arXiv:2209.02817v2 [astro-ph.HE] 8 Sep 2022

Deducing Neutron Star Equation of State Parameters Directly From Telescope Spectra with Uncertainty-Aware Machine Learning

Delaney Farrell,¹ Pierre Baldi,² Jordan Ott,² Aishik Ghosh,^{3,4} Andrew W. Steiner,^{5,6}

Atharva Kavitkar,⁷ Lee Lindblom,⁸ Daniel Whiteson,³ and Fridolin Weber^{1,8}

¹Department of Physics, San Diego State University, San Diego, CA 92115, United States

²Department of Computer Science, University of California Irvine, Irvine, California 92697, USA

³Department of Physics and Astronomy, University of California Irvine, Irvine, California 92697, USA

⁴Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

⁵Department of Physics and Astronomy, University of Tennessee, Knoxville, Tennessee 37996, USA

⁶Physics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

⁷Department of Computer Science, TU Kaiserslautern, Germany

⁸Center for Astrophysics and Space Sciences, University of California at San Diego, San Diego, CA 92093, United States

Neutron stars provide a unique laboratory for studying matter at extreme pressures and densities. While there is no direct way to explore their interior structure, X-rays emitted from these stars can indirectly provide clues to the equation of state (EOS) of the superdense nuclear matter through the inference of the star's mass and radius. However, inference of EOS directly from a star's X-ray spectra is extremely challenging and is complicated by systematic uncertainties. The current state of the art is to use simulation-based likelihoods in a piece-wise method, which first infer the star's mass and radius to reduce the dimensionality of the problem, and from those quantities infer the EOS. We demonstrate a series of enhancements to the state of the art, in terms of realistic uncertainty quantification and improved regression of physical properties with machine learning. We also demonstrate novel inference of the EOS directly from the high-dimensional spectra of observed stars, avoiding the intermediate mass-radius step. Our network is conditioned on the sources of uncertainty of each star, allowing for natural and complete propagation of uncertainties to the EOS.

I. INTRODUCTION

Neutron stars are the densest stellar objects, providing a unique laboratory for studying matter in physical conditions that cannot be replicated on Earth and are only found in these neutron-packed remnants of massive stars ($\gtrsim 8M_{\odot}$). Insights about the forms of matter which emerge under these extreme conditions can improve our understanding of two of the least well-understood fundamental forces, quantum chromodynamics, and gravity.

The neutron-rich matter within the inner regions of a neutron star can reach supranuclear densities of 10^{15} g/cm³, potentially leading to transitions to stable non-nucleonic states of strange matter in the form of hyperons [1–3], deconfined quark matter made of up, down, and strange quarks [4, 5], color superconducting phases [6, 7], or Bose-Einstein condensates made of negatively charged pions or K^- mesons [8–10]. The structure and behavior of matter at such extreme densities are one of the great mysteries in modern science, prompting decades of theoretical and experimental research into the interior composition of a neutron star.

The nature of matter within a neutron star is compactly summarized by its equation of state (EOS), the relationship between pressure P and energy density ϵ , which show starkly different behaviors under the various states of strange matter hypotheses described above. This relationship is determined by the microphysical interactions between various particles within the star, relative abundances of different particle species, as well as the star's temperature. Understanding the EOS of supranuclear density matter has been of interest to the nuclear and astrophysics communities for decades, resulting in many proposed phenomenological models for the equation of state of neutron star matter. Variations in models arise from a lack of precise knowledge of nuclear interactions between particles at such extreme conditions, as well as the wide range of densities and isospin asymmetries that are thought to exist within neutron stars [11].

While the internal pressure and density cannot be directly observed, the EOS of a star in static gravitational equilibrium determines stellar properties such as its mass and radius, which in turn determine observables such as the stellar X-ray spectrum. Conversely, the stellar spectra can be used to infer the masses and radii, which in principle allow inference of the EOS [12–17], though inversion of this second step is numerically very difficult. Additional challenges are due to the small number of neutron star observations, O(10), and the significant uncertainty of individual measurements. It is therefore vital that as much information as possible is extracted from each star, and that the uncertainties be propagated accurately, to provide the most complete information possible about the EOS.

At the same time, there has been a dramatic burst of progress in artificial intelligence, specifically deep learning [18], a modern re-branding of neural networks. This progress has led to breakthroughs not only in traditional areas such as natural language processing and computer vision but also in the natural sciences, including particle physics, often increasing the statistical power of difficultto-collect data [19] while allowing robust handling of uncertainties [20]. Where earlier neural networks were limited in size, computing progress especially in the form

 $\mathbf{2}$

of Graphical Processing Units (GPUs), has enabled the deployment of larger and deeper networks that can handle more complex and higher-dimensional data [21, 22], allowing direct analysis of data without requiring dimensional reduction, or other preprocessing steps, that can often sacrifice useful information. The full power of these techniques has not yet been brought to bear on many astrophysical tasks.

In the context of the inference of neutron star EOS, recent work by Fujimoto et al. [23, 24] demonstrated the ability of deep networks to regress the EOS directly from a set of stellar mass-radius pairs, without the need to extract the functional relationship between mass and radius. Their analysis used a toy model to describe the uncertainties in mass and radius, assuming uncorrelated Gaussian errors randomly drawn from ad-hoc priors. Related work [25] has demonstrated similar regression, again assuming Gaussian uncertainty on mass and radius values, but with clever efforts to reduce dependence on EOS parameterization. An alternative approach [26] uses both neural networks and support vector machines to regress the EOS from stellar radii and tidal deformations. What has received less attention are methods that directly access the observed telescope spectra, which underly the inferred stellar summary quantities such as mass and radius, and methods that allow for the full propagation of realistic uncertainties.

In this paper, we use neural network regression to infer the neutron star EOS from stellar mass and radius information derived from a realistic model of the underlying spectra and accounting for the associated underlying uncertainties. We subsequently introduce a network capable of directly analyzing high-dimensional neutron star spectra, performing inference of stellar mass and radii from telescope spectra as a demonstration of the capacity of deep networks. Finally, we perform a first-ofits-kind inference of the EOS parameters directly from a multi-star set of stellar spectra, without requiring the intermediate step of collapsing the information into mass and radius; see Figure 1. In each case, we allow for full propagation of uncertainties by conditioning the networks on the stellar nuisance parameters. Comparisons are made to benchmark methods, where possible.

The paper is organized as follows. In Section II, we provide background on the physics of the connection between the nuclear equation of state and the stellar observations. Section III describes the fundamentals of the machine learning concepts on which our studies rely. Details of the samples of simulated data are given in Section IV. Section V demonstrates inference of the EOS parameters from mass and radius, while Section VI describes how mass and radius parameters can be inferred directly from stellar spectra, and Section VII shows endto-end inference of EOS parameters from a set of neutron star spectra.

II. BACKGROUND

A. Equation of State for Dense Matter

Neutron star interiors present a unique opportunity to study matter under conditions beyond the reach of terrestrial laboratories: matter that is extremely high in density, relatively cold in temperature, and isospinasymmetric [27]. Perhaps the closest experimental constraints have come from ultrarelativistic heavy-ion collisions (as conducted at the RHIC [28], the LHC [29], and FAIR [30]), which probe the nature of hot, symmetric nuclear matter Extremely neutron-rich matter is more recently probed in studies of neutron skin or giant monopole and dipole resonances, but these studies are limited in nature [31]. Unfortunately, these experiments currently lack the temperature and density constraints present within the core of a neutron star - meaning the dense matter encountered within a neutron star cannot yet be replicated by experiment.

While properties of superdense matter cannot be derived directly from experiments, certain general principles from general relativity and quantum chromodynamics (QCD) guide the theoretical investigation into neutron star matter. [AWS: I reworded these next few sentences. Neutron star structure is controlled by the longrange gravitational force [32, 33] which holds the star together and short-range strong interactions between nucleons and nuclei which provide the pressure that prevents the star from collapsing. At low nuclear densities (below nuclear saturation), effective field theories based on QCD provide a systematic basis for nuclear forces, which offers good constraints on two-nucleon interactions [33–35]. At higher densities, the QCD framework predicts that baryonic matter (where quarks are confined within hadrons) will experience a phase transition to quark-gluon plasma (QGP), where quarks and gluons are freed from hadronic boundaries [35]. Other stable states of non-nucleonic matter may also occur, like the formation of hyperons, color superconducting phases of quark matter, or Bose-Einstein condensates of different mesons [8, 10]. Theoretical uncertainties have resulted in a wide range of proposed phenomenological models for the EOS of neutron star matter, which can then be tested by experiment or observation.

