

Inferring three-nucleon couplings from multi-messenger neutron-star observations

Received: 28 May 2025

Accepted: 24 September 2025

Published online: 06 November 2025

 Check for updates

Rahul Somasundaram  ^{1,2,9} , Isak Svensson  ^{3,4,5,9} , Soumi De  ², Andrew E. Deneris ⁶, Yannick Dietz  ^{3,4}, Philippe Landry ^{7,8}, Achim Schwenk  ^{3,4,5} & Ingo Tews  ²

Understanding the interactions between nucleons in dense matter is an important challenge in theoretical physics. Effective field theories have emerged as the dominant approach to address this problem at low energies, with many successful applications to the structure of nuclei and the properties of dense nucleonic matter. However, how far into the interior of neutron stars these interactions can describe dense matter is an open question. Here, we develop a framework that enables the inference of three-nucleon couplings in dense matter directly from astrophysical neutron star observations. We apply this formalism to the LIGO/Virgo gravitational-wave event GW170817 and the X-ray measurements from NASA's Neutron Star Interior Composition Explorer and establish direct constraints for the couplings that govern three-nucleon interactions in chiral effective field theory. Furthermore, we demonstrate how next-generation observations of a population of neutron star mergers can offer stringent constraints on three-nucleon couplings, potentially at a level comparable to those from laboratory data. Our work directly connects the microscopic couplings in quantum field theories to macroscopic observations of neutron stars, providing a way to test the consistency between low-energy couplings inferred from terrestrial and astrophysical data.

Obtaining a description of strong-interaction matter based on the underlying theory of quantum chromodynamics (QCD) is one of the most important challenges of theoretical physics. The development of effective field theories (EFTs) of QCD has revolutionized our understanding of nuclear forces^{1–3}. By employing a separation of scales between typical nucleon momenta and high-energy degrees of freedom, indicated by the breakdown scale of the theory, EFTs provide a systematic expansion for inter-nucleon interactions which can be improved order by order. Consequently, a natural hierarchy emerges among the interactions between two (NN), three (3N), and many

nucleons, the strengths of which are parameterized by low-energy couplings (LECs) that need to be determined from data or from the underlying theory. In the absence of direct calculations of the LECs from QCD, the LECs are commonly adjusted to reproduce nuclear experiments. For example, LECs that govern NN interactions are fit to NN scattering data, whereas 3N forces are fit to properties of light nuclei. The LECs governing the two-pion exchange contribution to the 3N forces⁴, c_1 and c_3 , describe pion-nucleon (πN) interactions and can be constrained using laboratory πN scattering data to $c_1 = -0.74 \pm 0.02 \text{ GeV}^{-1}$ and $c_3 = -3.61 \pm 0.05 \text{ GeV}^{-1}$ at one-sigma level;

¹Department of Physics, Syracuse University, Syracuse, NY, USA. ²Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA. ³Technische Universität Darmstadt, Department of Physics, Darmstadt, Germany. ⁴ExtreMe Matter Institute EMMI, GSI Helmholtzzentrum für Schwerionenforschung GmbH, Darmstadt, Germany. ⁵Max-Planck-Institut für Kernphysik, Heidelberg, Germany. ⁶Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, USA. ⁷Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, Ontario, Canada. ⁸Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada. ⁹These authors contributed equally: Rahul Somasundaram, Isak Svensson.

 e-mail: rsomasundaram@lanl.gov; isak.svensson@tu-darmstadt.de

see Hoferichter et al.⁵ and Siemens et al.⁶ for an analysis using Roy-Steiner equations. EFT-based interactions with LECs calibrated this way have enjoyed significant success in applications to nuclear structure and nuclear astrophysics^{7–18}. However, an open question is which data with $A \geq 3$ one should use to infer LECs describing three-nucleon interactions. For example, recently the calibration of certain LECs to medium-mass nuclei, such as oxygen isotopes, has resulted in increased success for nuclear structure phenomenology, enabling EFT-based Hamiltonians to successfully describe nuclei up to ^{208}Pb (see, e.g., Hu et al.¹² and Arthuis et al.¹⁹). These developments emphasize that it is key to identify critical experimental constraints for the calibration of powerful nuclear interactions.

Interestingly, the microscopic interactions that govern atomic nuclei also dictate the macroscopic properties of neutron stars. Neutron stars are fascinating systems that access higher densities and greater neutron excess than nuclei accessible in terrestrial experiments. Therefore, they offer an unparalleled laboratory for nuclear matter in regions of high density and isospin asymmetry (the ratio of neutrons to protons)^{20–27}. However, significant challenges arise in modeling the equation of state (EOS) of neutron stars, such as the involved computational cost of solving the quantum many-body problem. Initial steps have been taken to constrain 3N interactions from astrophysical observations^{28–30}, but none of these studies have attempted to calibrate the nuclear Hamiltonian directly to such data. While Maselli et al.²⁸ and Sabatucci et al.²⁹ explore the sensitivity of neutron star observables to a single parameter governing the strength of the short-range 3N contribution to the EOS, they do not solve the quantum many-body problem at each step in the sampling but instead make simplifying assumptions to connect the single parameter to neutron star observables. Furthermore, these studies do not fully model variations in the EOS from low to high densities, beyond the nuclear regime, where the density dependence can change from that dominated by 3N forces. For example, phase transitions to non-nuclear degrees of freedom are possible. Rose et al.³⁰ use astrophysical observations to distinguish between two Hamiltonians that model the 3N force in different ways but do not infer the LECs themselves.

Here, we develop a framework that allows us to constrain LECs directly from observations of neutron stars via Bayesian inference and, thus, explore EFT-based interactions for the densest neutron-rich systems in the cosmos. We consider a Hamiltonian at next-to-next-to-

leading order (N²LO) in the EFT expansion and focus on the leading (N²LO) 3N forces that provide a strong contribution to the EOS of neutron matter. As LEC values obtained from two entirely distinct sources (neutron star observations versus atomic nuclei) can be checked against each other for consistency, our framework will provide a unique test for the domain of applicability of nuclear interactions and of the convergence of the EFT expansion in dense matter.

Results

Machine-learning-based inference framework

Our Bayesian inference setup to constrain c_1 and c_3 from astrophysical data involves sampling over the microphysical LECs using a Markov chain Monte Carlo (MCMC) stochastic sampling algorithm. Every iteration of the sampler requires the solution of the many-body Schrödinger equation, which yields the neutron star EOS. The EOS is subsequently translated to astrophysical observables, such as radii or tidal deformabilities, by solving the Tolman-Oppenheimer-Volkoff (TOV) equations and the equations for a stationary quadrupolar tidal deformation^{31,32}. Given the large number of iterations required to sample the posterior distribution function— $\mathcal{O}(10^6)$ —this is a computationally intractable problem given that a single iteration requires $\mathcal{O}(10^2)$ CPU-h.

