich matter at such high densities as those present in a NS is still very much unknown, and our measurement accuracy of NS macroscopic parameters therefore remains limited.

We also briefly summarised some key pieces of work which aim to infer the NS equation of state from GW observations of BNS mergers. Since the observation of the first BNS merger event, GW170817, work has progressed from speculation of the possible levels of equation of state constraint from BNS merger events to how we can simplify the analysis to be non-parametric and unbiased as well as computationally efficient. Equation of state inference with next-generation GW detectors is therefore a challenge with many complex features to be considered, including the drastic increase in number of observations expected. Although detector sensitivity will be drastically improved when compared to current ground-based observatory sensitivity, new problems come to the fore-front of analysis, especially with regards to scale and precision of techniques. Primarily, methods to cope with equation of state inference from multiple events both quickly and efficiently will be necessary. To help solve this issue, we look to Machine Learning methods.

Chapter 3

An introduction to Machine Learning for scientific data analysis

We introduce ML, firstly with basic neural networks and the fundamentals of training and learning, before introducing more complex networks. Primarily, we introduce both Normalising Flows and autoencoders, two types of network which are pertinent for later chapters. We conclude with a literature review of current applications of ML in GW data analysis.

ML is the method of automating computational algorithms to identify patterns and trends within or to emulate large data sets. Once 'trained', i.e. once the algorithm has found its own optimal state in order to understand the data set to a certain degree of confidence, the algorithm can then be used on new data to identify trends and patterns quickly and effectively. The use of ML is now prevalent in a variety of scientific disciplines, including GW data analysis. In this chapter, we give a broad overview of the fundamentals of ML while focusing on its applicability to modern scientific data analysis.

As scientific instruments become more sensitive with capacity for vast amount of data to be stored, traditional methods to process this data become slow and computationally expensive [206]. ML offers a solution in such a scenario; models can

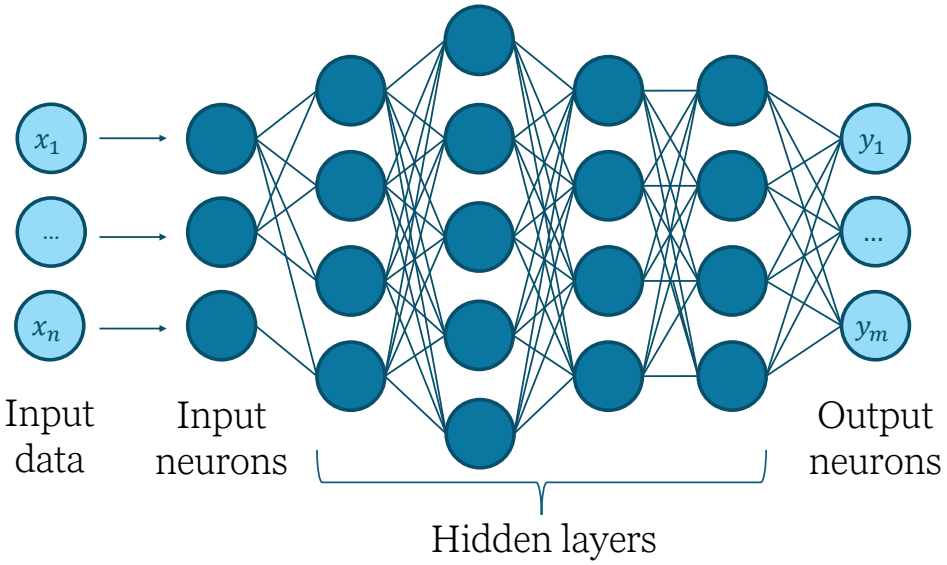


Figure 3.1.1: An example of a fully connected neural network architecture. Input data of n dimensions is passed to n input neurons before passing through a series of hidden layers of different numbers of neurons. Data is finally passed through to m output neurons.

parse vast data sets and interpret not only broad trends but intricate detail within the provided data. Additionally, most off-the-shelf ML tools are built to be compatible with graphics processing units (GPUs) for further parallelisation and acceleration.

Given that tools to apply a Machine-Learning-based method to your given data analysis problem are more abundant and accessible than ever, it is the task of the user to develop a model which is best suited to the task at hand. Various types of ML methods exist, the details of which we will cover in later sections before focussing on two select methods. We begin by introducing the most simple neural network architecture, the multi-layer perceptron, to discuss the fundamentals of building ML models for data analysis.

3.1 An introduction to neural networks

An artificial neural network is an algorithm which is built to be analogous to the human neural network, the series of connections which process information in the human brain. In an artificial neural network, this structure is emulated as a network of

interconnected *neurons*, which are operators on a given input, $\vec{x}$, to produce an output, $\vec{y}$. The simplest deep neural network architecture is that of a multi-layer perceptron (MLP), or fully-connected neural network. In an MLP, neurons are arranged in layers, where all neurons of a given layer are connected to those in the previous and subsequent layers. An illustration of this network architecture is presented in figure 3.1.1. The input layer of an MLP is the input data consisting of n parameters. This input layer then feeds to subsequent layers of different numbers of neurons, often called 'hidden' layers, before finally reaching m output neurons. The structure of an MLP can be designed relative to the problem at hand; more complex networks with more layers and more neurons per layer can perform more complex tasks. Any function can be approximated to any desired accuracy given that you have enough hidden layers within the network [207]. However, as functions increase in complexity, one can imagine the computational cost of building such networks increases significantly. Neural network layers designed to perform more complex tasks, or to take more complex inputs e.g. high dimensionality data sets, without dramatic increase in computational cost are discussed in section 3.1.3.

The computation performed on the j^{th} element of the i^{th} input layer x_{ij} to find the output layer $\vec{y_k}$ of the k^{th} layer, where $k = i + 1$, can be expressed as [206]

$$\vec{y_k} = \sigma\left(\sum_{j=1}^{n} w_{jk}\ x_{ij} + \vec{b_k}\right). \tag{3.1}$$

where we take the sum over all layer inputs, n, and multiply the input by the element w_{jk} of the weights matrix W_k associated to the k$^{\text{th}}$ output layer, length m, given by

$$W_k = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{nm} \end{bmatrix}. \tag{3.2}$$

Each element of W_k is a weight given to the input-output neuron pair, where a larger weight associates more importance to the input neuron and where b_k is a bias vector [208]. We also introduce the *activation function*, σ, which acts on the operation to add complexity and non-linearity [208]. Without the activation function, σ, the MLP

would remain a series of linear computations, regardless of how many layers were in the network.

A selection of activation functions are plotted in figure 3.1.2. Different activation functions operate to serve different purposes; a common function used between layers of neurons in a network being the Rectified Linear Unit (ReLU) [209, 210], $\mathrm{ReLU}(x) = max(0, x)$, a non-linear function which rids the network of negative values during training. This ensures stability and allows the neural network to perform more complex tasks. Activation functions are also often used on the output of the network to restrict the output to be within a given range. For example in a network built to perform binary classification (i.e. true or false) , one might wish to use a sigmoid activation function

$$\mathrm{sig}(x) = \frac{1}{1 + \exp(-x)} \tag{3.3}$$

on the output layer of the network. The result of applying the activation on input $\vec{x}$ is such that $\mathrm{sig}(\vec{x}) = [0, 1]$ represents the probability of the classification given some threshold, say 0.5, where $\mathrm{sig}(\vec{x}) < 0.5$ is false and $\mathrm{sig}(\vec{x}) \geq 0.5$ is true.

3.1.1 Training and learning

Neural networks can be built to serve a variety of purposes relative to the given input data and the desired function of the network. As such, neural networks can be trained using different methods. Three broad categories are:

- *Supervised learning:* N training data inputs x_i where $i = 1, \ldots, N$ are associated to a label or class, θ_i, that the model is then trained to predict or learn the association to. The performance of the model is determined as the ability of the model to learn the correct class relative to the true class, θ_i, in classification, or learn the true value of a given labelled quantity in regression. Once trained, it can classify previously unseen data.

- *Unsupervised learning:* the model is provided with training data inputs x_i and is trained to learn the general behaviour and features of the input data. The model is trained based on its understanding of the input training data set, and

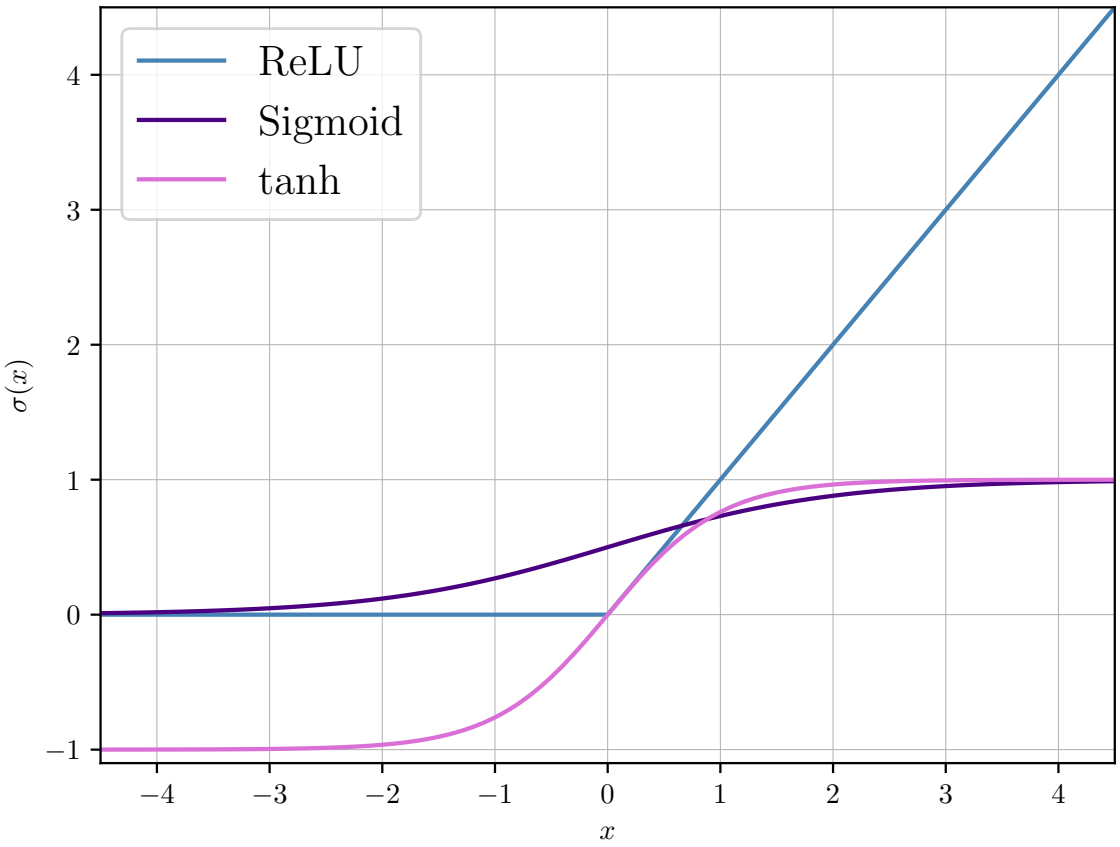


Figure 3.1.2: Example of three common neural network activation functions. Blue: ReLU, purple: sigmoid and pink: tanh.

once trained can be used to recognise patterns or associations among new data, without any introduction of labelling or naming. Examples of unsupervised learning include anomaly detection [211], data clustering [212] and dimensionality reduction, which is discussed in section 3.3.

- *Reinforcement learning:* an neural network, called an agent, is trained to interact with an environment and autonomously learns via trial and error. The agent can freely act within its environment and receives rewards or penalties for actions taken, where the goal of the agent is to accumulate the maximal number of rewards.

In this work we primarily cover supervised learning, and so will lean most heavily on this method in the next sections when discussing more concepts of neural network training.

Any neural network, the MLP included, is trained relative to a loss function which represents the performance of the network during training given its current input and status. The loss is a function of the network input and the neural network trainable parameters (as well as the target $\vec{y}$ in supervised learning), $\mathcal{L}(\vec{x}, (W, \vec{b}))$, and is calculated at the final layer of the network. Like the activation function, it is a chosen function relative to requirements of the network and the method used to train the network. One example of a commonly used loss function is binary cross-entropy (BCE). This is particularly useful for binary outputs from a network i.e. 'true' or 'false' like our previous example of a binary classification network. In this example network, θ_i is the true label associated to the i^{th} network input x_i, where $\theta_i \in (0, 1)$ and y_i is the predicted probability of the label of x_i from the network where $y_i = [0, 1]$. Therefore, the BCE loss is

$$\text{BCE} = -\frac{1}{N}\sum_{i=1}^{N} \theta_i \log(y_i) + (1 - \theta_i) \log(1 - y_i) \tag{3.4}$$

where we take the average BCE over N instances of training data [213]. As the BCE loss decreases, the closer the model's prediction y_i is to the true label θ_i.

The goal of training the neural network is to minimise the loss; this means that, in supervised learning, the output of the network is comparable to that of the given label or class that we are comparing to. Therefore, training a neural network is an optimisation problem of finding the parameters of the loss function (the optimal weights and biases) which minimise the loss. This optimisation routine is back-propagation [214]: we initialise the neural network with randomly selected weights w_{ij} and then update the weights throughout training with the goal of minimising the loss. As with regular function minimisation, the minimum of the loss function corresponds to a derivative of 0. We calculate the loss throughout training and its derivative with respect to all of the weights, W, in the network. The computational intensity of this calculation (noting we likely have millions of weights w_{ij} in our network) is lessened by back-propagation's chain rule calculation of the derivative of the loss with respect to all weights before updating them [215]. We therefore update the weights in such a way as to decrease the loss. This optimisation procedure is called gradient descent; we calculate the gradient of the loss with respect to the neural network parameters

continually throughout training, moving in the direction of the lower gradient, lowering the loss, aiming for the global minimum.

Updating the weights to reflect the update of the network output down towards lower losses is called a *backward pass*, whereas the preceding pass of the training data through the network to calculate the loss is the *forward pass*. The magnitude of step taken in the loss function is determined by the *learning rate*; large learning rates correspond to large steps through the neural network hyperparameter space, potentially avoiding fine features of the space. Alternatively, small learning rates may get stuck in local minima, never fully finding the truly optimal network configuration. As such, the learning rate is a hyperparameter which can be optimised for when testing network architecture, much like the number or depth of layers.

One full pass of training data is generally classed as an *epoch*, but this is not often done in one pass, as this becomes computationally expensive with lots of data, or impossible relative to memory constraints. Instead, the training data is often split into *batches*, and the gradient is calculated, back-propagated and the network optimised to minimise loss relative to each batch. The final loss after one epoch is the average loss over all batches; see equation 3.4, where we take the average over N instances of training data. This is repeated for N_b batches and the average loss over N_b batches found, however the network has been updated multiple times during each epoch relative to the decisions made after each batch. All together, this process is titled *stochastic gradient descent*, 'stochastic' referring to a random selection of data for each batch from the training data set [208]. Different optimisation procedures exist [216, 217] but we will focus on stochastic gradient descent for all following related discussion.

The network is trained over many epochs, so that the average loss per epoch gradually decays over training through optimisation of the weights and biases before plateauing, signifying that the network's potential for learning is saturated. This is an idealised scenario; the loss does not always decrease monotonically and fluctuations in the gradient can make for a noisy average loss per epoch. Additionally, the average

loss per epoch that is reached in saturation may not represent the global minimum, but may be a restriction of the depth and complexity of the network hyperparameters.

To summarise, a simple neural network is trained as follows: a forward pass of the training data through each of the layers within the network to the output is performed to calculate a loss, $\mathcal{L}(\vec{x}, (W, \vec{b}))$, at the output of the neural network, relative to a known truth or target if supervised. The gradient of this loss is calculated via back-propagation, and a backward pass updates the weights and biases in the network so to minimise the loss function. This process is repeated for a number of epochs, minimising the loss until training is completed.

3.1.2 Training data

The main ingredient for training a neural network, besides the network itself, is of course the training data. The number of input neurons of the network is typically associated to the number of parameters or attributes of our input data. In contrast, the number of output neurons is then typically associated to the goal of the neural network. For example, the previous network we proposed to perform binary classification may have a single neuron at the output, which is trained to output the classification probability which is a single number between $[0, 1]$.

Standard practice with a data set for neural network training is to split the training data set into three portions: training, validation and testing. The majority of the data is used for training the network as described above, with two smaller fractions retained for *validation* and *testing*. The entire training and validation sets are passed to the network independently on each epoch, with similarly independent training and validation losses calculated as a result. Unlike the training loss, the calculated validation loss does not influence the back-propagation stage and so does not inform the update of the trainable parameters of the network. This allows the validation data to serve as an independent portion of the data which only tests the network's performance on data from the same distribution but previously-unseen. This ensures the network is not *overfitting*, or memorising the specific details of the training data set. The testing data set is then used once the network has been trained and optimised

to test its performance on data that, again, is from the same overall distribution but has not been used in training or validation. This ensures we are testing the network on data that we know is in agreement with the overall distribution of training and validation data, but is independent.

It is also necessary to have an appropriate quantity of training data so to present the network with enough examples of what we would like it to learn. This quantity, again, is a function of the complexity of the problem. High-dimensional and complex problems demand lots of training data so to provide enough example data points within the training data space to provide a broad picture of the information we would like the network to learn. We also wish to present the network with enough data so that the network *generalises* well; we wish for our network to learn the general trends and patterns of the training data and associated classes. If too little data is used for training, the network runs the risk of learning individual features (for example, clusters of data points or local modes) or even memorising data points. Memorising individual features or data points would present as *overfitting* in the calculation of training and validation losses; training loss would continue to decrease as the network finds parameters which are capable of building a network which emulates the training data perfectly, while the validation data, an independent data set, is an outlier and so the validation loss instead increases. The result is a network which then struggles to generalise and which cannot perform well on new and previously unseen data.

3.1.3 Different types of neural network

An multi-layer perceptron (MLP) is one simple example of a neural network, however different network architectures can parse more complex input data with less computational expense. Two examples are:

Convolutional neural network: Commonly used with 2- or 3-dimensional image data sets, the input image is convolved through the use of a *kernel*, which is a filter which is applied to the (for the case of our example here) 2-dimensional input of shape $d \times d$. The kernel, a matrix of size $l \times l$ where $l < d$, passes over the image, where the elements of the matrix are multiplied by the pixel values in the image

[218] (the kernel does not have to be square, but we choose this for the sake of our example). The choice of the matrix elements within the kernel define the function of the kernel, but in the case of image classification, the kernel often functions to identify the key features of the image. Once features are identified, the resulting *feature map* [219, 218] is a representation of the image with features which have been amplified by the kernel highlighted. The goal is to abstracts the input image gradually through multiple convolutional layers so that the network learns the key features through which to classify or perhaps reconstruct the input image [220]. Convolutional networks often include some convolutional and some fully connected layers, so to control the dimensionality of the output of the network [221], alongside stages of *max pooling*, which acts to downsample the image via scanning a similar kernel to that described above across the image or feature map. Instead of highlighting features, the max pooling stage selects the maximum pixel value and reduces the size of the original image or feature map to a smaller output image, but one which still retains the main features.

Recurrent neural network: Designed for use with sequential training data, where the network is provided with the next piece of the sequence at a time [222]. The recurrent network includes, alongside its inputs and outputs, a 'memory' allowing it to retain information from previous stages of the series, identifying sequential behaviour [223, 224]. As well as being intuitively suited to time series data, where recurrent networks can be used to predict behaviour, they often also find use in language processing [225], where the series of language input can either be interpreted such that the network can predict the next steps or 'respond'. One variant of a recurrent network is a long short-term memory network [226], which adapts upon standard recurrent neural networks but with an improved short-term memory, allowing the network to identify patterns over longer sequences with better behaviour in training [227].

3.2 Normalising Flows

Our earlier example of a neural network designed to perform binary classification is an example of a network designed to perform a *discriminative* task; in this case, assigning the output y_i of the network a probability between $[0, 1]$. However the next network we introduce falls into the category of *generative* ML where a network instead learns a its own mapping of the training data input, otherwise referred to as the latent space $\mathcal{Z}$, which can then be sampled from. These new samples are from the same statistical distribution of the samples which were used to train the model but are entirely independent samples from the distribution, hence 'generative'. Examples of generative models include generative adversarial networks (GAN)s [228] and diffusion models [229].

A Normalising Flow is a type of generative ML model which is trained to perform both sampling and density estimation of potentially complex distributions [230]. To do so, the model is trained to perform the transform from latent space, $\mathcal{Z}$, described by a standard Gaussian probability distribution, $p_Z(\mathbf{z})$, to a more complex distribution, $p_X(\mathbf{x})$ in the data space $\mathcal{X}$. The transformations can be considered to be steps which are taken to evolve the Gaussian latent space to the more complex data distribution. The overall probability density of the spaces must be preserved in transforming from data to latent, and the reverse. Therefore we require the transformations to be *invertible* and *differentiable* [231]. The transformations are performed by neural networks parameterised by hyperparameters ϕ, where $x = f^{-1}(z, \phi)$ is the forward transform and $z = f(x, \phi)$ is the inverse transform for $\mathbf{x} = \{x_1, ..., x_n\}$ where $x_i \in \mathbb{R}$ and similarly $\mathbf{z} = \{z_1, ..., z_n\}$ where $z_i \in \mathbb{R}$.

In the case where we wish to perform a conditional transform from the data space $\mathcal{X}$ to the latent space $\mathcal{Z}$ given some complementary information - for example, a label or class, $\boldsymbol{\theta}$ - we use *conditional* Normalising Flows. The probability density of $\mathbf{x}$ given conditional information $\boldsymbol{\theta}$ can be expressed as a change of variables of two probability density functions [231]

$$p_{\mathcal{X}|\Theta}(\mathbf{x}|\boldsymbol{\theta}) = p_{\mathcal{Z}}(f(\mathbf{x}, \phi|\boldsymbol{\theta})|\boldsymbol{\theta}) \left|\det\left(\frac{\partial f(\mathbf{x}, \phi|\boldsymbol{\theta})}{\partial \mathbf{x}}\right)\right|, \tag{3.5}$$

where we are taking the absolute value of the determinant of the Jacobian, an $n \times n$ matrix given by

$$\frac{\partial f(\mathbf{x}, \phi)}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \cdots & \frac{\partial z_1}{\partial x_n} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & \cdots & \frac{\partial z_2}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial z_n}{\partial x_1} & \frac{\partial z_n}{\partial x_2} & \cdots & \frac{\partial z_n}{\partial x_n} \end{bmatrix}. \tag{3.6}$$

The Jacobian measures the change in volume from data to latent space and ensures the transformations are tractable. Therefore we can act in the simpler latent distribution, which is easy to sample from and evaluate the probability of, and then perform the inverse transform $\mathbf{x} = f^{-1}(\mathbf{z})$ to return samples from the target space $\mathcal{X}$ where $x \sim p_{\mathcal{X}}(\boldsymbol{x})$. Therefore the Jacobian determinant must be tractable and the neural network transforms must be invertible, both features which make Normalising Flows unique.

Given that the individual neural network transforms are invertible and differentiable, a *series* of such transforms will also have these properties [231]. Therefore, we can build more complex Flow models which overall perform more complex transforms by constructing the network from a series of k simpler transforms,

$$\boldsymbol{z}_k = f_k(x_k, \phi|\theta) \ \cdot \ f_{k-1}(x_{k-1}, \phi|\theta) \ \cdot \ \ldots \ \cdot \ f_1(x_1, \phi|\theta), \tag{3.7}$$

where $x_1 = \boldsymbol{x}$ and $x_k = f_{k-1}(x_{k-1}, \phi|\theta)$. The change of variables via the Jacobian determinant are repeatedly applied and so the requirements of invertibility and differentiability are upheld [232]. Repeated transforms and therefore repeated calculation of the change of variables means that we require a Jacobian determinant which is easy to compute [230]. We consider two main families of Normalising Flow (Flow) transforms in this work which have easy to compute Jacobian determinants, real non-volume preserving (RealNVP) and spline Flows.

3.2.0.1 Real non-volume preserving Flows

RealNVPs [233] are one type of affine transform, which transform the data space location x_i given scaling parameters βi and scale α_i

$$f(x_i, \alpha_i, \beta_i) = \alpha_i x_i + \beta i \tag{3.8}$$

where $\alpha \neq 0$ to preserve invertibility [231] in the i^{th} transform. To enforce this, we set $\alpha_i = \exp \alpha_i^*$ where α_i^* is not constrained such that

$$f(x_i, \alpha_i, \beta_i) = \exp(\alpha_i^*) x_i + \beta_i. \tag{3.9}$$

This transform is invertible, in that we can rearrange for x_i by performing the inverse transform $f^{-1}(z_i, \alpha_i, \beta_i)$. In both the forward and inverse transforms, the parameters which control scale and translation, α_i^* and β_i, remain the same and therefore are not inverted. These parameters can therefore be as complex as we like, and can be modelled by neural networks described by parameters ϕ. The Jacobian of the transform of equation 3.9 as defined by 3.6 is triangular such that each diagonal element is

$$\log \left| \det \frac{\partial f(x_i, \phi)}{\partial \mathbf{x}} \right| = \log |\alpha_i| = \alpha_i^*. \tag{3.10}$$

These transforms are, on their own, limited in there expressivity. Transforming a simple Gaussian distribution $p_{\mathcal{Z}}(z_i)$ via a single transformation like that of equation 3.9 would return a scaled but ultimately similarly Gaussian distribution [231]. Therefore, stacking multiple layers of affine transforms as in 3.8 as *affine coupling layers* [233] introduces more complexity, while the individual calculations themselves remain very simple: referring back to equations 3.7 and 3.10, this is simply

$$\begin{aligned} \log \left| \det \Big(f_k(x_k, \phi|\theta) \ \cdot \ f_{k-1}(x_{k-1}, \phi|\theta) \Big) \right| &= \log \left| \det \Big(f_k(x_k, \phi|\theta) \Big) \cdot \det \Big(\ f_{k-1}(x_{k-1}, \phi|\theta) \Big) \right| \\ &= \log \left| \det \Big(f_k(x_k, \phi|\theta) \Big) + \det \Big(\ f_{k-1}(x_{k-1}, \phi|\theta) \Big) \right| \\ &= \sum_k \alpha_k^* \end{aligned} \tag{3.11}$$

where the log Jacobian determinant is just a sum of the k diagonal terms, corresponding to the k-dimensional data and latent spaces.

3.2.0.2 Spline flows

Another method is to represent transforms as a series of monotonic spline functions [234, 235, 236]. This method leans on the possibility of having any function f to perform the transform in equation 3.5 given it is easily invertible and is differentiable.

Examples of spline or piecewise functions used include cubic splines [236], piecewise polynomials [235] and neural splines [234].

In the cubic spline case [236], each of the K segments of the splines which describe the transform f are monotonically increasing cubic polynomials. By defining $K+1$ points in the data and latent spaces $\{(x_k, z_k)\}_{k=0}^{K}$ as the joining pieces between the K segments, one can construct a differentiable cubic spline which passes through these points and is monotonic between points. The spline function is constructed of K piecewise segments, where each segment is a simple function which meet at define start x_k and end x_{k+1} points. A cubic spline transform can be expressed as

$$f_k = \alpha_{k_0} + \alpha_{k_1}(x - x_k) + \alpha_{k_2}(x - x_k)^2 + ... + \alpha_{k_K}(x - x_k)^K \tag{3.12}$$

such that the derivative is

$$\frac{df_k(x - x_k)}{dx} = \alpha_{k_1} + 2\alpha_{k_2}(x - x_k) + ... + (K+1)\alpha_{k_K}(x - x_k)^{K-1}. \tag{3.13}$$

Via equation 3.6, the log of the absolute value of the Jacobian for a series of transforms described by equation 3.13 is the sum of the logarithms of equation 3.13, akin to equation 3.11. Evaluating the data space $\mathcal{X}$ at location x involves finding the segment in which x lies and computing the Jacobian determinant of equation 3.13.

3.2.1 Training a normalising flow

We train a Normalising Flow to perform the *forward mapping* $z = f(x, \phi|\theta)$ from the more complex data space $\mathcal{X}$ to Gaussian latent space $\mathcal{Z}$. Likewise with the MLP, we train the flow to minimise a loss which is the measure of the difference between the Flow's learned data distribution $p_{\mathcal{X}}^{\text{flow}}(\boldsymbol{x}, \phi|\theta)$ and the true target distribution $p_{\mathcal{X}}^{\text{true}}(\boldsymbol{x})$. The difference between two probability distributions is measured by the KL divergence, such that

$$D_{\text{KL}}(p_{\mathcal{X}}^{\text{true}}(\boldsymbol{x})||p_{\mathcal{X}}^{\text{flow}}(\boldsymbol{x}, \phi|\theta)) = \int p_{\mathcal{X}}^{\text{true}}(\boldsymbol{x}) \log \frac{p_{\mathcal{X}}^{\text{true}}(\boldsymbol{x})}{p_{\mathcal{X}}^{\text{flow}}(\boldsymbol{x}, \phi|\theta)} \, dx. \tag{3.14}$$

In training, the Flow's probability distribution $p_{\mathcal{X}}^{\text{flow}}(\boldsymbol{x}, \phi|\theta)$ is attempting to model $p_{\mathcal{X}}^{\text{true}}(\boldsymbol{x})$ as accurately as possible. The KL divergence quantifies the difference in

information enclosed within each distribution and is measured in nats. If the two distributions are identical, the amount of information enclosed is the same and $D_{\mathrm{KL}} = 0$. We wish to minimise the KL divergence between $p_{\mathcal{X}}^{\mathrm{true}}(\boldsymbol{x})$ and $p_{\mathcal{X}}^{\mathrm{flow}}(\boldsymbol{x}, \boldsymbol{\phi}|\theta)$ such that

$$D_{\mathrm{KL}}(p_{\mathcal{X}}^{\mathrm{true}}(\boldsymbol{x})||p_{\mathcal{X}}^{\mathrm{flow}}(\boldsymbol{x}, \boldsymbol{\phi}|\theta)) = \left\langle \log p_{\mathcal{X}}^{\mathrm{true}}(\boldsymbol{x}) \right\rangle_{x\sim p_{\mathcal{X}}^{\mathrm{true}}(\mathbf{x})} - \left\langle \log p_{\mathcal{X}}^{\mathrm{flow}}(\boldsymbol{x}, \boldsymbol{\phi}|\theta) \right\rangle_{x\sim p_{\mathcal{X}}^{\mathrm{true}}(\mathbf{x})} \tag{3.15}$$

where the angle brackets take the Monte Carlo sum for the average probability over samples from the target distribution $p_{\mathcal{X}}^{\mathrm{true}}(\boldsymbol{x})$ given the target and Flow distributions respectively. The first term of equation 3.15 is constant, as the target probability distribution is static and is the probability distribution which we are training the flow relative to, and so can be neglected. As we train the Flow, the second angle bracket term evolves and we incorporate this into our loss

$$\mathcal{L} = D_{\mathrm{KL}}(p_{\mathcal{X}}^{\mathrm{true}}(\mathbf{x})||p_{\mathcal{X}}^{\mathrm{flow}}(\mathbf{x}, \boldsymbol{\phi}|\theta)). \tag{3.16}$$

Given equation 3.15 and neglecting the first angle bracket term, we find

$$\mathcal{L} \approx -\frac{1}{N}\sum_{i=1}^{N} \log p_{\mathcal{X}}^{\mathrm{flow}}(\mathbf{x}, \boldsymbol{\phi}|\theta). \tag{3.17}$$

where the Monte Carlo sum is approximated by finding the average loss over N discrete training data points from the data distribution for $x_i \in \{\boldsymbol{x}\}$ where $i = (1, ..., n)$. By taking the logs of both sides of equation 3.5 and then substituting in, we find

$$\mathcal{L} \approx -\frac{1}{n}\sum_{i=1}^{n} \log p_{\mathcal{Z}}(f(x_i, \boldsymbol{\phi}|\boldsymbol{\theta})|\boldsymbol{\theta}) \; + \log\left|\det\left(\frac{\partial f(x_i, \boldsymbol{\phi}|\boldsymbol{\theta})}{\partial \mathbf{x}}\right)\right|. \tag{3.18}$$

We minimise the KL divergence via back-propagation, as introduced in section 3.1.1, optimising the weights and biases of the network hyperparameters. The inclusion of the conditional term, $\boldsymbol{\phi}$, does not affect the optimisation of the network hyperparameters, $\boldsymbol{\phi}$. Once trained, the Flow then allows us to perform two main operations:

- *Sampling:* Sampling points in the data space $\mathcal{X}$ involves sampling points from the latent space $\mathcal{Z}$ where $z \sim p_{\mathcal{Z}}(\boldsymbol{z})$ and then performing the inverse transform $x = f^{-1}(z, \boldsymbol{\phi})$ to return samples $x \; p_{\mathcal{X}}(\boldsymbol{x})$.

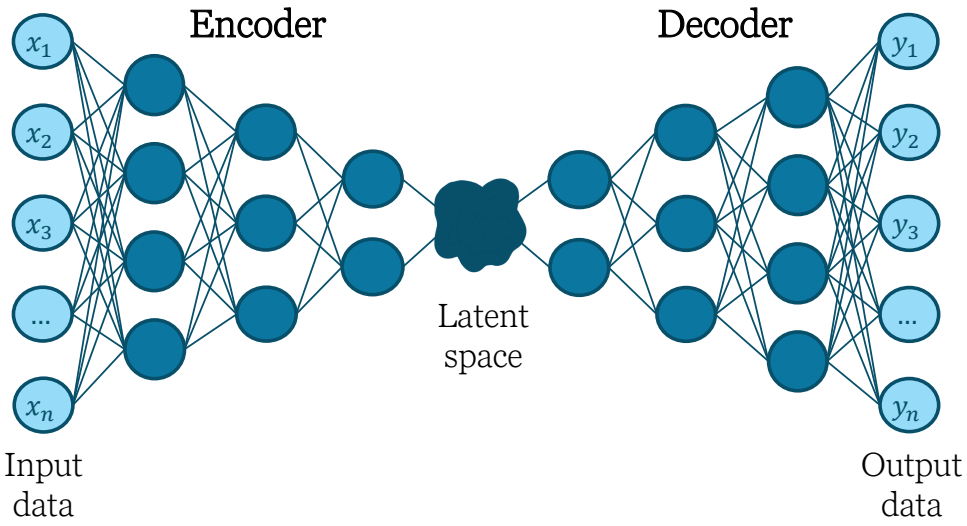


Figure 3.3.1: Input data with n attributes is passed into an encoder with initial layer size of n neurons. The number of neurons in subsequent layers decreases to a reduced-dimensionality latent representation of the input. The decoder then reconstructs the latent representation to an n dimensional output.

- *Evaluating the probability density of points:* evaluating the probability density of a point x in the data space $\mathcal{X}$ requires evaluating $p_{\mathcal{X}}(x, \phi|\theta)$. This is found via equation 3.5, where we compute the inverse transform $(f(\mathbf{x}, \boldsymbol{\phi}|\boldsymbol{\theta})|\boldsymbol{\theta})$ and the Jacobian determinant.

The two methods of *sampling* and *evaluating* with a Flow model are different operations in the same space. Clarifying the differences between these operations is particularly important for discussion in chapter 5.

3.3 Autoencoders

An autoencoder is a ML model which is a variant of the traditional fully connected neural network. A standard autoencoder consists of two main parts, an encoder and a decoder, where each are typically mirrored architectures of each other. This is illustrated in figure 3.3.1: the encoder takes training data of n parameters as n input neurons and condenses the data to a lower dimensionality latent space representation

on the output. The decoder then accepts points from the latent space as input and decodes the information to the same dimensionality as the input training data to the encoder. The goal of a standard autoencoder is to reconstruct the training data given as input as accurately as possible in a bottleneck-like workflow. The network, similar to previous discussion, is trained to minimise the loss; in this case it is the error on the output of reconstructing the input data. This is given as the mean-squared error loss [237] of a batch of training data

$$\text{MSE} = \frac{1}{D}\sum_{i=1}^{D}(y_i - x_i)^2 \tag{3.19}$$

where x_i is the i^{th} training data sample in a batch and y_i is the output of the autoencoder given the x_i, averaged over D total instances in a batch of training data. The loss of each data sample is averaged over all the dimensions of the input/output, returning a singular loss. In passing a batch of data to the autoencoder, we further average the loss. As discussed previously, the loss is back-propagated in order to optimise the weights and biases of the network to allow the network to reconstruct the input data as accurately as possible.

With linear fully-connected layers, an autoencoder essentially is an MLP with a bottle-neck-like structure designed for data compression. However, it is simple to incorporate layers of different architectures, for example convolutional or recurrent layers. This then makes the autoencoder suited to image compression or anomaly detection [211] of time series, for example.

Once trained, the latent space of a standard autoencoder can be used as a reduced dimensionality representation of the input, which non-linearly encodes the key information to be able to uniquely reconstruct any piece of data. Due to this, autoencoders are effective tools for simple data compression, due to their flexibility and the presence of non-linear activation functions. However, sampling any region of the autoencoder latent space without knowledge of its boundaries, structure or general shape may result in decoding regions of the space which returns an invalid or nonsense output. For this reason, an simple autoencoder is not by-nature generative.

A *variational autoencoder*, however, combines the generative mechanism with the autoencoder architecture [238]. Instead of training the encoder to output a latent space vector, a variational autoencoder is trained to output the mean μ and standard deviation σ of a Gaussian distribution $\mathcal{N}(\mu, \sigma)$. During training, the Gaussian that is defined by the latent outputs of the encoder is then sampled from $z \sim \mathcal{N}(\mu, \sigma)$ and the sample z is then passed to the decoder, as in a regular autoencoder, to be decoded. The loss function of a variation autoencoder is composed of two parts: the MSE quantifying the reconstruction error and the KL divergence between the learned latent distribution and a unit Gaussian

$$\mathcal{L}_{\text{VAE}} = \frac{1}{D}\sum_{i=1}^{D}(y_i - x_i)^2 + D_{KL}(\mathcal{N}(\mu, \sigma)||\mathcal{N}(0, 1)). \tag{3.20}$$

The latent space of a trained autoencoder is therefore Gaussian (or very close to this) such that when one encodes a series of data points to the autoencoder latent space, the latent space appears Gaussian, where similar inputs are encoded to similar regions of the autoencoder latent space. In a regular autoencoder, there is no restriction on the construction of the latent space by design, and so similar inputs may likewise cluster, but not in any defined way.

A conditional variational autoencoder is a further extension to this, where the latent space is again trained to be Gaussian, but latent vectors are encoded given a conditional label y [239]. The latent distribution is then sampled relative to this condition $z \sim \mathcal{N}(\mu, \sigma|y)$ and the sample z is decoded. While this remains a conditional generative model, a conditional variational autoencoder is not tractable [240] like a Normalising Flow, and one is not mapping from input to learned latent representation and back, but rather in one direction through two network structures (encoder and decoder), introducing more sources of error. Additionally, the conditional variational autoencoder latent space is limited to be a parameterised probability distribution with mean μ and standard deviation σ, or some other similarly parameterised distribution [241] . This can be restrictive for problems which are better suited to more complex or multi-modal latent representations.

3.4 Current applications of Machine Learning in gravitational wave science

We focus our review of the broader use of ML for data analysis particularly on GW data analysis and instrument science. Over the past decade, many ML pipelines have been introduced to either support and accelerate or offer a completely new method of performing GW data analysis. We discuss some key challenges of GW detection and analysis and introduce some ML techniques which aim to combat this.

GW detection is limited by fundamental sources of noise which are inherent to the detector design and hardware (discussed in chapter 1 section 1.2). Also limiting are the presence of detector glitches, bursts of noise in the detector which can appear signal-like or which crop-up at regular intervals with no known source [242]. These instances of noise have magnitudes comparable to that of the desired signal and appear in the detector's sensitivity band [243], impacting real signal detection. Gravity Spy [244] is a pipeline which utilises a data set provided by 'citizen scientists' - volunteers who label 'images' of detector glitches relative to some provided examples. The result is a labelled training data set of GW detector glitches in the time-frequency domain [245]. This data set is used to train a convolutional neural network which classifies glitches present in the GW detector. Once trained, the tool can be used to quickly sift through instances of excess power measured in the detector and to then identify glitches and their periodicity. Recent glitch investigations using other ML techniques including autoencoders have also investigated whether auxiliary channels of the interferometer (channels which record the behaviour of sub-systems of the interferometer, rather than the strain h) can improve our understanding of glitch sources and common theses amongst glitches [246].

Different GW signal types have their own challenges related to their detection. CBC signals are the only signal type to be detected thus far. As discussed in chapter 1 section 1.2, the general signal types - CBC, burst, continuous and stochastic - fall into two broad categories: modelled and unmodelled signals. Burst GWs are short-duration and are unmodelled due to the unknown or complicated astrophysical

processes associated to their source. The uncertainty associated to a signal of this type makes it difficult to confidently identify such a signal relative to noise. The burst low-latency pipeline MLy of [247] also uses a convolutional neural network to perform real-time analysis of unmodelled burst GWs from, for example, core-collapse supernovae and cosmic strings. Low latency analysis benefits from use of ML models due to rapid processing of large amounts of data and identification of generic patterns within said data. The convolutional network identifies generic signal morphology of the unmodelled signal while an additional statistic quantifying cross-correlation amongst ground-based detectors improves confidence in the result. The pipeline loosens strict requirements of well-modelled identification of signal and instead focusses on identification of excess noise, generic morphology and correlation amongst detectors. Their tool has competitive sensitivity to standard burst pipelines [248] and can operate on suitable timescales for low latency (around 1 second).

In contrast, CBC signals have been detected in vast numbers (at the time of writing, the LIGO-Virgo-KAGRA collaboration had declared more than 200 confident detections of GWs from CBCs [39]). Commonly, matched filtering techniques [249] are used to identify CBC signal amongst noise, but this requires a very large bank of template waveforms of all potential GW signals from sources with a variety of properties. As our detection of CBC signals becomes more sensitive and extends to broader regions of or makes for a higher dimensionality parameter space, especially in next generation ground-based detectors, matched filtering becomes a more expensive method for detection, as our template banks need to therefore increase in size and/or dimension. An application of ML to improving matched filtering has already been attempted [250]. Alternatively, full ML pipelines have been developed to have lightweight and rapid identification of CBC signal (once the network is trained) [251, 252, 253, 254, 195, 194]. The different pure ML methods developed vary in their details, but often suffer from biases which inhibit generalisation and therefore detection quality over the broad prior range of all potential GWs observed from CBCs [194].

Parameter estimation investigates the properties of the source which produced the GW signal. Pipelines like Bilby [255] and LALInference [**?**] calculate the posterior

probability of source parameters, for example masses and spins, through evaluation of a Bayesian likelihood. ML techniques for parameter estimation aim to speed up existing sampling methods (for example [256]) but also to complete a whole independent analysis [257, 258, 259]. The authors of [199] use a Normalising Flow to return a posterior distribution on parameters of a BBH merger in seconds, compared to traditional analysis time scales of ~hours. Their Flow is trained to map parameters of a GW signal to a Gaussian latent space conditioned on an embedding of the simulated waveform that is associated to those parameters with added detector noise. While training is expensive (~ 10 days on NVIDIA A100), PE can be repeatedly performed for individual events after this, with very quick turn-around time of around 20 s per GW event. This workflow has recently been extended to PE of BNS mergers in [258] with similar work completed in [3] including extension to neutron star equation of state (EOS) inference.

Astrophysical interpretation of cosmological parameters [260, 261] and GW population properties [262, 263, 264] also make use of ML tools, which broadly speed up not only individual analyses but also hierarchical methods [265]. Population analyses suffer from large computational cost due to handling whole populations of (simulated) GW observations - a cost which only increases as, again, detectors improve in sensitivity and we observe more events over a broader parameter space. ML techniques are particularly useful here, introducing flexibility in the ability to emulate a complex distribution of population and cosmological parameters at a considerably reduced computational cost. Once modelled, we can then which would be otherwise costly to evaluate.

Next generation detection and analysis of GW signals will need to consider improved detector sensitivity meaning longer signal duration [59] and a larger number of observed signals [183], with the potential of not only identifying new signal types but also of uncovering new physics. Investigations into how ML can help with such analyses are already underway [266, 267, 268, 269, 270].

3.5 Conclusion

We have introduced ML methods and how these may be applied to scientific data analysis. ML methods in the era of modern scientific discovery, with vast quantities of data and multi-messenger observation, offer efficient and rapid solutions. We introduced the fundamental building blocks of neural networks and what it means to train one - and how to train one well. This involves crafting a network which has appropriate structure relative to your problem, having plenty of training data to work with, and then optimising the network relative to figures of merit (primarily the training loss).

We also introduce two important network types: Normalising Flows and autoencoders. Normalising Flows are a generative ML model used for density estimation of complex data distributions. The neural network that underpins a Flow is trained to perform the forward mapping from complex multi-dimensional data space to simple Gaussian latent space of the same number of dimensions. By quickly and easily sampling from a multi-dimensional Gaussian, one can perform the inverse mapping using the Jacobian determinant to then return samples from the complex data distribution. We highlighted *conditional* Normalising Flows in particular, as the association of label to complex data samples will be particularly important in later chapters.

We introduce autoencoders next, a neural network with a bottle-neck like structure, trained to pass input data through a smaller dimensional latent space and out again, so to reconstruct the input data. These networks are adept at both anomaly detection and dimensionality reduction, which we will cover in particular in later chapters. We finally summarise the uses of ML in the field of GW data analysis. Not only has ML been prominent in the development of new tools to solve otherwise computationally intensive problems, it has additionally supplemented existing pipelines to accelerate analyses. In the next chapter, we demonstrate how a Normalising Flow can allow us to infer the neutron star equation of state very quickly using PE samples from observations of binary neutron star mergers. This is a very quick and model-agnostic tool while additionally allowing us to perform the inverse problem of inferring the equation

of state from individual neutron star mass and tidal deformability measurements. In later chapters, we demonstrate how an autoencoder can compress complex and highly-detailed neutron star equation of state data before applying the Flow method to hierarchical inference of the neutron star equation of state.

Chapter 4

Rapid neutron star equation of state inference with Normalising Flows

Observations of CBC signals containing at least one NS were, prior to the first Advanced LIGO and Virgo observing runs, thought to be the most likely candidates for detection [271], with the LIGO-Virgo collaboration predicting $0.0004 - 3$ BNS detections during the first observing run, assuming 'optimistic astrophysical rates' [272]. As discussed in chapter 2, prior to the detection of any GW signal, much effort was devoted to anticipating just how well we could constrain the equation of state of high density nuclear matter with incoming GW signals from BNS mergers, and how many signals would be necessary to do this to a significant degree. Much of this work was based on numerical simulations of BNS mergers and then subsequent Bayesian analysis assuming steady rates of observation of BNS mergers in the observing run of advanced LIGO discussed in chapter 2 section 2.2.1.

The first observing run [53] instead saw two BBH mergers, and the improvement in sensitivity of the Advanced LIGO and Virgo detectors prior to the second observing run [273] brought 3 more BBH signals, along with the first highly anticipated BNS merger [54]. As introduced in chapter 1, GW170817 was identified through matched filtering of the observed strain data with known template gravitational waveforms. The signal remained in the sensitive frequency range of the detectors for more than a

minute before eventual merger, with an SNR of 32.4 - the loudest GW event observed at the time [83]. The operation of a three-detector network at the time of observation (both Advanced LIGO detectors in North America as well as Advanced Virgo) allowed for precise sky localisation of the event - an area of 28 deg^2 - providing EM astronomers with the opportunity to search for associated counterpart signals. GW170817 was followed by a short gamma-ray burst (GRB) [88, 89] alongside emission in the x-ray, ultraviolet, optical and radio bands [176] originating from the host galaxy NGC 4993 [274]. The breadth of observations of this single event across different wavebands supported the existing evidence provided from the gravitational wave signal that GW170817 was the merger of two NSs.

PE of the GW signal, discussed in chapter 1 section 1.4.1, uncovered individual masses of the components to be $m_1 \in (1.36, 1.60)\ M_\odot$ and $m_2 \in (1.17, 1.36)\ M_\odot$ assuming spins of the component stars are low ($\chi \leq 0.05$), with total mass of the system $2.74^{+0.04}_{-0.01}\ M_\odot$, both measurements to 90% confidence [54]. These align with current measurements of NS masses [120], mostly dominated by radio observations of galactic NSs (discussed in chapter 1 section 1.4.2), and further suggest that the two components of the binary are indeed NSs. However, in the announcement of GW170817 [54], the authors do not exclude the potential of GW170817 being composed of more exotic compact objects [275], for example quark stars [276] or boson stars [277, 278]. For the entirety of this chapter, and indeed this thesis, we will assume GW170817 is the result of two coalescing NSs.

Announcement of the first observation of a BNS merger facilitated many independent investigations to infer the equation of state of nuclear matter ([279][280][181] for example). This included the LIGO-Virgo collaboration's complementary analysis to that of the original announcement [4], discussed in chapter 1 section 1.4.1. By assuming the astrophysical origin of the event and that the equation of state of NS matter is universal, they can more-so constrain the inference of the dimensionless tidal deformability and thus the equation of state through integration of the TOV equations. They also discuss another method involving the parameterisation of the equation of state; by direct estimation of equation of state parameters during PE,

one can then reconstruct a parameterised equation of state posterior. Both of these methods are discussed in detail in chapter 2 section 2.2.1.2. Both of their results, first assuming universal relations and inferring tidal parameters and secondly inferring equation of state parameters and enforcing physical constraints, prefer softer equations of state and suggest that the combined dimensionless tidal deformability parameter, $\tilde{\Lambda}$ (introduced in chapter 1 section 1.3.1) is small.

Inferring macroscopic parameters of the NSs in PE and integrating the TOV equations to return an equation of state posterior adds an additional computationally expensive and time-consuming stage to equation of state inference. PE of source parameters of a BNS system is already by nature expensive, relative to PE for a standard BBH signal, for example, due to the larger parameter space; typically, a BNS PE run requires 10^7 likelihood evaluations [125]. Along with integration of the TOV equations to return the equation of state posterior, which takes around $\sim$ 1 second per equation of state [281], the entire method becomes temporally and computationally expensive for large data sets posterior samples from GW events, typically 1000s of samples. Alternatively, inferring the equation of state parameters during PE returns a posterior on parameters of the spectral equation of state parameterisations alongside macroscopic parameters of the system, increasing the dimensionality of the likelihood space, and therefore increasing the computational complexity of the sampling. This also strictly represents the equation of state posterior in this relatively simple parameterisation, and does not incorporate more complex physics. Methods to infer the NS equation of state from GW170817 that followed [4] and their efficiency are reviewed in chapter 2 section 2.2.1.2.

In this work, we use ML methods to rapidly infer the NS equation of state once PE samples of the macroscopic parameters of the system are available from a GW observation of a BNS merger. We train a Flow to learn the mapping from equation of state space to standard Gaussian latent space conditioned on event information associated to each equation of state - namely the mass, m, and dimensionless tidal deformability, Λ, of each component. Once the Flow is trained, we can sample from the latent space conditioned on PE samples from a BNS event and return the equation

of state posterior in less than 1 second, significantly lowering the computation cost compared to non-ML based inference pipelines (see 2 section 2.2.1.2). In addition to returning an equation of state posterior quickly, the method is inherently flexible; the Flow can be trained on any equation of state parameterisation, or on a collection of multiple equation of state models. We additionally use a data compression scheme which decomposes the training equations of state into an independent set of parameters which do not depend on the equation of state parameterisation but rather the common morphology. Fast and model agnostic inference of the NS equation of state complements the current growth in low-latency PE as a method for informing potential EM follow-up of BNS merger events.

4.1 A recap of Normalising Flows

As introduced in chapter 3, a Normalising Flow is a generative ML model which learns the invertible and tractable mapping from a given N-dimensional complex distribution to a simpler latent distribution of the same dimensionality. This mapping is performed by a series of transforms, which are described by neural network hyperparameters, namely the number of neurons, transforms and residual blocks per transform. In training, a *conditional* Normalising Flow learns to map samples x from the data space $\mathcal{X}$ to a standard Gaussian latent distribution $\mathcal{Z}$, conditioned on label y, such that:

$$p_{\mathcal{X}|\mathcal{Y}}(x|y) = p_{\mathcal{Z}|\mathcal{Y}}(f(x|y)|y) \left|\det\left(\frac{\partial f(x|y)}{\partial x}\right)\right|, \tag{4.1}$$

where f is the function which describes the mapping from data to latent space, $f : \mathcal{X} \to \mathcal{Z}$. The Jacobian of the function evaluated at sample x in the data space is given by the partial derivative $\frac{\partial f(x|y)}{\partial x}$; we require the Jacobian determinant to remain tractable and easily computable for successful inverse sampling.

In this work, we train a conditional Normalising Flow to map a parameterisation of the equation of state plus additional parameters, constituting the data space $\mathbf{\Omega}$, conditioned on component masses and tidal deformabilities of a BNS system, $\theta = [m_1, m_2, \Lambda_1, \Lambda_2]$, such that we train

$$p_{\Omega|\Theta}(\Omega|\theta) = p_{\mathcal{Z}|\Theta}(f(\Omega|\theta)|\theta) \left|\det\left(\frac{\partial f(\Omega|\theta)}{\partial \Omega}\right)\right|. \tag{4.2}$$

The Flow is trained to learn the forward mapping - the transformation of a 10−dimensional reduced equation of state data space (which will be introduced in detail in section 4.2), $\mathbf{\Omega}$, to Gaussian latent space $\mathcal{Z}$, conditioned on event information associated to a given equation of state, θ. Once trained, the Flow performs the inverse transform; a latent space sample z is drawn from the 10-dimensional Gaussian latent space which is inversely mapped to an equation of state sample $\mathbf{\Omega}$ conditioned on event sample θ. This process is repeated $\forall\theta \in p(\theta|h)$, where h is GW strain, given that samples θ lie within our training prior bounds (discussed in detail in section 4.3.1) to build up an overall equation of state posterior, $p(\mathbf{\Omega}|\theta)$.

4.2 Data

The training data for the conditional Flow requires two data sets; one which represents the data space, $\mathbf{\Omega}$, which the Flow maps to the Gaussian latent space, $\mathcal{Z}$, and another for the conditional labelling of each mapping, θ. Our data space which we wish to map is that of the equation of state, where the conditional labels are event information associated to each equation of state. We will first discuss the details of the equation of state data and its compression before discussing the processing of the conditional labels.

The equation of state training data, presented in figure 4.2.1 consists of 10^5 equations of state varying in energy density on a fixed grid of pressure values. Each equation of state is a piecewise polytropic parameterisation [146] constructed of 3 pieces with transition densities at 5×10^{14} g cm^{-3} and 10^{15} g cm^{-3}. Each piecewise polytrope is stitched to a common crust equation of state at 6.7×10^{13} g cm^{-3}, defined by the SLy4 equation of state [143] (see chapter 2 section 2.1). The choice of piecewise polytropic equation of state in this analysis reflects the wide use of this equation of state parameterisation in the literature, however the analysis facilitates the use of any equation of state, given that the user is aware the choice of equation of state prior in this Bayesian analysis will ultimately influence the final equation of state posterior.

Each equation of state in our training data set has a maximum allowed density, ρ_{max},

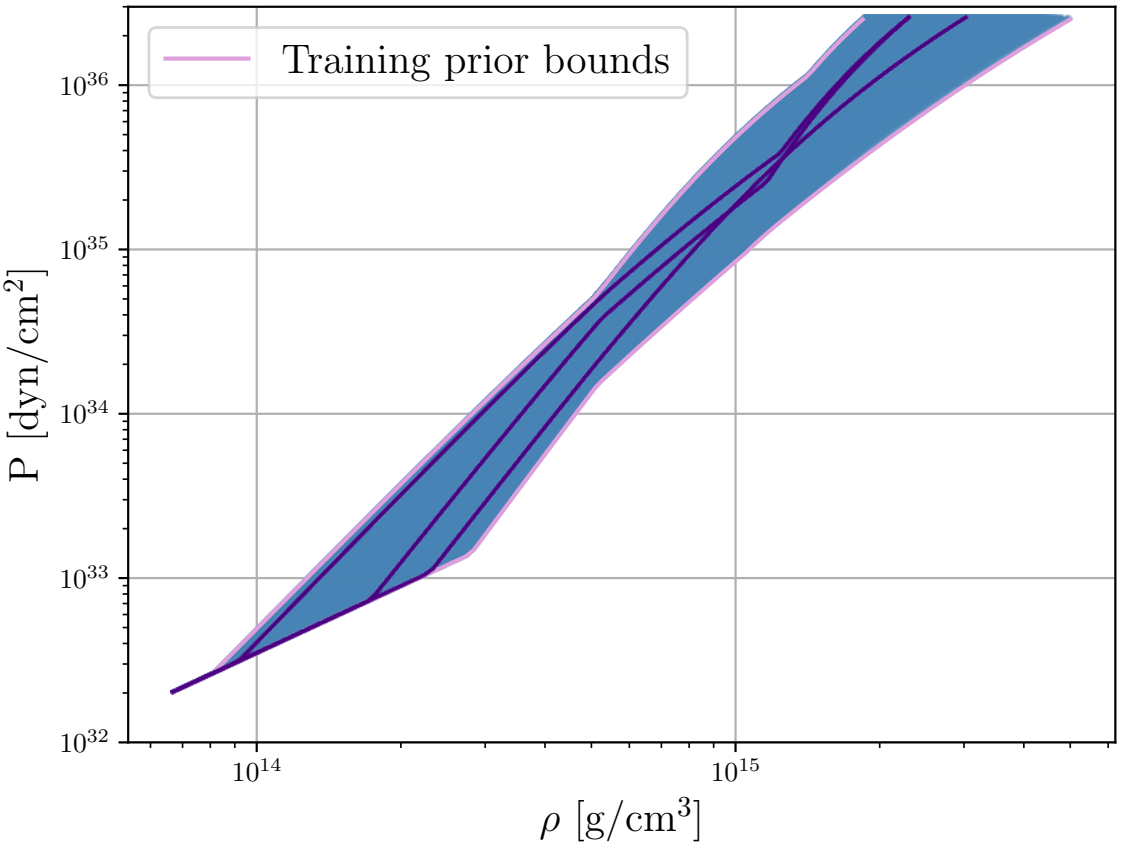


Figure 4.2.1: Plot of pressure vs. density of all 10^5 equations of state in the training data prior in blue, with highlighted examples in purple. The training prior bounds are in light pink, representing the stiffest (upper) and softest (lower) equations of state in our training data set. Each equation of state is stitched to the common low density crust equation of state SLy4 [143] at 6.7×10^{13} g cm^{-3} and is constructed of a 3-piece polytrope.

which defines the maximum density at which the equation of state is thermodynamically stable and causal (concepts discussed in the introduction of chapter 2). This maximum density is directly related to the maximum allowed mass, m_{max}, of the NS for this equation of state (see chapter 2 section 2.2 for discussion of maximum mass of a NS). To select an event associated to this equation of state, one uniformly samples component masses m_1 and m_2 between 0.5 $M_\odot$ and m_{max}, where $m_1 \geq m_2$. The choice of 0.5 $M_\odot$ as a lower limit was chosen so to match that of [4]. Setting these prior bounds define the region within which we can conditionally sample with the Flow: we cannot trust the Flow to behave well when evaluating out-of-distribution (OOD) data. This includes both evaluating the equation of state data space beyond the bounds of the training data distribution or conditionally sampling in-distribution equation of

state data given OOD conditional data.

Given a table of energy density for each equation of state in the training data set, we interpolate the relationship between energy density and mass (relationships of macroscopic parameters for each of the equations of state are found by solving the TOV equations for each equation of state in the data set). The maximum allowed density is defined via interpolation of the maximum allowed mass. Once found, the component masses m_1 and m_2 are sampled uniformly between a lower limit of 1 $M_\odot$ up to the allowed maximum mass for a given equation of state, and likewise interpolated with respect to energy density to find the central energy densities of each of the component NSs, ρ_1 and ρ_2. Similarly, the maximum allowed density ρ_{max} is found via interpolation of the maximum allowed mass and these three parameters are collected to form a set of *auxiliary parameters*, which will contribute towards the Flow's training data space, $\Omega = [\text{EOS}, \rho_1, \rho_2, \rho_{\max}]$. The energy density and tidal deformability are likewise interpolated and the central energy densities for each NS are mapped to the component tidal deformabilities, Λ_1 and Λ_2, where inversely to the mass convention, $\Lambda_1 \leq \Lambda_2$. The full set of conditional training data is presented in figure 4.2.2. We present the 4-dimensional conditional data space as $\theta = [m_1, m_2, \log \Lambda_1, \log \Lambda_2]$ to reduce the dynamic range of the tidal deformability parameters. The conventions are clear in the strict boundaries at $m_1 = m_2$ and $\Lambda_1 = \Lambda_2$, upheld by the mass convention and interpolated to the tidal deformability parameter.

This scheme is such that each equation of state is associated to a single BNS event $[m_1, m_2, \Lambda_1, \Lambda_2]$ for the purposes of training. Alternatively, one could generate single NSs associated to each equation of state, or data related to an NSBH (which would involve different mass priors for each component and a tidal deformability of 0 for the BH). The scheme also allows for simple data augmentation; one does not have to stop at 1 BNS event but can generate many, so to show the Flow many different events associated to each equation of state. Figure 4.2.3 shows what many events associated to 1 equation of state looks like in the $m - \Lambda$ plane; this is analogous to the pressure-density plane and is indicative of the type of relationship we are training the Flow to identify.

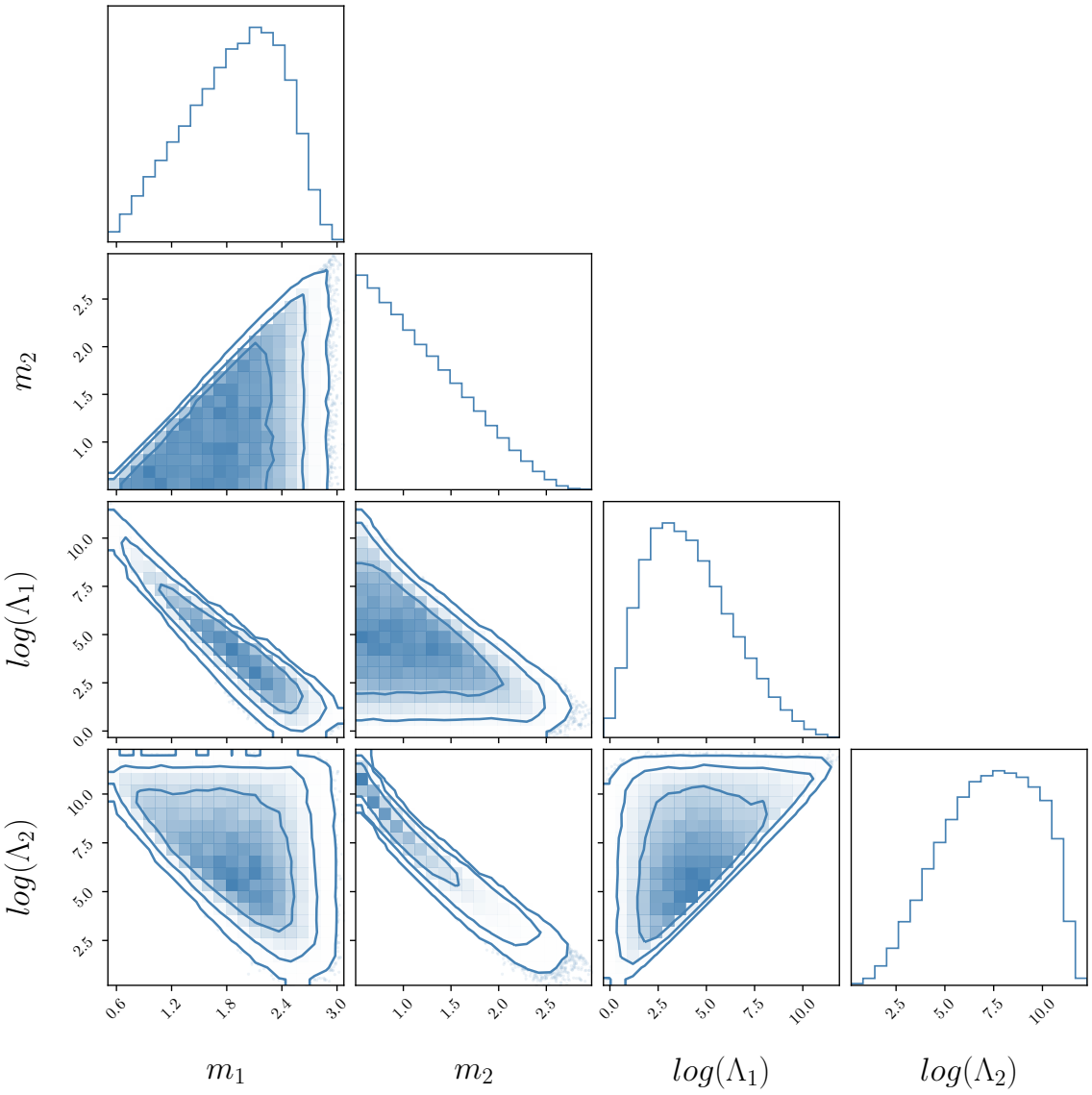


Figure 4.2.2: Corner plot of 1- and 2-dimensional marginalised distributions of $[m_1, m_2, log(\Lambda_1), log(\Lambda_2)]$ conditional training data. Extending outwards, each of the contours represent 68%, 95%, and 99.7% probability.

4.2.1 Using principle component analysis to compress the data

To reduce the complexity of the training data before passing to the Normalising Flow to learn the mapping, we wish to compress the equation of state data without losing unique information associated to each individual $P - \rho$ relationship and over-generalising. Reducing the dimensionality of the data space the Flow will be trained to learn has a direct impact on the training time and simplifies the input data

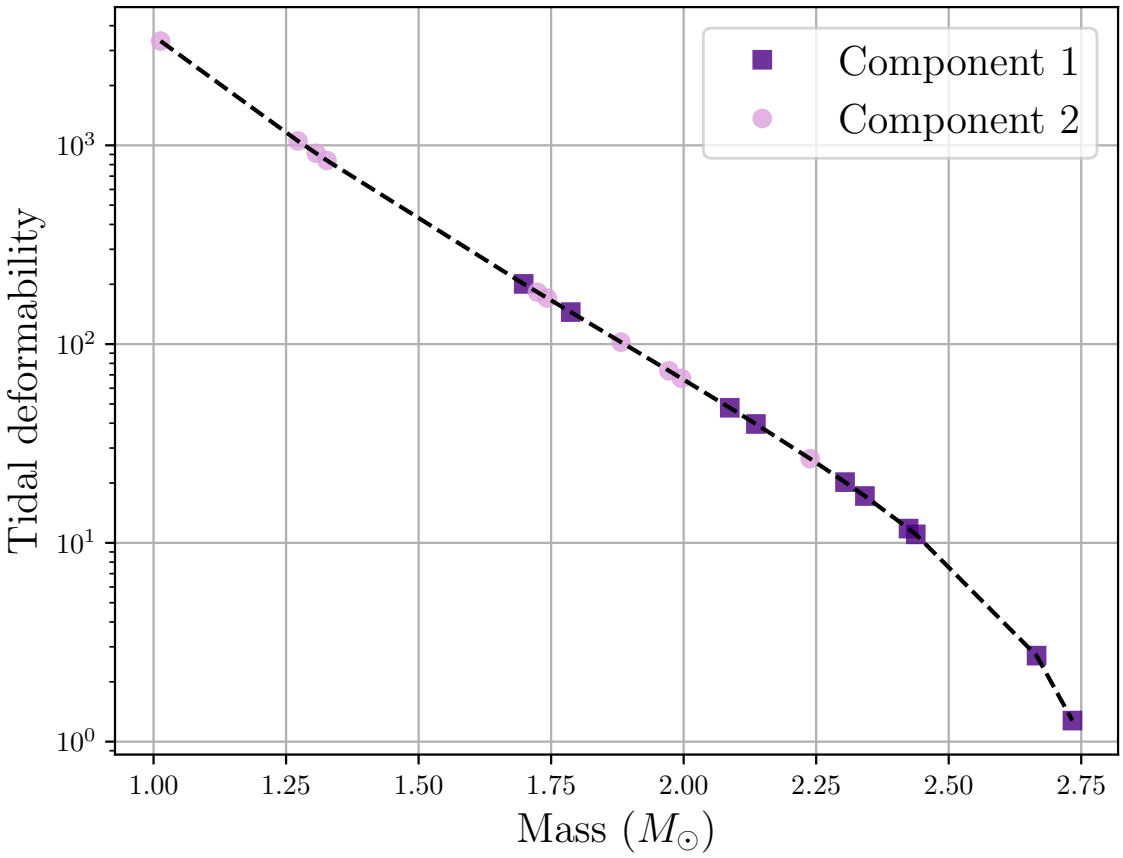


Figure 4.2.3: Plot of mass vs. tidal deformability of an example equation of state in the $m - \Lambda$ plane. Each purple square and pink circle represent component 1 and 2 of a BNS event respectively, where $m_1 \geq m_2$. There are 10 events presented, each generated to be associated to the same equation of state from the training data set via the interpolation scheme discussed in section 4.2, the relationship presented as a black dashed line.

such that the model is required only to learn necessary features of each equation of state. To reduce the equation of state training data to a size more conducive for the training of the Normalising Flow, we use principal component analysis (see [282] for a comprehensive introduction). This method allows us to decompose the multivariate training data set of M equations of state from N points in density (along the pressure grid) to d orthogonal principal components, where $d \ll N$.

We can define our equation of state training data as a matrix, $\mathbf{\Omega}$, with dimensions $M \times N$, where each row is an equation of state from our training data, and each column represents one of the 300 points along each of our equations of state in $P - \rho$. We find the covariance matrix, $\boldsymbol{C}$, of our training data $\mathbf{\Omega}$ which is a symmetric matrix

with dimensions $M \times M$, which describes the covariance of the i^{th} and j^{th} equations of state, $\boldsymbol{\Omega}_i$ and $\boldsymbol{\Omega}_j$, from the training data, s_{ij}, in upper and lower triangles. The matrix diagonal elements, s_i^2, describe the variance of the i^{th} equation of state [283] such that

$$\boldsymbol{C} = \begin{bmatrix} s_1^2 & s_{12} & \dots & s_{1j} \\ s_{21} & s_2^2 & \dots & s_{2j} \\ \dots & \dots & & \dots \\ s_{j1} & s_{j2} & \dots & s_M^2 \end{bmatrix}. \tag{4.3}$$

We find the M eigenvectors of the covariance matrix $\boldsymbol{C}$, where the eigenvalue associated to each eigenvector informs us how much variance is aligned with each eigenvector. The more variation quantified by a given eigenvalue, the less informative the eigenvector is. Therefore the eigenvalues associated to each eigenvector are inversely proportional to the quantity of information encoded in the eigenvector. The eigenvector with the largest eigenvalue is the **first principal component**; this component describes the most variance of the equation of state parameter space. The second principle component is the eigenvector with the second largest eigenvalue, and so on, until we have d principle components which describe the maximal variance of our training data.

We use the `scikit-learn` [284] implementation of Principal Component Analysis to compress the 105 points along each of the 10^5 equations of state in the highest density region (105 points of piecewise polytropic equations of state) to a linear combination of 7 orthogonal principal components, plotted in 4.2.4. These 7 principal components are the 7 most prominent eigenvectors to describe the distribution of the 10^5 training data points. The figure presents the data space distribution which will be modelled by the Flow, alongside three auxiliary parameters (not plotted here).

The number of principal components were selected relative to their respective explained variances; each principal component explains a magnitude of variance of the space. Figure 4.2.5 shows the cumulative explained variance ratio of the 7 principal components used for our equation of state training set. We demonstrate how one wishes to select a number of principal components which describe the parameter space

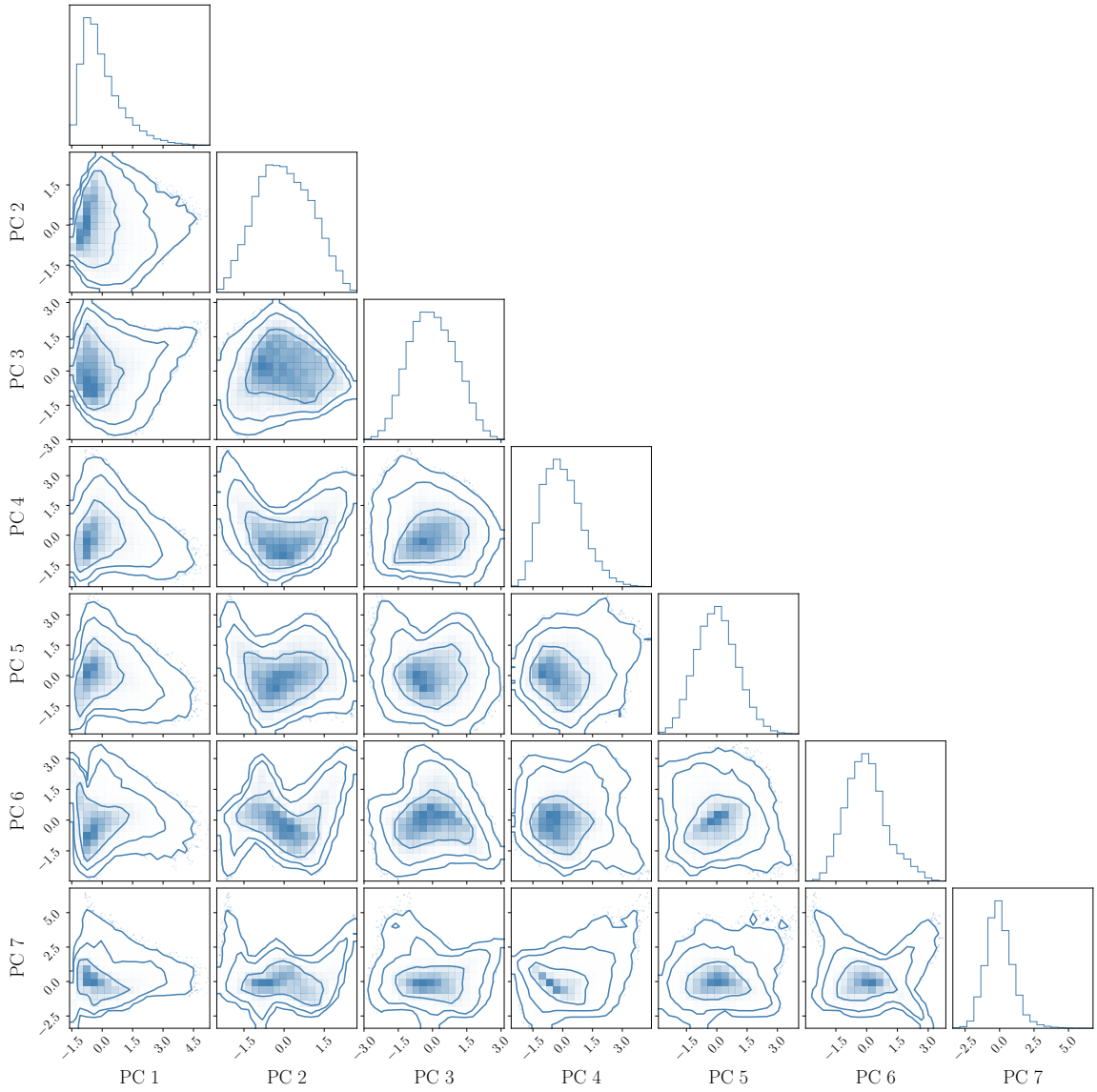


Figure 4.2.4: Corner plot of 1- and 2-dimensional marginalised distributions of the 7-dimensional PCA-compressed equation of state training data. Each component has been scaled to mean 0 and standard deviation 1. Extending outwards, the contours contain 68%, 95%, and 99.7% of the samples.

as fully as possible, without selecting so many as to overwrite the goal of compressing the high-dimensional space. The selection of 7 components in this case, was optimal for the magnitude of information preserved but also with regards to the input to the Normalising Flow. Together, the 7 principal components account for 99.975% of the variation of the training data set. Reducing to 6 components accounts for 99.96%, and increasing to 8 components accounts for 99.98%. We choose 7 as a middle-ground number of parameters which account for greater than 99.9% of variation in the data

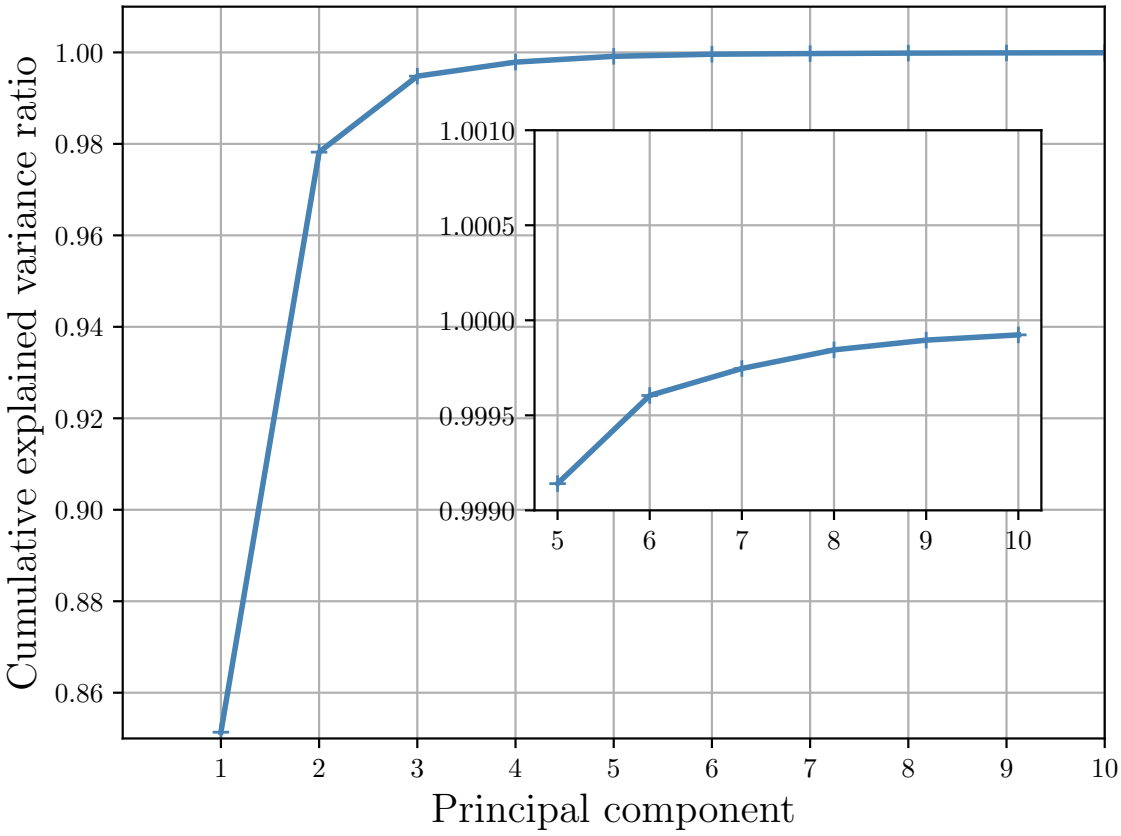


Figure 4.2.5: Number of principle components used to represent the equation of state training data vs. cumulative explained variance ratio for 1 to 10 principle components. The inset plot presents the cumulative explained variance ratio for principle components in the range $[5, 10]$ on a smaller y-scale. The 7 principal components used in the analysis a cumulative explained variance of 99.975% of the original training data.

