Neural simulation-based inference of the neutron star equation of state directly from telescope spectra

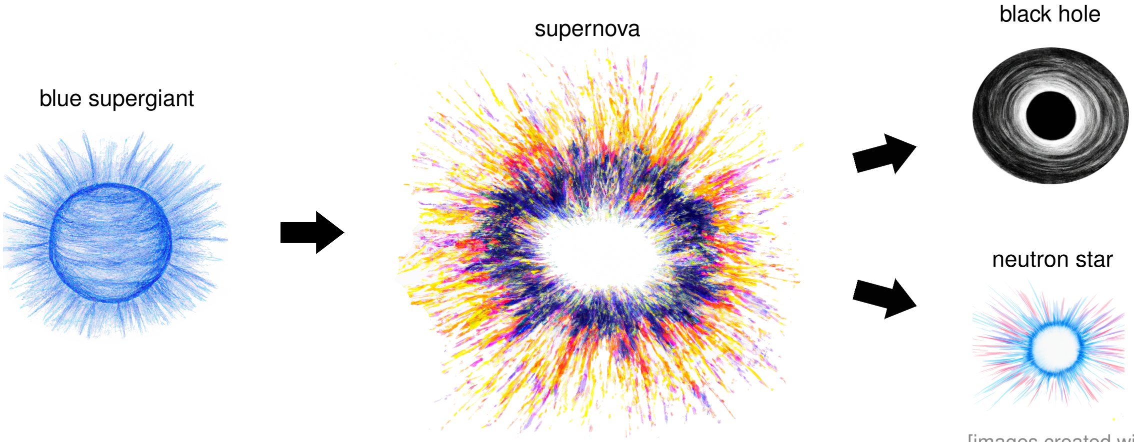
Continuous gravitational waves and neutron stars

Len Brandes, Chirag Modi, Aishik Ghosh, Delaney Farrell, Lee Lindblom,
Lukas Heinrich, Andrew W. Steiner, Fridolin Weber, Daniel Whiteson

19.06.2024

Technische Universität München

Neutron stars



[images created with DALL·E, openAI]

- Neutron stars weight $M \sim 1 - 2\,M_\odot$ and have radii $R \sim 11 - 13\,\text{km}$
- **High baryon densities** in core beyond terrestrial experiments

→ Constrain unknown physics of dense matter based on neutron star observations

[Brandes, Kaiser and Weise, Phys. Rev. D 108 (2023)]

Description of neutron stars

- Internal structure described by Tolman-Oppenheimer-Volkoff (TOV) equations

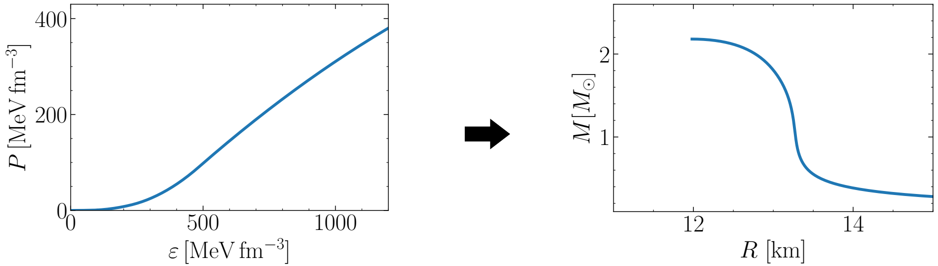
$$\frac{\partial P(r)}{\partial r} = -\frac{G_N}{r^2}\left(\varepsilon(r) + P(r)\right)\left(m(r) + 4\pi r^3 P(r)\right)\left(1 - \frac{2G_N m(r)}{r}\right)^{-1}$$

$$\frac{\partial m(r)}{\partial r} = 4\pi r^2 \varepsilon(r)$$

[Tolman, Phys. Rev. 55 (1939)] [Oppenheimer and Volkoff, Phys. Rev. 55 (1939)]

- Solution for given **equation of state (EoS)** $\boldsymbol{P(\varepsilon)}$ and central energy density $\varepsilon(r=0) = \varepsilon_c$