The EOS of neutron star matter is intrinsically linked to macroscopic characteristics like gravitational mass M and radius R through the general relativistic stellar structure equation known as the Tolman-Oppenheimer-Volkoff (TOV) equation [36, 37]. This equation assumes the object is spherically symmetric, non-rotating, and non-magnetic. The TOV equation is given by (assuming geometrized units where G = c = 1):

$$\frac{dP}{dr} = -\frac{(\epsilon + P)(m + 4\pi r^3 P)}{r^2 \left(1 - \frac{2m}{r}\right)} \tag{1}$$

where m is the gravitational mass enclosed within a sphere of radius r. The mass of the star can be solved



FIG. 1: Overview of the regression task, which involves either inferring stellar summary quantities such as mass and radius, which can then be used to deduce the equation of state as in earlier work [23–26] or inference of EOS directly from stellar spectra, as is demonstrated in this study.

for as:

$$\frac{dm}{dr} = 4\pi r^2 \epsilon \tag{2}$$

where the total gravitational mass M of a star with radius R is given by $M \equiv m(R) = 4\pi \int_0^R dr \ r^2 \epsilon$. Given an EOS, numerically solving the TOV equation for M and Ris straightforward. These equations create a one-to-one map from the EOS to the M - R relation [17]; the inverse form of this map can therefore provide constraints on the EOS from observable properties. To mathematically invert the TOV equation, at least two stars' mass and radius must be known exactly, a feat is not possible with current observational technology. Solving the inverse problem is therefore much more complicated, potentially even intractable without making significant numerical assumptions.

B. X-Ray Spectroscopy for Neutron Stars

Many reliable observations of neutron stars come from X-ray emission, either from electromagnetic radiation from pulsars or thermal emission in quiescent low-mass X-ray binaries (qLMXBs). qLMXBs are particularly desirable to place strong constraints on neutron star structure as they are likely to have low magnetic fields (10^{8-9} G) , resulting in minimal effects on the radiation transport or temperature distribution on the star's surface [38–40]. Additionally, these binaries are identified in globular clusters where distances, ages, and reddening are well-constrained [41]. The distinctive soft thermal spectra from these sources come from a long-lived thermal glow resulting from heat stored in the deep crust of the

neutron stars within the binary system during accretion, which is then re-radiated from the whole surface when accretion stops [42]. For the context of this work, the inference of EOS will come from simulated thermal spectra from qLMXBs.

Observation of neutron star emission, whether X-ray or gravitational wave, has long served as a way to constrain mass and radius for neutron stars (eg. [43, 44]), but uncertainties arise in the inference of these properties for a variety of reasons. In the case of X-ray radiation from qLMXBs, constraints on mass and radius are determined by fitting the emitted spectrum with an appropriate atmosphere model (where the surface composition is known or can be determined by the X-ray spectrum) and combining the spectroscopic measurements with the distance of the source. Models for thermal X-ray radiation are based on a light-element atmosphere, as the lightest element that is present in the atmosphere floats to the top due to rapid gravitational settling on neutron star surfaces [38]. Atmospheric models used on X-ray spectra from qLMXBs gave the first broad constraints on neutron star radius, and more modern analyses of X-ray spectra have provided tighter constraints on both radius and EOS.

he high-resolution imaging and spectroscopy of NASA's *Chandra X-ray Observatory* have provided powerful insight into neutron star properties like cooling [45, 46], mass and radius [38], and binary mergers of exotic stars [47]. *Chandra*'s telescope contains a system of four pairs of mirrors that focus incoming X-ray photons to the Advanced CCD Imaging Spectrometer (ACIS), which measures the energy of each incoming Xray. The observed spectrum, along with a corresponding instrument response, is then fit to a well-motivated parameterized model. Many such models for spectral fitting exist in XSPEC [48], an X-ray spectral fitting package distributed and maintained by the aegis of the GSFC High Energy Astrophysics Science Archival Research Center (HEASARC). XSPEC has been used numerous times in the past to analyze data from *Chandra* as well as other spectrometers like *NICER*, *Nustar*, and *XMM-Newton*, making it a valuable resource for inference of neutron star properties.

III. MACHINE LEARNING

Machine learning methods, in particular deep learning, aim to extract useful knowledge from data automatically and are rapidly being applied across many data-rich fields of science [18]. In regression tasks such as EOS inference, one is interested in constructing a function f whose inputs are the observed data and whose outputs are an estimate of some parameter of interest. The optimal function f is not known initially, but an approximation can be learned from a set of input-output example pairs.

In order to approximate f, machine learning methods first begin with a suitable class C of parameterized functions (e.g. polynomials of a certain degree, neural networks of a certain architecture) and then seek to find the best approximation to f within the class C. This is typically done through a stochastic gradient descent procedure that seeks to iteratively minimize the approximation error on the training set.

The well-known technique of linear regression is the most elementary form of regression, and can be viewed as a form of shallow learning (no hidden layers). Deep learning generalizes linear regression by using multi-layer neural networks as the class C and thus enabling the construction of sophisticated and flexible non-linear approximations. With sufficient training data and computing power, deep learning methods can handle large-scale problems with high dimensional data and avoid heuristic simplifications that lose information. It is not uncommon to deal with problems with input sizes in the range of up to 10^9 examples, each with dimensions of 10^{3-4} , with neural networks that can have up to 10^{11} free parameters. Training sets can range in size from 10^1 to 10^{10} or more. Unlike shallow learning and linear regression, deep learning does not require that the number of parameters be equal to the number of training examples [49]. More recent, attention-based architectures, such as transformers [50, 51], allow networks to take advantage of structures and symmetries in the data, and are applied in sections below.

When the interpretation of data depends on external unmeasured or poorly-known parameters, such as neutron star temperature or distance, it can be useful to apply *parameterized networks* [21]. Such networks learn a task as a function of the external parameter, allowing for evaluation of a fixed dataset under varying assumptions about the parameter [20].

IV. TRAINING SAMPLES

Samples of simulated neutron stars, used to train networks and evaluate their performance, are described below.

Each simulated star is described by two high-level summary quantities, mass and radius, which are drawn from the mass-radius relation determined by the EOS, as well as three nuisance parameters that are independent of the EOS and can vary from star to star. These five parameters are sufficient to determine the expected simulated Chandra telescope spectrum in the chosen NS theoretical model. In the case of EOS inference, sets of stars with consistent EOS are grouped to form training and testing sets. Details of each step of the generation are provided below.

A. Equation of State

The equation of state of the hadronic matter within the core is modeled with the relativistic non-linear mean field model GM1L [52]. The version used here only accounts for protons and neutrons but can be extended to include hyperons and Δ baryons [53]. The corresponding saturation properties of symmetric nuclear matter for the GM1L parametrization are shown in Table I [53, 54]. These properties include the nuclear saturation density n_0 , energy per nucleon E_0 , nuclear compressibility K_0 , effective nucleon mass m_N^*/m_N , asymmetry energy J, asymmetry energy slope L_0 , and the value of the nucleon potential U_N . The values of L_0 listed in Table I are in agreement with the value of the slope of the symmetry energy deduced from nuclear experiments and astrophysical observations [55].

The most commonly used constraints on K_0 come from experimental values of the giant monopole resonance, which lie in the range of 220 to 260 MeV [56, 57]. The analysis of [58], however, suggests a higher range of 250 to 315 MeV. The value of $K_0 = 300$ MeV considered in our paper falls into the latter category, but this will not dramatically impact the neutron-rich equation of state appropriate for neutron star interiors. The GM1L equation of state for the core is paired with two models for the crust. For the outer crust, which falls in the density range $10^4 - 10^{11}$ g/cm³, we use the Baym-Pethick-Sutherland (BPS) model [59]. For the inner crust, with densities in the range $10^{11} - 10^{14}$ g/cm³, we use the Baym-Bethe-Pethick (BBP) model [60].