.

set with a reasonable number of parameters. Chapter 5 discusses the quality of data compression via PCA compared to other Machine-Learning-based methods.

Once compressed, the reduced-dimensionality representation of the equation of state is collected alongside the auxiliary parameters. The auxiliary parameters are logged before being incorporated into the larger data set to reduce their dynamic range; by nature, ρ_1, ρ_2 and ρ_{max} are on the scale of 10^{14-15}, orders of magnitude different from the scaled principal component (PC)s (see figure 4.2.4). The 10-dimensional data set constitutes the data space, $\mathbf{\Omega} = [\text{PCA}_{1-7}, \rho_1, \rho_2, \rho_{\text{max}}]$, which we wish to map to the Gaussian latent space, $\mathcal{Z}$, with the conditional Normalising Flow. Each of the 10 dimensions of the data space are scaled to zero mean and unit standard

deviation, respectively, to have all dimensions of the data be on a comparable scale. This is standard good-practice and has been found to improve the Flow's ability to train. Similarly, each BNS event associated to each equation of state (a 4-dimensional vector consisting of $[m_1, m_2, \log \Lambda_1, \log \Lambda_2]$) likewise has each parameter independently scaled to zero mean and unit standard deviation, where $\Lambda_{1,2}$ are also logged firstly before being scaled to reduce their dynamic range.

4.2.2 Hyperparameter selection and optimisation

A Flow model was initially trained with 11 transforms, 4 blocks per transform and 50 neurons. The model was trained for 5000 epochs at an initial learning rate of 0.005 which was decayed to 0 over the course of training with Cosine Annealing. A batch size of 4096 was used. Over the course of training, the training and validation losses were both found to decay appropriately but with large KL divergence between the learned latent space and a standard Gaussian. Training and validation losses for this model, named the 'complex' model, alongside that of the simple model (discussed later), can be found in figure 4.2.6. In this figure, the purple training and validation losses reach a lower loss than of an alternative model, given in blue. This appears to suggest the Flow trained to produce the purple training and validation losses is performing better, however, further tests proved this not to be the case and rather that the complex Flow, was overfitting.

To find a new model whose performance improved upon that of the complex model in minimising the KL divergence and in testing (discussed more in the following sections), hyperparameter sweeps were performed using the hyperparameter optimisation software, Weights and Biases (or `wandb`) [285]. `wandb` runs training iterations of the Flow many times over with a different model architecture each time. The hyperparameters for each model are initially selected uniformly from a range defined by the user, and each subsequent model's hyperparameters are selected depending on the performance of a chosen figure of merit in the previous training iteration, for example validation loss. The method through which Weights and Biases selects the next set of hyperparameters relative to the user's chosen figure of merit can also be defined by the

	Hyperparameter range
Number of neurons	[128,256]
Number of transforms	[2,6]
Number of residual blocks	[2,4]

Table 4.2.1: Table of ranges of neurons, transforms and residual blocks sampled in hyperparameter optimisation with Weights and Biases.

user; this can be entirely random or can be selected via Bayesian optimisation. For this work, Bayesian optimisation was used with the goal of minimising the validation loss. At the end of a hyperparameter sweep, the user has access to figures of merit of interest for each model, so to compare overall performance and find an optimal model architecture. Hyperparameter sweeps through different model architectures are useful in helping the user to understand why a given model learns the data space, $\mathcal{X}$ (or learns the mapping of your data space to a unit Gaussian) better than another. Understanding, for example, why one data set may require more neurons in the Flow architecture as opposed to more transforms can promote interpretability of the model's understanding of the underlying data distribution, and also improve the user's understanding of the data.

A hyperparameter sweep of different models aiming to learn the conditional mapping from equation of state space to a standard Gaussian was performed with hyperparameters sampled uniformly in the ranges given in table 4.2.1 . The sweep was run with the aim of reducing the validation loss, so to avoid overfitting. Weights and Biases allows users to toggle what figure of merit is of interest to discover what hyperparameters are most important relative to this figure of merit. We investigated, for this purposes of finding a suitable model for this Flow's training, what models in the hyperparameter sweep performed best relative to minimising the KL divergence. It was found that the number of neurons and number of transforms were negatively and positively correlated with the KL divergence, respectively.

As a result of a hyperparameter sweep, a smaller network architecture was found to be optimal for our problem. As such, a Flow, hereon the 'simple Flow', with 2 blocks of 3 transforms with 151 neurons was trained. The network was, likewise with the

complex model, trained for 5000 epochs with an initial learning rate of 0.005, decayed over training to 0 with cosine annealing. The network was also found to benefit from the use of the `nflows` attribute `LULinear = 'lu'`; this accounts for correlation in the auxiliary parameters and avoids overfitting [286]. The use of this attribute in training the Flow was found to improve the minimisation of the *latent* KL divergence, between the learned Gaussian latent space and a true Gaussian distribution, as shown in figure 4.2.6 (b), which shows a smaller KL divergence of the simple model (blue) compared to the complex model (purple). Figure 4.2.6 (a) shows a larger overall training and validation loss for the simpler model (blue) but a smaller gap between training and validation loss than that of the complex model (purple), demonstrating improved generalisation (no gap between higher validation and lower training losses).

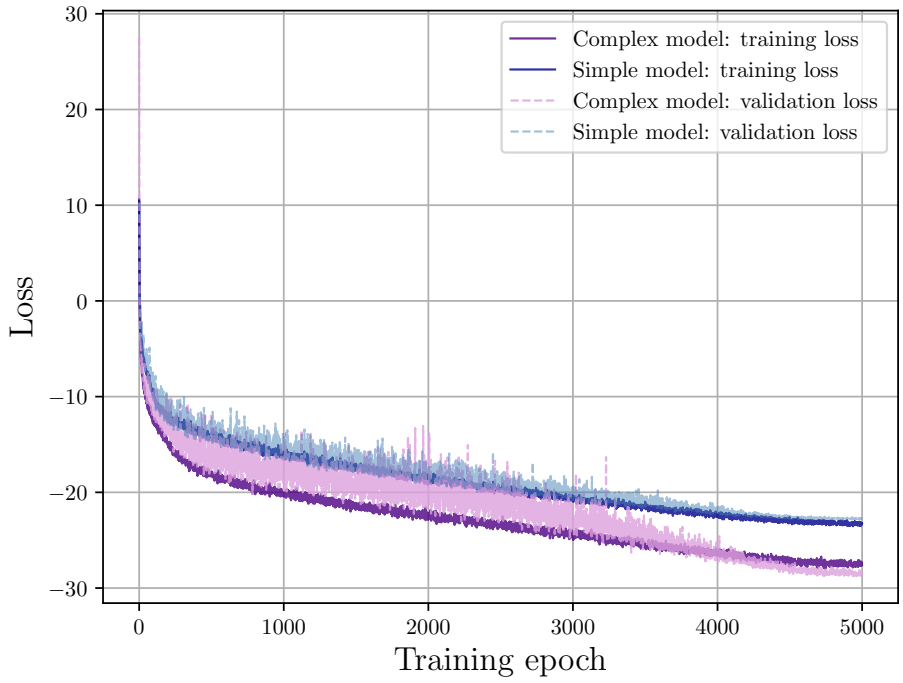
4.3 Inference

4.3.1 Convex hull around conditional data space

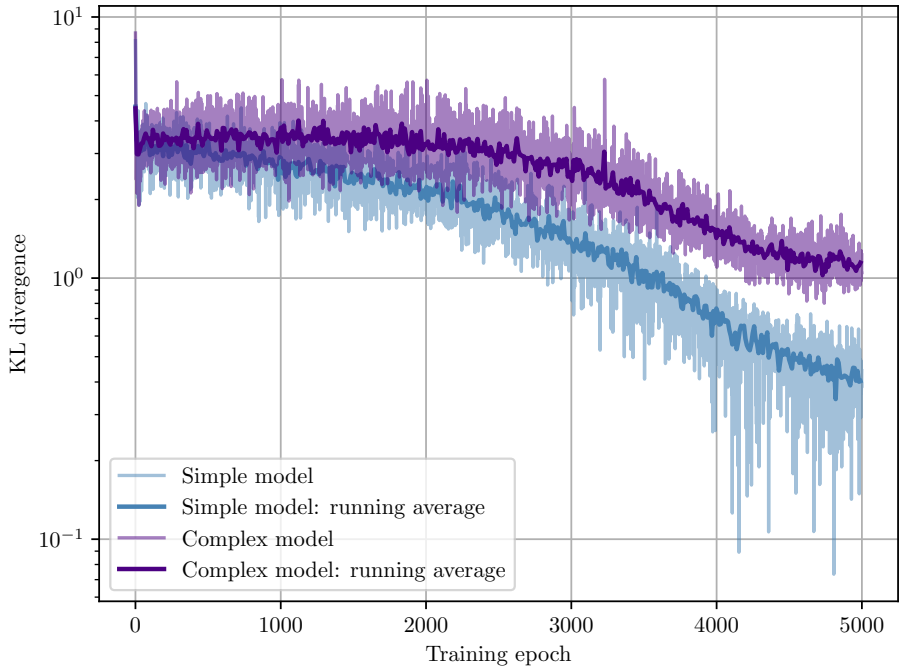
In using the Flow for inference and for conditionally sampling given some event information, we want to be sure the event information we are using to conditionally sample from the Flow are from within the training data prior bounds on the conditional space. In section 4.2, we discussed the prior bounds which we generated $\theta = [m_1, m_2, \log\Lambda_1, \log\Lambda_2]$ training data from; the subsequent conditional training data is presented in figure 1.3.3. In inference, we can then conditionally sample the Flow using any conditional label, $\Omega \sim p(\Omega|\theta)$, provided θ belongs within these bounds. These bounds represent a physical boundary, within which all possible BNS events lie. Therefore, this is a boundary we wish to enforce through use of a convex hull.

A convex hull describes an N-dimensional convex set in which all points that constitute the set and all lines that connect any two given points within the set all lie within the bounds of the set (hull). As such, one can walk between any two vertices (or points) within the hull without stepping outside of the hull. From [287], a convex shape is defined as:

Theorem 1. *A set $P \subseteq \mathbb{R}^d$ is convex if $\overline{pq} \subseteq P$, for any $p,q \in P$.*



(a) Training and validation loss



(b) Latent space KL divergence

Figure 4.2.6: Training and validation loss (a) and latent space KL divergence (b) as a function of training epoch for the simple (blue) and complex (purple) models.

By constructing a convex hull with the conditional training data set θ, we create a set which encloses all training data points and the space between them. We construct a convex set, H: the smallest convex set that encloses out training data set Θ, and is the convex hull of Θ. We use the `scipy.special` implementation to construct a convex hull around the $[m_1, m_2, \log\Lambda_1, \log\Lambda_2]$ training prior. Once constructed, $[m_1, m_2, \log\Lambda_1, \log\Lambda_2]$ event samples can be assessed as to whether they lie within the bounds of the convex hull or not (a boolean result). This ensures that conditional data we use in inference is physically motivated and so expect the Flow be able to sample within the training prior bounds.

4.4 Training the Normalising Flow

Different Flow architectures were trained and tested on a variety of figures of merit before settling on the optimal network architecture. As discussed in chapter 3 3.2, network performance is assessed relative to training and validation loss performance, the latent KL divergence between the learned latent representation and a true Gaussian, as well as the performance on the task the model has been trained to complete - in this case, conditionally sampling $\Omega \sim p(\Omega|\theta)$.

Two Flow architectures were investigated for the purpose of this work. The first architecture, the complex model, was found to struggle in sampling equations of state when given conditional data from certain regions of the mass-tidal deformability space. It became apparent that the complexity of the Flow meant that it overfit to the conditional data θ, such that sampling with conditional data which did not belong to the training or validation data returned out-of-distribution equations of state. A smaller network was built to tackle some of the sampling issues, which we refer to as the 'simple' model. In developing this model, various techniques were introduced to the inference scheme to reject unphysical results and ensure that we were sampling with physically relevant event information. These will be discussed in the coming subsections. The initial large model was likewise tested with these new features and with the same learning rate and batch size as the simple model. The different initialisations of these models and their development alongside testing on the

required task will be discussed here with results presented.

Once we have a well-trained model, we use the Flow to perform the conditional mapping: $\Omega \sim p(\Omega|\theta)$. The samples, θ, we wish to map from are PE samples from any BNS merger event, but first and foremost to test our analysis, we wish to conditionally samples from the BNS merger event GW170817. The PE samples used are those of the first Gravitational Wave Transient Catalogue, GWTC-1 [53] [288]. In our analysis, we use samples of $\theta = [m_1, m_2, \log \Lambda_1, \log \Lambda_2]$ associated to the PE performed with the frequency domain waveform model `IMRPhenomPv2_NRTidal` [289] with low-spin priors. As the analysis was performed post-EM follow-up, the re-analysis of PE for the catalogue was able to take advantage of the improved sky localisation using the precise position of the kilonova [176] and redshift associated to the source galaxy to obtain source-frame masses [290]. The GWTC-1 posteriors, however, do not assume a common equation of state and as such the mass and tidal deformability of each component can vary independently. This is accounted for in our pre-processing of the samples before use in the Flow; our constraints therefore reject around 50% of all posterior samples by enforcing the $m_1 \geq m_2$ boundary. Posterior samples from GWTC-1, before pre-processing is applied, are presented in figure 4.4.1. The posterior samples $p(\theta|h)$ included in the catalogue uphold our existing mass convention $m_1 \geq m_2$ but does not restrict each of the NSs to belong to the same equation of state, so that $\Lambda_{1,2}$ vary independently. This means that some of the posterior samples from the event are outwith our training prior bounds. We can see that, due to the loud nature of the event (discussed in chapter 1 section 1.4.1), the joint $m_1 - m_2$ joint posterior is well-constrained, whereas the $\log \Lambda_1 - \log \Lambda_2$ joint posterior is much broader.

The first equation of state inference performed by the LIGO and Virgo collaborations following GW170817 [4] published PE samples of component masses and tidal deformability [291]. There are three sets of posterior samples from this initial analysis, two of which are associated to their parametric approach to inferring the equation of state of GW170817 (one enforcing the maximum mass of each equation of state must be $\geq 2M_\odot$, and one not) and one associated to the approach which applied universal relations, introduced in 2 section 2.2.1.1. We compare our final result

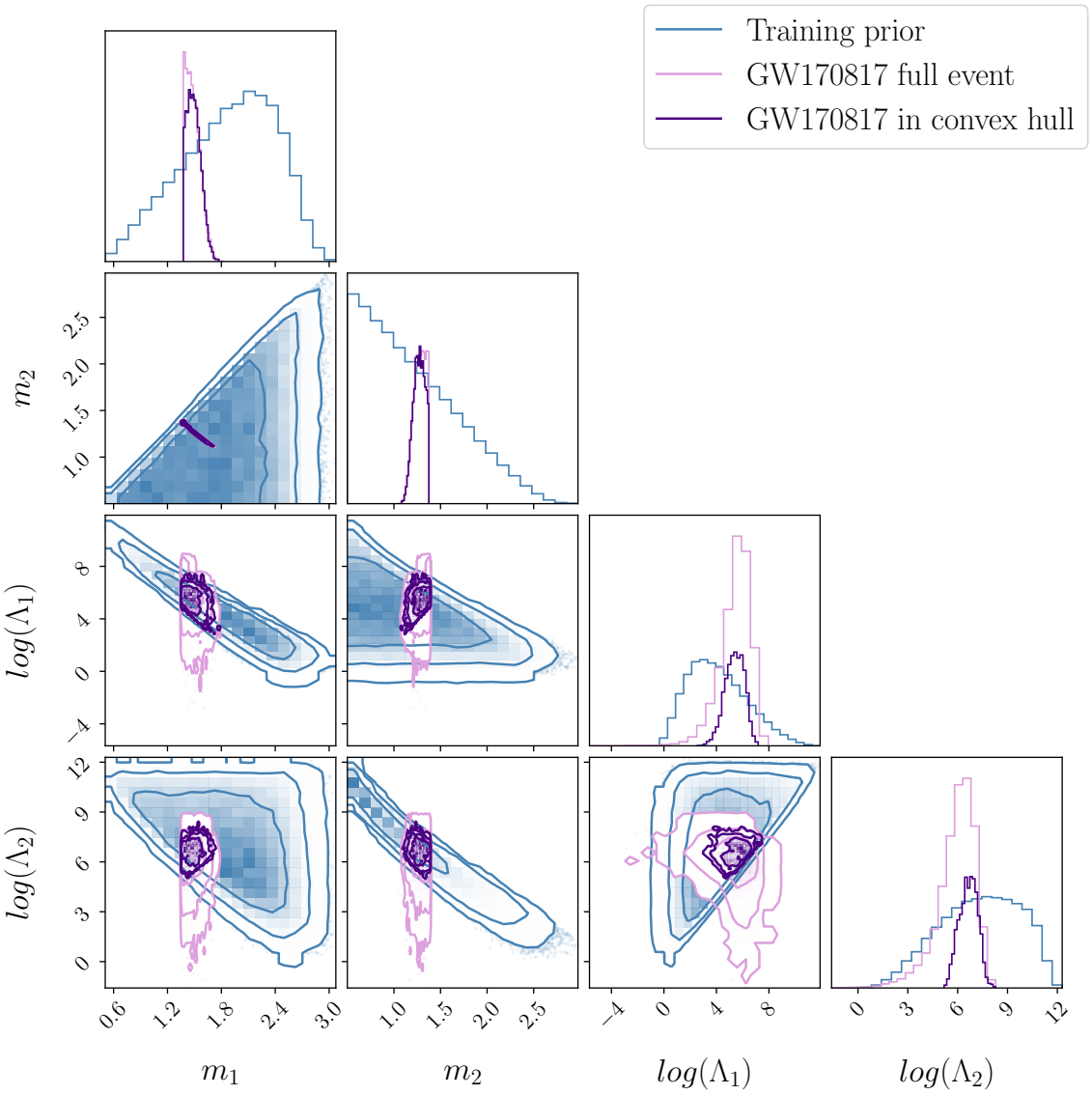


Figure 4.4.1: Corner plot of 1- and 2-dimensional marginalised posterior distributions of $[m_1, m_2, log(\Lambda_1), log(\Lambda_2)]$ from GWTC-1 for GW170817 in pink with samples within the convex hull in purple, plotted on the training data prior (blue). The contours contain 68%, 95%, and 99.7% of the samples.

to their parameterised approach with the maximum mass constraint so to compare directly to [4]. However, we choose to use the PE samples from GWTC-1 for use in our analysis as these are the least influenced by additional assumptions/constraints.

We keep only the PE samples from GW170817 which belong within the convex hull, introduced in section 4.3.1, that describe the bounds of our training data prior in the $m - \Lambda$ space. Figure 4.4.1 also shows samples from the event posterior that lie within the bounds of the convex hull. By cutting with respect to the convex hull we reject

4703 (58.2%) posterior samples, 3144 of which are associated to $\Lambda_1 \leq \Lambda_2$ violation (a convention which we uphold in inference). The remaining points lie beyond the bounds of which our Flow has been trained on. This could be due to being associated to regions of the parameter space which our equation of state prior training data does not support event information for, or may be beyond our lower mass bound.

4.4.1 Assessing inference quality

Table 4.4.1 presents the main differences between the complex and the simple model in training performance. We compare training time using the same graphics processing unit (GPU) hardware (NVIDIA GeForce RTX 3090), the difference in training times a consequence of the size of model trained, alongside the final latent space KL divergence and training and validation losses reached. We compare the relative model sizes, and see that the complex model is, indeed, more complex in that it consists of more transforms, neurons *and* residual blocks, and therefore takes longer to train. The learning rate and batch size used in training both models are the same.

Although the initial large model was found to train well with no evidence of overfitting, the latent space KL divergence was found to reach a minimum of 1.228 nats, where we expect this to be $\ll 1$. Some posterior samples returned by the Flow when conditionally sampling with data from GWTC-1 $\mathbf{\Omega}_{\text{GWTC-1}} \sim p(\mathbf{\Omega}|\theta_{\text{GWTC-1}})$ were found to have low outlier values of log-probability. Figure 4.4.2 (b) presents the log-probability of equation of state posterior samples $\mathbf{\Omega}_{\text{GWTC-1}}$ returned by the Flow when conditionally sampling with PE samples from GWTC-1, $\theta_{\text{GWTC-1}}$. There are clear outliers from the main high probability population of posterior sampling. Although the log-probability of samples returned by a Normalising Flow in inference is relative, and a negative log-probability posterior sample is not by definition a bad result, we set log-probability = 0 a threshold for rejection based on figure 4.4.2 (b), where there are clear outliers from the general trend of high probability samples. We use this metric as a complementary indicator of a badly-trained model alongside other statistically sound indicators (from table 4.4.1, discussed above).

The consequence of retaining low probability samples from $\mathbf{\Omega}_{\text{GWTC-1}}$ and projecting

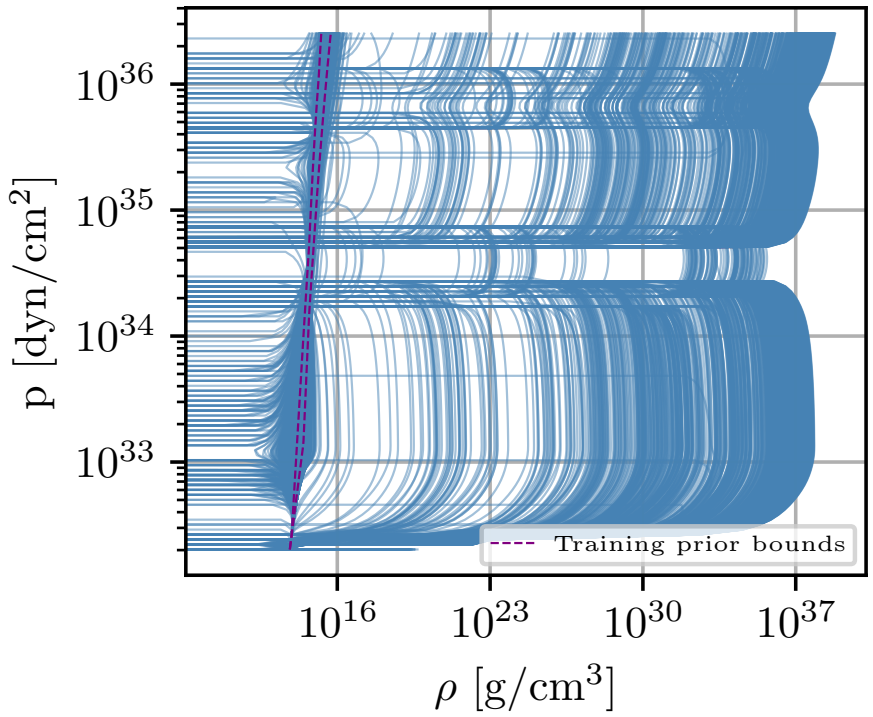
	Complex model	Simple model
Training time	2h 38m	1h 11m
Final latent KL div. (nats)	1.228	0.380
Final training loss	-27.556	-23.209
Final validation loss	-28.641	-22.789
# transforms	11	3
# blocks	4	2
# neurons	50	151
Learning rate	0.005	0.005
Batch size	4096	4096

Table 4.4.1: Table to compare the structures of and figures of merit related to the complex and the simple models respectively.

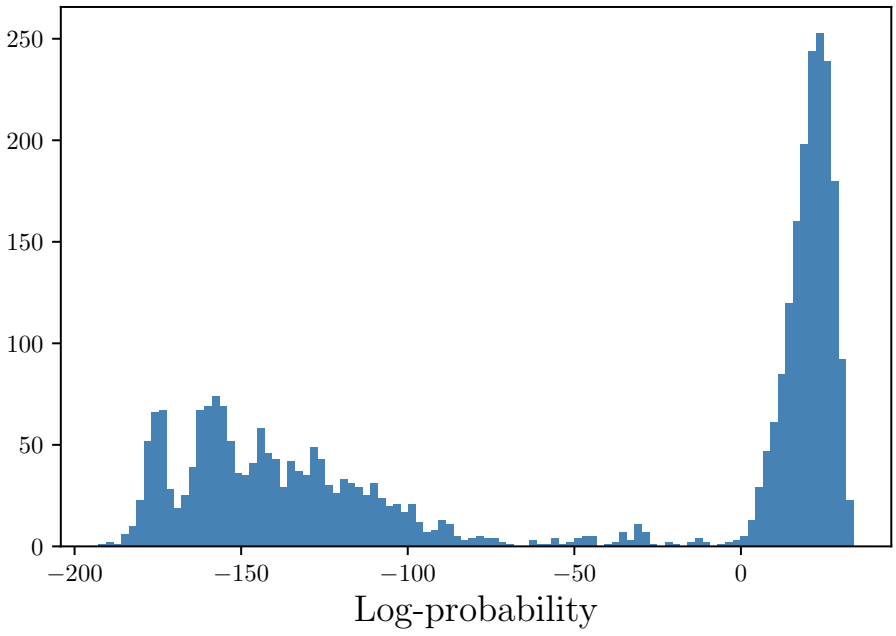
from the 10-dimensional equation of state space to the $P - \rho$ plane, is exhibited in 4.4.2 (a), which presents sampled equations of state $\Omega_{\text{GWTC-1}}$ projected to the $P - \rho$ plane with behaviour that can be labelled as unphysical. Unphysical equations of state are those which in the $P - \rho$ plane violate thermodynamic stability and/or causality or which extend beyond the prior training bounds (bounds in the $P - \rho$). The source of these unphysical equations of state are thus investigated and we test whether an equivalent analysis performed with the simple Flow network can minimise their quantity and influence on the overall posterior.

4.4.1.1 Testing edge-case conditional samples

We test the possibility that the aforementioned unphysical equations of state, presented in figure 4.4.2, are a result of mapping samples from the Flow latent space $\mathcal{Z}$ conditioned on event samples θ which are on the edge of the convex hull constructed around the conditional training data. We investigate the correlation between the log-probability of the equation of state sample returned by the Flow and the distance of the $\theta = [m_1, m_2, \log \Lambda_1, \log \Lambda_2]$ used to conditionally sample the equation of state space to the edge of the convex hull. This is calculated as the distance of the $m_1, m_2, \log \Lambda_1, \log \Lambda_2$ data point to the nearest line segment between two points belonging to the conditional training data set Ω, which was used to make the convex hull H. The convex hull is a hard boundary which essentially constructs a fence around the edges of the $m_1, m_2, \log \Lambda_1, \log \Lambda_2$ training prior. This is in contrast to the Nor-



(a) Equation of state posterior (no post-processing)



(b) Log-probability of all samples

Figure 4.4.2: (a) Plot of pressure vs. density for $\mathbf{\Omega} \sim p(\mathbf{\Omega}|\theta_{\text{GWTC-1}})$ from the complex model projected into the $P-\rho$ plane. The training prior bounds are given as dashed purple lines. (b) Histogram of probability of samples $\mathbf{\Omega}$.

malising Flow, which constructs an approximate distribution through transformations of a standard Gaussian, whose bounds extend to $[-\infty, \infty]$. It's not immediately clear how well edge cases of the training data set are mapped by the Flow, even though the convex hull would consider these to be completely viable.

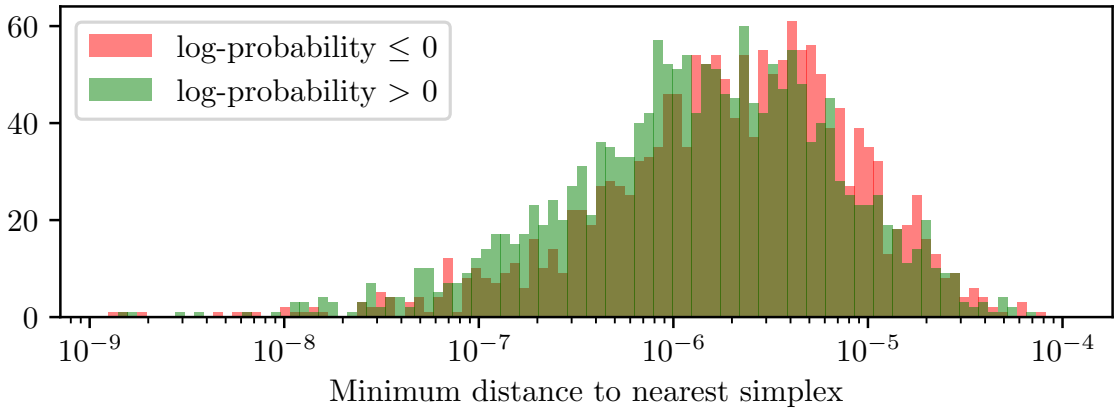
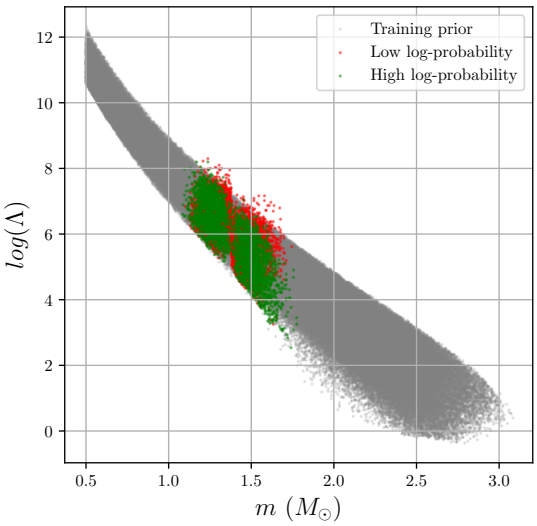


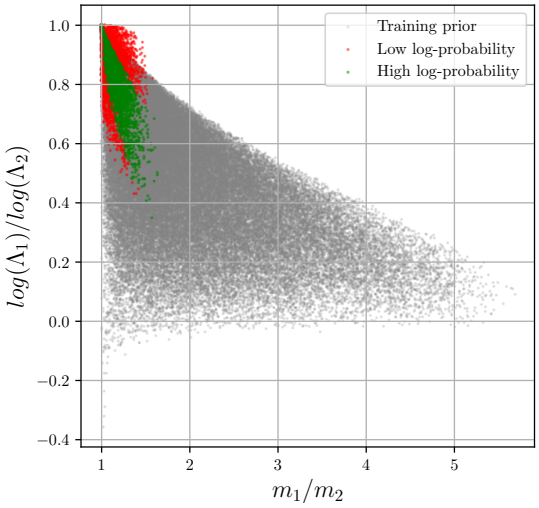
Figure 4.4.3: Histogram of distance to the nearest line segment for GW170817 event samples which match to a low log-probability equation of state sample in red or a high log-probability equation of state sample in green with the initial large Flow model.

Figure 4.4.3 presents two histograms of the distance of conditional samples $\theta_{\text{GWTC-1}}$ from their nearest simplex. The green histogram corresponds to samples $\mathbf{\Omega}_{\text{green}} \sim p(\mathbf{\Omega}|\theta_{\text{GWTC-1}})$ which have probability $p(\mathbf{\Omega}_{\text{green}}|\theta_{\text{GWTC-1}}) > 0$ and the red corresponds to $\mathbf{\Omega}_{\text{red}} \sim p(\mathbf{\Omega}|\theta_{\text{GWTC-1}})$ which have probability $p(\mathbf{\Omega}_{\text{red}}|\theta_{\text{GWTC-1}}) \leq 0$. We compare the outlier log-probability $\mathbf{\Omega}_{\text{red}}$ samples and high log-probability $\mathbf{\Omega}_{\text{green}}$ samples (presented in figure 4.4.2) to their corresponding conditional label's distance from its nearest line segment. The respective histograms overlap across the range of distances that points lie from their nearest line segments, with negative log probability equation of state samples $\mathbf{\Omega}_{\text{red}}$ being associated to conditional samples at further distances from line segments. We find such preference to be very slight, but importantly we do not find that the low log-probability equations of state samples are associated to conditional samples which are very close to line segments, in particular closer than their high log-probability counterparts.

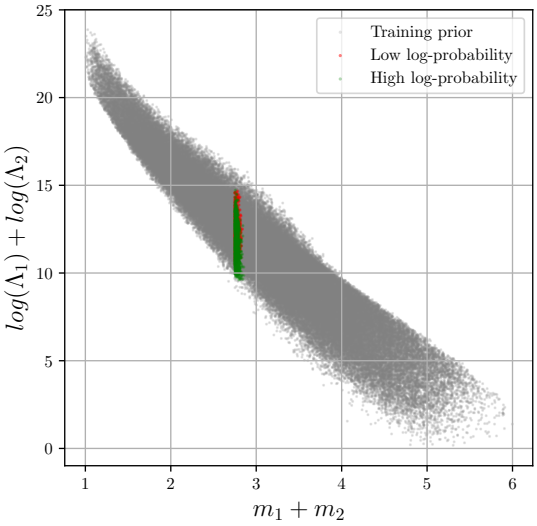
Therefore, it was proposed that there were regions of the $\theta = [m_1, m_2, \log \Lambda_1, \log \Lambda_2]$ conditional space which were not learned well by the Flow. These could be regions of the parameter space which are not well populated in training, and therefore not



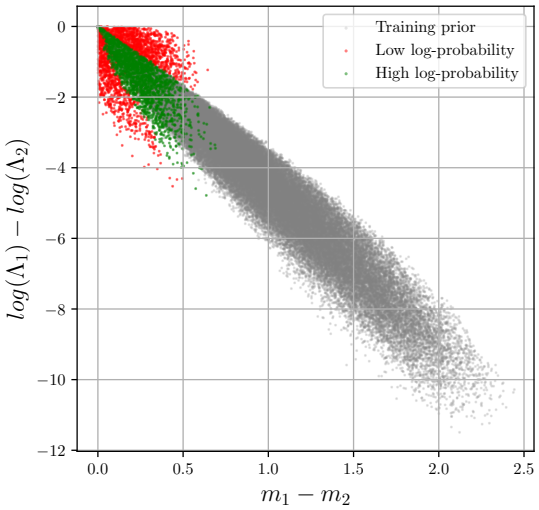
(a) Mass vs. tidal deformability



(b) m_1/m_2 vs. $log(\Lambda_1)/log(\Lambda_2)$



(c) $m_1 + m_2$ vs. $log(\Lambda_1) + log(\Lambda_2)$



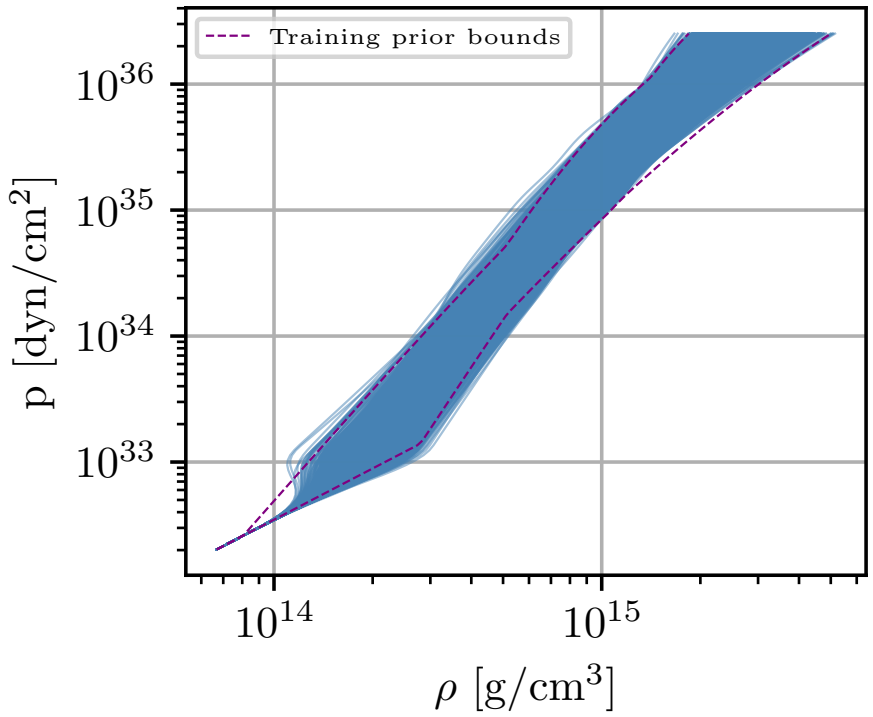
(d) $m_1 - m_2$ vs. $log(\Lambda_1) - log(\Lambda_2)$

Figure 4.4.4: 4 plots of different parameterisations of mass m and tidal deformability Λ for both components in the conditional training data set, in grey, alongside GW170817 event samples which map to low (high) probability equation of state posterior samples in red (green) with the complex model.

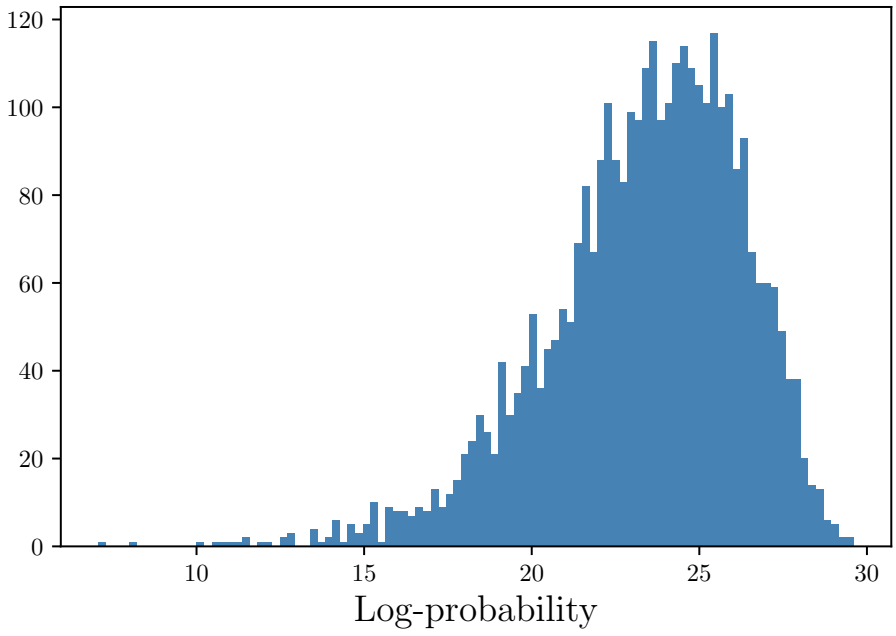
well understood by the Flow. Figure 4.4.4 presents different parameterisations of the condition parameter space, with the prior training points in grey alongside high and low log-probability points in green and red, returned from sampling the complex Flow with samples from GW170817, respectively. Different parameterisations of the space aim to present projections of the 4-dimensional space such that one can find the appropriate representation which makes clear the region which is not learned well or

has an abundance of low probability samples relative to high probability. Figure 4.4.4 demonstrates that, firstly, mass and tidal deformability samples which are beyond the space occupied by the prior $p(\theta)$, in grey, lead to low log-probability equation of state samples. This is an accepted truth and one which we use the convex hull to enforce and to mitigate the use of unphysical conditional samples beyond our training prior bounds. However, subfigures a, b and c of figure 4.4.4 all present high (green) and low (red) log-probability samples occupying overlapping regions of the training prior space. Therefore, none of these parameterisations display the low log-probability samples of $\theta_{\text{GWTC-1}}$ occupying a given region of the conditional parameter space *within* the prior bounds that high log-probability $\theta_{\text{GWTC-1}}$ does not. Subfigure d presents the low log(probability) samples lying *beyond* the bounds of the training prior and therefore in agreement with our existing knowledge. Therefore, particularly in plots (a) and (c) of figure 4.4.4, it appears there are unidentifiable regions of the conditional parameter space within the prior bounds which the Flow cannot sample confidently, indicating overfitting; the network appears to be over-complicating the problem and has overfit to the training data set, lacking the ability to interpolate over the space.

To simplify the problem, a smaller network was trained, the details of which are given in table 4.4.1. This smaller network trains quickly (in just over 1 hour compared to 2.5) and reaches a lower latent KL divergence than that of the complex model (see figure 4.2.6). While the overall training and validation losses reached at the end of training are greater than that of the more complex model, the training and validation losses are more in agreement and show no issue in generalising. With the smaller Flow model, the equations of state sampled with the GWTC-1 PE samples as conditions no longer have low log-probability outlier posterior samples, as seen in figure 4.4.5 (b), and can return almost entirely physically viable equations of state, see figure 4.4.5 (a), with only a few outlier equations of state beyond the prior bounds in $P - \rho$, which can likely we attributed to the intrinsic uncertainty associated to the Flow.



(a) Equation of state posterior (no post-processing)



(b) Log-probability of all samples

Figure 4.4.5: (a) Plot of pressure vs. density for $\mathbf{\Omega} \sim p(\mathbf{\Omega}|\theta_{\text{GWTC-1}})$ from the complex model projected into the $P-\rho$ plane. The training prior bounds are given as dashed purple lines. (b) Histogram of probability of samples $\mathbf{\Omega}$.

4.4.2 Removing unphysical results

Despite ensuring the conditional labels are within the bounds of our training data prior, we may still produce equations of state which lie beyond the training prior bounds of our equation of state data distribution (see figure 4.4.5 (a)). This is due to an intrinsic property of the Normalising Flow; we train the Flow to learn the mapping from 10-dimensional data distribution, $\mathbf{\Omega}$, which represents our equation of state and auxiliary parameter information, to a 10-dimensional unit Gaussian latent distribution, $\mathcal{Z}$. The unit Gaussian is not bounded and so extends $[-\infty, \infty]$, meaning that when one conditionally samples from the Flow, there is a potential of sampling the edges (low probability regions) of the latent Gaussian distribution $z \sim p(z)$ and mapping to a low probability region of the data space (beyond the bounds of the training distribution). The Flow is not trained with respect to hard boundaries in the 10-dimensional space, and so we should expect there to be some leakage.

Low probability posterior samples in the edges of the 10-dimensional equation of state prior distribution are entirely reasonable outputs from the Flow but can produce unphysical results when projected back to the $P - \rho$ plane. To prevent this projection, we wish to apply a boundary in the 10-dimensional reduced equation of state space which will retain samples that lie within the bulk of the posterior and exclude outliers, essentially enforcing a step function prior which does not invalidate the use of the Flow in generating samples.

4.4.2.1 Gaussian Mixture Model

Gaussian mixture modelling is a probabilistic method used to represent a multimodal data set as a superposition of unimodal Gaussians [292]. The result is a GMM, a probability density function of a weighted sum of Gaussian component densities. This method provides a non-parametric method to model the general shape of a distribution and identify regions of high/low probability. It is classed as a supervised ML method.

Given dataset $\boldsymbol{x}$, we may represent the data set with a GMM composed of M

Gaussians, expressed as

$$p(\boldsymbol{x}) = \sum_{i=1}^{M} w_i \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu_i}, \Sigma_i) \tag{4.4}$$

where w_i are the individual weights of each Gaussian component, normalised such that $\sum_i^M w_i = 1$. The Gaussian distribution representing the i^{th} Gaussian $\mathcal{N}(\boldsymbol{x}|\mu_i, \Sigma_j)$ has mean $\boldsymbol{\mu_i}$ and covariance matrix Σ_i [293]. Each Gaussian component is represented by the function

$$g(\boldsymbol{x}|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu_i})^T \Sigma_i^{-1}(\boldsymbol{x} - \mu_i)\right\}. \tag{4.5}$$

By iterating through sets of GMM parameters to find the optimal set of parameters which maximise the probability of the GMM at the equation of state training data points, $p(\boldsymbol{\Omega}_{\text{training}}) = \sum_{i=1}^{M} w_i \mathcal{N}(\boldsymbol{\Omega}_{\text{training}}|\boldsymbol{\mu_i}, \Sigma_i)$, we find the optimal 10-dimensional GMM to represent our 10-dimensional equation of state training data [294].

We used the `scikit-learn` implementation of GMM to model the 10-dimensional reduced equation of state training prior distribution, $\boldsymbol{\Omega}$, with a superposition of 22 Gaussian distributions. In using GMM to model this distribution, we can set a probability threshold defined by the least probable training prior point in the 10-dimensional space, $\min(p(\boldsymbol{\Omega}_{\text{training}}))$, which has a given probability density as determined by equation 4.4. This sets a threshold, beyond which samples are lower probability than our most extraneous training samples, and therefore are outside of our region of interest. We chose to use 22 Gaussian components to model the training prior equation of state distribution so to ensure the modelling of the space was detailed and all features were accounted for. An equivalent result could have likely been achieved with a smaller number of Gaussian components, however the analysis remains lightweight. Figure 4.4.6 shows the training prior for equation of state plus auxiliary parameter data in blue alongside the GMM in purple. The 1−dimensional histograms show that GMM emulates the equation of state training data accurately.

The GMM was evaluated via equation 4.4 at the equation of state posterior samples $\boldsymbol{\Omega}_{\text{GW170817}}$ returned by the Flow when conditionally sampling via $\boldsymbol{\Omega}_{\text{GW170817}} \sim p(\boldsymbol{\Omega}|\theta_{\text{GW170817}})$ using both the complex model and simple models. The probability

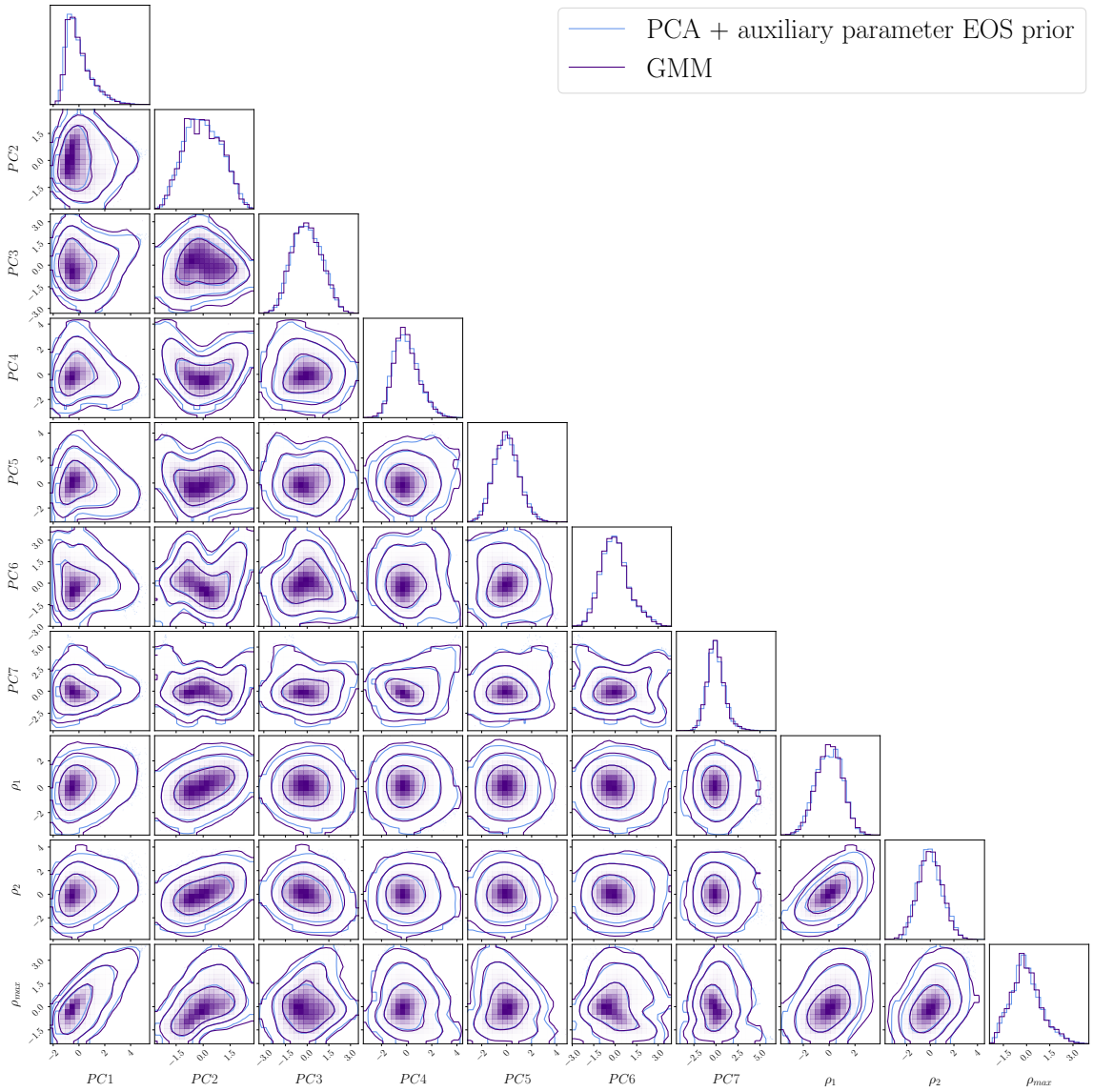


Figure 4.4.6: Corner plot of 1- and 2-dimensional marginalised distributions of the 10-dimensional equation of state training data distribution (blue) and the GMM (purple), composed of the superposition of 22 Gaussians, constructed to emulate this. The contours contain 68%, 95%, and 99.7% of the probability.

threshold with respect to which samples are evaluated are the same for both models, as this is set relative to the underlying training data distribution. Table 4.4.2 presents figures of merit which have been evaluated with respect to both the simple and complex models. As can be found in table 4.4.2, 82.25% of equation of state plus auxiliary parameter posterior samples $\Omega\ p(\Omega|\theta_{\text{GWTC-1, in hull}})$ were rejected relative to the GMM probability threshold for the complex model. This demonstrates the complex Flow is returning equation of state posterior samples of low probability,

beyond the bounds of our prior training data distribution. When projected back to the $P - \rho$ plane in which we normally present the equation of state, we see the result of retaining low probability samples. Figure 4.4.2 (a) again highlights the result of converting the output of the Flow from scaled PCs to equations of state in the $P - \rho$ plane. It is clear that low probability samples are overwhelming the equation of state posterior distribution and push beyond the training prior bounds; these present as equations of state which disobey the physical requirements of an equation of state, namely thermodynamic stability which is maintained by having an equation of state which is continuous and where pressure increases monotonically with density.

In contrast, when using the simple model, the GMM threshold rejects 15.64% of the equation of state posterior samples generated when conditionally sampling with GWTC-1 $[m_1, m_2, \log \Lambda_1, \log \Lambda_2]$ samples, a significantly smaller quantity compared to the complex model, suggesting the simpler model finds it easier to conditionally sample from the Flow within the training prior bounds defined by the equation of state training data which is modelled by the GMM.

Samples close to the GMM boundary may also have relatively low log-probability relative to the general distribution of equation of state posterior samples. Again, this is inherent to the nature of the Flow and the lack of hard threshold in the 10-dimensional equation of state data space, which is instead modelled loosely by a GMM probability threshold. As a result, relatively low log-probability equation of state posterior samples may produce equations of state in the $P - \rho$ plane which subtly extend beyond the training prior bounds. As discussed previously, these are the bounds of what we believe to be physically valid equations of state, and so any equation of state which extends beyond these bounds can also be rejected. Therefore, we perform one final stage of sample rejection which consider any equation of state in the $P - \rho$ plane which extends beyond the most stiff and soft equations of state (the pink bounds in figure 4.2.1). This accounts for the lack of explicit boundary in the 10-dimensional equation of state data representation. From the edges of the equation of state prior training data distribution in the $P - \rho$ plane, 15.69% of remaining (post-GMM) samples from the complex model are rejected from beyond the edges of

the training prior space, and 6.85% of samples from the simple model.

Each model returns less equation of state posterior samples than $[m_1, m_2, \Lambda_1, \Lambda_2]$ event samples that were used in the analysis, however the initial large model returns equations of state from only 26% of the original $\theta_{\text{GWTC-1}}$ set and the current model returns 75%. It is evident that, without the post-processing steps, the initial large model would have returned an almost entirely invalid equation of state posterior distribution, with respect to known physical constraints. This was curtailed by training of a simpler model, which better understood the general features of the equation of state space and whose equation of state posterior samples survive post-processing steps. However, an additional merit of the Flow method is that, even if one is limited by the number of conditional samples from a given event, the Flow can return any number of equation of state posterior samples. Each conditional sampling of the latent space is unique, and so one can sample the latent space distribution N times with the same conditional sample θ, and return N independent equation of state posterior samples from the given distribution, as you can sample as many times as you wish from the learned distribution (sampling the same number of times for each condition).

	Complex model	**Simple model**
# rejections via GMM	2776 (82.25%)	528 (15.64%)
# rejections via prior bounds	94 (15.69%)	195 (6.85%)

Table 4.4.2: Table to compare complex and simple models on the number of samples retained after pre- and post-processing stages.

4.5 Results

Results below are presented using the simple model. We present the equation of state posterior when conditioning on GWTC-1 posterior samples from GW170817 (these are samples which lie within the bounds of the convex hull only), along with injection studies to test the Flow's ability to recover the true equation of state.

4.5.1 GW170817

The GWTC-1 samples from GW event GW170817 make no prior assumptions with regards to a common equation of state or enforce the constraint that $m_1 \geq m_2$. Similarly, no universal relations are incorporated in the PE, so we assume no inherent correlation between the tidal parameters. We collect the mass and tidal parameters for use in the Flow; for this one event, we have ≈ 8000 samples in total which are then passed through the convex hull, such that we keep all samples which belong within the prior bounds over which our Flow was trained. We also enforce the convection that $m_1 \geq m_2$ and therefore that $\Lambda_1 \leq \Lambda_2$.

Equation of state posterior samples are returned by conditionally sampling with the Flow $\Omega_{\text{GW170817}} \sim p(\Omega|\theta_{\text{GWTC-1, in hull}})$. Equation of state samples Ω_{GW170817} are then evaluated with respect to the probability threshold set by the GMM and some are rejected. Finally, equations of state are then projected into the $P - \rho$ plane and assessed with respect to the training prior bounds in the $P - \rho$ plane. The result is 2500 valid equations of state which have been collectively sampled by the Flow in $\mathcal{O}(0.1)$ s. Once sampled in the reduced equation of state representation, it takes < 1 s to convert the posterior samples to equation of state curves in the $P - \rho$ plane.

The final equation of state posterior is presented in figure 4.5.1, presented alongside the result published by the LIGO-Virgo collaborations in [4]. We present results as 50% and 90% confidence intervals of density as a function of pressure, alongside cumulative probability densities ρ_1, ρ_2 and ρ_{max} and by interpolation P_1, P_2 and P_{max}. The confidence intervals derived from our equation of state posterior presented here agrees well with the result presented in [4], with only slightly tighter constraints in the mid-density region. The deviation of our result from that presented in [4] appears to be most constraining at the region of density where the central densities of the NS components are most likely to lie, in the range $[8 \times 10^{14}, 1 \times 10^{15}]$ g cm^{-3}.

The slight discrepancy between the results could possibly be due to the difference in the training prior distribution; the lower boundary on component mass prior is identical in both analyses, but differs in the maximum mass upper boundary [83]. The

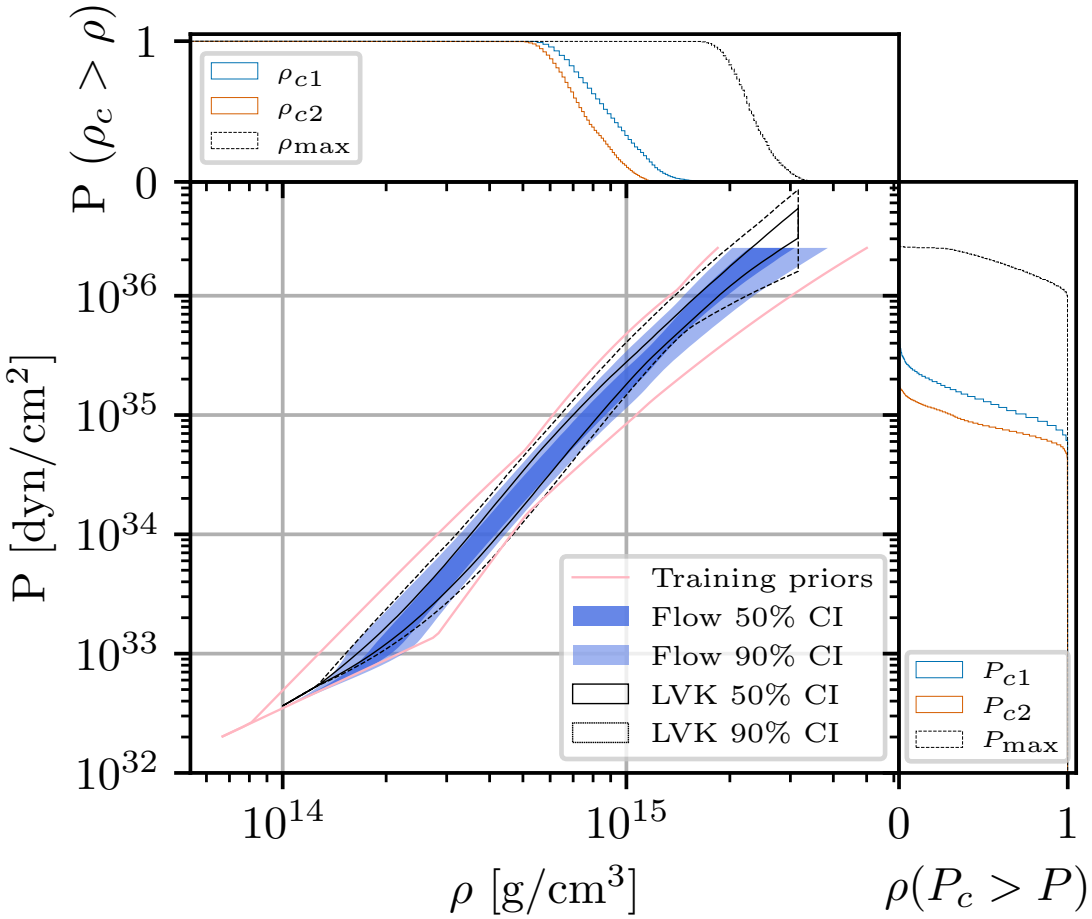


Figure 4.5.1: Plot of density ρ vs. pressure P of the equation of state posterior returned by the Flow, compared to the result of the LIGO-Virgo collaboration. The plot shows 50% and 90% confidence intervals in dark and light blue for the equation of state posterior returned by the current Flow model when conditionally sampling with $m - \Lambda$ posterior samples in mass and tidal deformability from GW170817 that lie within the bounds of our convex hull. The spectral equation of state result from [4] is given in black solid and dotted lines for the 50% and 90% confidence bounds. The cumulative probability densities in component central and maximum density and pressure from the Flow analysis are plotted above and to the right, respectively. The equivalent information is not available from the LIGO-Virgo analysis. The training prior bounds, representing the stiffest and softest equations of state in our training prior are given in light pink.

bounds on the uniform mass prior used by the LIGO-Virgo collaboration in recovering the mass and tidal deformability posterior samples are 0.5 $M_\odot \leq m_1$ and $m_2 \leq 7.7\ M_\odot$. The result we compare to in 4.5.1 is that which involves sampling the parameters of a spectral parameterisation of the equation of state directly in PE alongside the masses, tidal deformability, and other parameters $\theta_{\text{ext.}}$ (which we consider extrinsic for this

analysis), for the joint posterior probability $p(\theta_{\text{ext.}}, m_1, m_2, \gamma_0, \gamma_1, \gamma_2, \gamma_3)$. The upper mass boundary in this analysis is defined by the maximum allowed mass for each equation of state in the training data set. The maximum of these maxima therefore define our maximum mass boundary as 3.07 $M_\odot$. Therefore, the mass priors for both analyses are different. For this analysis, we keep only the GW170817 samples that belong within the bounds of the training prior distribution, whereas the analysis performed by the LIGO-Virgo collaboration do not make explicit cuts; their physical requirements placed upon the equations of state are inbuilt into the equation of state parameterisation they sample the posterior of, so does not need to be explicitly enforced [4], and so this introduces further variability between the the analyses.

We note the speed and computational efficiency of the analysis: once posterior samples associated to an event are available, the Flow analysis can return an equation of state posterior in the $P - \rho$ plane in < 1 second. This does not add additional complexity or computational demand to the existing sampling process of macroscopic parameters that are standard data products of a GW observation. The training process of the Flow is the main temporal and computational demand of the workflow, however only needs to be completed once for use in all subsequent BNS merger events.

4.5.2 Testing on perfect measurements

When we have a distribution of event samples $p(\theta|h)$ from PE from a BNS merger, we carry the error of the measurement through the Flow and present the equation of state posterior with an error which is due to the Flow's uncertainty but also that of the PE itself. By using an infinitely precise measurement, we present only the uncertainty due to the Flow. Therefore, to assess the quality of the Flow's inference, we present the result of conditionally sampling the Flow given a single previously-unseen $[m_1, m_2, \Lambda_1, \Lambda_2]$ posterior sample. By sampling with one condition only (but repeatedly), we mimic the scenario of an infinitely precise measurement of the parameters of a BNS merger.

We generate an event associated to a true underlying equation of state from our validation data set; the Flow has not been trained on this data and as such we can

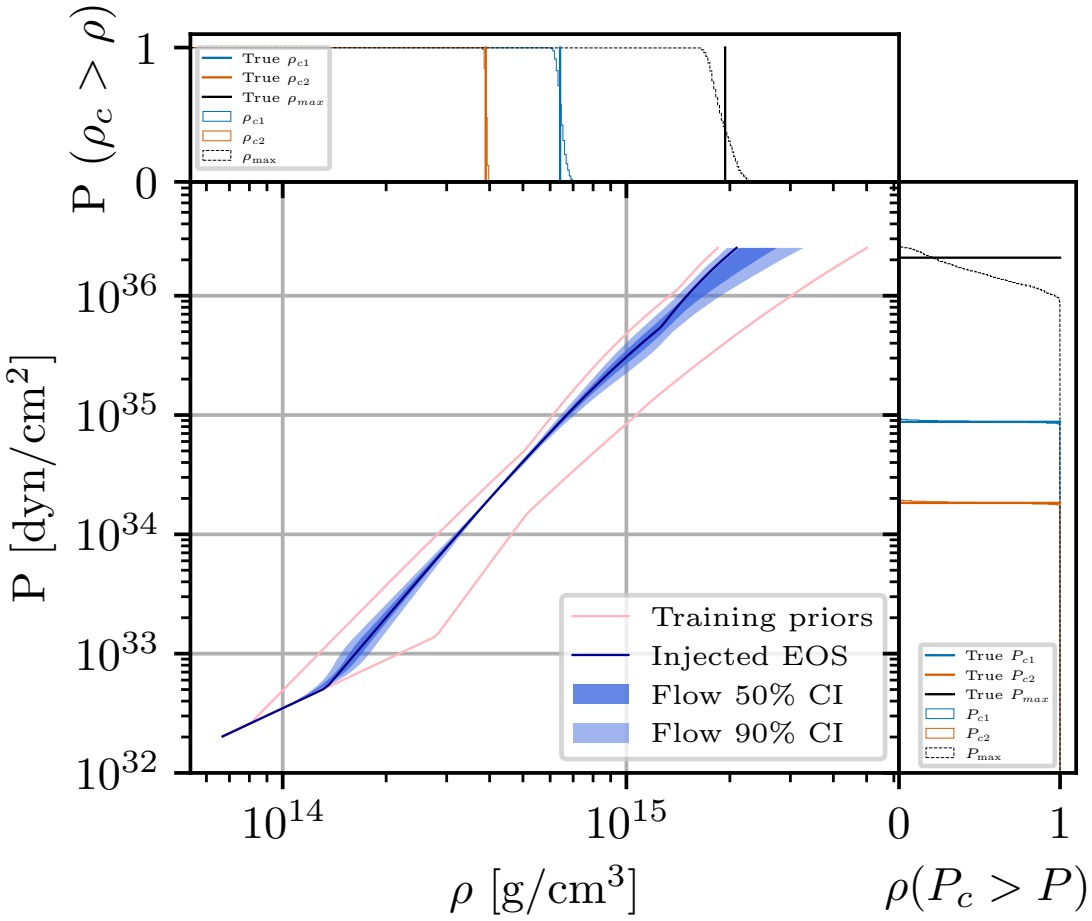


Figure 4.5.2: Plot of density ρ vs. pressure P of the equation of state posterior returned from repeatedly conditionally sampling with a single $[m_1, m_2, \Lambda_1, \Lambda_2]$ sample. 50% and 90% confidence intervals are presented in dark and light blue for the equation of state posterior. The known equation of state, given in dark blue. The cumulative probability densities in component central and maximum density and pressure are plotted above and to the right, respectively. The training prior bounds, representing the stiffest and softest equations of state in our training prior are given in light pink.

use it for validating performance. Figure 4.5.2 presents the result of evaluating the Flow with a 'perfect' measurement of a BNS merger. The true underlying equation of state that the event is associated to is presented in dark blue, and this remains in the 90% credible intervals for the extent of the pressure grid of the equation of state. The equation of state posterior, presented in credible intervals, is found to constrain much more at a mid-range density region (around $4 - 5 \times 10^{14}$ g cm^{-3}) as opposed to at high density; upwards of 1×10^{15} g cm^{-3}. This demonstrates the Flow's uncertainty here, and at which densities current measurements are most constraining. It is not

currently expected that single event analyses will be able to constrain the high density equation of state particularly strongly, and as such the combination of multiple events is necessary to constrain more precisely. It is therefore evident that the Flow can recover the true equation of state with no GW measurement uncertainty present in the measurement.

4.5.2.1 Probability-probability plot

To validate the statistical consistency of the analysis, we present a p-p plot in figure 4.5.3. The figure presents 10 curves, each representing a dimension of the reduced equation of state space which the Flow returns a posterior in (7 principle components and 3 auxiliary parameters). The plot presents the fraction of true parameter values that lie within a given confidence interval as a function of confidence interval. Lines which trace the diagonal within the bounds of expected confidence intervals demonstrate statistical consistency of the analysis.

The plot is constructed by running the Flow with conditional data, $[m_1, m_2, \Lambda_1, \Lambda_2]$, which is associated to a known true equation of state. As a result, all tests involved in constructing the p-p plot are independent of noise on measurement of the parameters of the BNS system from the GW signal and allow us to verify the quality and constraining power of the Flow independently. Each sample y of event information is passed to the Flow in order to return 2000 draws from the equation of state posterior, which are then compared to the known truth. This process is repeated 100 times and constructs the results which are presented in figure 4.5.3, giving a combined p-value of 0.7301. The p-p plot presented in figure 4.5.3 indicates no biases as all parameters trace the diagonal (with some small level of noise) and all individual p-values are statistically consistent.

4.5.3 Injection study with simulated event

To further validate the ability of the Flow in returning an accurate and, in particular, statistically consistent equation of state posterior distributions, we simulate a BNS merger event associated to a true underlying equation of state. We wish to mimic

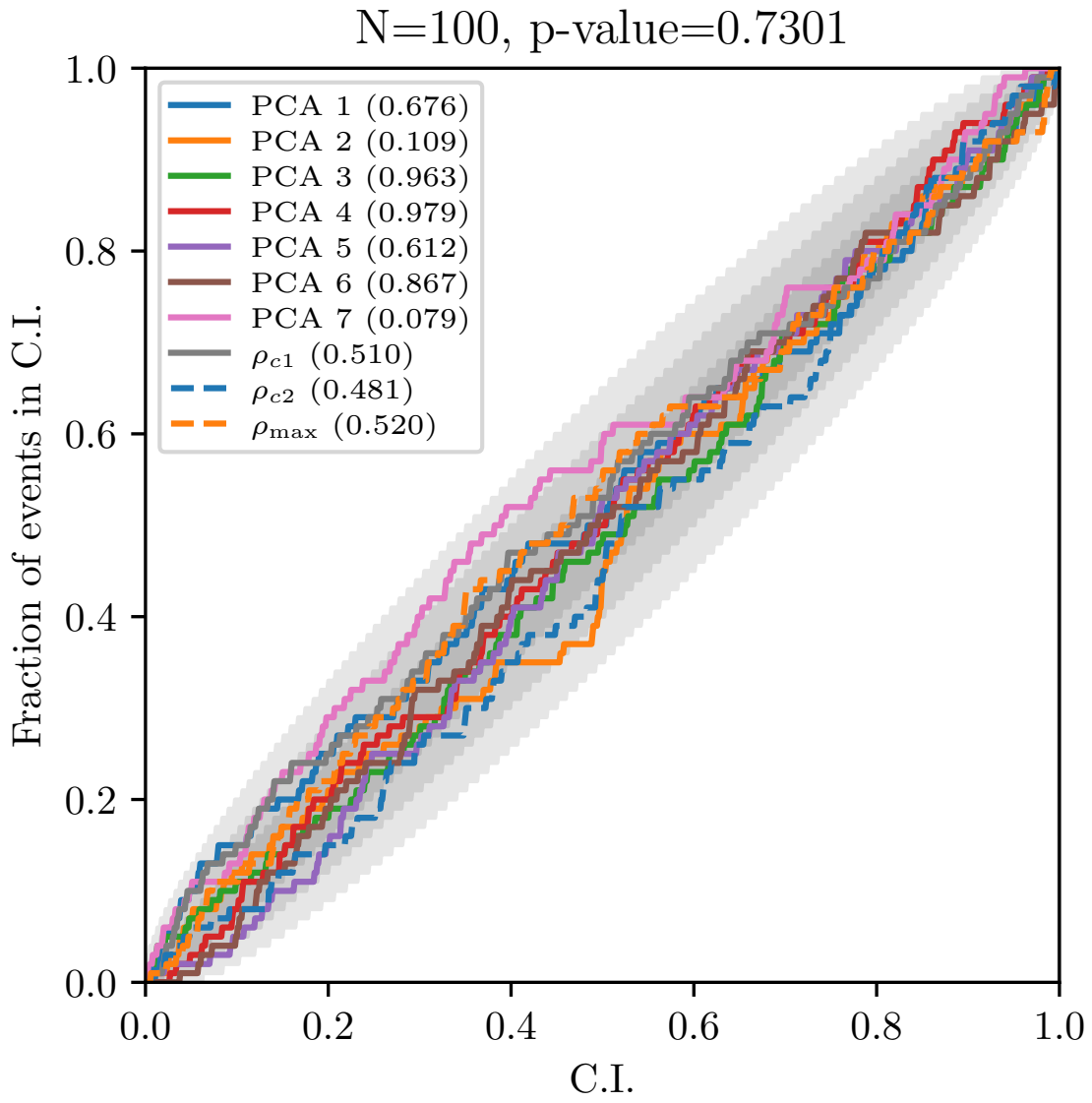


Figure 4.5.3: A p-p plot of the 10 parameters inferred by the Flow, which are the 7 principle components representing the equation of state and the 3 auxiliary parameters. The p-value for each individual parameter is given, alongside the overall p-value for the result of 0.7301. This result was found by running 100 different instances of the Flow with 'true' conditional data from the validation data set. The grey regions show the 1, 2 and 3σ confidence intervals for the curves.

the PE of a true GW event, where we know the true equation of state associated to the event, so to assess whether the Flow can identify the underlying equation of state, particularly when we have noise associated to the GW event. We process the simulated merger event using standard PE tools to provide conditional data such that the Flow can recover the true equation of state. Using the validation data set, we select an equation of state and set of auxiliary parameters and sample an associated

Waveform parameter	True value
Component spins, $\chi_{1,2}$	0
Inclination angle, θ_{jn}	0.4
Orbital phase at coalescence, ϕ_c	1.3
Polarisation angle, ψ_c	2.659
Time of coalescence, t_c	1126259642.413
Luminosity distance, D_L (Mpc)	70 Mpc
Right ascension, α	1.375
Declination, δ	-1.2108

Table 4.5.1: Table of true parameters used to simulate a GW signal in order to perform PE in component masses and tides.

set of event information $[m_1, m_2, \Lambda_1, \Lambda_2]$. We use the event information as an injection in GW PE with the Bayesian PE software `bilby` [295]. We provide bilby with the macroscopic parameters associated to our true equation of state $[m_1, m_2, \Lambda_1, \Lambda_2]$ alongside additional parameters which are required for the injection but which are equation of state-insensitive so are selected for the convenience of the sampling. These are presented in table 4.5.1 and are chosen to match that of a standard BNS example injection provided by `bilby` software [295, 296]. We initialise `bilby` to analyse a 4 second signal centred on the merger time at a frequency of 4096 Hz. The choice of a 4 second signal lies in the assumption that the majority of our information from the GW merger of the two NSs comes from the late stages of BNS inspiral, where the tidal parameters are influential the gravitational waveform (see chapter 1 section 1.3.1). The injection parameters, presented in table 4.5.1, are used to generate a waveform model in the `IMRPhenomD_NRTidal` [297] waveform family which is injected into simulated detector noise. Assuming a 3 detector network, `bilby` then performs PE using, in this case, the nested sampling algorithm `nessai` [298, 256] to return posteriors on the parameters of the system (presented in figure 4.5.2). This simulated signal is of particularly high mass, especially in m_1, and has been processed to be within the bounds of the convex hull, defined by the training prior bounds in $[m_1, m_2, log(\Lambda_1), log(\Lambda_2)]$, represented by the purple distribution in figure 4.5.4. As such, the posterior samples from the simulated event remain within the bounds of the conditional training data prior, as can be seen in the reduction of volume between the pink and purple distributions.

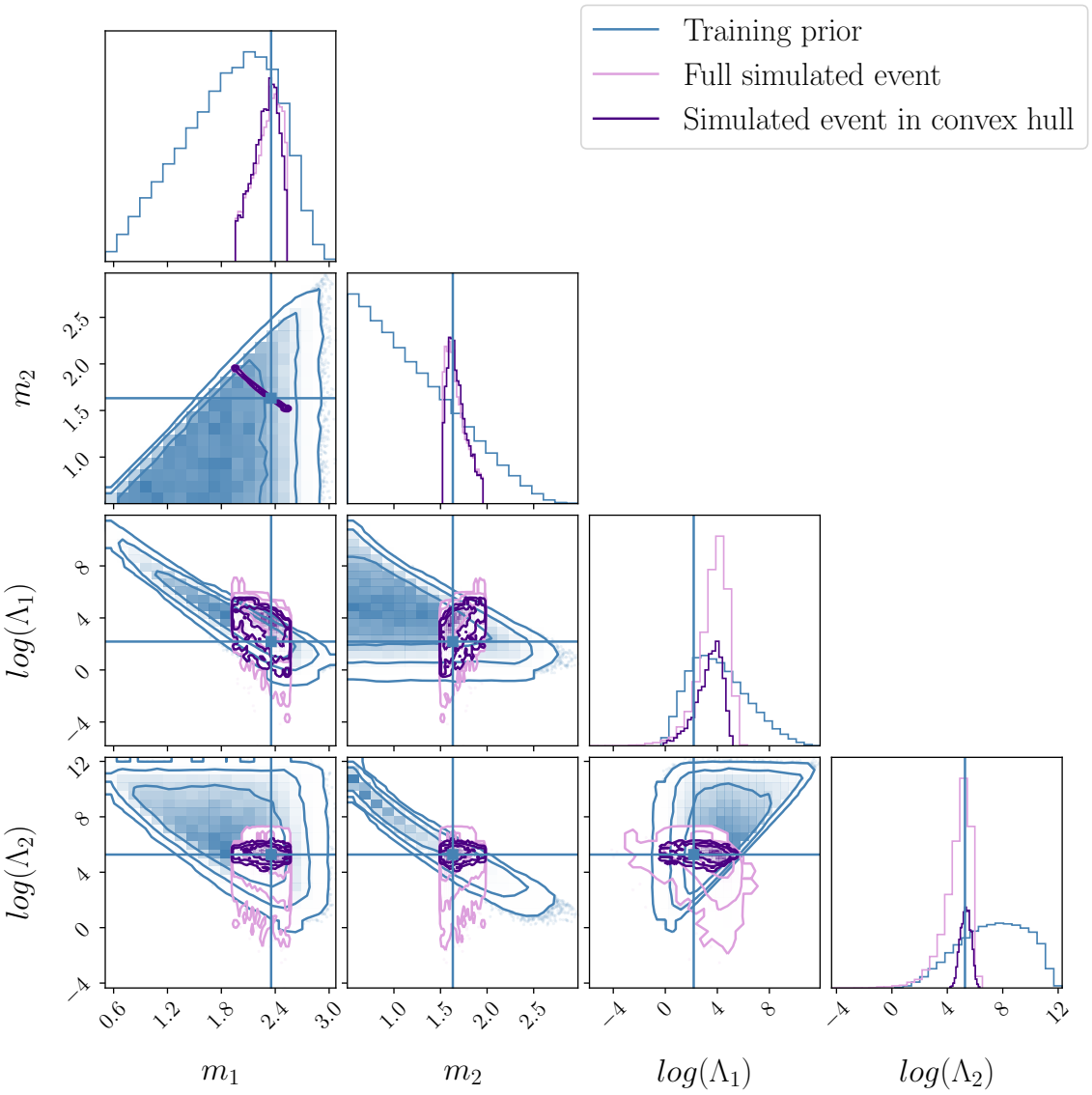


Figure 4.5.4: Corner plot of 1- and 2-dimensional marginalised distributions of $(m_1, m_2, log(\Lambda_1), log(\Lambda_2))$ conditional training data (blue) and all posterior samples (pink) from a simulated event associated to a known equation of state alongside samples within the bounds defined by the convex hull (purple). The contours contain 68%, 95%, and 99.7% of the samples.