We overcome this challenge by employing recent advances in machine-learning-based algorithms that act as surrogate models to more complex high-fidelity calculations. First, for rapid calculation of the EOS, we employ the recently proposed parametric matrix model (PMM)^{33,34}, which is trained on third-order many-body perturbation theory (MBPT) calculations¹⁸ of the neutron-matter EOS. In the left panel of Fig. 1, we show results for 70 validation samples for a PMM trained on 30 high-fidelity MBPT calculations. We find emulator uncertainties to be well under control, with predictions of the EOS at nuclear saturation density $n_{\text{sat}} = 0.16 \text{ fm}^{-3}$ differing from the high-fidelity results by 0.04% on average. This emulator uncertainty is much smaller than the LEC variation, indicated by the spread of samples, as well as the uncertainty in the MBPT calculations themselves (see Methods). We then extend these neutron-matter results to neutron star matter in beta equilibrium. We model the neutron star EOS using three different parameterizations (see Methods), the first and simplest of which only uses the emulator results based on c_1 and c_3 to characterize the EOS up to $10n_{\text{sat}}$.

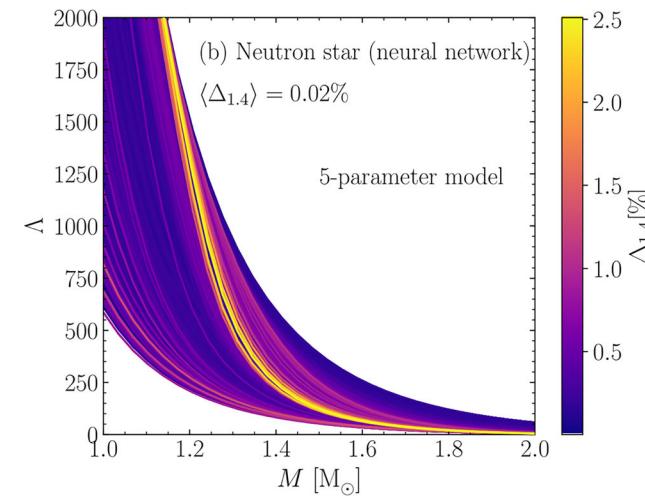
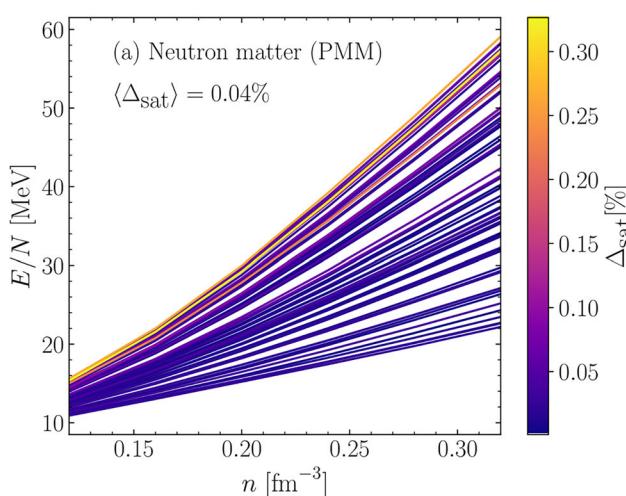


Fig. 1 | Validation results for our machine-learning-based emulators. In both panels, the color of the curves corresponds to the percentage uncertainty in the emulators' prediction. **a** Shows 70 results for the neutron-matter EOS up to $2n_{\text{sat}}$ predicted by a PMM built using a training set of 30 different samples. At nuclear saturation density, individual validation uncertainties Δ_{sat} are below 0.3%, with the average uncertainty being 0.04%. **b** Shows ~60,000 tidal deformability-mass

curves predicted by an ensemble of 50 neural networks trained on the 5-parameter EOS model. The uncertainty at $1.4M_{\odot}$, $\Delta_{1.4}$, is at most 2.5% with an average of 0.02%. We note that less than 0.1% of EOSs have emulator uncertainties of around 2%, making these very rare occurrences. Source data for this figure are provided as a Source Data file.

The nucleonic description of dense matter eventually breaks down, and exotic phenomena such as QCD phase transitions in high-density matter might appear. Therefore, in the following, we assume the validity of the EFT expansion up to $2n_{\text{sat}}$ ^{18,35,36}, and hence, the LECs c_1 and c_3 determine the EOS only up to this density. An important aspect of our framework is to find a suitable parameterization and marginalization over uncertainties in the high-density EOS. Here, we use two different models based on the speed of sound that allow for a physics-agnostic extension of the EOS above $2n_{\text{sat}}$ ³⁷. These models employ either 5 or 7 parameters, including the two LECs. Comparing these models with the 2-parameter model illustrates the importance of the marginalization over the high-density EOS.

For a given EOS model, we use an ensemble of neural networks to predict the tidal deformability of neutron stars³⁸. Results obtained from validating our emulator on ~60,000 samples are shown in the right panel of Fig. 1 for the 5-parameter model, but the emulator performance is similar for the other EOS models. We find that both the PMM and neural-network emulators provide highly accurate and rapid emulation of the underlying complex calculations that are required to compute neutron star properties starting from microscopic LECs. The use of machine-learning algorithms in this manner is a key part of our analysis that allows us to sample complex posterior distribution functions in our astrophysical Bayesian analysis framework.

Inference based on GW170817 and NICER data

We condition the LECs c_1 and c_3 on the LIGO/Virgo Collaboration's first GW observation of a binary neutron star merger, GW170817^{39,40}, as well as X-ray observations of three pulsars made by NASA's Neutron Star Interior Composition Explorer (NICER) mission: PSR J0030+0451^{41,42}, PSR J0740+6620^{43,44}, and PSR J0437-4715⁴⁵. For these observations, we employ likelihood functions of the source parameters (neutron star masses, binary tidal deformabilities, radii, etc.) that are proportional to the posteriors computed in Abbott et al.⁴⁰ for GW170817, and Riley et al.⁴¹, Salmi et al.⁴⁶, and Choudhury et al.⁴⁵ for the three NICER observations. With wide uniform priors between zero and twice the mean of their laboratory values on the LECs c_1 and c_3 , the posterior conditioned on GW170817 is sampled using MCMC, and the posterior samples are weighted according to the product of the NICER likelihoods. Each likelihood evaluation is carried out by evaluating our two emulators consecutively, thus converting microscopic LECs to macroscopic neutron star observables.

The marginalized two-dimensional posterior distribution on c_1 and c_3 is shown in Fig. 2. We find essentially no influence of the neutron star observations on the parameter c_1 since the effect of c_1 on the EOS of neutron matter is subdominant compared to that of c_3 ⁴⁷. On the other hand, given the GW and X-ray observations, we see a clear preference for less repulsive 3N forces. We attribute this to c_3 being negatively correlated with the pressure inside neutron stars¹³. Since GW170817, as well as the X-ray observation of PSR J0470 favor more compact neutron stars, this translates to smaller pressures in neutron star interiors. The upper bound of the posterior is set by the prior boundary at $c_3 = 0 \text{ GeV}^{-1}$; however, this boundary is physically motivated since positive values of c_3 correspond to an attractive 3N force, which would lead to the collapse of neutron matter. Although the posterior median of $c_3 = -2.52 \text{ GeV}^{-1}$ deviates from the laboratory value of $c_3 = -3.61 \pm 0.05 \text{ GeV}^{-1}$, the laboratory and astrophysical determinations are consistent at the 90% confidence level, albeit with the latter constraint currently having large uncertainties. Existing neutron star observations do not offer high-precision constraints on the LECs, given the significant statistical uncertainties present in these observations.