To limit the number of parameters the networks must learn, the essential features of the high-density portion of the EOS needed to be represented efficiently by just a few values. This can be done accurately by constructing parametric representations based on spectral fits, formed as generalizations of the Fourier series used to represent periodic functions [15]. An EOS, defined as $P = P(\epsilon)$ or $\epsilon(P) = \epsilon$, can be represented as a linear combination of

TABLE I: Parameters of the model used to select example equations of state for the generation of simulated data samples. Shown are properties of the symmetric nuclear matter at saturation density for the GM1L parametrization of neutron star interiors[52]; see text for details.

Saturation Property	Value	Units
n_0	0.153	fm^{-3}
E_0	-16.3	MeV
K_0	300.0	MeV
m_N^*/m_N	0.70	-
J	32.5	MeV
L_0	55.0	MeV
U_N	-65.5	MeV

basis functions $\epsilon_k(\phi)$:

$$\epsilon(p) = \sum_{k} \epsilon_k \Phi_k(p) \tag{3}$$

where $\phi_k(p)$ can be any complete set of functions. The EOS is therefore determined by the spectral coefficients ϵ_k , making $\epsilon_k = \epsilon(p, \epsilon_k)$. There are two important conditions that a physical EOS must satisfy to ensure microscopic stability. The first is that the EOS must be non-negative, or $p(\epsilon) \ge 0$, and the second is that pressure must be monotonically increasing with density [15]. Because these conditions are not naturally respected by arbitrary basis functions in a spectral representation, representing an EOS with a straightforward spectral expansion will likely produce data that violates microscopic stability and is therefore erroneous. To ensure these two conditions are met, we instead turn to a faithful construction of spectral representations of the EOS; the process for constructing these is outlined in detail in [61] and [15].

The spectral representation of GM1L is formed from representing the EOS in terms of the relativistic enthalpy, h, where the EOS can be rewritten as a pair of equations P = P(h) and $\epsilon = \epsilon(h)$. The enthalpy can be defined as

$$h(P) = \int_0^P \frac{dP'}{\epsilon(P')c^2 + P'} \tag{4}$$

where c is the speed of light [61]. Inverting Eq. 4 obtains the equation P = P(h), which can recover the EOS $\epsilon(P)$ as $\epsilon(h) = \epsilon[P(h)]$. The pair of equations P = P(h) and $\epsilon = \epsilon(h)$ can be expressed from a reduction to quadrature:

$$P(h) = P_0 + (\epsilon_0 c^2 + P_0) \int_{h_0}^h \mu(h') dh', \qquad (5)$$

$$\epsilon(h) = -P(h)c^{-2} + (\epsilon_0 + p_0 c^{-2})\mu(h).$$
(6)

The function $\mu(h)$ is defined as

$$\mu(h) = \exp\left\{\int_{h_0}^h [2 + \Gamma(h')]dh'\right\}$$
(7)

which is dependent on the sound speed or velocity function $\Gamma(h)$ and constants $P_0 = P(h_0)$ and $\epsilon_0 = \epsilon(h_0)$ [61]. Similar to Eq. 3, the velocity function can be represented as a spectral expansion:

$$\Gamma(h,\nu_k) = \exp\left[\sum_k \nu_k \Phi_k(h)\right] \tag{8}$$

where $\Phi_k(h)$ is any complete set of basis functions on the domain $[h_0, h_{max}]$.

We constructed enthalpy-based causal spectral fits for the GM1L EOS with up to 10 parameters. Figure 2 shows that just two spectral parameters produce parameterizations with a mean relative error of only 10%, and additional parameters can reduce the error to 5% or lower. A small number of parameters is preferred due to the increased complexity of learning multiple parameters, and the danger of Runge's phenomenon when applying our networks to current neutron star observations with accurate readings of mass and radius, which are still relatively few. Runge's phenomenon arises when attempting to fit equispaced data points with polynomials of high degree; increasing the order of the polynomial interpolation can result in issues with convergence or divergence rates for certain functions [62]. When applied to neutron star observation, attempting to fit a small number of data points (with varying accuracy in observation method) with a model having many parameters may result in fits that accurately fit the data, but are very poor representations of the actual physics. Based on the reasons listed above, we chose to use two spectral parameters to represent the EOS, hereafter referred to as λ_1 and λ_2 .



FIG. 2: Relative error in spectral parameterization of the equation of state, shown as a function of the number of parameters used.

To create many samples needed for training and testing, spectral parameters were then constructed from the expression:

$$\lambda_{\text{generated}} = \lambda_{\text{true}} \cdot (1 + 2 \cdot scale(-0.5 + ran2))$$

where $\lambda_{\text{generated}}$ represents the newly constructed spectral parameter, λ_{true} is the best fit (true) spectral parameter of GM1L, and *scale* is a scaling parameter set

to 0.05. ran2 are uniformly distributed random numbers in the range 0 to 1 generated by the ran2 function given in [63]. This process was repeated to create 10^4 different EOS variations. Each EOS variation was used to generate a coinciding M - R relation using equations 1 2, examples of which are seen in Figure 3, from which 100 (M, R) pairs are selected, each representing stellar parameters consistent with that EOS. Due to the random component of our EOS generation, some models have a mass peak below the current observed mass limit, 2.1 M_{\odot} . All models have a minimum mass of at least 1 M_{\odot} . The physicality of predicted results will be discussed in further detail VIII.



FIG. 3: Examples of training data. Top, 100 samples in EOS spectral parameter space (λ_1, λ_2) randomly selected from the full set of 10,000 EOS spectral pairs. Bottom, neutron star mass-radius curves determined by the selected EOS parameters.

B. Modeling X-ray Spectra

The relation between stellar parameters (M, R) is determined by the EOS, and samples from the allowed curve are used as input to generate simulated X-ray Chandra spectra, such as the Chandra observation of the quiescent low-mass X-ray binary (qLMXB) X7 in the globular cluster 47 Tuc [38].

The XSPEC program [48], which can be used for spectral fitting, is also capable of generation of simulated spectra, via the **fakeit** command when a NS model and telescope response matrix are provided.

The NS theoretical model NSATMOS [64] selected includes a hydrogen atmosphere model with electron conduction and self-irradiation. The Chandra telescope response specified in Ref. [64] was also used to describe the instrument response and telescope effective area.

C. Nuisance Parameters

The NSATMOS model has five parameters to describe each star: gravitational mass M in units of $M \odot$, radius R in units of km, and three additional parameters related to observation. For the context of M - R and subsequent EOS inference, only M and R are parameters of interest, whose values come from those generated by the GM1L EOS and so provide information relevant to the physical question. The remaining three nuisance parameters are the effective temperature of the surface, $T_{\rm eff}$, the distance to the star, d, and the hydrogen column, N_H which parameterizes the reddening of the spectrum by the interstellar medium. These parameters influence the observed spectrum of a given neutron star. Lack of knowledge of these values is a leading source of uncertainty in the inference of mass and radius, and hence EOS.

Using Table 1 in Ref. [44] as a guide, we find that distances typically range between 2 and 10 kpc, and hydrogen columns lie between 0.2 and 5×10^{21} cm⁻². While neutron stars with larger distances and larger hydrogen columns exist, they are sufficiently distant as to be difficult to obtain spectral information. From Table 3 in Ref. [65], effective temperatures at the surface typically lie between 50 and 200 eV, or from 6×10^5 and 2.4×10^6 K. Note that core temperatures are typically a few orders of magnitude larger. Again colder neutron stars most certainly exist but are more difficult to observe.

Examples of generated spectra for varying stellar parameters are shown in Figure 4. The generated spectra are very sensitive to the effective surface temperature, with lesser sensitivity to other parameters.

The networks detailed below provide estimates of either the neutron star mass and radius or the EOS parameters, conditioned on NP values. Uncertainty in regressed parameters of interest due to uncertainties in the NP can then be fully propagated via variation of the NPs used during regression. To demonstrate the impact of NP uncertainties, we define three example scenarios of uncertainties, dubbed "true", "tight", and "loose", which describe the quality of prior information on the NP values for each star.

In the "true" scenario, the NPs are set to the true value used to generate the spectra, such that the NP prior is essentially a delta function. In the "tight" scenario, the uncertainty is described as a narrow Gaussian



FIG. 4: Examples of simulated stellar spectra expected for several values of stellar parameters. Each pane shows the expected rate of photons in Chandra per energy bin, for variations of the parameters of interest (mass M, radius R) as well as for variations of the nuisance parameters (N_H , log($T_{\rm eff}$), distance). The dashed black line has the same parameters in each pane.