  → Solution for different ε_c yields $\boldsymbol{(M, R)}$**-relation**

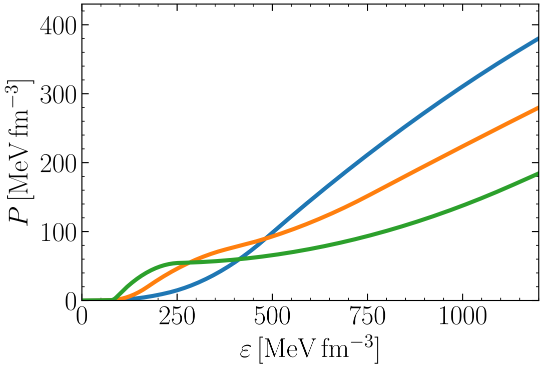


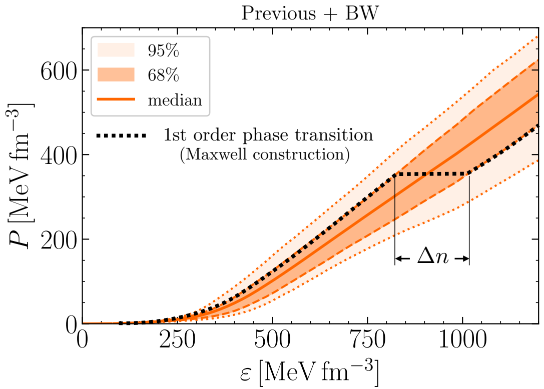
Equation of state

- EoS depends on the internal phases: phase transitions, crossover, ... [Kojo, AAPPS Bull. 31 (2021)]
- Introduce **general parametrization $P(\varepsilon, \lambda)$** that can describe variety of scenarios

  → Here: spectral parametrization with two parameters (λ_1, λ_2) [Lindblom, Phys. Rev. D 82 (2010)]
- **Bayesian inference of λ based on neutron star data $\mathscr{D}$**

$$p(\lambda|\mathscr{D}) \propto p(\mathscr{D}|\lambda)\, p(\lambda)$$





[Brandes and Weise, Symmetry 16 (2024)]

Thermal X-ray spectra

- Thermal **X-ray spectra *s*** of low-mass binaries in quiescence (qLMXBs)

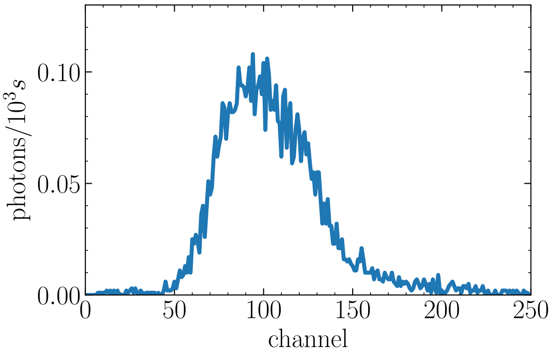
  → Depend on **mass-radius (M, R)** and **nuisance parameters** $\boldsymbol{\nu} = (N_H, d, \log(T_{\text{eff}}))$

  → Information on ν from other observations

- **Likelihood $p(s|\lambda, \nu)$ analytically intractable**, inference in two steps:
  [Riley, Raaijmakers and Watts, MNRAS 478 (2018)]

  1. Infer (M, R) from telescope spectrum s using XSPEC package
     [Arnaud, ASP Conf. Ser. 17 (1996)]
  2. *Use posteriors $p(M, R|s)$ as likelihoods to infer EoS*

$$p(s|\lambda) \propto \int dM\,dR\ \ p(M, R|s)\,\delta(R - R(M, \lambda))$$



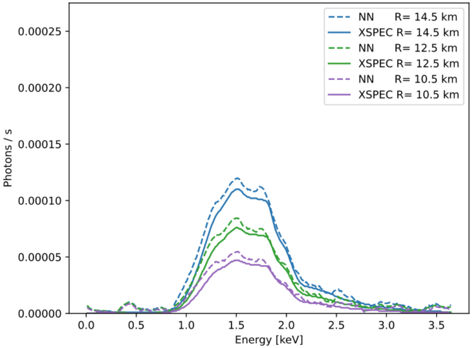
Inference directly from telescope spectra

- Can **sample from likelihood**: for given λ_i sample (M_i, R_i) and (independently) ν_i

  → **Simulate spectrum $s_i \sim p(s|\lambda_i, \nu_i)$** with XSPEC and add noise