Constraints from next-generation GW observatories

In the next decade, two upcoming next-generation ground-based GW detectors are expected to begin operations, namely the Einstein Telescope (ET) in Europe^{48,49} and Cosmic Explorer (CE) in the United

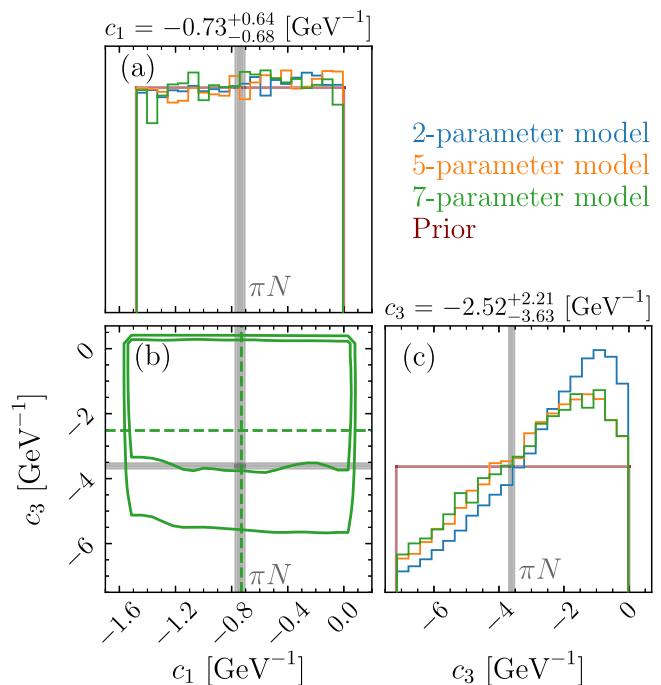


Fig. 2 | Constraints from GW170817 and NICER data. **a** Shows the posterior distribution function of the LEC c_1 . **c** Shows the posterior distribution function of the LEC c_3 . **a, c** Results are shown for the 2- (blue), 5- (orange), and 7-parameter (green) models. The quoted errors are for the 7-parameter model evaluated at the 90% confidence level. **b** Depicts the correlation between c_1 and c_3 and displays iso-probability contours at the 68% and 90% confidence levels. The laboratory values for c_1 and c_3 from pion-nucleon (πN) scattering experiments are shown in gray. Source data for this figure are provided as a Source Data file.

States^{50,51}. These detectors will provide a sensitivity that is improved by an order of magnitude over current GW detectors. Consequently, they are expected to observe $\mathcal{O}(100)$ events with signal-to-noise (SNR) ratios above 100 per year^{52–57}. Here, we demonstrate how such observations, at the level of populations of events, can provide stringent constraints on 3N couplings.

We simulate a population of neutron star merger events that can potentially be observed within a year-long observing run by a network of three next-generation detectors. Our detector network consists of two CEs, each with a 40 km arm length, and one ET with its fiducial 10 km triangular design. For our population model, we assume a uniform distribution in the range $1\text{--}2 M_{\odot}$ for the component masses and a random pairing into binary systems. Furthermore, we assume a constant local merger rate of $170 \text{ Gpc}^{-3} \text{ yr}^{-1}$, which is consistent with the merger rate inferred in Abbott et al.⁵⁸, and a uniform distribution of sources in co-moving volume. This results in a total of approximately 400 events within a redshift z of 0.2 observed by the detector network within one year. A lower merger rate would only affect our results by increasing the time required to observe the same number of events. There are many additional events with $z > 0.2$ that are detected with lower SNR; however, here we focus on the loudest signals in the detected population. The underlying EOS explored by our population, i.e., the EOS that relates the component masses of our simulated binaries to the corresponding component tidal deformabilities, is generated from our 5-parameter model with $c_3 = -3.68 \text{ GeV}^{-1}$, which is in agreement with the laboratory value at the 90% confidence level⁶.

From the detected events in our simulated population, we select the $N = 20$ highest-SNR events to analyze. This most informative subset is large enough that its chirp masses span the entire range allowed by the population model, with at least one event in every chirp mass bin of width $0.1 M_{\odot}$. This choice amounts to a selection cut at an SNR of about

225. Because the SNR depends primarily on the chirp mass, whose distribution is independent of the EOS in our assumed population model, this selection cut does not bias the recovery of EOS parameters. GW selection effects can thus be neglected in our inference.

The evolution of the uncertainties in c_3 as a function of the number N of observed events, represented at the 90% confidence level, is shown in Fig. 3. The N events are analyzed using a Bayesian hierarchical approach^{59,60} for $N=1, 3, 5, 10, 15$, and 20, within the Fisher matrix approximation^{61,62}. For $N=20$, the analysis is repeated using parameter estimation on zero-noise injections^{63,64}. The zero-noise realization of Gaussian noise is, in the statistical sense, the most likely realization⁵⁶ and, furthermore, the effect of any non-zero Gaussian noise generally weakens drastically with the SNR. The result matches well with the corresponding Fisher matrix result (see Fig. 3). This is consistent with other findings in the literature, see, for example, Vallisneri⁶⁵, which demonstrate that, for a four-dimensional Fisher matrix analysis, an SNR above 20 is typically sufficient for robust parameter estimation. We find that a single event among the 20 loudest ones observed by next-generation GW detectors in a year decreases uncertainties by only, on average, a factor of two compared with uncertainties obtained from present astrophysical data. However, as the number of detections increases, we find that the statistical uncertainties in c_3 decrease approximately as $1/\sqrt{N}$, and thus converge remarkably well to the injected value. We see how inference performed at the level of populations of events can potentially provide high-precision constraints on nuclear interactions within a year, competitive with and complementary to terrestrial laboratory data. These constraints would improve further if events with lower SNR are also considered. In contrast, the constraints could weaken in the same time frame if, for example, the observed merger rate is lower than expected.

Figure 3 also demonstrates the importance of the marginalization over uncertainties in the high-density EOS, implemented in our framework in the 5- and 7-parameter models. We find excellent

agreement between the results obtained from these two models. However, the very simple 2-parameter model—which does not account for such uncertainties—is in significant tension with the injected value, as it converges toward an incorrect c_3 . This underscores the importance of allowing for general high-density extensions in order to avoid systematic uncertainties in the inference of LECs.