TABLE II: Description of "true", "tight", and "loose" nuisance parameter (NP) scenarios. Shown are the width of each Gaussian distribution representing the prior knowledge of each NP. For distance and N_H , width is relative; for log(T_{eff}), it is absolute. See text for details and references.

Nuis. Param.	True	Tight	Loose
Distance	exact	5%	20%
Hydrogen Column ${\cal N}_H$	exact	30%	50%
$\log(T_{\rm eff})$	exact	± 0.1	± 0.2

for each NP, with distance having a width of 5%, hydrogen column having a width of 30%, and $\log(T_{\rm eff})$ having a width of 0.1. In the "loose" scenario, the uncertainties are described by a wider Gaussian, with distance having a width of 20%, hydrogen column having a width of 50%, and $\log(T_{\rm eff})$ having a width of 0.2. These ranges are shown in Table II. The sensitivity to NP values is reflected in the performance of the networks below.

V. INFERENCE OF EOS FROM MASS AND RADIUS

Previous applications of machine learning to the task of inferring the equation of state have begun from the stellar mass and radii [23–25], or equivalent parameters [26], though with simple ad-hoc descriptions of the uncertainty on stellar mass and radius values, often modeled as two-dimensional uncorrelated Gaussian's rather than propagating the underlying uncertainties. In this section, we tackle the same problem, but where the stellar data are more realistic and the underlying uncertainties are fully propagated to the EOS estimation. Specifically, the best estimates of stellar mass and radius are derived using state-of-the-art tools that extract them from realistic stellar spectra, which include the impact of stellar nuisance parameters and limited observation time. In addition, this mass-radius estimation is conditioned on the nuisance parameters, such that variations in those nuisance parameters lead to variations in the mass and radius estimates. This connects directly to the neural network regression of EOS parameters from mass-radius values parameterized in the nuisance parameters, allowing for the direct propagation of the underlying uncertainties to give a measure of the resulting uncertainty on the regressed EOS parameters.

Below, we describe the extraction of realistic mass and radius values with XSPEC and their subsequent use in NN regression of the EOS parameters and the estimation of the uncertainty. For comparison, we also provide a demonstration of the regression of EOS parameters using polynomial regression. In subsequent sections, we consider an alternative extraction of mass and radius using a NN, as well as end-to-end regression of EOS directly from stellar spectra.

A. Mass and Radius inference by XSPEC

Sample stellar spectra are generated as described above, including Poisson noise corresponding to an observation time of 100 ks, and nuisance parameters variations as specified in Table II.

Given a sample observed X-ray spectrum, the XSPEC code scans the mass and radius parameter space, searching for values that best describe it. For each mass-radius pair, the expected spectrum is calculated using the chosen model and telescope response function, identical to those used to generate the sample spectra being fit. The fitted values are those which minimize a bin-wise χ^2 , and reported errors are those which generate a fixed increase in the χ^2 metric.

To propagate the uncertainty due to the lack of knowledge of the NP values, the fit on a given spectrum is performed several times with varying assumed values of the NPs drawn from the appropriate prior. The variation in the resulting fitted values then describes the impact of uncertainty on the NPs. For this reason, during each XSPEC fit, the NP values are not allowed to vary, but are frozen. Figure 5 shows examples for individual stars, demonstrating the variation of the stellar parameter estimates with varying NP values. For illustrative comparison to ad-hoc models of uncertainty, the standard deviation in mass and radius are used to define the widths of a 2d error ellipse, though it is clear that this fails to capture the complex nature of the impacts of the underlying uncertainties; these simple error models are not used in our analysis.

XSPEC is also capable of *floating* the nuisance parameters, varying their values to improve an individual fit, and reporting an uncertainty envelope in the mass-radius plane. This can be helpful in the case where the mass and radius and their envelope are the final targets. However, to propagate the uncertainty downstream requires that we have the full posterior in the mass-radius plane or samples from it. An estimate and envelope do not provide that capacity, though they can allow for ad-hoc parameterizations of the prior as have been performed previously. We condition on the nuisance parameters to allow full propagation of the NP uncertainty through to EOS estimation, as we do below.

Performance of XSPEC regression of mass and radius is shown in Figure 6, where the residuals increase as expected with wider priors on the nuisance parameters. In addition, note that in the case of the "loose" priors, there is a small fraction of cases, where XSPEC fails to converge on an estimate, as the nuisance parameters are fixed to a value far from the value used to generate the spectrum.

B. Inference of EOS from mass and radius

In this section, we demonstrate the inference of EOS parameters from the stellar mass and radius data, using



FIG. 5: Estimation of the mass and radius of a neutron star from the underlying stellar spectra, by XSPEC. Each pane represents one star, and shown (green) are estimates for several independent values of the nuisance parameters drawn from the associated priors, and the mean value (red). Top two cases have loose priors, bottom two have tight. The dashed ellipse, whose widths are set to the standard deviation of the mass and radius estimates, is a demonstration of the inadequacy of a simple uncertainty model.



FIG. 6: Performance of XSPEC inference of neutron star mass and radius, as measured by residuals between the fitted ("pred") and true values under three treatments of the nuisance parameters (NPs). In the "true" case, the NPs are fixed to their true values; in the "tight" and "loose" cases, they are drawn from narrow or wide priors, respectively; see text for details. Cases in which XSPEC fails to converge are shown as large negative residuals.

neural network regression parameterized in the NPs to allow propagation of the uncertainty. In addition, we build polynomial regression models which serve as a benchmark, following Ref. [23].

1. Neural Network regression

Deep feed-forward neural networks are trained to provide the EOS parameters given a collection of ten stars, each represented by their mass and radius. Each network has two outputs, λ_1 and λ_2 .

We train three networks, one for each of the true, tight, and loose NP scenarios. All networks have identical architecture, 10 hidden layers with 32 nodes each followed by an output layer with 2 nodes. Rectified linear units are used as activation functions for the hidden layers while linear activations are used for the output layers. They were trained up to 1000 epochs with a mean squared error loss and an Adam optimizer [66], and the performance is evaluated on independent validation data. The performance was not found to be highly sensitive to hyperparameter tuning, so permutation symmetry preserving architectures were not explored. The networks were implemented using **Tensorflow 2.7.0** on a single NVIDIA RTX A5000 GPU.

2. Polynomial regression

Following the example of Ref [23], as a performance benchmark we also construct a polynomial regression model to regress EOS parameters λ_1 and λ_2 from stellar mass and radius information. The input for each model is a 20 x 1 vector containing the following information from each of ten stars: mass M, radius R.

Each 20 x 1 input to the polynomial regression network represents a set of stars, all of which are chosen from the same EOS. We construct two multivariate polynomial regression models of degree two, following the form:

$$\lambda_i = \beta_0 + \sum_{i=1,2}^N \beta_i x_1^i$$

where each β are coefficients and N = 20 to represent the cluster size, or 10 mass-radius pairs. Polynomial features were created using the machine learning toolbox scikit-learn [67], and subsequently fit to a linear regression model. This model uses optimization in the form of ordinary least squares, which takes the form:

$$\min_{\beta} ||X\beta - y||_2^2$$

where X is the N = 20 input vector, and y is the target EOS parameter, either λ_1 or λ_2 .

3. Estimation of Uncertainty

The uncertainty in the underlying NS nuisance parameters has a significant impact on the estimation of EOS parameters. Because the mass and radius estimation is conditioned on the NPs, leading to variations in the mass and radius, (eg see Figure 5), those variations can be propagated through the EOS estimation. The significant uncertainties in stellar nuisance parameters and the small number of stars observed to date make the treatment of those uncertainties vital. The validity of the final result is only as powerful as the validity of its uncertainties. In Figure 7, examples of the variation of the EOS estimates are shown, where the underlying stellar spectra are fixed. Thus, this provides a measure of the uncertainty in the EOS due to the uncertainty in the NPs.

4. Performance

Performance of neural network regression of the EOS parameters λ_1 and λ_2 are compared to polynomial regression of the same quantities using identical datasets via comparison of the residuals, the difference between the true and regressed values. As seen in Figures 8 and 9, while PR is able to achieve narrower residuals in the true case, the network regression is more robust in cases with larger uncertainties.