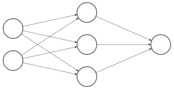
  → Here use `NSATMOS` model and Chandra response function

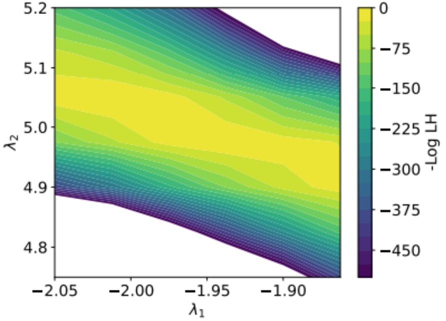
  [Heinke *et al.*, Astrophys. J. 644 (2006)]



- **Train neural network to regress (λ_1, λ_2)** from spectrum and nuisance parameters

  [Farrell *et al.*, JCAP 02 (2023)]

$\begin{bmatrix} s \\ \nu \end{bmatrix}$  λ



- **Train neural network to predict spectra *s*** for given (M, R, ν)

  → Approximate likelihood $p(s|M, R, \nu)$

  → Compute $p(s|\lambda, \nu)$ with second neural network

  [Farrell *et al.*, JCAP 12 (2023)]

Neural likelihood estimation

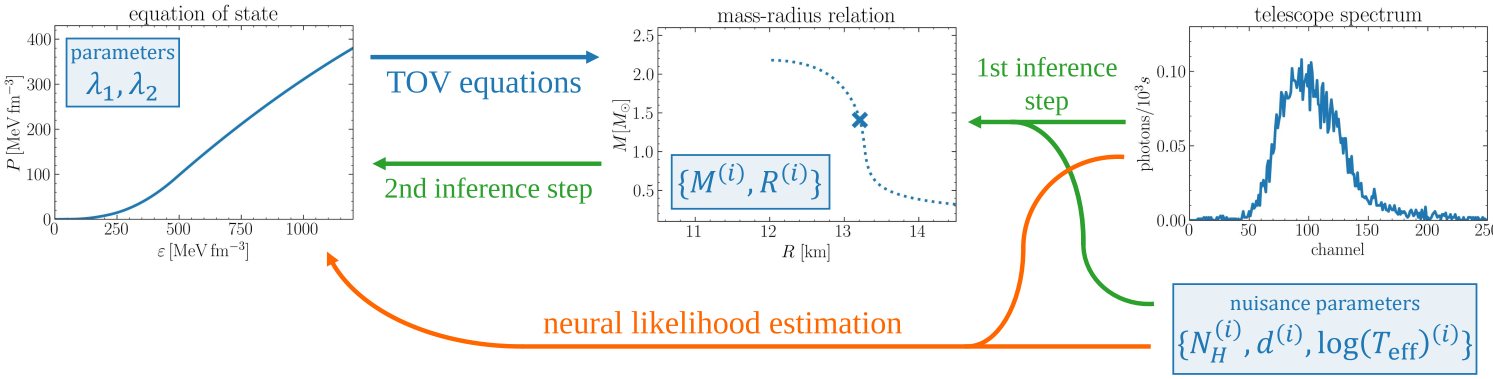
- Previous approaches only point estimates of EoS parameters or **expensive likelihood evaluation**

  → Marginalized over nuisance parameters

- **Neural likelihood estimation:** *train neural density estimator to approximate likelihood*

  $$q_\Phi(s|\lambda,\nu) \approx p(s|\lambda,\nu)$$

[Papamakarios, Sterratt and Murray, AISTAT (2019)]



Neural likelihood estimation

- **Normalizing flows can model complicated probability distributions** in high dimensions [Dinh, Sohl-Dickstein and Bengio, arXiv:1605.08803 (2016)]