Discussion

In this paper, we introduce a framework that allows for the inference of couplings describing microscopic 3N interactions from astrophysical observations of neutron stars. Naively, one would assume that the complex multi-physics calculations required to compute neutron star observables starting from microscopic Hamiltonians are too computationally expensive to allow for a full stochastic sampling of the posterior distribution. We have overcome this challenge by enhancing our Bayesian inference approach with machine learning. In particular, we employ two machine-learning methods—the parametric matrix model^{33,34} and the ensemble neural-network method^{38,66}—that circumvent this challenge by drastically speeding up our likelihood evaluations.

The inference of LECs from neutron star observations provides several benefits. While LECs can be adjusted to data on atomic nuclei or scattering, our novel approach makes it possible to constrain interactions using the densest and most neutron-rich system in the cosmos. Furthermore, it is possible that improved nuclear interactions involve not only higher orders in the EFT expansion, but the addition of new degrees of freedom, such as the Δ resonance⁶⁷ or hyperons⁶⁸. Even existing EFTs that include such degrees of freedom suffer from a lack of experimental data for their calibration and verification. For all these cases, astrophysical data provide a viable option for the calibration of the LECs.

While we have chosen to study only the 3N LECs c_1 and c_3 , our framework is applicable to any sector of the Hamiltonian, provided that it correlates sufficiently well with the EOS of neutron-rich matter. This enables consistency checks between constraints from neutron star observations and from nuclear experiments, which provides a nontrivial verification of the EFT expansion and its applicability to calculations of neutron star observables. Our framework also enables sensitivity analyses of neutron-rich matter that will allow for the identification of the most important LECs in chiral Hamiltonians for neutron star observables. In Fig. 3, the extracted value of c_3 is in good agreement with laboratory data by choice. Such a result with actual GW data would demonstrate consistency of the EFT approach between nuclei and dense matter. However, a hypothetical convergence of c_3 to a value incompatible with laboratory data—similar to the result for the 2-parameter model—might indicate the breakdown of the EFT and the appearance of new physics in neutron stars. Our framework also permits us to compare different implementations of EFTs as well as various many-body methods for strongly correlated quantum systems. With the advent of next-generation neutron star observations in the near future, we therefore expect our framework to be a key tool for inferring LECs and improvements to the EFTs themselves. Our work thus also demonstrates the importance of building a global network of next-generation ground-based gravitational-wave observatories.

Methods

High-fidelity EOS model

We require EOS models, connecting the nuclear Hamiltonian to neutron star structure, to generate training data for the machine learning-based algorithms employed in our Bayesian inference framework. In this work, we keep the NN part of the Hamiltonian fixed using the N²LO NN interactions of Entem, Machleidt, and Nosyk (EMN) with cutoff $\Lambda_c=450$ MeV⁶⁹, as this part is very well determined from NN scattering experiments. At N²LO, a given nuclear Hamiltonian for neutron matter is then parameterized by the values of the LECs c_1 and c_3 . To generate

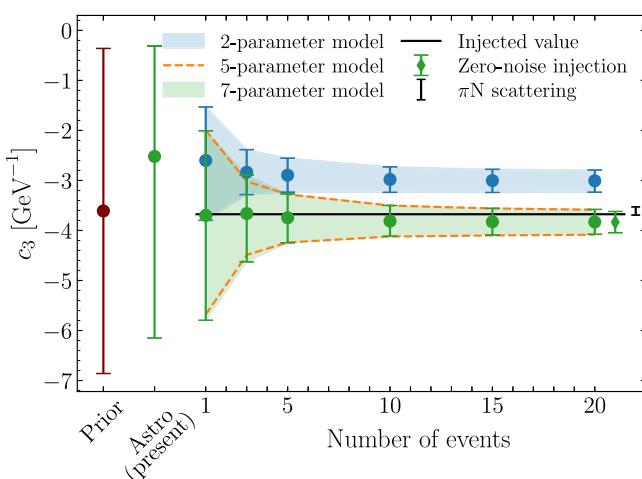


Fig. 3 | Evolution of the obtained constraints on c_3 . We show the prior on c_3 (purple) as well as the present astrophysical constraints, which include GW170817 and the NICER X-ray observations of millisecond pulsars but exclude simulated data from next-generation GW observatories. The inferred c_3 for our population study is then shown as a function of the number of events observed by a network of next-generation GW detectors for the 2-parameter (blue), 5-parameter (orange-dashed), and 7-parameter (green) EOS models. We observe that the constraints are well converged at 20 observations. For the 7-parameter model, the diamond error bar is obtained by analyzing 20 observations using parameter estimation on zero-noise injections, whereas the circle error bars are obtained within a Fisher matrix approximation. We compare the resulting constraints to the injected value and to the uncertainties in the laboratory result from πN scattering (black). All error bars represent 90% confidence levels. Source data for this figure are provided as a Source Data file.

the EOS from this Hamiltonian, the first step is to solve the many-body Schrödinger equation. Here, we use MBPT at third order^{16,18} to calculate the energy per particle of neutron matter between 0.12 fm^{-3} and 0.32 fm^{-3} in steps of 0.04 fm^{-3} for 100 randomly selected sets of LEC combinations. For symmetric nuclear matter, which is not as important but is needed to include the small fraction of protons, we use empirical properties. We estimate the uncertainty of the MBPT approach by comparing the difference in results at the second and third orders. The average (maximum) size of the third-order MBPT correction is 252 (790) keV across all densities and for all LEC combinations, which is much smaller than the overall range of the energy per particle; see Fig. 1. The correction generally increases with density away from saturation density. Based on this, we conclude that the MBPT calculations are sufficiently converged so as not to significantly impact our results. For this work, we use an MBPT code that is based on 3N operators in a single-particle basis, as in Drischler et al.¹⁶. This makes it computationally feasible to perform the calculations required for training our emulators.

Using an MBPT calculation of the EOS of neutron matter, we compute the EOS for neutron star matter in beta equilibrium up to $2n_{\text{sat}}$ ($10n_{\text{sat}}$) using the metamodel of Margueron et al.^{70,71} for our 5- and 7- (2-) parameter models. The metamodel provides a smooth density-functional-based interpolation between the discrete density-grid points of the MBPT calculation and allows for a straightforward extrapolation from the EOS of neutron matter to the EOS of matter in beta equilibrium, including leptons. The model parameters of the metamodel are the so-called nuclear empirical parameters (NEPs) that govern the density dependence of the EOSs of neutron matter and symmetric matter via a Taylor expansion about nuclear saturation density, see Margueron et al.⁷⁰ for more details. Here, we fix the iso-scalar NEPs to be $E_{\text{sat}} = -16 \text{ MeV}$, $n_{\text{sat}} = 0.16 \text{ fm}^{-3}$, $K_{\text{sat}} = 230 \text{ MeV}$, and $Q_{\text{sat}} = Z_{\text{sat}} = 0 \text{ MeV}$, i.e., we do not consider variations in the EOS of symmetric matter in this work. On the other hand, the iso-vector parameters are fit to the MBPT calculation of the energy per particle up to $2n_{\text{sat}}$ in neutron matter. For the 5- and 7-parameter models, the fit includes the iso-vector parameters E_{sym} , L_{sym} , K_{sym} , and Q_{sym} , i.e., all the iso-vector NEPs excluding Z_{sym} , which is fixed at 0 MeV, in order to reliably reproduce the MBPT calculation. We found that the difference between the MBPT calculation and the metamodel result, averaged over density and different LEC samples, is -10 keV. For the 2-parameter model, the NEP Q_{sym} is fixed to $Q_{\text{sym}} = 500 \text{ MeV}$, which results in a slight loss in the quality of the fit but allows for an extension of the neutron-star EOS up to $10n_{\text{sat}}$ using the metamodel alone while ensuring that the symmetry energy remains positive up to this density. Finally, for all three of our EOS models, the crust EOS of Douchin and Haensel⁷² is added below 0.08 fm^{-3} following the approach of Koehn et al.⁷³.