This result confirms what has been seen in earlier studies of NN regression from mass-radius pairs [23], but our study extends previous work by using realistic values of the mass and radius inferred from realistic simulated spectra, as well as by demonstrating uncertainty quantification, via full propagation of the underlying uncertainties due to nuisance parameters.



FIG. 7: Neural network regression of the EOS parameters λ_1 and λ_2 of a set of 10 neutron stars from from their masses and radii as estimated by XSPEC from each stars spectrum. Each pane represents an example dataset of 10 simulated stars, and shown (green) are EOS estimates for several independent values of the stellar nuisance parameters drawn from the associated priors, and the mean value (red). Top two cases have loose priors, bottom two have tight.



regression and polynomial regression of EOS parameter λ_1 from mass radius pairs inferred by XSPEC from stellar spectra. Shown is the residual, the difference between the predicted and true values for each of three treatments of the stellar nuisance parameters. In the "true" case, the NPs are fixed to their true values; in the "tight" and "loose" cases, they are drawn from narrow or wide priors, respectively; see text for details.



FIG. 9: Comparison of the performance of NN regression and polynomial regression of EOS parameter λ_2 from mass radius pairs inferred by XSPEC from stellar spectra. Shown is the residual, the difference between the predicted and true values for each of three treatments of the stellar nuisance parameters. In the "true" case, the NPs are fixed to their true values; in the "tight" and "loose" cases, they are drawn from narrow or wide priors, respectively; see text for details.

VI. INFERENCE OF MASS AND RADIUS FROM SPECTRA

Previous applications of machine learning to neutron star datasets focus on analysis of mass-radius pairs, as demonstrated above, rather than direct analysis of the stellar spectra by neural networks.

A potential obstacle to direct analysis of spectra by neural networks is that the spectra are high-dimensional, often with $\mathcal{O}(10^3)$ bins of photon energy. However deep learning methods combined with GPUs have no trouble analyzing data with similarly high dimensionality [21], opening up new opportunities to tackle this important topic. While the mass and radius are powerful summaries of the information in the lower-level spectra which is relevant to the equation of state, direct ML analysis of the spectra themselves may allow for the extraction of additional information or provide more robust propagation of uncertainties. As an initial step, we begin by estimating the stellar mass and radius from a single stellar spectrum before moving on to end-to-end inference of EOS parameters directly from a set of spectra in the next section.

In this section, we apply machine learning to the task of extracting the mass and radius from the stellar spectra, training a network we refer to as MR_Net . This serves as a demonstration of the capabilities of ML to grapple with high-dimensional datasets, allows us to harmonize the treatment of nuisance parameters end-to-end from spectra to EOS, and potentially extract more relevant information.

A. MR_Net Method

We build a network whose inputs correspond to the bins of the stellar X-ray spectrum, and whose outputs are the estimates of the star's mass and radius. In addition, the mass and radius regressor is parameterized on the stellar nuisance parameters (distance, N_H , log($T_{\rm eff}$)), which allows the results to be conditioned on the nuisance parameters.

This architecture is composed of two input branches, one to process the star's spectra and another to process the corresponding nuisance parameters. Each branch contains a series of layers that process its inputs in isolation. Following these initial layers, the output from the branches is combined, forming a single vector containing all the information. This vector is then passed to a final series of layers to predict the star's mass and radius.

Mass-radius regression is formulated as a supervised learning problem, where the network learns to minimize the error between the true mass and radius and its predictions. The network weights are updated by stochastic gradient descent using backpropagation. The Huber loss function is used and the Adam optimizer computes gradients and schedules the backward passes.

B. MR_Net Performance in Mass, Radius

We begin with the best-case scenario in which the nuisance parameters are known with zero uncertainty, referred to as "true NP" above. Figure 10 shows the performance of MR_Net given neutron star spectra with statistical noise corresponding to 100,000 seconds (100 ks) of observation time, as well as for spectra without statistical noise. This demonstrates the contribution of statistical noise to the residual and demonstrates the network's ability to digest the spectral information and understand the dependence on mass and radius.



FIG. 10: Performance of the MR_Net regression of a neutron star mass (top) and radius (bottom) from its stellar X-ray spectrum. Shown is the residual, the difference between the true and predicted values, for spectra with statistical noise corresponding to an observation time of 100ks, and for spectra without statistical noise, which demonstrates the capacity of the network. Nuisance parameters are fixed to their true values

What is clear is that MR_Net is capable of extracting the mass and radius values of the star directly from the spectrum. We emphasize that the network is trained on examples, but does not benefit from the knowledge of the theoretical model used to generate these stars, while XSPEC requires precise specification of the theoretical model. This lack of requirement of a theoretical model opens new possibilities, such as training MR_Net to interpolate smoothly between theoretical models, by providing a mixed or parameterized training set.

The MR_Net is conditioned on the nuisance param-



FIG. 11: Performance of the MR_Net regression of a neutron star mass from its stellar X-ray spectrum, compared to regression using XSPEC. Shown is the residual, the difference between the true and predicted values, for three scenarios of nuisance parameter

uncertainties. In the "true" case, the NPs are fixed to their true values; in the "tight" and "loose" cases, they are drawn from narrow or wide priors, respectively; see text for details.

eters, allowing for propagation of the NP uncertainties through to the regression target as was done with XSPEC estimates. To assess the impact of NP uncertainty, we draw from priors on the NPs under the "Tight" and "Loose" scenarios defined earlier. The residuals widen, as expected. Figures 11 and 12 show the mass and radius residuals, respectively, under each NP scenario for MR_Net and XSPEC. Table III shows the mean and width of each distribution, as well as the combined width.

MR_Net is capable of analyzing the stellar spectrum directly and extracting stellar parameters in a robust manner that allows for propagation of NP uncertainties.



FIG. 12: Performance of the MR_Net regression of a neutron star radius from its stellar X-ray spectrum, compared to regression using XSPEC. Shown is the residual, the difference between the true and predicted

values, for three scenarios of nuisance parameter uncertainties. In the "true" case, the NPs are fixed to their true values; in the "tight" and "loose" cases, they are drawn from narrow or wide priors, respectively; see text for details.

C. Network uncertainty on mass and radius

Conditioning the extraction of M and R on the nuisance parameters allows for the natural propagation of the corresponding uncertainties. As was done for XSPEC mass and radius estimates, we propagate the NP uncertainty through to mass and radius uncertainty by sampling from the stellar NP priors several times for a given stellar spectrum, performing the mass and radius regression multiple times. Figure 13 demonstrates this for several example stars. TABLE III: Performance of the regression of neutron star mass and radius for XSPEC as well as our neural

network regression, MR_Net , which lacks any knowledge of the theoretical model. Shown are the mean (μ) and standard deviation (σ) of the residuals

under three scenarios of nuisance parameter uncertainties. In the "true" case, the NPs are fixed to their true values; in the "tight" and "loose" cases, they are drawn from narrow or wide priors, respectively; see text for details. The combined column is a quadrature sum of the standard deviations of radius and mass.

	Nuis.	Mass	Radius	Combined
Method	Params	μ σ	μ σ	σ
XSPEC	True	-0.01 0.50	0.23 1.44	1.51
MR_Net	True	-0.14 0.93	-0.07 2.80	2.99
XSPEC	Tight	-0.06 0.73	$0.24 \ 2.61$	2.69
MR_Net	Tight	$0.17 \ 1.06$	$0.06 \ 3.52$	2.76
XSPEC	Loose	$0.18 \ 0.86$	-0.06 4.32	4.40
MR_Net	Loose	$0.28\ 1.29$	$0.14 \ 4.93$	5.10

D. Revisiting EOS regression

The question driving the analysis of neutron star spectra is not a desire to measure their masses and radii, but to use those to determine the equation of state parameters. In this section, we push the results of MR_Net through our NN regression of EOS parameters to analyze the performance of spectra $\rightarrow (M, R) \rightarrow \text{EOS}$ regression. In later sections, we remove the intermediate step and perform direct spectra $\rightarrow \text{EOS}$ regression.

Performance of the EOS regression using stellar mass and radius information from MR_Net is shown in Figures 17 and 18, and compared to regression from mass and radius information by XSPEC. Examples of uncertainty propagation through to EOS estimates are shown in Figure 14.