From bilby, we obtain 9230 posterior samples θ, where mass and tidal parameters are expressed as chirp mass $\mathcal{M}$, mass ratio q, combined dimensionless tidal deformability $\tilde{\Lambda}$ and a parameter which describes the difference in the individual dimensionless tidal deformabilities $\delta\tilde{\Lambda}$. This is a common choice in PE for more efficient sampling. We undo this parameterisation and retain only the component masses and tidal parameters $[m_1, m_2, \Lambda_1, \Lambda_2]$. These posterior samples from the simulated event are passed through the convex hull, so to retain only those which lie within the training

prior bounds of our Flow in the $[m_1, m_2, log(\Lambda_1), log(\Lambda_2)]$ space. The latent space, $\mathcal{Z}$, is sampled conditionally given posterior samples θ from the simulated GW event to return equation of state posterior samples in the data space, $\mathbf{\Omega} \sim p(\mathbf{\Omega}|\theta)$. Posterior samples outwith the training prior bounds in the equation of state space (relative to the GMM and the training prior bounds in the $P - \rho$ space) are rejected and we present the final equation of state posterior in figure 4.5.5. Similar to figure 4.5.1, the equation of state posterior is presented in confidence intervals of 50% and 90% alongside the true equation of state in dark blue. The true equation of state remains within the 90% confidence intervals for the entirety of the density range that all equations of state span, and the level of constraint remains comparable to that of the GW170817 posterior in figure 4.5.1. Likewise, the 1-dimensional cumulative posterior densities for ρ_1, ρ_2 and $\rho_{\max}$ agree with the true values, given by vertical lines.

This test demonstrates that the Flow understands the mapping from macroscopic parameters to equation of state, and can recover the true underlying equation of state, which gives confidence in the validity of the equation of state posterior presented for true events.

4.6 Conclusions

This work demonstrates the success of using Normalising Flows to rapidly infer the NS equation of state using observations of BNS mergers. Observations of BNS mergers provide us with the opportunity to measure the tidal deformability of each component of the system, which is a direct measure of the presence of matter in the system, which can be used to complement existing electromagnetic analyses to infer the ultra-high density nuclear equation of state. To complement collaborative discovery, rapid analyses are necessary so to communicate findings quickly and make electromagnetic astronomers aware of potential follow-up opportunities. This analysis makes possible the rapid inference of the equation of state of single BNS merger events through Machine Learning methods.

The Flow is trained to learn the distribution of plausible equations of state

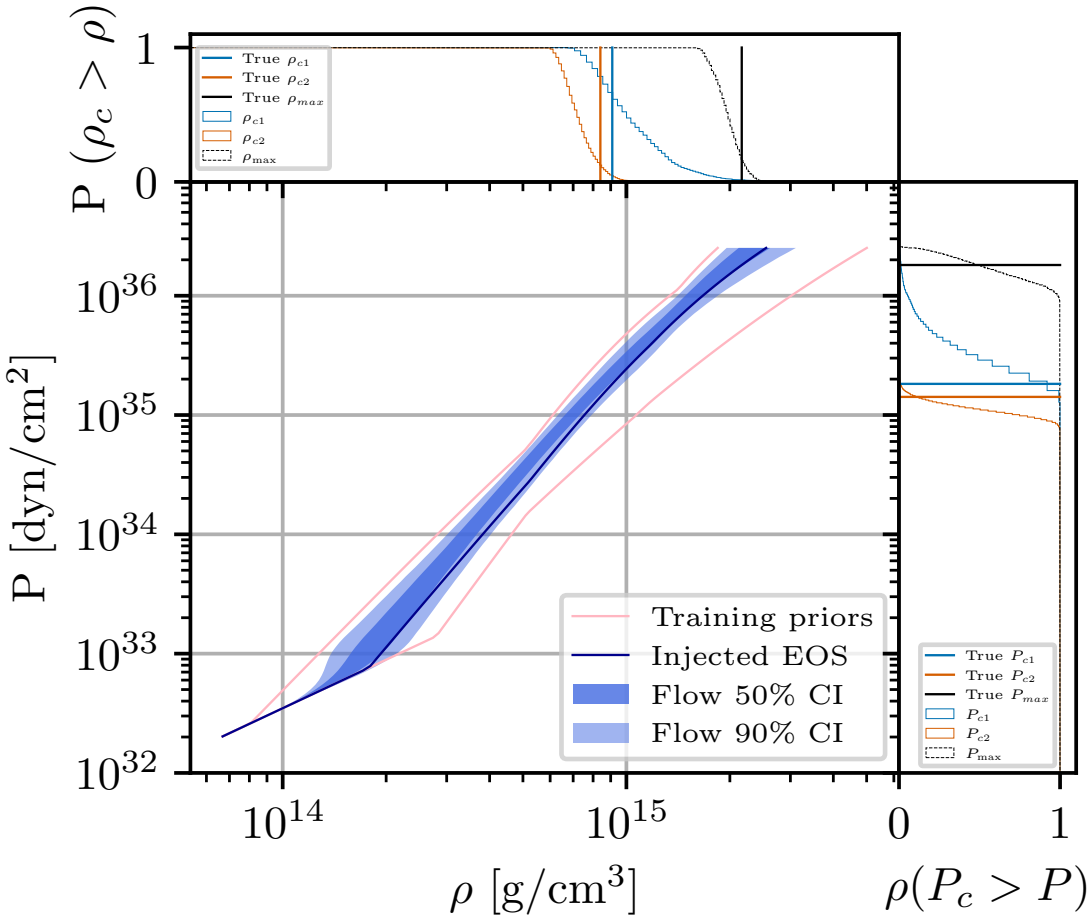


Figure 4.5.5: Plot of density ρ vs. pressure P of the equation of state posterior returned from conditionally sampling with simulated PE samples in $[m_1, m_2, \Lambda_1, \Lambda_2]$ associated to a true equation of state. 50% and 90% confidence intervals are plotted in dark and light blue for the equation of state posterior returned by the current Flow model and the true equation of state is given in blue. The cumulative probability densities in component central and maximum density and pressure are plotted above and to the right, respectively. The training prior bounds, representing the stiffest and softest equations of state in our training prior are given in light pink.

conditioned on event information associated to each equation of state, and as such can be used in inference to return $\mathcal{O}(2500)$ physically feasible equation of state posterior samples in ≈ 0.1 seconds. Pre- and post-processing steps are applied to ensure that both the conditional inputs to the Flow and the subsequent output are within the bounds of what is believed to be physically reasonable.

The method is explicitly model-agnostic: although the Flow is trained on piecewise polytropic equation of state data, the Flow is not trained to learn the parameters

of this chosen parameterisation and as such the output is model independent. The Flow equivalently could be trained on a different parameterisation or a mixture of many, and it is this choice of equation of state training data which defines the prior and as such can be as broad or as constraining as the user wishes. Additionally, the use of PCA to reduce the equation of state before passing to the Flow allows for model-independent compression and only compresses relative to the morphology of the equation of state and does not require equation of state-specific parameterisation of the data.

The Flow model was applied to the first gravitational wave observation of the merger of two NSs, GW170817. Future work could consider the comparison of the Flow result to that of standard stochastic sampling, so to truly assess the quality of the Flow in conditions consistent with the rest of the analysis pipeline. While the equation of state constraint in the $P - \rho$ plane using information from this event is small relative to the breadth of the prior, the constraints of this analysis broadly agrees with existing analysis published by the LIGO-Virgo collaborations [4] and is produced very rapidly. This also constitutes the result from a single BNS merger event: it is expected that future observing runs of current ground based observatories will bring observations of more BNS events [299] and with this we expect tighter constraints on the ultra-high density nuclear equation of state through the combination of information from these events. Normalising Flows are an ideal potential tool for the development of hierarchical inference methods, as will be demonstrated in the following chapters.

Chapter 5

Equation of state data: modelling and compression for inference

In chapter 2 section 2.1, we introduced different methods of modelling the NS equation of state and in section 2.2 we discussed how the quality of NS equation of state inference is influenced by the choice of model employed in the analysis. We now consider appropriate equation of state modelling and data compression methods with regards to NS equation of state inference given many observations of GWs from BNS mergers in the third generation of GW observation. In chapter 2 section 2.2.2, we introduced the likely increase of number of BNS mergers in third-generation detectors due to increased sensitivity; there are expected to be $\sim 10^5$ BNS mergers per year observed by ET alone [300]. The increase in number of observations would allow for unprecedented accuracy of inference of the NS equation of state, and calls for effective hierarchical analysis to combine information from all events into one overall equation of state inference from GW observations.

In third-generation GW observation with ET or CE, we expect to be able to improve our sensitivity to individual GW events such that we can resolve not only the combined dimensionless tidal deformability, $\tilde{\Lambda}$, well but also the difference in dimensionless tidal deformability, $\delta\tilde{\Lambda}$, in parameter estimation of the gravitational wave signal, which is not possible with current GW detectors [300]. See chapter 1

section 1.3.1 for the introduction of these parameters. Improved measurement of these parameters means that we can then resolve the individual tidal deformabilities of each of the component neutron stars more effectively, allowing for substantially improved inference of the NS equation of state with single events, which only improves as we accumulate many. In advance of this next generation of detection, it is pertinent to ensure that our inference schemes incorporate well-considered modelling methods of the NS equation of state, leaving room for the widely unknown high density region and allowing for flexibility in the extremities without making sweeping assumptions.

In current equation of state inference, phenomenological high density equation of state models are an acceptable choice in inference schemes due to its simplicity and therefore minimal number of parameters [146, 148]. These generic models account for our current lack of understanding of the high density equation of state and do not assume any physically motivated model. Phenomenological equations of state are usually stitched to a single unique crust equation of state, modelled by a physically motivated equation of state. However, for future equation of state inference with third generation gravitational wave detection the assumption of a fixed crust equation of state is no longer valid; in future gravitational wave detection, the sensitivity of detection has been suggested to be comparable to the error accrued with inconsistent crust-core equation of state matching [2]. Therefore, we need an equation of state model which allows for broad and generic treatment of the high density equation of state and makes no assumptions about the crust equation of state. In this chapter we employ the unified equation of states of Davis et al [2] which employ piecewise polytropic equations of state at high density to their own maxima and a nuclear metamodel at low density, down to the crust-core boundary at 4×10^{11} g cm^{-3}, described in section 5.1.

In order to employ this equation of state parameterisation, we wish to compress the equation of state data for use in inference methods. For use in a Normalising Flow framework similar to that of chapter 4 for example, we compress the equation of state data to reduce the number of parameters (or dimensions) we are performing inference over to make our computational methods more efficient. In this chapter, we introduce

using an autoencoder to compress the unified equation of state data. Previous work employed PCA to compress piecewise-polytropic-only equation of state data, but we find with the increased complexity of the unified equation of state data, we require more complex compression methods which will be discussed in section 5.2.

The goal of this chapter is as follows: we wish to employ our new equation of state data set and compress this using an autoencoder with convolutional layers. We then train a conditional Normalising Flow to learn the mapping from the complex multi-dimensional compressed equation of state data space to a multi-dimensional unit Gaussian, conditioned on BNS merger event parameters, m_1, m_2, Λ_1 and Λ_2. The Flow training follows that of chapter 4 but instead is used to conditionally *evaluate* the learned Flow data space instead of sampling, a distinction which will be discussed in section 5.4.

We find that Normalising Flows struggle to generalise in learning the compressed equation of state data space and appears to learn unforeseen artificial underlying structure. The following analysis focusses on the structure found in the compressed unified equation of state data, however we note that the structure was also present when using a different compression of this data and also when using the purely piecewise polytropic equations of state of chapter 4. We present only the results relevant to the newly introduced data, but note this is not a unique feature of this data and assert caution generally. We present the results of tests which aim to identify the artificial structure the Flow has learned and outline the attempts to mitigate artificial structure learning. In an effort to improve the generalisability of the Flow and mitigate the artificial structure learning, we introduce a simple augmentation method which uses an additional Normalising Flow during training to improve training quality. In summary, we present the need for caution when training a Normalising Flow on an expensive and potentially sparse training data set and demonstrate the flexibility of Normalising Flows for supporting hierarchical inference of the neutron star equation of state, performed in chapter 6.

5.1 Unified equation of state data

In this work, we choose to use unified equations of state of [2]. These equations of state are described by a meta model at low density and a piecewise polytrope at high density and are constructed with Crust Unified Tool for Equation-of-state Reconstruction (CUTER). The authors account for the current uncertainty in the high density equation of state by employing a standard approach of the widely accepted piecewise polytropic parameterisation of Read et al. [146]. In this work, we use piecewise polytropes with 3 pieces which are represented by 3 polytropic parameters, Γ_0, Γ_1 and Γ_2 and 2 joining densities ρ_1 and ρ_2, plotted in figure 5.1.2, where $\rho_2 > \rho_1$. In order to have a 'complete description' of the NS EOS while allowing for uncertainty in the high density equation of state - that is, each equation of state is individually unique over its full density range and is described by the same nuclear model throughout - the authors extract nuclear parameters from the high density equation of state near the crust-core transition density to be used in constructing consistent low density equations of state with a meta model [301, 149], matched at a baryon number density of n_{match}. A set of $\sim 10^5$ 3-piece polytropic equations of state (introduced in chapter 2 section 2.1) are generated with transition densities randomly to be between n_{match} and $10n_{\text{sat}}$, 10 times nuclear saturation density expressed in terms of the baryon number density, n_B. The polytropic indices are chosen randomly in the range $[0, 8]$.

The nuclear parameters (likewise introduced in chapter 2 section 2.1) associated to the piecewise polytropic equations of state are extracted below but near n_{sat}. At these low densities, we can make safe assumptions about the nucleonic content of the matter and therefore define the energy per baryon e_{nuc} of nucleonic matter at zero T as an expansion of the asymmetry parameter δ

$$e_{\text{nuc}}(n_B, \delta) = e_{\text{is}}(n_B) + e_{\text{iv}}(n_B)\delta^2 + t^*_{\text{FG}}(n_B, \delta) \tag{5.1}$$

where

$$e_{\text{is}}(n_B) = (n_B, n_{\text{sat}}, E_{\text{sat}}, K_{\text{sat}}, Q_{\text{sat}}, Z_{\text{sat}}, m^*_{\text{sat}}, \Delta m^*_{\text{sat}}, b) \tag{5.2}$$

$$e_{\text{iv}}(n_B) = (n_B, n_{\text{sat}}, J_{\text{sym}}, K_{\text{sym}}, L_{\text{sym}}, Q_{\text{sym}}, Z_{\text{sym}}, m^*_{\text{sat}}, \Delta m^*_{\text{sat}}, b) \tag{5.3}$$

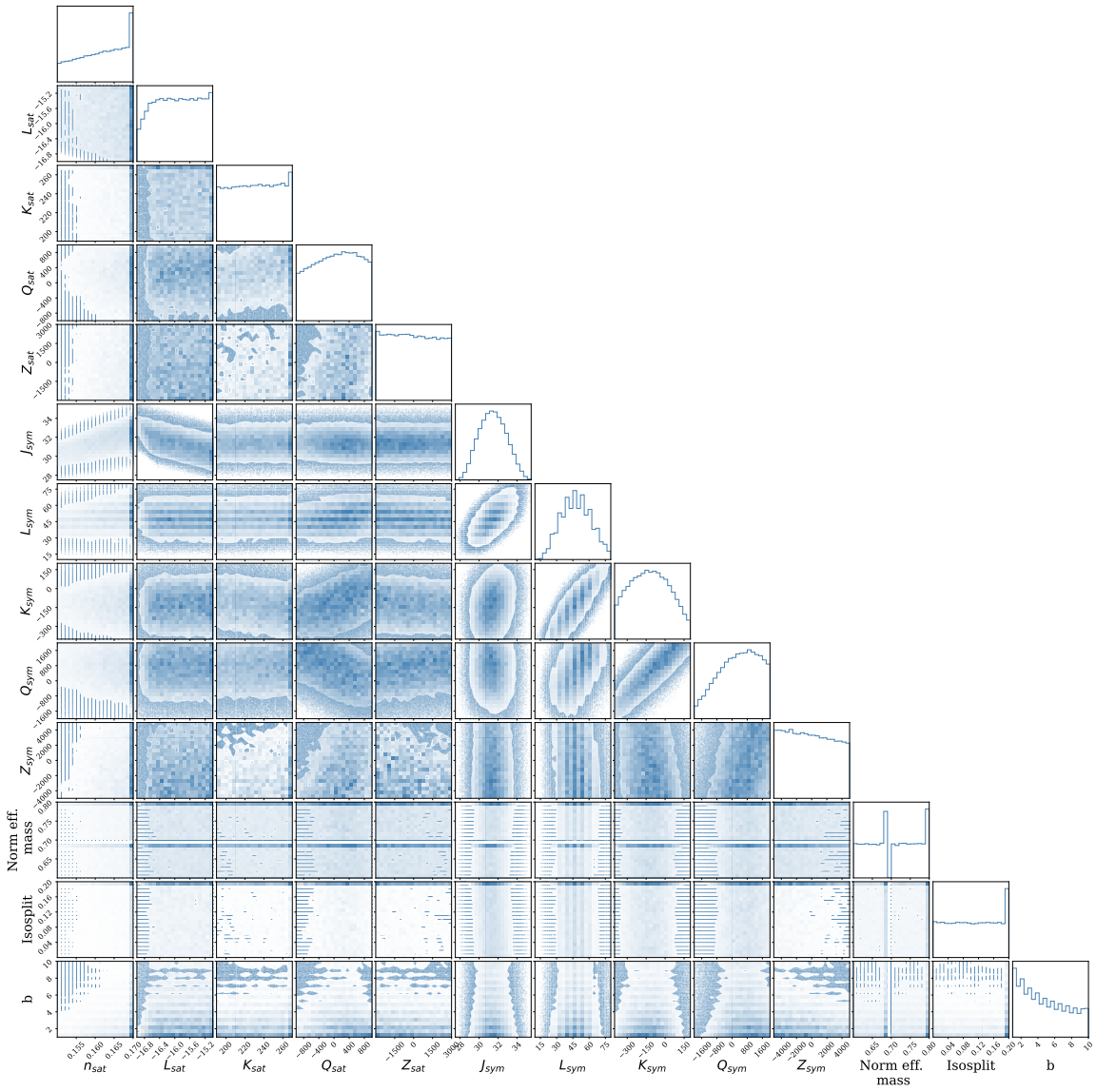


Figure 5.1.1: Corner plot of 1- and 2-dimensional marginalised distributions of the nuclear parameters used to construct the low density region of the neutron star equation of state training data as described in [2].

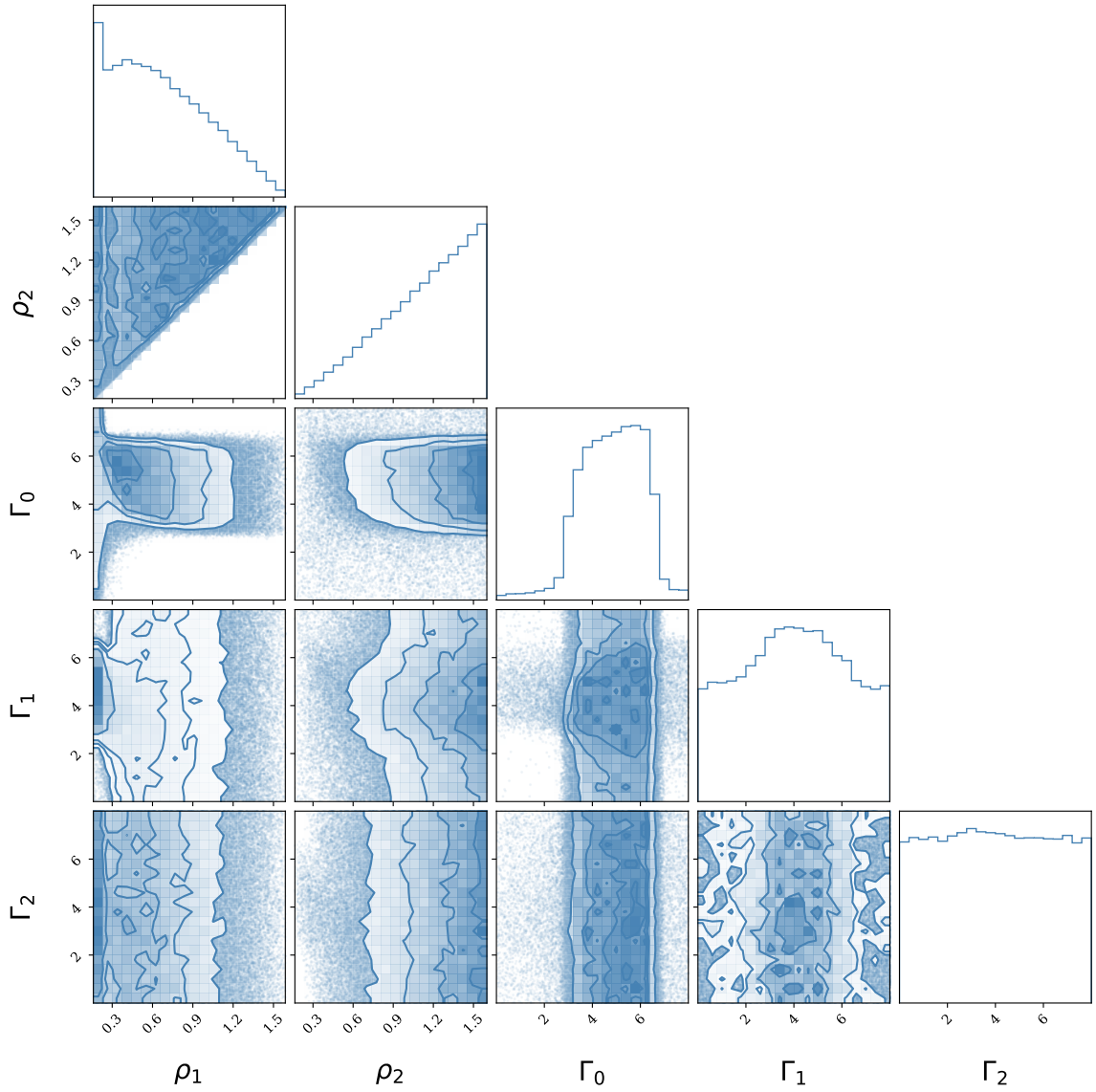


Figure 5.1.2: Corner plot of 1- and 2-dimensional marginalised distributions of the polytropic parameters used to construct the high density region of the neutron star equation of state training data using the method outlined in [2].

are the isoscalar and isovector energies respectively, and $t^*_{\rm FG}$ is a kinetic energy term [2]. The terms in equation 5.1 are functions themselves of the baryon number density n_B and nuclear parameters presented in figure 5.1.1: $m_{\rm sat}$ and $\Delta m^*_{\rm sat}$ being the effective mass in symmetric matter and the isosplit in neutron matter, respectively, $E_{\rm sat}$ the energy of symmetric matter, $K_{\rm sat}$ the isoscalar compressibility, $Q_{\rm sat}$ the isoscalar skewness, $Z_{\rm sat}$ the isoscalar kurtosis, J or $J_{\rm sym}$ the symmetry energy, $L_{\rm sym}$ the slope of the symmetry energy, $K_{\rm sym}$ the isovector incompressibility, $Q_{\rm sym}$ the isovector skewness, $Z_{\rm sym}$ the isovector kurtosis and b which ensures the zero density limit [159]. These parameters include those which can be inferred from nuclear experiment (introduced in chapter 2 section 2.2.1) and are used to construct detailed nuclear-physics-informed equations of state. For details on these parameters and how they are used for equation of state construction we refer the reader to [2, 149, 301]. We wish to know the energy per baryon $e_{\rm nuc}$ at different levels of asymmetry, δ, where

$$\delta = \frac{n_n - n_p}{n_B}. \tag{5.4}$$

for n_n and n_p, the number density of neutrons and protons respectively. After extraction of these parameters from the high density piecewise polytropic equations of state at $n < n_{\rm match}$ in the ranges provided in table 5.1.1, one can construct the set of low density equations of state using the meta-model of [149] up to $n_{\rm match}$. At the stitching point, the monotonicity of pressure is ensured [2]. The density of $n_{\rm match}$ is chosen to be above $n_{\rm sat}$, or around the crust-core transition density so not to allow for phase transitions (not restricted from the polytropic equations of state) at high density.

The result is a training data set of 97 000 'consistent and unified crust-core' [2] equations of state, presented in figure 5.1.3. The production of this data takes $O(1)$ day and we are limited by off-site production of said data. Each equation of state in the figure consists of 300 points in pressure on a fixed density grid beginning at 4×10^{11} g cm^{-3}, signifying the boundary between the inner and outer crust, given the outer crust of the neutron star is currently well understood and able to be investigated experimentally (see chapter 2 section 2.2). Each equation of state has its own minimum pressure $P_{\rm min}$, maximum allowed pressure $P_{\rm max}$, and maximum allowed density $\rho_{\rm max}$.

Parameter	Range
$n_{\rm sat}$	[0.15,0.17]
$E_{\rm sat}$	[-17.0-15.0]
$K_{\rm sat}$	[190.0,270.0]
$Q_{\rm sat}$	[-1000.0,1000.0]
$Z_{\rm sat}$	[-3000.0,3000.0]
$J_{\rm sym}$	[26.0,38.0]
$L_{\rm sym}$	[10.0,80.0]
$K_{\rm sym}$	[-400.0,200.0]
$Q_{\rm sym}$	[-2000.0,2000.0]
$Z_{\rm sym}$	[-5000.0,5000.0]
m^*/m	[0.6,0.8]
$\Delta m^*/m$	[0.0,0.2]
b	[1,10]

Table 5.1.1: Table of ranges sampled for nuclear parameters used to construct the low density meta-model equation of state.

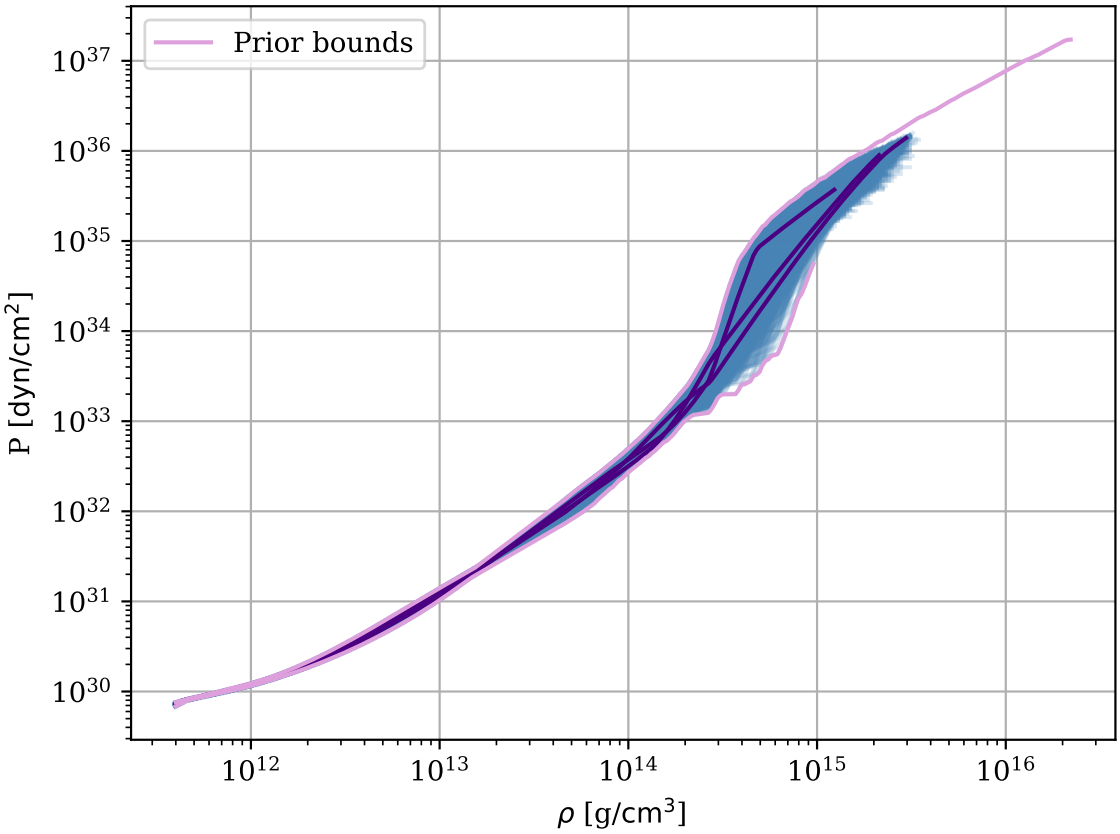


Figure 5.1.3: Training data set of 97 000 unified equations of state in pressure on a fixed grid of density in blue, beginning at 4×10^{11} g cm^{-3} extending to each equation of state's maximum allowed density and pressure. Three example equations of state are presented in purple. The softest and stiffest equations of state in the training data set are given by the lower and upper training prior bounds, respectively, in pink.

5.2 Compression of equation of state data

In chapter 4, when working with piecewise polytropic equations of state in high density only, we used PCA to compress the equations of state to a vector of 7 principle components and then performed the inference on the principle components (and auxiliary parameters) with the Normalising Flow (see section 4.2). The current unified training data equations of state are extended over a larger density range with more expressive behaviour - particularly at the transition from inner crust to outer core - and unique maximum allowed pressures and densities. We find compression via PCA smoothed out features of the equations of state, and did not retain the desired level of detail. Figure 5.2.1 shows the reconstruction of a given equation of state from the training data set after compression of the equation of state via PCA with numbers of PCs in the range $[7, 19]$. Despite the number of PCs used, the reconstruction fails to match the true equation of state at densities around 10^{14} g cm^{-3}. Figure 5.2.2 presents the residuals of the true equation of state and the same reconstructions for the range of numbers of PCs used. The left hand plot is presented at a different scale to the right hand so to show the quality of recovery at different density ranges of the equation of state. All reconstructions have a difference in pressure of 10^{32} dyn cm^{-2} from the true equation of state at a density of $\sim 10^{14}$ g cm^{-3}, a difference which manifests as a clear deviation from the truth. The right hand plot shows errors in pressure on the scale of 10^{34} dyn cm^{-2} at densities of $\sim 7 \times 10^{14}$ g cm^{-3}, with the largest error given by the smallest number of PCs, as expected. The maximum number of components tested (19), demonstrates the same error at low densities as the minimal number of components used and still has $\sim 10\%$ error at the highest density and pressure region. The reduction in error mostly appears to be due to the PCs not being able to express sharp changes in the equations of state, like phase transitions or boundaries, with small numbers of PCs, as would be desirable in this analysis. In principle, one could express infinite detail with infinite PCs, however to use PCA as a data compression method, we require $n_{\text{PCA}} \ll n_{\text{data points}}$.

To have high quality data compression with a lightweight and flexible method we instead use an autoencoder. Introduced in chapter 3 section 3.3, an autoencoder is

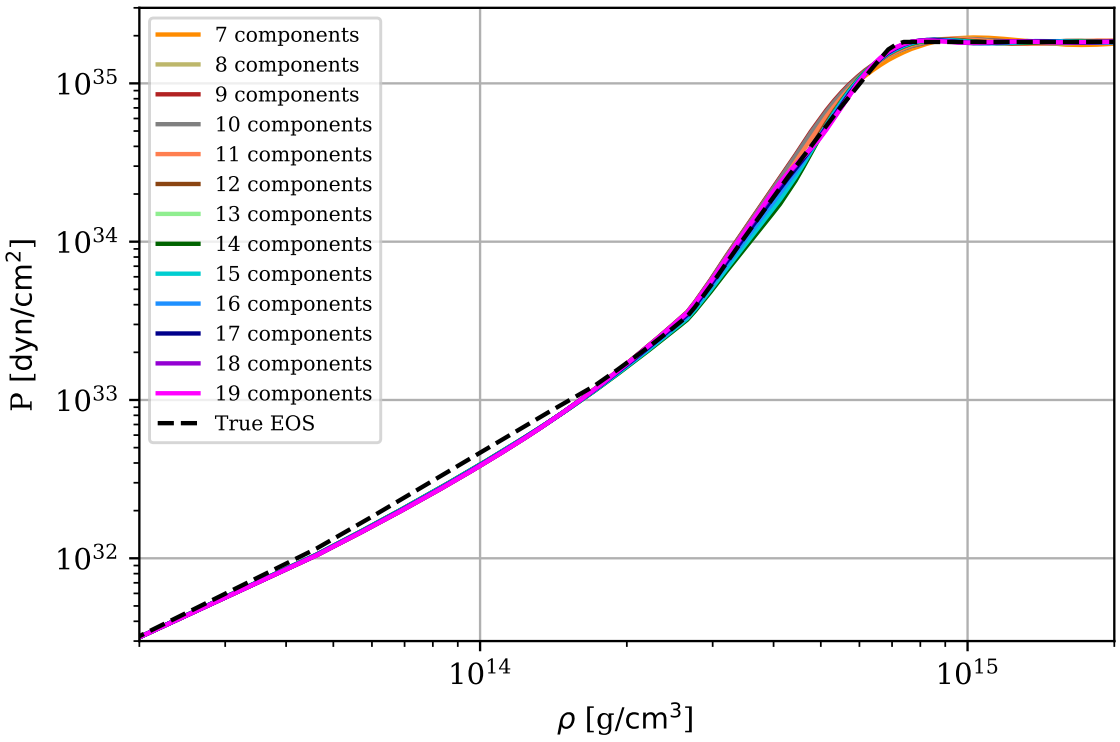


Figure 5.2.1: Plot of pressure as a function of density for an example equation of state from the training data set in the dashed black line and the reconstruction of the same EOS after PCA compression with different numbers of principle components in different colours.

a neural network architecture built in a bottleneck-like structure, with two discrete parts: an encoder and a decoder. Training data is passed to the encoder and is compressed through a series of hidden layers to a latent space of a user-defined size, usually smaller than the input data size. From the latent space, the data is passed to the decoder where the network then again expands back out in size, commonly in a path mirrored to the encoder, with the aim of reconstructing the network input. The network is trained to minimise the MSE loss between the input and the reconstructed output. In dimensionality reduction, we are interested in training the autoencoder to be able to encode our training data to a reduced-dimensionality representation of the input in the latent space. We can use the reduced representation for analysis and then use the decoder to express our result in the original data format.

We use an autoencoder with 2 initial convolutional layers in the encoder and compress the 300 points of each equation of state through the convolutional layers followed

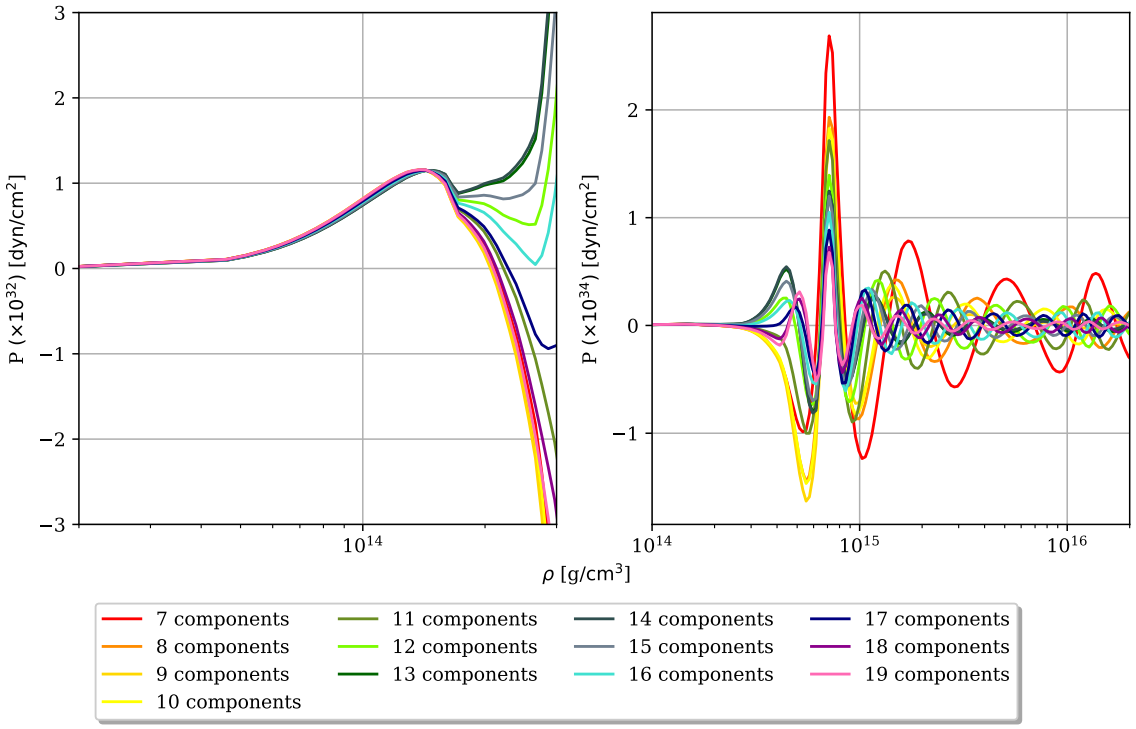


Figure 5.2.2: Plot of pressure as a function of density for the residuals of reconstructed equations of state after PCA compression with different numbers of principle components relative to the true equations of state, presented at different density regions and pressure scales. The LHS presents the residuals at the low density equation of state below 1.5×10^{14} g cm^{-3} and the RHS plot presents the residuals above 1.5×10^{14} g cm^{-3}. The pressure scales are different in each of the plots to reflect the different density ranges covered.

by a series of linear layers, gradually decreasing in size, to a final autoencoder latent space of 12 dimensions. We use ReLU activation functions between the convolutional and linear layers of the network network up to the final layer of neurons before the latent space, or bottleneck. Once trained, we can pass the original set of training data to the autoencoder and return the output of the encoder, presenting the training data in its reduced form. In order to compress the training data consistently, we first take the log of each equation of state in pressure along the log density grid to reduce the dynamic range. We then scale all equations of state to [0,1] along a [0,1] mock density grid by scaling each equation of state by its $P_{\rm min}$, $P_{\rm max}$, $\rho_{\rm max}$ and the global $\rho_{\rm min}$. These parameters are retained in the analysis so to recover the same initial level of detail. We discuss incorporating these parameters into the analysis later in this section. Finally, we subtract the mean equation of state from the [0,1] scaled set to

remove generic behaviour and allow the autoencoder access to reconstruct fine detail of each equation of state. This reduced training data set is then autoencoded to a latent representation of 12 dimensions.

The autoencoder was trained to minimise the MSE loss, discussed in chapter 3 section 3.3, between the input scaled training equation of state and the reconstructed output. We trained for 2000 epochs and achieved a final MSE loss of 10^{-6}, representing the error between the true and reconstructed equation of state averaged over the whole density range and the training data batch. A more intuitive error is that of the RMSE, where RMSE = $\sqrt{\text{MSE}} = 10^{-3}$, inferring we have an average error of 10^{-3} on each equation of state in the [0,1] scale. The training and validation losses are plotted in figure 5.2.3 where there is no evidence of over-fitting. We have very slight evidence of underfitting, which may be due to an over-simplified model or a lack of training data to fully describe the 12-dimensional space. This discrepancy is small, however, in relation to the losses achieved so we do not investigate further. Figure 5.2.4 presents the reconstruction of an example equation of state from the validation data set using the trained 12−dimensional autoencoder. In comparison to the reconstruction from the PCA-compressed equations of state, the autoencoder manages to reconstruct the equation of state to a higher scale of accuracy which is quantified by the MSE loss. The true equation of state is uniquely matched across the entire density range of the equation of state, including at the areas around 10^{14} which were problematic in the PCA cases. The extremities are also matched to the truth, including at it's unique maximum pressure and density, parameters which are carried alongside the 12 autoencoder latent space points.

The main advantage to the autoencoder approach is, like the PCA approach, its non-parametric and model-agnostic nature, which is consistent with later stages of the analysis. Autoencoding a training data set of a different parameterisation of the neutron star equation of state (not the unified training data set we have introduced above) would follow the exact same method, the only requirement being optimising the network training relative to the new data. We choose to train with a training data set of 10^5 different unified equations of state. However, alternatively, one could

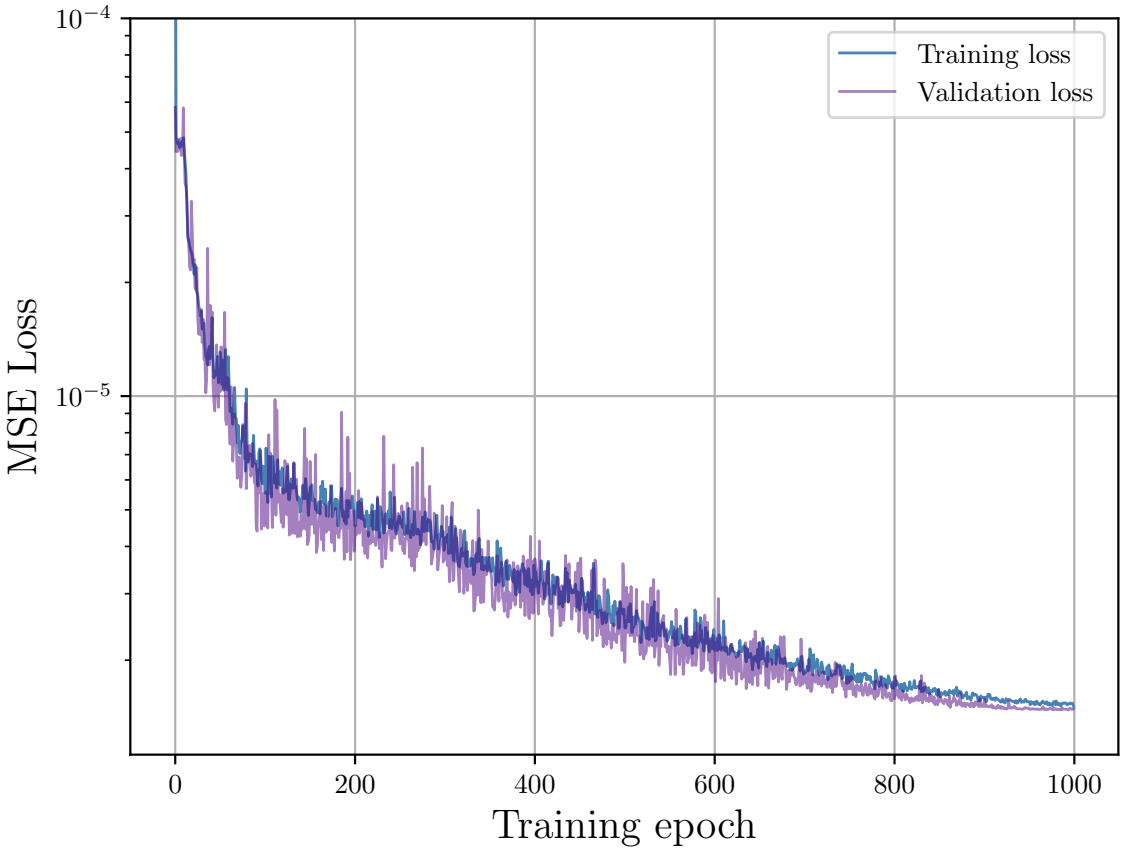


Figure 5.2.3: Training (blue) and validation (purple) MSE loss for the 12-dimensional autoencoder.

autoencode a training data set of a mix of neutron star equations of state from a variety of parameterisations (for example, piecewise polytropic *and* spectral at the high density) so to allow for a latent space which may reconstruct EOSs belonging to different or made up of a mix of parameterisations.

By performing inference in the 12D autoencoder latent space, one can then sample regions of the autoencoder latent space that do not belong explicitly to the training data set but are within the bounds of the training data, potentially with a unique morphology. As stated above, we carry P_{min}, P_{max}, and ρ_{max} as 3 additional parameters in inference. This means our inferred equations of state can be projected entirely from the autoencoder latent space to the original $P-\rho$ representation presented in figure 5.1.3 and we can infer the extremities of each equation of state as well as the morphology and behaviour over the density range. Therefore, our overall equation of state data space that we perform inference in consists of 15 dimensions and is

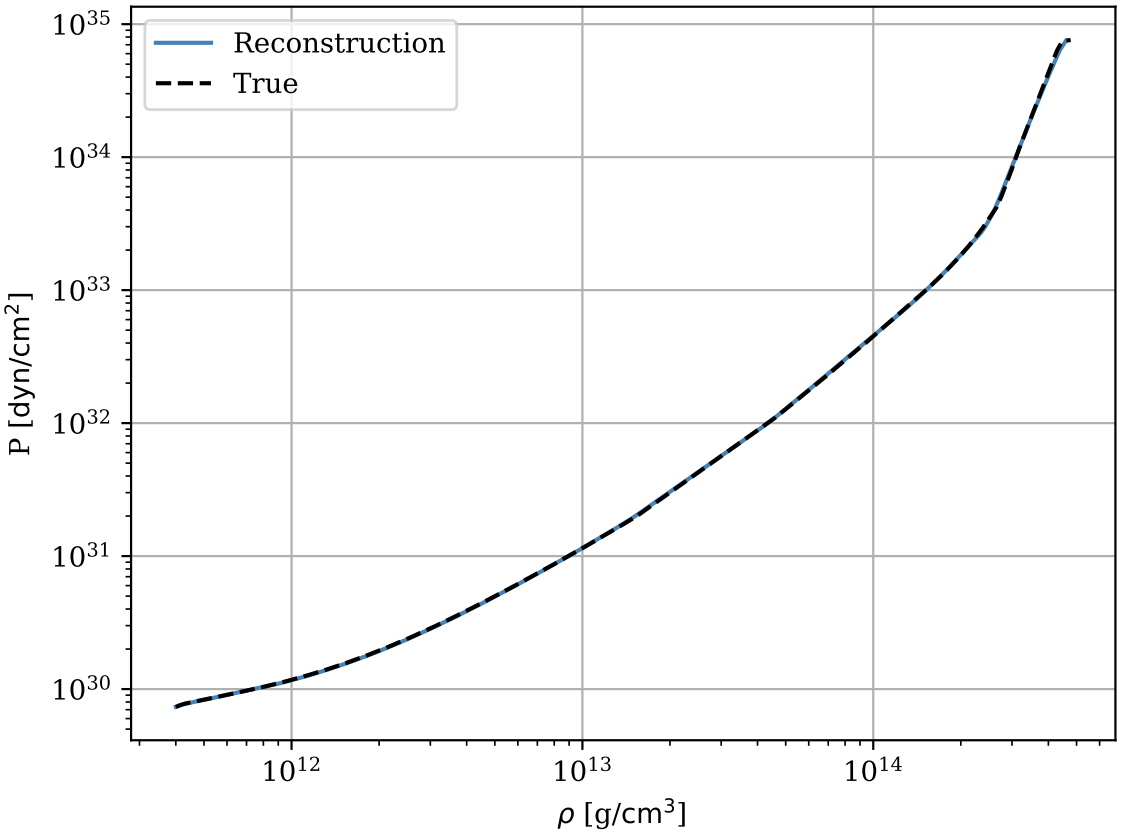


Figure 5.2.4: Plot of P vs. ρ of an example reconstruction of an equation of state from the validation data set using the 12−dimensional autoencoder.

presented in figure 5.2.5. Each dimension has been scaled relative to its own mean and standard deviation to have mean 0 and standard deviation 1. The first 12 dimensions from the left are the autoencoded latent space dimensions and the final 3 on the right are the 3 additional maxima and minima for P and ρ.

5.3 Hierarchical Bayesian inference

Given we expect the number of observed BNS events in the next generation of gravitational wave detection to increase dramatically, our goal is to develop an inference scheme which can combine the information of all N events together into one combined inference of the NS equation of state. Therefore, we develop a hierarchical Bayesian framework where we can adapt the Flow method from chapter 4 to instead allow us to calculate and combine the likelihood through Bayesian methods. We build on the introduction from chapter 2 section 2.2.2.1.

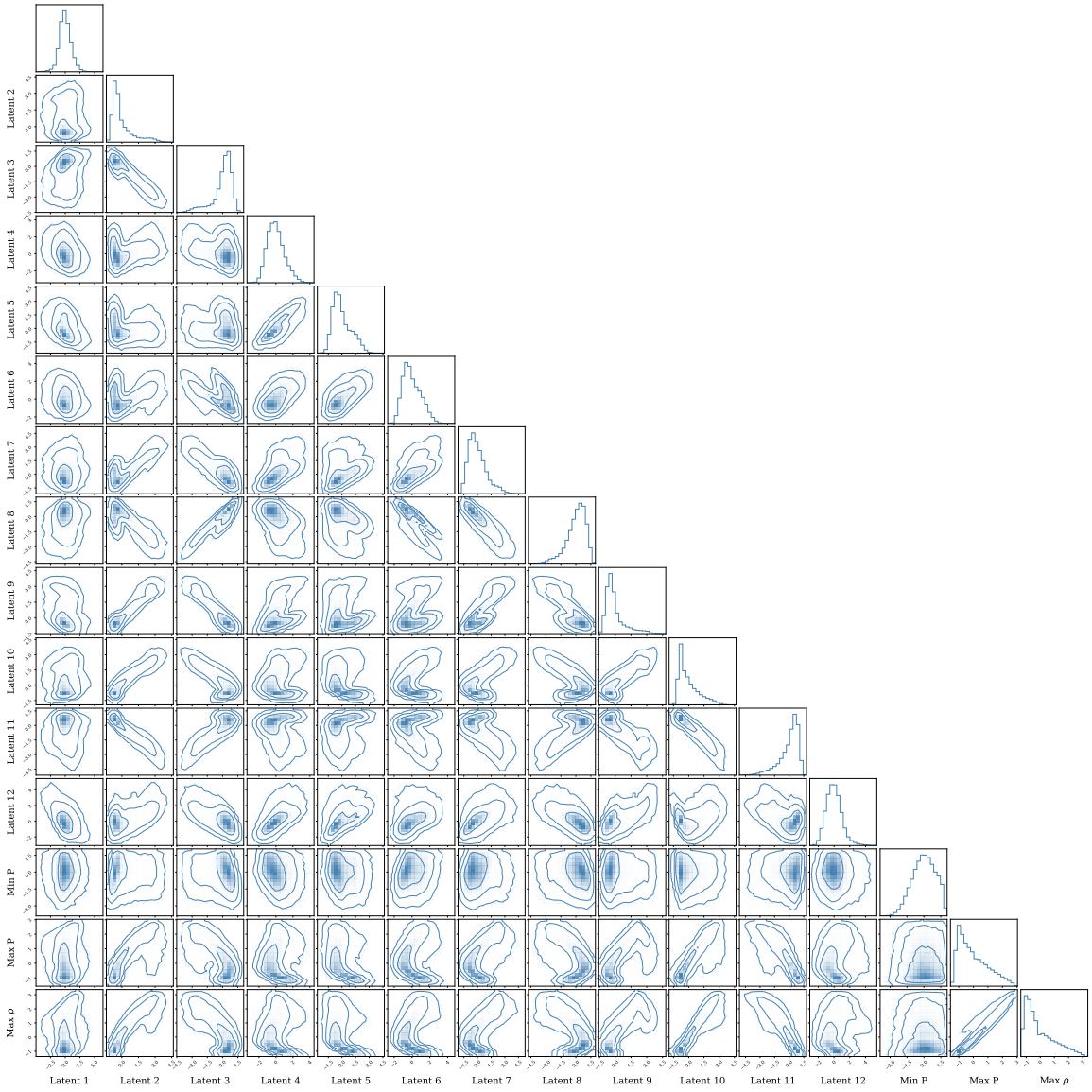


Figure 5.2.5: 15-dimensional compressed equation of state training data space consisting of 12 autoencoded latent dimensions and 3 additional scaling parameters. The contours represent 68%, 95%, and 99.7% of the samples respectively.

If Ω represents our 15D compressed equation of state data space, we wish to express equation 2.6 in terms of the observed parameters θ, which includes m_1, m_2, Λ_1 and Λ_2 (or alternative parameterisations of) among others. By marginalising over θ, we express equation 2.6 in terms of the parameters measured given event i:

$$p(\Omega|h) \propto p(\Omega) \prod_{i=1}^{N} \int p(h_i|\Omega,\theta_i)\ p(\theta_i|\Omega)\ d\theta_i. \tag{5.5}$$

which simplifies to

$$p(\Omega|h) \propto p(\Omega) \prod_{i=1}^{N} \int p(h_i|\theta_i)\ p(\theta_i|\Omega)\ d\theta_i. \tag{5.6}$$

where $p(\Omega)$ is the equation of state prior probability density. The strain data of event i does not depend on the EOS, only the measured parameters of event i, θ_i. We apply Bayes' theorem again such that

$$p(\Omega|h) \propto p(\Omega) \prod_{i=1}^{N} \int \frac{p(\theta_i|h_i)p(h_i)}{p(\theta_i)}\ \frac{p(\Omega|\theta_i)p(\theta_i)}{p(\Omega)}\ d\theta_i. \tag{5.7}$$

The prior used to compute the posterior samples, $p(\theta_i)$, is exactly the same prior used to train the Flow, and therefore the above expression simplifies to

$$p(\Omega|h) \propto \frac{p(\Omega)}{p(\Omega)^N} \prod_{i=1}^{N} \int p(\theta_i|h_i)\ p(\Omega|\theta_i)\ d\theta_i. \tag{5.8}$$

By taking the Monte-Carlo sum over the posterior samples of event i, we find that

$$p(\Omega|h) \propto p(\Omega)^{(1-N)} \prod_{i=1}^{N} \Big\langle p(\Omega|\theta_i) \Big\rangle_{p(\theta_i \sim p(\theta_i|h_i))}. \tag{5.9}$$

Therefore, our overall posterior probability density on the equation of state space given information from N gravitational wave events can be found by taking the Monte-Carlo average over posterior samples for each of the N events and then taking the product over N events and including the equation of state prior $(1-N)$ times. By applying Bayes' theorem to the likelihood $p(h_i|\theta_i)$ of equation 5.6 to return the posterior $p(\theta_i|h_i)$, we can then use samples θ_i from the posterior of event i.

The goal is to adapt the method of chapter 4 to train a Flow to map the 15D equation of state data space, Ω conditioned on event parameters $\theta = [m_1, m_2, \Lambda_1, \Lambda_2]$ via equation 3.5 for a conditional Flow given in chapter 3 section 3.2:

$$p(\Omega|\theta) = p(f(\Omega|\theta)|\theta) \left|\det\left(\frac{\partial f(\Omega|\theta)}{\partial \Omega^T}\right)\right|. \tag{5.10}$$

Once trained, we wish to use the Flow to return the probability density of $\mathbf{\Omega}$ given information from the i^{th} event's parameters, θ_i, $p(\mathbf{\Omega}|\theta_i)$. This requires us to be able to propose equations of state, $\mathbf{\Omega}$, in the Flow's data distribution, evaluate the likelihood given θ_i and repeat. This process is usually automated by sampling algorithms like MCMC (introduced in chapter 1 section 1.5.1.1) that, when given access to a likelihood and a prior function, can efficiently traverse the multi-dimensional data distribution and return a posterior probability given some information. The simpler the likelihood surface is, the easier these algorithms find the process and the more computationally efficient they are. We next discuss the complexity of the 15-dimensional data space for this problem, and in which way the presence potential artificial structure apparent through extensive testing.

5.3.1 Training the Normalising Flow

Following the method employed in chapter 4 section 4.4, the Normalising Flow was trained to perform the forward mapping from compressed equation of state data space $\mathbf{\Omega}$ to latent space $\mathcal{Z}$, conditioned on properties of mergers of neutron stars associated to each of the equations of state, θ. Training data sets for both the data space $\mathbf{\Omega}$ and conditional space θ were generated following the methods employed in chapter 4 section 4.2. To recap, this involves selecting an equation of state $\mathbf{\Omega}_i$ from the training data set and selecting its relevant TOV information - both the equation of state in pressure P_i on the fixed grid of density ρ and the associated relationships between macroscopic parameters (mass, m, and tidal deformability, Λ, amongst others) are standard products of the CUTER [2] analysis. By finding the maximum allowed mass for the selected equation of state, $m_{\text{max},i}$, two component masses are uniformly sampled $m_{i,1}, m_{i,2} \sim U[1, m_{\text{max},i}]$ $\text{M}_\odot$. These two masses, $m_{i,1}$ and $m_{i,2}$, represent two neutron stars which belong to the i^{th} equation of state and are part of a binary merger system. Through interpolation of the component masses, one can obtain allowed central densities ρ_{c1}, ρ_{c2} of the neutron stars and, subsequently, tidal deformabilities, Λ_1, Λ_2. The allowed component masses and dimensionless tidal deformability parameters for two neutron stars are collected into $\theta_{\text{i}} = [m_1, m_2, \Lambda_1, \Lambda_2]$ and used as conditional information in training the Normalising Flow.

	Hyperparameter range
Number of neurons	32
Number of transforms	2
Number of residual blocks	1
Number of epochs	1000
Learning rate	0.005
Batch size	4096

Table 5.3.1: Table of hyperparameters and settings for training the conditional Flow.

A single binary neutron star merger system with parameters θ_i is generated for each of the 10^5 equations of state $\mathbf{\Omega}_i$ and collectively this constitutes the training data used to train the Normalising Flow. The Flow is a Coupling Neural Spline Flow [234] trained with hyperparameters and settings given in table 5.3.1. Training takes ~ 30 minutes on an NVIDIA GeForce RTX 3090 GPU. We choose to keep the Flow lightweight with small numbers of transforms and neurons per transform so to reduce the computational cost of training the model and to have a more lightweight model when it comes to repeatedly calling the trained model in inference. A lightweight model with small numbers of hyperparameters also aims to achieve a more generic understanding of the mapping from data space $\mathbf{\Omega}$ to latent space $\mathcal{Z}$. Further discussion in this chapter will make the reasons behind this decision more apparent.

5.4 Evidence of learned artificial structure

In chapter 4, we used the trained Normalising Flow to conditionally sample the Flow latent space given posterior samples from single BNS merger events. This returns a posterior probability distribution in the equation of state data space, $\mathbf{\Omega}$. We performed this analysis using PE samples from the real gravitational wave event GW170817 to produce our main result (see chapter 4 figure 4.5.1). We use the same GW event posterior samples for testing how the Flow evaluates the probability of proposed equations of state for consistency and to directly compare the quality of sampling vs. proposing equations of state. As introduced in chapter 3 section 3.2, a trained Flow can be used for two main functions:

- *Sampling* the data distribution: this involves conditionally sampling the data space to produce samples $\Omega \sim p(\Omega|\theta_{\text{event}})$ through randomly sampling the Gaussian latent space $\mathcal{Z}$ and performing the conditional inverse mapping to return samples from the Flow posterior $p(\Omega|\theta)$.

- *Evaluating* the probability of defined points in the data distribution: points Ω_{prop} in the data distribution are proposed and their conditional probability $p(\Omega_{\text{prop}}|\theta_{\text{event}})$ evaluated by performing the forward mapping to the latent space $\mathcal{Z}$ and evaluating the relative probability in the simple Gaussian latent space.

As stated above, our method in chapter 4, circulated around using the Flow to sample the learned posterior $\Omega \sim p(\Omega|\theta_{\text{event}})$. Now that we wish to evaluate the probability of equations of state given information from multiple events via the methods discussed in 5.3, we want to *propose* regions of the equation of state space, Ω_{prop} and then *evaluate* the probability of the proposed points, $p(\Omega_{\text{prop}}|\theta_{\text{event}})$. This is what sampling methods like MCMC or Nested Sampling will do; survey the space via their own respective algorithm and evaluate the probability of selected regions and then make decisions on where to survey next based on what they find. We test the Flow's ability to evaluate the probability of proposed equations of state. We summarise each stage of testing at the end of their respective sections to make the conclusions clear.

5.4.1 Proposing equations of state

We test the trained Normalising Flow in evaluating the probability of points Ω_{prop} in the equation of state data space $p(\Omega_{\text{prop}}|\theta)$, given the same conditions used in sampling $\Omega_{\text{samp}} \sim p(\Omega|\theta)$. We propose equations of state in the 15-dimensional compressed space from a Gaussian distribution, Ω_{Gauss}, centred at $\mu = 0$ with standard deviation $\sigma = 0.5$. This distribution is proposed here such that it aligns with the scaled 15-dimensional data distribution that the Flow has been trained to map. The proposed points Ω_{Gauss} are presented in figure 5.4.1 in pink alongside the sampled distribution of points, $\Omega_{\text{samp}} \sim p(\Omega|\theta_{170817})$ given when sampling with PE samples from GW170817, θ_{170817}, in purple. The training data prior distribution is given in blue. The distribution of proposed points Ω_{Gauss} overlaps with the sampled distribution

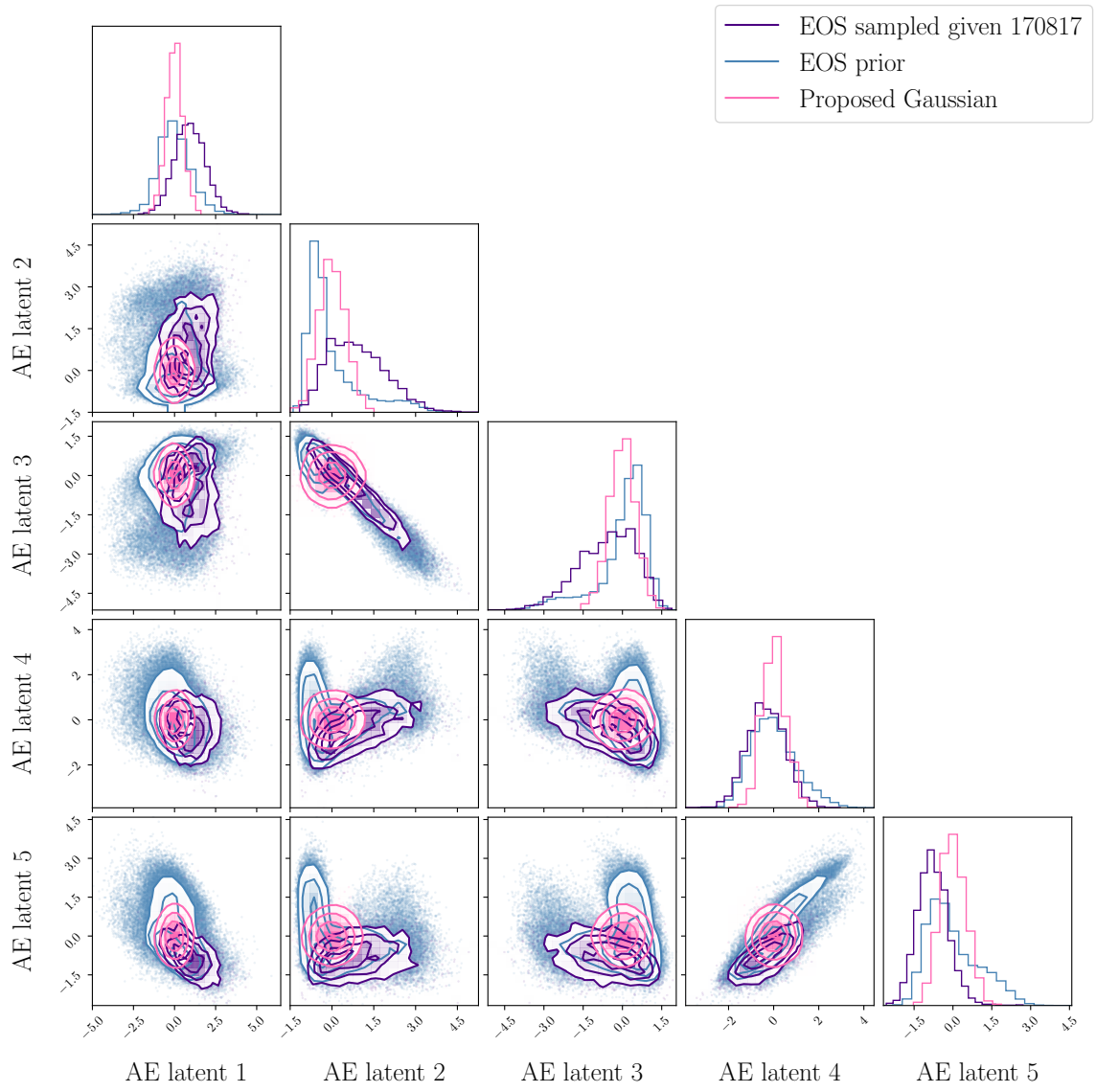


Figure 5.4.1: Corner plot of 1- and 2-dimensional marginalised distributions of training equations of state (blue), sampled points from the conditional Normalising Flow (purple) and proposed equations of state (pink) from a Gaussian distribution with $\sigma = 0.5$. We present 5 out of the total 15 dimensions. The full figure can be found in the appendix section 7.1 figure A1.

$\mathbf{\Omega}_{\text{samp.}}$ in the vast majority, with some extension of $\mathbf{\Omega}_{\text{Gauss}}$ beyond the prior bounds in some dimensions due to highly correlated 2-dimensional projections. However, we expect the probability of $p(\mathbf{\Omega}_{\text{Gauss}}|\theta_{170817})$ to be for the most part comparable to the probability of $p(\mathbf{\Omega}_{\text{samp.}}|\theta_{170817})$, allowing for some outliers.

However, we find that the proposed distribution $p(\mathbf{\Omega}_{\text{Gauss}}|\theta_{170817})$ has very low log(probability) compared to the points which are sampled. Figure 5.4.2 presents the

log(probability) distribution of the sampled equations of state from the trained Flow given conditions from GW170817, $p(\Omega_{\text{samp.}}|\theta_{170817})$, on the right and the log(probability) distribution of proposed Gaussian equations of state given the same condition, $p(\Omega_{\text{Gauss}}|\theta_{170817})$ on the left. Sampling from the Flow given conditional information produces points that lie within the prior bounds of the 15D space, but when $\log p(\Omega_{\text{Gauss}}|\theta_{170817})$ is evaluated with the Flow given the same condition, the log(probability) of these points are significantly different. We compute the difference

$$\Delta_{\log \text{ prob.}} = \left|\max(\log p(\Omega_{\text{samp.}}|\theta_{170817})) - \max(\log p(\Omega_{\text{prop}}|\theta_{170817}))\right| \tag{5.11}$$

which, for $\Omega_{\text{prop}} = \Omega_{\text{Gauss}}$ with $\sigma = 0.5$ gives $\Delta_{\log \text{ prob.}} \simeq 1052$. The most probable equation of state from the Gaussian distribution has log(probability) 1052.13 smaller than that of the maximum probability sampled point, or the most probable samples from Ω_{Gauss} is exp(1052) less likely than the most probable sample from $\Omega_{\text{samp.}}$. These log(probabilities) suggest that Ω_{Gauss} are all situated outwith the training prior distribution, while it does not appear to be entirely beyond the bounds of either the training prior or the sampled distributions $\Omega_{\text{samp.}}$ in figure 5.4.1. Points proposed from such a distribution as this Gaussian mimics points that would likely be proposed in early stages of a sampling algorithm, suggesting that an algorithm would struggle to explore this space, as many, if not all, of the proposed regions early on in sampling would be disfavoured.

To ensure that points belong within the bounds of the prior training distribution, we instead propose equations of state around the maximum probability point from the sampled equation of state distribution, $\Omega_{\text{max sample}}$, where $p(\Omega_{\text{max sample}}|\theta_{170817}) = \max(p(\Omega_{\text{samp.}}|\theta_{170817}))$. Figure 5.4.3 presents again the prior training distribution and $\Omega_{\text{samp}} \sim p(\Omega|\theta_{170817})$, but now alongside a new proposed Gaussian distribution $\Omega_{\text{Gauss at max}}$ centred at mean $\mu = \Omega_{\text{max sample}}$ and with standard deviation $\sigma = 0.1$. This proposed distribution is much smaller in volume and is centred on a region of the parameter space which the Flow has deemed to be very probable given the information from the event GW170817. We evaluate the probability $p(\Omega_{\text{Gauss at max}}|\theta_{170817})$.

Figure 5.4.4 presents a histogram of $\log p(\Omega_{\text{Gauss at max}}|\theta_{170817})$ alongside the prob-

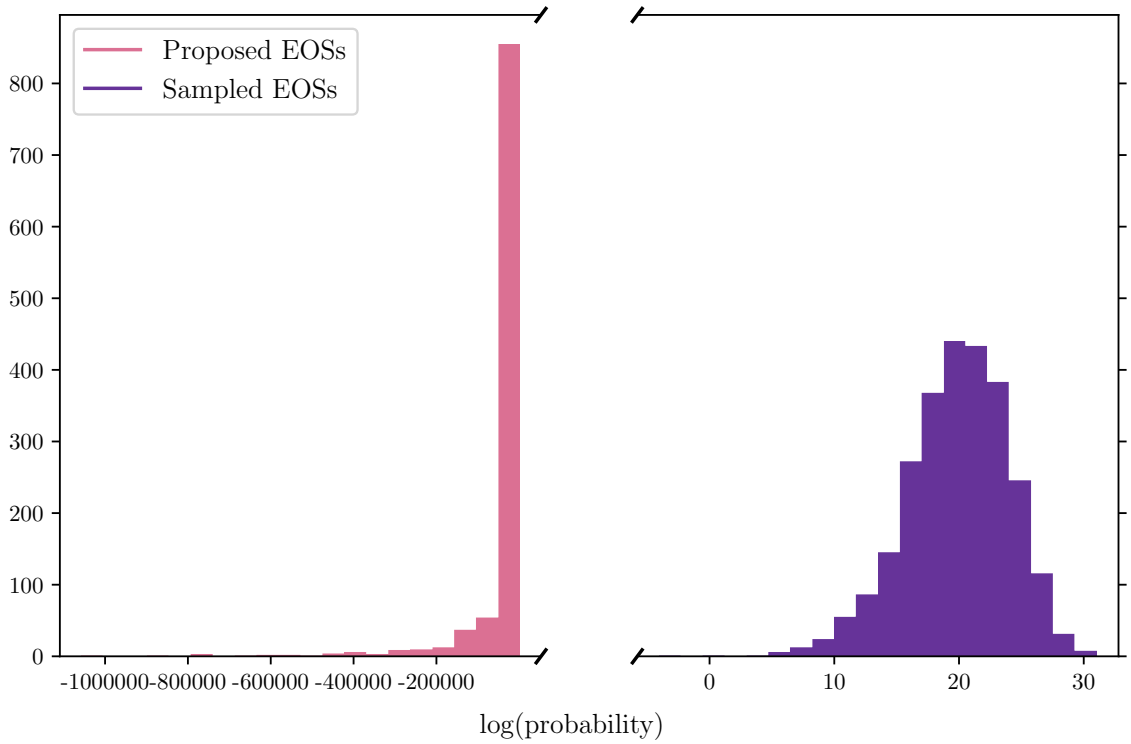


Figure 5.4.2: Histograms in log(probability) of sampled equations of state given GW170817 (purple) and Gaussian proposed equations of state given the same condition (pink).

abilities $p(\Omega_{\text{samp.}}|\theta_{170817})$. Again, we find the log(probabilities) of proposed equations of state evaluated relative to the same conditional information are significantly lower and less probable than $\log p(\Omega_{\text{samp.}}|\theta_{170817})$. The proposed equations of state around a maximum probability region in a small volume are around 4 orders of magnitude smaller in log(probability) than that of the local sampled point. For the small Gaussian proposed around $\Omega_{\text{Gauss at max}}$, we find $\Delta_{\log \text{ prob.}}$ to be ~ 440, such that $\max(\log p(\Omega_{\text{Gauss at max}}|\theta_{170817}))$ is $\exp(440)$ less likely than $\log p(\Omega_{\text{max samp.}}|\theta_{170817})$. This would normally indicate that $\Omega_{\text{Gauss at max}}$ is proposing samples in a very unlikely region of parameter space, which it is not relative to how the Flow has sampled the Ω space. In a 15-dimensional parameter space, which is intuitively difficult to visualise, the distribution of points within each dimension is often not as one would imagine, but for a small volume of proposed samples around a high-probability region, we would expect some points to be of comparable probability, and definitely within the bounds of the training prior distribution.

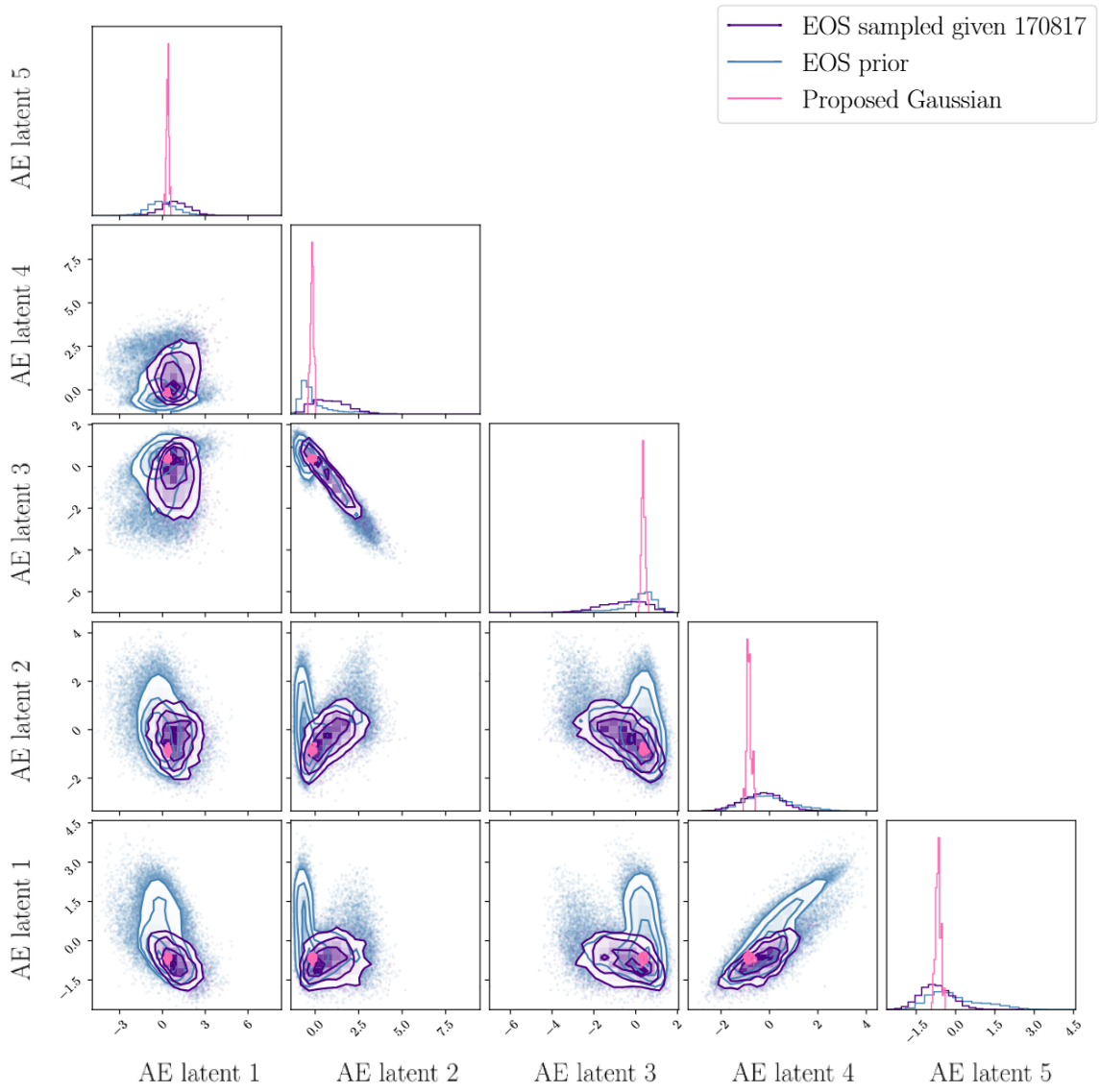


Figure 5.4.3: Corner plot of 1- and 2-dimensional marginalised distributions of training equations of state (blue), sampled points from the conditional Normalising Flow (purple) and proposed equations of state (pink) from a Gaussian distribution with $\sigma = 0.1$. We present the first 5 out of the total 15 dimensions. The full figure can be found in the appendix section 7.1 figure A2.

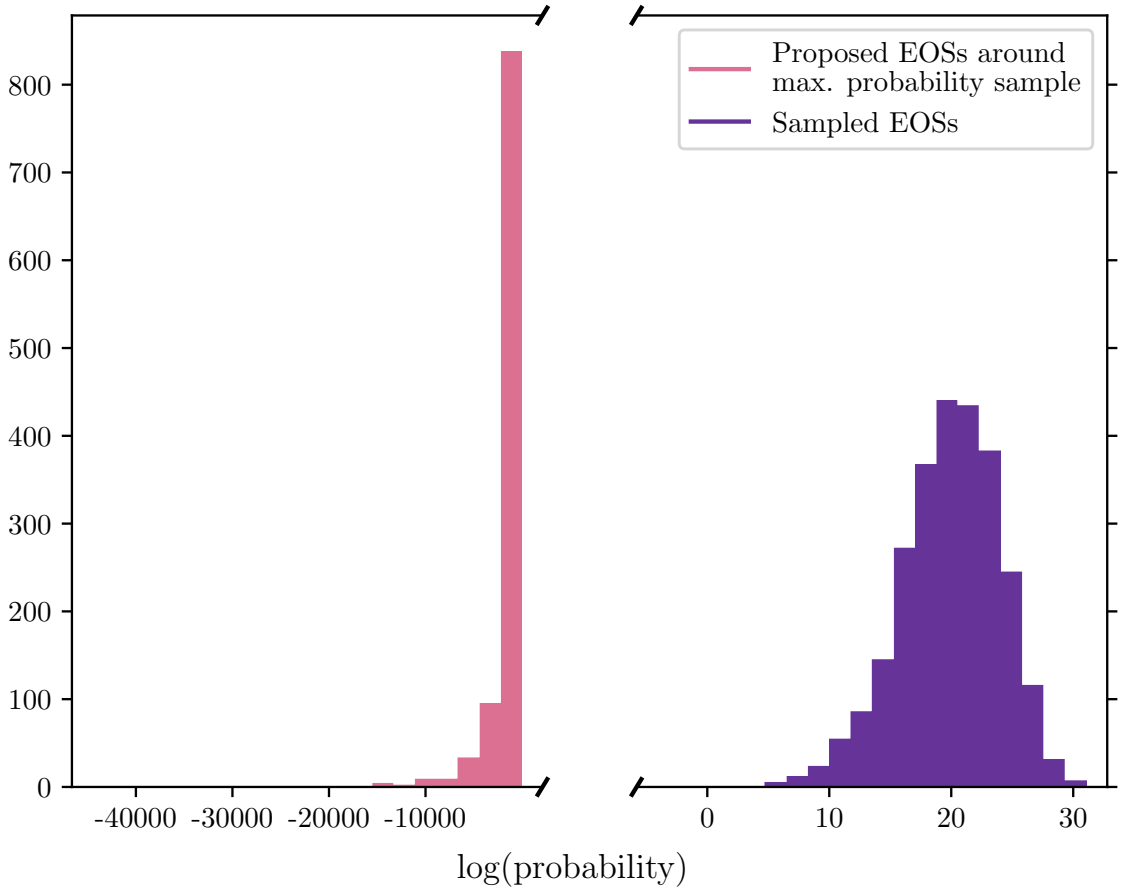


Figure 5.4.4: Histograms in log(probability) of sampled equations of state given GW170817 (purple) and Gaussian proposed equations of state around $\mathbf{\Omega}_{\text{max sample}}$ given the same condition (pink).