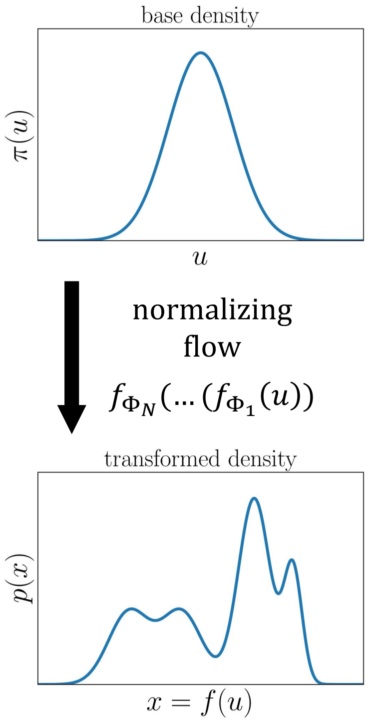
  → Represent probability distribution via invertible and differentiable transformations $f_{\Phi_1} \circ \cdots \circ f_{\Phi_N}$

- **Minimizing statistical distance** is equivalent to maximizing log probability

$$\min_{\Phi} D_{\mathrm{KL}}\left(p(s|\lambda,\nu) || q_{\Phi}(s|\lambda,\nu)\right) = \max_{\Phi} \int \mathrm{d}s\ p(s|\lambda,\nu)\ [\log p - \log q_{\Phi}]$$
$$= \max_{\Phi} \sum_{s_i \sim p(s|\lambda_i,\nu_i)} \log q_{\Phi}(s_i|\lambda_i,\nu_i)$$

- Train 100 flows, use **ensemble average** over $N = 5$ best-performing

$$\log p(s|\lambda,\nu) \approx \frac{1}{N} \sum_j \log q_{\Phi_j}(s|\lambda,\nu)$$



Neural likelihood estimation

- Multiply with priors and **sample from posterior** $p(\lambda, \nu | s) \propto p(s|\lambda, \nu) p(\lambda) p(\nu)$

  → Uniform prior on EoS parameters $p(\lambda)$ $\lambda_1 \in [4.75, 5.25]$ $\lambda_2 \in [-2.05, -1.85]$

  → **Different scenarios** for prior information on nuisance parameters $p(\nu)$

parameter	true	tight	loose
d	exact	5%	20%
N_H	exact	30%	50%
$\log(T_{\text{eff}})$	exact	± 0.1	± 0.2

- Scaling to multiple observations

$$p(\lambda, \nu_{1 \dots J} | s_{1 \dots J}) \propto \left(\prod_j p(s_j | \lambda, \nu_j) p(\nu_j) \right) p(\lambda)$$

- High dimensional parameter space

  → Likelihood differentiable $\nabla_{\lambda,\nu} p(s|\lambda, \nu)$, can use **Hamiltonian Monte Carlo** (HMC) sampling

Example posterior distribution

- **Full posterior $p(\lambda, \nu | s)$ directly from telescope spectra**

  → Here: based on 10 simulated spectra for example EoS

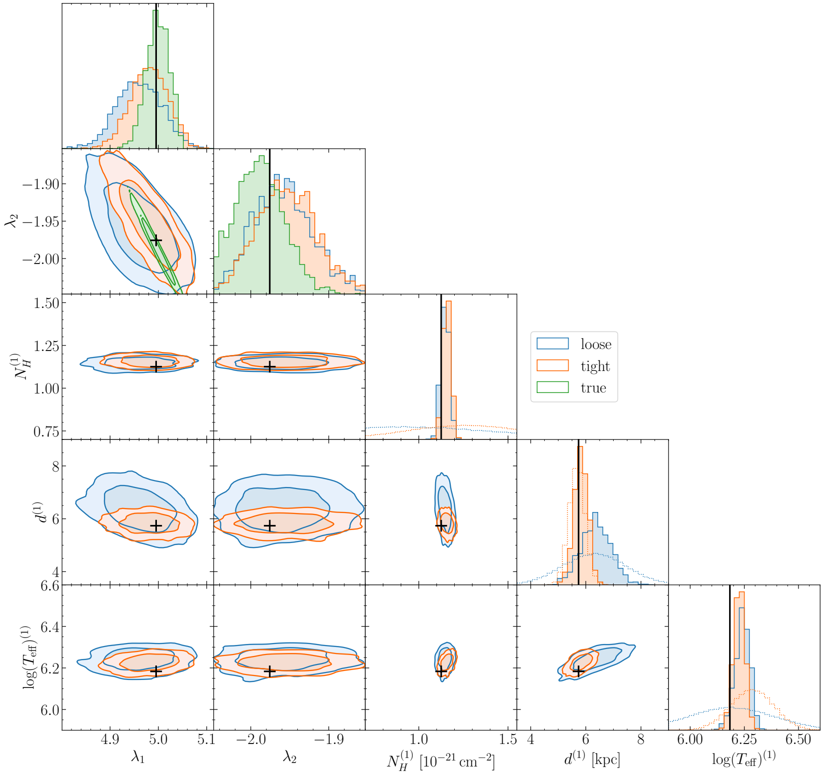
- Three scenarios: true, tight and loose