FIG. 13: Estimation of the mass and radius of a neutron star from the underlying stellar spectra, by MR_Net . Each pane represents one star, and shown (green) are estimates for several independent values of the nuisance parameters drawn from the associated priors, and the mean value (red). Top two cases have loose priors, bottom two have tight. The dashed ellipse, whose widths are set to the standard deviation of the mass and radius estimates, is a demonstration of the inadequacy of a simple uncertainty model.



FIG. 14: Neural network regression of the EOS parameters λ_1 and λ_2 of a set of 10 neutron stars from from their masses and radii as estimated by MR_Net from each stars spectrum. Each pane represents an example dataset of 10 simulated stars, and shown (green) are EOS estimates for several independent values of the stellar nuisance parameters drawn from the associated priors, and the mean value (red). Top two cases have loose priors, bottom two have tight.

VII. INFERENCE OF EOS FROM SPECTRA

In the previous section, we connected the two ML models, MR_Net (spectra $\rightarrow M, R$) and our EOS regression NN ($M, R \rightarrow EOS$). This required the somewhat artificial collapse of the full information and dependence on nuisance parameters into two numbers per star. At the same time, MR_Net demonstrated that it is possible to regress physics quantities directly from high-dimensional stellar spectra. In this section, rather than connecting two networks, we use a single network to perform endto-end regression of EOS parameters from a set of stellar spectra, avoiding the information collapse, keeping the full information, and allowing for robust propagation of the stellar nuisance parameters into uncertainty quantification for the EOS parameters of interest.

In addition to a demonstration of the power of networks to directly analyze low-level data, this allows us to probe the question of whether the mass and radius are sufficient statistics, and whether they contain *all* of the information relevant to the problem. There are many examples in the literature in which such well-motived highlevel heuristics fail to capture the complete information contained in lower-level data. In this case, while in principle the mass and radius are all that are required to infer the EOS, such information is never without uncertainty. Full propagation of the dependence on nuisance parameter uncertainty may allow for more accurate and robust estimates.

Furthermore, there are properties of neutron stars that can be deciphered from spectra beyond simply mass and radius. Quantities like temperature inhomogeneities [68] may impact a star's equation of state but are not captured by the mass and radius.

A. Architecture

Many neural network architectures operate on sequences of vectors, rather than set of vectors. For instance, in natural language processing, the input may be a sentence where each word is converted to a vector and the ordering of the vectors matters. However, in the case of neutron stars and other problems, we need neural networks that operate on sets of vectors, such as the independent spectra observations for multiple stars.

One architectures with invariant properties with respect to permutations of the input vectors is the transformer architecture [51, 69]. Perhaps surprisingly, transformers were originally developed for problems in natural language processing, thus requiring the addition of positional information bits to the vector encoding each word in order to recover the sequential dimension. More recently they have been used in other areas, including physics [70, 71] in order to leverage their permutation invariance properties. Transformer architectures typically consists of stacks of encoder modules followed by decoder modules. The structure of each encoder module and each



FIG. 15: Schematic diagram of the transformer neural network used to determine EOS coefficients λ_1 and λ_2 from an input of spectra and NPs.

decoder module is similar, thus we describe only a typical encoder module. A transformer encoder module accepts an unordered set of inputs and produces a set of outputs. The transformer employs a mechanism called *self-attention*, which allows it to compare each element in the set against every other. This mechanism allows the network to attend to important features in the set while computing an output prediction. Briefly, self attention (Equation 9) operates on an input matrix X, with N rows and D columns. Three matrices are produced from the projection of X with differing, trainable, weight matrices: $Q = XW_Q, K = XW_K, V = XW_V$ (termed Query, Key, and Value respectively).

$$S = D(Q, K, V) = \operatorname{softmax}\left(\frac{QK^T}{\sqrt{d_q}}\right)V \qquad (9)$$

Thus in short each output corresponds to a different convex combination of the Value vectors, where each convex combination depends on the degree of similarity between the corresponding Query and Key vectors. The similarity is computed by taking dot products between corresponding Query and Key vectors, and then applying a softmax to yield a convex combination (see [51, 69] for additional details).

In the neutron star application, the transformer architecture takes as input spectra and corresponding nuisance parameters for each star in the set. In the results shown below, the network is given a set of 10 stars, though the structure of the network allows it to accept larger or smaller datasets with minimal modification. The final output of the network is the two EOS parameters.

B. Training

The parameters of the network architecture and the learning algorithm, the hyperparameters, were optimized with Sherpa [72], a Python library for hyperparameter tuning. The algorithm used is based on a random search and has the advantage of making no assumptions about the structure of the hyperparameter search problem and thus is well suited for exploring a variety of parameter settings. An initial exploratory search was conducted on a subset of the data to find appropriate hyperparameters.

Following this exploratory phase, the network is trained for 1,000 epochs on the full dataset. The Adam optimizer [66] is used for gradient descent, with early stopping monitoring of the validation loss to prevent overfitting.

C. Results

Figure 16 shows the performance of spectra \rightarrow EOS regression in the best-case scenario, where the nuisance parameters are perfectly known. Shown is the performance with statistical noise corresponding to 100 ks of observation time per star, as well as for spectra without statistical noise. While there is significant width to the residuals, this is dominated by the statistical uncertainty, not the network's ability to digest the spectral information and understand the dependence on mass and radius. This clearly demonstrates the network's capacity is sufficient for the regression task.

We next analyze the performance of the direct regression in cases where the nuisance parameters are not perfectly determined. Figures 17 and 18 shows the residuals in the EOS parameters for the end-to-end regression, as compared to regression from mass and radius information provided by MR_Net or XSPEC from the stellar spectra. Table IV summarizes the performance for each method.

As the full network is again conditioned on the NPs, we can propagate this uncertainty directly through our regression. Figure 19 demonstrates how variations of the NPs, drawn from the appropriate priors, provide a measure of the uncertainty on the final result.

VIII. DISCUSSION

Networks that regress the EOS parameters directly from a set of stellar spectra, without any knowledge of the theoretical model used to generate the simulated data, perform comparably to those which regress the EOS parameters from mass and radius information extracted from XSPEC fits which rely on knowledge of those theoretical stellar models. Interestingly, the performance depends on the uncertainty in the underlying nuisance parameters, where the end-to-end regression networks outperform those relying on XSPEC fits when the uncertainties are large, the more realistic scenario.



FIG. 16: Performance of the neural network regression of the neutron star EOS parameters λ_1 (top) and λ_2 (bottom) directly from a set of stellar X-ray spectra, without intermediate prediction of the mass and radius. Shown is the residual, the difference between the true and predicted values, for spectra with statistical noise corresponding to an observation time of 100k seconds per star, and for spectra without statistical noise, which demonstrates the capacity of the network. Nuisance parameters are fixed to their true values.

When the data are simulated and drawn from a known theoretical model, one cannot achieve more statistical power than directly calculating the likelihood. But even powerful theoretical models for spectral fitting still rely on a variety of assumptions about the spectrum's source. Having a network that is flexible to inputs and demonstrates a robust propagation of uncertainties is vital.

As a further visualization, Figure 20 shows several example curves in the mass-radius plane fitted to the same stellar spectra with varying nuisance parameters. TABLE IV: Performance of the regression of neutron star EOS parameters λ_1 and λ_2 using direct regression from spectra, as compared to NN regression from mass and radius (M, R) information extracted via MR_Net or XSPEC. Shown are the mean (μ) and standard deviation (σ) of the residuals under three scenarios of nuisance parameter uncertainties; distributions are given in Figures 17 and 18. In the "true" case, the NPs are fixed to their true values; in the "tight" and "loose" cases, they are drawn from narrow or wide priors, respectively; see text for details. The combined column is a quadrature sum of the standard deviations of λ_1 and λ_2 .