For the 5- and 7-parameter models, we account for phenomena beyond the EFT breakdown scale—such as the appearance of non-nucleonic degrees of freedom at high densities—by extending the EOS above $2n_{\text{sat}}$ using the speed-of-sound parameterization^{15,37}. This parameterization is agnostic towards the composition of matter; the model parameters are the values of the squared speed of sound, defined on a discrete density grid. For the 5- (7-) parameter model, we use three (five) grid points located at $[3, 5, 7] n_{\text{sat}}$ ($[3, 4, 5, 6, 7] n_{\text{sat}}$). Then, for a given set of sound speeds on the specified grid, the full sound-speed curve is obtained by interpolating linearly between the grid points. In both models, the speed of sound at $n > 7n_{\text{sat}}$ is taken to be a constant equal to its value at $7n_{\text{sat}}$. The full density-dependent speed of sound can then be integrated to obtain the pressure, energy density, and baryon chemical potential^{15,37}.

Using the neutron star EOS generated in this manner, we obtain the neutron star radius R and tidal deformability Λ as a function of the neutron star mass by solving the TOV equations and the equation for the quadrupolar tidal perturbation^{31,32}.

Machine learning the EOS model

The sequence of steps required to convert a set of EOS parameters to neutron star observables, such as the tidal deformability, is computationally expensive and requires $\mathcal{O}(10^2)$ CPU-h for a single evaluation. This is prohibitively expensive given that the total number of iterations required in this work is $\mathcal{O}(10^9)$. We overcome this bottleneck by using machine-learning approaches to emulate the two steps necessary to obtain the high-fidelity EOS model and connect it to neutron stars.

The first and most expensive step of generating an EOS model is the computation of the neutron matter EOS using MBPT. To emulate this computation, we use the PMM^{33,34} that combines ideas of both machine learning and reduced-order modeling. In our implementation of the PMM, the energy per particle in neutron matter at a given density is represented by the lowest eigenvalue of a 2×2 matrix \mathbf{M} given as

$$\mathbf{M} = \mathbf{M}_0 + c_1 \mathbf{M}_1 + c_3 \mathbf{M}_3, \quad (1)$$

where \mathbf{M}_0 is a diagonal matrix, \mathbf{M}_1 and \mathbf{M}_3 are symmetric matrices, and c_1 and c_3 are the two LECs of interest. While the dependence of \mathbf{M} is inspired by the theory of reduced-order modeling and the affine structure of chiral Hamiltonians, the matrix elements of \mathbf{M}_0 , \mathbf{M}_1 , and \mathbf{M}_3 need to be learned from data. To achieve this, we sample 100 different LEC combinations from their uniform prior distribution and calculate the EOS of neutron matter using MBPT. Then, we divide the set of 100 samples into 30 training and 70 validation samples, and fit the real matrix elements of \mathbf{M}_0 , \mathbf{M}_1 , and \mathbf{M}_3 at each density such that the lowest eigenvalues of \mathbf{M} agree with the MBPT results for the training set. Results from validating the PMM on the remaining 70 samples are shown in Fig. 1. To verify the accuracy of our PMM in the posterior parameter space, we averaged the emulator uncertainty using only those samples that lie within the span of the posterior on c_3 and found an average of 0.03%. The average (maximum) deviation between the PMM and the MBPT calculation, averaged over density and the 70 validation samples, is 15 keV (158 keV).

The second step is to solve the TOV equations and the equation for the quadrupolar tidal perturbation in order to calculate neutron star observables. This process takes only $\mathcal{O}(1)$ seconds for each EOS sample. This computational cost, while far cheaper than the MBPT calculations, remains a significant hurdle in Bayesian inference frameworks. To overcome the hurdle, we follow the approach of Reed et al.³⁸ and use an ensemble of feed-forward neural networks⁶⁶ to emulate solutions to the stellar-structure equations. To generate training and validation data, we first sample a set of 200,000 model parameters from their uniform priors and generate the neutron star EOS for each EOS model. During this process, we employ the PMM to calculate the corresponding EOSs of neutron matter for the various draws for c_1 and c_3 and use the meta- and sound-speed models for the neutron star EOSs. For each of the 200,000 EOS models, we solve the TOV equations using a high-fidelity solver. The resulting data is then used to train the neural-network emulator, since each sample corresponds to a well-defined input (the EOS model parameters) and output (the neutron star tidal deformabilities). In this work, the neural-network emulator is only used to predict the Λ - M curve and not the M - R curve. The latter does not require an emulator, as we perform significantly fewer calculations of this quantity.

We evenly split our set of 200,000 samples into training and validation data and discard all samples that do not reach $2 M_{\odot}$, i.e., the neural-network emulator is trained to predict the Λ - M curve only for those EOS parameters for which the maximum mass $M_{\text{TOV}} \geq 2M_{\odot}$. For all EOS models, the network architecture is such that it accepts the EOS parameters as input via the input layer, and then outputs the Λ - M curve on a grid of 30 points uniformly spaced between $1-2 M_{\odot}$. We use five hidden layers to emulate the 5- and 7-parameter models, whereas we use two hidden layers for the 2-parameter model. All hidden layers

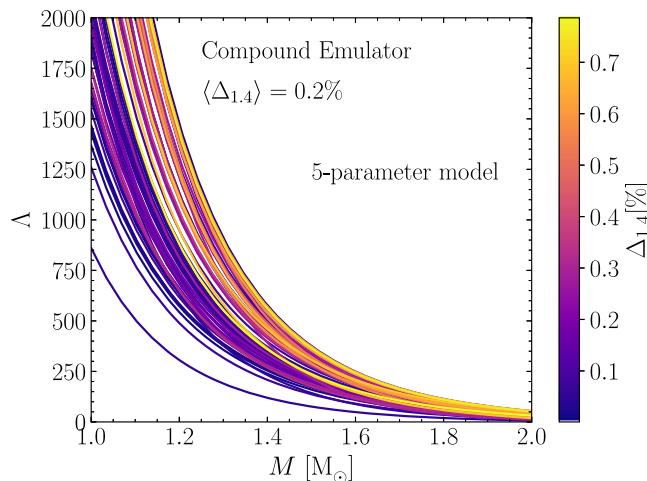


Fig. 4 | Compound emulator uncertainty. The figure shows the accuracy of our emulators, where the validation is performed using both high-fidelity solvers: MBPT and the full TOV solver. The quoted average error for $1.4M_\odot$, $\langle \Delta_{1.4} \rangle = 0.2\%$, is an estimate of the total compound emulator uncertainty, including both the PMM and the neural network, instead of only the latter. Source data for this figure are provided as a Source Data file.

consist of 64 nodes. Similar to Reed et al.³⁸, we use an ensemble of 50 (5) neural networks for the 5- and 7-parameter (2-parameter) models, with all neural networks trained independently and the final output obtained by averaging over all neural networks in the ensemble.