  → Uncertainties much smaller in true case

- Constrain N_H and $\log(T_{\text{eff}})$ much **better than prior information**

  → In tight case d posterior similar to prior

- Transform into $P(\varepsilon)$ and $R(M)$ constraints using TOV equations



[LB *et al.*, arXiv:2403.00287 (2024)]

Example posterior distribution

- **Full posterior $p(\lambda, \nu | s)$ directly from telescope spectra**

  → Here: based on 10 simulated spectra for example EoS

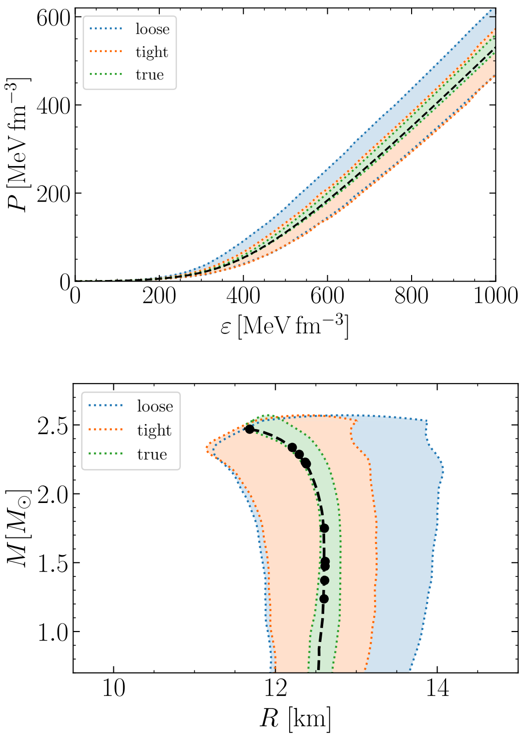
- Three scenarios: true, tight and loose

  → Uncertainties much smaller in true case

- Constrain N_H and $\log(T_{\rm eff})$ much **better than prior information**

  → In tight case *d* posterior similar to prior

- Transform into $P(\varepsilon)$ and $R(M)$ constraints using TOV equations



[LB *et al.*, arXiv:2403.00287 (2024)]

Example posterior distribution

- **Full posterior $p(\lambda, \nu | s)$ directly from telescope spectra**

  → Here: based on 10 simulated spectra for example EoS

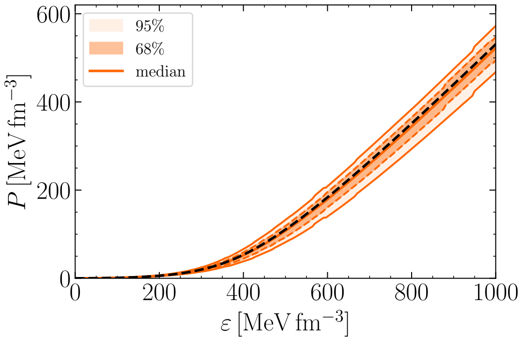
- Three scenarios: true, tight and loose

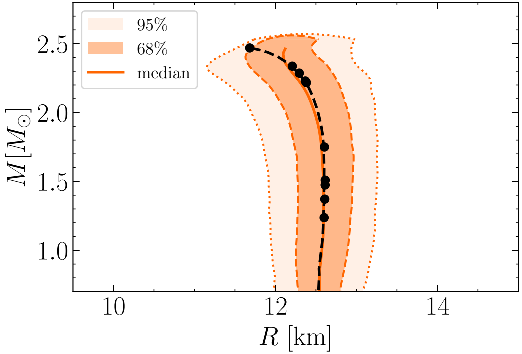
  → Uncertainties much smaller in true case

- Constrain N_H and $\log(T_{\text{eff}})$ much **better than prior information**

  → In tight case *d* posterior similar to prior

- Transform into $P(\varepsilon)$ and $R(M)$ constraints using TOV equations





[LB *et al.*, arXiv:2403.00287 (2024)]