	Nuis.	λ_1	λ_2	Combined
				-
Method	Params.	μ σ	μ σ	σ
NN(Spectra)	True	-0.02 0.066	$0.01 \ 0.075$	0.075
$NN(M, R \text{ via } MR_Net)$	True	-0.03 0.089	-0.02 0.068	0.068
NN(M, R via XSPEC)	True	-0.03 0.065	$0.01 \ 0.055$	0.055
NN(Spectra)	Tight	$0.02 \ 0.085$	-0.02 0.077	0.077
$NN(M, R \text{ via } MR_Net)$	Tight	$0.00 \ 0.104$	$0.02 \ 0.072$	0.072
NN(M, R via XSPEC)	Tight	-0.03 0.081	$0.01 \ 0.056$	0.056
NN(Spectra)	Loose	-0.03 0.131	-0.01 0.078	0.152
$NN(M, R \text{ via } MR_Net)$	Loose	-0.01 0.135	-0.02 0.078	0.159
NN(M, R via XSPEC)	Loose	-0.03 0.123	$0.01 \ 0.058$	0.163

IX. CONCLUSIONS

We have demonstrated the network regression of EOS parameters from realistic neutron star mass and radius estimates drawn from simulated stellar spectra. Our approach of conditioning each step on nuisance parameters allows us to fold in the NP uncertainty via multiple sampling from priors and permits full propagation of the uncertainty through to the final regression targets. In addition, we have shown that networks can analyze high-dimensional telescope data directly, including sets of multiple stars, and achieve comparable performance to methods that assume perfect knowledge of the theoretical model used to generate the simulated samples. In realistic cases where the nuisance parameter uncertainties are significant, the proposed end-to-end network regression achieves greater precision than the state of the art.

These results suggest many future directions. Our networks are parameterized in the nuisance parameters, allowing for the propagation of prior uncertainties which are implicitly derived from auxiliary data. But the stellar spectra may also offer information that constrains the NP uncertainty. Profiling over the nuisance parameters could reduce this uncertainty, though it may be computationally very expensive without neural likelihood estimation techniques [73].

Alternatively, rather than employing regression to directly produce estimates of the EOS parameters, one might train a generative model to operate as a surrogate of the likelihood [74], allowing for fast evaluation of the likelihood as a function of the EOS parameters and potentially direct profiling.

Other future directions for the networks described in this paper would be to test more exotic neutron star equations of state, including those with phase transitions. Including additional parameters will play a key role in conducting similar research using alternative models in XSPEC that rely on different nuisance parameters, like a Helium atmospheric model. Even more interesting would be extending this type of EOS inference to other compact objects like white dwarfs.

X. ACKNOWLEDGEMENTS

DW and AG are supported by The Department of Energy Office of Science. LL was supported by NSF Grant No. 2012857 to the University of California at San Diego. AWS was supported by NSF AST 19-09490, PHY 21-16686, and the Department of Energy Office of Nuclear Physics. DF and FW are supported by the National Science Foundation (USA) under Grant No. PHY-2012152. The authors are grateful to C. O. Heinke and W. C. G. Ho for their assistance in understanding and navigating the XSPEC software.





0.0

 $\lambda_{1,true} - \lambda_{1,pred}$

-ó.2

NN from Spectra, True NP

NN from MR Net, True NP

NN from XSPEC, True NP

0.2

0.4

10

101

100

10-

10

-0.4

Fraction



FIG. 17: Performance of the regression of neutron star EOS parameter λ_1 using direct regression from spectra, as compared to regression from mass and radius information extracted via MR_Net or XSPEC. Shown are the residual distributions, the difference between the true and predicted values, under three scenarios of nuisance parameter uncertainties. See Table IV for quantitative analysis. In the "true" case, the NPs are fixed to their true values; in the "tight" and "loose" cases, they are drawn from narrow or wide priors, respectively (see text for details).



NN from Spectra, True NP

NN from MR Net, True NP

■ NN from XSPEC, True NP

FIG. 18: Performance of the regression of neutron star EOS parameter λ_2 using direct regression from spectra, as compared to regression from mass and radius information extracted via MR_Net or XSPEC. Shown are the residual distributions, the difference between the true and predicted values, under three scenarios of nuisance parameter uncertainties. See Table IV for quantitative analysis. In the "true" case, the NPs are fixed to their true values; in the "tight" and "loose" cases, they are drawn from narrow or wide priors, respectively (see text for details).



FIG. 19: Neural network regression of the EOS parameters λ_1 and λ_2 of a set of 10 neutron stars directly from the set of stellar spectra. Each pane represents an example dataset of 10 simulated stars, and shown (green) are EOS estimates for several independent values of the stellar nuisance parameters drawn from the associated priors, and the mean value (red). Top two cases have loose priors, bottom two have tight.



FIG. 20: Demonstration of the impact of nuisance parameters on the regression of EOS parameters for three sets of observed stellar spectra. In red are the true mass and radius of the stars, drawn from the mass-radius curve determined by the true EOS parameters. In blue dots are the values of the mass and radius deduced by MR_Net for each star given a set of stellar NPs drawn from the priors; the blue line shows the mass-radius curve corresponding to the EOS

parameters deduced directly from the stellar spectra and NPs by the proposed end-to-end regression. The results of MR_Net are not used in the EOS regression, and only appear to aid the visualization. Brown and green are similar to blue, but for independent draws of the NPs from the same stellar priors.

- L. Tolos and L. Fabbietti, Prog. Part. Nucl. Phys. 112, 103770 (2020), 2002.09223.
- [2] J. J. Li, A. Sedrakian, and F. Weber, Physics Letters B 783, 234 (2018).
- [3] W. M. Spinella and F. Weber, Astronomische Nachrichten 340, 145 (2019).
- [4] K. Fukushima and C. Sasaki, Prog. Part. Nucl. Phys. 72, 99 (2013), 1301.6377.
- [5] M. Orsaria, H. Rodrigues, F. Weber, and G. Contrera, Physical Review C 89, 015806 (2014).
- [6] M. G. Alford, A. Schmitt, K. Rajagopal, and T. Schäfer, Rev. Mod. Phys. 80, 1455 (2008), 0709.4635.
- [7] J. Zdunik and P. Haensel, Astronomy & Astrophysics 551, A61 (2013).
- [8] G. Baym, Physical Review Letters **30**, 1340 (1973).
- D. Kaplan and A. Nelson, Physics Letters B 175, 57 (1986), ISSN 0370-2693, URL https: //www.sciencedirect.com/science/article/pii/ 037026938690331X.
- [10] P. J. Ellis, R. Knorren, and M. Prakash, Physics Letters B 349, 11 (1995).
- [11] J. L. Friedman and J. R. Ipser, Astrophys. J. **314**, 594 (1987).
- [12] R. E. Rutledge, L. Bildsten, E. F. Brown, G. G. Pavlov, and V. E. Zavlin, Astrophys. J. **514**, 945 (1999), astroph/9810288.
- [13] C. O. Heinke, G. B. Rybicki, R. Narayan, and J. E. Grindlay, Astrophys. J. **644**, 1090 (2006), astroph/0506563.
- [14] J. M. Lattimer and M. Prakash, Astrophys. J. 550, 426 (2001), URL https://doi.org/10.1086/319702.
- [15] L. Lindblom, Phys. Rev. D 82, 103011 (2010), URL https://link.aps.org/doi/10.1103/PhysRevD. 82.103011.
- [16] A. W. Steiner, J. M. Lattimer, and E. F. Brown, Astrophys. J. **722**, 33 (2010), 1005.0811, URL http://doi. org/10.1088/0004-637X/722/1/33.
- [17] L. Lindblom and N. M. Indik, Phys. Rev. D 89, 064003 (2014), [Erratum: Phys.Rev.D 93, 129903 (2016)], 1310.0803.
- [18] P. Baldi, <u>Deep Learning in Science</u> (Cambridge University Press, Cambridge, UK, 2021).
- [19] P. Baldi, P. Sadowski, and D. Whiteson, Nature Communications 5 (2014).
- [20] A. Ghosh, B. Nachman, and D. Whiteson, Phys. Rev. D 104, 056026 (2021), 2105.08742.
- [21] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, Eur. Phys. J. C 76, 235 (2016), 1601.07913.
- [22] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, Phys. Rev. D 94, 112002 (2016), 1607.08633.
- [23] Y. Fujimoto, K. Fukushima, and K. Murase, Phys. Rev. D 101, 054016 (2020), 1903.03400.
- [24] Y. Fujimoto, K. Fukushima, and K. Murase, JHEP 03, 273 (2021), 2101.08156.
- [25] F. Morawski and M. Bejger, Astron. Astrophys. 642, A78 (2020), 2006.07194.
- [26] M. Ferreira and C. Providência (2019), 1910.05554.
- [27] J. M. Lattimer, Universe 5 (2019), ISSN 2218-1997, URL https://www.mdpi.com/2218-1997/5/7/159.
- [28] B. Müller and J. L. Nagle, Annu. Rev. Nucl. Part. Sci.