Summary of section: proposed equations of state from both a Gaussian distribution which spans the prior volume and one centred around the maximum likelihood sample spanning a fraction of the prior volume have log(probabilities) orders of magnitude less than that of the equations of state which are sampled by the Normalising Flow. The Normalising Flow appears to evaluate points within the prior space as having very different probabilities depending on if they are *proposed to* or *sampled by* the Flow.

5.4.2 Assessing training and validation data performance

We further test the Flow's ability to generalise to non-sampled points by assessing performance of evaluating the conditional probability of both training and validation data points, $p(\mathbf{\Omega}_{\text{train}}|\theta)$ and $p(\mathbf{\Omega}_{\text{val.}}|\theta)$. We plot the training and validation losses

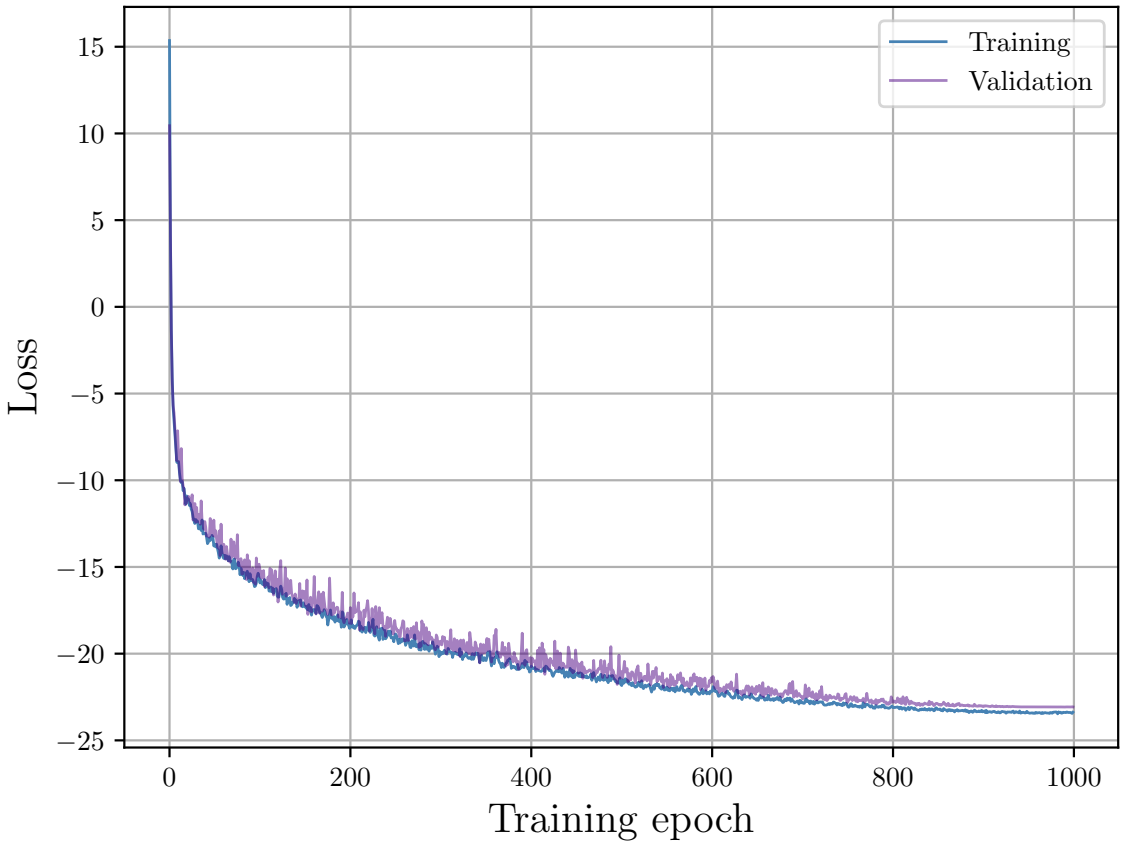


Figure 5.4.5: Training (blue) and validation (purple) loss for the conditional Flow trained as per equation 5.10 with unified equation of state data.

from the training of the conditional Flow in figure 5.4.5; the training and validation data sets are selected from the same data distribution of the entire 97 000 equation of state data set, where 25% are set aside for validation. The training and validation data sets are independent of one another, with no duplicated samples, but stem from the same underlying distribution. Their respective data sets are plotted alongside one another in the appendix section 7.1 figure A3. Throughout training the validation loss steadily decreases alongside the training loss showing no evidence of traditional over-fitting. Therefore, it appears the Flow evaluates well both the training and validation data sets throughout training.

Evidence of over-fitting would normally appear as an increase in validation loss over the course of training as training loss continues to decrease. This signifies that the Flow is over-specifying to the training data points themselves and cannot generalise well enough to perform well on the validation data also, given the current scenario

we have in which they stem from the same distribution. This can happen with too little training data where there is so little that the Flow can instead memorise the individual data points or trends of the training data and therefore performs badly when testing on validation data, unseen during training. We plot the training and validation data together in figure 5.4.6, where both data sets clearly occupy the same distribution in all 15 dimensions, which is in agreement with the loss plot of figure 5.4.5, as the training and validation performance throughout the training routine is consistent with one another.

We likewise plot the training and validation data in the conditional space, θ_{train} and θ_{val} consisting of parameters $[m_1, m_2, \Lambda_1, \Lambda_2]$, in figure 4.4.1. The gravitational wave event PE samples for GW170817, θ_{170817}, are also plotted in pink contours and have been passed through the convex hull so to be within the bounds of the training data prior (see chapter 4 section 4.4.1.1 for details). In section 5.4.1, we found vastly different log(probabilities) between equations of state that had been proposed and sampled, but each relative to the same conditional term, $p(\mathbf{\Omega}_{\text{prop}}|\theta_{170817})$ and $p(\mathbf{\Omega}_{\text{samp.}}|\theta_{170817})$ respectively. Therefore, we wish to assess $\mathbf{\Omega}_{\text{prop}}$ from other distributions of data, namely training and validation equations of state, given θ_{train}, $\theta_{\text{val.}}$, and θ_{170817}.

We use the Flow to evaluate the probability of training equations of state given training $p(\mathbf{\Omega}_{\text{train}}|\theta_{\text{train}})$, validation $p(\mathbf{\Omega}_{\text{train}}|\theta_{\text{val}})$ and event $p(\mathbf{\Omega}_{\text{train}}|\theta_{170817})$ conditions and present the results in figure 5.4.7. We vary the conditional term to test the potential of over-fitting conditionally, i.e. if the Flow can evaluate $\mathbf{\Omega}_{\text{train}}$ with training data conditions θ_{train} but struggles with conditions outside of the training data set, including θ_{170817}. We find the distributions of $p(\mathbf{\Omega}_{\text{train}}|\theta_{\text{train}})$, $p(\mathbf{\Omega}_{\text{train}}|\theta_{\text{val}})$ and $p(\mathbf{\Omega}_{\text{train}}|\theta_{170817})$ are comparable suggesting all are similarly valid given a set of conditions. This demonstrates we are not over-fitting with respect to the conditional term.

If over-fitting is instead occurring in the equation of state data space, then evaluating the probability of proposed validation data should achieve low log(probabilities)

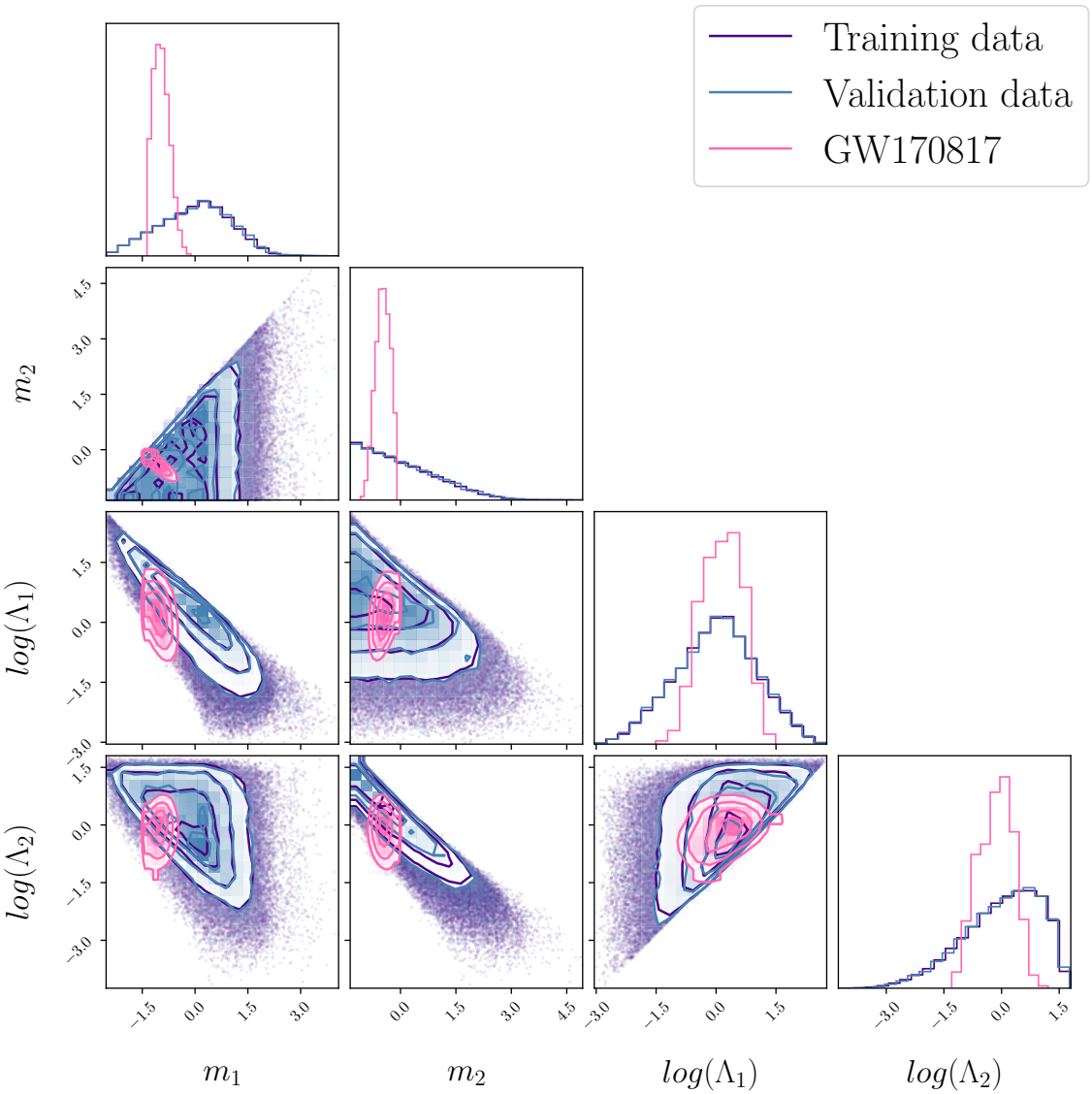


Figure 5.4.6: Corner plot of 1- and 2-dimensional marginalised distributions of training (purple) and validation (blue) data in $[m_1, m_2, \log(\Lambda_1), \log(\Lambda_2)]$ alongside the PE samples from gravitational wave event GW170817 (pink). Contour plot of The contours represent 68%, 95%, and 99.7% of the samples respectively.

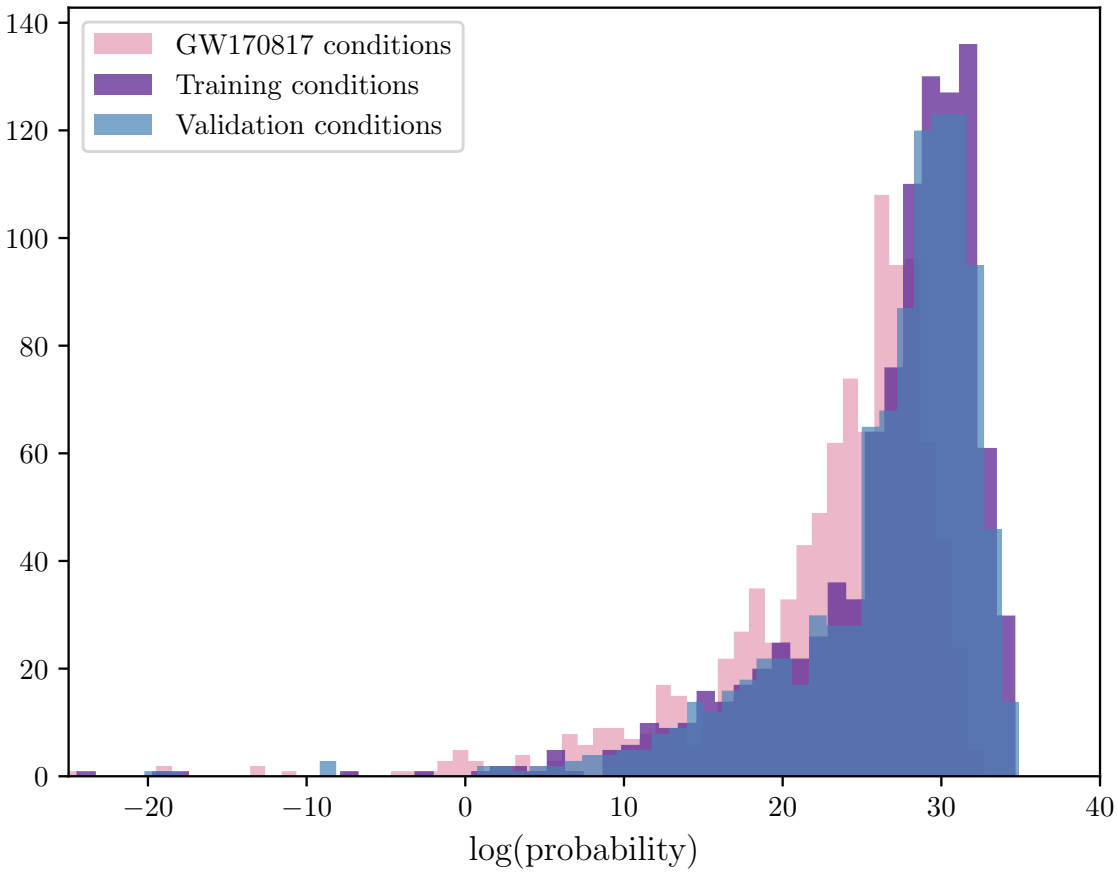


Figure 5.4.7: Histograms of log(probabilities) of proposed validation equations of state evaluated relative to conditions from training, validation and event data sets.

by the Flow, even though it appears the loss plot in figure 5.4.5 that this is not the case. Figure 5.4.8 presents the result of evaluating validation equations of state relative to training $p(\mathbf{\Omega}_{\rm val}|\theta_{\rm train})$, validation $p(\mathbf{\Omega}_{\rm val}|\theta_{\rm val})$ and event $p(\mathbf{\Omega}_{\rm val}|\theta_{170817})$ conditional samples. We find no evidence of outlier log(probabilities) when evaluating non-training data equations of state with a variety of conditions, which aligns with the loss plot presented in figure 5.4.5. We therefore conclude that the training and validation data belong to the same data distribution which the Flow has learned well and instead the lack of generalisability is in moving away from this distribution.

To test the ability of the Flow to generalise to data beyond that of the training and validation data distributions, we add noise on a small scale to the validation data and perform the same test of evaluating the probability of these proposed equations of state given a variety of conditions. Gaussian noise is added to the validation data

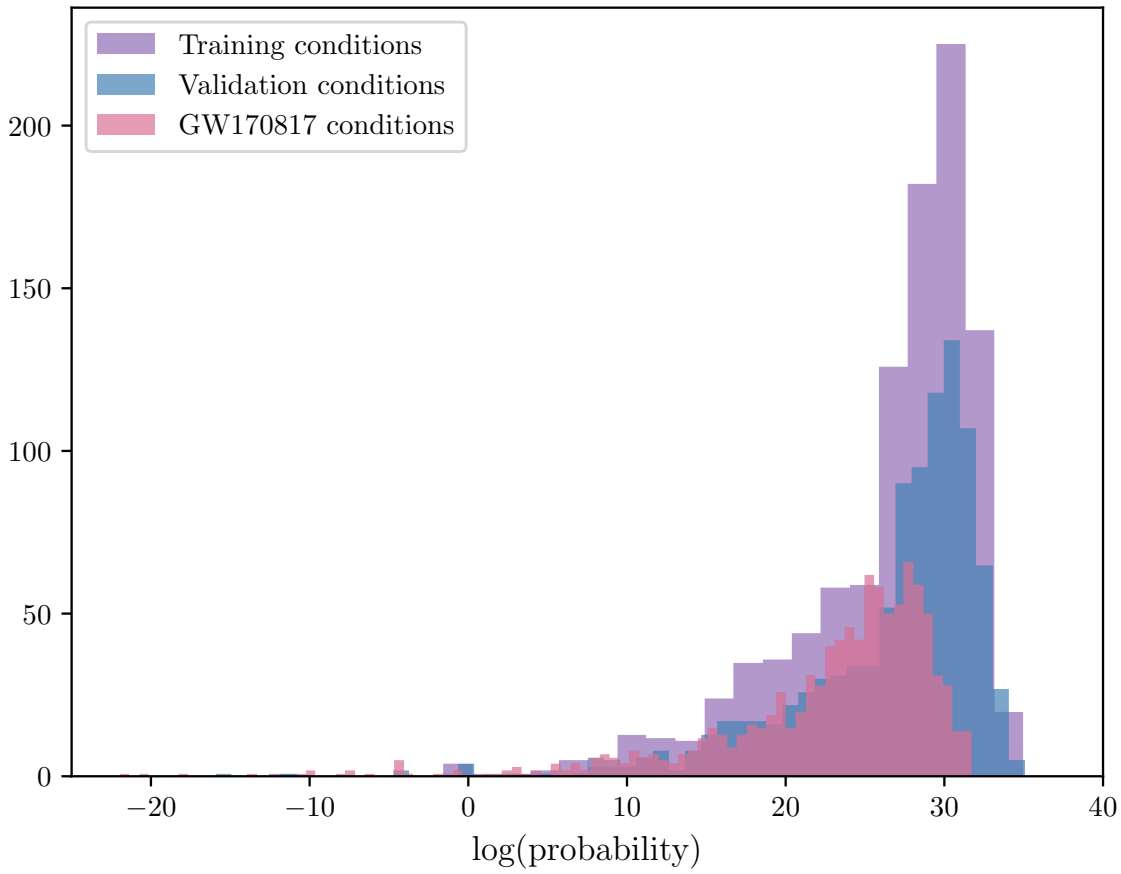


Figure 5.4.8: Histograms of log(probabilities) of proposed validation equations of state evaluated relative to conditions from training, validation and event data sets.

on a scale of 1% of the standard deviation of the data set (scaled to have standard deviation $\sigma = 1$). The original validation data $\mathbf{\Omega}_{\text{val}}$ is plotted alongside the noise-added validation data $\mathbf{\Omega}_{\text{noisy val}}$ in the appendix, section 7.1 figure A4. It is clear that $\mathbf{\Omega}_{\text{noisy val}}$ differs on a very small scale to $\mathbf{\Omega}_{\text{val}}$, so much so that when plotted alongside each other they are almost irresolvable.

Figure 5.4.9 presents the results of evaluating the probability of noisy validation equations of state given training $p(\mathbf{\Omega}_{\text{noisy val}}|\theta_{\text{train}})$, validation $p(\mathbf{\Omega}_{\text{noisy val}}|\theta_{\text{val}})$ and event $p(\mathbf{\Omega}_{\text{noisy val}}|\theta_{170817})$ conditions. The noise-added validation equations of state are far less probable than that of the original validation data set, even though the data sets differ from one another by a very small percentage of added noise. Regardless of the conditional term, a significant fraction of the noise-added equations of state are around exp(100) less probable than the original validation equations of state. It

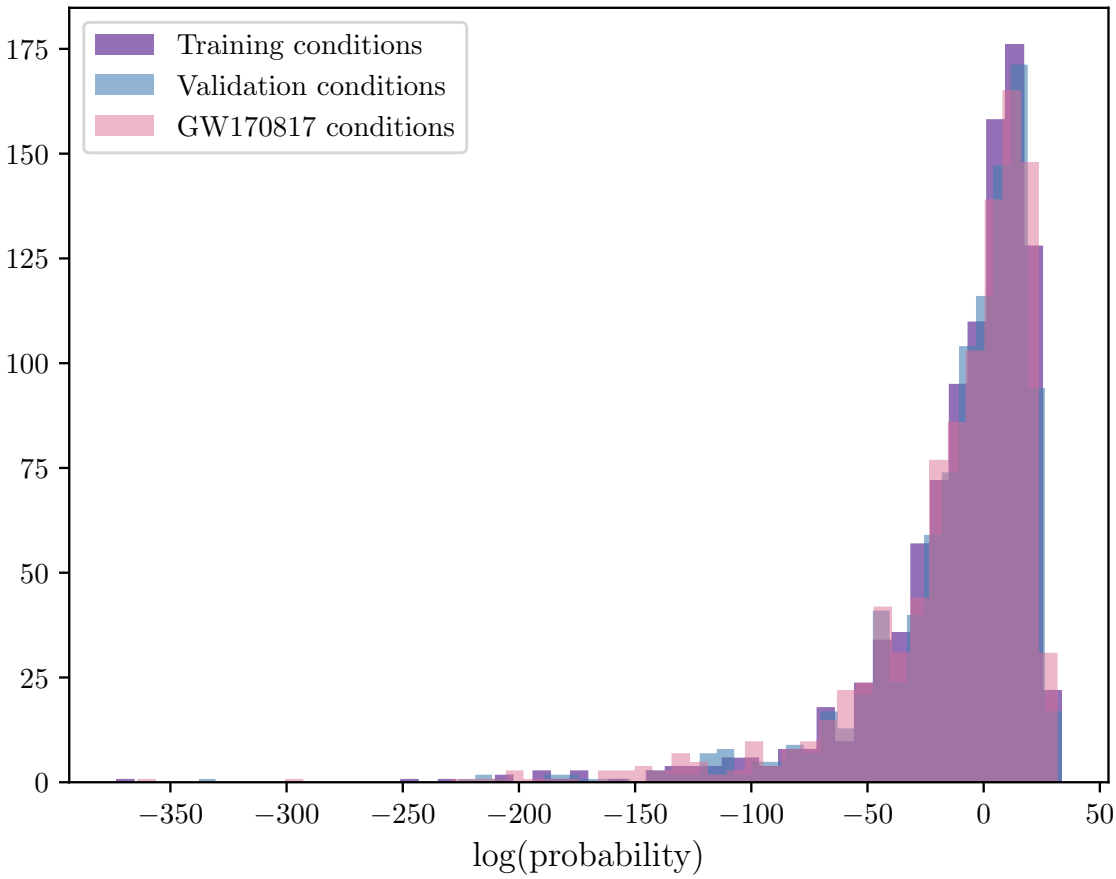


Figure 5.4.9: Histograms of log(probabilities) of proposed noise-added validation equations of state evaluated relative to conditions from training, validation and event data sets.

therefore appears that the validation data belongs to the same structure or overall distribution as the training data that the Flow finds easy to learn. Adding a small magnitude of noise to the distribution moves the data beyond the structure which the Flow has learned and the Flow struggles as a result.

To further demonstrate the alignment of validation data on the grid that the training data belongs to, we add the same 1% Gaussian noise to the validation data during training. Figure 5.4.10 presents the first 100 epochs of training a Flow with added Gaussian noise on the scale of 1% noise of the standard deviation of the validation data space. This Flow is identical to the original Flow trained for this analysis, the only difference in the training regime being the addition of noise to the validation data. We see clear and extreme over-fitting, where the validation loss increases as the training loss continues to decrease. This suggests the validation

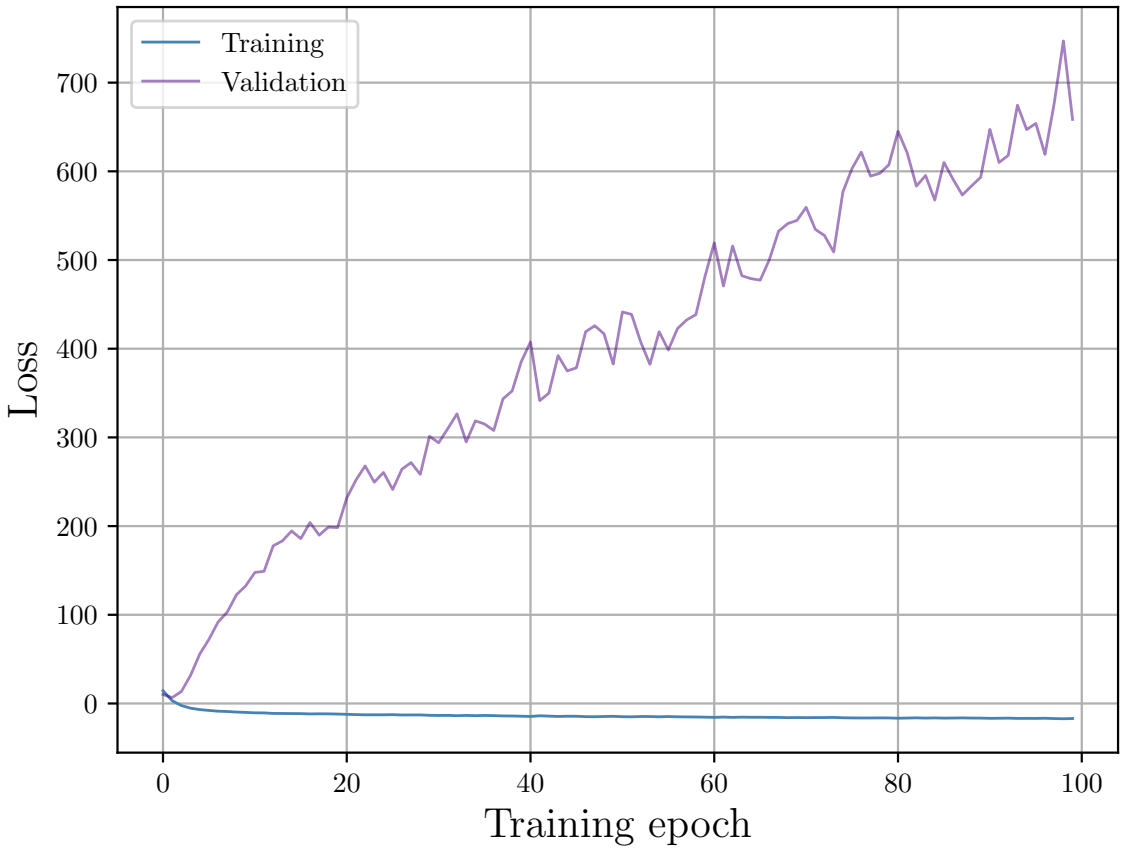


Figure 5.4.10: Training (blue) and validation (purple) loss plotted as a function of training epoch for training the conditional Normalising Flow with noise on validation data.

data with such small scale noise is starkly different to the training data distribution, and strongly suggests that the training and validation data sets belong to a shared underlying distribution which the Flow can learn. When conditionally sampling $\mathbf{\Omega}_{\text{samp}} \sim p(\mathbf{\Omega}|\theta)$, the Flow can then sample from the structure it has learned which is associated to high log(probabilities). Evaluating the probability of regions of the 15-dimensional data space with randomly proposed equations of state $\mathbf{\Omega}_{\text{prop}}$ given the same sets of conditions is difficult, as we are not likely to propose equations of state from the underlying structure, which we are as of yet blind to.

Summary of findings: The trained conditional Normalising Flow evaluates the log(probability) of training and validation data points as comparably probable (regardless of the conditional term) even though the validation data is not seen by the Flow during training. By adding a small magnitude of noise to the validation data,

we find the Flow determines the noise-added validation points to be far less probable, even though they are visually irresolvable. By training with noise-added validation data, we find severe overfitting. We propose the potential of artificial structure which the Flow has learned in the equation of state training data which the validation data likewise belongs to.

5.4.3 Training an unconditional Normalising Flow

We have thus far demonstrated the ability of a conditional Flow to over-fit to an underlying data structure, regardless of the conditional term. We suggested that changing the conditional term, θ, did not affect the overall magnitude of log(probability) in section 5.4.2, but to fully test this, we remove the conditional term entirely to remove any dependency of it from the Flow. We train an unconditional Normalising Flow with the same architecture as the original conditional Flow (hyperparameters given in table 5.3.1) to map the 15-dimensional equation of state prior prior training data space, $\mathbf{\Omega}$, to a 15-dimensional unit Gaussian via

$$p(\mathbf{\Omega}) = p(f(\mathbf{\Omega})) \left|\det\left(\frac{\partial f(\mathbf{\Omega})}{\partial \mathbf{\Omega}}\right)\right|, \tag{5.12}$$

noting the absence of the conditional term, θ. Once trained, we can sample from the Normalising Flow $\mathbf{\Omega}_{\text{samp}} \sim p(\mathbf{\Omega})$ to return samples from the learned prior distribution. Likewise with the trained conditional Flow, we can also evaluate the prior probability of proposed equations of state $p(\mathbf{\Omega}_{\text{prop}})$.

Figure 5.4.11 presents the 15-dimensional training data distribution $\mathbf{\Omega}_{\text{train}}$ alongside 1000 randomly sampled points, $\mathbf{\Omega}_{\text{samp.}}$ from the learned unconditional Flow. We propose a Gaussian distribution of equations of state $\mathbf{\Omega}_{\text{prop}}$ centred at $\mu = 0$ and with a standard deviation of $\sigma = 0.5$. We see the unconditional Flow has learned the data distribution well and that the sampled points $\mathbf{\Omega}_{\text{samp.}}$ align well with the training data $\mathbf{\Omega}_{\text{train}}$.

Figure 5.4.12 plots the relative log probabilities of the training data equations of state $\mathbf{\Omega}_{\text{train}}$, those randomly sampled from the trained Flow $\mathbf{\Omega}_{\text{samp.}}$ and the Gaussian proposed equations of state $\mathbf{\Omega}_{\text{prop}}$. Both the training equations of state and those

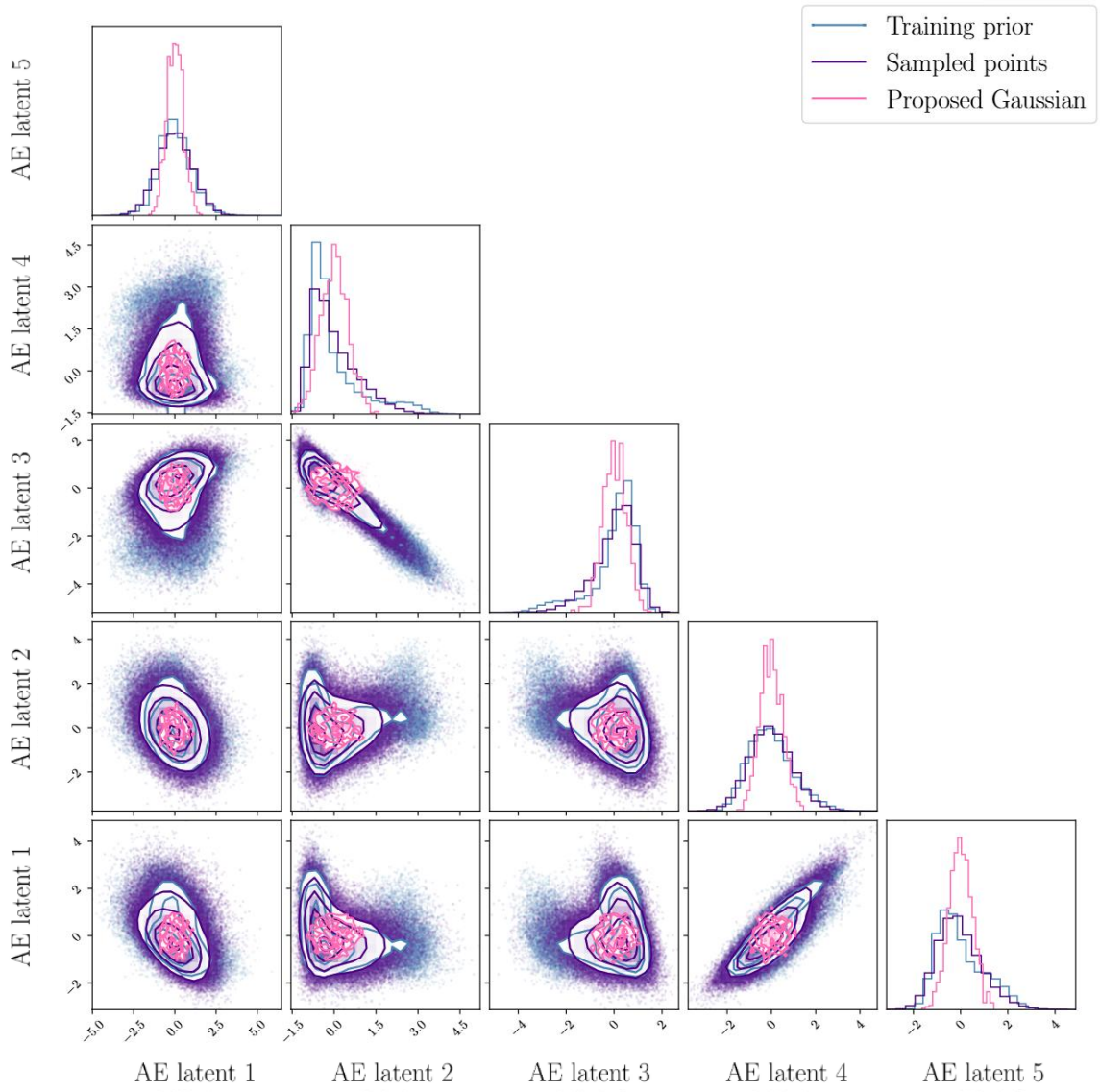


Figure 5.4.11: Corner plot of 1- and 2-dimensional marginalised distributions of training equations of state (blue), sampled points from the unconditional Normalising Flow (purple) and proposed equations of state (pink) from a Gaussian distribution. We present the first 5 out of the total 15 dimensions. The full figure can be found in the appendix section 7.1 figure A5.

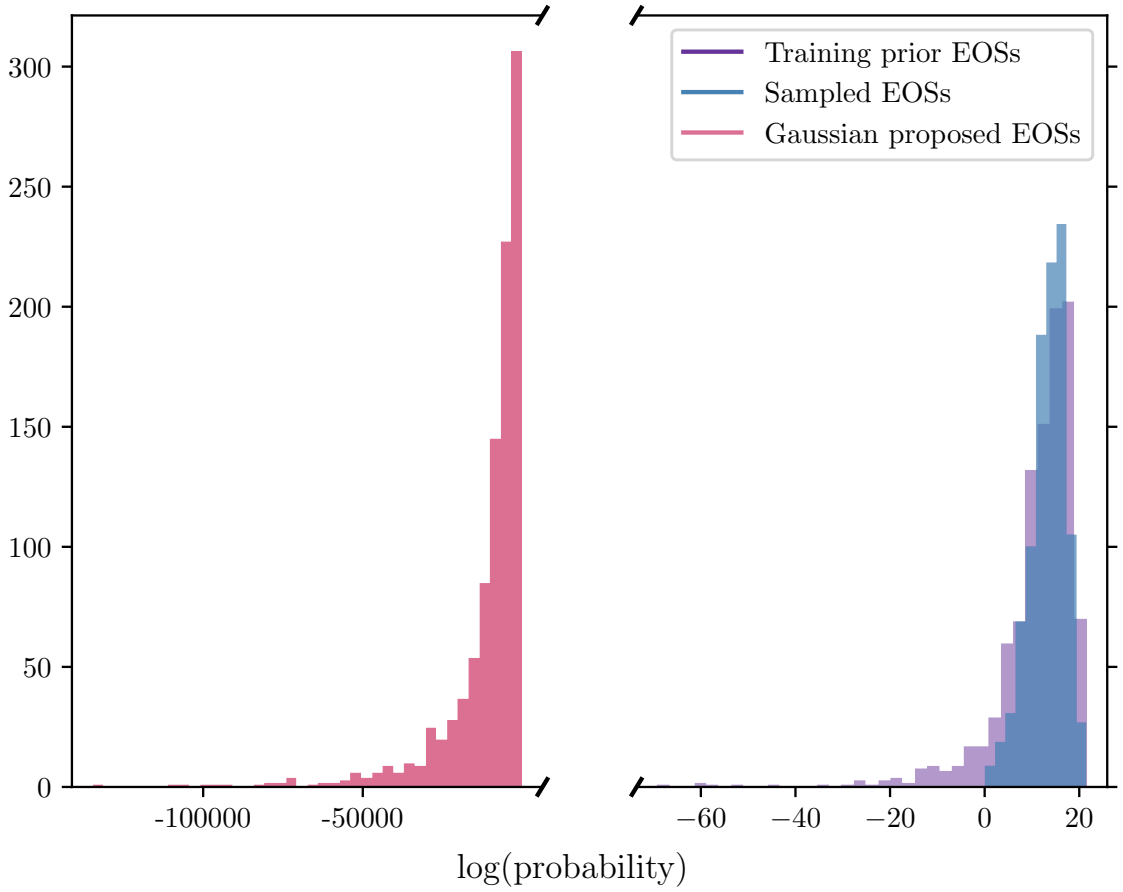


Figure 5.4.12: Histograms of log(probabilities) of proposed equations of state from a Gaussian distribution (pink), training equations of state (purple) and equations of state sampled by the unconditional Flow (blue).

sampled from the Flow have comparable and reasonable log(probabilities), however, the Gaussian proposed equations of state are significantly less probable, with $\Delta_{\text{log prob.}} = 286$ from equation 5.11. Again, this means the most probable proposed equation of state is around exp(286) less probable than that of the most probable sampled equations of state $\mathbf{\Omega}_{\text{samp.}}$, with the distribution of the log(probabilities) of $\mathbf{\Omega}_{\text{prop}}$ extending beyond -10^6, suggesting these are severely improbable equations of state which are on a scale of $\exp(10^6)$ less probable than that of any $\mathbf{\Omega}_{\text{samp.}}$.

As discussed previously, our Gaussian proposal $\mathbf{\Omega}_{\text{prop}}$ is not perfect and we expect some equations of state from this distribution to be beyond the reaches of the edge of the training data prior, especially in a high dimensionality space. However, log(probabilities) of this magnitude are not expected. This further suggests that a

Flow, whether conditional or not, can learn the training data distribution well, even in 15 dimensions, and the underlying structure it belongs to. The Flow then struggles to generalise beyond this and operate in regions of unknown parameter space, which may be between training and validation data points.

Summary of findings: an unconditional Normalising Flow trained to map the equation of state prior data space experiences the same behaviour; the Flow assigns proposed and sampled equations of state with vastly different log(probabilities). This suggests the structure is inherent to the compressed equation of state data distribution and how the Flow maps it.

5.4.4 Identifying artificial data structure

We now focus on trying to identify the structure which it appears the training and validation data belong to, which the Flow has been able to identify. Previously, we investigated whether issues arose due to handling data which lay outwith the training data distribution i.e. beyond the edges or physically outside the distribution which was learned. We learned this was not the case, but rather that there are specific points *within* the bounds on the training data distribution which have vastly different log(probability) to other points within the bounds. We therefore propose the presence of some structure within the learned distribution which the Flow has identified. Therefore, when we use the Flow to sample from this space $\mathbf{\Omega} \sim p(\mathbf{\Omega}|\theta)$, the Flow produces high probability samples $\mathbf{\Omega}_{\text{samp.}}$ which belong to the structure but evaluating randomly placed samples $\mathbf{\Omega}_{\text{prop}}$ which do not belong to the underlying 'grid' are of low probability. We note that by 'grid' we mean some underlying structure in 15 dimensions, not necessarily a uniform spacing of points in each dimension (which does not appear to be visible in 1− and 2−dimensional projections of the autoencoded training data).

These features of the data space make sampling methods very difficult. Evaluating the 15-dimensional space is already difficult for sampling methods, and this is only exacerbated by having a space which has inherent but unknown structure. Sampling methods will find it difficult to find the high probability regions of the space when the

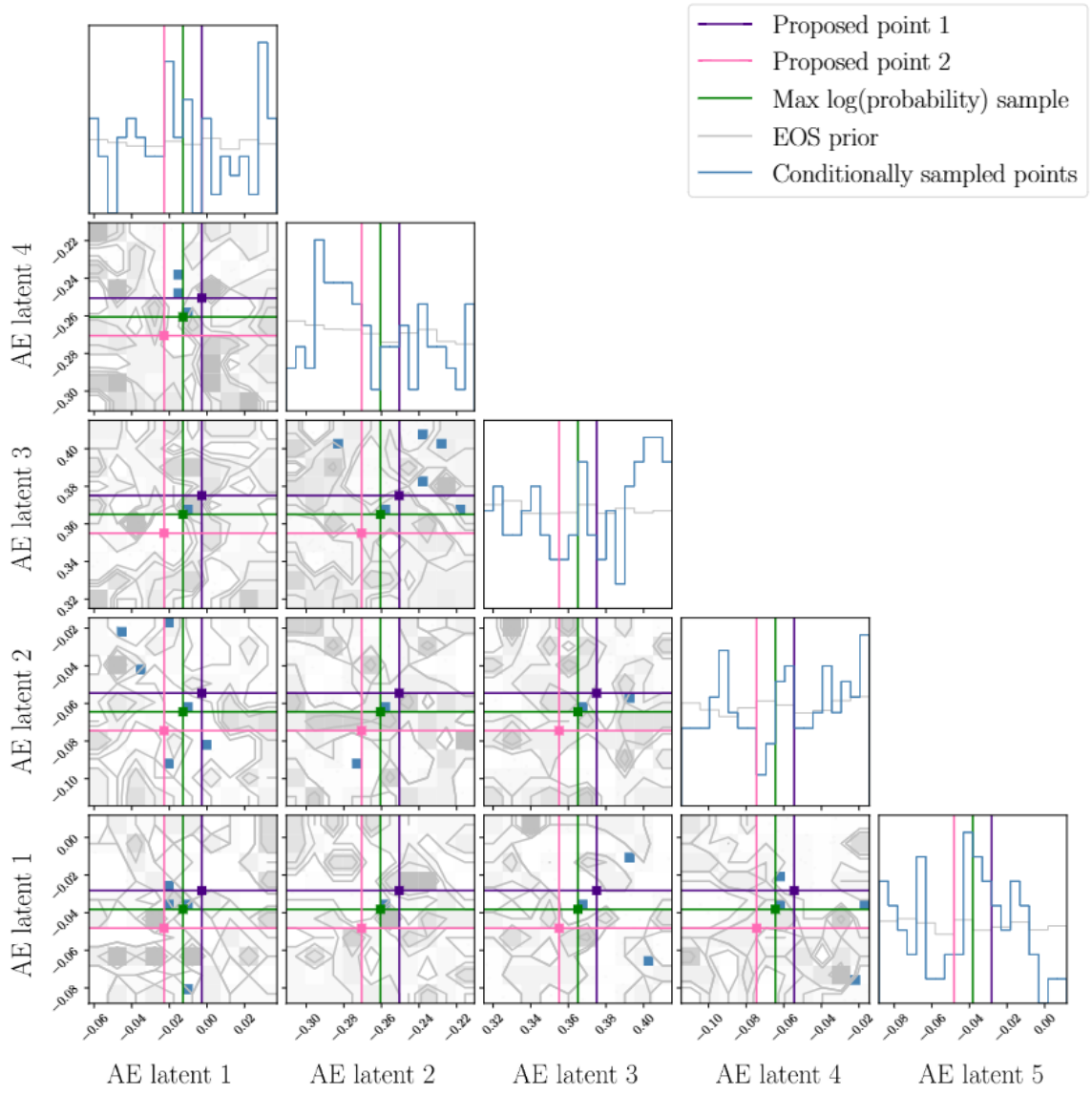


Figure 5.4.13: Corner plot of 1- and 2-dimensional marginalised distributions of points from the equation of state training data prior (grey), sampled points from the Normalising Flow (blue) and two proposed equations of state (pink and purple) which we traverse between. We present the first 5 out of the total 15 dimensions. The full figure can be found in the appendix section 7.1 figure A6.

surface is not (for the most part) smooth but are instead on sharp peaks surrounded by multi-dimensional troughs, or voids, of low probability space. We demonstrate the spike-like features by passing through the 15-dimensional space from one sampled location, $\mathbf{\Omega}_{\text{samp.}}$, to another. To do, we randomly select two sampled equation of state points, $\mathbf{\Omega}_{\text{samp, 1}}$ and $\mathbf{\Omega}_{\text{samp, 2}}$ and plot a 15−dimensional straight line path between these points. We take 100 total steps between $\mathbf{\Omega}_{\text{samp, 1}}$ and $\mathbf{\Omega}_{\text{samp, 2}}$ in 15−dimensions, evaluating the probability of points $\mathbf{\Omega}_{\text{steps}}$ along our path.

Figure 5.4.13 presents a corner plot of a small region of the first 5 out of the total 15 of the equation of state data space. The region plotted is populated by support from both the prior and points sampled by the Flow $\mathbf{\Omega}_{\text{samp}}$ and so there is no evidence we are near the edges of the space. We focus in around the most probable sampled point $\mathbf{\Omega}_{\text{max sample}}$ when sampling with θ_{170817} such that $p(\mathbf{\Omega}_{\text{max sample}}|\theta_{170817}) = \max(p(\mathbf{\Omega}_{\text{samp.}}|\theta_{170817}))$, as introduced previously. We wish to walk through regions of the equation of state data space which return high and low log(probability). To do so, we propose equations of state around the maximum probability sample $\mathbf{\Omega}_{\text{max sample}}$ from the Flow given conditions θ_{170817}. In this example, we propose equations of state in a 15-dimensional diagonal line through the maximum probability sample, $\mathbf{\Omega}_{\text{steps}}$, beginning and ending ± 0.01 in all 15 dimensions around the maximum log(probability) sample $\mathbf{\Omega}_{\text{max sample}}$. For each proposed equation of state from the diagonal line of $\mathbf{\Omega}_{\text{steps}}$, we find the average log(probability) given all samples from the event θ_{170817}, given the Monte Carlo sum over posterior samples θ in equation 5.9.

In figure 5.4.14, we show the average log(probability) given θ_{170817} of 100 steps taken through the most probable location $\mathbf{\Omega}_{\text{max sample}}$ in the 15-dimensional space. Over the course of a very small distance of ± 0.01 in 15 dimensions away from $\mathbf{\Omega}_{\text{max sample}}$, the log(probability) drops by almost 2 orders of magnitude in both directions, suggesting the region directly surrounding the most probable sample is highly improbable, or exp(100) less probable. However, in the surrounding regions of the most probable sample, there are other sampled points beyond a distance of ± 0.01 which we are aware are of comparable probability to the maximum probability sample, as demonstrated by figure 5.4.2. With no explicitly visible structure to the sampled points $\mathbf{\Omega}_{\text{samp.}}$ in this small region of space, this suggests artificial structure on a scale that is not visible in our current projections.

We plot samples z_{steps} from the Flow latent space, $\mathcal{Z}$, for this journey through the maximum probability sample $\mathbf{\Omega}_{\text{max sample}}$ in the Flow data space $\mathbf{\Omega}$ in figure 5.4.15. The figure presents four different 2−dimensional projections of 2 out of the 15 total dimensions of the Flow latent space. We plot a unit Gaussian for comparison, and demonstrate that sampled points in the data space $\mathbf{\Omega}_{\text{GW170817}} \sim p(\mathbf{\Omega}|\theta_{\text{GW170817}})$

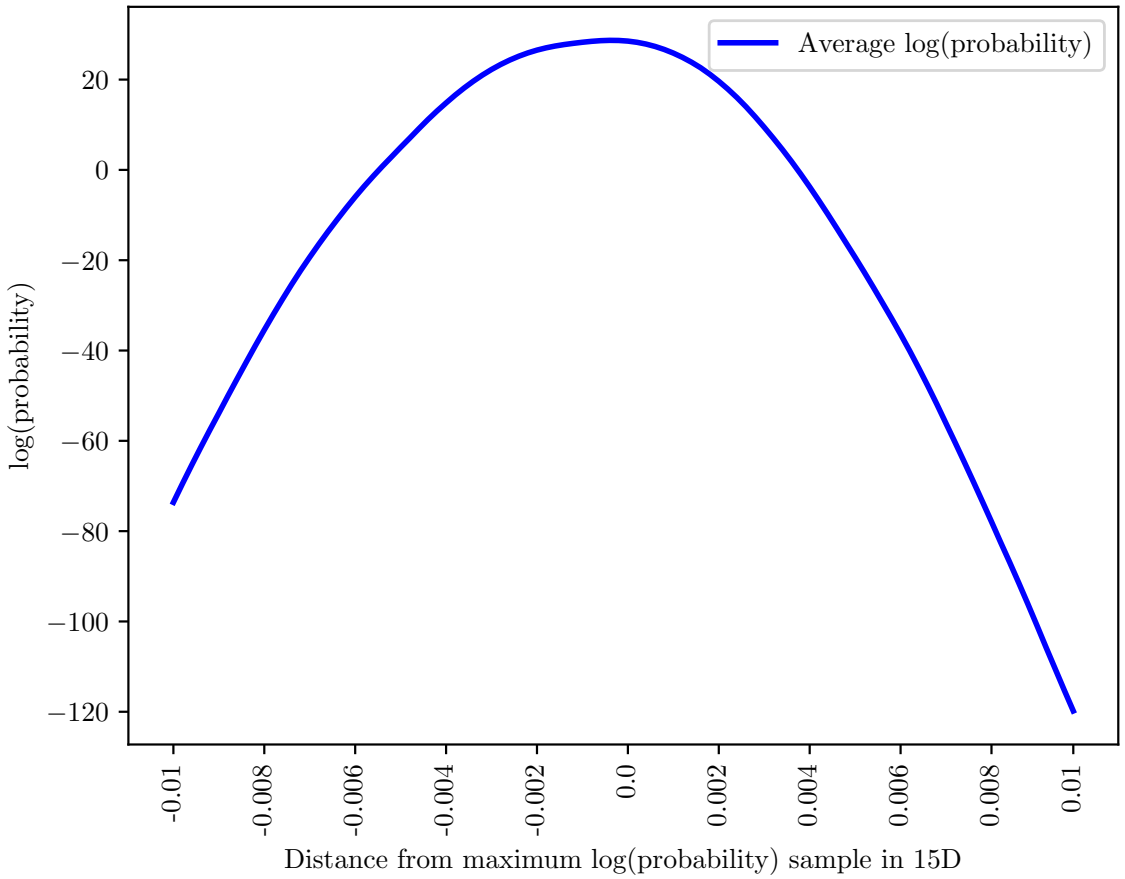


Figure 5.4.14: Average log(probability) over samples of path taken through ±0.01 in 15 dimensions walking through the maximum probability sample as given by the Flow.

which have undergone the forward transform to the Flow latent space, are Gaussian, as expected. The blue points which represent the latent samples z_{GW170817} generally overlap with the samples from the unit Gaussian in grey. The figure additionally presents the samples from the data space $\mathbf{\Omega}_{\mathrm{steps}}$ which are associated to the steps taken along the path between two sampled points. Their equivalent samples in the Flow latent space, z_{steps}, present a snake-like path which extends beyond the breadth of the unit Gaussian and learned latent space. For a walk along a 15−dimensional straight line through the Flow data space forward-transformed into the Flow latent space, a path of points in the latent space would still be expected, but we would expect this path to remain within the bounds of the unit Gaussian, somewhere near the centre. In some extreme cases, for example dimensions 12 and 14, the path extends up to 20 latent units beyond where would be expected. This further justifies that

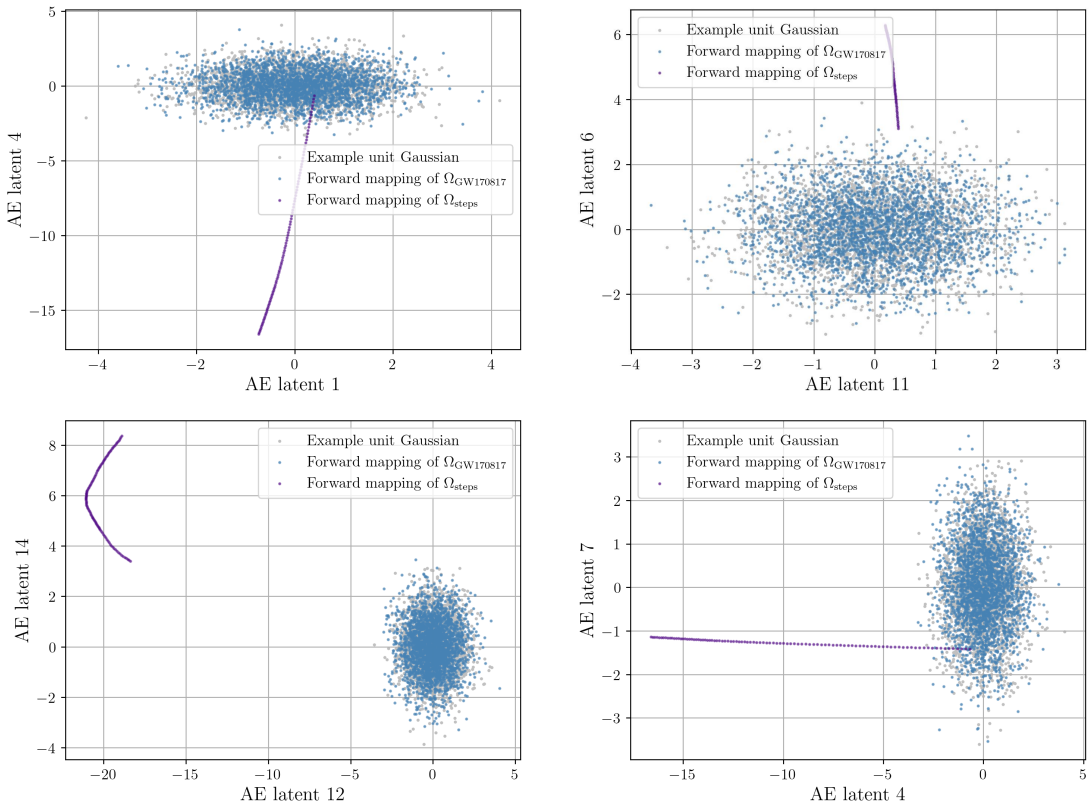


Figure 5.4.15: 2−dimensional scatter plots of the Normalising Flow latent space given the forward mapping of $\mathbf{\Omega}_{\rm GW170817}$ (blue), $\mathbf{\Omega}_{\rm steps}$ (purple) and an example unit Gaussian (grey).

the proposed regions of the equation of state data space $\mathbf{\Omega}_{\rm prop}$ lie beyond the Flow's understanding of the training data such that the mapping from data space to latent space cannot be performed as expected.

We additionally walk between 4 sampled points – $\mathbf{\Omega}_{\rm samp.\ 1}$, $\mathbf{\Omega}_{\rm samp.\ 2}$, $\mathbf{\Omega}_{\rm samp.\ 3}$ and $\mathbf{\Omega}_{\rm samp.\ 4}$ – in the 15-dimensional equation of state data space, $\mathbf{\Omega}$, to demonstrate the spike-like features of sampled points and low log(probability) space elsewhere. Figure 5.4.16 shows the first 5 out of the total 15 dimensions of the equation of state data space $\mathbf{\Omega}$ alongside the training data prior and points sampled by the Flow $\mathbf{\Omega}_{\rm samp.}$ conditioned on $\theta_{\rm GW170817}$. The four different colours of crosshairs represent the 4 randomly selected points: $\mathbf{\Omega}_{\rm samp.\ 1}$, $\mathbf{\Omega}_{\rm samp.\ 2}$, $\mathbf{\Omega}_{\rm samp.\ 3}$ and $\mathbf{\Omega}_{\rm samp.\ 4}$. We walk from point 1 in order until we reach point 4.

Figure 5.4.17 presents the log(probability) evolution as we follow the path of

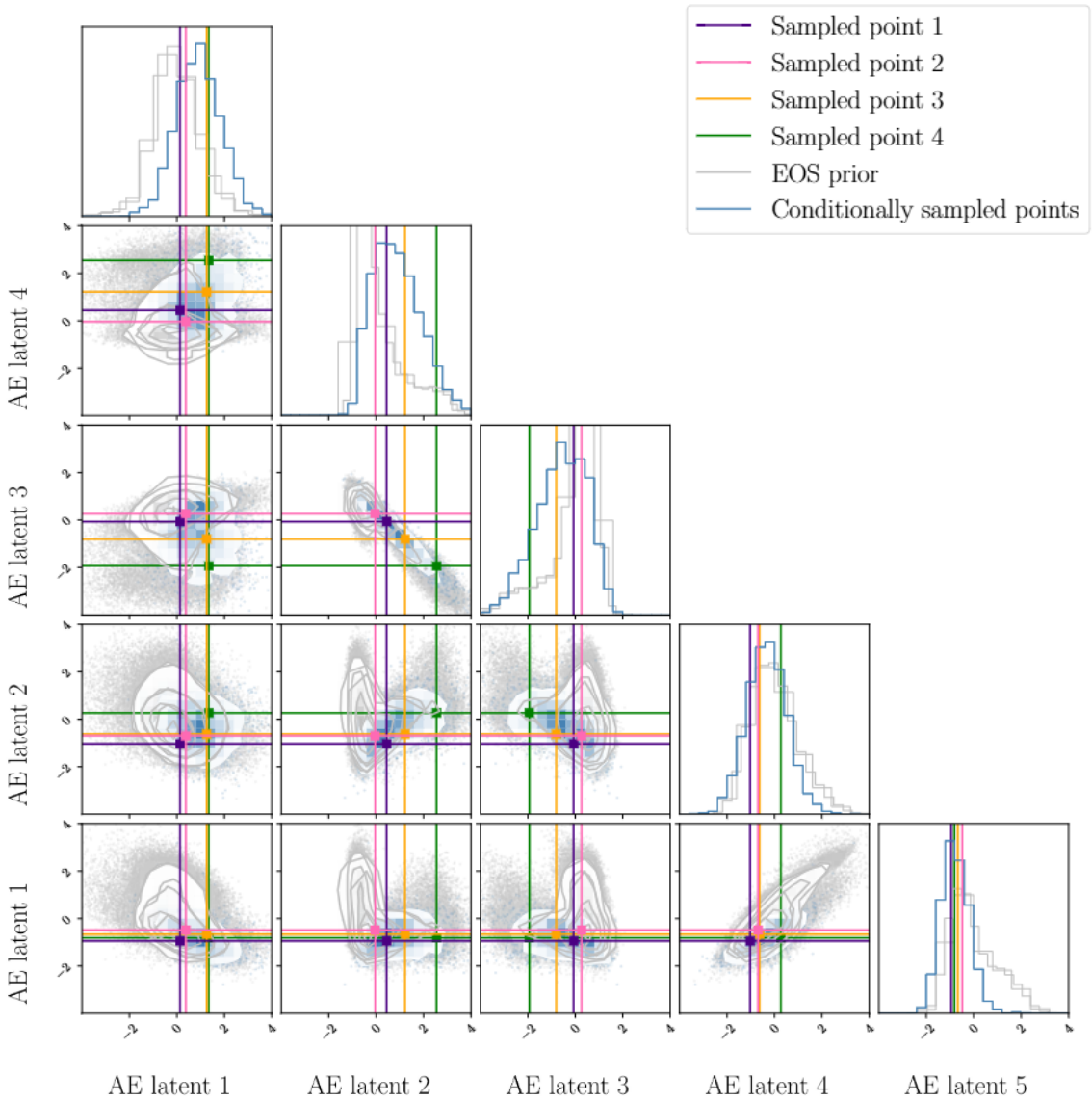


Figure 5.4.16: Corner plot of 1- and 2-dimensional marginalised distributions of points from the equation of state training data prior (grey), sampled points from the Normalising Flow (blue) and four proposed equations of state (pink, purple, green and yellow) which we traverse between. We present the first 5 out of the total 15 dimensions. The full figure can be found in the appendix section 7.1 figure A7.

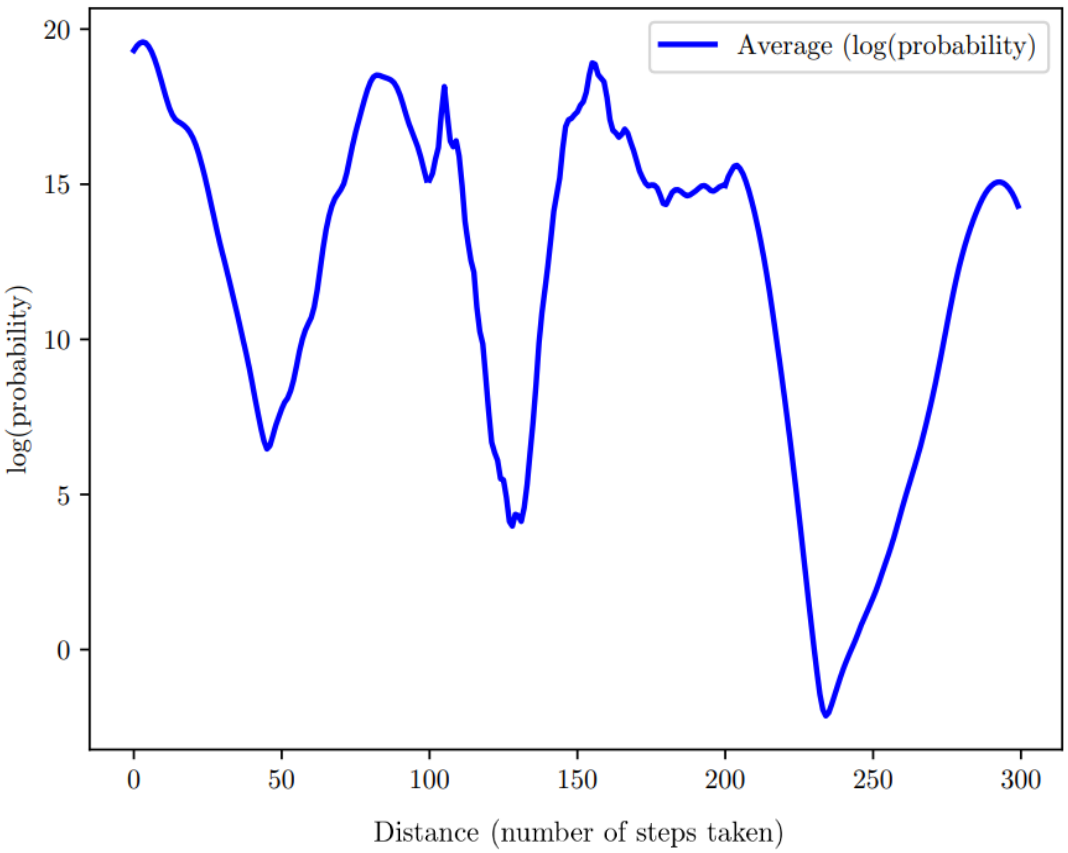


Figure 5.4.17: Evolution of log(probability) of proposed equations of state between points conditionally sampled by the Flow. Sampled points occur at steps 0, 100, 200, 300.

$\Omega_{\text{samp. 1}}$, $\Omega_{\text{samp. 2}}$, $\Omega_{\text{samp. 3}}$ to $\Omega_{\text{samp. 4}}$. We take 100 steps in the 15-dimensional space between each of these sampled points, and so steps 0, 100, 200, and 300 in figure 5.4.17 land on the sampled points. It is between these sampled points that the log(probability) drops dramatically, potentially by more than an order of magnitude, meaning a difference in probability of exp(10), which is especially present between points 2 and 3. We note that this difference in magnitude of log(probability) is less than that shown in figure 5.4.14, where steps are taken over a smaller distance through the 15−dimensional space. This suggests the presence of fine features with very low log(probability) around high log(probability) spikes, relative to broader steps across the space. Referring to figure 5.4.17, we see that these points have support from the prior (as expected and as previously clarified in figure 5.4.1) and their path, too, remains within the bounds of the training prior, as far as we can tell from 1-D and

2-D projections. Figure 5.4.17 demonstrates the spike- vs. trough-like behaviour that we anticipated. Sampling algorithms traversing the space looking for regions of high probability will struggle to find the spike-like features of the space and likewise struggle to leave the vast swathes of low log(probability) samples.

In attempting to identify the artificial structure learned by the Flow, we consider potential sources of the structure. While there is no clear grid-like structure of the points $\Omega_{\text{samp.}}$ which are sampled by the Flow, we must consider that we view the sampled points as 1− and 2−dimensional projections of a 15−dimensional space, where 12− of the dimensions are an autoencoded latent representation. In the original training data (which is scaled and then autoencoded), there is no evidence of structure in the $P - \rho$ plane presented in figure 5.1.3, such as branching, clustering, or repeated equations of state on a grid. However, we consider that perhaps the 12−dimensional autoencoder is not optimal for representing our $P - \rho$ training data. Therefore, we explore different number of dimensions for compression via the autoencoder.

Summary of findings: By following a straight line in 15 dimensions through a small region around the maximum log(probability) sample, we find a spike of log(probability). We repeat this process but walking between 4 sampled points, where we see repeated drops in log(probability) of $\mathcal{O}(10)$ between sampled points. When performing the forward transform of a walk through high and low log(probability) regions of the equation of state space, we find the walk extends beyonds the bounds of the unit Gaussian we expect all points in the data distribution to map to. This suggests the Flow has not learned a generic representation of the training data distribution as it cannot map any region to the latent Gaussian.

5.4.5 Low-dimensionality representation

It was expected that we were perhaps over-representing the EOS data by encoding the equation of state data to a higher number of latent dimensions than were necessary. If over-representing the data with too many latent dimensions, this may result in a lower-dimensional manifold existing in a higher dimensional space, e.g. a 2−dimensional plane in a 3−dimensional box. When we then evaluate the probability of a given

point in the high dimensional space, we then are likely to sample beyond the plane of information present. If the Flow has learned the presence of the low-dimensionality manifold, then points beyond this (in a higher-dimensional cloud around the plane) are very likely to be considered improbable, as we have previously found. We consider the possibility that perhaps encoding the 300 points in $P-\rho$ to an autoencoded latent space of 12−dimensions is too many. Perhaps the autoencoder struggles to stretch the information from the equations of state training data to 12−dimensions, or the Flow identifies a lower-dimensionality representation better suited to the data. We investigate the possibility that a smaller number of latent space dimensions would eradicate this problem, by training an autoencoder as outlined in section 5.2 to encode the same training data set to latent spaces of dimensions d where $d \in [2, 11]$. It is thought that there potentially exists a lower dimensional representation which is more suited to the problem and that using additional dimensions to describe the equation of state data becomes redundant.

Figure 5.4.18 presents training and validation loss as a function of training epoch for each of the trained autoencoders with latent spaces of dimension d where $d \in [2, 11]$. Each autoencoder was trained with identical training data (described in sections 5.1 and 5.2) for 1000 epochs with learning rate 0.005 and batch size 1024. The autoencoders trained also have architecture identical to the initial autoencoder, introduced in section 5.2; this ensures that our test of the different autoencoders in performing the tasks we will outline below only considers the change in size of latent space. Of course, testing different autoencoder architectures to find the most optimal for the task at hand using its given latent space dimensionality is desirable, but for the current tests we choose simplicity and neglect optimising the hyperparameters of the autoencoder and instead test changing the latent space size only. We find with figure 5.4.18 that autoencoders with latent spaces of dimensions 4, 8, 9, 10, and 11 have comparable losses to that of the 12-dimensional autoencoder (loss presented in figure 5.2.3). We choose to investigate these models further. Autoencoders with latent spaces of size 3 and 5 both struggle to train effectively and have losses which do not decay over time but rather remain stagnant at a significantly higher loss than the other models. We therefore neglect testing autoencoders of this size further, as they would require

further investigation and optimisation. We additionally find that autoencoders with latent spaces of size 2 and 6 also train with no evidence of overfitting but with final losses which are approximately an order of magnitude greater than those listed earlier.

We compare the quality of equation of state reconstruction with each of the autoencoders with latent spaces of d dimensions where $d = 2, 4, 6, 7, 8, 9, 10, 11$ in figure 5.4.19. To do so, we select an equation of state from the training data set and use each of the trained autoencoders to encode and decode this equation of state (one full pass of the trained autoencoder), comparing the decoded equation of state to that of the true input. We additionally present the residual of the reconstruction in the lower figure. We find that autoencoders with latent spaces of size $d = 2$ and $d = 6$ give equations of state which differ from the true equation of state (black dotted line) most, which is in agreement with the increased MSE loss presented in figure 5.4.18. The reconstruction differs from the true equation of state most at its kinks and extremities i.e. where the true equation of state deviates from the generic diagonal line trend. We find in particular than an autoencoder with 4 latent dimensions has comparable (and in some regions, better) reconstruction quality to autoencoders with $d = 7$ dimensions or more. It appears beneficial to use an autoencoder of reduced latent space dimensionality for future analysis as it can reconstruct equations of state at a comparable accuracy to autoencoders with many more dimensions. Additionally, a smaller latent space size is beneficial for later stages of analysis, where we can train Flow to learn the mapping of a smaller data space and, in performing hierarchical analysis, have sampling methods operate in a smaller number of dimensions, which allows for more efficient sampling.

Further, we train a conditional Normalising Flow as outlined in section 5.3.1 to learn the forward mapping from $d + 3$ dimensional equation of state plus scaler data space to $d + 3$ dimensional Gaussian conditioned on event information, as per equation 5.10. Once trained, we perform the tests we have previously used to assess the performance of the 12−dimensional autoencoder in section 5.4 to assess the performance of the conditional Flows trained to map $d + 3$ dimensional autoencoded latent spaces. The primary test we perform is assessing the log(probability) distribution

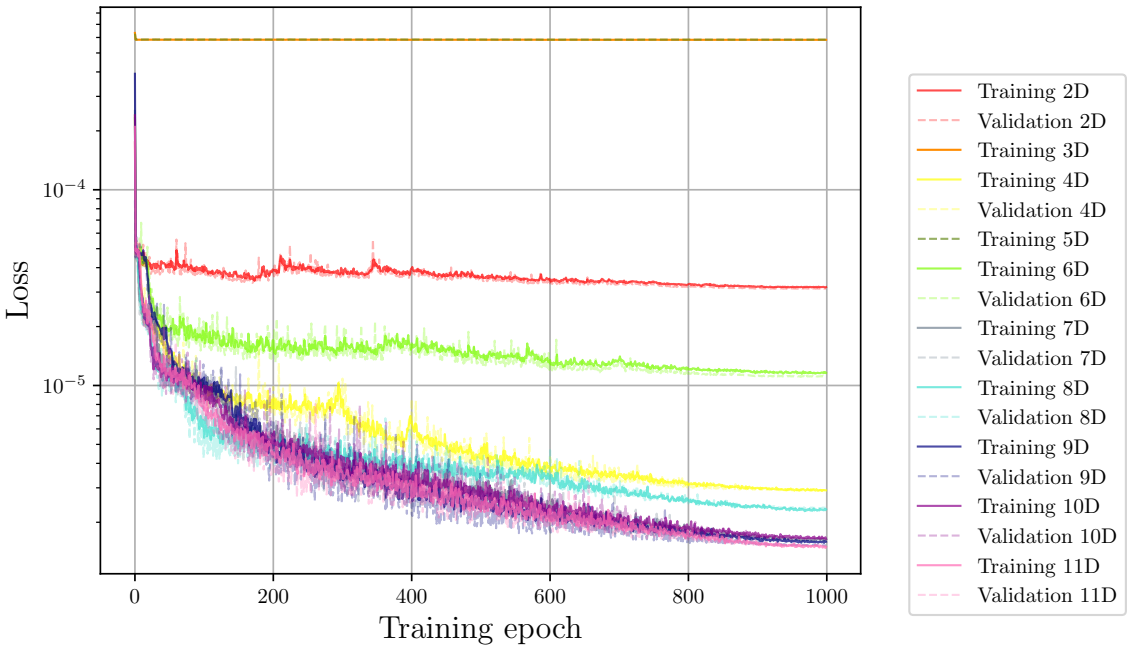


Figure 5.4.18: MSE training (solid) and validation (dashed) loss plotted as a function of training epoch for autoencoders with variety of latent space dimensions d where $d \in [2, 11]$.

of $p(\Omega_{\text{Gauss at max}}|\theta_{170817})$ for each of the $d+3$ dimensional Flows, the result of which is presented in figure 5.4.20. For autoencoders with $d = 9, 10, 11$, we find the log(probabilities) are very large and negative and are on a comparable magnitude to that of the original 12-dimensional autoencoder (see figure 5.4.4). The results demonstrate that there is support for $p(\Omega_{\text{Gauss at max}}|\theta_{170817})$ with log(probability) at a magnitude comparable to that of $p(\Omega_{\text{samp.}}|\theta_{170817})$ when using the Flows trained on the autoencoded spaces with $d = 6, 8$ but with significant outliers. The Flow trained using the autoencoded latent space with $d = 4$, however, shows a entire log(probability) distribution which is very comparable to that of the log(probabilities) of $p(\Omega_{\text{samp.}}|\theta_{170817})$ with a $d+3$ dimensional conditional Normalising Flow. With a higher dimensionality data space, one would expect a larger and more negative log(probability) distribution when assessing the probabilities of randomly selected points within the space. However, it is clear that for spaces with $d = 9$ and above, there are much more significant outliers to the distribution, compared to spaces of much smaller dimensionality.

Finally, we use each of the conditional Flows to 'walk' around the maximum likeli-

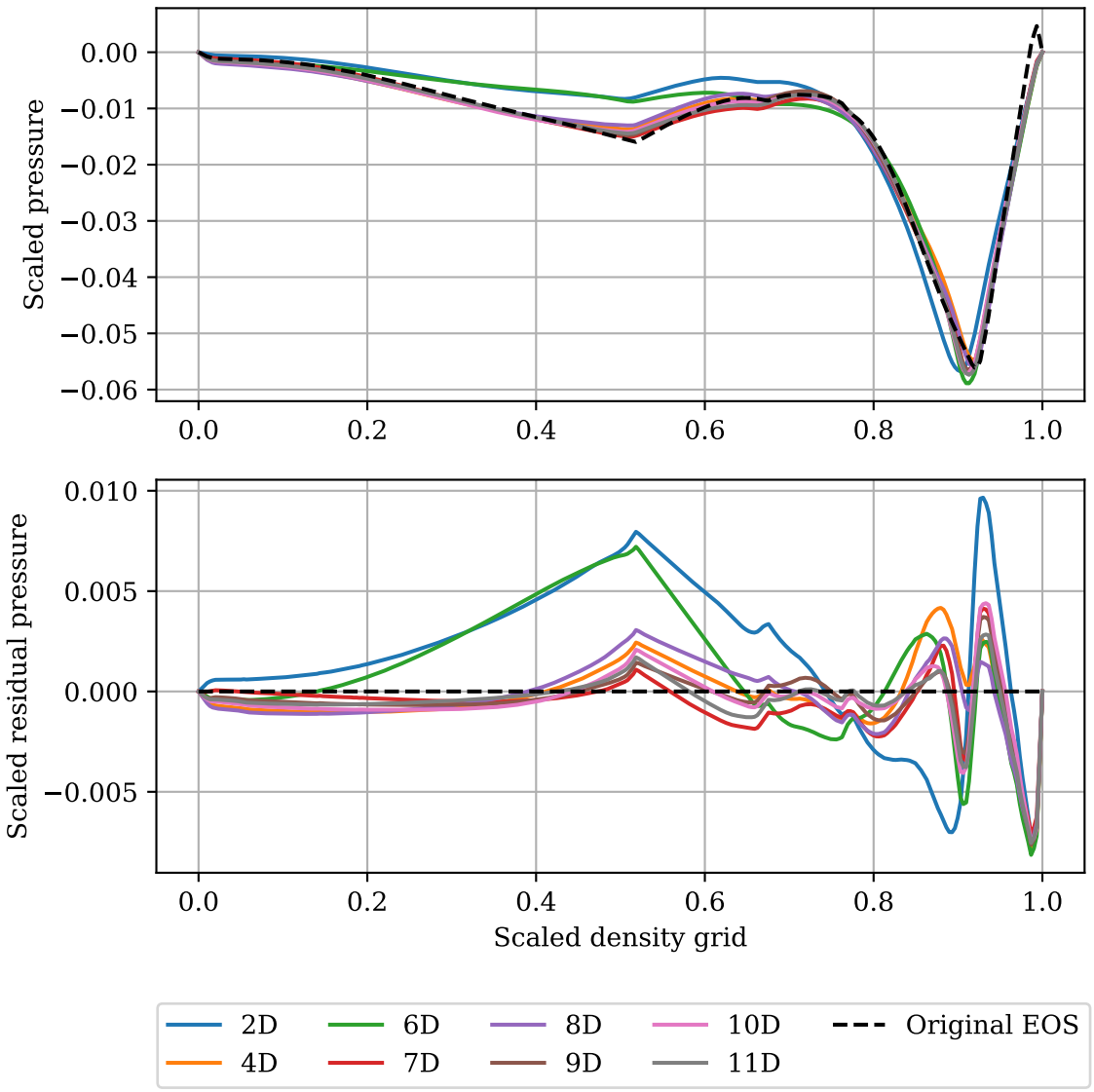


Figure 5.4.19: Plot of scaled pressure vs. scaled density of an equation of state from the training data set (black dashed line) and the reconstruction of this equation of state using autoencoders with different numbers of latent space dimensions (various colours, upper) and the residual of the reconstruction (lower).

hood sample $\Omega_{\text{max samp.}}$ given by conditionally sampling each of the $d+3$ dimensional Flows given θ_{170817}. Figure 5.4.21 presents the log(probability) range spanned for all Flows trained on autoencoded latent spaces of $d = 2, 4, 6, 7, 8, 9, 10, 11$. Flows trained to perform the conditional mapping from autoencoded spaces with $d = 11$ and $d = 10$ experience a change in log(probability) when travelling through the small volume of space of upwards of 3 orders of magnitude, a difference in probability on a scale of exp(1000). In the inset of figure 5.4.21, we visualise this figure again in region of log(probability) around 0. Flows trained with autoencoded latent spaces of sizes $d = 4, 6$ and 7 (plus scaling parameters) have a change in log(probability) which is more comparable with what one would expect in such a small volume of space. In particular, the latent space of $d = 4$ dimensions shows almost no change in log probability across the walk which is more realistic for such a test, and is indicative of a lack of peaks and troughs of probability in this small volume of $d+3$-dimensional space.