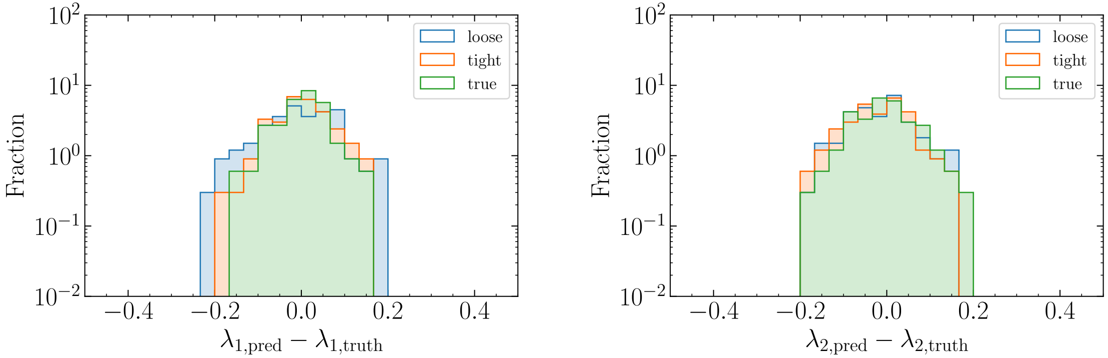
Average performance on test set

- Compute **maximum-a-posteriori estimate ($\lambda_{1,\text{pred}}, \lambda_{2,\text{pred}}$)** for 100 random EoS

  → Again 10 simulated spectra for each EoS

- Compare to **ground-truth values ($\lambda_{1,\text{true}}, \lambda_{2,\text{true}}$)**

  → Distribution centered around 0, **no systematic bias**



[LB *et al.*, arXiv:2403.00287 (2024)]

Average performance on test set

- Compare total standard deviation to previous approaches

$$\sigma_{\text{tot}} = \sqrt{\sigma_{\lambda_1}^2 + \sigma_{\lambda_2}^2}$$

→ **Better accuracy than previous approaches**

$p(\nu)$	method	σ_{tot}
true	NN(Spectra)	0.099
	ML-Likelihood	0.096
	NLE	**0.090**
tight	NN(Spectra)	0.115
	ML-Likelihood	0.112
	NLE	**0.097**
loose	NN(Spectra)	0.152
	ML-Likelihood	0.120
	NLE	**0.113**

[LB *et al.*, arXiv:2403.00287 (2024)]

- For the first time posterior for nuisance parameters

Increasing number of observations

- Amortization: need to train normalizing flows only once

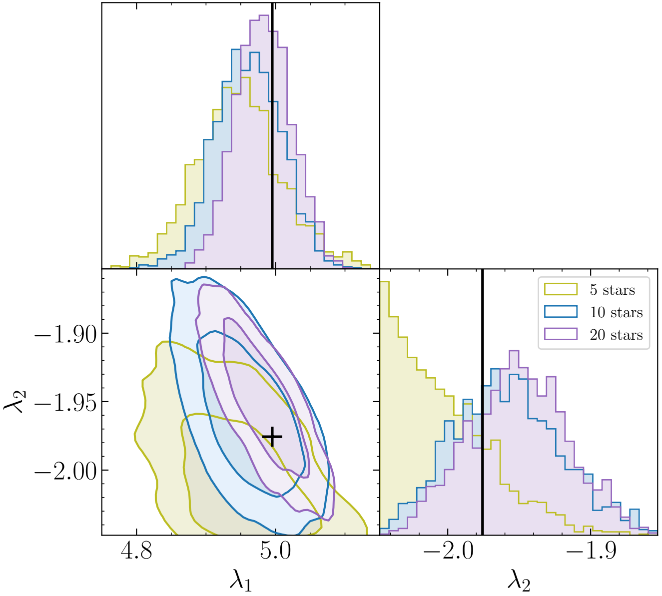
  → **Inexpensive likelihood evaluation**, inclusion of additional observations straightforward

- **Many more measurements in the future**
  [Iacovelli *et al.*, Phys. Rev. D 108 (2023)]

  → Not feasible to use two-step method with integrations

- *Extend neural likelihood estimation to other neutron star data*
  [Dax *et al.*, Phys. Rev. Lett. 127 (2021)]