56, 93 (2006).

- [29] G. Aad, E. Abat, J. Abdallah, A. Abdelalim, A. Abdesselam, B. Abi, M. Abolins, H. Abramowicz, E. Acerbi, B. Acharya, et al., Journal of instrumentation 3 (2008).
- [30] P. Spiller and G. Franchetti, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 561, 305 (2006).
- [31] X. Roca-Maza, M. Centelles, X. Viñas, and M. Warda, Phys. Rev. Lett. **106**, 252501 (2011), URL https:// link.aps.org/doi/10.1103/PhysRevLett.106.252501.
- [32] N. Glendenning, <u>Compact Stars: Nuclear Physics, Particle Physics as</u> Astronomy and Astrophysics Library (Springer New York, 2012), ISBN 9781468404913.
- [33] A. L. Watts, N. Andersson, D. Chakrabarty, M. Feroci, K. Hebeler, G. Israel, F. K. Lamb, M. C. Miller, S. Morsink, F. Özel, et al., Rev. Mod. Phys. 88, 021001 (2016), URL https://link.aps.org/doi/10. 1103/RevModPhys.88.021001.
- [34] R. Machleidt and D. R. Entem, Phys. Rept. 503, 1 (2011), 1105.2919.
- [35] E. S. Fraga, A. Kurkela, and A. Vuorinen, Eur. Phys. J. A 52 (2016), URL https://link.springer.com/article/ 10.1140/epja/i2016-16049-6.
- [36] J. R. Oppenheimer and G. M. Volkoff, Phys. Rev. 55, 374 (1939), URL https://link.aps.org/doi/10.1103/ PhysRev.55.374.
- [37] R. C. Tolman, Phys. Rev. 55, 364 (1939), URL https: //link.aps.org/doi/10.1103/PhysRev.55.364.
- [38] S. Bogdanov, C. O. Heinke, F. Özel, and T. Güver, Astrophys. J. 831, 184 (2016), 1603.01630.
- [39] S. Campana, M. Colpi, S. Mereghetti, L. Stella, and M. Tavani, The Astronomy and Astrophysics Review 8, 279 (1998).
- [40] A. Y. Potekhin, Physics-Uspekhi 57, 735 (2014).
- [41] C. O. Heinke, J. E. Grindlay, P. M. Lugger, H. N. Cohn, P. D. Edmonds, D. A. Lloyd, and A. M. Cool, The Astrophysical Journal 598, 501 (2003), URL https: //doi.org/10.1086/378885.
- [42] E. F. Brown, L. Bildsten, and R. E. Rutledge, The Astrophysical Journal 504, L95 (1998).
- [43] K. Hebeler, J. Lattimer, C. J. Pethick, and A. Schwenk, The Astrophysical Journal 773, 11 (2013).
- [44] A. W. Steiner, C. O. Heinke, S. Bogdanov, C. Li, W. C. G. Ho, A. Bahramian, and S. Han, Mon. Not. Roy. Astron. Soc. 476, 421 (2018), 1709.05013, URL https://doi.org/10.1093/mnras/sty215.
- [45] P. Slane, D. Helfand, and S. Murray, Astrophysical Journal - ASTROPHYS J 571 (2002).
- [46] R. Wijnands, N. Degenaar, and D. Page, Journal of Astrophysics and Astronomy 38, 1 (2017).
- [47] H. Sun, Y. Li, B.-B. Zhang, B. Zhang, F. Bauer, Y. Xue, and W. Yuan, The Astrophysical Journal 886, 129 (2019).
- [48] K. A. Arnaud, in <u>Astronomical Data Analysis Software and Systems</u> edited by G. H. Jacoby and J. Barnes (1996), vol. 101 of <u>Astronomical Society of the Pacific Conference Series</u>, p. 17.
- [49] B. Neyshabur, S. Bhojanapalli, D. McAllester, and N. Srebro, Advances in neural information processing systems **30** (2017).

- [50] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, in <u>Advances in Neural Information Processing Systems 30</u>, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017), pp. 5998–6008, URL http://papers.nips.cc/paper/ 7181-attention-is-all-you-need.pdf.
- [51] P. Baldi and R. Vershynin (2022), submitted. Also arXiv:2202.08371.
- [52] S. Typel, G. Röpke, T. Klähn, D. Blaschke, and H. H. Wolter, Phys. Rev. C 81, 015803 (2010), URL https: //link.aps.org/doi/10.1103/PhysRevC.81.015803.
- [53] W. M. Spinella and F. Weber, Dense Baryonic Matter in the Cores of Neutron Stars (World Scientific, 2020), chap. 4, pp. 85–152.
- [54] G. Malfatti, M. G. Orsaria, G. A. Contrera, F. Weber, and I. F. Ranea-Sandoval, Phys. Rev. C 100, 015803 (2019), URL https://link.aps.org/doi/10. 1103/PhysRevC.100.015803.
- [55] M. Oertel, M. Hempel, T. Klähn, and S. Typel, Rev. Mod. Phys. 89, 015007 (2017), URL https://link.aps. org/doi/10.1103/RevModPhys.89.015007.
- [56] S. Shlomo, V. M. Kolomietz, and G. Colo, **30**, 23 (2006).
- [57] U. Garg and G. Colò, Progress in Particle and Nuclear Physics 101, 55 (2018), URL https://doi.org/ 10.1016%2Fj.ppnp.2018.03.001.
- [58] J. R. Stone, N. J. Stone, and S. A. Moszkowski, Phys. Rev. C 89, 044316 (2014).
- [59] G. Baym, C. Pethick, and P. Sutherland, Astrophys. J. 170, 299 (1971).
- [60] G. Baym, H. A. Bethe, and C. J. Pethick, Nuclear Physics A 175, 225 (1971), ISSN 0375-9474, URL https://www.sciencedirect.com/science/ article/pii/0375947471902818.
- [61] L. Lindblom, Phys. Rev. D 97, 123019 (2018), URL https://link.aps.org/doi/10.1103/PhysRevD. 97.123019.
- [62] B. Fornberg and J. Zuev, Computers & Mathematics with Applications 54, 379 (2007).

- [63] W. Η. Press, S. Α. Teukolsky, W. Т. Vetterling, and В. Р Flannery, Numerical Recipes 3rd Edition: The Art of Scientific Computing (Cambridge University Press, 2007),3rd ed., ISBN 0521880688, URL http://www.amazon.com/ Numerical-Recipes-3rd-Scientific-Computing/ dp/0521880688/ref=sr_1_1?ie=UTF8&s=books&qid= 1280322496&sr=8-1.
- [64] C. O. Heinke, G. B. Rybicki, R. Narayan, and J. E. Grindlay, The Astrophysical Journal 644, 1090 (2006), URL https://doi.org/10.1086/503701.
- [65] J. M. Lattimer and A. W. Steiner, Astrophys. J. 784, 123 (2014), 1305.3242, URL http://doi.org/10.1088/ 0004-637X/784/2/123.
- [66] D. P. Kingma and J. Ba, <u>Adam: A method for stochastic optimization</u> <u>URL https://arxiv.org/abs/1412.6980.</u> (2014),
- [67] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Journal of Machine Learning Research 12, 2825 (2011).
- [68] K. Elshamouty, C. Heinke, S. Morsink, S. Bogdanov, and A. Stevens, The Astrophysical Journal 826, 162 (2016).
- [69] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, in <u>Advances in neural information processing systems</u> (2017), pp. 5998–6008.
- [70] M. J. Fenton, A. Shmakov, T.-W. Ho, S.-C. Hsu, D. Whiteson, and P. Baldi, Physical Review D 105, 112008 (2022), also arXiv:2010.09206.
- [71] A. Shmakov, M. J. Fenton, T.-W. Ho, S.-C. Hsu, D. Whiteson, and P. Baldi, SciPost Phys. 12, 178 (2022), URL https://scipost.org/10.21468/ SciPostPhys.12.5.178.
- [72] L. Hertel, J. Collado, P. Sadowski, J. Ott, and P. Baldi, SoftwareX 12, 100591 (2020).
- [73] L. Heinrich (2022), 2203.13079.
- [74] K. Cranmer, J. Brehmer, and G. Louppe, Proc. Nat. Acad. Sci. 117, 30055 (2020), 1911.01429.