It is for the reasons outlined above that we choose to move forward in using the $d = 4$ dimensional autoencoded latent space as our equation of state data space for future work. We therefore use the conditional Normalising Flow trained to map a $d+3$ (7−dimensional once we include $\rho_{\text{max}}, P_{\text{min}}, P_{\text{max}}$) space equation of state conditioned on θ as this Flow has provided evidence that it is the least afflicted by learned artificial structure, but is not necessarily entirely absent. Therefore we must act to mitigate other potential sources of learned artificial structure.

Summary of findings: We reduce the number of autoencoder latent space dimensions which we then use to represent the equation of state data space to $d = 4$. We find an improvement in the variation in magnitude of log(probability) over the space, and less evidence of the features of peaks and troughs, or learned structure.

5.5 Data augmentation with Normalising Flows

We consider that the performance of the Flow may also be limited by our quantity of data. The Flow has been trained with 97 000 equations of state in the Flow data

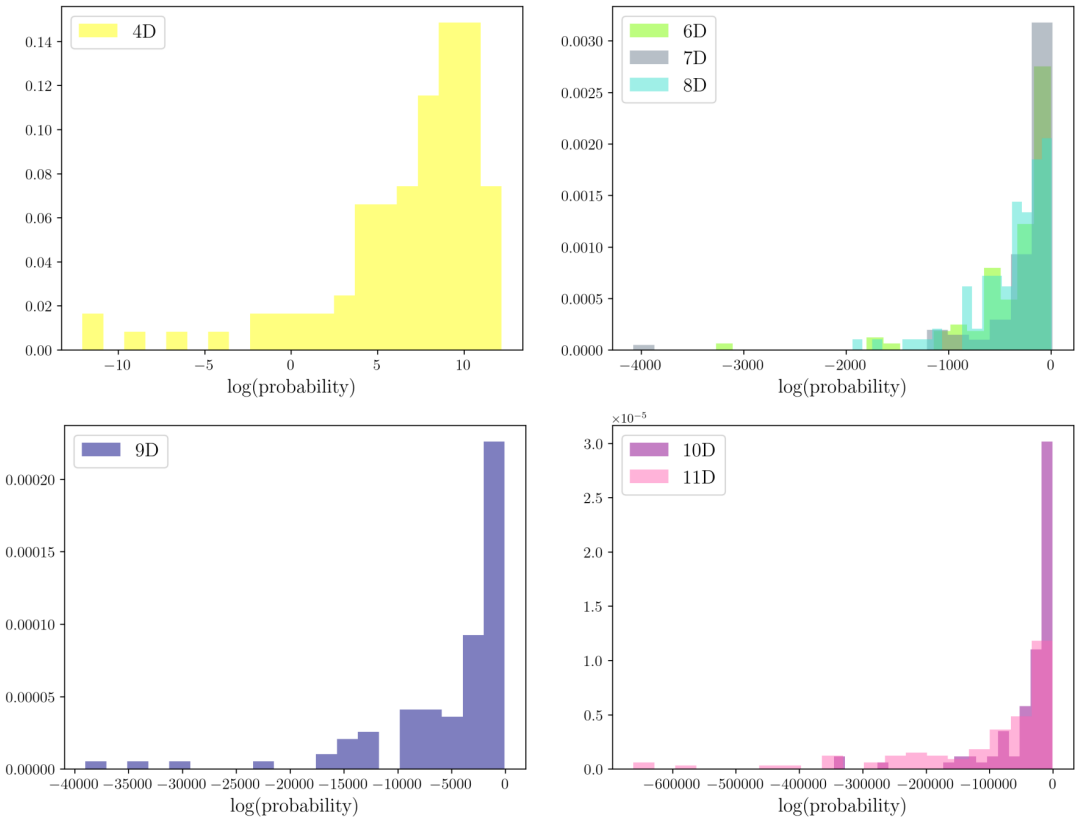


Figure 5.4.20: Histograms in log(probability) of proposed equations of state given GW170817 for Flows trained using $d + 3$ dimensional autoencoded data, where $d = 4, 6, 7, 8, 9, 10, 11, 12$. Each of the four figures shows histograms which have a similar log(probability) range.

space with one $[m_1, m_2, \Lambda_1, \Lambda_2]$ label per equation of state as a conditional term θ. It has been considered that in high numbers of dimensions, this quantity of data was not enough for the Flow to generalise to, especially if it contains or emulates some unknown artificial structure which the Flow can learn easily. However, equation of state data is expensive to generate; the process of randomly generating a high density phenomenological equation of state, detaching the nuclear parameters to construct the meta-model and then solving the TOV equations for the macroscopic parameters takes ~ 1 second per equation of state [281]. Although this is ~ 1 day of computation, we require external bodies to produce the data required for the analysis. Additionally, increasing the bulk quantity of training data that is processed per epoch makes the training process lengthier and more expensive.

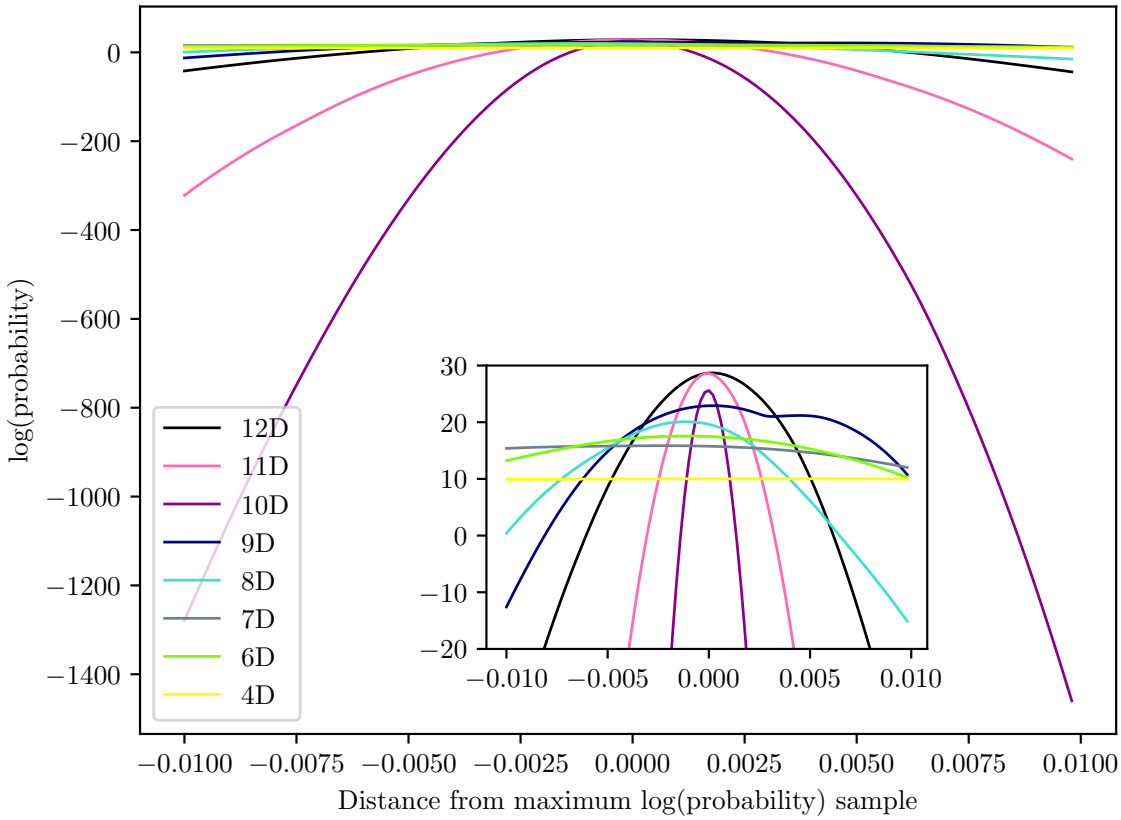


Figure 5.4.21: Log(probability) evolution walking through maximum probability sample using Flows trained using $d+3$ dimensional autoencoded latent spaces where $d = 4, 6, 7, 8, 9, 10, 11, 12$. The inset plot shows the same paths but at a focussed range of log(probability) around 0.

We introduce a Normalising Flows method which allows us to achieve our goal of presenting the Flow with more training data examples over the course of training while keeping the training costs down. The method allows for cheap and effective data production that can be done locally and rapidly with little impact to the current training routine. Firstly, we construct a combined data set of 11 dimensions consisting of the 7 equation of state data dimensions ($d = 4$ latent dimensions from the autoencoder and 3 scaling parameters) and the 4 conditional dimensions, $[m_1, m_2, \Lambda_1, \Lambda_2]$, presented in figure 5.5.2. We then train an unconditional Normalising Flow to model this combined data space

$$p(\Omega, m, \Lambda) = p_Z(\mathbf{z}) \left| \det \frac{\partial f(\Omega, m, \Lambda)}{\partial \Omega, m, \Lambda} \right|, \tag{5.13}$$

such that we can then randomly sample from the Flow and return samples from this combined data space. Sampling from the Flow is very efficient; we can produce 10^5

random samples in around 1 second, which consist of 10^5 equations of state in the 7-dimensional data space and one event associated to each equation of state in the $[m_1, m_2, \Lambda_1, \Lambda_2]$ space.

During training we can then continuously sample from this Flow, henceforth the 'regeneration Flow' to generate training data. Once sampled, the 11 dimensions are separated into the 7 data dimensions $\mathbf{\Omega}$ and the 4 conditional dimensions θ, and training of the conditional Flow resumes normally as described in section 5.3.1. We sample from the regeneration Flow 10^5 times on every training epoch, of which there are 1000 in total. By the end of training, the Flow has now seen 10^8 discrete training data samples in both the data $\mathbf{\Omega}$ and conditional θ spaces. Sampling from the regeneration Flow during training adds ~ 1000 seconds on to Flow training time (sampling from the regeneration Flow once per epoch), but at a significant gain; by sampling from the regeneration on each epoch, we expose the Flow to orders of magnitude more samples from the training data distribution, so the increased time is outweighed by the gain in total number of pieces of training data. We validate on a random selection of 25 000 equations of state from the real training data set, to ensure the regeneration Flow remains to produce samples that are consistent with the true data set.

The results of training the regeneration Flow are presented in figures 5.5.1 and 5.5.2. Firstly, figure 5.5.1 presents the training and validation losses for training the regeneration Flow with an architecture of 1 residual block, 2 transforms and 32 neurons per transform. The entire training routine of 500 epochs took less than 10 minutes with a batch size of 4096 and at a learning rate of 0.001 which decays during training with cosine annealing. Figure 5.5.2 presents a set of 10^4 samples from the regeneration Flow in blue on top of the 11-dimensional combined equation of state $\mathbf{\Omega}$ and event data θ training set in purple. We see the samples from the regeneration Flow agree well with the training distribution, further clarifying the Flow has learned the 11-dimensional data distribution well.

We employ the regeneration Flow in training the conditional Normalising Flow

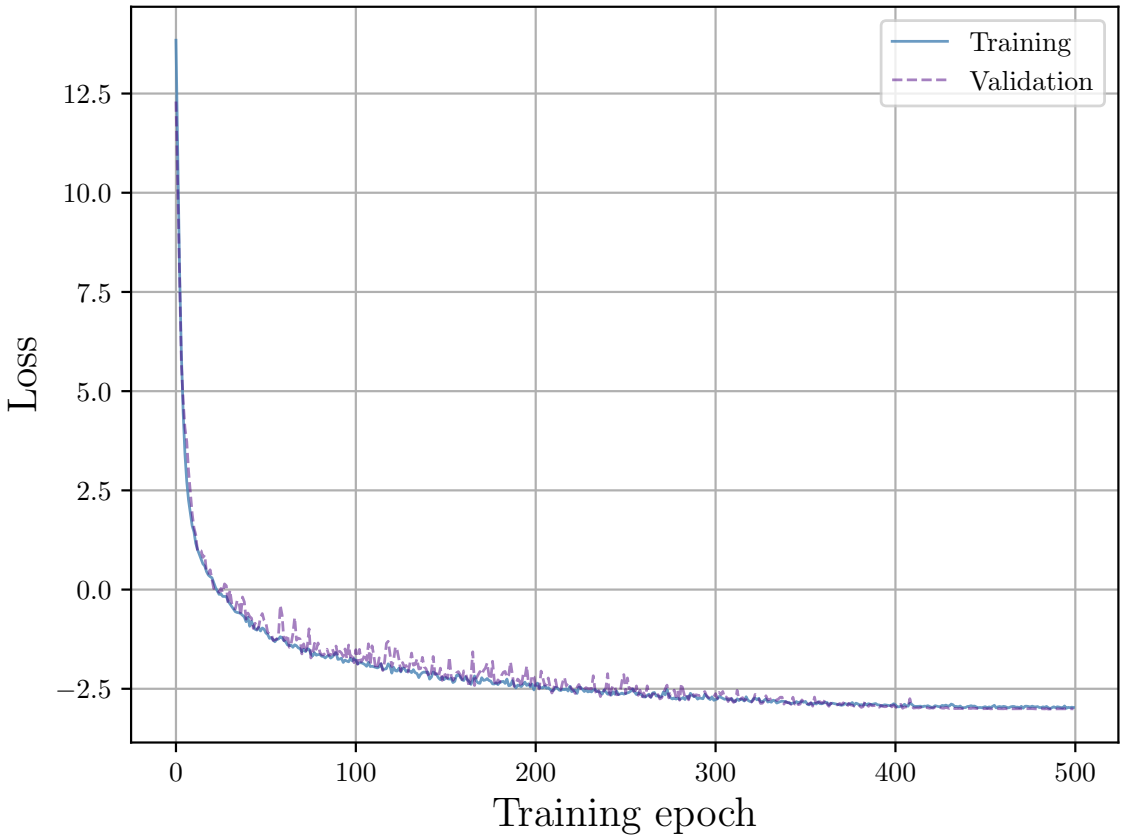


Figure 5.5.1: Training (blue) and validation (dashed purple) loss for the 11-dimensional regeneration Flow trained as per equation 5.13.

by sampling from the regeneration Flow on each training epoch in order to provide the training data. We train the conditional Flow with the same architecture as the regeneration Flow to maintain simplicity and remove sources of potential over-complication. The plot of training loss over time for the conditional Flow trained using the regeneration Flow is presented in figure 5.5.3. We see the Flow learns consistently throughout and the training loss evolves well relative to the validation loss, which is the loss evaluated given a data set of size 10^4 which is sampled from the regeneration Flow once at the beginning of training and which is then used throughout training as validation data and is never replaced.

However, our initial tests of training and sampling performance of the original conditional Flow (section 5.4.2) did not show any concerning results, especially in loss performance. Therefore we complete the same tests of evaluating the probability of proposed samples and traversing the space through regions of known high probability

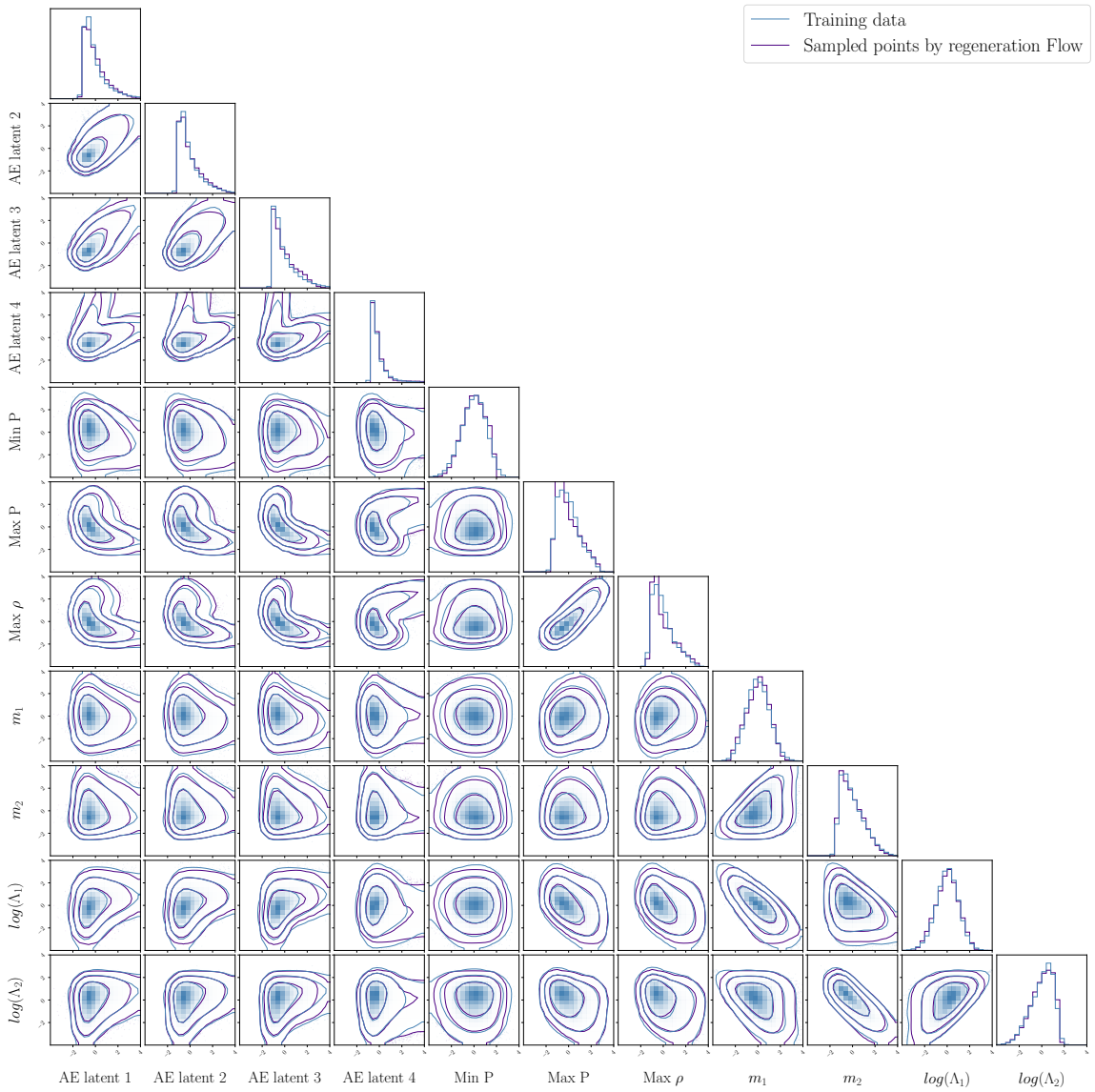


Figure 5.5.2: Corner plot of 1- and 2-dimensional marginalised distributions of both $\mathbf{\Omega}$ and θ training data (blue) alongside points sampled from the 11-dimensional regeneration Flow (purple).

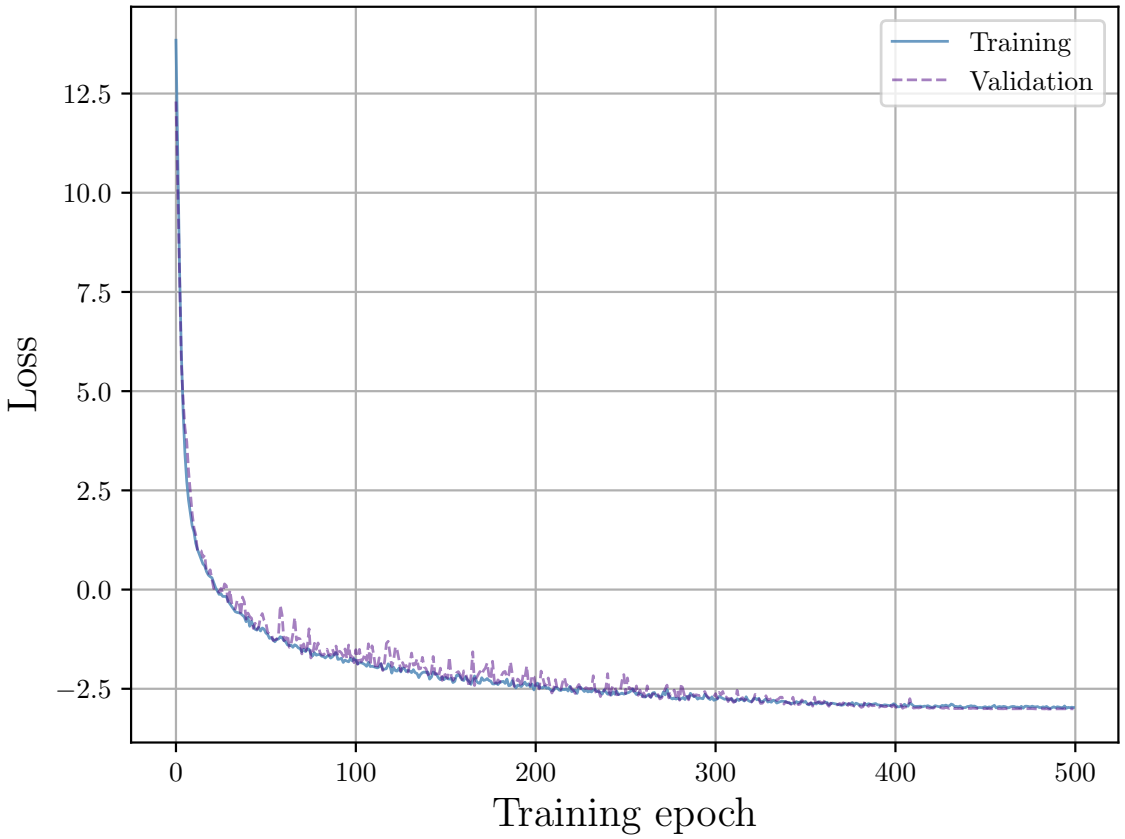


Figure 5.5.3: Training (blue) and validation (purple) loss for the conditional Flow trained using the 11-dimensional regeneration Flow.

to fully test the result of data augmentation.

5.5.1 Testing the augmented conditional Flow

We complete the same test introduced in section 5.4.1 of evaluating the log(probability) of $p(\mathbf{\Omega}_{\text{Gauss at max}}|\theta_{170817})$, equations of state $\mathbf{\Omega}$ proposed from a small Gaussian distribution centred around the maximum likelihood sampled point, conditioned on event samples θ from GW170817. We plot the corner plot of these proposed samples $\mathbf{\Omega}_{\text{Gauss at max}}$ alongside those sampled by the Flow given event information $\mathbf{\Omega} \sim p(\mathbf{\Omega}|\theta_{\text{GW170817}})$, which we discovered the Flow attributed vastly different magnitudes of log(probability) to, along with the prior training data in section 5.4.1 figure 5.5.4. Again, we do not find that the proposed samples $\mathbf{\Omega}_{\text{Gauss at max}}$ are on the extremities of the space and find the point selected which the Gaussian has been constructed around (the most probable sample returned by the conditional Flow given event samples

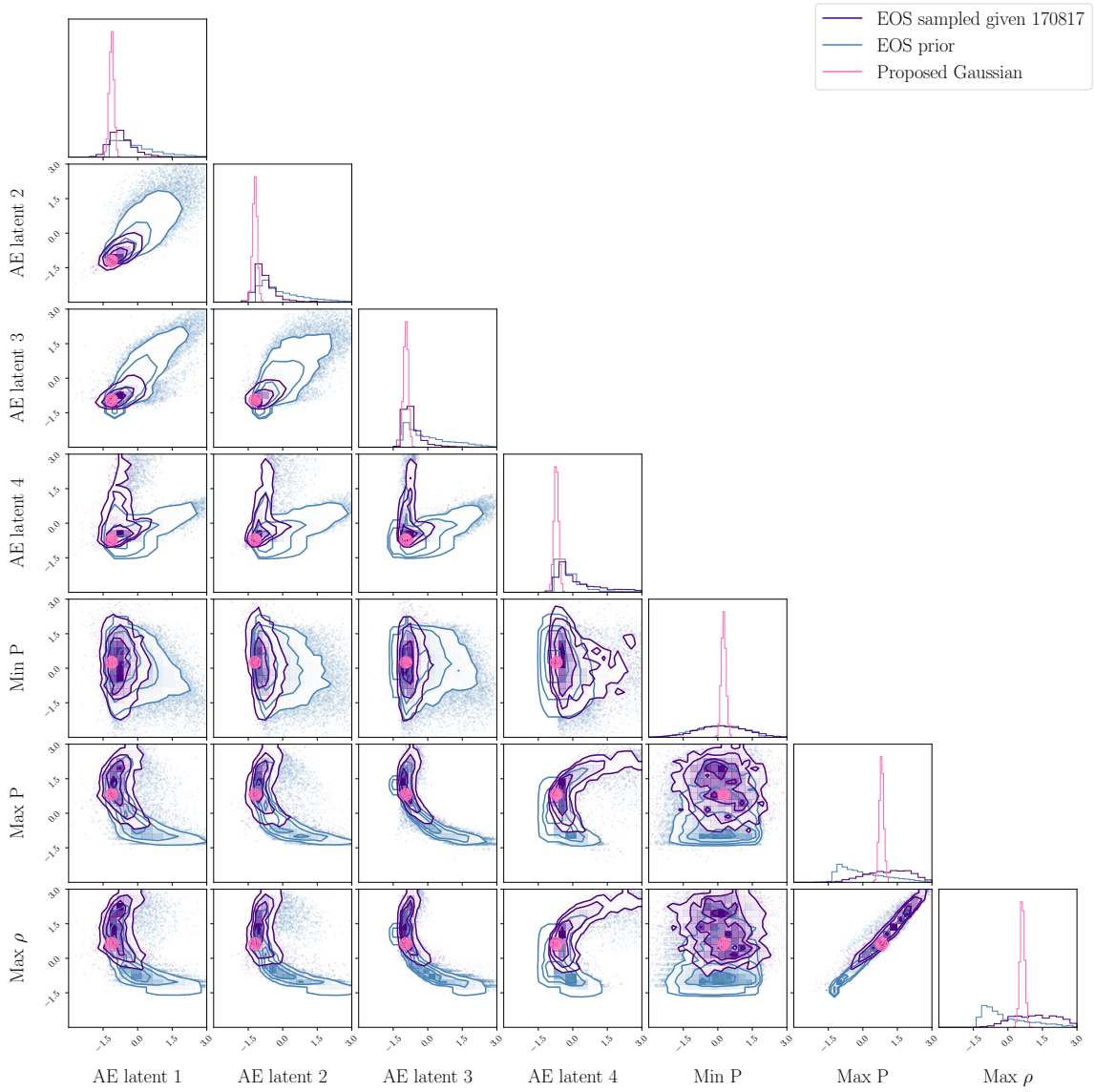


Figure 5.5.4: Corner plot of 1- and 2-dimensional marginalised distributions of training equations of state (blue), sampled points from the regeneration-Flow-trained conditional Normalising Flow (purple) and proposed equations of state (pink) from a Gaussian distribution.

$\mathbf{\Omega}_{\text{max prob}} = \max(p(\mathbf{\Omega}|\theta_{\text{GW170817}}))$ to be surrounded by other $\mathbf{\Omega}_{\text{samp.}}$ sampled by the Flow which we know to be of comparable probability.

The histogram of log(probabilities) of proposed equations of state around the maximum likelihood point are presented in figure 5.5.5. We find the log(probabilities) of the proposed equations of state $\mathbf{\Omega}_{\text{Gauss at max}}$ are of comparable magnitude to those sampled $\mathbf{\Omega} \sim p(\mathbf{\Omega}|\theta_{\text{GW170817}})$ with significant overlap in both of the distributions. Additionally, we find that the log(probability) distribution of $p(\mathbf{\Omega}_{\text{Gauss at max}}|\theta_{\text{GW170817}})$

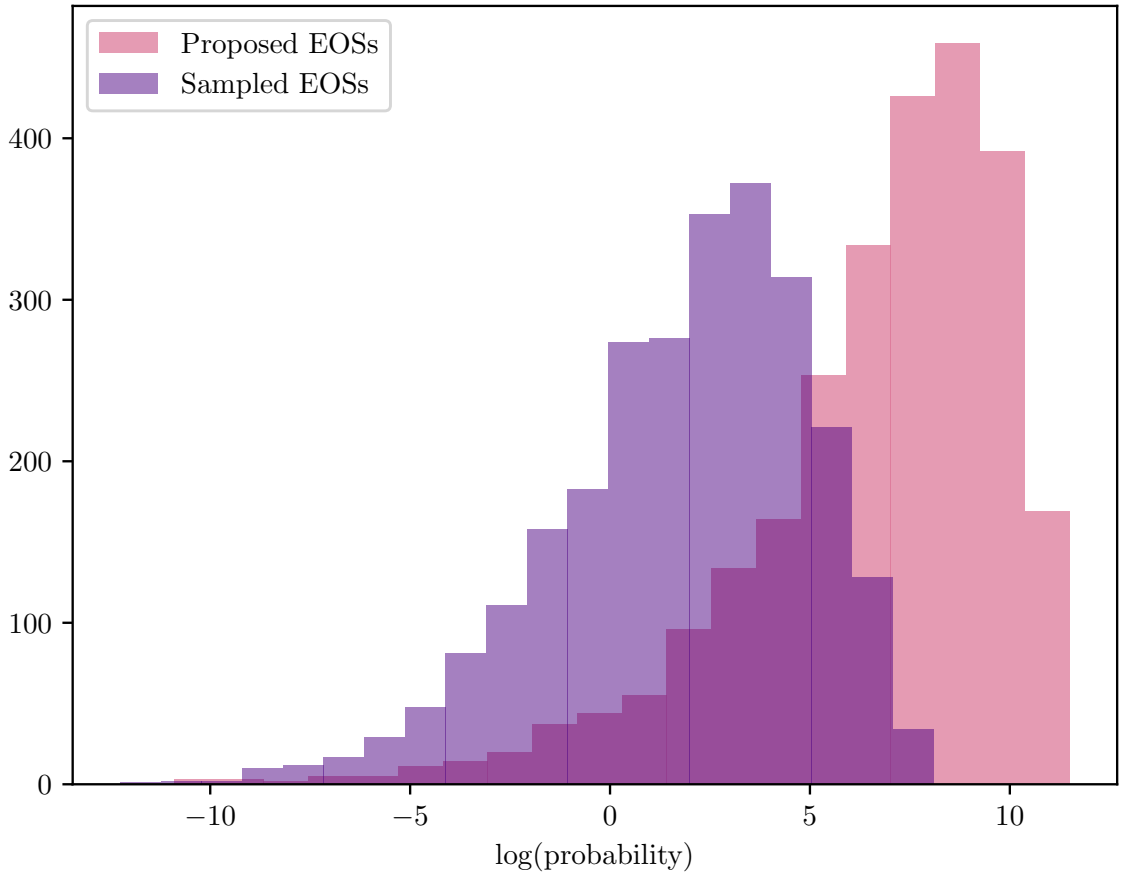


Figure 5.5.5: Histograms in log(probability) of sampled equations of state given GW170817 for the regeneration-Flow-trained conditional Flow (purple) and Gaussian proposed equations of state around $\Omega_{\text{max sample}}$ given the same condition (pink).

extends beyond the maximum of log(probability) of points $\Omega \sim p(\Omega|\theta_{\text{GW170817}})$. This suggests the Flow has found the region of maximum probability in the 7-dimensional space, and find that our chosen $\Omega_{\text{max prob}}$ is not necessarily the true maximum, and that points nearby have higher log(probability). In sampling $\Omega \sim p(\Omega|\theta_{\text{GW170817}})$, we are sampling a finite and relatively small number ($\sim$ 2000) of points in the 7-dimensional space, and therefore do not necessarily sample the true maximum. This suggests the 7-dimensional space does not have the same level of learned structure as perviously found and instead finds a smooth and comparably probable surface.

We next evaluate the probability of regions of the space in a path through the maximum likelihood point as previously performed in section 5.4.4 . Similar to previous tests, we evaluate the probability of points ±0.01 in 7 dimensions around $\Omega_{\text{max prob}}$. Figure 5.5.6 presents the log(probability) evolution over this path, aver-

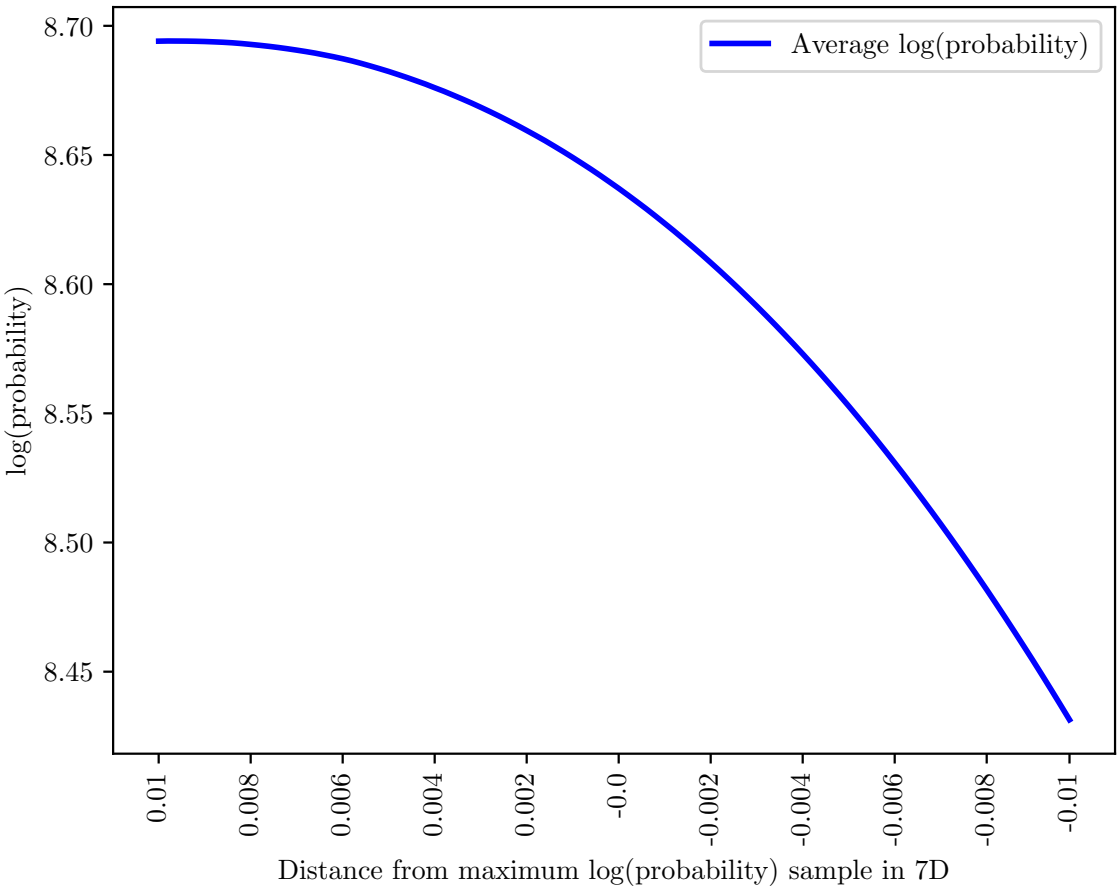


Figure 5.5.6: Log(probability) evolution walking through the maximum probability sample using a conditional Flow trained using the 11-dimensional regeneration Flow.

aged over all event samples θ_{170817}. We see the log(probability) evolve on a scale of 0.3 log(probability) which is starkly different to our previous change of 100 in log(probability). We also see a result which is in agreement with the findings from figure 5.5.5, where the log(probability) of $\mathbf{\Omega}_{\text{max prob}}$ given θ_{170817} is not the true maximum, and that points at a distance of 0.01 from $\mathbf{\Omega}_{\text{max prob}}$ in 7-dimensions are actually more probable. This is expected from a smooth 7-dimensional volume of points which we sample finitely from.

This method of data augmentation has improved the Flow's understanding of the space and has constructed a mapping which appears to be free of evidence of learned artificial structure. The augmentation process is a cheap and effective solution for problems which are limited by the quantity of data available. It allows the user to generate more samples from an existing data distribution. Note, these additional

samples are not 'new' samples from a posterior, but are additional synthetic samples from the posterior we currently have. Providing the conditional Normalising Flow with additional samples for training inhibit the Flow's ability to learn any structure or memorise as there is a larger quantity of training samples and also introduces the regeneration Flow's own intrinsic noise on learning the true training data. A Normalising Flow will not learn the data perfectly, and so even if the regeneration Flow learns 11-dimensional artificial structure, when sampling the Flow's learned mapping of the 11-dimensional space, we sample given the Flow's own intrinsic uncertainty which obscures the learned structure. This has proven, alongwith the reduction in dimensionality of the autoencoded latent space from 12 to 4 dimensions, to be helpful in allowing the Flow to learn a generic understanding of the mapping.

Summary of findings: We introduce a 'regeneration' Flow, an unconditional Normalising Flow trained to learn to map the joint $\mathbf{\Omega}$ and θ space. Once trained, the Flow can be sampled from during training of the conditional of equation of state prior Flows to augment the training data and provide more training data examples, obscuring the structure present with the Flow's intrinsic noise. As a result, the Flow learns the generic behaviour of the equation of state data space.

5.6 Conclusion

This chapter discusses the equation of state data we wish to use for future hierarchical inference of the neutron star equation of state, before introducing the compression methods used. Subsequently, we discussed evidence of learned artificial structure that arose when using the Normalising Flow in a new capacity of evaluating the probability of regions of the equation of state data space rather than conditionally sampling the space directly, with the ultimate goal of hierarchical inference of the neutron star equation of state.

The new data introduced is a set of unified equations of state which are solved via the TOV equations to obtain their respective mass and tidal deformability parameters. These consist of piecewise polytropic equations of state at high density, each of which

is adjoined continuously to a low density equation of state described by a nuclear meta model. Each equation of state has a unique minimum pressure, maximum pressure and maximum density and is more expressive than a simple polytrope and spans a broader density range. This ensures the analysis is next-generation-proof, where we can no longer assume a fixed crust equation of state in equation of state inference. However, the more complex equations of state require a more involved compression method, which retains the unique features of each of the individual equations of state to allow for high quality inference. It is for this reason we choose to employ an autoencoder with convolutional layers. We compress the equations of state to a latent space of 12 dimensions, retaining an additional 3 scaling parameters such that we can reconstruct the equations of state fully after compression. This also allows us to maintain as many unique features of individual equations of state as possible.

Hierarchical analysis aims to infer the underlying true equation of state given multiple observations of neutron star observables. In this case, we wish to combine information from multiple GW observations of BNS mergers for an overal inference of the neutron star equation of state.

However we find that in using the conditional Normalising to evaluate the probability of proposed equations of state instead of sampling like was done in the previous chapter, that the conditional Normalising Flow severely struggles. This is potentially detrimental for hierarchical inference because we want to use sampling algorithms to survey the parameters space and evaluate the likelihood at given points. We find the Flow shows no signs of traditional overfitting but struggles to generalise to equations of state not from the data set we use to train the Flow. We test proposal equations of state from a variety of distributions and also try to walk through the space to visualise the proposed 'structure' that the Flow has managed to over-specify to.

We find that reducing the dimensionality of the compressed equation of state space from 12− dimensions to 4 helps significantly. We do so by retraining the autoencoder to instead compress the 300 points along each equation of state to only 4 dimensions. Doing so allows the conditional Normalising Flow which is trained using this data to

perform much better on tests such as evaluating the probability of regions of space around sampled equations of state and also in the lack of variation in log(probability) found when walking through regions of parameter space.

Additionally, we introduce a Normalising Flow method to augment our training data set. We train a conditional Normalising Flow to model the data and conditional spaces together. During training, we can then sample from this unconditional Flow. We find this helps the conditional Flow to learn a smooth surface with no evident structure using our tests. This is helpful for high-dimensionality problems where the amount of training data can be critical in the quality of your tool but may be expensive to generate.

We find that the main evidence of structure most likely was due to training a Flow model with too many dimensions than were necessary for the problem at hand. This means we may have had dimensions involved in the problem which did not contain any valuable information. In asking the Flow to then evaluate randomly proposed regions of the space which were beyond the valid region of information for this dimension (for example, a plane within a higher-dimensional volume of space), the Flow cannot perform well. Higher dimensionality problems are difficult to visualise and also to present in 2 or 3 dimensions which is what we are limited to in trying to find and present the cause of the problem. Our regeneration Flow method also allows us to generate more training data examples to improve training, and have supported evidence that the artificial structure in the space is no longer present.

To conclude, we wish to express the need for caution when applying Flow models to new sets of training data, especially in high numbers of dimensions or where training data is limited. By exploring issues that arose early during our application of hierarchical analysis, we were able to work around some issues in dimensionality reduction and employ a new but simple method Normalising Flows for data augmentation. We can now perform hierarchical analysis of the neutron star equation of state.

Chapter 6

Hierarchical inference of the NS EOS

With the framework developed in chapter 5, we now wish to perform hierarchical inference of the NS equation of state. In chapter 5 section 5.3.1, a Normalising Flow was trained to map a compressed representation of our equation of state training data conditioned on BNS merger properties. Our NS equation of state training data, introduced in section 5.1, consists of unified equations of state with nuclear meta-model low density and piecewise polytropic high density. The equations of state each span an region covering the NS inner crust to the inner core of the NS, where each equation of state has its own unique crust. These specifications of the equation of state training data are particularly pertinent when considering hierarchical analysis of the NS equation of state in next generation GW detection. As previously mentioned in chapter 2 section 2.3, both improved tidal deformability measurement and the expected increase in number of BNS merger observations with next generation detectors are expected to improve the quality of equation of state inference, so much so that we can no longer assume a fixed crust equation of state.

In the previous chapter, we laid the groundwork for hierarchical inference of the NS equation of state, but identified some problematic behaviour in how the conditional Flow learned the mapping to the equation of state training data, $\mathbf{\Omega}$. Through several stages of testing in section 5.4, artificial features were identified in the learned $\mathbf{\Omega}$ space and manifested as peaks and troughs of probability, making any future sampling of the

space very difficult. By reducing the dimensionality of the autoencoded latent space which constitutes $\mathbf{\Omega}$ in section 5.4.5, we removed problematic probability evaluations of regions of $\mathbf{\Omega}$ by the Flow. We additionally introduced the 'regeneration Flow' in section 5.5, an unconditional Normalising Flow which is trained to learn the combined data and conditional training data spaces such that we can sample from this and augment the training data sets.

By building on the the findings of chapter 5, we wish to perform hierarchical inference of the NS equation of state. We firstly reintroduce hierarchical Bayesian inference in section 6.1 and set up the Flow-based probability function for combining information from multiple GW events via MCMC in section 6.1.1. We demonstrate the performance of the pipeline in section 6.2 in combining information from the two currently known real BNS merger events before introducing the simulated GW event data. We discuss previous use of this event data in repeating methods introduced in chapter 4 to infer the NS equation of state from a single simulated events, and then discuss expanding this to multiple events. We present the results of combining information from multiple events with the previously affirmed methods in section 6.3.2 and discuss a different method in section 6.3.3. We present results for multiple simulated equations of state next, before discussing computational requirements and time taken to perform the analysis in section 6.4.

6.1 A recap of hierarchical Bayesian inference

In chapter 5 section 5.3, we introduced a Bayesian framework for combining information from multiple observations of GWs from BNS merger events into one overall inference of the NS equation of state. Here, we briefly recap the key points. The equation of state data space $\mathbf{\Omega}$ - discussed at length throughout chapter 5 - is represented by a compressed 7−dimensional data space, consisting of an autoencoded representation of the equations of state and their respective scaling parameters. Each equation of state is associated to a BNS merger event, which is represented by parameters $\theta = [m_1, m_2, \Lambda_1, \Lambda_2]$. We combine information from multiple observed BNS merger events to infer the true equation of state through hierarchical Bayesian inference.

Through manipulation of Bayes theorem, covered fully in chapter 5 section 5.3, we arrive at the posterior probability density of $\mathbf{\Omega}$ given observed GW strain data h

$$p(\mathbf{\Omega}|h) \propto p_\pi(\mathbf{\Omega})^{1-N} \prod_{i=1}^{N} \left\langle p_{\text{cond}}(\mathbf{\Omega}|\theta_i) \right\rangle_{p(\theta_i \sim p(\theta|h_i))}. \tag{6.1}$$

where $p_\pi(\mathbf{\Omega})$ is the equation of state prior probability density. It is through this framework that we can use our trained Normalising Flow to calculate the posterior probability density $p_{\text{cond}}(\mathbf{\Omega}|\theta_i)$ of a given equation of state for a given set of posterior samples from the i^{th} event, and then combine the information from multiple events. Equation 5.10 in chapter 5 section 5.3 describes the Normalising Flow which is trained to do so, mapping the equation of state data space $\mathbf{\Omega}$ to Gaussian latent space $\mathcal{Z}$ conditioned on BNS merger event parameters θ for rapid probability estimation. A Normalising Flow can evaluate $\mathcal{O}(10^5)$ probabilities in < 1 second and so we can evaluate many iterations of $p(\mathbf{\Omega}|\theta_i)$ at once very quickly and therefore explore the posterior probability density efficiently.

6.1.1 Summary of method

To perform the analysis as outlined above, and as discussed in chapter 5 section 5.3, we wish to propose possible equations of state $\mathbf{\Omega}$ to the Normalising Flow and evaluate the probability density given conditions; posterior samples in $[m_1, m_2, \Lambda_1, \Lambda_2]$ as measured with GWs from BNS merger events. The proposal of equations of state to the Normalising Flow for probability evaluation is performed using MCMC which, given access to a probability function can propose equations of state from a user-defined proposal distribution – in this case the prior $p(\mathbf{\Omega})$ – and can then traverse and survey this space with the provided probability function and propose new points in this space.

In chapter 1 section 1.5.1.1, we introduced MCMC as a stochastic method of sampling a probability distribution. We use `emcee` [134], a particular python implementation of MCMC, which performs ensemble sampling of the equation of state posterior probability density. A series of *walkers* evaluate the probability of multiple locations of the equation of state in parallel relative to a probability function. In this

case, the probability function is finding the posterior probability density $p_{\text{cond}}(\mathbf{\Omega}|\theta)$. We run MCMC for a defined number of steps through the equation of state data space, after an initial burn-in stage. By evaluating the probability of equations of state $\mathbf{\Omega}$ conditioned on θ (posterior samples from multiple individual events) – as defined in equation 6.1 – using the Normalising Flow on a GPU, individual posterior probability evaluations are quick. While each Markov Chain is inherently sequential, we evaluate the probability of $\mathbf{\Omega}$ given a whole set of posterior samples θ from a given event and find the average, before repeating for multiple events. We can use the GPU effectively so to perform as many of these evaluations and calculations in parallel, before then moving to the next stage of the Markov Chain. The ensemble sampling of `emcee` allows for parallelisation of all walkers' probability evaluations. As such the overall sampling routine is cost-efficient.

The workflow we follow for this work is presented in figure 6.1.1. The conditional Normalising Flow $p_{\text{cond}}(\mathbf{\Omega}|\theta)$ is trained as is described in chapter 5 section 5.5. To recap, we train the conditional Normalising Flow using the 'regeneration Flow' $p_{\text{RG}}(\mathbf{\Omega}, \theta)$, an 11−dimensional unconditional Flow trained to learn the joint equation of state $\mathbf{\Omega}$ and event θ parameter space. Once trained, the regeneration Flow can be sampled from via $\mathbf{\Omega}, \theta \sim p_{\text{RG}}(\mathbf{\Omega}, \theta)$ where $\theta = m_1, m_2, \Lambda_1, \Lambda_2$ and the equation of state data $\mathbf{\Omega}$ detached from the event data so to train the conditional Normalising Flow $p_{\text{cond}}(\mathbf{\Omega}|m_1, m_2, \Lambda_1, \Lambda_2)$. In chapter 5 section 5.5, we introduced how the introduction of the regeneration Flow reduced the effect of artificial structure in the Flow's learned data space. This allows us to propose new equations of state and evaluate their probability $p(\mathbf{\Omega}|\theta)$ without falling into gaps within the learned structure. This phenomenon is discussed extensively in chapter 5.

Our evaluation of the posterior probability $p(\mathbf{\Omega}|h)$ requires consideration of the equation of state prior probability, $p(\mathbf{\Omega})$, which we similarly model using a Normalising Flow, $p_\pi(\mathbf{\Omega})$ as introduced in chapter 5 section 5.4.3 equation 5.12. Once trained, samples can be drawn from the Flow $\mathbf{\Omega} \sim p_\pi(\mathbf{\Omega})$ accompanied by their prior probability densities. This unconditional Flow is likewise trained using the regeneration Flow $p_{\text{RG}}(\mathbf{\Omega}, \theta)$ in order for the training data prior for both the unconditional prior Flow

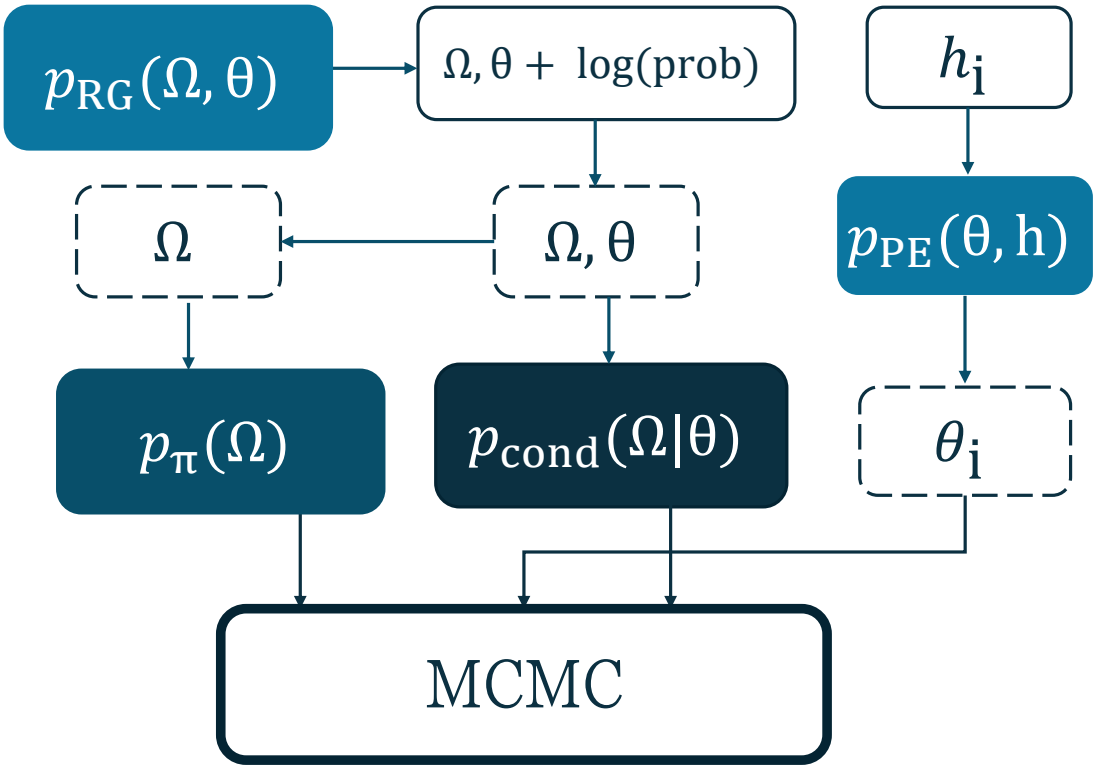


Figure 6.1.1: Block diagram of workflow of the original Flow-based method for hierarchical analysis. This includes the use of regeneration, conditional, prior and PE Flows. Samples θ_{i} are generated from the PE Flow for use in MCMC while the conditional and prior Flows are used in MCMC to evaluate probability of proposed $\mathbf{\Omega}$ given θ_{i}.

and the conditional Flow to remain consistent. With regards to the prior Flow $p_{\pi}(\mathbf{\Omega})$, training with the (first 7 dimensions of the) regeneration Flow incorporates a small amount of noise into $p_{\pi}(\mathbf{\Omega})$ without fundamentally changing the behaviour of the data space which is it learning the mapping of, while additionally providing more training examples of $\mathbf{\Omega}$ to learn from. Chapter 5 found that the use of the regeneration Flow $p_{\mathrm{RG}}(\mathbf{\Omega}, \theta)$ in training both $p_{\pi}(\mathbf{\Omega})$ and $p_{\mathrm{cond}}(\mathbf{\Omega}|\theta)$ also solved issues surrounding artificial structure so we use the regeneration Flow to more safely model the prior, as well as to train $p_{\mathrm{cond}}(\mathbf{\Omega}|\theta)$. Using a Normalising Flow to represent our prior probability density allows for very quick sampling of the prior itself but also for very fast evaluation of the prior probability of randomly proposed points, in this case equations of state $p_{\pi}(\mathbf{\Omega})$, which we have no alternative analytic way to evaluate. We

perform the analysis by taking data from observations of GW events from N BNS mergers in $[m_1, m_2, \Lambda_1, \Lambda_2]$, evaluating the probability of equations of state given data from these N events in a MCMC to return posterior samples from the posterior probability density in the 7−dimensional compressed equation of state space.

6.2 Applications to real events

We firstly apply this analysis to combine information from the real BNS merger events GW170817 [54] and GW190425 [95]. As discussed in chapter 1 section 1.3, these two events are currently the only two observations of GWs from BNS mergers. These events had vastly different SNRs from one another and therefore different precision in the measurement of mass and tidal deformability of each of the component NSs. We introduce GW190425 to this analysis to demonstrate hierarchical inference using Normalising Flows, and compare to the previous single event study of GW170817 made in chapter 4 presented in figure 4.5.1.

6.2.1 Data

We randomly select 256 samples from the full set of posterior samples from Third Gravitational-Wave Transient Catalog (GWTC-3) and Gravitational-Wave Transient Catalog from the First Half of the Third Observing Run (GWTC-2.1) for the GW events GW170817 and GW190425, respectively. We choose to use 256 posterior samples per event as this is the minimal number of samples we need so as to represent the posterior without biasing the result with too little samples but also not too many as to become a computational burden. Appendix B presents the plots in justification of using 256 samples. The posterior samples for both events that are within the convex hull – defined by the boundary of the conditional training data θ – are presented in figure 6.2.1 alongside training data θ. In chapter 1 section 1.4.1 we discussed how GW190425 had a much smaller SNR of 12.4 compared to that of GW170817 with 32.4, the loudest observed event by the global ground-based detector network for a long time before being surpassed in the fourth observing run [39]. This is apparent in figure 6.2.1 which shows GW190425 with a broader event posterior probability

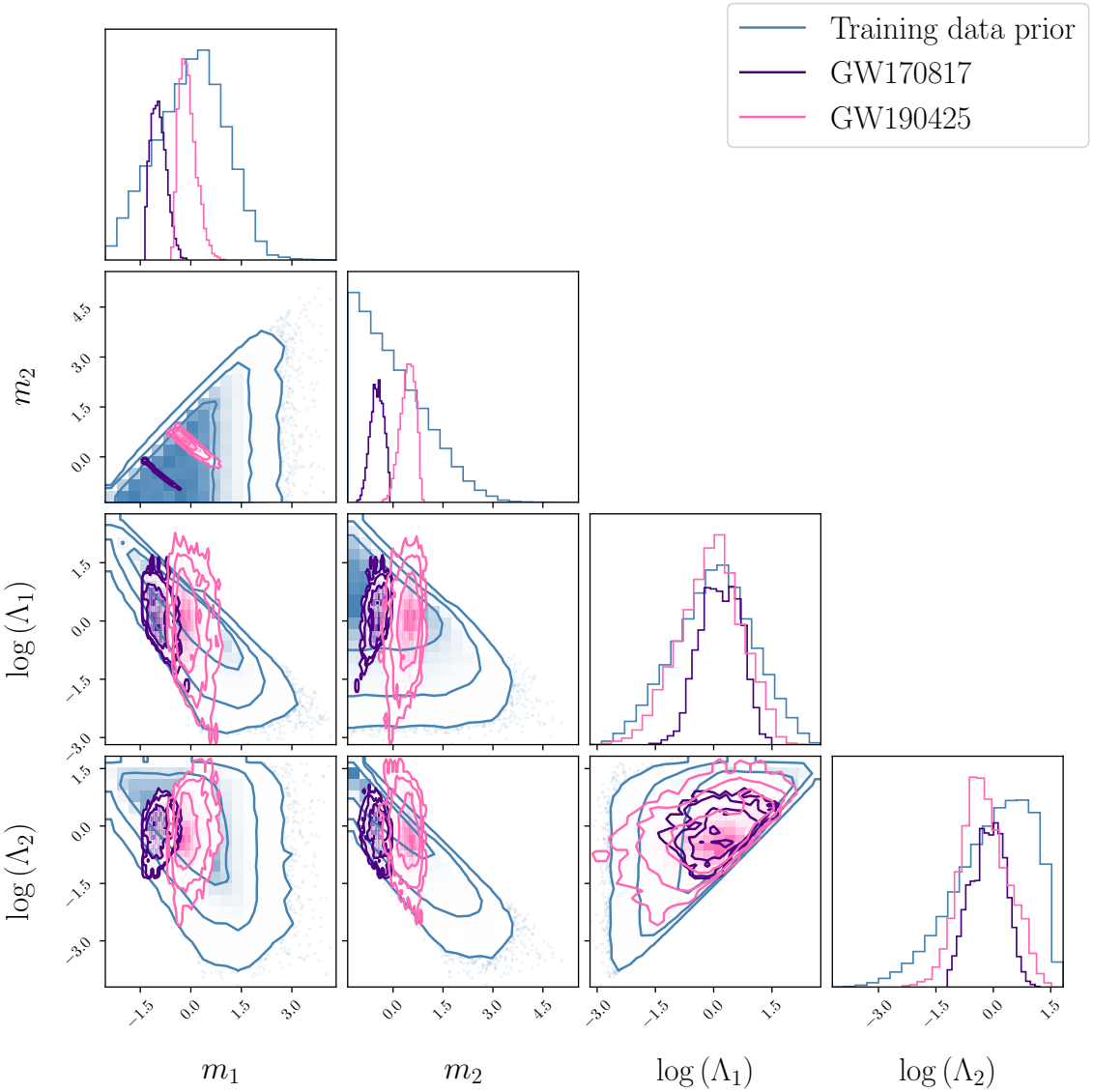


Figure 6.2.1: 1- and 2-dimensional marginalised distributions of posterior samples of $[m_1, m_2, \log \Lambda_1, \log \Lambda_2]$ from GW events GW170817 (purple) and GW190425 (pink) on the training data prior (blue). The contours represent 68%, 95%, and 99.7% of the samples respectively.

distribution than that of GW170817, especially in the spread of the outermost contour of the joint distribution of $\log \Lambda_1$ and $\log \Lambda_2$, which defines for 99.97% of all posterior samples. Therefore, we expect a reduced level of constraint in the equation of state space from single event analysis with GW190425 than we do for that of GW170817. We pass 256 posterior samples from each event in $[m_1, m_2, \log \Lambda_1, \log \Lambda_2]$ from both GW170817 and GW190425 to the hierarchical analysis. We run MCMC for 5×10^3 steps with 200 walkers with a 10^3 step burn-in stage. This analysis took just over 2 hours to combine the information from the two events into a combined 7−dimensional equation of state posterior using the batching method described in section 6.1.1 which efficiently packs data onto the GPU for processing.

6.2.2 Results

We plot the result of the combined analysis in figure 6.2.2 alongside the individual event posteriors for both GW170817 and GW190425. These were found following the method applied in chapter 4 of using a trained conditional Normalising Flow to sample the individual events' equation of state posterior given posterior samples from a single BNS merger event. The purple and pink distributions therefore represent $\mathbf{\Omega}_{170817} \sim p(\mathbf{\Omega}|\theta_{\text{GW170817}})$ and $\mathbf{\Omega}_{190425} \sim p(\mathbf{\Omega}|\theta_{\text{GW190425}})$ respectively. As expected, given the wider spread of posterior samples θ due to the much smaller SNR of the event, the single event equation of state posterior associated to GW190425 is much broader than that of GW170817, and in some cases leaks beyond the edges of the prior. The single event posteriors presented in figure 6.2.2 have not gone through any post-processing stages and so some samples lie beyond the training data prior. This is a manifestation of the intrinsic uncertainty of the Flow, which has, as previously mentioned, been training using the regeneration Flow. These multiple stages that construct the Flow introduces uncertainty which do not fundamentally change what the Flow is learning but does introduce some noise.

In contrast, the result of the MCMC analysis in blue represents posterior samples from $p(\mathbf{\Omega}|h_{\text{GW170817}}, h_{\text{GW190425}})$. The combination of information from both GW events returns a smaller, shared volume of the purple and pink distributions as appears

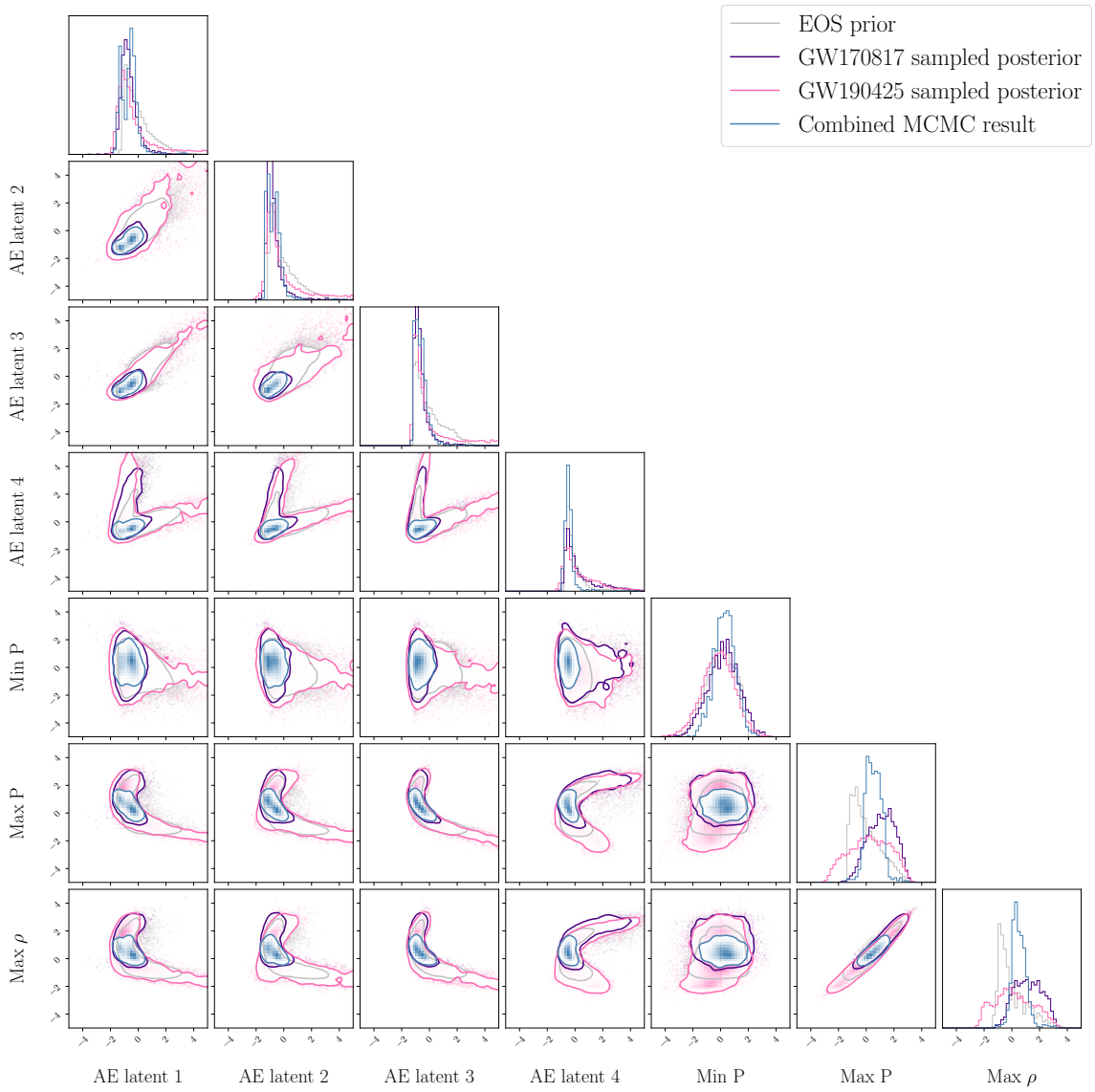


Figure 6.2.2: 1- and 2-dimensional marginalised distributions of posterior samples in the compressed equation of state space given information from both GW170817 and GW190425 (blue) alongside their single event posteriors (purple and pink, respectively) alongside the training data prior (grey). The contour for each figure represents 90% of the respective total number of samples.

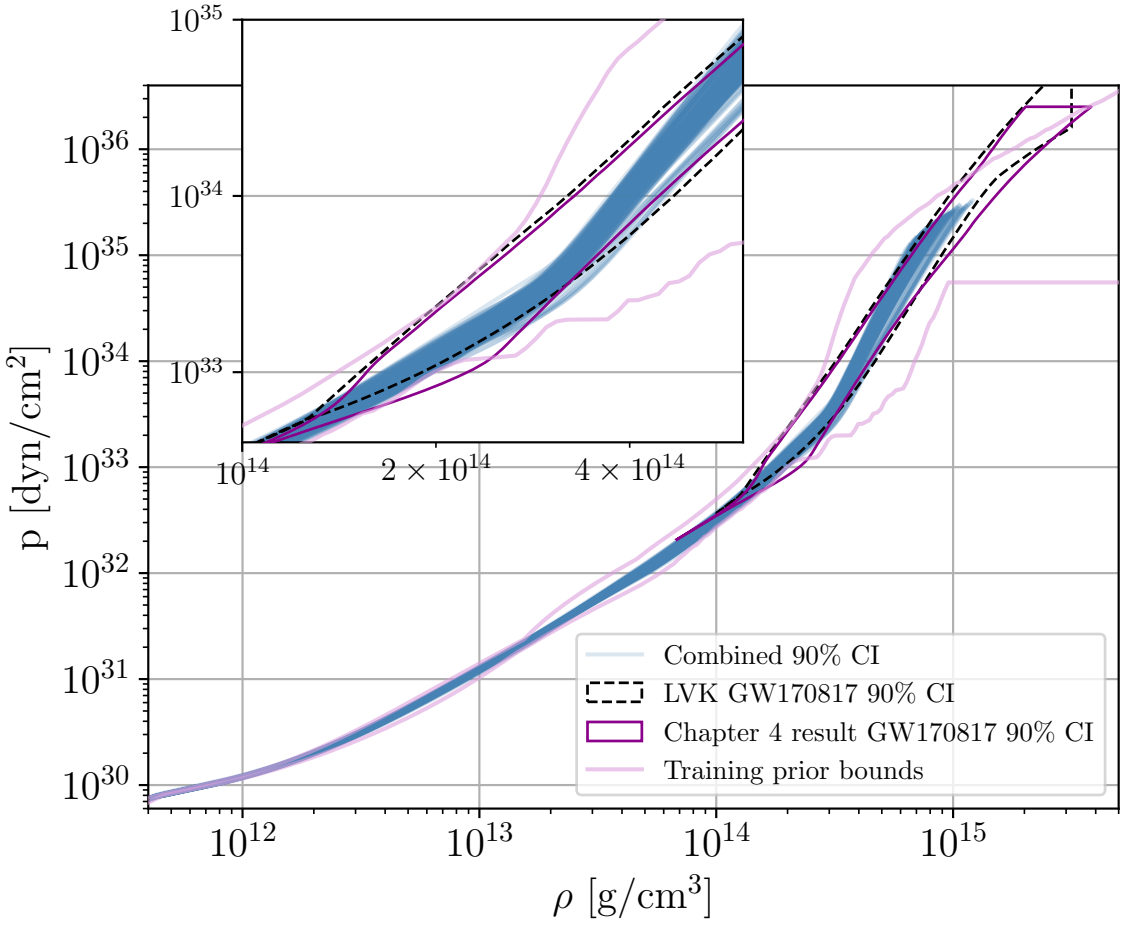


Figure 6.2.3: Plot of pressure vs. density of a projection of the combined equation of state posterior given GW170817 and GW190425 in blue. The single event GW170817 result is presented as a purple boundary. The LIGO-Virgo-KAGRA (LVK) result from GW170817 is presented in the black dashed boundary. The training data prior for the combined analysis is given by the pink prior bounds. The inset displays the zoomed-in posterior in the region of $[1, 5] \times 10^{14}$ g cm^{-3}.

in the 1− and 2−dimensional projections in figure 6.2.2. This demonstrates that the walkers are finding regions of the 7−dimensional space which is commonly favoured by both GW170817 and GW190425 given evaluation of the probability of these regions using the conditional Normalising Flow.

We also express the result of combining information from these two events in figure 6.2.3 in the $P - \rho$ plane. We compare the combined GW170817 and GW190425 result to that of chapter 4 figure 4.5.1 and, likewise in the chapter 4 result, the equation of state inference performed by the LVK collaboration in [4]. Each of these previously declared results express their posterior in the $P - \rho$ plane in intervals of

50% and 90% confidence, however their respective choices of presentation differs from the current result. Both the LVK and chapter 4 results determine the regions of confidence in the $P - \rho$ plane, with the LVK result slicing in pressure and chapter 4 slicing in density. Only the 90% boundary for both of the aforementioned results are presented in figure 6.2.3. For the result of this Flow-based combined analysis, the 90% confidence equations of state in blue are determined from their 7−dimensional $\mathbf{\Omega}$ representation. To do so, a KDE is constructed around the 7−dimensional equation of state posterior, from which the samples within the 90% confidence interval are determined. These samples in the 7−dimensional equation of state space are then decoded and the results are presented in figure 6.2.3 in blue. The workflow discussed in this chapter for the combined analysis (and introduced in chapter 5 section 5.2) allows for each equation of state to be presented to its own maximum density and pressure. Therefore, the upper density region of the equation of state posterior differs from the LVK and chapter 4 results, which restrict each equation of state in reaching a common maximum density or pressure, respectively. The workflow discussed here allows for a more flexible equation of state posterior, and does not restrict the equation of state within a common band.

We find that the hierarchical inference result given both GW170817 and GW190425 constrains more in $P - \rho$ when compared to the single event results from chapter 4 and that of the LVK collaboration in [4]. We find the addition of GW190425 in the analysis is particularly powerful in the density range of $[1, 4] \times 10^{14}$ g cm^{-3}. The component masses of GW190425 are both more massive than that of GW170817 (see figure 6.2.1) with comparable tidal deformabilities. For a given tidal deformability, a higher mass suggests a stiffer equation of state, which is reflected in figure 6.2.3 where the equations of state from the combine result appear to have a steeper gradient than the bounds on the chapter 4 result suggest. Additionally, there is less support at high densities and pressures compared to the GW170817 result from chapter 4 and that of the LVK. We also find the constraint extends below the lower density limit of both the chapter 4 and LVK results, with equation of state constraints within the prior bounds from 2×10^{13} g cm^{-3} onwards. We note the change in prior on both the conditional training data θ between the three different analyses and the

equation of state $\mathbf{\Omega}$. This accounts for the difference in upper density/pressure limit, and the shape that the equations of state take in this region. For the LVK result presented in figure 6.2.3 from [4], the prior on equation of state is defined by their choice of equation of state parameterisation. As introduced in chapter 1 section 1.4.1 and chapter 2 section 2.2.1, [4] use the spectral parameterisation (introduced in 2 section 2.1) to represent their equations of state, and choose a prior range over the spectral parameters so to represent a broad range of physically valid equations of state [4]. The analysis here differs from that of chapter 4 purely in the training data used, as both analyses are non-parametric and model-agnostic. But we note both analyses also use different compression methods so the compressed data spaces which inference is performed in is also different. This limit was a particular choice made by each of the analyses, and is not simple to account for so to make the analyses compatible for comparison purposes. The prior on the analysis performed here in the combination of both events, however, is broader and extends a larger density range than the aforementioned analyses. While the shape of the upper pressure limit varies and is softened slightly above 10^{15} g cm^{-3}, we believe the change in training data is beneficial to improving the result for third-generation analyses.

6.3 Simulation studies

In addition to the combination of information from the two current BNS merger events, we discuss the performance of the analysis in the combination of many more simulated events. We wish to perform simulation studies of hierarchical inference of three different equations of state of varying softness/stiffness. We do so by simulating a multitude of GW events associated to this underlying equation of state and performing the hierarchical analysis with PE samples associated to the simulated events. We wish to test the analysis on inferring three different equations of state from the validation data set, presented in figure 6.3.1.

The chosen equations of state are presented in the $P - \rho$ plane in figure 6.3.1. We demonstrate three randomly selected equations of state which feature 3 distinct examples of possible behaviour. The stiffest equation of state in the density range

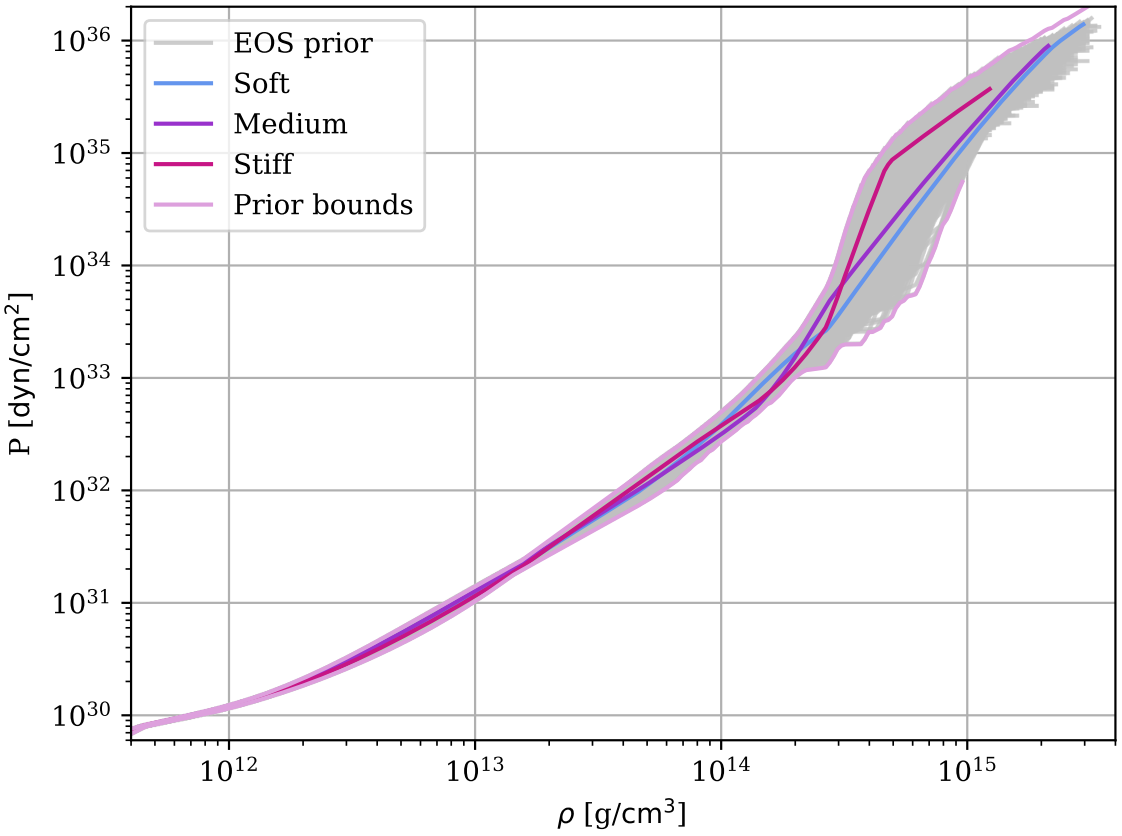


Figure 6.3.1: Plot of pressure vs. density for three simulated equations of state from the validation data set alongside the prior training data in grey and the bounds on the prior training data in pink.