- **Apply to real data and include mass in likelihood** $\boldsymbol{q_\Phi(\lambda, \nu, M|s)}$
  [Steiner *et al.*, MNRAS 476 (2018)]



[LB *et al.*, arXiv:2403.00287 (2024)]

Summary

- **Bayesian inference of neutron star EoS based on thermal X-ray spectra**

  → Likelihood analytically intractable → Inference in two steps

- **Neural likelihood estimation: train normalizing flows to approximate likelihood**

  $$q_\phi(s|\lambda, \nu) \approx p(s|\lambda, \nu)$$

  → Full posterior directly from telescope spectra

  → Better accuracy than previous methods

- **Inexpensive inclusion of additional data**

  → More scalable to growing number of observations expected in coming years

Supplementary material

Comparison with two-step method

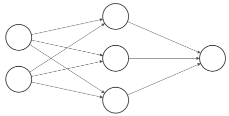
- Use XSPEC to **predict mass and radius of neutron stars**

  → Repeat for multiple ν_i from $p(\nu)$ to estimate ΔM and ΔR

- **Train neural network to predict λ:**

[Farrell *et al.*, JCAP 02 (2023)]

$$\begin{bmatrix} M \\ \Delta M \\ R \\ \Delta R \end{bmatrix}$$



$$\lambda$$

- Compare accuracy:

$p(\nu)$	method	σ_{tot}
true	NN(M, R via XSPEC)	**0.085**
	NLE	0.090
tight	NN(M, R via XSPEC)	0.098
	NLE	**0.097**
loose	NN(M, R via XSPEC)	0.136
	NLE	**0.113**

[LB *et al.*, arXiv:2403.00287 (2024)]

Spectral EoS parametrization

- Begin with **relativistic mean-field model** GM1L [Typel *et al.*, Phys. Rev. C 81 (2010)]

→ Describe by **spectral expansion** [Lindblom, Phys. Rev. D 82 (2010)] [Lindblom, Phys. Rev. D 97 (2018)]

$$\Upsilon(P) = \frac{1 - c_s^2}{c_s^2} \qquad \text{with} \quad c_s^2 = \frac{\partial P}{\partial \varepsilon}$$

- Expand in effective enthalpy $h = \int_0^P dP' \ (\varepsilon(P') + P')^{-1}$

$$\Upsilon(h) = \exp\left[\sum_k \lambda_k \Phi_k(h)\right] \qquad \text{with} \quad \Phi_k = \left[\log\left(\frac{h}{h_0}\right)\right]^k$$

→ Second-order expansion already sufficient to describe EoS up to 10% [Lindblom, Phys. Rev. D 82 (2010)]

- **Vary parameters (λ_1, λ_2)** for 10^3 EoS with ~ 100 stars (M, R) each

Hamiltonian Monte Carlo

- Generate **chain of samples of target distribution $\pi(q)$**

  → With gradients can reduce correlations and improve over random walks

- Potential and kinetic term inspired by classical mechanics

$$H(p,q) = -\log \pi(q) + \frac{1}{2} p^T M^{-1} p$$

[Betancourt, arXiv:1701.02434 (2017)]

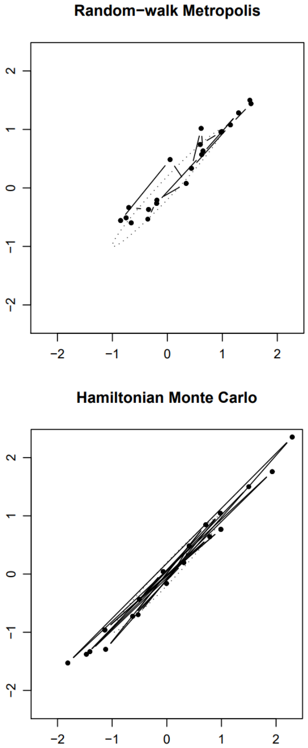
- Sample auxiliary 'momentum' based on covariance $p \sim \mathcal{N}(0, M^{-1})$

  → Solve **Hamiltonian equations** with leap frog integrator for (q', p')

$$\frac{dq}{dt} = \frac{\partial H}{\partial p} \qquad \frac{dp}{dt} = -\frac{\partial H}{\partial q}$$