Validation results for our neural-network-based emulation are shown in the right panel of Fig. 1. However, since the validation data is computed using the PMM emulator and not high-fidelity MBPT calculations, the results of Fig. 1 capture only the accuracy of the neural-network emulator alone and not the full compound emulation uncertainty that also includes the PMM. Therefore, in Fig. 4, we have computed the compound uncertainty using the 70 MBPT samples that were not used when training the PMM. Each of these samples was extended to higher densities using the 5-parameter model, and the Λ – M sequence was subsequently obtained using our high-fidelity TOV solver. We have checked that our results do not change if the 2- or 7-parameter models are used instead. The error in the predicted tidal deformability at $1.4M_\odot$, $\Delta_{1.4}$, while somewhat larger than in Fig. 1, is sufficiently small for our machine-learning-based analyses to remain unbiased. Quantitatively, an event with an SNR of ~4000—far higher than any event considered here—would be required for the average compound error of 0.2% to be detectable with 90% confidence.

Since the emulator is trained on only those EOS parameter values that satisfy $M_{\text{TOV}} \geq 2M_\odot$, and therefore capable of making predictions only for those samples, we require an additional machine-learning tool, trained and validated on the same sets as above, to identify whether a sampled set of EOS parameters satisfies this criterion. We train an ensemble of ten neural networks to perform this binary classification, where each network accepts the EOS parameters as input and outputs the probability that the required constraint is satisfied. The final classification is obtained by averaging over the outputs of all ten neural networks. We found that for the 5-parameter and 7-parameter models, our classifier successfully identifies samples that satisfy $M_{\text{TOV}} \geq 2M_\odot$ with an accuracy of 99.9%. In the very few cases where the classifier fails, the maximum mass is usually very close to the $2M_\odot$ threshold, and the emulator nevertheless provides a reasonable emulation. The 2-parameter model does not require such a classifier since the maximum mass constraint is satisfied for all values of c_1 and c_3 by construction (because the density dependence of the symmetry energy for the 2-parameter model is stiff enough to support $2M_\odot$ stars).

Bayesian analysis of GW170817 and X-ray pulse-profile data

To obtain the posterior $P(\boldsymbol{\theta}|\mathbf{d})$ presented in Fig. 2, we use Bayes' theorem:

$$P(\boldsymbol{\theta}|\mathbf{d}) \propto [\mathcal{L}(d_{\text{J0030}}|\boldsymbol{\theta})\mathcal{L}(d_{\text{J0740}}|\boldsymbol{\theta})\mathcal{L}(d_{\text{J0437}}|\boldsymbol{\theta})\mathcal{L}(d_{\text{GW170817}}|\boldsymbol{\theta})]\pi(\boldsymbol{\theta}). \quad (2)$$

Here, $\boldsymbol{\theta}$ is a set of EOS parameters, $\pi(\boldsymbol{\theta})$ is the corresponding prior, and the likelihood function is obtained as the product of the likelihoods for different sources of data \mathbf{d} : the NICER observations of PSRs J0030, J0740, and J0437, as well as the gravitational-wave observation GW170817. We do not consider the event GW190425 given its low SNR⁷⁴ and negligible impact on the EOS (see, e.g., Dietrich et al.²¹).

As priors, we use uncorrelated uniform distributions for all EOS parameters. For c_1 and c_3 , the uniform priors span the range between 0 and twice the laboratory value obtained using the Roy-Steiner analysis⁶. For the speed-of-sound parameters, the uniform priors span the range between 0 and the speed of light c . Before analyzing the gravitational-wave and NICER data, we additionally impose the constraint $M_{\text{TOV}} \geq 2M_\odot$ as a hard cut, i.e., $\pi(\boldsymbol{\theta}) \propto \Theta(M_{\text{TOV}}(\boldsymbol{\theta}) - 2M_\odot)$, where Θ denotes a step function. This prior constraint accounts for the radio observations of heavy pulsars with masses around $2M_\odot$ ^{75–77} and is implemented implicitly when training the neural-network emulator for the stellar-structure equations.

In order to evaluate Eq. (2), we first sample the distribution $\mathcal{L}(d_{\text{GW170817}}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ using MCMC⁷⁸. The relevant likelihood is given by

$$\mathcal{L}(d_{\text{GW170817}}|\boldsymbol{\theta}) = \int dm_1 dm_2 \mathcal{L}(d_{\text{GW170817}}|m_1, m_2, \Lambda_1(m_1, \boldsymbol{\theta}), \Lambda_2(m_2, \boldsymbol{\theta})) \times \pi(m_1, m_2). \quad (3)$$

The prior on the component masses is taken to be uniform in the 1 – $2M_\odot$ range. We do not perform a parameter estimation for GW170817 starting from the GW strain data, as sampling over the LEC parameters would dramatically increase the computational cost of such an analysis, and instead take the likelihood $\mathcal{L}(d_{\text{GW170817}}|m_1, m_2, \tilde{\Lambda})$ —where $\tilde{\Lambda}$ is the binary tidal deformability—to be proportional to the posterior computed in Abbott et al.⁴⁰. We use a Gaussian kernel density estimator (KDE) to construct the probability density function from the samples calculated in Abbott et al.⁴⁰. The use of a KDE might introduce uncertainties as the probability density function is not always reconstructed accurately by this method. However, we did not see any significant change in our results when the bandwidth of the KDE was varied to within 30%. This is consistent with the general findings in the literature that KDE-induced uncertainties are subdominant for events that are only moderately informative, such as GW170817^{59,79}. At each iteration of the MCMC sampler, the factor $\Theta(M_{\text{TOV}}(\boldsymbol{\theta}) - 2M_\odot)$ is evaluated by the neural-network classifier, and the dependence of the binary tidal deformability $\tilde{\Lambda}$ on $\boldsymbol{\theta}$ in the above equation is determined by the emulator of the TOV equations.