2×10^{14} g cm^{-3} to 4×10^{14} g cm^{-3} is given by the pink-red line, with the steepest gradient and therefore the larger increase in pressure across this density. In contrast, at the lower end of this density range, the soft equation of state, in blue, has the the smallest change in pressure for a given change in density out of the three equations of state highlighted. The behaviour of each equation of state evolves beyond these definitions out of this density region; the training data chosen for this analysis allows for the equations of state to be expressive over a broad density range and to have unique maximum pressures and densities. The 'medium' equation of state merely gives a third example, which is neither extremely soft or stiff. The equations of state given in figure 6.3.1 are equivalently plotted in the $m - \Lambda$ plane in figure 6.3.2. In this figure, we instead plot the individual events which are simulated so to be associated to each of the stiff, soft or 'medium' equations of state. As discussed in chapter 1, each equation of state in $P - \rho$ has an equivalent relationship in $m - \Lambda$, where Λ evolves

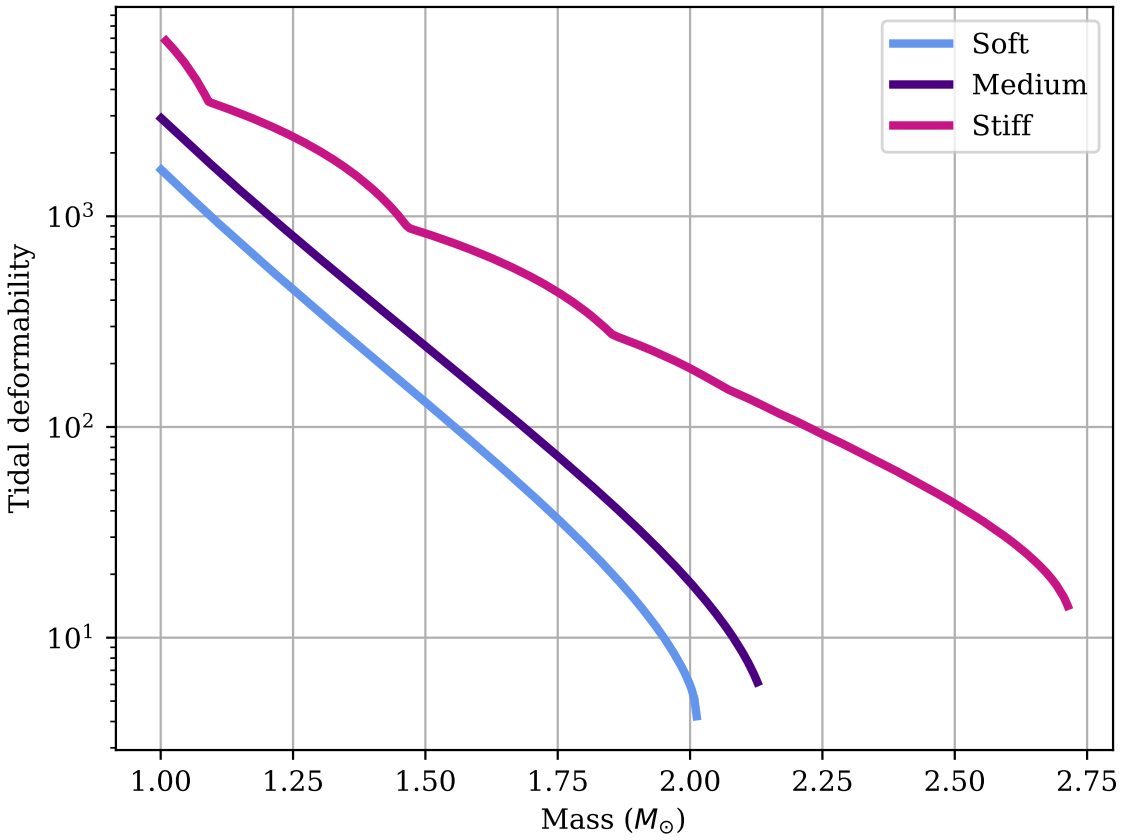


Figure 6.3.2: Mass vs. tidal deformability for three example equations of state from the validation data set. Dots along each equation of state represent a single NS with the darker shade of a given colour of dot associated to the first component star of the merger and the lighter shade being the second component star of the merger.

with m as $\Lambda \sim C^{-5}$, where $C = M/R$ is the compactness, a measure of the amount of mass M per unit radius R. It is clear from this figure that the stiff equation of state has more support for high mass NSs and supports higher tidal deformability at a given mass. This is because for stiffer equations of state, there is more pressure support against the gravitational force, and so NSs have a larger radius. As radius R increases, the tidal deformability therefore increases dramatically, due to the fifth power on radius. A larger tidal deformability means the star is easier to distort via tidal forces. Alternatively, the softer the equation of state, the smaller the tidal deformability. Soft equations of state are associated to less massive NSs which are more compacted and therefore have a smaller radius. Thus, given our previous argument for stiff stars, soft equations of state support smaller tidal deformability for a given mass indicating they are less easily deformed.

6.3.1 Parameter estimation data

For the simulation studies introduced in section 6.1.1, we require a method to simulate PE for multiple observations of GWs from BNS merger events. In order to test the success of our pipeline, we wish for each of these events to be associated to a true underlying equation of state that we can infer through combining information from multiple BNS merger events. Methods to simulate PE of mock GW merger events already exist for the current era of GW observation. We previously introduced the PE software `bilby` [295] which, for simulation purposes, allows the user to specify true mass and tidal deformability parameters for a given BNS merger event which may be associated to any equation of state. Once these parameters are defined, a waveform is produced relative to the defined parameters, and has noise added associated to a given PSD (introduced in chapter 1 section 1.2.3). With the noise-added waveform, one can then perform PE as normal, via stochastic sampling methods, as is discussed in chapter 1 section 1.5.1.

However, simulation studies for the third generation of GW detectors must take into consideration the change in how long a BNS signal will be observable for in the detectors sensitive frequency range; for ET at design sensitivity, a BNS merger signal may be in-band and observable for multiple hours [302]. PE of this signal with current stochastic sampling methods would potentially take thousands of central processing unit (CPU) hours [200]; computational cost increases alongside increasing SNR and signal length. This is a computational burden which will become prohibitive as numbers of observations increase. Alongside the increased cost of PE comes the subsequent equation of state inference, which adds additional expense and time requirement (see chapter 2 section 2.2.1.2 for discussion), all of which takes place before the information from multiple events is combined together in hierarchical inference.

Hu et al. [3] present a ML-based workflow of performing PE for 3G-era BNS signals and subsequent inference of the equation of state together within $\mathcal{O}(1$ second). With a method similar to that of [258] but applied to long-duration BNS mergers in

particular, Hu et al. use a conditional Normalising Flow, $p_{\rm PE}(\theta|h)$ to perform full PE on long-duration precessing BNS merger signals detected by 3G detectors and subsequently infer the equation of state following the work of chapter 4.

Hu et al. train a conditional Normalising Flow, $p_{\rm PE}(\theta|h)$ following the notation of chapter 3 section 3.2, on a selected region of BNS parameter space, θ which will be observable with 3G detectors. They find it difficult to train a Flow which covers the entire BNS signal parameter space, so instead divide the parameter space into the low SNR regime which covers the range $[20, 50]$ and the high SNR regime with SNRs in the range $[200, 500]$ and train a different Flow model for each. We consider only the low SNR model for our work here as this covers the majority of observable BNS mergers with 3G detectors [3]. To train the conditional Normalising Flow $p_{\rm PE}(\theta|h)$, the authors first generate training data by sampling chirp mass, $\mathcal{M}$ uniformly in the range $[2, 2.1]$ which corresponds to GW signals which are in the low SNR regime. Additionally $\tilde{\Lambda}$ is sampled uniformly in the range $[0, 1600]$ and $\delta\tilde{\Lambda}$ is determined by component masses and $\tilde{\Lambda}$. The authors note the relatively tight range of $\tilde{\Lambda}$ used for training; the upper limit is chosen to reflect the findings from GW170817 [54] but note that a wider prior in $\tilde{\Lambda}$ would benefit for application of the model to lower mass systems. They note this as a potential avenue of improvement for the model but use the current upper limit as a proof-of-concept. Other parameters required to construct the waveform are sampled relative to common choices [3].

The prior on parameters of interest θ is consistent with the training data prior on θ for the equation of state conditional Flow, $p_{\rm cond}(\mathbf{\Omega}|\theta)$. The specifics of the prior on the conditional term θ for this Flow are pertinent to the training process only. Beyond this, we only look to ensure that any conditional terms lie within this prior when sampling or evaluating using the equation of state Flow, which is the case for the θ samples generated using the PE Flow. We enforce this by applying the convex hull defined by the training data prior boundary on θ to ensure event samples belong within the bounds of the θ training prior (introduced in chapter 4 4.3.1). Figure 6.3.3 presents an example set of posterior samples in θ generated using the PE Flow $p_{\rm PE}(\theta|h)$. We demonstrate that the example event posterior samples, while broad

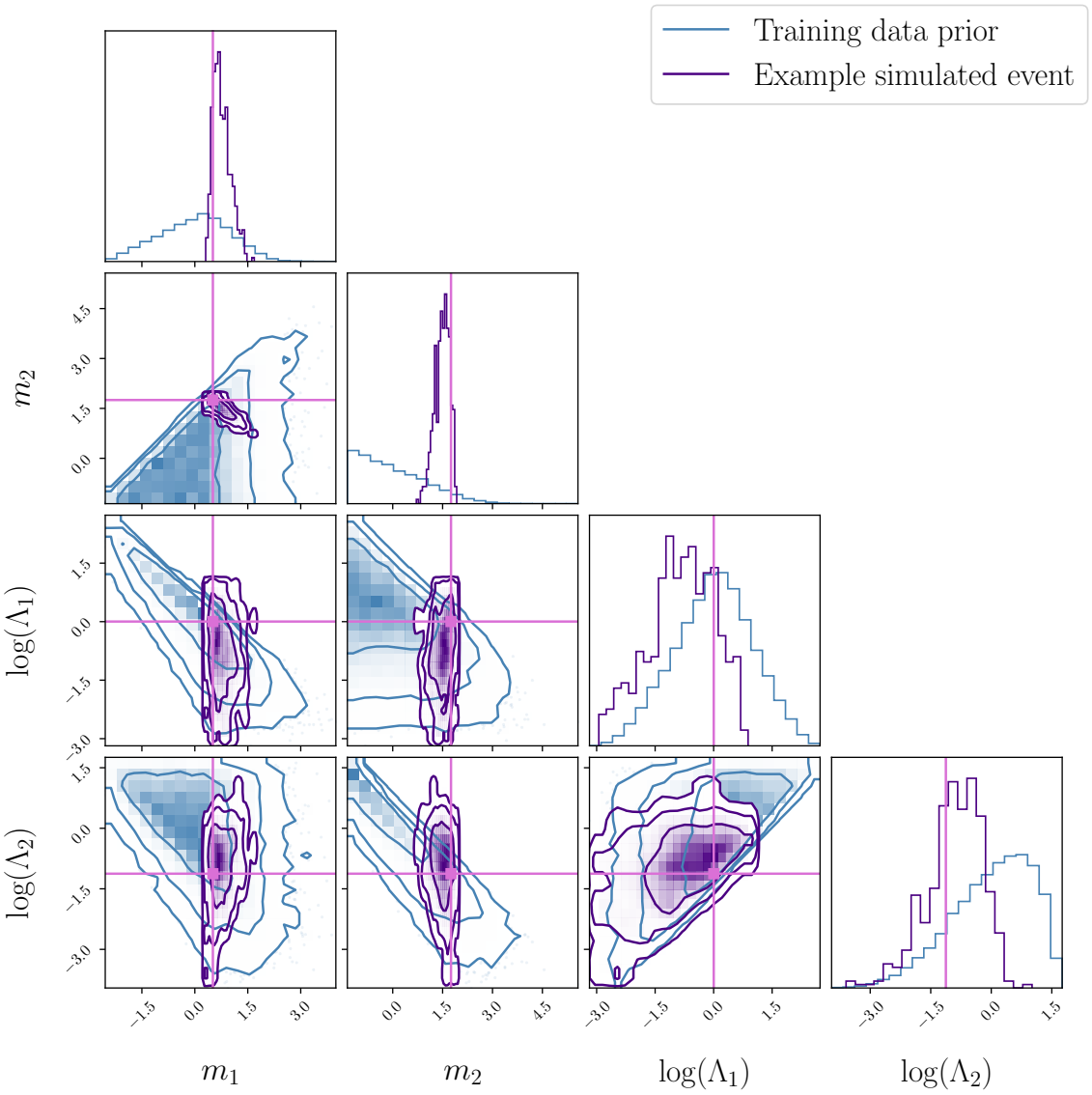


Figure 6.3.3: 1- and 2-dimensional marginalised distributions of posterior samples in $(m_1, m_2, \log\Lambda_1, \log\Lambda_2)$ from an example simulated ET-observed GW event on the training data prior (blue) with the true parameters of the event highlighted in pink. The contours represent 68%, 95%, and 99.7% of the samples respectively.

over the prior range in the 2−dimensional projections, are within the bounds of the training data prior and obey the harsh bounds on the space, particularly in upholding conventions of $m_1 \geq m_2$ and $\Lambda_1 \leq \Lambda_2$.

Once parameters are sampled, a gravitational waveform can be generated. However, long-duration BNS waveforms, observable from frequencies as low as 5 Hz until merger, sampled at standard sampling rates used in current PE analyses may have up to 10s of millions of data points. Therefore, compression techniques are necessary to

compress the waveform to something more suitable for the Flow analysis. The authors employ a multi-banding (the process of adapting the sampling rate depending on the frequency interval of the waveform you are sampling [303, 304]), singular value decomposition (SVD) (analogous to PCA, a compression of the waveform onto linear bases) and finally two neural network architectures to compress the long-duration GW signal observed in multiple channels of an arrangement with both CE and ET into a single vector of length 128. The authors note that plenty of training data is required to avoid overfitting, likely due to the high dimensionality of the problem (17−dimensional data space and 128−dimensional conditional space). Training takes $\mathcal{O}$(weeks) on a NVIDIA A100 GPU.

In inference of real events in the third generation of GW detection, one can then pass the observed long-duration BNS waveform to the Flow as a condition and return posterior samples from the Flow's data space. The authors additionally employ importance sampling [305] to improve confidence in the result by resampling the network output to cover for potential network inaccuracy [306]. The samples from the Flow output are the proposal samples for importance sampling, which are then reweighted based on the ratio of the 'true' analytical and Flow-determined probabilities. An example set of posterior samples are presented in figure 6.3.3. The posterior samples θ for a third-generation simulated event are comparable in volume to the posterior of the event GW170817 (see figure 6.2.1), which was an exceptionally high SNR event. Relative to current-generation posterior samples (again, relative to figure 6.2.1), we find the posterior deviates from the prior in Λ_1 and Λ_2, indicating more information being present in the tidal measurements.

6.3.1.1 Validating with single event equation of state inference

We firstly demonstrate the equation of state inference quality and discuss the computational cost of a single event simulation study performed in [3]. For this work, we firstly simulate BNS mergers associated to a true equation of state using the method outlined in chapter 4 section 4.5.3. With a given event's parameters, we use the pipeline described here to generate mock PE associated to the event, by firstly

generating the BNS merger waveform, adding noise, and then performing PE using the conditional Flow for PE $p_{\mathrm{PE}}(\theta|h)$. All other parameters required to build the gravitational waveform are not defined by the equation of state. Instead, these are sampled uniformly from bounds discussed in [3] so to be within the bounds of what the PE Flow $p_{\mathrm{PE}}(\theta|h)$ has been trained on. The process of generating $\mathcal{O}(1000)$ PE samples for a simulated event takes $\mathcal{O}(1$ second).

In order to infer the equation of state of the simulated event, PE samples are passed to another conditional Normalising Flow $p_{\mathrm{cond.}}(\Omega|\theta)$ trained as is outlined in chapter 4. To recap, a conditional Normalising Flow is trained to perform the mapping from compressed equation of state data space conditioned on parameters $\theta = m_1, m_2, \Lambda_1, \Lambda_2$ of a BNS event associated to each equation of state. In inference, one can then conditionally sample the $p_{\mathrm{cond.}}(\Omega|\theta)$ with PE samples θ from a given BNS merger event and return samples from the equation of state posterior $\Omega \sim p_{\mathrm{cond.}}(\Omega|\theta)$. In Hu et al. [3], the conditional Normalising Flow for equation of state inference $p_{\mathrm{cond.}}(\Omega|\theta)$ is trained as is outlined in chapter 5 section 5.3.1 with equation of state training data which was introduced in section 5.1.

Two BNS events were simulated associated to a single equation of state from the validation data set and GW signals were simulated for these events with SNRs of 39 and 390, and using the low and high SNR PE Flow models $p_{\mathrm{PE}}(\theta|h)$ respectively, PE samples are produced. The PE samples for each event are then passed to the equation of state Flow $p_{\mathrm{cond.}}(\Omega|\theta)$ for equation of state inference, and the results are presented in 6.3.4. The equation of state posterior for both a low and high SNR event are presented in light and dark blue bands, respectively. While both event equation of state posteriors agree with the true equation of state, the high SNR event is more constraining, as one would expect. Both posteriors additionally agree on the general trend of the equation of state, with the constraint in increasing SNR tightening in both low and high density, remaining consistent with the true equation of state. While increasing the SNR of a single BNS merger event, does improve the level of constraint in $P-\rho$, the result does highlight that hierarchical analyses which combine information from multiple BNS merger events are necessary to truly constrain the NS equation of

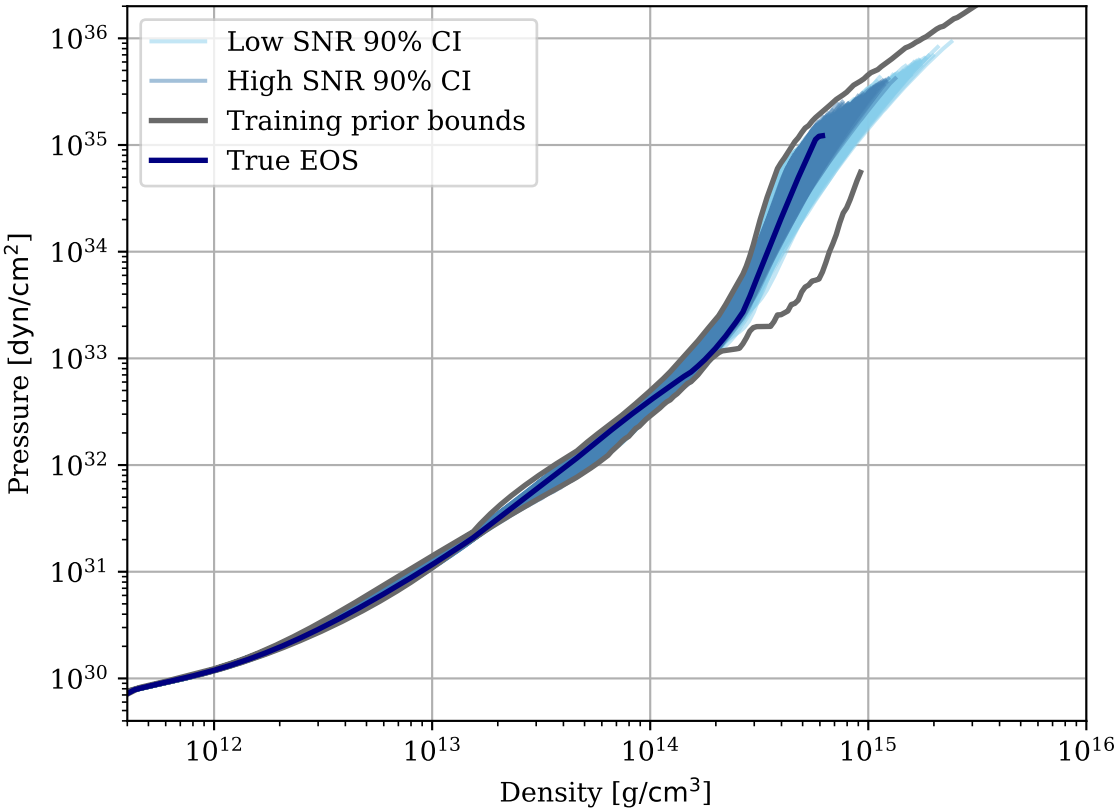


Figure 6.3.4: Plot of P vs ρ for the equation of state posterior associated to two simulated BNS merger events to 90% confidence. The confidence intervals are defined in the 7−dimensional compressed equation of state data and projected to $P - \rho$. The light blue band corresponds to the SNR 39 measurement, while the dark blue corresponds to the SNR 390 measurement. The true equation of state is given by the bold blue line and the prior boundaries are presented in grey.

state.

6.3.2 Results of hierarchical inference

We test the validity of the analysis firstly on the 'medium' simulated equation of state so to introduce the stages of the analysis and demonstrate it's effectiveness in section 6.3.2.1. We discuss the results of the hierarchical analysis through the 'original' method, introduced in section 6.1.1, and discuss a new avenue through which to perform the analysis in section 6.3.3, before demonstrating the agreement of the two methods in section 6.3.3.1. We additionally discuss the performance of the new method on inference of 'stiff' and 'soft' equations of state in sections 6.3.3.2 and 6.3.3.3 respectively.

6.3.2.1 Medium equation of state result

We first perform hierarchical inference using PE samples from up to 128 events, using our 'original' method, which have been generated using the method outlined in section 6.3.1 and are associated to the 'medium' equation of state presented in figure 6.3.1. This equation of state was selected randomly from the validation data set for training the 11−dimensional regeneration Flow. We assign 'medium' as a label as, out of the three equations of state randomly selected, it is neither the softest or the stiffest. Figure 6.3.2 reflects this labelling, as it allows for NSs with tidal deformability up to ~ 3000 and an upper limit on mass of around $\sim 2.12\ M_{\odot}$. These limits on the allowed parameters defined by this equation of state are relatively standard (see chapter 1 section 1.4 for discussion of our current knowledge of limits on NS macroscopic parameters).

Data is collected from all 128 events and 256 PE samples – which belong within the bounds of the convex hull defined by the training data prior on θ – are randomly selected from each event. The reason for using 256 samples per event is justified in section 6.2.1 with further discussion in appendix B. We perform MCMC using information from 4, 32 and 128 events. Walkers are initially proposed as samples from the prior Flow $\mathbf{\Omega}_{\text{prop}} \sim p_{\text{prior}}(\mathbf{\Omega})$. The algorithm is run for a total of 5×10^3 steps after an initial 10^3 burn-in steps. The conditional Flow evaluates the probability of a proposed location $\mathbf{\Omega}_{\text{prop.}}$ given posterior samples θ_i from the i^{th} GW event where $\theta_i \sim p_{\text{PE}}(\theta|h_i)$. The trained prior flow $p_\pi(\mathbf{\Omega})$ (introduced in equation 5.12 of chapter 5 section 5.4.3) evaluates the prior probability of the proposed sample $p_\pi(\mathbf{\Omega}_{\text{prop}})$ and, the posterior probability $p(\mathbf{\Omega}_{\text{prop}}|h)$ is found via equation 6.1.

After running for a total of 5×10^3 steps with 200 walkers, there are 10^6 total samples from the posterior. Standard MCMC algorithms expect between 20% and 50% unique total probabilities [134], however we find that in evaluating the posterior probability $p(\mathbf{\Omega}|h)$ with information from 4 BNS merger events given equation 6.1 and performing MCMC, that we return 9.96% unique samples. As we increase the number of events that we combine information from, the fraction of unique samples decreases

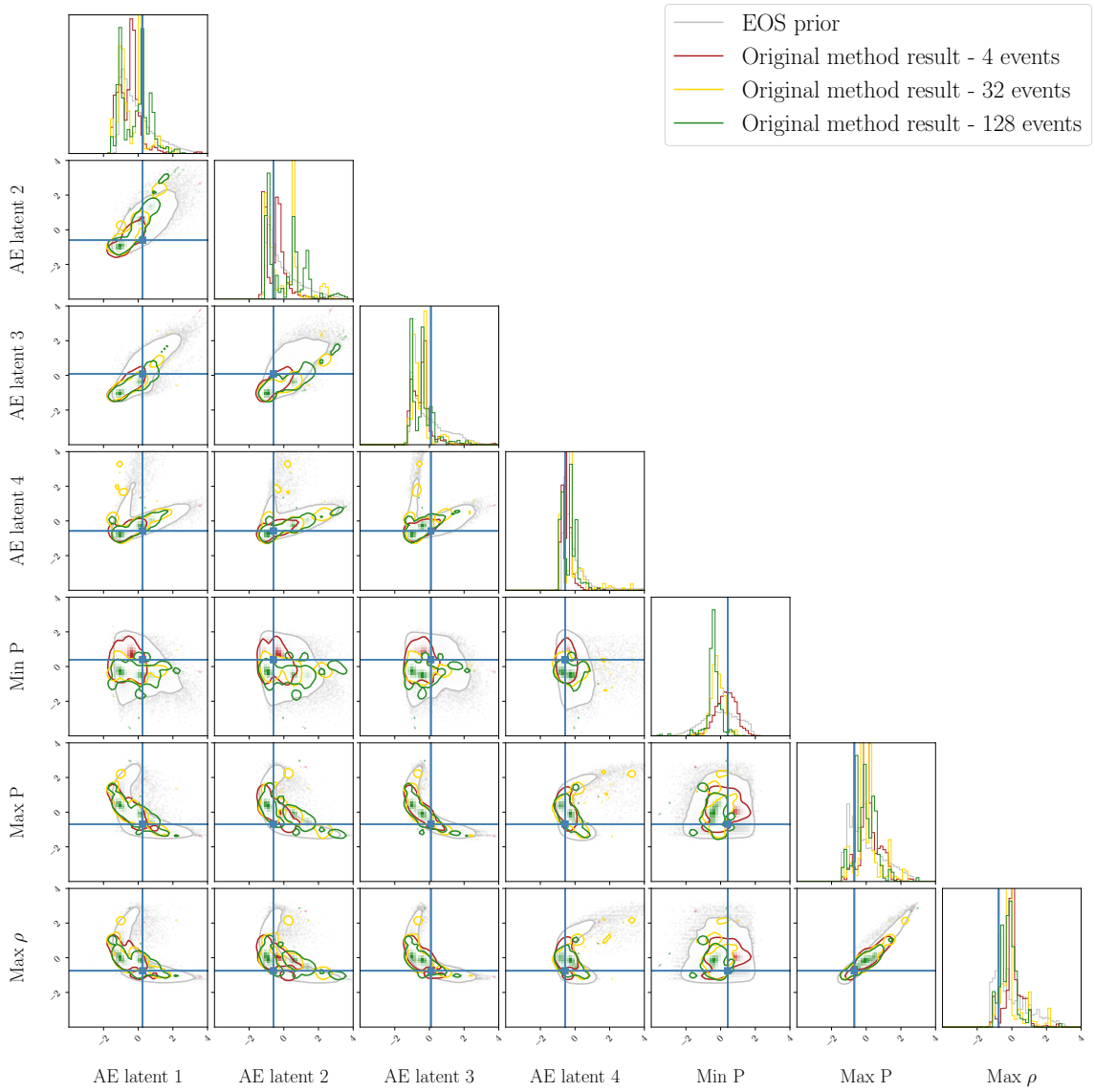


Figure 6.3.5: 1- and 2-dimensional marginalised distributions of posterior samples in the compressed equation of state space given information from 4 (red), 32 (yellow) and 128 (green) GW events associated to a true 'medium' equation of state on the training data prior (grey). The contour represents 90% of the samples.

further and the results from $4, 8, 16, 32, 64$ and 128 events using this (which we refer to as 'original') method are presented in figure 6.3.13. In attempting to combine results from 64 events or more, the algorithm accepts less than 3% of the proposed steps, suggesting the walkers, once proposed from the prior at the beginning, are static for 97% of the total 5×10^3 steps.

We present posterior samples from these MCMC runs in figure 6.3.5. From the total 10^6 posterior samples from all steps of all 200 chains, we select 10 uniformly

spaced samples from the chain of each walker, resulting in 2000 total posterior samples. After a 10^3 step burn-in, we expect the chains to have had the opportunity to search for the bulk of the posterior and now to be uniquely moving around the $\mathbf{\Omega}$ space on a scale consistent with the uncertainty on the posterior. The statistical properties of the walkers should not be changing as we continue to sample the rest of the chain. Therefore, selecting uniformly spaced samples from each of the chains should return independent posterior samples from the posterior probability density. However, with less than 10% unique steps, it suggests the chains are finding it difficult to move to new regions of the parameter space and are preferring to remain static. This suggest there are significant regions of the parameter space which are highly improbable.

Figure 6.3.5 presents the 2000 samples that have been selected from the chains for results from the combination of 4, 32 and 128 events. We find the 2−dimensional 90% contour for the 32 and 128 event results agree on the same region of parameter space in most projections but are more constrained than that of the 4−dimensional result. However, in increasing from 4 events, the posterior for 4 and 32 events tends further from the truth and becomes much more multimodal. Given we also found that the number of unique steps taken by the walkers in the 32 and 128 event cases are 4.60% and 2.73% comparatively, it suggests the posteriors are similar because statistically the runs are performing similarly and not evolving as desired. It's also clear that as we increase the number of events, the equation of state posterior becomes more multimodal, suggests the walkers become associated to peaks of probability in the parameter space and struggle to explore elsewhere, which would be in agreement with the low sample acceptance rate.

Individual event likelihood evaluations with the conditional equation of state Flow $p_{\text{cond}}(\mathbf{\Omega}|\theta)$ with equation 6.1 are weakly informative, as we know that the information provided from an individual or select set of BNS merger events provides support for much of the equation of state parameter space and so we expect to need many events to begin to truly infer the equation of state posterior. This may be suggestive of multiple things, first and foremost being a badly trained conditional Flow $p_{\text{cond}}(\mathbf{\Omega}|\theta)$ which produces noisy posteriors. However, the results of chapter 5 suggest this has

been mitigated. Alternatively, more samples θ per event could improve the calculation of the average in equation 5.9; this was tested by increasing the number of samples beyond 256 up to the maximum number of samples per event, however, there was no change in the statistical behaviour of the walkers. Running MCMC, however, without the removal of any prior information via equation 6.1 improves the statistical behaviour of the run significantly, but is mathematically incorrect given what we wish to achieve. We briefly state these findings so not to labour the issue, but in summary we find that the division of N evaluations of the equation of state prior $p_\pi(\Omega)$ in equation 6.1 significantly affects the movement options of the walkers. In order to tackle both sampling a potentially multimodal equation of state posterior and accounting for an influential prior given equation 6.1, which may also have features which alter the sampling quality upon division many times over, we look towards a new approach which makes use of previous ways we have used the Flow which are less at-risk of potential structure.

6.3.3 Gaussian Mixture Model method

The issues discussed above in using a Normalising Flow to evaluate the seemingly multimodal posterior probability density of proposed regions of the equation of state space given conditional information complement the findings of chapter 5. With multiple sources which now suggest that using the Flow for evaluation of defined regions of what appears to be a very complex space, we wish to now combine information from multiple events in a way that does not require us to *evaluate* regions of the equation of state space $p(\Omega_{\text{prop.}}|\theta)$ but rather via sampling $\Omega \sim p(\Omega|\theta)$.

We therefore wish to follow the method of chapter 4 and use the conditional Normalising Flow $p_{\text{cond}}(\Omega|\theta)$ to sample individual event posteriors $\Omega_{\text{event}} \sim p_{\text{cond}}(\Omega|\theta_{\text{event}})$. In MCMC, proposed equations of state, or the steps of the individual walkers, are instead evaluated relative to each of the individual event posterior probability distributions, and the information from multiple events is then combined together. In order to remove any potential for learned artificial structure, or small artificially enhanced modes of information that may be modelled by the Normalising Flow, we

wrap each event's equation of state posterior with a GMM (see chapter 4 section 4.4.2.1 for background). This ensures that each event's equation of state posterior is defined purely by a superposition of Gaussians, as is defined in chapter 4 section 4.4.2.1. Therefore, the posterior probability density of a proposed equation of state $\mathbf{\Omega}$ as determined by the information from multiple BNS mergers is

$$p(\mathbf{\Omega}|h) \propto p_{\pi}(\mathbf{\Omega})^{1-N} \prod_{i=1}^{N} g_{\mathrm{i}}(\mathbf{\Omega}), \tag{6.2}$$

where

$$g_{\mathrm{i}}(\mathbf{\Omega}) = \sum_{j=1}^{M} w_{ij} \mathcal{N}(\mathbf{\Omega}|\boldsymbol{\mu_{ij}}, \Sigma_{ij}). \tag{6.3}$$

Relative to equation 6.1, we replace the Normalising Flow used to evaluate $p_{\mathrm{cond}}(\mathbf{\Omega}|\theta)$ with the GMM $g_{\mathrm{i}}(\mathbf{\Omega})$ given in equation 6.3, which models the equation of state posterior for the i^{th} event. However, we retain the Normalising Flow $p_{\mathrm{cond}}(\mathbf{\Omega}|\theta)$ to generate the samples $\mathbf{\Omega}$ to which we fit the GMM. This workflow is illustrated in figure 6.3.6. We additionally model the equation of state prior $p(\mathbf{\Omega})$ with a GMM. This is to remove additional subtraction of features which may be learned by the equation of state prior Flow and enforces that the prior is consistent with the method through which we have chosen to collate information from multiple events, now performed by GMMs.

6.3.3.1 Inferring the medium equation of state with a Gaussian Mixture Model method

We repeat the method applied in section 6.3.2.1 combining information from multiple GW events, but instead evaluating the posterior probability given in equation 6.2 through combining information from individual equation of state event posteriors, each modelled by a GMM as in equation 6.3. We likewise take 10^3 burn-in steps followed by 5×10^3 steps by 200 walkers. We again select 2000 total posterior samples by selecting 10 evenly spaced samples from each walker's chain, assuming the walkers are statistically independent.

Figure 6.3.7 presents the equation of state posterior after combining information from 4 GW events associated to the 'medium' equation of state, marked as crosshairs

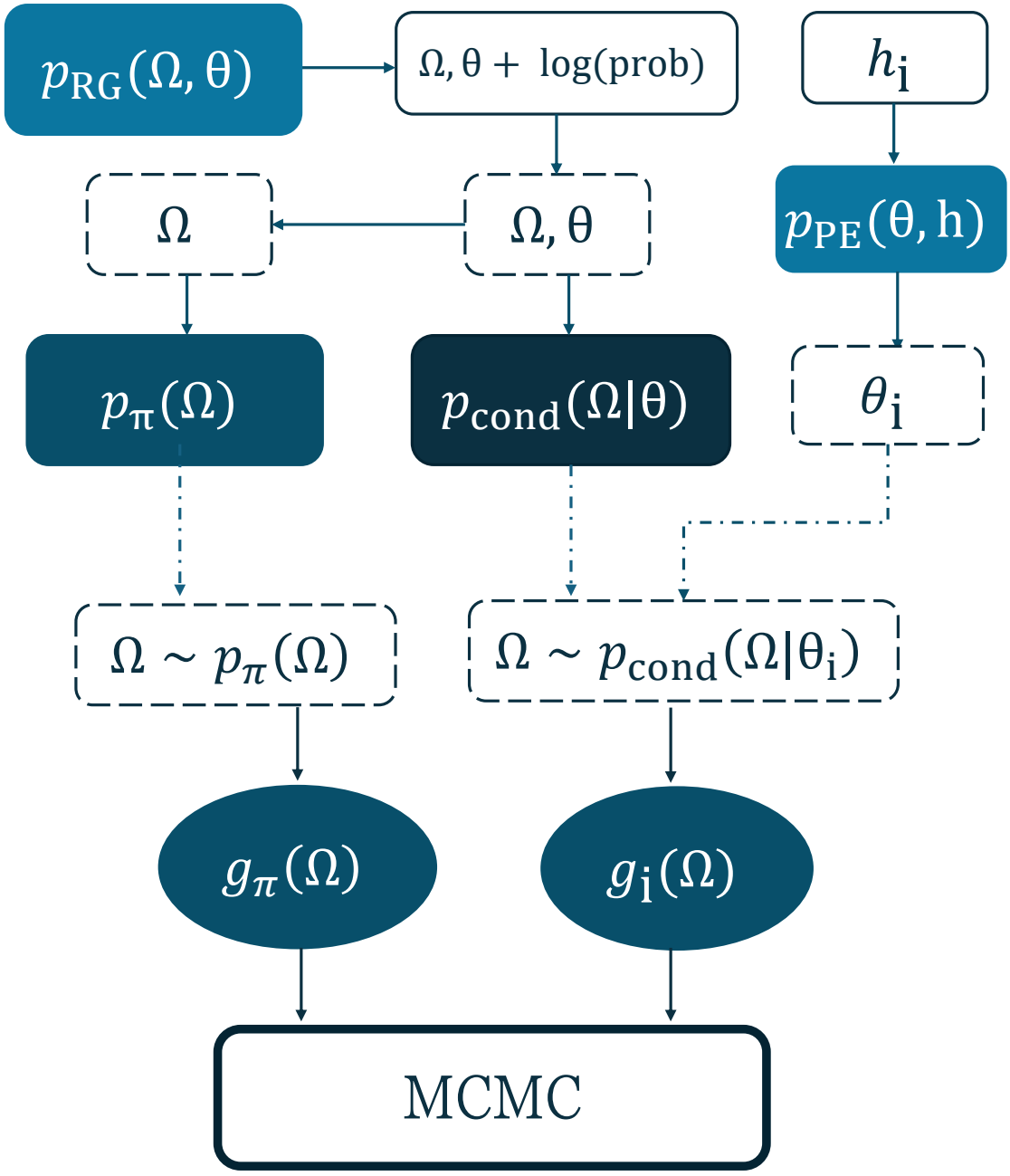


Figure 6.3.6: Block diagram of the workflow of the GMM method for hierarchical analysis. This includes the use of regeneration, conditional, prior and PE Flows. Samples θ_i are generated from the PE Flow and are used to sample $\Omega_i \sim p_{cond}(\Omega|\theta_i)$ to construct event-wise GMMs. The prior Flow is likewise sampled such that $\Omega \sim p_\pi(\Omega)$ to construct a GMM to model the equation of state prior. GMMs which represent each event's equation of state posterior and the prior are then used in MCMC for the multi-event analysis.

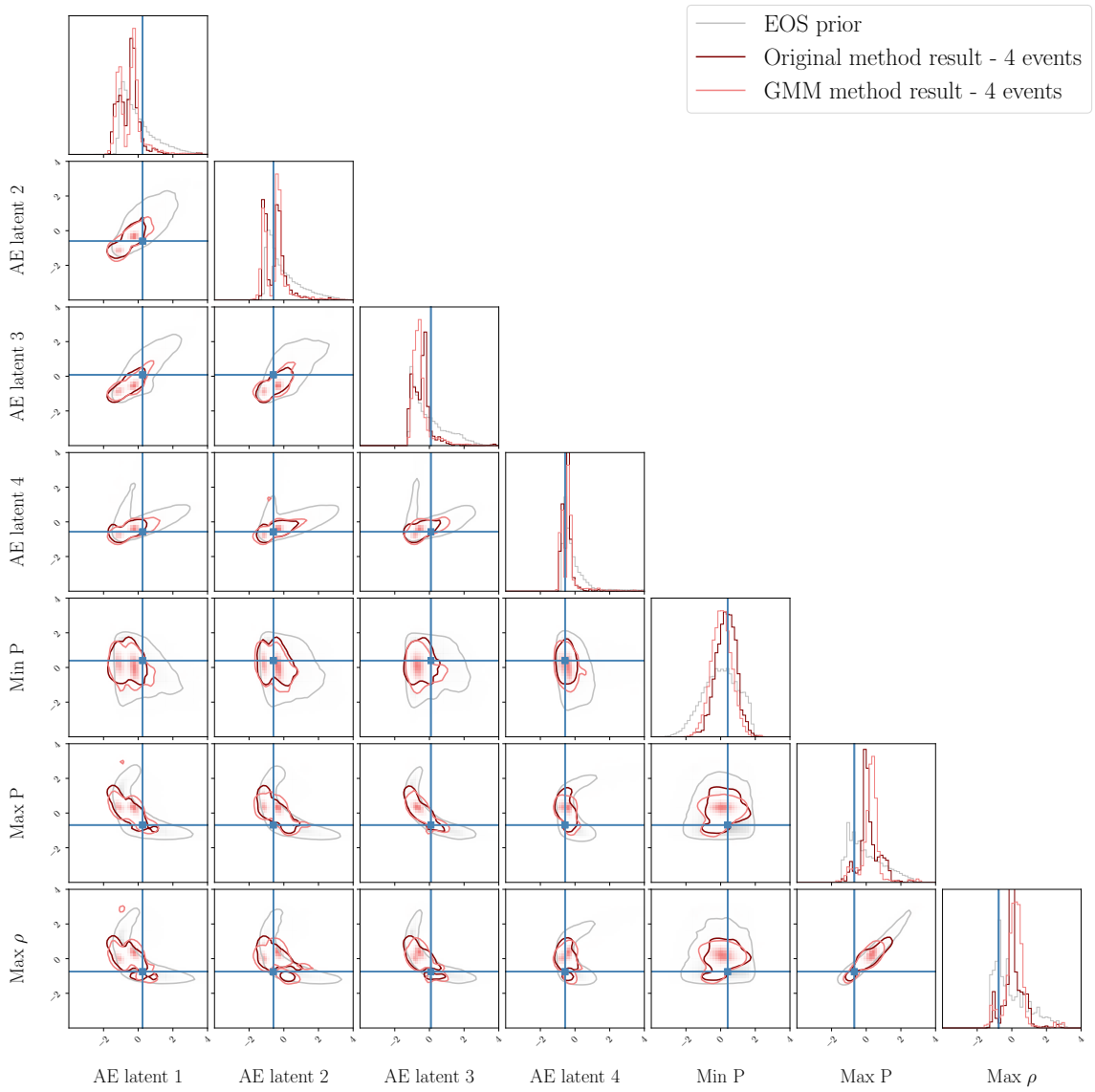


Figure 6.3.7: 1- and 2-dimensional marginalised distributions of posterior samples in the compressed equation of state space given information from 4 GW events associated to a true 'medium' equation of state using both the original method (pale red) and the GMM method (dark red) on the training data prior (grey). The contour represents 90% of the samples.

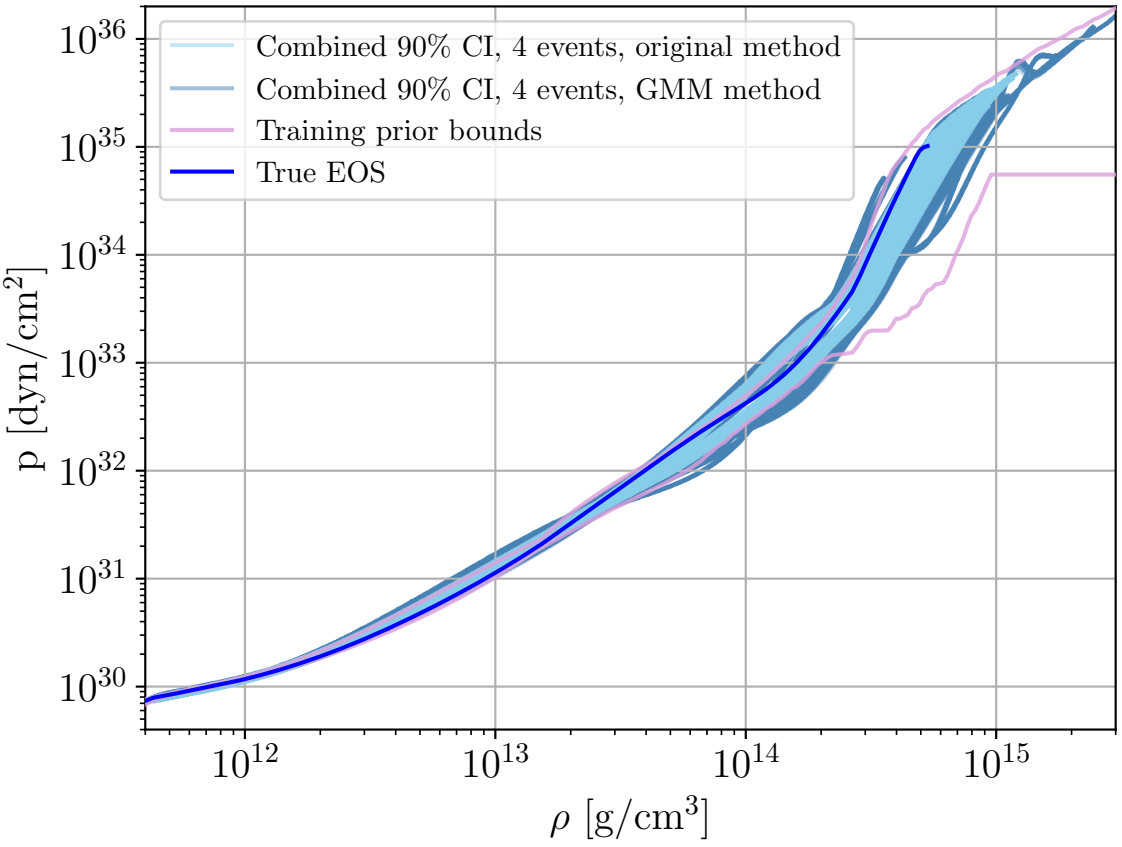


Figure 6.3.8: Plot of pressure P against density ρ of the equation of state posterior to 90% confidence given the combination of 4 GW events from an 'medium' equation of state using the original method (dark blue) and the GMM method (light blue). True equation of state plotted in bold blue line. Training prior bounds in $P - \rho$ are plotted in pink.

in the figure. The figure presents both the result using the original conditional Flow-based method and the result when the GMM method was applied instead. In both the 2−dimensional and 1−dimensional projections, the results for both methods broadly agree. This gives confidence to our new approach. In most dimensions, the true equation of state belongs within the 90% contour. In the maximum ρ and AE2 dimensions, the truth appears to lie between two modes which are present in both dimensions posteriors.

Figure 6.3.8 presents the 90% confidence interval (CI) result for 4 events from both the original and GMM methods in $P - \rho$ alongside the true equation of state which we are aiming to infer (see previous KDE method discussion). It appears that the result from both methods find similar regions of the equation of state space

desirable. Both results span a broad range of the prior, where the result for the original method is slightly more constrained than that of the GMM method, which has more outlier equations of state. The GMM method additionally returns equations of state which violate the requirement of a monotonically increasing function of pressure with density. This is not an enforced requirement in the analysis, but is not allowed in the training data. The bounds in the $P - \rho$ plane are defined purely in the 2-dimensional projection, and this boundary is not defined in the 7−dimensional space. To account for the training prior boundary in the 7−dimensional space, we follow the method used in chapter 4 which applies a GMM which models the training data prior in the compressed equation of state space. The least probable sample from the training data set as evaluated by the GMM then defines a probability threshold. If equations of state are sampled with probability less than this threshold, they are rejected. This prohibits severe outliers from the 7−dimensional space being projected into $P - \rho$. All equations of state for both methods in figure 6.3.8 have been passed through such a threshold. Therefore, equations of state which are beyond the training prior bounds in $P - \rho$ are from probable regions of the equation of state data space relative to the prior training data. This reminds us that in using either a Normalising Flow or GMM to model the autoencoder latent space, the autoencoder does not enforce that the latent space has any physical meaning on requirement on its shape, unlike a variational autoencoder. Regions of the autoencoder latent space which are not training data points themselves can't be guaranteed to produce anything valid. We take this into consideration for future inference results.

Additionally, the equations of state at 50% confidence appear to split into two branches, a result which may be a manifestation of the multimodality of the 7−dimensional space. By this, we mean that independent modes in the 7−dimensional space are decoded from clusters in 7 dimensions to clustered 2−dimensional $P - \rho$ equations of state. We assume therefore that the branching is artificial and does not make any claim about a potential invalid gap between valid regions of the equation of state posterior. Given this assumption, we calculate the pressure range at nuclear saturation density, $\rho_{\rm sat}$ to be 5.99×10^{33} dyn cm^{-2} to 90% confidence and 2.41×10^{33} dyn cm^{-2} to 50% confidence, which demonstrates the increased contraction around the truth as

we increase the number of GW events we combine information over.

We compare the number of unique steps for both methods in combining information from 4 GW events. Figure 6.3.13 presents the number of unique steps taken by the walkers in MCMC using both the original method of evaluating the probability of a given equation of state (using the conditional Flow) and the GMM method associated to the 'medium' equation of state. The evolution of number of unique steps for both the GMM and original methods are very similar for the same simulated equation of state. This suggests both methods are finding the equation of state space difficult to sample and are finding very improbable regions of the equation of state space. In the GMM analysis, we model the equation of state posterior associated to each event with a GMM, and likewise model the equation of state prior probability distribution with a GMM. In doing so, we generalise these spaces with a superposition of Gaussians; 4 Gaussians were used to make both the event-specific and the prior GMMs. This aims to remove any fine features or intricate modelling of either the event equation of state posteriors or the equation of state prior. For sampling statistics like those presented for the GMM method in figure 6.3.13, this suggests the posterior is inherently difficult to sample, and the division of the equation of state prior N times is a strong effect. This is in agreement with the result discussed in section 6.3.2.1, where we stated that in using the original equation of state pure-Flow based method but not dividing any prior via 6.1 resulted in more statistically valid sampling. We find the same result in removing the division of the prior in equation 6.2; we find that the walkers accept far more proposed steps and the MCMC runs are much more statistically valid. We compare the combination of information from more GW events to compare the quality of the GMM results further.

Figures 6.3.10 and 6.3.12 present the result of inferring the same true 'medium' equation of state with information from 8 and 16 GW events, respectively, using both the original and GMM methods. With more events, the results from the two methods begin to differ, with the original method providing more support broadly across the entire prior training space. In the 1−dimensional marginalised distributions, the original method appears more multimodal. The GMM result appears more consistent

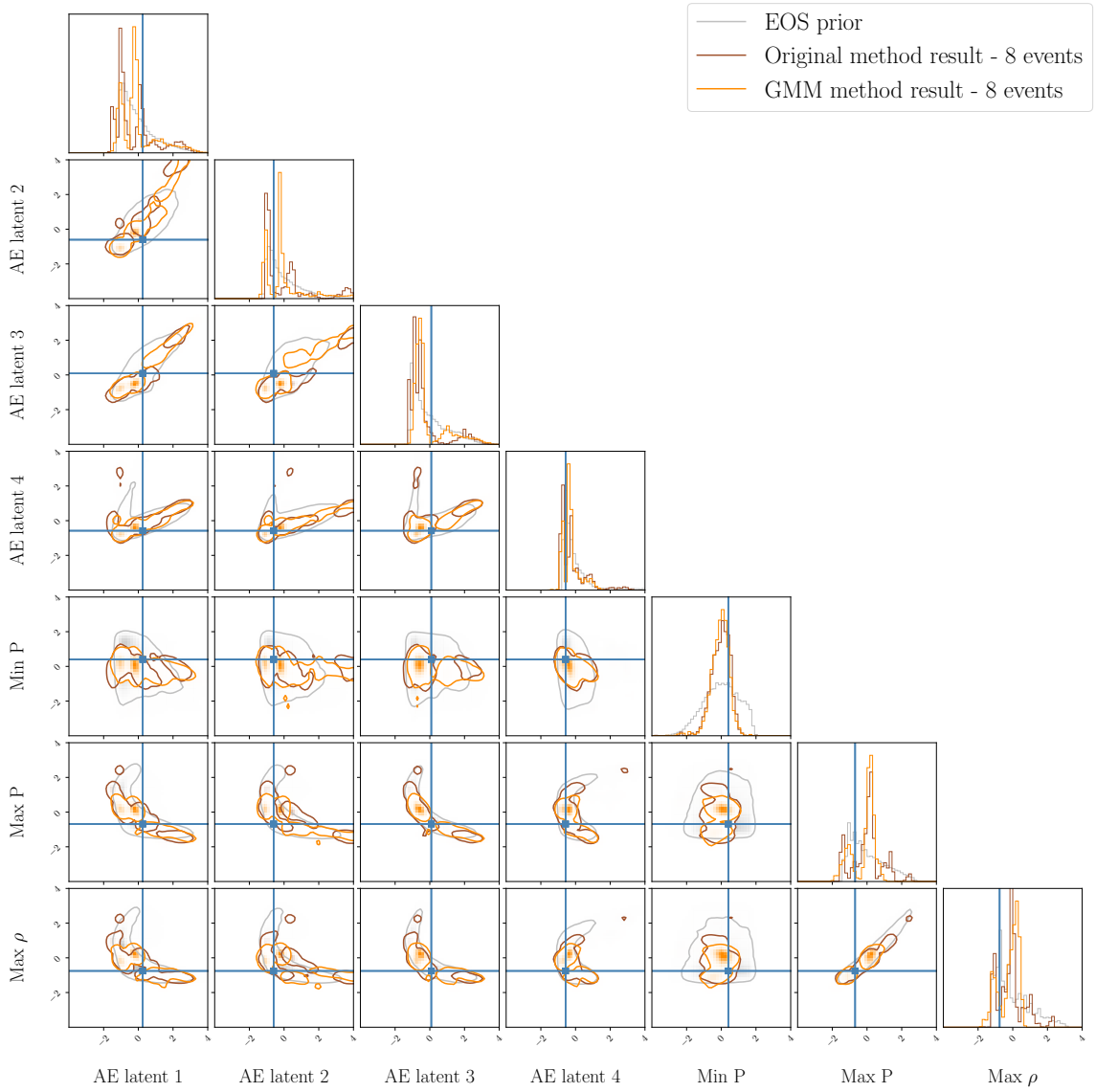


Figure 6.3.9: 1- and 2-dimensional marginalised distributions of posterior samples in the compressed equation of state space given information from 8 GW events associated to a true 'medium' equation of state using the original method (brown) and the GMM method (orange). The contour represents 90% of the samples.

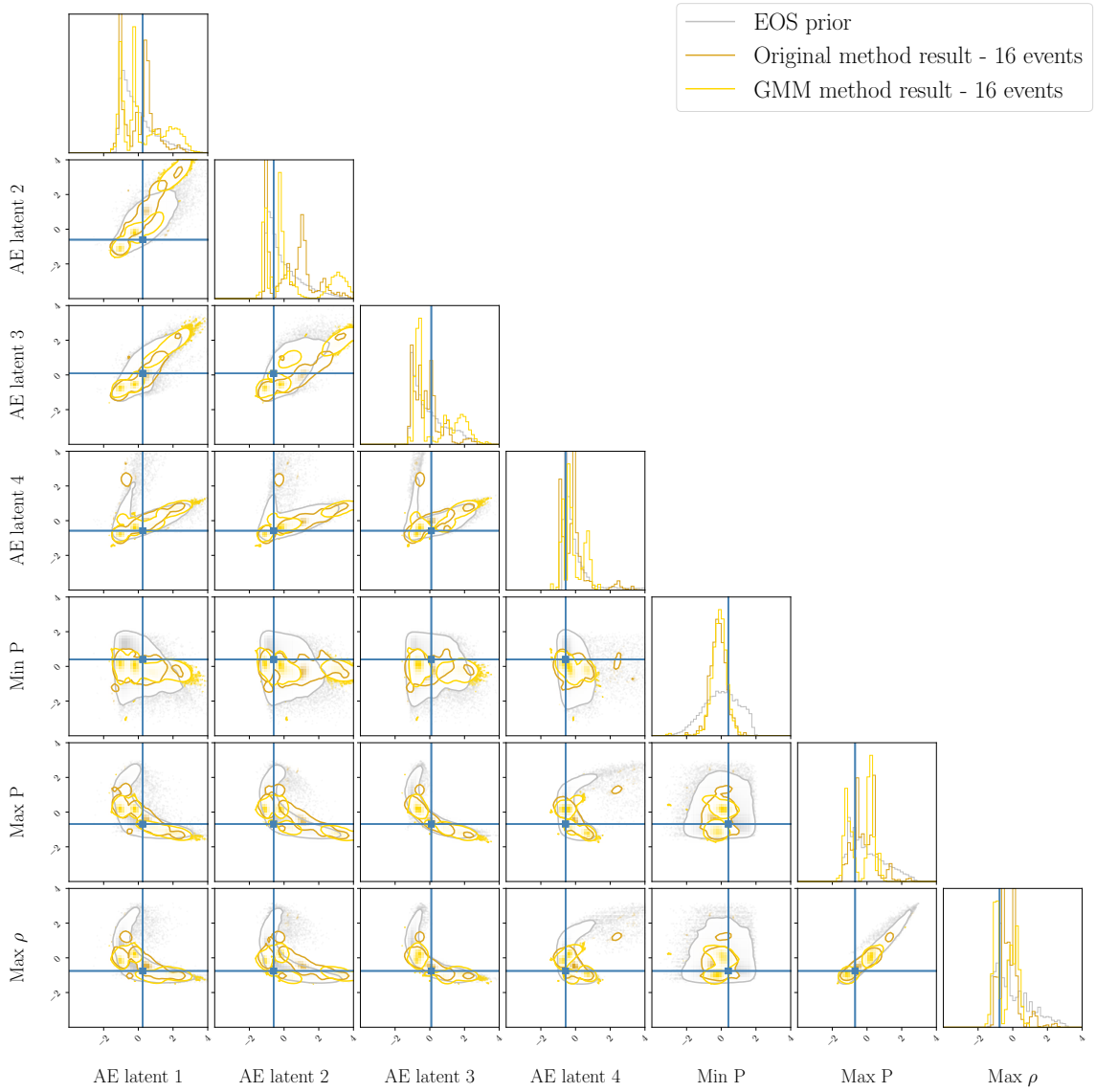


Figure 6.3.10: 1- and 2-dimensional marginalised distributions of posterior samples in the compressed equation of state space given information from 8 GW events associated to a true 'medium' equation of state using the original method (brown) and the GMM method (orange). The contour represents 90% of the samples.

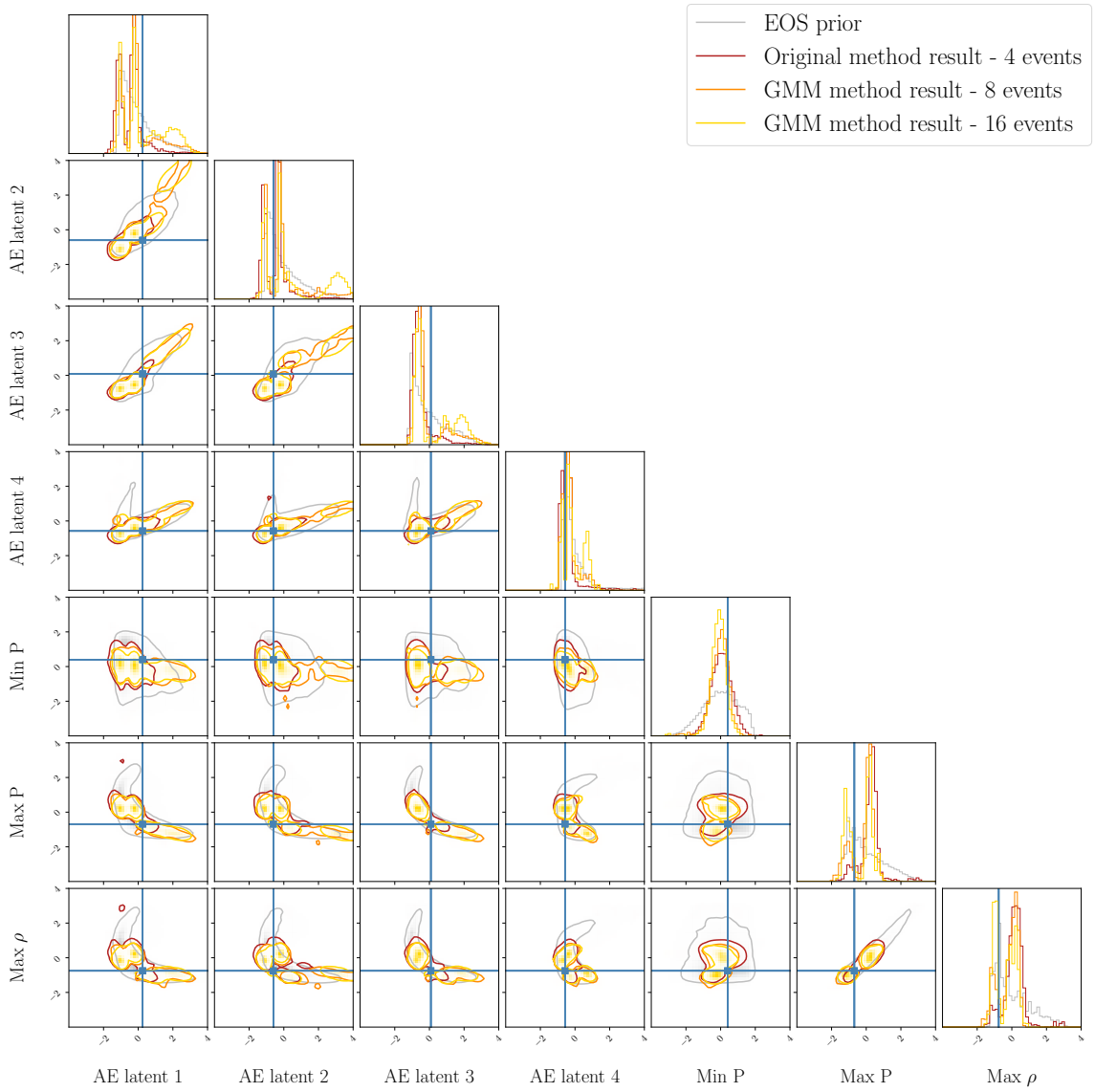


Figure 6.3.11: 1- and 2-dimensional marginalised distributions of posterior samples in the compressed equation of state space given information from 4 (red), 8 (orange) and 16 (yellow) GW events associated to a true 'medium' equation of state using the GMM method on the training data prior (grey). The contour represents 90% of the samples.

with the truth than the original method result in the 2−dimensional distributions; in most projections, the truth lies within or on the 90% probability contour. For the original method result, there are multiple small regions bounded by the 90% contour, especially on the edges of the training prior. Due to the unstable nature of the original method result, we choose to move forward with the GMM approach and test how many events we can combine over and remain stable.

We present the results of combining information from 4, 8 and 16 events associated to the same 'medium' equation of state using the GMM method in figure 6.3.11. While the result of combining 4 events differs from the others with a broader posterior which fills more of the prior, the result of combining 8 and 16 events are very similar in their 2−dimensional contours. In the 1−dimensional marginalised histograms, it is clear that in increasing from 8 to 16 events, more weight is placed in secondary modes and the posterior becomes more multimodal. In figure 6.3.13, we find that the increase from 8 to 16 events also means a reduction in number of unique steps from 6% to 4%. While this number aligns with the increase in number of modes in the posterior, both of these quantities are very low, and are much lower than would be desirable for an MCMC run. For these reasons, we produce the results in $P - \rho$ for combinations of up to 16 events for the following simulation studies. It appears that inference of the equation of state using more than 16 events is uninformative due to the properties of our method which presents itself as low quality sampling statistics. We leave inference of more events to future work, once the multimodality of the equation of state posterior is better understood and/or the sampling performance is improved. For the next and fifth observing run of the current ground based detector network, the number of observed BNS events is expected to be 28^{+44}_{-21} [299], and so we demonstrate the performance of our pipeline for these purposes; while 16 is less than the expected median value of BNS observations, it does lie within the expected range of $[7, 72]$.

The equation of state posterior after the combination of information 16 GW events is projected into $P - \rho$ and presented in figure 6.3.12 (see previous KDE method discussion). The 90% confidence equations of state span the prior range and interestingly, the increase from 4 to 16 events has removed the non-thermodynamically-stable equations of state from the higher pressure and density regions. This suggests these equations of state in the 4 event result, presented in figure 6.3.8 were on the edges of the parameter space which still had support in the 4 event case, whose support is no longer present in the 16 event result. Given 16 GW events, we find the compression of the equation of state posterior at nuclear saturation density $\rho_{\rm sat}$ to 90% and 50% confidence to be 5.29×10^{33} dyn cm^{-2} and 3.35×10^{33} dyn cm^{-2} respectively.

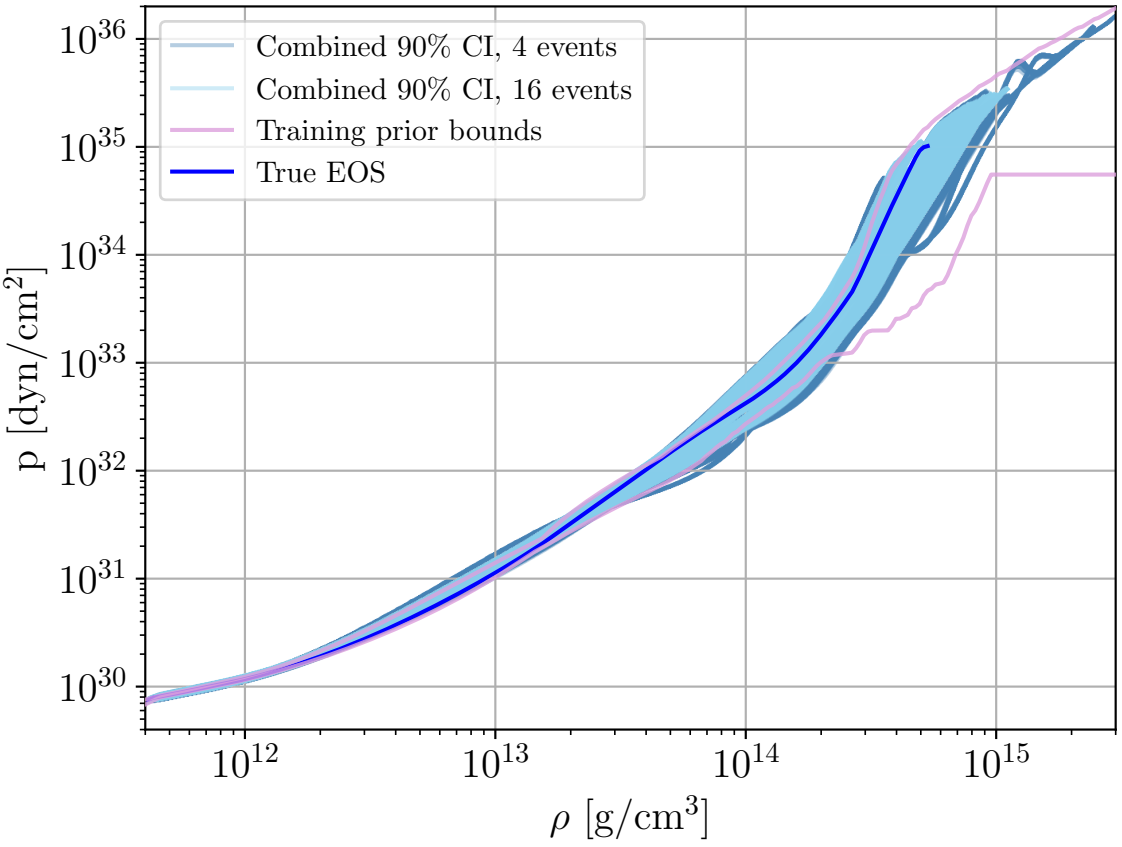


Figure 6.3.12: Plot of pressure P against density ρ of the equation of state posterior to 50% (dark blue) and 90% (light blue) confidence given the combination of 16 GW events from an 'medium' equation of state using the GMM method. True equation of state plotted in bold blue line. Training prior bounds in $P - \rho$ are plotted in pink.

Compared to the pressure contraction to 50% and 90% confidence given 4 GW events, we find the increase from 4 to 16 events has marginally decreased the error on the inference of the equation of state at saturation density at the 90% confidence level, while the 50% confidence result is more significantly constrained in the 16 event result compared to 4.

6.3.3.2 Inferring the stiff equation of state with Gaussian Mixture Model method

We also present results of combining information from multiple events associated to a true 'stiff' equation of state, presented in dark pink in figure 6.3.1. Out of the three equations of state presented, this equation of state presents the steepest gradient in the $10^{14} - 10^{15}$ g cm^{-3} density range. Figure 6.3.2 presents the $m - \Lambda$ relationship for

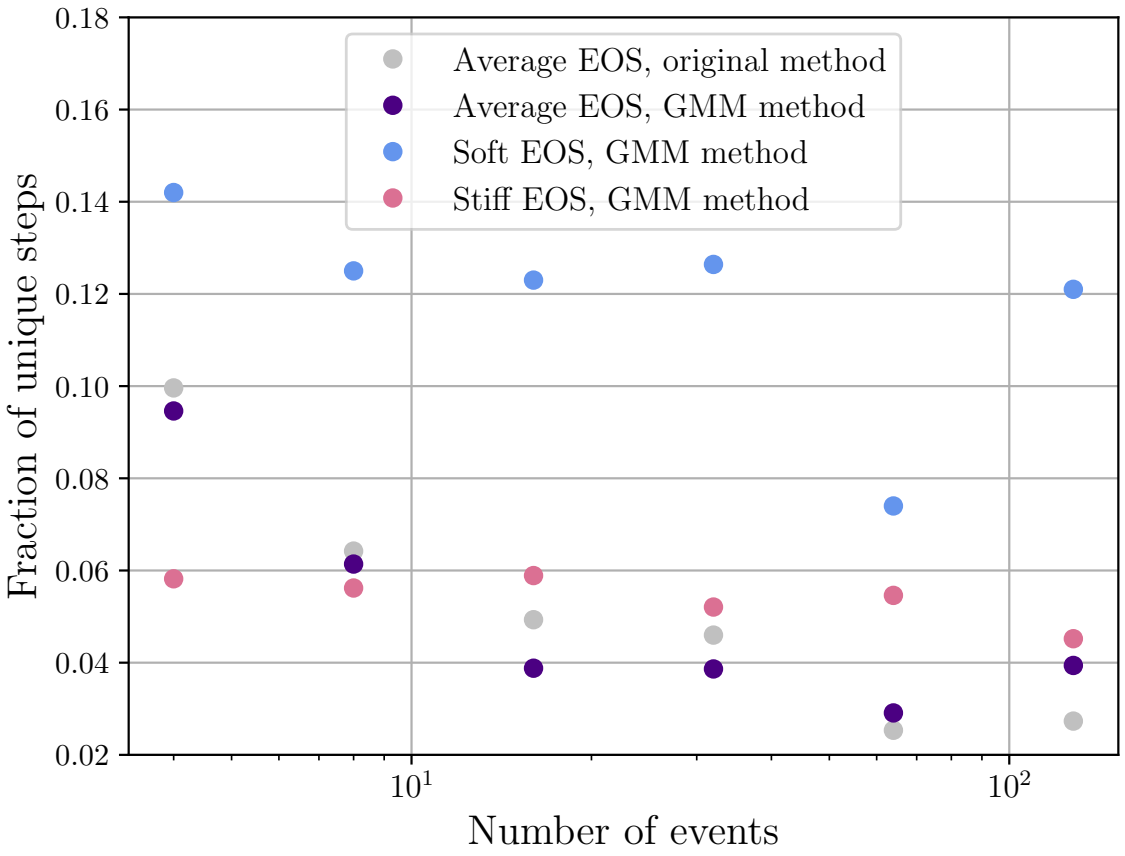


Figure 6.3.13: Plot of fraction of unique steps vs. number of events for different MCMC runs for various simulation studies associated to the 'medium', 'soft' and 'stiff' simulated equations of state using the GMM method. We additionally include the inference of the 'medium' equation of state using the original purely Flow-based method.

this equation of state, where we see support for larger Λ for a given mass, implying a more extended star with larger radius and therefore is more tidally deformable. The equation of state also supports higher mass NSs, compared to the other equations of state discussed, and supports very high tidal deformability for NSs of mass 1 $M_{\odot}$. Additionally, the $m - \Lambda$ relationship shows some kink-like behaviour which is thought to represent compactification of stiff equations of state [137], where an equation of state has a phase transition (introduced in chapter 2 section 2.1) where the pressure drops suddenly at a given density, or at the edge of a region of NS material. A NS with such an equation of state becomes more compacted as a result, relative to a NS associated to an equation of state without a pressure drop. These pressure drops are not necessarily perceptible in the $P - \rho$ relationship in figure 6.3.1.

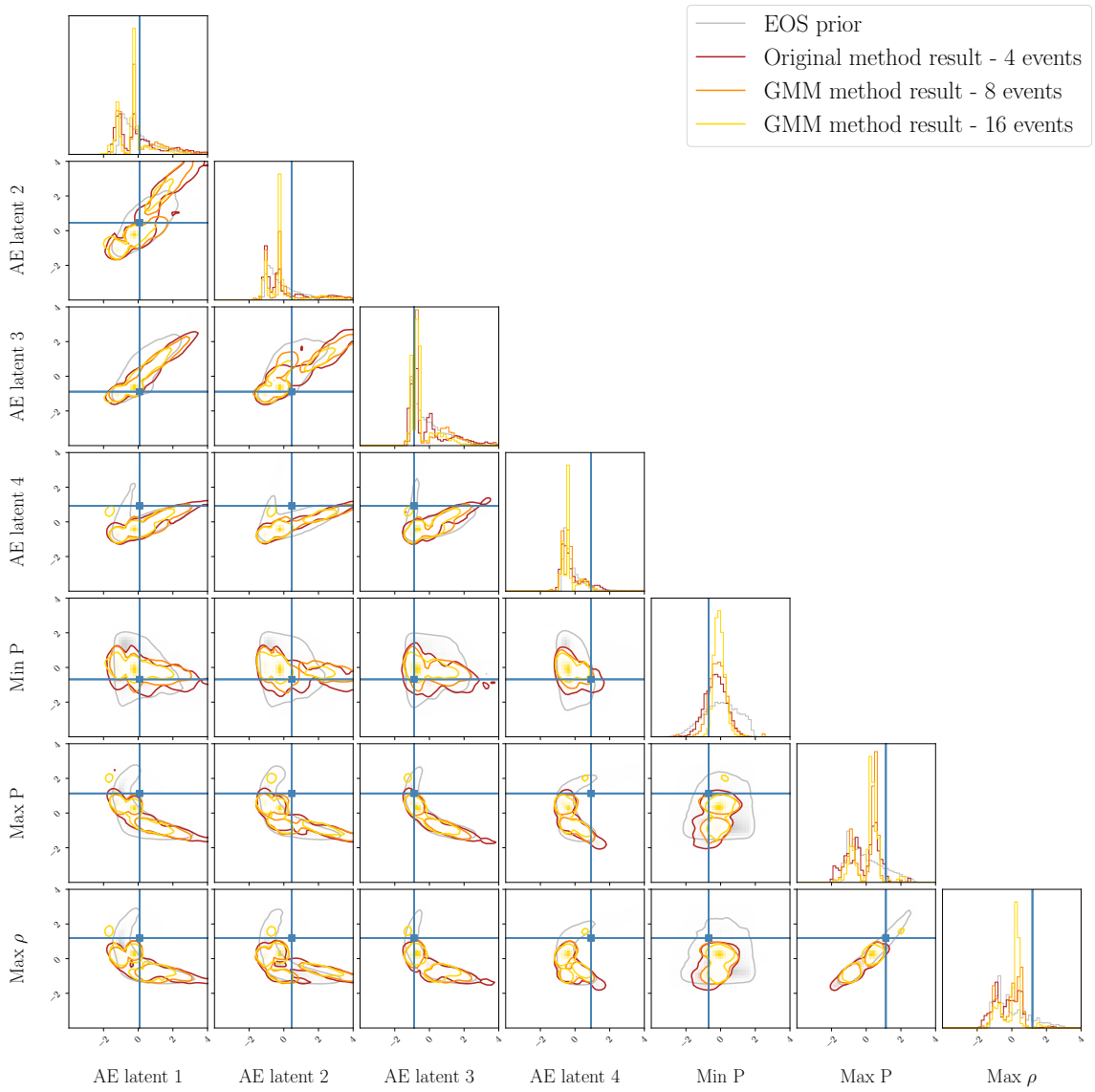


Figure 6.3.14: 1- and 2-dimensional marginalised distributions of posterior samples in the compressed equation of state space given information from 4 (red), 8 (orange) and 16 (yellow) GW events associated to a true 'stiff' equation of state using the GMM method. on the training data prior (grey). The contour represents 90% of the samples.

The equation of state posterior from the combination of information from 4, 8 and 16 events is presented in figure 6.3.14. Relative to the result for the 'medium' equation of state, we find that the 4 event result does not differ as much from either of the 8 or 16 event results. In the 1−dimensional marginalised histograms, the result for 4 events appears more multimodal than that of the 'medium' equation of state in figure 6.3.11, and the results for higher numbers of events remain multimodal with different levels of support for different modes. In most dimensions, the truth remains on the edge of or outside the 90% contour of the posterior, regardless of the number of events we are combining information from. The posteriors from the combination of 4, 8 and 16 events being similar is in agreement with the number of unique steps taken by walkers in each of the analyses, presented in figure 6.3.13. For the stiff equation of state, the number of unique walkers is consistent, regardless of the number of events we are combining information from. This suggests that for this particular simulated equation of state, the approximation to the posterior determined by MCMC is multimodal from early combination of events and does not become any more significantly multimodal, at least up to ~ 100 events. This is in contrast to the previous simulation of the medium equation of state in section 6.3.3.1, where the multimodality appeared to get worse with increasing number of events.

Figure 6.3.15 presents the $P - \rho$ projection of the equation of state posterior after combination of 4 and 16 events to 90% confidence (see previous KDE method discussion). The figure presents a decrease in spread of pressure at a given density when combining 16 instead of 4 events. The result for 16 events additionally appears to separate into two main modes, one of which is in agreement with the true equation of state up to around 2×10^{14} g cm^{-3} and then disagrees. Notably, the stiffest region of the equation of state posterior is in the region of $[2, 4] \times 10^{14}$ g cm^{-3}, at the top end of which the two modes appear to recombine.