- **Metropolis-Hasting**: accept new state q' with probability

$$\min\left[1, \exp\left(-H(q',p') + H(q,p)\right)\right]$$



[Neal, in *Handbook of MCMC*, Chap. 5 (2011)]

Normalizing flows

- Represent **probability distribution $p(x)$** via *invertible and differentiable transformations* f_Φ of base distribution $\pi(u)$
  [Papamakarios *et al.*, J. Mach. Learn. Res. 22 (2021)]

- Choose simple **base distribution $\pi(u)$**, i.e. standard Gaussian

- Generate samples

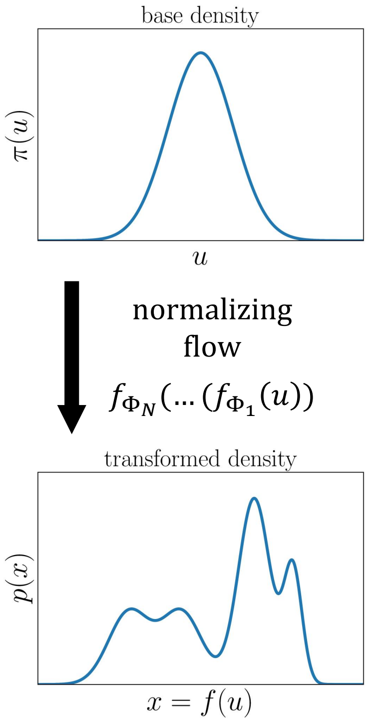
$$u \sim \pi(u) \qquad x = f_\Phi(u)$$

- Compute probability density

$$p(x) = \pi(f_\Phi^{-1}(x)) \left| \det \left(\frac{\partial f_\Phi^{-1}}{\partial x} \right) \right|$$

→ Transformation should be easily invertible and have fast computable Jacobian determinant

- Stack multiple transformations $\qquad f_\Phi = f_{\Phi_1} \circ \cdots \circ f_{\Phi_N}$



Masked Autoregressive Flows

- Each dimension x of $(x_1, \dots, x_D)$ **sequentially transformed** based on previous $x_{1:d-1} = (x_1, \dots, x_{d-1})$

$$p(x) = \prod_{d=1}^{D} p(x_d | x_{1:d-1})$$

[Papamakarios, Pavlakou and Murray, NeurIPS 30 (2017)]

- Choose **Gaussian marginals** $p(x_d | x_{1:d-1}) = \mathcal{N}(x_d | \mu_d, \sigma_d^2)$

→ Mean and standard deviation computed by NN based on previous values

→ Sampling is equivalent to **scaling and shifting standard Gaussian**

$$x_d = \mu_d + \sigma_d u_d \qquad u_d \sim \mathcal{N}(0, 1)$$

[Kingma *et al.*, NeurIPS 29 (2016)]

- For inverse do not need to invert NN, autoregressive structure leads to triangular Jacobian

- Choose binary masks to compute all u_d in single pass

→ Very efficiently obtain $p(x)$, less efficient sampling x

Low-mass binaries in quiescence

- Low-mass donor star transferring material onto neutron star

  → Minimal X-ray radiation during **low-accretion periods** [Heinke *et al.*, Astrophys. J. 598 (2003)]

- Luminosity given by distance and bolometric flux

  $$L = 4\pi d^2 F$$

  → Flux partly absorbed by interstellar medium depending on N_H

- Assume uniform temperature → **Stefan-Boltzmann law**

  $$L = \sigma_{\mathrm{SB}} 4\pi R^2 T_{\mathrm{eff}}^4$$

  → Needs to account for gravitational redshift and light-bending (depending on M/R)

- Radius from eight qLMXBs:

  $$R(1.4\,M_\odot) = 12.0^{+1.1}_{-0.6}\,\mathrm{km}$$

  [Steiner *et al.*, Mon. Not. Roy. Astron. Soc. 476 (2018)]

Multimodal masses

- Parameters close to border of prior $(\lambda_1, \lambda_2) = (4.77, -1.91)$ in true scenario

  → Some chains **converge at wrong** $M^{(2)}$ and wrong λ
- **Radius has larger impact** on spectrum than mass
- Likelihood of false mode 10^{-4} times smaller