To account for the NICER observations, the samples obtained from the MCMC sampling above are given a weight proportional to $\prod_i \mathcal{L}(d_i|\boldsymbol{\theta})$, where i indexes the pulsar. The individual NICER likelihoods are given as

$$\mathcal{L}(d_i|\boldsymbol{\theta}) = \int dM \mathcal{L}(d_i|M, R(M, \boldsymbol{\theta}))\pi(M), \quad (4)$$

where the prior $\pi(M)$ is taken to be uniform between $1M_\odot$ and the maximal TOV mass M_{TOV} . The function $R(M, \boldsymbol{\theta})$ is determined by solving the TOV equations using a high-fidelity solver. We approximate the likelihood functions $\mathcal{L}(d_i|M, R)$ to be proportional to the M - R posteriors for PSRs J0030⁴¹, J0740⁴⁶, and J0437⁴⁵, respectively, and we use the hot-spot models corresponding to the main result in each

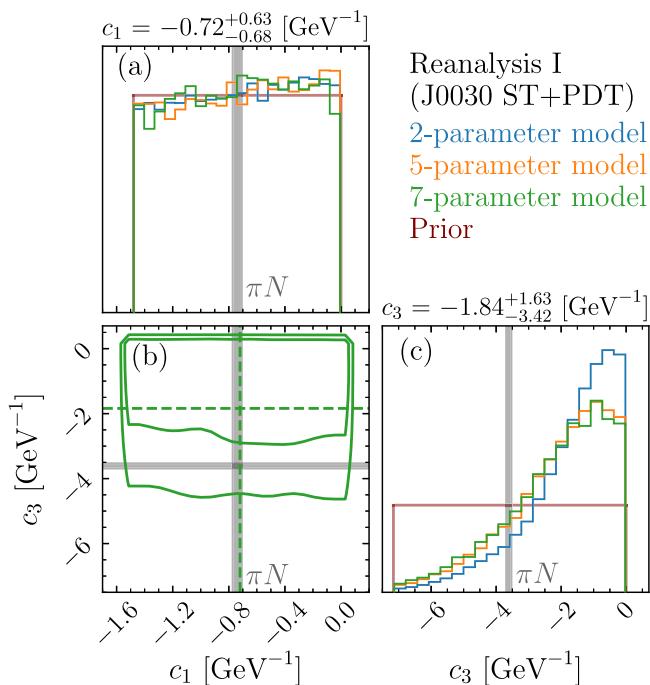


Fig. 5 | Updated constraints from NICER reanalysis I. **a** Shows the posterior distribution function of the LEC c_1 . **c** Shows the posterior distribution function of the LEC c_3 . **b** Depicts the correlation between c_1 and c_3 and displays iso-probability contours at the 68% and 90% confidence levels. These distributions are obtained by applying constraints from GW170817 and results from the recent reanalysis of NICER X-ray observations of PSR J0030⁸⁰. Here, we use the mode with mass and radius of $[-1.4M_\odot, 11.5 \text{ km}]$, i.e., using the hot-spot model ST+PDT. Source data for this figure are provided as a Source Data file.

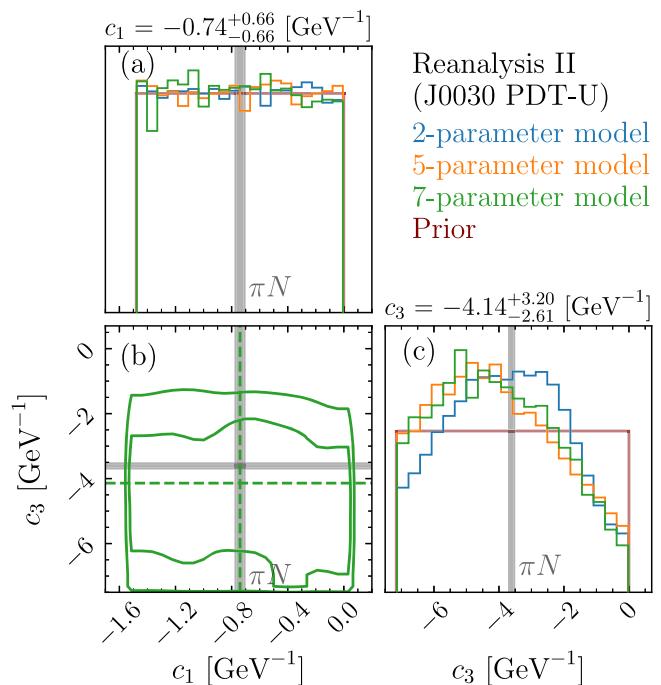


Fig. 6 | Updated constraints from NICER reanalysis II. **a** Shows the posterior distribution function of the LEC c_1 . **c** Shows the posterior distribution function of the LEC c_3 . **b** Depicts the correlation between c_1 and c_3 and displays iso-probability contours at the 68% and 90% confidence levels. These distributions are obtained by applying constraints from GW170817 and results from the recent reanalysis of NICER X-ray observations of PSR J0030⁸⁰. Here, we use the mode with mass and radius of $[-1.7M_\odot, 14.5 \text{ km}]$, i.e., using the hot-spot model PDT-U. Source data for this figure are provided as a Source Data file.

publication. Note that this approximation is not strictly valid for J0740 because a Gaussian prior on the pulsar's mass, informed by radio observations, was used to obtain the M - R posterior in Salmi et al.⁴⁶. Consequently, the existence of heavy pulsars is effectively accounted for twice in our analysis: first by applying a hard cut in $\pi(\boldsymbol{\theta})$ and then by the term $\mathcal{L}(d_{j0740}|M, R)$. However, since the Gaussian prior on the pulsar's mass used in Salmi et al.⁴⁶ lies above $2 M_\odot$ at the 68% CL and our hard-cut term is constant in this region, we conclude that the double counting of the pulsar's mass measurement has only a minor effect on our results. Our main results for the marginal posteriors on c_1 and c_3 obtained in this manner are shown in Fig. 2.

Recently, the NICER collaboration published a reanalysis of PSR J0030 using additional data⁸⁰. In Figs. 5 and 6, we present results on c_1 and c_3 using two results of this reanalysis: one with the hot-spot model ST+PDT with mass and radius of about $[1.4M_\odot, 11.5 \text{ km}]$ (reanalysis I) and the other with the hot-spot model PDT-U with a mean at about $[1.7M_\odot, 14.5 \text{ km}]$ (reanalysis II). While the results of reanalysis I are in agreement with Fig. 2, reanalysis II leads to a notable difference in the posterior on c_3 . However, when considering the 90% CL given in Fig. 3 for the present astrophysical data, this only changes the allowed c_3 range from $[-6.15, -0.31] \text{ GeV}^{-1}$ to $[-6.75, -0.94] \text{ GeV}^{-1}$.

Hierarchical analysis of next-generation events

For our study of the impact of next-generation gravitational-wave networks, we analyze the N loudest simulated events in a hierarchical approach using Bayes' theorem^{59,60}, i.e.,

$$P(\boldsymbol{\theta}|d) \propto \left[\prod_{i=1}^N \int dm_{1,i} dm_{2,i} \mathcal{L}(d_i|m_{1,i}, m_{2,i}, \tilde{\Lambda}_i(m_{1,i}, m_{2,i}, \boldsymbol{\theta})) \pi(m_{1,i}, m_{2,i}) \right] \pi(\boldsymbol{\theta}). \quad (5)$$

In such an approach, EOS parameters are simultaneously informed by all of the events, and the full multidimensional posterior is calculated to properly account for correlations between the various inferred parameters. While one in principle also has to infer the parameters of the source population simultaneously with $\boldsymbol{\theta}$, we here circumvent this issue by choosing the priors on the component masses $\pi(m_1, m_2)$ to be the same as the distribution used to generate the population of events, i.e., a uniform distribution in the range $1-2 M_\odot$ for both m_1 and m_2 . Because the population model is fixed and independent of the EOS, the GW selection bias towards detecting more massive systems has no impact on the recovery of the EOS parameters and can be neglected⁵⁹.