We believe the feature of the two modes or branches in figure 6.3.15 to be an artifact, rather than something physically relevant. The branching of the equations of state in $P - \rho$ appear to be a direct result of multimodality in the 7−dimensional equation of state space. Modes or clusters of samples in a given region of $P - \rho$

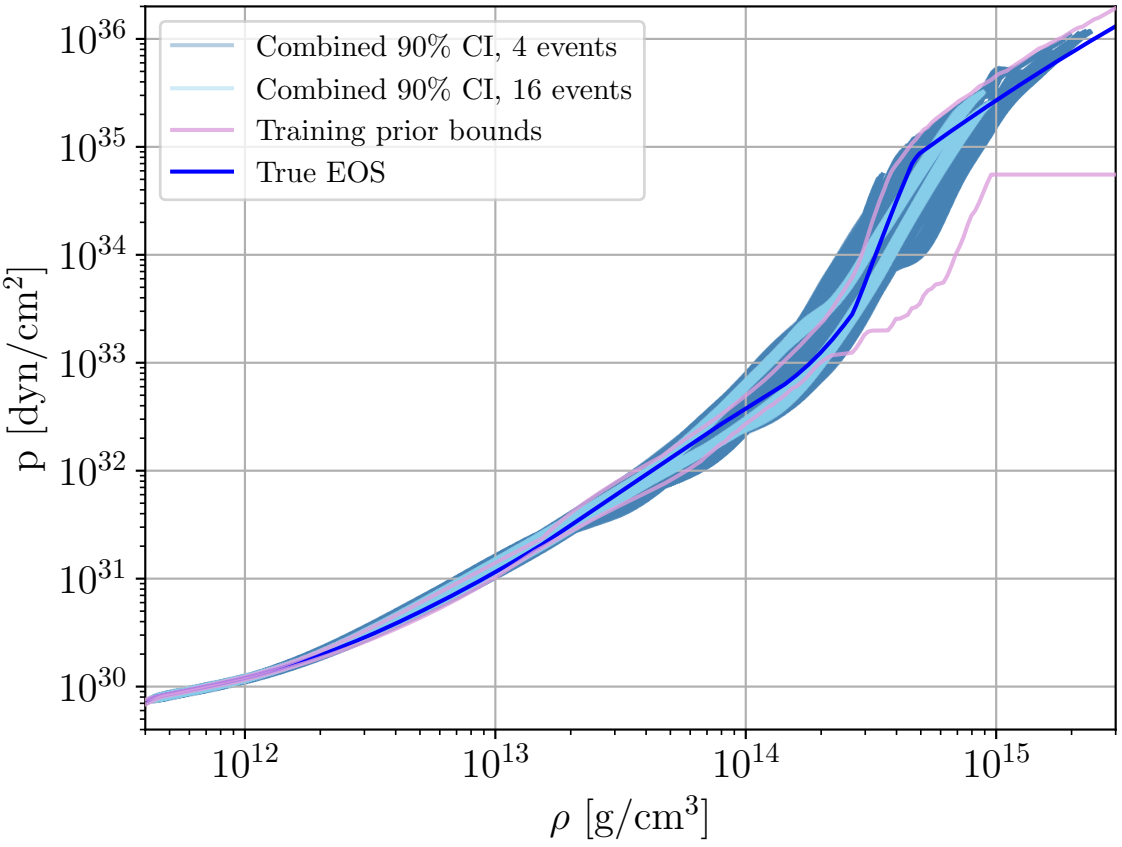


Figure 6.3.15: Plot of pressure P against density ρ of the equation of state posterior to 90% confidence given the combination of 4 (dark blue) and 16 (light blue) GW events from a 'stiff' equation of state. True equation of state plotted in bold blue line. Training prior bounds in $P-\rho$ are plotted in pink.

space appear to decode to clusters or branches of very similar equations of state. We consider two independent branches of equations of state to suggest that there is support which spans that given region of density, so we choose to ignore the gap and therefore calculate the uncertainty at given density values as if the gap was not present. The span of pressure at ρ_{sat} is 1.02×10^{34} dyn cm^{-2} at 90% confidence for 4 events compared to 3.37×10^{33} dyn cm^{-2} for 16 events. The breadth of the equation of state posterior in $P-\rho$ at 90% confidence for the 4 event result for the stiff equation of state is broader compared to the medium equation of state result, which has width of around half of the width quoted for the stiff result. However at 16 events, the compression is more comparable to that of the medium result, suggesting there is more significant contraction of the posterior for the stiffer equation of state compared to the medium. In moving from 4 to 16 events the equation of state posterior in

$P-\rho$ does compress beyond the training prior bounds at the 90% confidence level and does compress in a way that is broadly consistent with the truth. We note that our choice of events we combine information over is randomly selected from the GW event prior, so some events may be more influential than others. In future work, we suggest combining information over multiple iterations of randomly selected events from the GW prior to test the true scale of compression possible. Additionally, with a more stable analysis tool, one could calculate to what level we would expect constraint on the equation of state at saturation density given the expected number of events in the fifth observing run or in next generation detection. We leave this for future work.

6.3.3.3 Inferring the soft equation of state with Gaussian Mixture Model method

Finally, we discuss a 'softer' equation of state. The chosen simulated equation of state is soft in the sense that it is an equation of state with the least steep gradient we are considering, but with a larger allowed max pressure. We perform the same analysis, except now with 4, 8 and 16 events associated to this equation of state. We can see from figure 6.3.2 that this equation of state supports smaller tidal deformability for a given mass of NS. Therefore, this equation of state is less easily deformed under tidal forces and has a smaller radius, suggesting a NS which is small in volume and is well compacted. This equation of state has the lowest maximum allowed mass out of all equations of state discussed here due to the smaller volume and therefore less material contained within the star, owing to less total mass.

Figure 6.3.16 presents the equation of state posterior after the combination of information from 4, 8 and 16 GW events. Similar to previous results, the 4 event result is broader in the 2−dimensional contours than that of the 8 and 16 event results, however the truth lies on the edges of the 90% contour after the combination of 4 events in most dimensions, and beyond the posterior when combining information from more events. In combining information from 4 compared to 8 events, the posterior contracts in volume and removes support from the edges of the equation of state prior and instead focusses on the primary mode. In doing so, this removes support from

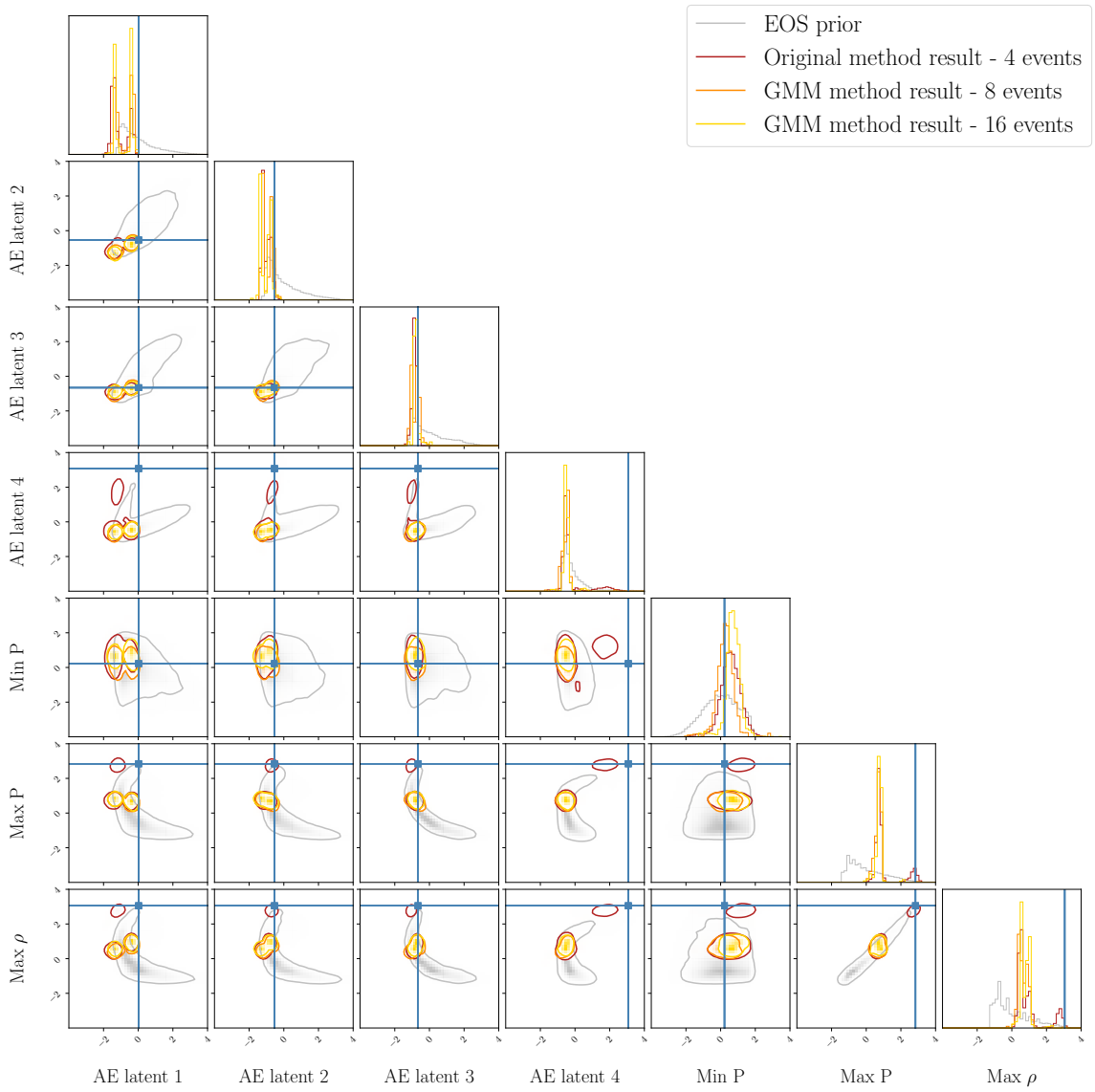


Figure 6.3.16: 1- and 2-dimensional marginalised distributions of posterior samples in the compressed equation of state space given information from 4 (red), 8 (orange) and 16 (yellow) GW events associated to a true 'soft' equation of state using the GMM method on the training data prior (grey). The contour represents 90% of the samples.

regions of the posterior which is actually consistent with the truth. This is particularly clear in the 2−dimensional maximum pressure and maximum density plane; the result which combines information from 4 events shows a clear second mode which agrees with the true equation of state. However, as the number of events increase, support for the second mode decreases and eventually shifts away from the truth.

Figure 6.3.13 presents the number of unique steps associated to the MCMC runs for this simulation study for the 'soft' equation of state. When combining information over 4 events, the number of unique steps is $4-8\%$ higher than the same metric for the runs which aim to infer the other equations of state we have introduced. As the number of events increase, this number does not decrease steadily but remains relatively consistent (apart from the significant dip at 64 events which we attribute to issues with computational resources, see section 6.4). This suggests that the equation of state posterior surface which is being explored in this run as opposed to the others has more features of comparable probability, which the walkers find it easier to move around amongst or between. We see the support for the second mode around the true equation of state wanes as we increase the number of events. If the posterior we are looking for in this case is on a sharp peak which the MCMC algorithm finds difficult to identify, as the walkers explore during the run they may instead find broader regions of the equation of state posterior space which is 'probable enough', instead of the true peak, leading to inconsistency with the truth. Alternatively, the likelihood model, modelled by the Flow or GMM, may not be accurate enough and accumulation of error over 8 or 16 events is enough to return an incorrect probability function.

We plot the projection of the equation of state posterior from the 7−dimensional result into the $P-\rho$ plane and present the projected posteriors after combination of information from 4 and 16 events to 90% confidence in figure 6.3.17 (see previous KDE method discussion). Likewise with figure 6.3.15, the posterior for 16 events is a more compressed result than that of 4 events, but with some similar branching features. The posterior in figure 6.3.16 for 16 events is multimodal – especially visible in dimensions 1 and 2 – and therefore the most logical reason for the branching is a direct reflection of the multimodality of the 7−dimensional posterior. Given that the

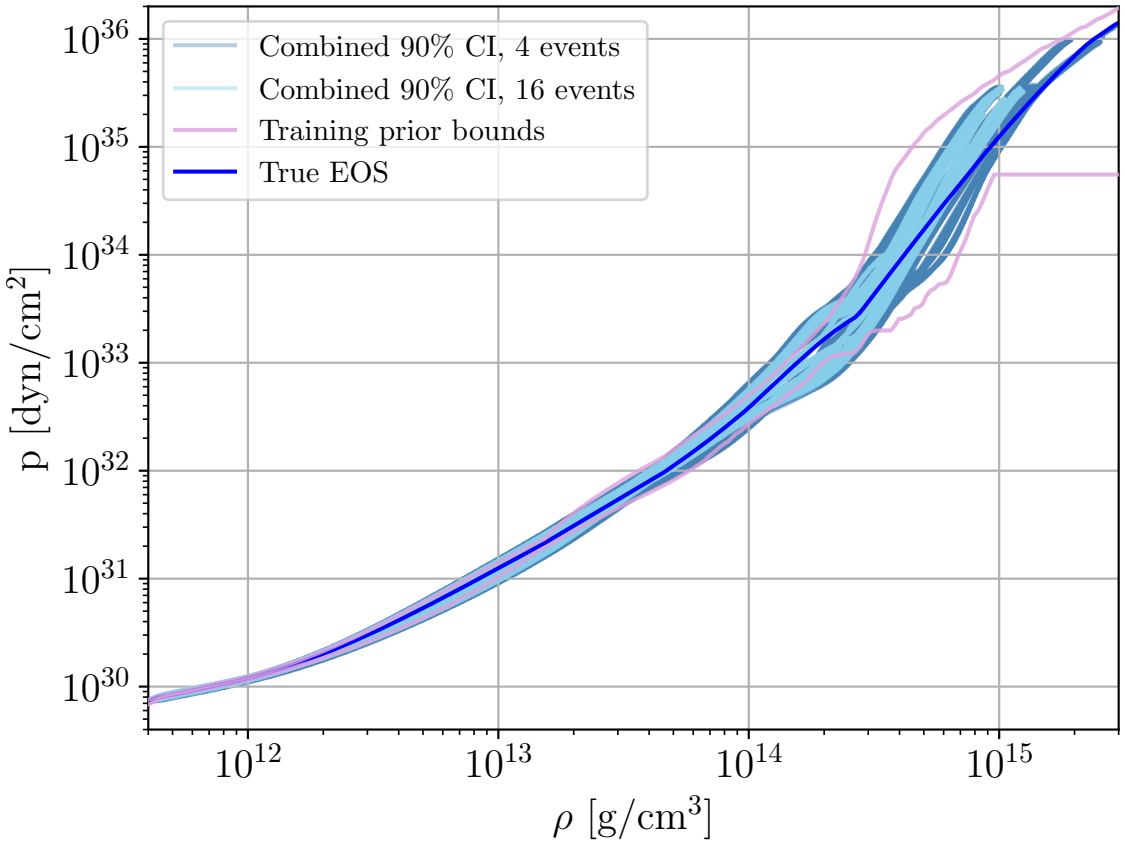


Figure 6.3.17: Plot of pressure P against density ρ of the equation of state posterior to 90% confidence given the combination of 4 (dark blue) and 16 (light blue) GW events from a 'soft' equation of state. True equation of state plotted in bold blue line. Training prior bounds in $P-\rho$ are plotted in pink.

posterior for 16 events in figure 6.3.16 does not find the true maximum pressure and density, it is therefore clear why the 16 event result in figure 6.3.17 does not extend the full breadth of the true equation of state. Given that we follow the previously described logic of assuming the branching features are artificial, we therefore place errors on the equation of state at ρ_{sat} of 3.69×10^{33} dyn cm^{-2} and 3.28×10^{33} dyn cm^{-2} at 90% and 50% confidence respectively. We assert that the 16 event result includes the true equation of state to 90% confidence up to a density of around 6×10^{14} g cm^{-3}, beyond which it no longer agrees and tends to a stiffer result. Both the 4 and 16 event results project to similar regions of the $P-\rho$ space to 90% confidence. Relative to figure 6.3.16, both posteriors in 7 dimensions are very similar and find a common (if incorrect) mode in the 7−dimensional space. This mode projects to a branch in figure 6.3.17 which is disfavoured in moving from 4 to 16 events, which agrees with

the result of increasing the number of events in the 7−dimensional result. Otherwise, the posteriors are only slightly different between the two results, and therefore the very little change in the error on pressure agrees with what we find in both figures.

6.3.4 Discussion of potential structure

Akin to the results from chapter 5, the above results of sections 6.3.2 and 6.3.3 present multimodality in the 1− and 2−dimensional projections of the 7−dimensional equation of state posteriors. We consider in the discussion of the above results if the spike-like modes are physical. The spikes are evident not only in the projected figures (for example, figure 6.3.11) but also in the statistical qualities of the MCMC sampling itself. The sampling struggles to accept steps that are proposed, suggesting that much of the parameter space is of very low log(probability) relative to some 'spikes' in log(probability). This emulates the findings of chapter 5, suggesting that the cause of this may then be a feature of Normalising Flows which is not so easily avoided. Alternatively, the spikes in the parameter space are physically motivated and are manifestations of the true equation of state posterior; however it is unlikely that the true equation of state posterior after the combination of 4 GW events is a multimodal distribution with such sharp features. We know that individual GW events with current GW detector sensitivity are relatively uninformative relative to the equation of state (see chapter 4 section 4.5 for an example of a single event result) and so we expect much of the equation of state posterior to be comparably probable. This is in contrast to the small select regions of high probability and vast swathes of severely improbable equations of state found in sections 6.3.2 and 6.3.3.

In order to test whether the features found in the combined equations of state posterior are real or are features of the Normalising Flow, we suggest tests which constitute future work. One example, would be to test such an analysis on an over-simplified equation of state training data set. Chapter 5 introduced a new training data set, and while the findings of this chapter were not unique to this data set (the artificial structure discussed in chapter 5 was also found when using an equation of state training data set of a different, and much more simplified, parameterisation), one

could over-simplify the equation of state data so to test if it is the Flow's modelling of the training data which introduces complication. One could introduce a simple $n-$dimensional Gaussian distribution as the equation of state data and perform hierarchical analysis or, use straight line equations of state with minimal number of parameters. This would simplify the equation of state data so that one can then learn where the complication is introduced to the hierarchical analysis and where the Flow begins to struggle. Alternatively, one could perform hierarchical analysis with standard stochastic sampling techniques absent of the Flow and investigate whether the multimodality is still present. While this would be computationally more expensive and take more time, this analysis would be a definitive comparison of the same conditions with and without the Flow.

6.4 Computational expense

We calculate the computational expense of such an analysis, including both the original method discussed in section 6.3.2.1 and the later GMM results. Figure 6.4.1 presents the time taken in minutes to perform the MCMC run for inference of different injected equations of state given information from different numbers of GW events. We compare the performance of the original and GMM methods using the same simulated equation of state and associated events and the performance of the GMM method on different simulations. The figure shows an upwards trend in the time required for all GMM analyses as the number of events increases. The time required for the original method for small numbers of events ($4, 8, 16$ events) is substantially more than that of the GMM method, highlighting the intrinsic computational burden for the initial method in using a given number of samples per event and using the Flow and a GPU for probability evaluation.

Figure 6.4.1 presents only the time taken to run the user-defined number of steps of the MCMC algorithm. Therefore, the results for different simulations applying the same GMM method should, in theory, be very similar given the MCMC has access to the same probability function and is making the same evaluations of probability albeit with different data. However the grey dots and the purple dots compare the

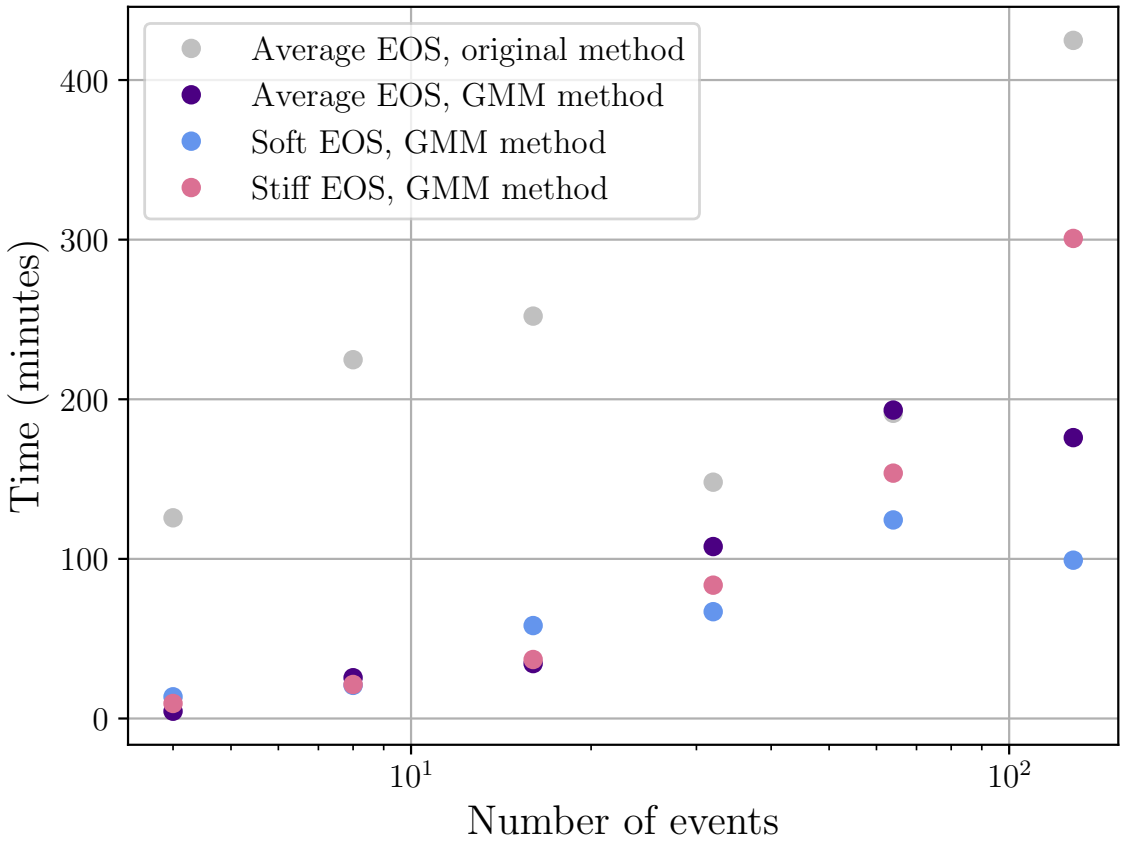


Figure 6.4.1: Plot of time taken vs. number of events for different MCMC runs for various simulation studies associated to the 'medium', 'soft' and 'stiff' simulated equations of state. We additionally include the inference of the 'medium' equation of state using the original method.

time taken for both the original and GMM methods given the same information, respectively. The time taken to combine information from 4 events for the original and GMM methods are starkly different with around 100 minutes of difference between the different analyses. This trend continues up to 16 events, the results of which have been presented here. For the original analysis, the time taken to combine information from 32 and 64 events is less than that of combining information from 8 events, a discrepancy which can be afforded to using a different and faster GPU for this analysis. Combining 128 events again used the same GPU as the combination of 32 events, and so we note the clear difference in performance depending on the GPU used.

The computational time required to perform the GMM-based analysis follows a consistent upwards trend with number of events. This analysis does not require the

Flow training	
Normalising Flow model	Time taken (minutes)
$p_{\mathrm{RG}}(\mathbf{\Omega}, \theta)$	15.78
$p_{\mathrm{cond}}(\mathbf{\Omega}\|\theta)$	25.95
$p_{\mathrm{prior}}(\mathbf{\Omega})$	15.22

Table 6.4.1: Table of time taken in minutes to train regeneration (RG), conditional (cond) and prior Normalising Flows for equation of state inference. We do not include the training time of the PE Flow $p(\theta|h)$ as this is an independent analysis.

GMM construction	
No. of events	Time taken (minutes)
4	0.12
8	0.16
16	0.29
32	0.46
64	0.94
128	1.80

Table 6.4.2: Time taken in minutes to construct all GMMs for each equation of state event posterior for different numbers of events.

Full pipeline inc. MCMC		
	Time taken (minutes)	
No. of events	Flow (original) method	GMM method
4	125.73	4.55
8	224.79	25.55
16	252.10	34.40
32	158.05	107.72
64	191.20	193.23
128	424.85	175.97

Table 6.4.3: Table of time taken for all required parts of the pipeline for both the original and GMM methods for different numbers of events. For the Flow (original) method, this includes the time taken to train all the Flow models required (table 6.4.1) and to run the MCMC for the given number of events. For the GMM method this includes the time taken to train the equation of state conditional Normalising Flow, construct GMMs for the required number of single event equation of state posteriors 6.4.2 run MCMC.

use of a GPU, and therefore as the number of events we are combining information from increases, the number of GMMs we use the evaluate the probability of a given $\mathbf{\Omega}_{\mathrm{prop.}}$ increases and the number of items we take the sum over likewise increases. Therefore the increase in time is the increased computational load, which is overall lightweight, and does not make use of a Flow.

In the original method, a discrete quantity of posterior sample data from each of the GW events is provided to the probability function, and therefore the GPU at once. The GPU processes this data simultaneously, and can do so for any quantity of data up to the memory limit of the GPU. Therefore we would expect a relatively constant time taken for all runs with different numbers of events, especially in the regime of relatively low numbers of events (and samples). Inference over 128 GW events would mean $\sim 2.3 \times 10^5$ numbers are passed to the GPU at once to then have their probability evaluated relative to the Flow. The 128 event analysis was performed on an Tesla V100 GPU, which has a memory limit of 32 GB. Therefore, processing of $\sim 2.3 \times 10^5$ numbers (assuming we used `float16` numbers which are 2 bytes each) would require 0.00046 GB, which is a fraction of the capability of the GPU. We do, however, consider that analyses were performed on different GPUs, which is difficult to be avoided in shared pools of computing resources. We discuss this next relative to the time taken for different runs.

Discussion of computational cost until now has been based purely on the MCMC algorithm. We present the computational costs of different required parts of the pipeline in tables 6.4.1 and 6.4.2. Both methods use the regeneration and conditional Flows, the training times given in table 6.4.1. Only the original Flow-based method requires the prior Flow, likewise given in 6.4.1. Since all trained models for this analysis are relatively lightweight, the training time is not a significant expense. All models were trained using an NVIDIA GeForce RTX 3090 GPU. The time taken to make a GMM which encompasses a single event posterior, given in table 6.4.2, is very small, as each GMM was designed to be constructed of only 4 Gaussians, likewise for the GMM which models the equation of state prior (this is not included in the table but takes 0.01 seconds). While no Normalising Flow is used in the GMM method to evaluate probabilities, mathematical operations in equation 6.2 were performed on the GPU through default `PyTorch` operations. The full times required to train the required Flows or GMMs for each analysis and then perform MCMC are given in table **??**. The times taken for the MCMC stage only for different workflows are presented in 6.4.1. In the original method, there is an upward trend in the time taken for 4 events to 16 with then a dramatic drop off. We attribute this to a change in GPU, as the

original method MCMC runs for combination of 32 and 64 events were completed on an NVIDIA GeForce RTX 3090 GPU while all other runs were completed on either NVIDIA GeForce RTX 2080 GPU (16 events) or Tesla V100 (128 and 8 events). Due to computational constraints, it was not possible to complete all runs on the same GPU with no competition for resources to truly compare the time taken, so we suggest this as potential future work.

We compare the computational requirements of our result to that of similar analyses. In [189] the authors build a pipeline for hierarchical inference of the NS equation of state using a KDE method. They reuse PE samples from single event analysis (a by-product of a GW observation), and then performs repeated equation of state inference for each new BNS event observed, regularly updating the prior after each observation. This details of this workflow is discussed in chapter 2 section 2.2.1.2. The primary comparison here is to the time required for the analysis. The authors quote that for inference of ~ 10 BNS merger events, their pipeline takes around 20 hours to 1 day to infer the overal equation of state. Alternatively [187] applies a similar method of using a GPU to model single event posterior samples in θ and then compute the likelihood, which they combine given multiple sets of PE samples from multiple BNS observations. By combining this information in the θ space, they wish to find relationships between macroscopic observables and then constrain the equation of state in the $m - \Lambda$ space. They demonstrate their results by combining the information from 37 BNS merger events and, assuming their analysis completes around 10^6 likelihood evaluations in nested sampling, their pipeline is estimated to take around 13 hours. Our analysis is therefore highly competitive with other pipelines in the time required to perform NS equation of state inference given observation of $\mathcal{O}(10)$ BNS merger events, with our pipeline able to perform in a fraction of the time.

Given we have presented valid results for the combination of up to 8 and 16 events for different simulations, we place this analysis to be valid for use in the upgraded current generation of GW observatories. The current expectation of the number of BNS events which will be observed during the fifth observing run of the current ground based detector network is 28^{+44}_{-21} [299]. Our pipeline is therefore suited to

operating during this observing run to combine information from the 10s of events expected. We quote above the computational time and resources required to complete the combination of information from 4 up to 128 GW events, such that, in future work tackling the multimodality and difficult in sampling of the algorithm, one can be aware of the computational requirements and expectations in advance.

6.5 Conclusion

Building on the results of chapter 5, we have demonstrated the use of a pipeline for hierarchical inference of the NS equation of state given multiple observations of GWs from BNS mergers. We demonstrate the performance of the pipeline in combining information from the first two real BNS merger events, GW170817 and GW190425, before applying to simulated events associated to known equations of state. We also discuss the computational requirements of such an analysis relative to future ground-based detector observing runs and compare to existing pipelines.

Chapter 5 demonstrated how we overcame issues with apparent artificial structure in the learned mapping of the equation of state data by the Normalising Flow. What appeared as peaks and troughs of probability in the equation of state data space were tempered by, firstly, reducing the dimensionality of the autoencoded equation of state data space and then subsequently by augmenting the training data through use of an additional Normalising Flow. This was titled the 'regeneration Flow' and serves as a cheap and effective way to generate more training data samples from a learned distribution quickly using a Flow. The results of this chapter suggested that the artificial features in the learned equation of state space has been eradicated through these additions to the workflow and as such we had a Normalising Flow model which could evaluate the probability of valid regions of the equation of state data space with rational results.

In this chapter, we firstly demonstrate the results of the workflow in combining information from the two real BNS merger events GW170817 and GW190425. While GW170817 is the much louder event with the more constrained event posterior in θ,

and therefore has the most potential for significant equation of state constraints, we present results of the combined constraint given information from both events. We present results in the 7−dimensional equation of state space, where we can see the common volume of space both events' posteriors occupy and the subsequent combined posterior in agreement. We also project results into the $P-\rho$ plane and find the result occupies a smaller region of the plane than that of GW170817 results only, but assert caution due to the difference in priors between comparative analyses.

We next introduce our method for inferring the equation of state of simulated GW events associated to know equations of state. Firstly, we introduce data from the work of Hu et al [3] who introduce a method to perform PE of long-duration BNS signals, suited to the next generation of GW observation. Their work employs another conditional Normalising Flow pipeline such that, once hours-long signals from GWs of a BNS merger are observed, one can conditionally sample for posterior samples of macroscopic observables θ. Their work also applies the method of chapter 4 in performing single event equation of state inference; through the rapid PE and subsequent equation of state inference, both using Normalising Flows, the full analysis from observation to equation of state inference can be performed in $\mathcal{O}(1)$ second. We present the result of performing single event equation of state inference using the full pipeline given two simulated signals of very different SNRs associated to the same known equation of state. We demonstrate the results are in agreement with each other and with the true equation of state, and find we achieve improved equation of state constraint for the higher SNR signal. However, due to the similarity in the results regardless of the significant difference in SNR of each of the measurements, we highlight the need for hierarchical inference for any significant equation of state constraint.

We therefore perform hierarchical inference using the Flows-based method given up to 128 simulated BNS merger events associated to a known equation of state. We find the sampling efficiency of MCMC is low in the combination of 4 events and decreases further as the number of events increases. The sampling efficiency beyond 16 events is 5% or less, suggesting the equation of state posteriors for the combination of any

more than 16 events have not evolved beyond their initial proposition of walkers from the equation of state prior.

Given the complications detailed in chapter 5 and the apparent complexity and multimodality of the equation of state posterior given multiple events, we introduce a new method which applies the findings from chapter 4 in sampling single event equation of state posteriors as opposed to using the flow to evaluate proposed equations of state. We therefore use a GMM to wrap single event equation of state posteriors as well as the equation of state prior. Therefore, in hierarchical analysis, we can evaluate the probability of a given proposed equation of state relative to each event's posterior and the simplified equation of state prior. Each event's posterior and the prior is represented by a superposition of only 4 Gaussians, so to be as generic as possible and offer no opportunity for feature learning. The choice of using 4 Gaussians was made to rapidly construct the GMM and remain generic. Using more Gaussians, perhaps to model the single event posteriors more accurately, is suggested for future work. This removes some ambiguity and complexity which is introduced by the nature of the Normalising Flows; even though our Flow models were purposely built to be simple and lightweight without the capacity for feature learning, is is clear that Flows are still flexible in the case of complex problems. However, we find the GMM results are somewhat comparable to that of the Flow result, but with improved stability. It is therefore apparent that the equation of state posterior given combination of multiple BNS merger events is naturally multimodal, and the equation of state prior removal from single event results is a prominent effect, regardless how accurately it is modelled. Given the improved stability and transparency of the GMM analysis compared to that of the Flow, we choose to move forward with this tool.

We present results up to the combination of up to 16 events from 3 different true equations of state (medium, stiff and soft), presenting the results in the 7−dimensional equation of state result and projected into $P - \rho$. We find for the medium equation of state, the equation of state posterior after the combination of 16 events occupies a smaller volume of the 7−dimensional equation of state space than that of the 4 event result, both of which are in agreement with the truth. In the $P - \rho$ representation,

the 16 event result appears to project the multimodal 7−dimensional posterior into multiple equation of state 'branches' which we deem not to be physically relevant but a consequence of the multimodality of our estimate of the equation of state posterior. While the branches are projections of real modes in our 7−dimensional equation of state posterior, these are an artefact of our approach. The inferred posterior for the stiff equation of state from the combination of 4 and 16 events are more comparable in their 7−dimensional projection, but their projected equations of state to 90% confidence in the $P - \rho$ plane show more significant contraction in the posterior when increasing to 16 events. Finally, the application to a 'soft' equation of state shows disagreement with the true equation of state when increasing the number of events beyond 4 events. The disagreement is present in both the 7−dimensional and $P - \rho$ representations, and appears to be a discrepancy due to the multimodality of the problem.

We finally discuss the computational requirements of the workflow. While the original method and GMM method are very similar in their components, the lack of requirement of a Normalising Flow in the GMM method makes the analysis significantly more efficient, where information from 4 events can be combined in less than 10 minutes. The time required increases linearly with the number of events we combine information over, owing to the number of simple computations required (equal to the number of events). We use a GPU in the analysis to be able to accelerate these operations. The original method requires the use of the trained Flow which we likewise access via the GPU. This allows us to parallelise evaluation of probabilities, which scales as number of events $\times$ number of posterior samples per events $\times$ dimensions of the equation of state space. Therefore, the number of probability evaluations required with minimal number of events is already significantly higher than the GMM method. However, in the increase in number of events to order 100s, we expect the original method to be competitive if not more efficient than that of the GMM method. As the number of events increases, the number of operations performed in the GMM method will increase linearly while the batch of operations performed on the GPU will get larger, but the time taken to perform these operations on the GPU will not change, due to the large memory and parallelisation.

We finally compare the computational requirements of the GMM method to existing analyses. Both the analyses that we compare to quote results from the combination of $\mathcal{O}(10)$ events, a similar scale of investigation to the results presented here. Our workflow is substantially quicker than the other pipelines we compare to, but its validity is restricted currently to 16 events. Other examples need from 10 hours to 1 day to combine information from order 10 events, whereas the pipeline here needs only ~ 1 hour to do so. Therefore, we project that, once sampling difficulties and Flow complications are set aside, that the use of Normalising Flows as a tool for hierarchical inference of the NS equation of state has much potential.

Chapter 7

Conclusions

Observations of NSs with GWs, while a relatively recent discovery, have offered us a new method through which to learn about NS matter. Our prior understanding of high density supranuclear matter relied on the observation of EM emission from NSs, most of which to date have been radio observations. Stemming from supernovae, NSs are the most dense matter in the universe that we know of; we believe there to be a solid external crust of nuclei arranged in a lattice structure, below which are layers of neutron-rich matter, with ever-increasing density. GWs emitted from the merger of a BNS system allows us to observe two NSs as they inspiral towards each other, gradually pulling each other apart. In the late stages of merger, we can determine to what degree the material in each of the neutron stars has been perturbed by measuring the tidal deformability of the neutron star. By measuring this parameter, alongside the masses of the two neutron stars, we can directly measure the effect of matter being present in the system. To date, there have been two observations of such an event; while the field of GW astronomy as a whole is fairly young, there have been almost 300 observations of GWs in total from the global ground based detector network, the overwhelming majority of which being the mergers of two black holes. In chapter 1, we discussed the methods of GW signal modelling and how we employ Bayesian analysis for GW PE.

In chapter 2, we introduced the concept of the NS equation of state, the universal

relationship between pressure and density inside a NS. Through the equation of state, we can communicate our understanding of the physics of neutron star matter, and relate it to the macroscopic observables of a neutron star through the TOV equations. Through an application of GR, we can solve equations of state to return relationships between $\boldsymbol{M}$ and $\boldsymbol{R}$ of a NS and can solve for the tidal deformability, Λ, and can integrate to do the reverse. We can model the NS equation of state through physical theories of ultra dense nuclear matter or otherwise phenomenologically model the equation of state, instead modelling the general shape and trend of the equation of state. In using seeking to understand the equation of state, we have different means through which we can explore it: nuclear experiment informs us about low density physics, up to the maximal densities we can support in terrestrial experiments, and we rely on astrophysical observation at higher densities. EM and GW observations are our only method of directly observing ultra dense nuclear matter, however the quality of our measurement of macroscopic observables is currently limited, and so our resultant understanding of the NS equation of state is likewise limited. To conclude chapter 2, we reviewed literature which applies GW observations of BNS mergers to inferring the NS equation of state, and discussed expectations pre-GW170817 and what has been possible post-detection.

To set the scene for the analysis performed in this thesis, we introduced ML method for scientific data analysis in chapter 3. In modern astrophysical data analysis, ML methods offer efficient solutions to handling large data sets and inferring broad trends amongst the vast amounts of information. We introduced the building blocks of neural network architectures and the best practices to follow in training and validation network performance. We introduce conditional Normalising Flows, our main type of neural network we discuss in this work. Flows are a type of generative ML model for density estimation of complex data sets. By learning an invertible mapping between a complex data space and an easy-to-sample-from distribution (like a standard Gaussian) one can then sample from the simple distribution and conditionally map this to a sample from the complex distribution. We also introduce autoencoders, which we primarily use in this work for dimensionality reduction. An autoencoder is a neural network with a bottleneck-like structure, which learns to encode training data to a

reduced-dimensionality representation and then decode it back to its original form. Once trained, one can operate in the latent representation at a reduced computational cost, and then decode an inferred result back to the original data format. Machine Learning has been used widely in GW data analysis in many areas of the field. It has been used to supplement existing analyses in detection of GW signals themselves and afterwards, inferring their parameters through PE. More recently, full Machine Learning pipelines have been developed for detection and subsequent PE which accelerate what is otherwise a time consuming and computationally heavy task. These pipelines set the scene for next-generation GW detection, where the increased number of signals observed will become a big constraint with current analysis methods.

Chapter 4 presented the first results of using a Normalising Flow for rapid inference of the neutron star equation of state. From GW observations of BNS mergers, we can measure the masses and tidal deformabilities of the two components of the merger, a standard by-product of any PE performed on a BNS observation. As previously noted, these parameters are direct measures of matter in the BNS system, and therefore are parameters through which we can infer the equation of state. We trained a Normalising Flow to map a compressed representation of phenomenologically modelled equation of state data to a unit Gaussian, conditioned on neutron star observables, mass and tidal deformability. Once trained, one can conditionally sample the Normalising Flow using PE samples from GW observations of BNS mergers and return the equation of state posterior associated to the observation. It takes ~ 0.1 second to conditionally sample the Flow with one event's worth of PE samples (on the order of 1000) to return the relevant equation of state posterior. Alongside pre- and post-processing steps, one can then return full equations of state in $P - \rho$ associate to a single BNS event in $\mathcal{O}(1)$ second. We applied the tool to inferring the equation of state given observation of the first observation of GWs from a BNS merger, GW170817 and found that the result broadly agrees with that of the LVK. We also demonstrated the performance of the Flow via two simulation studies; one using a perfect BNS measurement and another with simulated PE samples; results agree with the truth in both cases. Our method is explicitly model agnostic; we choose to train on a given equation of state morphology but could equivalently train on another, or on a mix of

multiple models. The Normalising Flow does not learn parameters of the model itself but rather a general behaviour of the equation of state, and the conditional mapping. While single event equation of state inference is relatively uninformative with regards to the equation of state, we have demonstrated a method which can facilitate rapid communication with EM astronomers after GW observation. For improved equation of state inference we require hierarchical inference.

To prepare for hierarchical analysis of the NS equation of state in next generation observation of GWs in chapter 5, we firstly introduced new training data more suited to the analysis. We introduced a set of unified equations of state with piecewise polytropic high density and metamodel low density, with no assumption of a fixed crust equation of state, to allow for the potential of inference of the crust equation of state, as the literature states may be possible with next-generation detectors like ET. Each equation of state in the training set also extends to its own maximum pressure and density, allowing for full flexibility. These complex equations of state, however, require a more sophisticated compression technique, and for this we employed autoencoders. We train an autoencoder with two convolutional layers to map our equation of state training data to a latent representation of 12 dimensions, allowing us to maintain as many unique features of each equation of state as possible. We wish to demonstrate that we can use a Normalising Flow also to hierarchically infer the NS equation of state and introduce how we do so with hierarchical Bayesian inference. For this, we require the Flow to be able to return the likelihood of a given equation of state; given data from multiple GW events, we want to determine the probability of a given equation of state. We used sampling algorithms to survey our 12−dimensional equation of state space and evaluate the probability of regions of the space given our data from multiple observations.

However, we found a significant discrepancy in using the Flow to evaluate the probability of proposed equations of state, compared to the previous chapter where we asked the Flow only to sample. We demonstrated the discrepancy on different test cases, and demonstrated that it is not an issue of simple over-fitting of the network or out-of-distribution data. Rather we determined that the Flow can identify some

structure within the training and validation data, or the compressed representation of it, the source of which is not clear. The structure appeared as peaks and troughs of probability in the 12−dimensional space, and is a feature of both the conditional Flow which maps equation of state data conditioned on macroscopic parameters and the prior Flow, which maps the equation of state space only. By reducing the dimensionality of the autoencoded latent space to 4 dimensions instead of 12, we reduced the apparent effect of these features and found a significant drop in the difference in probability between sampled points and proposed equations of state. In addition to a possible over representation of the training data with too many dimensions, we believe that perhaps we were limited in the amount of training data. Therefore, we employed a regeneration Flow, an unconditional Flow trained to learn the joint data and conditional spaces. Once trained, we can sample from this Flow repeatedly during training and generate new training data samples. This allowed for a theoretically infinite quantity of training data examples, but also data which has some noise added which is intrinsic to the Flow. This further supports the reduction in probability variation over the equation of state surface which the Flow has learned, meaning the artificial structure is gone. By exploring some unexpected features which made themselves known early in our investigations into hierarchical inference, we could identify a serious behavioural problem in applying Flows to this problem which, once solved, allowed for the use of Flows in hierarchical inference.

We reintroduced our hierarchical inference method and how we planned to apply the Normalising Flow in providing the likelihood evaluation in hierarchical Bayesian inference in chapter 6. We also used MCMC as a sampling method, such that we can survey the now 7−dimensional equation of state space and used the Flow to find the posterior probability at proposed regions of the space. We firstly presented results on applying this method to the two BNS events GW170817 and GW190425. We compared the result to the single event posterior of GW170817 returned by sampling using methods from chapter 4. We found the combined posterior returns a smaller volume of the equation of state posterior as expected, a result which is dominated by GW170817 due to the higher SNR of the signal. For our simulation study, we introduced data from the work of Hu et al. [3], who use a conditional Normalising

Flow for rapid PE of BNS signals from next-generation detectors. For simulation studies, we generated BNS events associated to known equations of state. Through the pipeline of Hu et al., we can then generate PE associated to each of these equations of state within ~ 1 second per event. We demonstrated the validity of the samples and the simulation study by first following the method of chapter 4 and using the conditional Flow sample the equation of state posterior given one event's PE samples. We found the equation of state posterior is in agreement with the truth, in both the low and high SNR simulations.

We then performed hierarchical inference as previously outlined for three simulated equations of state of varying softness/stiffness. For the mid-range injection, we found the sampling statistics of MCMC are low in combining 4 events and decreases as the number of events increases. The walkers in the algorithm appeared to find it difficult to move across the space, and the low statistical quality of the sampling alongside the 1− and 2−dimensional projections of the equation of state posterior suggest it is severely multimodal. Given individual likelihoods are uninformative (as previously shown, individual events do not strongly infer the equation of state), this suggests the correct application of the equation of state prior for each event is a strong effect, given that it gets worse with increased number of events.

Given the statistical figures of merit and the findings of chapter 5, we therefore introduced a new method which goes back to the method of chapter 4 in using the Flow to sample instead of evaluate the probability at points in the equation of state space. By wrapping each event's equation of state posterior from single event sampling with a GMM, we made the event posteriors as generic as possible by representing them by a superposition of Gaussians. We also represent the equation of state prior with a GMM so to remain generic, and therefore evaluate the hierarchical posterior via multi-event GMMs. The results of this method were comparable to that of the Flow method results and appear to be more stable. Given this method cannot have any hidden features or learning of structure as it is constructed purely of a small number of Gaussians, we moved forward with this method. We found the combined result for the average equation of state when combining information from 16 events

shows a contracted posterior compared to that of the 4 event result, both of which are in agreement with the truth. In projecting to $P - \rho$ we found some structure in the equations of state present which is a reflection of the multimodality of the result. Assuming this is an artificial feature, we found the error in pressure at saturation density decreases with increasing number of events. The stiff equation of state showed different statistical behaviour but a result mostly in agreement with the truth, whereas the soft equation of state was a more extreme example case, whose result does not agree with the truth, especially as the number of events are increased. We owed the disagreement with the truth to the difficulty of the sampling in the multimodal approximation to the equation of state posterior. At its root, the problems are likely to be due to the inaccuracy of the the Flow and GMM models.

However, the time required to complete such an analysis is competitive to similar analyses in the literature. With the GMM method, the time required to combine information over 4 events is around 45 minutes, for the full pipeline, and 50 minutes for 16 events. While these results make use of a GPU to accelerate simple computation, the result for the full pipeline – including the training of different Normalising Flows – the time required is significantly less for similar analyses which combine the same order of magnitude of events. However given the nature of the analysis we believe there is ample room for future work and development.

7.1 Future work

The future of GW astronomy has much to offer in improving our understanding of the NS equation of state. As numbers of observations of GWs from BNS mergers increase – additionally at improved levels of sensitivity – the work presented here may be applied and improved to support this cause. The work of chapter 4 successfully applies Normalising Flows to inferring the NS equation of state giving observation of GWs from single BNS merger events. The rapid output of equation of state posteriors in $P - \rho$ once PE samples from the event have been released facilitates communication with EM astronomers for follow-up studies. For improved rapid inference, one can adapt on the work of chapter 6 section 6.3.1.1 in combining the pipeline of chapter

4 with a ML-based PE pipeline, which can return both PE samples and subsequent equation of state inference in $\mathcal{O}(1)$ second. Rapid equation of state inference after a BNS merger event may inform astronomers of the type of EM emission expected post-merger. By inferring the equation of state pre-merger, one can then make EM astronomers aware of potential kilonovae or emission in a particular frequency band. This allows for efficient follow-up and appropriate allocation of resources. The pipeline can additionally be altered to account for equation of state inference of NSBH systems, such that equation of state information from these types of events is not overlooked, given the current limited number of BNS events. Finally, the training of the pipeline can be improved to account for a larger range of equation of state parameterisations. This would ensure the Flow model learns a broad variety of equation of state morphologies and ensures the inference is flexible.

Chapter 5 presents new findings of applying Flows to high dimensional data of limited quantity with potential artificial structure. If time had permitted, the initial goal of further work would have been in applying different Flow architectures to the problem. For example, Flows which learn to map lower-dimensional manifolds in high dimensional spaces or continuous Normalising Flows are interesting alternatives to consider when dealing with a complicated data set. Alternatively, it would be interesting to run the workflow on a test case of mock equations of state, such as straight lines with defined gradient and $y-$intercept, to investigate which feature the Flow finds it easy to specify to. In the same vein, one could investigate different data compression methods, which includes changing the architecture of the autoencoder. While section 5.4.5 investigates different dimensionalities for data compression via the autoencoder, more time to investigate this would have allowed for proper optimisation of each of the lower dimensionality autoencoders. Chapter 5 also introduces the regeneration Flow in section 5.5. Significant future work could be devoted to applying this method to other astrophysical data analysis problems which are limited by data quantity. Additionally, this tool could be beneficial to problems which have a potential ML model in mind but whose training is limited by shortage of data, or whose training does not seem optimal. The regeneration Flow offers a low-cost alternative to data generation for virtually any problem.

With regards to chapter 6, the primary goal of future work will be to improve the efficiency and quality of the sampling methods. One option is to investigate alternative sampling methods, perhaps those which are more sophisticated in nature and are suited to high dimensional spaces. One example sampling method to test would be MCMC with parallel tempering, which is suited to sampling in highly multi-modal spaces. This would reduce the discrepancy in probability between the most and least probable regions at the start of sampling, and gradually re-introduce the variability after the walkers have had the opportunity to explore the space. Alternatively, due to the flexibility of Normalising Flows as both a density estimation and a sampling tool, one could investigate a method which does not require sampling algorithms at all, but rather makes use of the efficiency of the Flow. This is extensive additional work. Alternatively, in relation to chapter 5, one could investigate how best to represent the equation of state and if the way the space is currently represented is optimal for our sampling method.

Once the sampling efficiency is improved, there are a number of avenues for future work with the tool introduced in chapter 6. These are highly important with regards to the prospects of equation of state inference during the next observing run and subsequently generation of GW detection. Once sampling methods or the accuracy of the Flow/GMM models are improved, one can combine information from ~ 100 up to 1000 events quickly and easily. In chapter 6 section 6.4 we discussed the logic for the computational requirements of the tool (if functional beyond ~ 16 events) making use available computational resources (GPUs). We found that the extension of the Flow-based method to higher orders of magnitude of events would be possible within comparable timescales. Tests can be done to estimate the computational time required for such an analysis, and what resources are best applied. Once completed, we can therefore make a claim as to the quality of equation of state inference with orders of magnitude more events than have been combined before. This allows us to test the quality of different stages of the analysis; how well does the quality of our data compression method matter in projecting a highly constrained equation of state in $P - \rho$? Is it necessary to include the crust equation of state in inference? Are we limited by the PE samples we have and the restrictions of our chosen pipeline?

How many events do we need to make a statement of whether the equation of state is either soft or stiff?

Throughout this work we have demonstrated the power of ML tools in allowing us to infer the NS equation of state quickly and efficiently. As we look to the future of GW observation, it is crucial that we consider the efficiency and flexibility that these tools may offer us in allowing us to learn from our potentially vast numbers of future GW events, and communicate with our peers for further discovery.

Appendix A

We present the full 15−dimensional figures of those presented in chapter 5.

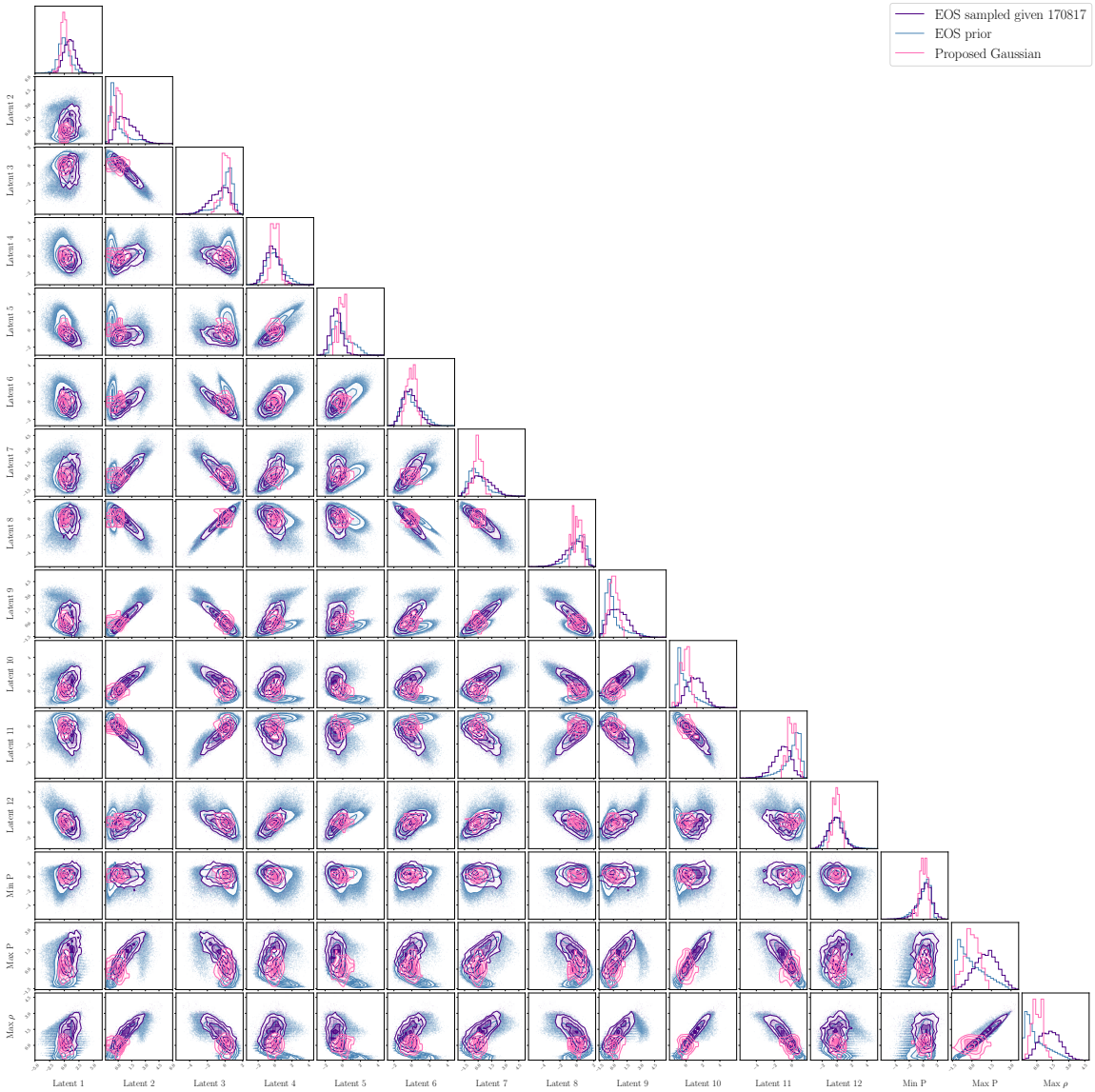


Figure A1: Corner plot of 1- and 2-dimensional marginalised distributions of training equations of state (blue), sampled points from the conditional Normalising Flow (purple) and proposed equations of state (pink) from a Gaussian distribution with $\sigma = 0.5$.

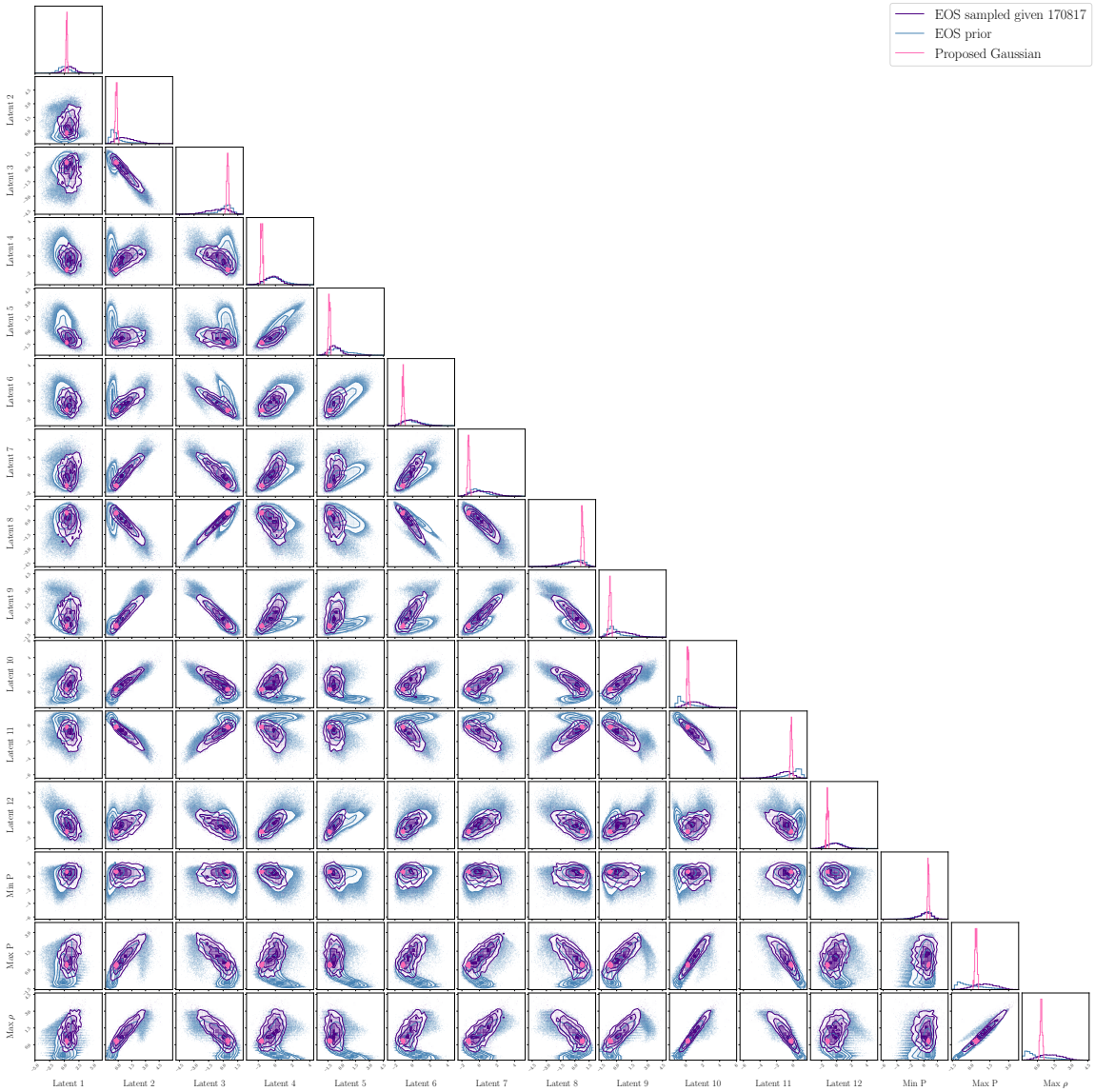


Figure A2: Corner plot of 1- and 2-dimensional marginalised distributions of training equations of state (blue), sampled points from the conditional Normalising Flow (purple) and proposed equations of state (pink) from a Gaussian distribution with $\sigma = 0.1$. We present the first 5 out of the total 15 dimensions.

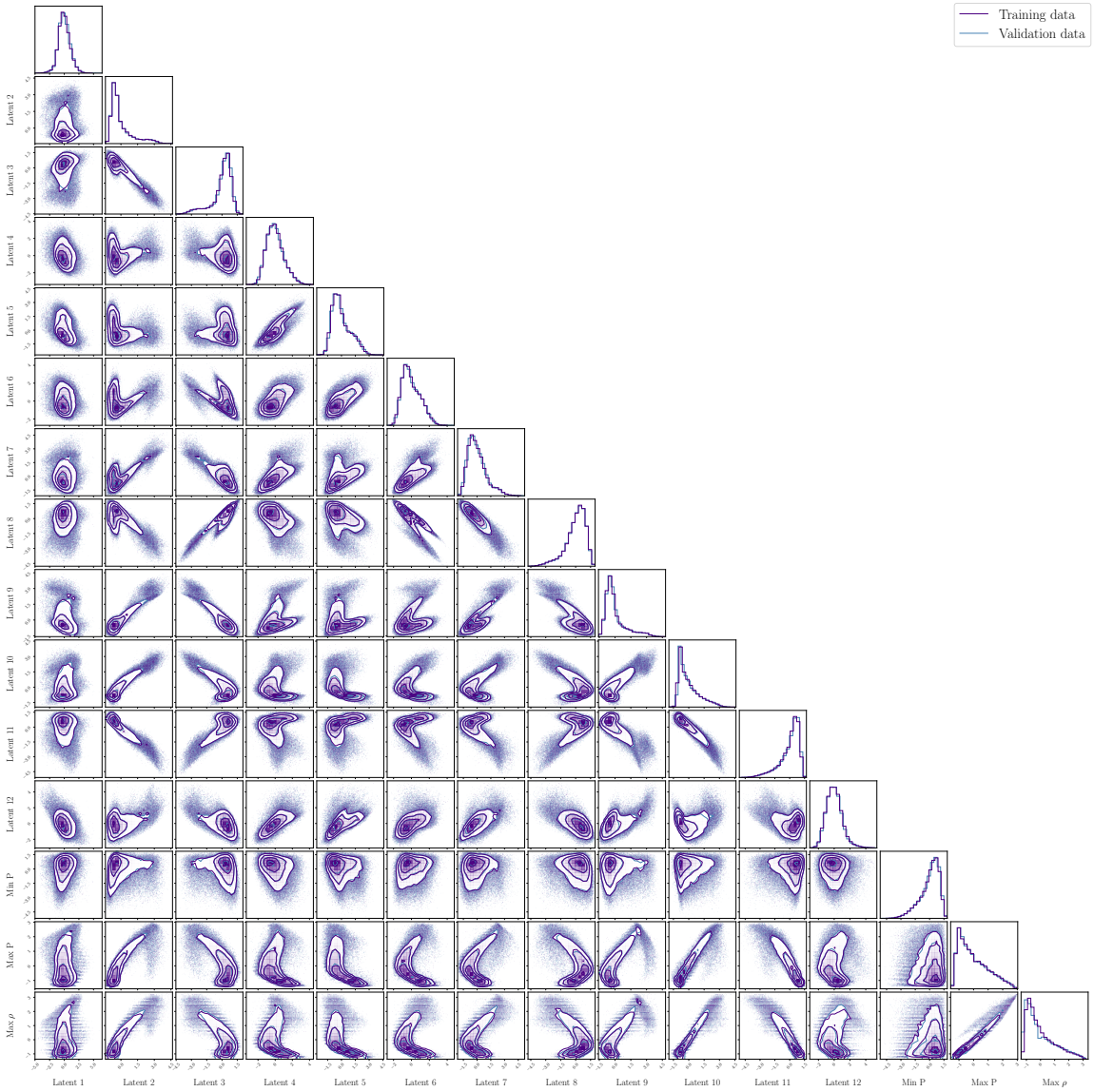


Figure A3: Corner plot of 1- and 2-dimensional marginalised distributions of the 15-dimensional training (purple) and validation (blue) data distributions used to train the Normalising Flow. The contours represent 68%, 95%, and 99.7% of the samples respectively.

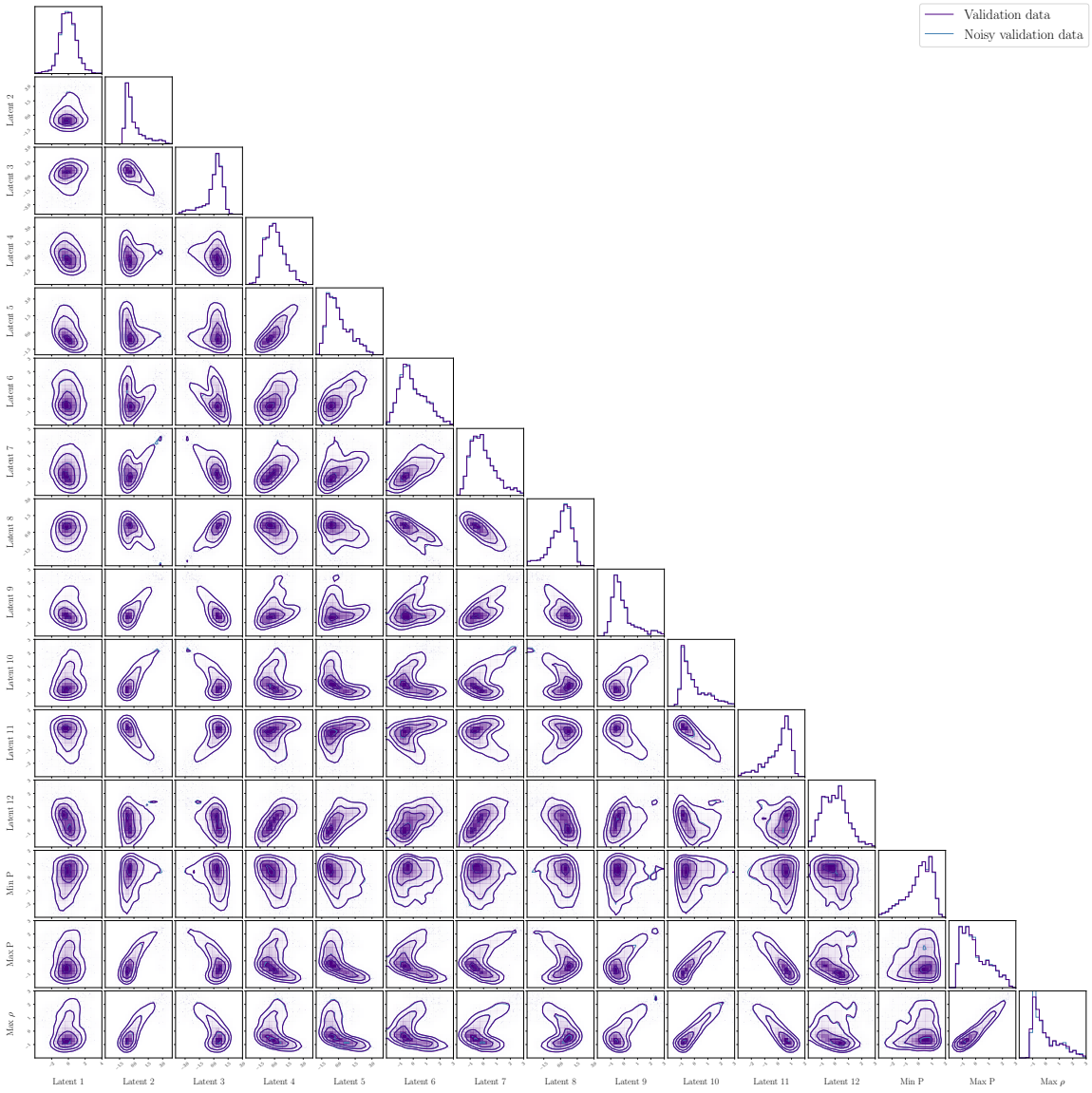


Figure A4: Corner plot of 1- and 2-dimensional marginalised distributions of validation equation of state data and noise-added validation equation of state data.

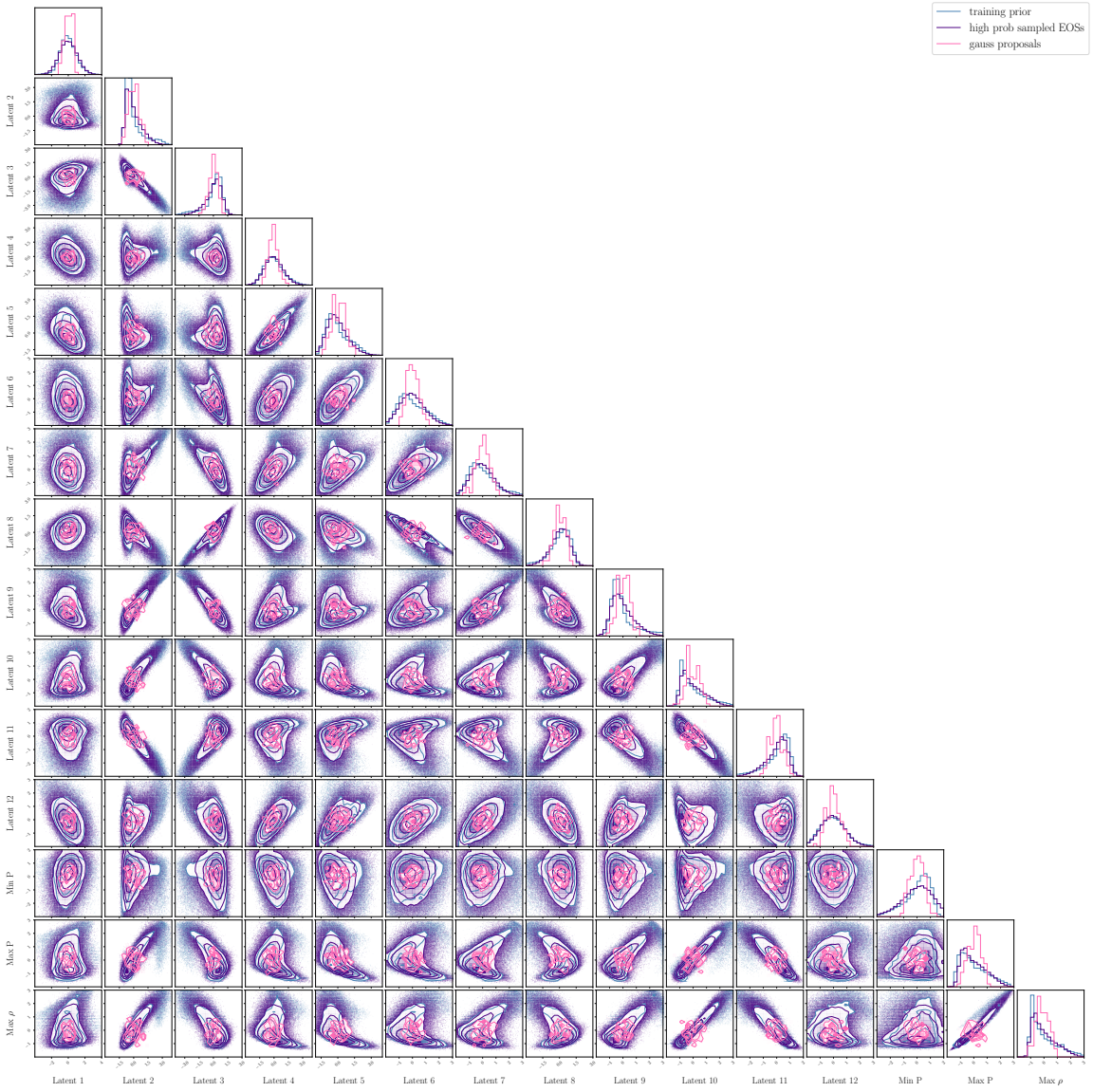


Figure A5: Corner plot of 1- and 2-dimensional marginalised distributions of validation equation of state data and noise-added validation equation of state data.

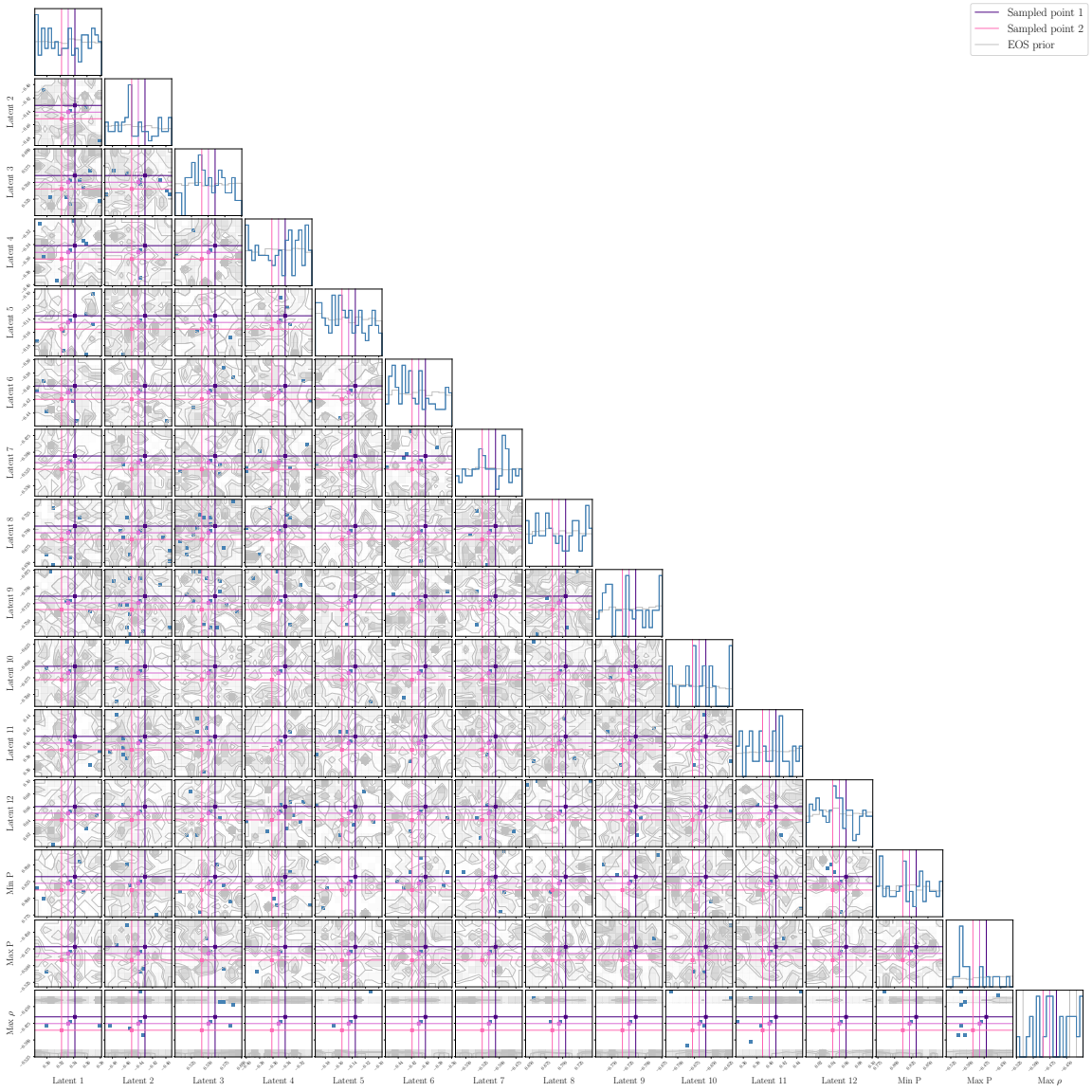


Figure A6: Corner plot of 1- and 2-dimensional marginalised distributions of a small region of equation of state data space including region of maximum likelihood sample. The grey contours are the training data prior, the blue points are points conditionally sampled by the Flow. The light pink crosshair is the maximum likelihood sample returned the Flow. The space is travelled from ±0.01 around this point from the dark pink to the purple crosshairs.

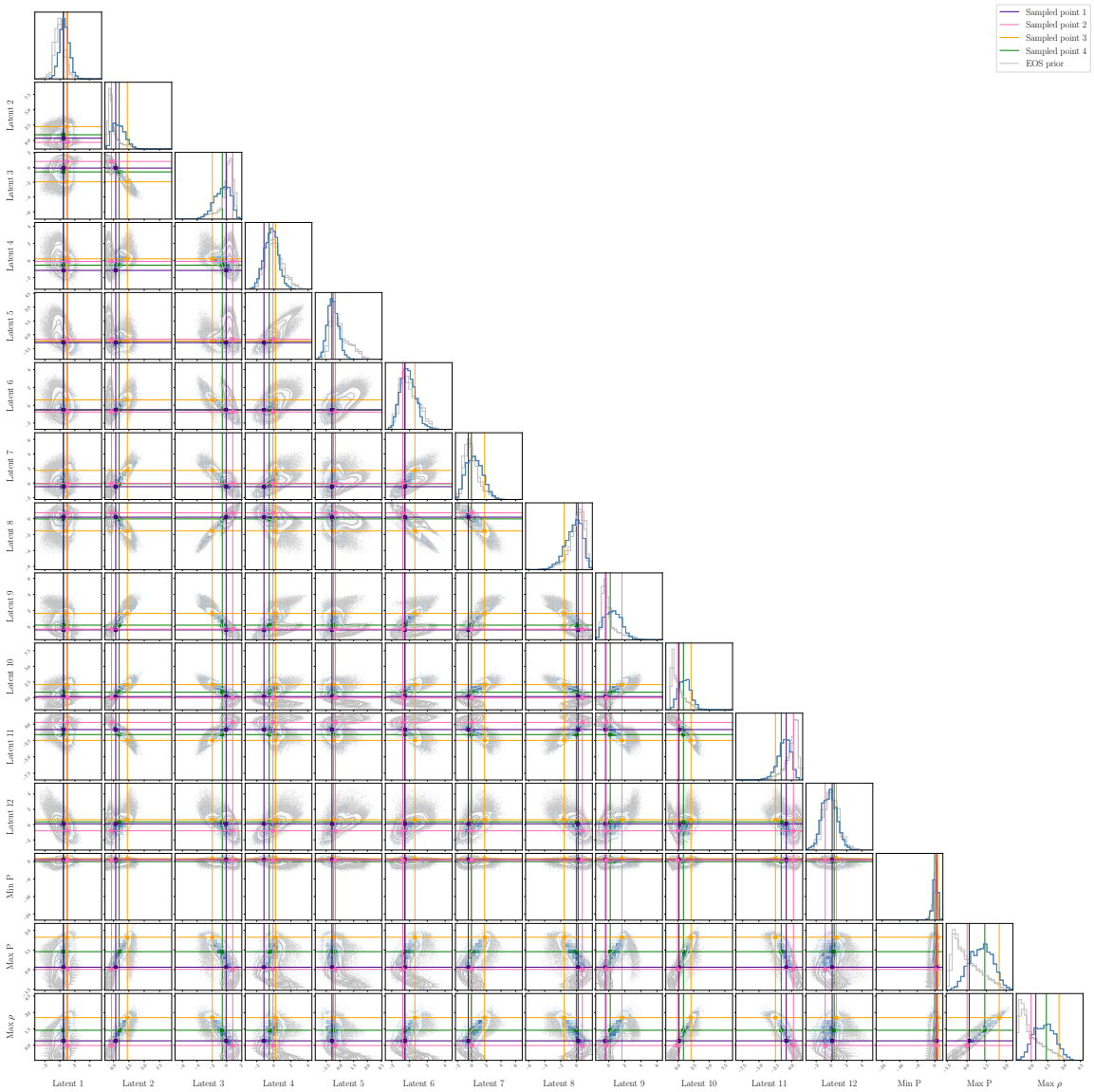


Figure A7: Corner plot of 1- and 2-dimensional marginalised distributions of points from the equation of state training data prior (grey), sampled points from the Normalising Flow (blue) and four proposed equations of state (pink, purple, green and yellow) which we traverse between.

Appendix B

In chapter 6 section 6.3, we describe the method for combining information from multiple GW events using the conditional Normalising Flow for equation of state inference. MCMC is used to traverse the equation of state parameter space, and at each proposed location, the probability $p(\Omega|\theta)$ is evaluated for multiple values of θ relative to multiple GW merger events. Given equation 6.1, we find the Monte-Carlo average over the number of samples per GW event, as well as taking the product over the number of events. Therefore, we must use enough samples per GW merger event as to appropriately represent the true single event equation of state posterior, but not too many so to become a computational burden.

Figure B1 shows how average log(probability) differs as we increase the number of PE samples used per GW event. The orange line shows the average log(probability) found when using all samples in the GW event. For almost all tests of different numbers of samples, the range of average log(probability) values returned is consistent with the average log(probability) if using all samples. We include the result of using as little as 32 samples per GW event so to demonstrate the potential spread in error from various different iterations of 32 samples from the full posterior set, which contains 10528 samples in total. Using less than 128 samples likely leads to an incomplete GW posterior in the 4 dimensions that are of interest to this analysis. As expected, the error in average log(probability) decreases in line with the truth as we increase the number of samples used per event. We find the error in using 256 samples is reduced from that of using 128 but is comparable to that of 512. The smallest amount of error occurs beyond 2048 samples. However, as stated previously, we also have to be cautious about the computational burden of using more sample per event.

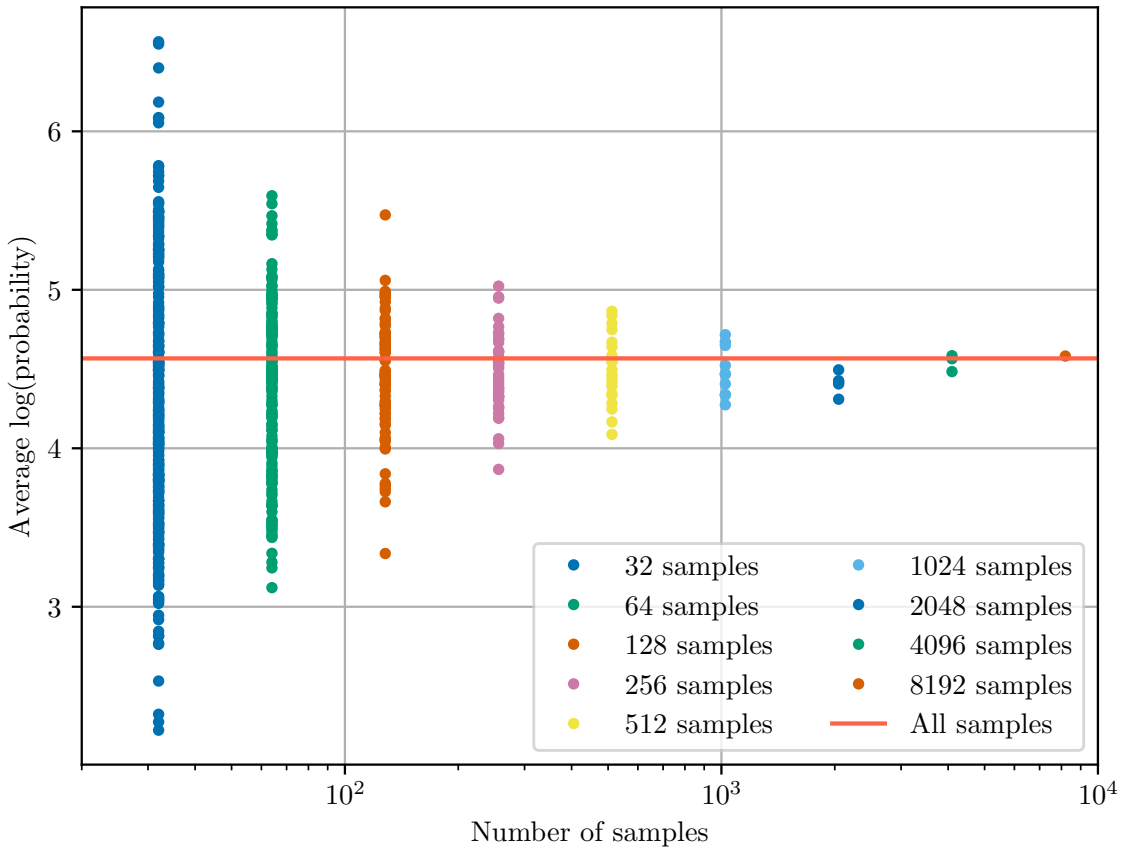


Figure B1: Plot of average log(probability) of $\mathbf{\Omega}_{\text{samp}}$ vs. number of samples when sampling $\mathbf{\Omega}_{\text{samp}} \sim p_{\text{cond}}(\mathbf{\Omega}|\theta)$ given different random iterations of numbers of PE samples θ associated to a simulated equation of state of medium stiffness. The average log(probability) found when using all ($\sim 10^4$) PE samples for the event is given by the solid horizontal line.

Figure B2 presents the time taken to perform MCMC in combining information from 4 GW merger events using a variety of numbers of PE samples per event. It is clear the time required to combine information from more than 10^3 events become prohibitive, as the analysis would take 3 times as long for a given number of events. To parallelise the analysis and make it as efficient as possible, we batch the GPU with all possible combinations of proposed equation of states and samples per event possible. This allows us to parallelise all probability evaluations that the Flow must perform on the GPU so the speed up the analysis. This process is discussed in chapter 6 section 6.1.1. However, this scales as number of events × number of samples × number of dimensions. Therefore, if we increase the number of samples per event drastically, we therefore sacrifice efficiency, and limit the number of events we can

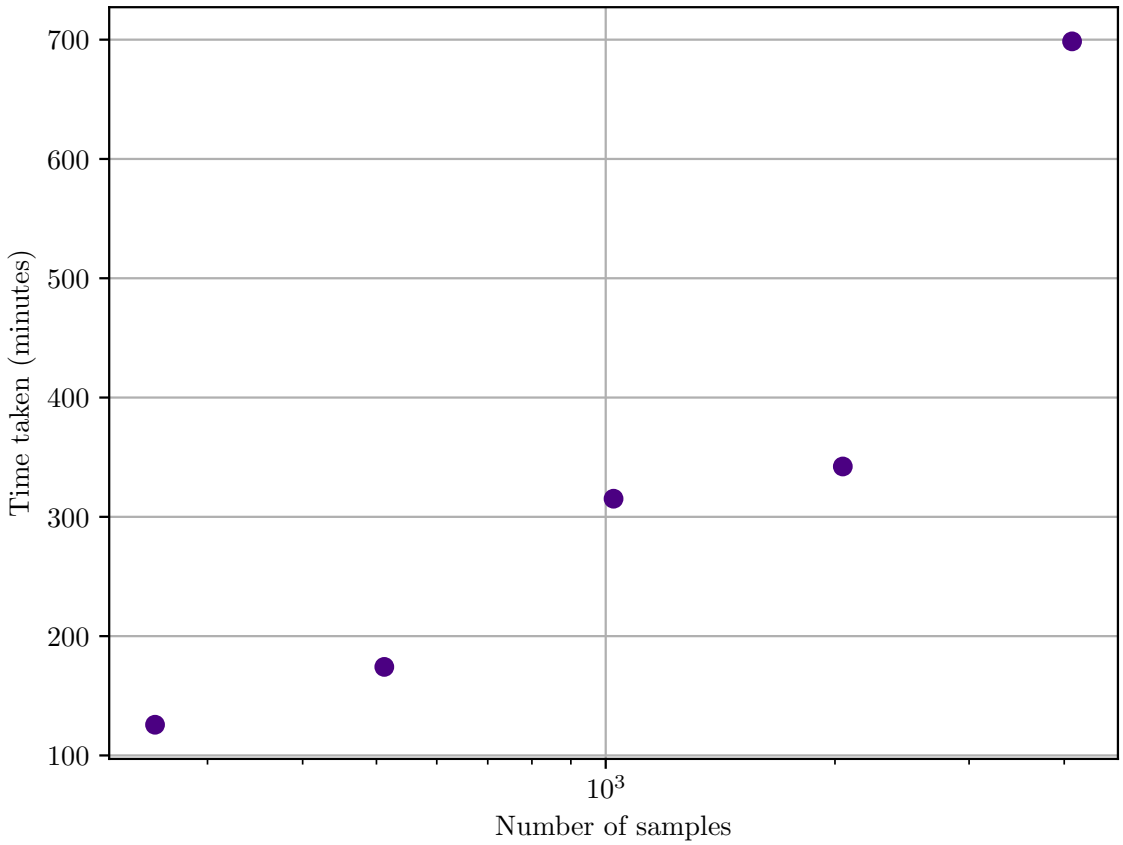


Figure B2: Plot of time taken vs. number of samples for various MCMC runs associated to the inference of a simulated equation of state of medium stiffness given information from 4 GW merger events with different numbers of samples per event.

combine information from quickly. This is of particular importance when we reach orders of magnitude of events upwards of 10^3. Therefore, we choose to use 256 samples per event so to remain as efficient as possible without inducing significant error. We find 512 samples per event may have also been acceptable for the analysis, but choose 256 so to maximise efficiency, without any significant loss in error induced in the Flow's probability evaluation.

Appendix C

Figure C1 illustrates the block diagrams of inputs and outputs to each of the 4 Flows used in chapter 6. These are: regeneration (RG), conditional (cond), prior (π) and PE. This is supplementary to the workflow diagrams in figures 6.1.1 and 6.3.6.

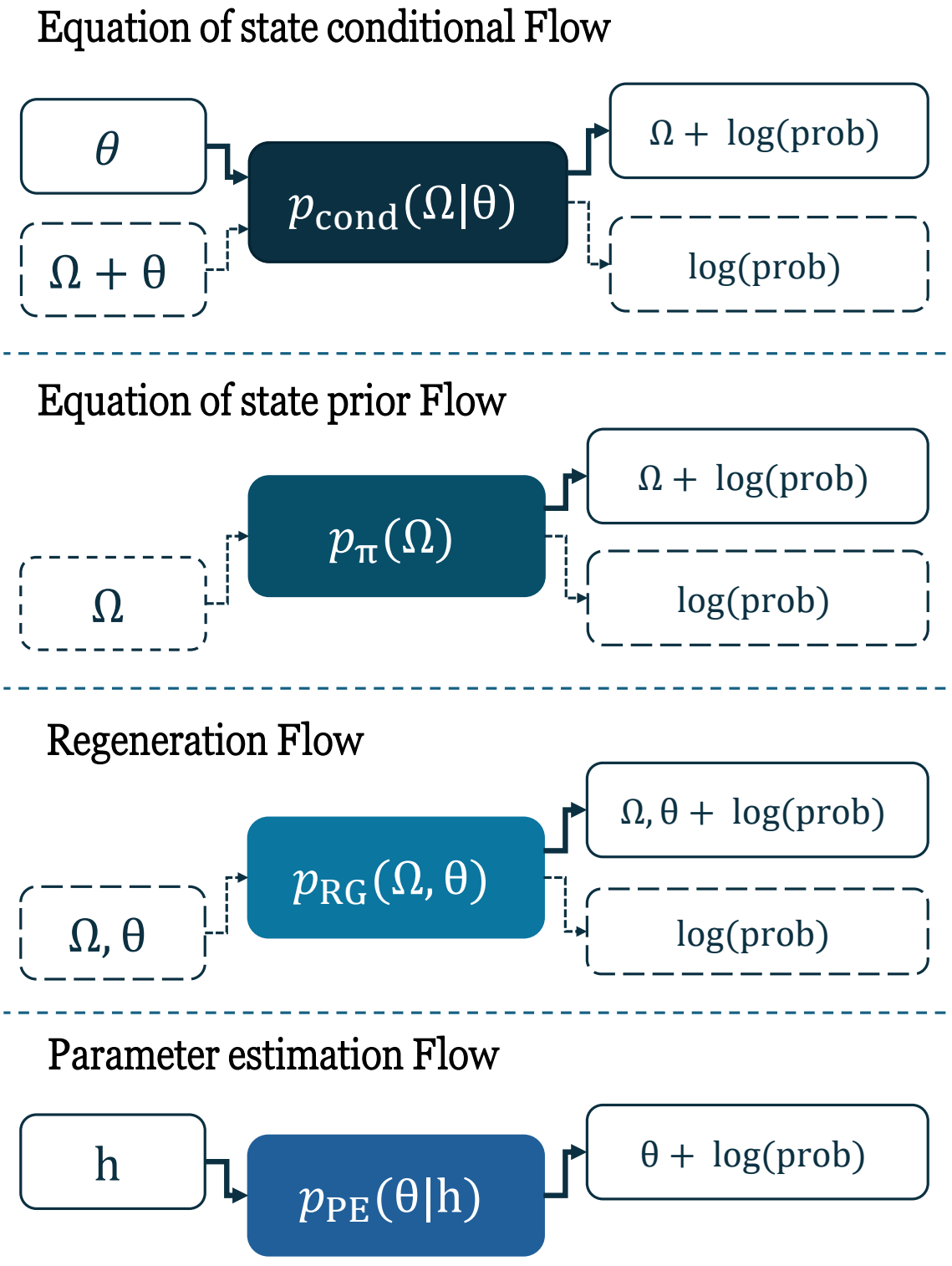


Figure C1: Block diagrams of inputs and outputs of the 4 Normalising Flows used in chapter 6. For conditional Normalising Flows, dotted lines represent evaluation of a data space location given a condition, while solid lines represent conditional sampling. For unconditional Flows, the bold line represents randomly sampling the Flow, and dotted lines represent evaluation of a given data space location.

Bibliography

[1] J. McGinn, A. Mukherjee, J. Irwin, C. Messenger, M. J. Williams, and I. S. Heng, "Rapid neutron star equation of state inference with normalising flows," 2024.

[2] P. J. Davis, H. Dinh Thi, A. F. Fantina, F. Gulminelli, M. Oertel, and L. Suleiman, "Inference of neutron-star properties with unified crust-core equations of state for parameter estimation," *Astronomy amp; Astrophysics*, vol. 687, p. A44, June 2024.

[3] Q. Hu, J. Irwin, Q. Sun, C. Messenger, L. Suleiman, I. S. Heng, and J. Veitch, "Decoding long-duration gravitational waves from binary neutron stars with machine learning: Parameter estimation and equations of state," 2024.

[4] B. P. Abbott et al, "GW170817: Measurements of Neutron Star Radii and Equation of State," *Phys. Rev. Lett.*, vol. 121, p. 161101, Oct 2018.

[5] J. Chadwick, "The existence of a neutron," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 136, no. 830, pp. 692–708, 1932.

[6] W. Baade and F. Zwicky, "Remarks on super-novae and cosmic rays," *Phys. Rev.*, vol. 46, pp. 76–77, Jul 1934.

[7] L. D. Landau, "On the theory of stars," *Phys. Z. Sowjetunion*, vol. 1, p. 285, 1932.

[8] J. M. Irvine, *Neutron Stars*. Oxford: Clarendon Press, 1978.

[9] D. G. Yakovlev, P. Haensel, G. Baym, and C. Pethick, "Lev landau and the concept of neutron stars," *Physics-Uspekhi*, vol. 56, p. 289–295, Mar. 2013.

[10] G. Shaviv, *The Life of Stars: The Controversial Inception and Emergence of the Theory of Stellar Structure*. Springer Berlin, 01 2009.

[11] A. Einstein, "Die grundlage der allgemeinen relativitätstheorie," *Annalen der Physik*, vol. 354, no. 7, pp. 769–822, 1916.

[12] M. Ishak, "Testing general relativity in cosmology," *Living Reviews in Relativity*, vol. 22, Dec. 2018.

[13] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation*. San Francisco: W. H. Freeman, 1973.

[14] P. A. Dirac, *General Theory of Relativity*. Princeton: Princeton University Press, 1996.

[15] F. J. Fattoyev, "Neutron stars in general relativity and scalar-tensor theory of gravity," *Arabian Journal of Mathematics*, vol. 8, p. 293–304, July 2019.

[16] B. Li, C. Ji, W. Yang, J. Wang, K. Yang, R. Xu, W. Liu, Z. Cai, J. Chen, and H. kwang Mao, "Diamond anvil cell behavior up to 4 mbar," *Proceedings of the National Academy of Sciences*, vol. 115, no. 8, pp. 1713–1717, 2018.

[17] S. Y. Gus'kov, P. A. Kuchugov, R. A. Yakhin, and N. V. Zmitrenko, "Gigabar shock wave driven by laser-accelerated electron stream," *Plasma Physics and Controlled Fusion*, vol. 64, p. 045001, feb 2022.

[18] J. R. Oppenheimer and G. M. Volkoff, "On massive neutron cores," *Phys. Rev.*, vol. 55, pp. 374–381, Feb 1939.

[19] R. C. Tolman, "Static solutions of einstein's field equations for spheres of fluid," *Phys. Rev.*, vol. 55, pp. 364–373, Feb 1939.

[20] S. Tsuruta, "Temperature of neutron stars," in *14th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories*, vol. 1, pp. 160–179, 2017.

[21] L. Lindblom, "Determining the Nuclear Equation of State from Neutron-Star Masses and Radii," , vol. 398, p. 569, Oct. 1992.

[22] J. Piekarewicz, *Neutron Star Matter Equation of State*, pp. 1–20. Cham: Springer International Publishing, 2016.

[23] L. S. Rocha, J. E. Horvath, L. M. de Sá, G. Y. Chinen, L. G. Barão, and M. G. B. de Avellar, "Mass distribution and maximum mass of neutron stars: Effects of orbital inclination angle," 2023.

[24] M. Large, A. Vaughan, and B. Mills, "A pulsar supernova association?," *Nature*, vol. 220, pp. 340–341, 1968.

[25] V. Doroshenko, V. Suleimanov, G. Pühlhofer, and A. Santangelo, "A strangely light neutron star within a supernova remnant," *Nature Astronomy*, vol. 6, pp. 1444–1451, 2022.

[26] J. Piekarewicz, "The nuclear physics of neutron stars," 2022.

[27] S. B. Rüster, M. Hempel, and J. Schaffner-Bielich, "Outer crust of nonaccreting cold neutron stars," *Phys. Rev. C*, vol. 73, p. 035804, Mar 2006.

[28] K. Iida and K. Oyamatsu, "Symmetry energy, unstable nuclei and neutron star crusts," *The European Physical Journal A*, vol. 50, Feb. 2014.

[29] M. E. Caplan, A. S. Schneider, and C. J. Horowitz, "Elasticity of nuclear pasta," *Phys. Rev. Lett.*, vol. 121, p. 132701, Sep 2018.

[30] M.-a. Hashimoto, H. Seki, and M. Yamada, "Shape of nuclei in the crust of neutron star," *Progress of Theoretical Physics*, vol. 71, pp. 320–326, 02 1984.

[31] D. G. Ravenhall, C. J. Pethick, and J. R. Wilson, "Structure of matter below nuclear saturation density," *Phys. Rev. Lett.*, vol. 50, pp. 2066–2069, Jun 1983.

[32] J. A. Pons, D. Viganò, and N. Rea, "A highly resistive layer within the crust of x-ray pulsars limits their spin periods," *Nature Physics*, vol. 9, p. 431–434, June 2013.

[33] M. Oertel, M. Hempel, T. Klähn, and S. Typel, "Equations of state for supernovae and compact stars," *Reviews of Modern Physics*, vol. 89, Mar. 2017.

[34] N. Chamel and P. Haensel, "Physics of neutron star crusts," *Living Reviews in Relativity*, vol. 11, no. 10, 2008.

[35] V. M. Kaspi, J. R. Lackey, and D. Chakrabarty, "A glitch in an anomalous x-ray pulsar," *The Astrophysical Journal*, vol. 537, p. L31, jun 2000.

[36] B. Haskell and K. Schwenzer, "Gravitational waves from isolated neutron stars," 2021.

[37] B. P. Abbott et al., "Observation of gravitational waves from a binary black hole merger," *Phys. Rev. Lett.*, vol. 116, p. 061102, Feb 2016.

[38] R. Abbott et al., "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run," *Physical Review X*, vol. 13, Dec. 2023.

[39] The LIGO Scientific Collaboration and the Virgo Collaboration and the KAGRA Collaboration and A. G. Abac et al., "GWTC-4.0: Updating the Gravitational-Wave Transient Catalog with Observations from the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run," 2025.

[40] D. F. Styer, "The geometrical significance of the laplacian," *American Journal of Physics*, vol. 83, no. 12, pp. 992–997, 2015.

[41] B. S. Sathyaprakash and B. F. Schutz, "Physics, astrophysics and cosmology with gravitational waves," *Living Reviews in Relativity*, vol. 12, no. 2, pp. 1–78, 2009.

[42] N. Williams, *Gravitational Wave Modelling and Analysis for Binary Neutron Star Inspirals*. Phd thesis, University of Birmingham, 2025.

[43] J. Aasi et al., "Advanced ligo," *Classical and Quantum Gravity*, vol. 32, p. 074001, Mar. 2015.

[44] F. Acernese, "Advanced virgo: a second-generation interferometric gravitational wave detector," *Classical and Quantum Gravity*, vol. 32, p. 024001, Dec. 2014.

[45] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto, "Interferometer design of the kagra gravitational wave detector," *Phys. Rev. D*, vol. 88, p. 043007, Aug 2013.

[46] D. Shoemaker, "Advanced ligo reference design," Tech. Rep. LIGO-M060056-v2, LIGO Laboratory / LIGO Scientific Collaboration, March 2011. Internal working note.

[47] B. P. Abbott at al, "Ligo: the laser interferometer gravitational-wave observatory," *Reports on Progress in Physics*, vol. 72, p. 076901, June 2009.

[48] B. J. Meers, "Recycling in laser-interferometric gravitational-wave detectors," *Phys. Rev. D*, vol. 38, pp. 2317–2326, Oct 1988.

[49] G. Billingsley, H. Yamamoto, and L. Zhang, "Characterization of advanced ligo core optics," in *Proceedings of the American Society for Precision Engineering (ASPE) Spring Topical Meeting*, vol. 66, pp. 78–83, 2017. LIGO Document Number: LIGO-P1700029-v5.

[50] *Gravitational-Wave Detectors*, ch. 6, pp. 197–267. John Wiley Sons, Ltd, 2011.

[51] LIGO Laboratory, "What is an interferometer?." https://www.ligo.caltech.edu/page/what-is-interferometer. Accessed: 2025-04-29.

[52] A. Kembhavi and P. Khare, *Gravitational Waves: A New Window to the Universe*. 01 2020.

[53] B. P. Abbott et al., "GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs," *Phys. Rev. X*, vol. 9, p. 031040, Sep 2019.

[54] B. P. Abbott et al, "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral," *Phys. Rev. Lett.*, vol. 119, p. 161101, Oct 2017.

[55] The LIGO Scientific Collaboration and the Virgo Collaboration and the KAGRA Collaboration and A. G. Abac et al., "Observation of gravitational waves from the coalescence of a 2.5-4.5 $m_\odot$ compact object and a neutron star," 2024.

[56] C. S. Unnikrishnan , "Indigo and ligo-india: Scope and plans for gravitational wave research and precision metrology in india," *International Journal of Modern Physics D*, vol. 22, no. 01, p. 1341010, 2013.

[57] K. Ackley et al., "Neutron star extreme matter observatory: A kilohertz-band gravitational-wave detector in the global network," *Publications of the Astronomical Society of Australia*, vol. 37, Nov. 2020.

[58] M. Maggiore, C. V. D. Broeck, N. Bartolo, E. Belgacem, D. Bertacca, M. A. Bizouard, M. Branchesi, S. Clesse, S. Foffa, J. García-Bellido, S. Grimm, J. Harms, T. Hinderer, S. Matarrese, C. Palomba, M. Peloso, A. Ricciardone, and M. Sakellariadou, "Science case for the Einstein telescope," *Journal of Cosmology and Astroparticle Physics*, vol. 2020, p. 050, mar 2020.

[59] M. Evans, R. X. Adhikari, C. Afle, S. W. Ballmer, S. Biscoveanu, S. Borhanian, D. A. Brown, Y. Chen, R. Eisenstein, A. Gruson, A. Gupta, E. D. Hall, R. Huxford, B. Kamai, R. Kashyap, J. S. Kissel, K. Kuns, P. Landry, A. Lenon, G. Lovelace, L. McCuller, K. K. Y. Ng, A. H. Nitz, J. Read, B. S. Sathyaprakash, D. H. Shoemaker, B. J. J. Slagmolen, J. R. Smith, V. Srivastava, L. Sun, S. Vitale, and R. Weiss, "A horizon study for cosmic explorer: Science, observatories, and community," 2021.

[60] T. Islam, J. Roulet, and T. Venumadhav, "Factorized parameter estimation for real-time gravitational wave inference," 2022.

[61] F. Pretorius, "Evolution of binary black-hole spacetimes," *Physical Review Letters*, vol. 95, Sept. 2005.

[62] J. G. Baker, J. R. van Meter, S. T. McWilliams, J. Centrella, and B. J. Kelly, "Consistency of post-newtonian waveforms with numerical relativity," *Phys. Rev. Lett.*, vol. 99, p. 181101, Oct 2007.

[63] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, J. T. Whelan, B. Brügmann, P. Diener, J. Gonzalez, M. Hannam, S. Husa, M. Koppitz, D. Pollney, L. Rezzolla, L. Santamaría, A. M. Sintes, U. Sperhake, and J. Thornburg, "A phenomenological template family for black-hole coalescence waveforms," *Classical and Quantum Gravity*, vol. 24, p. S689–S699, Sept. 2007.

[64] P. Schmidt, "Gravitational waves from binary black hole mergers: Modeling and observations," *Frontiers in Astronomy and Space Sciences*, vol. Volume 7 - 2020, 2020.

[65] S. Ossokine, A. Buonanno, S. Marsat, R. Cotesta, S. Babak, T. Dietrich, R. Haas, I. Hinder, H. P. Pfeiffer, M. Pürrer, C. J. Woodford, M. Boyle, L. E. Kidder, M. A. Scheel, and B. Szilágyi, "Multipolar effective-one-body waveforms for precessing binary black holes: Construction and validation," *Physical Review D*, vol. 102, Aug. 2020.

[66] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, "Findchirp: An algorithm for detection of gravitational waves from inspiraling compact binaries," *Phys. Rev. D*, vol. 85, p. 122006, Jun 2012.

[67] S. Klimenko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. Prodi, C. Lazzaro, K. Ackley, S. Tiwari, C. Da Silva, and G. Mitselmakher, "Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors," *Physical Review D*, vol. 93, Feb. 2016.

[68] K. Chatziioannou, "Neutron-star tidal deformability and equation-of-state constraints," *General Relativity and Gravitation*, vol. 52, Nov. 2020.

[69] K. S. Thorne, "Tidal stabilization of rigidly rotating, fully relativistic neutron stars," *Phys. Rev. D*, vol. 58, p. 124031, Nov 1998.

[70] T. Hinderer, "Tidal love numbers of neutron stars," *The Astrophysical Journal*, vol. 677, p. 1216–1220, Apr. 2008.

[71] T. Binnington and E. Poisson, "Relativistic theory of tidal love numbers," *Physical Review D*, vol. 80, Oct. 2009.

[72] R. Abbott, "Diving below the Spin-down Limit: Constraints on Gravitational Waves from the Energetic Young Pulsar PSR J0537-6910," *The Astrophysical Journal Letters*, vol. 913, p. L27, May 2021.

[73] F. Gittins, "Gravitational waves from neutron-star mountains," *Classical and Quantum Gravity*, vol. 41, p. 043001, Jan. 2024.

[74] A. G. Abac, "Search for Continuous Gravitational Waves from Known Pulsars in the First Part of the Fourth LIGO-Virgo-KAGRA Observing Run," *The Astrophysical Journal*, vol. 983, p. 99, Apr. 2025.

[75] J. Aasi et al., "Narrow-band search of continuous gravitational-wave signals from crab and vela pulsars in virgo vsr4 data," *Physical Review D*, vol. 91, Jan. 2015.

[76] R. Abbott et al, "Searches for gravitational waves from known pulsars at two harmonics in the second and third ligo-virgo observing runs," *The Astrophysical Journal*, vol. 935, p. 1, aug 2022.

[77] E. Abdikamalov, G. Pagliaroli, and D. Radice, *Gravitational Waves from Core-Collapse Supernovae*, p. 1–37. Springer Singapore, 2021.

[78] M. Sieniawska and M. Bejger, "Continuous gravitational waves from neutron stars: Current status and prospects," *Universe*, vol. 5, no. 11, 2019.

[79] W. Del Pozzo, T. G. F. Li, M. Agathos, C. Van Den Broeck, and S. Vitale, "Demonstrating the feasibility of probing the neutron-star equation of state with second-generation gravitational-wave detectors," *Phys. Rev. Lett.*, vol. 111, p. 071101, Aug 2013.

[80] B. D. Lackey and L. Wade, "Reconstructing the neutron-star equation of state with gravitational-wave detectors from a realistic population of inspiralling binary neutron stars," *Phys. Rev. D*, vol. 91, p. 043002, Feb 2015.

[81] É. Flanagan and T. Hinderer, "Constraining neutron-star tidal love numbers with gravitational-wave detectors," *Physical Review D*, vol. 77, Jan. 2008.

[82] K. Yagi and N. Yunes, "Approximate universal relations among tidal parameters for neutron star binaries," *Classical and Quantum Gravity*, vol. 34, p. 015006, dec 2016.

[83] B. P. Abbott et al, "Properties of the binary neutron star merger gw170817," *Phys. Rev. X*, vol. 9, p. 011001, Jan 2019.

[84] J. E. Horvath and R. Valentim, *The Masses of Neutron Stars*, p. 1317–1330. Springer International Publishing, 2017.

[85] J. M. Lattimer, "The nuclear equation of state and neutron star masses," *Annual Review of Nuclear and Particle Science*, vol. 62, no. Volume 62, 2012, pp. 485–515, 2012.

[86] A. Gupta, D. Gerosa, K. G. Arun, E. Berti, W. M. Farr, and B. S. Sathyaprakash, "Black holes in the low-mass gap: Implications for gravitational-wave observations," *Phys. Rev. D*, vol. 101, p. 103036, May 2020.

[87] T. Narikawa, N. Uchikata, K. Kawaguchi, K. Kiuchi, K. Kyutoku, M. Shibata, and H. Tagoshi, "Reanalysis of the binary neutron star mergers GW170817 and GW190425 using numerical-relativity calibrated waveform models," *Physical Review Research*, vol. 2, Oct. 2020.

[88] A. Goldstein et al., "An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A," *The Astrophysical Journal Letters*, vol. 848, no. 2, p. L14, 2017.

[89] V. Savchenko, C. Ferrigno, E. Kuulkers, A. Bazzano, E. Bozzo, S. Brandt, J. Chenevez, T. J.-L. Courvoisier, R. Diehl, A. Domingo, L. Hanlon, E. Jourdain, A. von Kienlin, P. Laurent, F. Lebrun, A. Lutovinov, A. Martin-Carrillo, S. Mereghetti, L. Natalucci, J. Rodi, J.-P. Roques, R. Sunyaev, and P. Ubertini, "Integral detection of the first prompt gamma-ray signal coincident with the gravitational wave event gw170817," *The Astrophysical Journal Letters*, vol. 848, no. 2, p. L15, 2017.

[90] B. P. Abbott et al, "Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A," *The Astrophysical Journal Letters*, vol. 848, p. L13, oct 2017.

[91] M. W. Coughlin, T. Dietrich, B. Margalit, and B. D. Metzger, "Multimessenger bayesian parameter inference of a binary neutron star merger," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 489, p. L91–L96, Aug. 2019.

[92] B. Margalit and B. D. Metzger, "Constraining the Maximum Mass of Neutron Stars from Multi-messenger Observations of GW170817," *The Astrophysical Journal Letters*, vol. 850, p. L19, Nov. 2017.

[93] C. L. Fryer, K. Belczynski, E. Ramirez-Ruiz, S. Rosswog, G. Shen, and A. W. Steiner, "The fate of the compact remnant in neutron star mergers," *The Astrophysical Journal*, vol. 812, p. 24, Oct. 2015.

[94] M. Nicholl, B. Margalit, P. Schmidt, G. P. Smith, E. J. Ridley, and J. Nuttall, "Tight multimessenger constraints on the neutron star equation of state from GW170817 and a forward model for kilonova light-curve synthesis," *Monthly Notices of the Royal Astronomical Society*, vol. 505, pp. 3016–3032, 05 2021.

[95] B. P. Abbott et al, "GW190425: Observation of a Compact Binary Coalescence with Total Mass 3.4 M," *The Astrophysical Journal Letters*, vol. 892, p. L3, Mar. 2020.

[96] C. A. Raithel, "Constraints on the neutron star equation of state from GW170817," *The European Physical Journal A*, vol. 55, no. 80, 2019.

[97] V. Kalogera, C. Kim, D. R. Lorimer, M. Burgay, N. D'Amico, A. Possenti, R. N. Manchester, A. G. Lyne, B. C. Joshi, M. A. McLaughlin, M. Kramer, J. M. Sarkissian, and F. Camilo, "The cosmic coalescence rates for double neutron star binaries," *The Astrophysical Journal*, vol. 614, p. L137–L138, Oct. 2004.

[98] R. O'Shaughnessy, C. Kim, V. Kalogera, and K. Belczynski, "Constraining population synthesis models via empirical binary compact object merger and supernova rates," *The Astrophysical Journal*, vol. 672, p. 479–488, Jan. 2008.

[99] N. Pol, M. McLaughlin, and D. R. Lorimer, "Future Prospects for Ground-based Gravitational-wave Detectors: The Galactic Double Neutron Star Merger Rate Revisited," , vol. 870, p. 71, Jan. 2019.

[100] A. Hewish, S. J. Bell, J. D. H. Pilkington, P. F. Scott, and R. A. Collins, "Observation of a rapidly pulsating radio source," *Nature*, vol. 217, pp. 709–713, 1968.

[101] K. Chatziioannou, H. T. Cromartie, S. Gandolfi, I. Tews, D. Radice, A. W. Steiner, and A. L. Watts, "Neutron stars and the dense matter equation of state: from microscopic theory to macroscopic observations," 2024.

[102] R. A. Hulse and J. H. Taylor, "Discovery of a pulsar in a binary system.," , vol. 195, pp. L51–L53, Jan. 1975.

[103] T. J.-L. Courvoisier, *The Hulse-Taylor Pulsar and Gravitational Radiation*, pp. 225–233. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

[104] The Royal Swedish Academy of Sciences, "Press release: The 1993 Nobel Prize in Physics." https://www.nobelprize.org/prizes/physics/1993/press-release/, 1993. Accessed: 2025-04-08.

[105] S. E. Thorsett and D. Chakrabarty, "Neutron Star Mass Measurements. I. Radio Pulsars," , vol. 512, pp. 288–299, Feb. 1999.

[106] P. B. Demorest, T. Pennucci, S. M. Ransom, M. S. E. Roberts, and J. W. T. Hessels, "A two-solar-mass neutron star measured using shapiro delay," *Nature*, vol. 467, p. 1081–1083, Oct. 2010.

[107] H. Cromartie, E. Fonseca, S. Ransom, P. Demorest, Z. Arzoumanian, H. Blumer, P. Brook, M. DeCesar, T. Dolch, J. Ellis, R. Ferdman, E. Ferrara, N. Garver-Daniels, P. Gentile, M. Jones, M. Lam, D. Lorimer, R. Lynch, M. McLaughlin, C. Ng, D. Nice, T. Pennucci, R. Spiewak, I. Stairs, K. Stovall, J. Swiggum, and W. Zhu, "Relativistic shapiro delay measurements of an extremely massive millisecond pulsar," *Nature Astronomy*, vol. 4, pp. 72–76, 2020.

[108] M. Kramer, I. H. Stairs, R. N. Manchester, M. A. McLaughlin, A. G. Lyne, R. D. Ferdman, M. Burgay, D. R. Lorimer, A. Possenti, N. D'Amico, J. M. Sarkissian, G. B. Hobbs, J. E. Reynolds, P. C. C. Freire, and F. Camilo, "Tests of general relativity from timing the double pulsar," *Science*, vol. 314, p. 97–102, Oct. 2006.

[109] J. G. Martinez, K. Stovall, P. C. C. Freire, J. S. Deneva, F. A. Jenet, M. A. McLaughlin, M. Bagchi, S. D. Bates, and A. Ridolfi, "Pulsar j0453+1559: A double neutron star system with a large mass asymmetry," *The Astrophysical Journal*, vol. 812, p. 143, Oct. 2015.

[110] P. Podsiadlowski, J. D. M. Dewi, P. Lesaffre, J. C. Miller, W. G. Newton, and J. R. Stone, "The double pulsar j0737-3039: testing the neutron star equation of state," *Monthly Notices of the Royal Astronomical Society*, vol. 361, pp. 1243–1249, 08 2005.

[111] D. J. Nice, E. M. Splaver, I. H. Stairs, O. Löhmer, A. Jessner, M. Kramer, and J. M. Cordes, "A 2.1 M_{solar} Pulsar Measured by Relativistic Orbital Decay," , vol. 634, pp. 1242–1249, Dec. 2005.

[112] P. C. C. Freire, A. Wolszczan, M. van den Berg, and J. W. T. Hessels, "A massive neutron star in the globular cluster m5," *The Astrophysical Journal*, vol. 679, p. 1433–1442, June 2008.

[113] R. W. Romani, D. Kandel, A. V. Filippenko, T. G. Brink, and W. Zheng, "Psr j09520607: The fastest and heaviest known galactic neutron star," *The Astrophysical Journal Letters*, vol. 934, p. L17, July 2022.

[114] R. P. Breton, S. A. Rappaport, M. H. van Kerkwijk, and J. A. Carter, "Koi 1224: A fourth bloated hot white dwarf companion found with kepler," *The Astrophysical Journal*, vol. 748, p. 115, mar 2012.

[115] A. Y. Potekhin, "Atmospheres and radiating surfaces of neutron stars," *Physics-Uspekhi*, vol. 57, p. 735–770, Aug. 2014.

[116] H. Sotani, K. Iida, K. Oyamatsu, and A. Ohnishi, "Mass and radius formulas for low-mass neutron stars," *Progress of Theoretical and Experimental Physics*, vol. 2014, p. 051E01, 05 2014.

[117] F. Özel, "Soft equations of state for neutron-star matter ruled out by exo 0748-676," *Nature*, vol. 441, pp. 1115–1117, 2006.

[118] J. Van Paradijs, "Possible observational constraints on the mass-radius relation of neutron stars," *ApJ*, vol. 234, 1979. Cited by: 70.

[119] J. M. Lattimer and M. Prakash, "The equation of state of hot, dense matter and neutron stars," *Physics Reports*, vol. 621, pp. 127–164, 2016. Memorial Volume in Honor of Gerald E. Brown.

[120] F. Özel and P. Freire, "Masses, radii, and the equation of state of neutron stars," *Annual Review of Astronomy and Astrophysics*, vol. 54, no. Volume 54, 2016, pp. 401–440, 2016.

[121] 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," *The Astrophysical Journal Letters*, vol. 887, p. L24, dec 2019.

[122] T. E. Riley, A. L. Watts, S. Bogdanov, P. S. Ray, R. M. Ludlam, S. Guillot, Z. Arzoumanian, C. L. Baker, A. V. Bilous, D. Chakrabarty, K. C. Gendreau, A. K. Harding, W. C. G. Ho, J. M. Lattimer, S. M. Morsink, and T. E. Strohmayer, "A nicer view of psr j0030+0451: Millisecond pulsar parameter estimation," *The Astrophysical Journal Letters*, vol. 887, p. L21, dec 2019.

[123] T. E. Riley, A. L. Watts, P. S. Ray, S. Bogdanov, S. Guillot, S. M. Morsink, A. V. Bilous, Z. Arzoumanian, D. Choudhury, J. S. Deneva, K. C. Gendreau, A. K. Harding, W. C. G. Ho, J. M. Lattimer, M. Loewenstein, R. M. Ludlam,

C. B. Markwardt, T. Okajima, C. Prescod-Weinstein, R. A. Remillard, M. T. Wolff, E. Fonseca, H. T. Cromartie, M. Kerr, T. T. Pennucci, A. Parthasarathy, S. Ransom, I. Stairs, L. Guillemot, and I. Cognard, "A nicer view of the massive pulsar psr j0740+6620 informed by radio timing and xmm-newton spectroscopy," *The Astrophysical Journal Letters*, vol. 918, p. L27, sep 2021.

[124] A. L. Watts, N. Andersson, D. Chakrabarty, M. Feroci, K. Hebeler, G. Israel, F. K. Lamb, M. C. Miller, S. Morsink, F. Özel, A. Patruno, J. Poutanen, D. Psaltis, A. Schwenk, A. W. Steiner, L. Stella, L. Tolos, and M. van der Klis, "Colloquium: Measuring the neutron star equation of state using x-ray timing," *Rev. Mod. Phys.*, vol. 88, p. 021001, Apr 2016.

[125] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade, "Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library," *Physical Review D*, vol. 91, Feb. 2015.

[126] E. Thrane and C. Talbot, "An introduction to bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models," *Publications of the Astronomical Society of Australia*, vol. 36, 2019.

[127] T. B. Littenberg and N. J. Cornish, "Bayesian inference for spectral estimation of gravitational wave detector noise," *Phys. Rev. D*, vol. 91, p. 084034, Apr 2015.

[128] E. T. Jaynes, *Ignorance priors and transformation groups*, p. 372–396. Cambridge University Press, 2003.

[129] S. J. Magnall, C. Ecker, L. Rezzolla, P. D. Lasky, and S. R. Goode, "Physics-informed priors improve gravitational-wave constraints on neutron-star matter," 2025.

[130] P. T. H. Pang, T. Dietrich, M. W. Coughlin, M. Bulla, I. Tews, M. Almualla, T. Barna, R. W. Kiendrebeogo, N. Kunert, G. Mansingh, B. Reed, N. Sravan,

A. Toivonen, S. Antier, R. O. VandenBerg, J. Heinzel, V. Nedora, P. Salehi, R. Sharma, R. Somasundaram, and C. V. D. Broeck, "An updated nuclear-physics and multi-messenger astrophysics framework for binary neutron star mergers," *Nature Communications*, vol. 14, no. 1, p. 8352, 2023.

[131] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[132] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 04 1970.

[133] S. Sharma, "Markov chain monte carlo methods for bayesian data analysis in astronomy," *Annual Review of Astronomy and Astrophysics*, vol. 55, p. 213–259, Aug. 2017.

[134] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "<tt>emcee</tt>: The mcmc hammer," *Publications of the Astronomical Society of the Pacific*, vol. 125, p. 306–312, Mar. 2013.

[135] S. Altiparmak, C. Ecker, and L. Rezzolla, "On the sound speed in neutron stars," *The Astrophysical Journal Letters*, vol. 939, p. L34, nov 2022.

[136] P. Bedaque and A. W. Steiner, "Sound velocity bound and neutron stars," *Phys. Rev. Lett.*, vol. 114, p. 031103, Jan 2015.

[137] R. Essick, I. Legred, K. Chatziioannou, S. Han, and P. Landry, "Phase transition phenomenology with nonparametric representations of the neutron star equation of state," *Phys. Rev. D*, vol. 108, p. 043013, Aug 2023.

[138] K. Hebeler, J. M. Lattimer, C. J. Pethick, and A. Schwenk, "Equation of state and neutron star properties constrained by nuclear physics and observation," *The Astrophysical Journal*, vol. 773, p. 11, ju 2013.

[139] S. K. Greif, G. Raaijmakers, K. Hebeler, A. Schwenk, and A. L. Watts, "Equation of state sensitivities when inferring neutron star and dense matter properties,"

Monthly Notices of the Royal Astronomical Society, vol. 485, pp. 5363–5376, 03 2019.

[140] N. Jokela, M. Järvinen, and J. Remes, "Holographic qcd in the veneziano limit and neutron stars," *Journal of High Energy Physics*, vol. 2019, Mar. 2019.

[141] S. Typel, G. Röpke, T. Klähn, D. Blaschke, and H. H. Wolter, "Composition and thermodynamics of nuclear matter with light clusters," *Phys. Rev. C*, vol. 81, p. 015803, Jan 2010.

[142] N. Jokela, M. Järvinen, G. Nijs, and J. Remes, "Unified weak and strong coupling framework for nuclear matter and neutron stars," *Phys. Rev. D*, vol. 103, p. 086004, Apr 2021.

[143] E. Chabanat, P. Bonche, P. Haensel, J. Meyer, and R. Schaeffer, "A skyrme parametrization from subnuclear to neutron star densities part ii. nuclei far from stabilities," *Nuclear Physics A*, vol. 635, no. 1, pp. 231–256, 1998.

[144] M. Bender, P.-H. Heenen, and P.-G. Reinhard, "Self-consistent mean-field models for nuclear structure," *Rev. Mod. Phys.*, vol. 75, pp. 121–180, Jan 2003.

[145] Douchin, F. and Haensel, P., "A unified equation of state of dense matter and neutron star structure," *AA*, vol. 380, no. 1, pp. 151–167, 2001.

[146] J. S. Read, B. D. Lackey, B. J. Owen, and J. L. Friedman, "Constraints on a phenomenologically parametrized neutron-star equation of state," *Phys. Rev. D*, vol. 79, p. 124032, Jun 2009.

[147] M. F. Carney, L. E. Wade, and B. S. Irwin, "Comparing two models for measuring the neutron star equation of state from gravitational-wave signals," *Phys. Rev. D*, vol. 98, p. 063004, Sep 2018.

[148] L. Lindblom, "Spectral representations of neutron-star equations of state," *Phys. Rev. D*, vol. 82, p. 103011, Nov 2010.

[149] J. Margueron, R. Hoffmann Casali, and F. Gulminelli, "Equation of state for dense nucleonic matter from metamodeling. i. foundational aspects," *Phys. Rev. C*, vol. 97, p. 025805, Feb 2018.

[150] F. Collaboration, W. Reisdorf, Y. Leifels, A. Andronic, R. Averbeck, V. Barret, Z. Basrak, N. Bastid, M. L. Benabderrahmane, R. Caplar, P. Crochet, P. Dupieux, M. Dzelalija, Z. Fodor, P. Gasik, Y. Grishkin, O. N. Hartmann, N. Herrmann, K. D. Hildenbrand, B. Hong, T. I. Kang, J. Kecskemeti, Y. J. Kim, M. Kirejczyk, M. Kis, P. Koczon, M. Korolija, R. Kotte, T. Kress, A. Lebedev, X. Lopez, T. Matulewicz, M. Merschmeyer, W. Neubert, M. Petrovici, K. Piasecki, F. Rami, M. S. Ryu, A. Schuettauf, Z. Seres, B. Sikora, K. S. Sim, V. Simion, K. Siwek-Wilczynska, V. Smolyankin, M. Stockmeier, G. Stoicea, Z. Tyminski, K. Wisniewski, D. Wohlfarth, Z. G. Xiao, H. S. Xu, I. Yushmanov, and A. Zhilin, "Systematics of azimuthal asymmetries in heavy ion collisions in the 1 a gev regime," 2011.

[151] P. Russotto et al, "The ASY-EOS experiment at GSI: investigating the symmetry energy at supra-saturation densities," *Journal of Physics: Conference Series*, vol. 420, p. 012092, Mar. 2013.

[152] J. M. Mammei, "The pb radius experiment (prex)," *AIP Conference Proceedings*, vol. 1560, pp. 620–622, 10 2013.

[153] D. Adhikari et al., "Precision determination of the neutral weak form factor of ^{48}Ca," *Phys. Rev. Lett.*, vol. 129, p. 042501, Jul 2022.

[154] M. B. Tsang, J. R. Stone, F. Camera, P. Danielewicz, S. Gandolfi, K. Hebeler, C. J. Horowitz, J. Lee, W. G. Lynch, Z. Kohley, R. Lemmon, P. Möller, T. Murakami, S. Riordan, X. Roca-Maza, F. Sammarruca, A. W. Steiner, I. Vidaña, and S. J. Yennello, "Constraints on the symmetry energy and neutron skins from experiments and theory," *Phys. Rev. C*, vol. 86, p. 015803, Jul 2012.

[155] J. M. Lattimer, "Constraints on the nuclear symmetry energy from experiments, theory and observations," *Journal of Physics: Conference Series*, vol. 2536, p. 012009, jun 2023.

[156] H. Shen, F. Ji, J. Hu, and K. Sumiyoshi, "Effects of symmetry energy on the equation of state for simulations of core-collapse supernovae and neutron-star mergers," *The Astrophysical Journal*, vol. 891, p. 148, mar 2020.

[157] C. J. Horowitz and J. Piekarewicz, "Neutron star structure and the neutron radius of $^{208}\boldsymbol{pb}$," *Phys. Rev. Lett.*, vol. 86, pp. 5647–5650, Jun 2001.

[158] X. Roca-Maza, M. Centelles, X. Viñas, and M. Warda, "Neutron skin of ^{208}Pb, nuclear symmetry energy, and the parity radius experiment," *Phys. Rev. Lett.*, vol. 106, p. 252501, Jun 2011.

[159] S. Ng, I. Legred, L. Suleiman, P. Landry, L. Traylor, and J. Read, "Inferring the neutron star equation of state with nuclear-physics informed semiparametric models," 2025.

[160] F. Özel and D. Psaltis, "Reconstructing the neutron-star equation of state from astrophysical measurements," *Phys. Rev. D*, vol. 80, p. 103003, Nov 2009.

[161] J. Antoniadis, P. C. C. Freire, N. Wex, T. M. Tauris, R. S. Lynch, M. H. van Kerkwijk, M. Kramer, C. Bassa, V. S. Dhillon, T. Driebe, J. W. T. Hessels, V. M. Kaspi, V. I. Kondratiev, N. Langer, T. R. Marsh, M. A. McLaughlin, T. T. Pennucci, S. M. Ransom, I. H. Stairs, J. van Leeuwen, J. P. W. Verbiest, and D. G. Whelan, "A massive pulsar in a compact relativistic binary," *Science*, vol. 340, no. 6131, p. 1233232, 2013.

[162] P. B. Demorest, T. Pennucci, S. M. Ransom, M. S. E. Roberts, and J. W. T. Hessels, "A two-solar-mass neutron star measured using shapiro delay," *Nature*, vol. 467, no. 7319, pp. 1081–1083, 2010.

[163] S. Guillot, M. Kerr, P. S. Ray, S. Bogdanov, S. Ransom, J. S. Deneva, Z. Arzoumanian, P. Bult, D. Chakrabarty, K. C. Gendreau, W. C. G. Ho, G. K. Jaisawal, C. Malacaria, M. C. Miller, T. E. Strohmayer, M. T. Wolff, K. S. Wood, N. A. Webb, L. Guillemot, I. Cognard, and G. Theureau, "Nicer x-ray observations of seven nearby rotation-powered millisecond pulsars," *The Astrophysical Journal Letters*, vol. 887, p. L27, dec 2019.

[164] A. Kedia, R. O'Shaughnessy, L. Wade, and A. Yelikar, "Exploring hidden priors when interpreting gravitational wave and electromagnetic probes of the nuclear equation of state," 2025.

[165] J. Golomb, I. Legred, K. Chatziioannou, and P. Landry, "Interplay of astrophysics and nuclear physics in determining the properties of neutron stars," *Phys. Rev. D*, vol. 111, p. 023029, Jan 2025.

[166] H. Koehn, H. Rose, P. T. H. Pang, R. Somasundaram, B. T. Reed, I. Tews, A. Abac, O. Komoltsev, N. Kunert, A. Kurkela, M. W. Coughlin, B. F. Healy, and T. Dietrich, "From existing and new nuclear and astrophysical constraints to stringent limits on the equation of state of neutron-rich dense matter," *Phys. Rev. X*, vol. 15, p. 021014, Apr 2025.

[167] P. Landry and J. S. Read, "The mass distribution of neutron stars in gravitational-wave binaries," *The Astrophysical Journal Letters*, vol. 921, p. L25, nov 2021.

[168] M. Favata, "Systematic parameter errors in inspiraling neutron star binaries," *Phys. Rev. Lett.*, vol. 112, p. 101101, Mar 2014.

[169] L. Wade, J. D. E. Creighton, E. Ochsner, B. D. Lackey, B. F. Farr, T. B. Littenberg, and V. Raymond, "Systematic and statistical errors in a bayesian approach to the estimation of the neutron-star equation of state using advanced gravitational wave detectors," *Phys. Rev. D*, vol. 89, p. 103012, May 2014.

[170] K. Yagi, L. C. Stein, G. Pappas, N. Yunes, and T. A. Apostolatos, "Why i-love-q: Explaining why universality emerges in compact objects," *Phys. Rev. D*, vol. 90, p. 063010, Sep 2014.

[171] G. Martinon, A. Maselli, L. Gualtieri, and V. Ferrari, "Rotating protoneutron stars: Spin evolution, maximum mass, and i-love-q relations," *Phys. Rev. D*, vol. 90, p. 064026, Sep 2014.

[172] K. Yagi and N. Yunes, "I-love-q: Unexpected universal relations for neutron stars and quark stars," *Science*, vol. 341, no. 6144, pp. 365–368, 2013.

[173] D. Bandyopadhyay, S. A. Bhat, P. Char, and D. Chatterjee, "Moment of inertia, quadrupole moment, love number of neutron star and their relations with strange-matter equations of state," *The European Physical Journal A*, vol. 54, no. 26, 2018.

[174] J. S. Read, C. Markakis, M. Shibata, K. b. o. Uryū, J. D. E. Creighton, and J. L. Friedman, "Measuring the neutron star equation of state with gravitational wave observations," *Phys. Rev. D*, vol. 79, p. 124033, Jun 2009.

[175] K. Chatziioannou, K. Yagi, A. Klein, N. Cornish, and N. Yunes, "Probing the internal composition of neutron stars with gravitational waves," *Phys. Rev. D*, vol. 92, p. 104008, Nov 2015.

[176] B. P. Abbott *et al.*, "Multi-messenger observations of a binary neutron star merger," *The Astrophysical Journal Letters*, vol. 848, no. 2, p. L12, 2017.

[177] B. P. Abbott et al, "Estimating the Contribution of Dynamical Ejecta in the Kilonova Associated with GW170817," *The Astrophysical Journal Letters*, vol. 850, p. L39, Dec. 2017.

[178] A. Albert et al., "Search for High-energy Neutrinos from Binary Neutron Star Merger GW170817 with ANTARES, IceCube, and the Pierre Auger Observatory," *The Astrophysical Journal*, vol. 850, p. L35, Nov. 2017.

[179] P. Landry and R. Essick, "Nonparametric inference of the neutron star equation of state from gravitational wave observations," *Phys. Rev. D*, vol. 99, p. 084049, Apr 2019.

[180] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 11 2005.

[181] R. Essick, P. Landry, and D. E. Holz, "Nonparametric inference of neutron star composition, equation of state, and maximum mass with GW170817," *Phys. Rev. D*, vol. 101, p. 063007, Mar 2020.

[182] P. Landry, R. Essick, and K. Chatziioannou, "Nonparametric constraints on neutron star matter with existing and upcoming gravitational wave and pulsar observations," *Phys. Rev. D*, vol. 101, p. 123007, Jun 2020.

[183] B. S. Sathyaprakash, B. F. Schutz, and C. Van Den Broeck, "Cosmography with the Einstein Telescope," *Classical and Quantum Gravity*, vol. 27, p. 215006, sep 2010.

[184] I. Mandel, "Parameter estimation on gravitational waves from multiple coalescing binaries," *Phys. Rev. D*, vol. 81, p. 084029, Apr 2010.

[185] J. Veitch and A. Vecchio, "Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network," *Phys. Rev. D*, vol. 81, p. 062003, Mar 2010.

[186] J. Abadie et al., "Predictions for the rates of compact binary coalescences observable by ground-based gravitational-wave detectors," *Classical and Quantum Gravity*, vol. 27, p. 173001, July 2010.

[187] J. Golomb and C. Talbot, "Hierarchical inference of binary neutron star mass distribution and equation of state with gravitational waves," *The Astrophysical Journal*, vol. 926, p. 79, feb 2022.

[188] C. Talbot, R. Smith, E. Thrane, and G. B. Poole, "Parallelized inference for gravitational-wave astronomy," , vol. 100, p. 043030, Aug. 2019.

[189] A. Ray, M. Camilo, J. Creighton, S. Ghosh, and S. Morisaki, "Rapid hierarchical inference of neutron star equation of state from multiple gravitational wave observations of binary neutron star coalescences," *Phys. Rev. D*, vol. 107, p. 043035, Feb 2023.

[190] R. Huxford, R. Kashyap, S. Borhanian, A. Dhani, I. Gupta, and B. S. Sathyaprakash, "Accuracy of neutron star radius measurement with the next generation of terrestrial gravitational-wave observatories," *Phys. Rev. D*, vol. 109, no. 10, p. 103035, 2024.

[191] K. Chatziioannou, "Uncertainty limits on neutron star radius measurements with gravitational waves," *Phys. Rev. D*, vol. 105, p. 084021, Apr 2022.

[192] J. S. Read, L. Baiotti, J. D. E. Creighton, J. L. Friedman, B. Giacomazzo, K. Kyutoku, C. Markakis, L. Rezzolla, M. Shibata, and K. Taniguchi, "Matter effects on binary neutron star waveforms," *Phys. Rev. D*, vol. 88, p. 044042, Aug 2013.

[193] E. Belgacem, Y. Dirian, S. Foffa, E. J. Howell, M. Maggiore, and T. Regimbau, "Cosmology and dark energy from joint gravitational wave-grb observations," *Journal of Cosmology and Astroparticle Physics*, vol. 2019, p. 015, aug 2019.

[194] N. Nagarajan and C. Messenger, "Identifying and mitigating machine learning biases for the gravitational-wave detection problem," 2025.

[195] T. D. Gebhard, N. Kilbertus, I. Harry, and B. Schölkopf, "Convolutional neural networks: A magic bullet for gravitational-wave detection?," *Phys. Rev. D*, vol. 100, p. 063015, Sep 2019.

[196] M. B. Schäfer, O. c. v. Zelenka, A. H. Nitz, H. Wang, S. Wu, Z.-K. Guo, Z. Cao, Z. Ren, P. Nousi, N. Stergioulas, P. Iosif, A. E. Koloniari, A. Tefas, N. Passalis, F. Salemi, G. Vedovato, S. Klimenko, T. Mishra, B. Brügmann, E. Cuoco, E. A. Huerta, C. Messenger, and F. Ohme, "First machine learning gravitational-wave search mock data challenge," *Phys. Rev. D*, vol. 107, p. 023021, Jan 2023.

[197] B. Zackay, L. Dai, and T. Venumadhav, "Relative binning and fast likelihood evaluation for gravitational wave parameter estimation," 2018.

[198] N. Leslie, L. Dai, and G. Pratten, "Mode-by-mode relative binning: Fast likelihood estimation for gravitational waveforms with spin-orbit precession and multiple harmonics," *Phys. Rev. D*, vol. 104, p. 123030, Dec 2021.

[199] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, "Real-time gravitational wave science with neural posterior estimation," *Phys. Rev. Lett.*, vol. 127, p. 241103, Dec 2021.

[200] Q. Hu and J. Veitch, "Costs of bayesian parameter estimation in third-generation gravitational wave detectors: a review of acceleration methods," 2024.

[201] A. Bauswein, S. Goriely, and H.-T. Janka, "Systematics of dynamical mass ejection, nucleosynthesis, and radioactively powered electromagnetic signals from neutron-star mergers," *The Astrophysical Journal*, vol. 773, p. 78, jul 2013.

[202] C. Zhao, Y. Lu, Q. Chu, and W. Zhao, "The luminosity functions of kilonovae from binary neutron star mergers under different equation of states," *Monthly Notices of the Royal Astronomical Society*, vol. 522, pp. 912–936, 04 2023.

[203] W.-Z. Qiumu, M.-H. Chen, Q.-H. Chen, and E.-W. Liang, "Kilonova emission from neutron star mergers with different equations of state," 2025.

[204] L. Suleiman and J. Read, "Quasiuniversal relations in the context of future neutron star detections," *Phys. Rev. D*, vol. 109, p. 103029, May 2024.

[205] L. Suleiman, M. Fortin, J. L. Zdunik, and P. Haensel, "Influence of the crust on the neutron star macrophysical quantities and universal relations," *Phys. Rev. C*, vol. 104, p. 015801, Jul 2021.

[206] *PATTERN RECOGNITION: STATISTICAL, STRUCTURAL AND NEURAL APPROACHES*. Wiley India Pvt. Limited, 2007.

[207] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[208] F. Chollet, *Deep Learning with Python*. USA: Manning Publications Co., 1st ed., 2017.

[209] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics* (G. Gordon, D. Dunson, and M. Dudík, eds.), vol. 15 of *Proceedings of Machine Learning Research*, (Fort Lauderdale, FL, USA), pp. 315–323, PMLR, 11–13 Apr 2011.

[210] V. Nair and G. Hinton, "Rectified linear units improve restricted boltzmann machines vinod nair," vol. 27, pp. 807–814, 06 2010.

[211] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, Mar. 2021.

[212] G. Hinton and S. Roweis, "Stochastic neighbor embedding," in *Proceedings of the 16th International Conference on Neural Information Processing Systems*, (Cambridge, MA, USA), p. 857–864, MIT Press, 2002.

[213] D. Jurafsky and J. H. Martin, "Logistic regression," in *Speech and Language Processing (3rd ed. draft)*, ch. 5, Prentice Hall, 2023.

[214] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, 1986.

[215] P. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, pp. 1550 – 1560, 11 1990.

[216] S. Khirirat, H. R. Feyzmahdavian, and M. Johansson, "Mini-batch gradient descent: Faster convergence under data sparsity," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 2880–2887, 2017.

[217] Y.-q. Zhao, Z.-Y. Qiu, L. Wang, C. Wang, and Z.-H. Guo, "Nesterov momentum based optimization algorithm for deep learning," in *2024 10th International Conference on Computer and Communications (ICCC)*, pp. 127–131, 2024.

[218] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[219] F. J. Huang and Y. LeCun, "Large-scale learning with svm and convolutional for generic object categorization," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 284–291, 2006.

[220] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in Neural Information Processing Systems* (D. Touretzky, ed.), vol. 2, Morgan-Kaufmann, 1989.

[221] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.

[222] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," vol. 1, p. 270–280, June 1989.

[223] R. Williams and J. Peng, "An efficient gradient-based algorithm for on-line training of recurrent network trajectories," *Neural Computation*, vol. 2, pp. 490–501, 12 1990.

[224] R. J. Williams and D. Zipser, "Gradient-based learning algorithms for recurrent networks and their computational complexity," in *Back-propagation: Theory, Architectures and Applications*, 1995.

[225] T. Du, S. Ji, L. Shen, Y. Zhang, J. Li, J. Shi, C. Fang, J. Yin, R. Beyah, and T. Wang, "Cert-rnn: Towards certifying the robustness of recurrent neural networks," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, CCS '21, (New York, NY, USA), p. 516–534, Association for Computing Machinery, 2021.

[226] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, p. 1735–1780, Nov. 1997.

[227] R. C. Staudemeyer and E. R. Morris, "Understanding lstm – a tutorial into long short-term memory recurrent neural networks," 2019.

[228] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[229] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," 2020.

[230] D. J. Rezende and S. Mohamed, "Variational inference with normalizing flows," 2016.

[231] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing flows for probabilistic modeling and inference," *Journal of Machine Learning Research*, vol. 22, no. 57, 2021.

[232] E. Tabak and E. Vanden-Eijnden, "Density estimation by dual ascent of the log-likelihood," *Communications in Mathematical Sciences - COMMUN MATH SCI*, vol. 8, 03 2010.

[233] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real nvp," 2017.

[234] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, "Neural spline flows," 2019.

[235] Thomas Müller and Brian McWilliams and Fabrice Rousselle and Markus Gross and Jan Novák, "Neural importance sampling," 2019.

[236] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, "Cubic-spline flows," 2019.

[237] A. Jadon, A. Patil, and S. Jadon, "A comprehensive survey of regression based loss functions for time series forecasting," 2022.

[238] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Foundations and Trends in Machine Learning*, vol. 12, no. 4, p. 307–392, 2019.

[239] A. A. Pol, V. Berger, G. Cerminara, C. Germain, and M. Pierini, "Anomaly detection with conditional variational autoencoders," 2020.

[240] C. Zhang, R. Barbano, and B. Jin, "Conditional variational autoencoder for learned image reconstruction," 2021.

[241] F. Janjos, M. Hallgarten, A. Knittel, M. Dolgov, A. Zell, and J. M. Zollner, "Conditional unscented autoencoders for trajectory prediction," 2024.

[242] D. Davis et al., "Ligo detector characterization in the second and third observing runs," *Classical and Quantum Gravity*, vol. 38, p. 135014, jun 2021.

[243] B. P. Abbott et al., "Characterization of transient noise in advanced ligo relevant to gravitational wave signal gw150914," *Classical and Quantum Gravity*, vol. 33, p. 134001, jun 2016.

[244] M. Zevin, S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, K. Crowston, A. K. Katsaggelos, S. L. Larson, T. K. Lee, C. Lintott, T. B. Littenberg, A. Lundgren, C. Østerlund, J. R. Smith, L. Trouille, and V. Kalogera, "Gravity spy: integrating advanced ligo detector characterization, machine

learning, and citizen science," *Classical and Quantum Gravity*, vol. 34, p. 064003, feb 2017.

[245] F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, "Omicron: A tool to characterize transient noise in gravitational-wave detectors," *SoftwareX*, vol. 12, p. 100620, 2020.

[246] P. Laguarta, R. van der Laag, M. Lopez, T. Dooney, A. L. Miller, S. Schmidt, M. Cavaglia, S. Caudill, K. Driessens, J. Karel, R. Lenders, and C. V. D. Broeck, "Detection of anomalies amongst ligo's glitch populations with autoencoders," 2023.

[247] V. Skliris, M. R. K. Norman, and P. J. Sutton, "Real-time detection of unmodelled gravitational-wave transients using convolutional neural networks," 2024.

[248] B. P. Abbott et al., "All-sky search for short gravitational-wave bursts in the second advanced ligo and advanced virgo run," *Phys. Rev. D*, vol. 100, p. 024017, Jul 2019.

[249] S. A. Usman, A. H. Nitz, I. W. Harry, C. M. Biwer, D. A. Brown, M. Cabero, C. D. Capano, T. D. Canton, T. Dent, S. Fairhurst, M. S. Kehl, D. Keppel, B. Krishnan, A. Lenon, A. Lundgren, A. B. Nielsen, L. P. Pekowsky, H. P. Pfeiffer, P. R. Saulson, M. West, and J. L. Willis, "The pycbc search for gravitational waves from compact binary coalescence," *Classical and Quantum Gravity*, vol. 33, p. 215004, oct 2016.

[250] S. Schmidt, B. Gadre, and S. Caudill, "Gravitational-wave template banks for novel compact binaries," *Phys. Rev. D*, vol. 109, p. 042005, Feb 2024.

[251] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, "Matching matched filtering with deep networks for gravitational-wave astronomy," *Phys. Rev. Lett.*, vol. 120, p. 141103, Apr 2018.

[252] D. George and E. Huerta, "Deep learning for real-time gravitational wave detection and parameter estimation: Results with advanced ligo data," *Physics Letters B*, vol. 778, pp. 64–70, 2018.

[253] P. Nousi, A. E. Koloniari, N. Passalis, P. Iosif, N. Stergioulas, and A. Tefas, "Deep residual networks for gravitational wave detection," *Phys. Rev. D*, vol. 108, p. 024022, Jul 2023.

[254] M. B. Schäfer and A. H. Nitz, "From one to many: A deep learning coincident gravitational-wave search," *Phys. Rev. D*, vol. 105, p. 043003, Feb 2022.

[255] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov, F. H. Vivanco, J. Harms, M. E. Lower, G. D. Meadors, D. Melchor, E. Payne, M. D. Pitkin, J. Powell, N. Sarin, R. J. E. Smith, and E. Thrane, "Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy," *The Astrophysical Journal Supplement Series*, vol. 241, p. 27, apr 2019.

[256] M. J. Williams, J. Veitch, and C. Messenger, "Nested sampling with normalizing flows for gravitational-wave inference," *Phys. Rev. D*, vol. 103, no. 10, p. 103006, 2021.

[257] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, "Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy," *Nature Physics*, vol. 18, pp. 112–117, Jan 2022.

[258] M. Dax et al., "Real-time inference for binary neutron star mergers using machine learning," *Nature*, vol. 639, pp. 49–53, Mar 2025.

[259] A. J. K. Chua and M. Vallisneri, "Learning bayesian posteriors with neural networks for gravitational-wave inference," *Phys. Rev. Lett.*, vol. 124, p. 041102, Jan 2020.

[260] F. Stachurski, C. Messenger, and M. Hendry, "Cosmological inference using gravitational waves and normalizing flows," *Phys. Rev. D*, vol. 109, p. 123547, Jun 2024.

[261] K. Leyde, S. R. Green, A. Toubiana, and J. Gair, "Gravitational wave populations and cosmology with neural posterior estimation," *Phys. Rev. D*, vol. 109, p. 064056, Mar 2024.

[262] K. W. K. Wong, G. Contardo, and S. Ho, "Gravitational-wave population inference with deep flow-based generative network," *Phys. Rev. D*, vol. 101, p. 123005, Jun 2020.

[263] A. Ray, I. M. Hernandez, S. Mohite, J. Creighton, and S. Kapadia, "Nonparametric inference of the population of compact binaries from gravitational-wave observations using binned gaussian processes," *The Astrophysical Journal*, vol. 957, p. 37, oct 2023.

[264] S. Colloms, C. P. L. Berry, J. Veitch, and M. Zevin, "Exploring the evolution of gravitational-wave emitters with efficient emulation: Constraining the origins of binary black holes using normalising flows," 2025.

[265] D. Ruhe, K. Wong, M. Cranmer, and P. Forré, "Normalizing flows for hierarchical bayesian analysis: A gravitational wave population study," 2022.

[266] M. Punturo et al., "The einstein telescope: a third-generation gravitational wave observatory," *Classical and Quantum Gravity*, vol. 27, p. 194002, sep 2010.

[267] W. Alhassan, T. Bulik, and M. Suchenek, "Detection of einstein telescope gravitational wave signals from binary black holes using deep learning," *Monthly Notices of the Royal Astronomical Society*, vol. 519, pp. 3843–3850, 12 2022.

[268] F. Santoliquido, U. Dupletsa, J. Tissino, M. Branchesi, F. Iacovelli, G. Iorio, M. Mapelli, D. Gerosa, J. Harms, and M. Pasquato, "Classifying binary black holes from population iii stars with the einstein telescope: A machine-learning approach," *Astronomy amp; Astrophysics*, vol. 690, p. A362, Oct. 2024.

[269] Q. Meijer, M. Lopez, D. Tsuna, and S. Caudill, "Gravitational-wave searches for cosmic string cusps in einstein telescope data using deep learning," *Phys. Rev. D*, vol. 109, p. 022006, Jan 2024.

[270] P. Mukherjee, R. Shah, A. Bhaumik, and S. Pal, "Reconstructing the hubble parameter with future gravitational-wave missions using machine learning," *The Astrophysical Journal*, vol. 960, p. 61, dec 2023.

[271] "Upper limits on the rates of binary neutron star and neutron star–black hole mergers from advanced ligo's first observing run," *The Astrophysical Journal Letters*, vol. 832, p. L21, nov 2016.

[272] T. L. S. Collaboration, the Virgo Collaboration, and the KAGRA Collaboration, "Prospects for observing and localizing gravitational-wave transients with advanced ligo, advanced virgo and kagra," 2020.

[273] T. L. S. Collaboration, the Virgo Collaboration, and the KAGRA Collaboration, "Prospects for observing and localizing gravitational-wave transients with advanced ligo and advanced virgo," *Living Reviews in Relativity*, vol. 19, Feb. 2016.

[274] J. Hjorth, A. J. Levan, N. R. Tanvir, J. D. Lyman, R. Wojtak, S. L. Schrøder, I. Mandel, C. Gall, and S. H. Bruun, "The Distance to NGC 4993: The Host Galaxy of the Gravitational-wave Event GW170817," , vol. 848, p. L31, Oct. 2017.

[275] V. Cardoso, S. Hopper, C. F. B. Macedo, C. Palenzuela, and P. Pani, "Gravitational-wave signatures of exotic compact objects and of quantum corrections at the horizon scale," *Phys. Rev. D*, vol. 94, p. 084031, Oct 2016.

[276] N. Itoh, "Hydrostatic equilibrium of hypothetical quark stars," *Progress of Theoretical Physics*, vol. 44, pp. 291–292, 07 1970.

[277] D. J. Kaup, "Klein-gordon geon," *Phys. Rev.*, vol. 172, pp. 1331–1342, Aug 1968.

[278] S. L. Liebling and C. Palenzuela, "Dynamical boson stars," *Living Reviews in Relativity*, vol. 15, May 2012.

[279] E. R. Most, L. R. Weih, L. Rezzolla, and J. Schaffner-Bielich, "New constraints on radii and tidal deformabilities of neutron stars from gw170817," *Phys. Rev. Lett.*, vol. 120, p. 261103, Jun 2018.

[280] S. De, D. Finstad, J. M. Lattimer, D. A. Brown, E. Berger, and C. M. Biwer, "Tidal Deformabilities and Radii of Neutron Stars from the Observation of GW170817," *Phys. Rev. Lett.*, vol. 121, p. 091102, Aug 2018.

[281] B. T. Reed, R. Somasundaram, S. De, C. L. Armstrong, P. Giuliani, C. Capano, D. A. Brown, and I. Tews, "Towards accelerated nuclear-physics parameter estimation from binary neutron star mergers: Emulators for the tolman-oppenheimer-volkoff equations," 2024.

[282] I. Jolliffe, *Principal Component Analysis*. New York: Springer, 2nd ed., 2002.

[283] J. E. Jackson, *A User's Guide to Principal Components*. New York: John Wiley & Sons, 1991.

[284] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Duchesnay, "scikit-learn: Machine learning in python," 2011.

[285] L. Biewald, "Experiment tracking with weights and biases," 2020. Software available from wandb.com.

[286] M. J. Williams, *Accelerating Gravitational-Wave Inference with Machine Learning*. PhD thesis, University of Glasgow, 2023. Accessed: 2025-02-08.

[287] B. Gärtner and M. Hoffmann, "Computational geometry lecture notes hs 2013." https://ti.inf.ethz.ch/ew/Lehre/CG13/lecture/cg-2013.pdf, 2013. Accessed: 2025-02-08.

[288] L. S. Collaboration and V. Collaboration, "Parameter estimation sample release for gwtc-1." https://dcc.ligo.org/LIGO-P1800370/public, 2019. LIGO-P1800370-v5.

[289] T. Dietrich, S. Bernuzzi, and W. Tichy, "Closed-form tidal approximants for binary neutron star gravitational waveforms constructed from high-resolution numerical relativity simulations," *Phys. Rev. D*, vol. 96, p. 121501, Dec 2017.

[290] A. J. Levan, J. D. Lyman, N. R. Tanvir, J. Hjorth, I. Mandel, E. R. Stanway, D. Steeghs, A. S. Fruchter, E. Troja, S. L. Schrøder, K. Wiersema, S. H. Bruun, Z. Cano, S. B. Cenko, A. d. U. Postigo, P. Evans, S. Fairhurst, O. D. Fox, J. P. U.

Fynbo, B. Gompertz, J. Greiner, M. Im, L. Izzo, P. Jakobsson, T. Kangas, H. G. Khandrika, A. Y. Lien, D. Malesani, P. O'Brien, J. P. Osborne, E. Palazzi, E. Pian, D. A. Perley, S. Rosswog, R. E. Ryan, S. Schulze, P. Sutton, C. C. Thöne, D. J. Watson, and R. A. M. J. Wijers, "The Environment of the Binary Neutron Star Merger GW170817," *The Astrophysical Journal Letters*, vol. 848, p. L28, oct 2017.

[291] L. S. Collaboration and V. Collaboration, "GW170817: Measurements of neutron star radii and equation of state," *Physical Review Letters*, vol. 121, no. 16, p. 161101, 2018.

[292] L. Smith, S. Ghosh, J. Sun, V. Gayathri, I. S. Heng, and A. Pai, "Enhancing search pipelines for short gravitational-wave transients with gaussian mixture modeling," *Phys. Rev. D*, vol. 110, p. 083032, Oct 2024.

[293] V. Gayathri, D. Lopez, P. R. S., I. S. Heng, A. Pai, and C. Messenger, "Enhancing the sensitivity of transient gravitational wave searches with gaussian mixture models," *Phys. Rev. D*, vol. 102, p. 104023, Nov 2020.

[294] D. Reynolds, *Gaussian Mixture Models*, pp. 659–663. Boston, MA: Springer US, 2009.

[295] G. Ashton *et al.*, "BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy," *Astrophys. J. Suppl.*, vol. 241, no. 2, p. 27, 2019.

[296] LIGO Scientific Collaboration, "Bns eos example — bilby gw injection example." https://git.ligo.org/lscsoft/bilby/-/blob/master/examples/gw_examples/injection_examples/bns_eos_example.py, 2025. Accessed: 2025-07-25.

[297] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, "Frequency-domain gravitational waves from nonprecessing black-hole binaries. ii. a phenomenological model for the advanced detector era," *Phys. Rev. D*, vol. 93, p. 044007, Feb 2016.

[298] M. J. Williams, "nessai: Nested sampling with artificial intelligence," Feb. 2021.

[299] O. S. Salafia, "Detection probability of light compact binary mergers in future observing runs of the current ground-based gravitational wave detector network," 7 2025.

[300] A. Abac et al., "The science of the einstein telescope," 2025.

[301] R. Essick, P. Landry, A. Schwenk, and I. Tews, "Detailed examination of astrophysical constraints on the symmetry energy and the neutron skin of ^{208}Pb with minimal modeling assumptions," *Phys. Rev. C*, vol. 104, p. 065804, Dec 2021.

[302] C. Ma, X. Yu, Z. Cao, and M. Jia, "Extraction of binary neutron star gravitational wave waveforms from einstein telescope using deep learning," *Phys. Rev. D*, vol. 111, p. 123053, Jun 2025.

[303] S. Vinciguerra, J. Veitch, and I. Mandel, "Accelerating gravitational wave parameter estimation with multi-band template interpolation," *Classical and Quantum Gravity*, vol. 34, p. 115006, may 2017.

[304] S. Morisaki, "Accelerating parameter estimation of gravitational waves from compact binary coalescence using adaptive frequency resolutions," *Phys. Rev. D*, vol. 104, p. 044062, Aug 2021.

[305] S. T. Tokdar and R. E. Kass, "Importance sampling: a review," *WIREs Computational Statistics*, vol. 2, no. 1, pp. 54–60, 2010.

[306] M. Dax, S. R. Green, J. Gair, M. Pürrer, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, "Neural importance sampling for rapid and reliable gravitational-wave inference," *Phys. Rev. Lett.*, vol. 130, p. 171403, Apr 2023.