The single-event likelihoods $\mathcal{L}(d_i|m_{1,i}, m_{2,i}, \tilde{\Lambda}_i)$ are computed within the Fisher matrix approximation^{61,62,81}, which is valid for high-SNR events, using the TaylorF2 waveform model^{82,83} and the next-generation detector network described in the main text. This results in uncorrelated Gaussian likelihoods on the chirp masses, symmetric mass ratios, and the binary tidal deformabilities. Given that the quietest event in our sample has an SNR of about 225, we do not expect the results of our inference—performed at the level of a population of events—to be impacted significantly by the Fisher matrix approximation. In order to verify this, we re-computed all 20 single-event likelihoods (proportional to the single-event posteriors) with Bayesian parameter estimation⁶³. For each event, a zero-noise injection was created, and the binary parameters were recovered using the heterodyne likelihood model⁸⁴ implemented in PYCBC⁶⁴. In Fig. 3, we see that the results obtained using the Fisher matrix approximation (circle) are very similar to those obtained using parameter estimation on zero-noise injections (diamond) for the 7-parameter model at $N=20$. This indicates the validity of employing the Fisher matrix approximation for the 20 single-event likelihoods.

As in the analysis of GW170817 and the NICER data, the prior $\pi(\boldsymbol{\theta})$ is taken to be uniform on all the EOS parameters, along with the constraint that $\pi(\boldsymbol{\theta}) \propto \Theta(M_{\text{Tov}}(\boldsymbol{\theta}) - 2M_{\odot})$. The posterior distribution defined in Eq. (5) is sampled using MCMC for different values of N , with the relation between $\bar{\Lambda}$ and $\boldsymbol{\theta}$ in all 20 single-event likelihoods being evaluated by the neural-network emulator.

In Fig. 3, our results for the three EOS models are shown for $N = 1, 3, 5, 10, 15$, and 20. The marginal posteriors on c_3 are obtained by integrating Eq. (5) over all EOS parameters except c_3 . We average over a large number of orderings of our N events, i.e., for a given N , the MCMC sampling of Eq. (5) is repeated for 100 different permutations of N events drawn from our pool of 20 events. Upon performing this average, we find that the statistical uncertainties of the posteriors approximately obey a $1/\sqrt{N}$ behavior, similar to the findings of Landry et al.⁵⁹ and Kunert et al.⁸⁵.

We found that, even at $N=20$, the marginal posterior on c_1 is uninformative and is fully determined by the prior in the case of the 5- and 7-parameter models. This highlights that c_1 cannot be constrained from astrophysical data, even with next-generation detectors, given the lack of sensitivity of the EOS of neutron matter to c_1 . For the 2-parameter model, we found a small but noticeable shift of the posterior towards smaller values of c_1 . However, this is a consequence of the limitations of the 2-parameter model and further highlights the importance of the marginalization over the high-density EOS.

Finally, we note that our constraints on c_3 are sensitive to the assumed value of the EFT breakdown density. In Fig. 7, we have recomputed our constraints by decreasing the breakdown density from $2n_{\text{sat}}$ to $1.5n_{\text{sat}}$. The employed high-density model is similar to the 7-parameter model, but with an additional parameter for the squared sound speed at $2n_{\text{sat}}$. We find that, when only 20 events are used, the marginal posterior on c_3 significantly broadens, compared to the case with a breakdown density of $2n_{\text{sat}}$. However, the constraint tightens as more events are added. When analyzing almost all 400 events observed within one year, we find that the posterior shows a remarkable convergence towards the $2n_{\text{sat}}$ result. This indicates that a lower value of the EFT breakdown density still allows for significant constraints on c_3 , only with more events required, and that no major qualitative shift in EFT behavior occurs between the two breakdown densities. This approach can potentially allow us to constrain the breakdown density in future work.

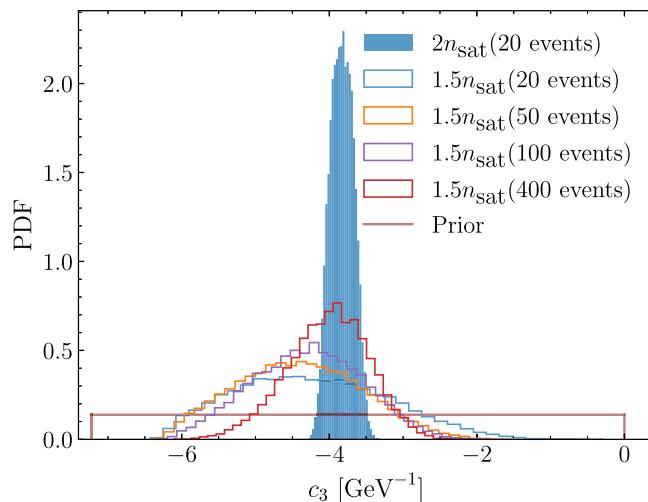


Fig. 7 | Results with lower EFT breakdown scale. A reanalysis of our projected constraints for c_3 in the era of third-generation GW detectors, but assuming the validity of the EFT expansion up to $1.5n_{\text{sat}}$ rather than $2n_{\text{sat}}$ with a different number of events. Source data for this figure are provided as a Source Data file.

Data availability

The posterior samples for GW170817 used in our analysis can be found at <https://dcc.ligo.org/LIGO-P1800061/public>⁸⁷. The mass-radius posteriors from the NICER analyses are available at <https://zenodo.org/records/3386449>⁸⁸ (PSR J0030, 2019), <https://zenodo.org/records/8239000>⁸⁹ (PSR J0030, 2024), <https://zenodo.org/records/6827537>⁹⁰ (PSR J0740), and <https://zenodo.org/records/12703175>⁴⁵ (PSR J0437). The datasets generated in this study, including all posterior samples and emulator training/validation data, have been deposited in this [Zenodo repository](https://zenodo.org/)⁹¹. Source data are provided with this paper.

Code availability

The package `GWBENCH`⁸¹ used to construct the Fisher matrix likelihoods can be found at <https://gitlab.com/sborhanian/gwbench>. The injection and parameter estimation-based recovery runs were performed using `PYCBC`⁶⁴, a publicly available software package that can be found at <https://github.com/gwastro/pycbc>. All codes developed for the analyses in this paper can be found in this [GitHub repository](https://github.com)⁹².

References

1. Weinberg, S. Nuclear forces from chiral Lagrangians. *Phys. Lett. B* **251**, 288–292 (1990).
2. Epelbaum, E., Hammer, H.-W. & Meißen, U.-G. Modern theory of nuclear forces. *Rev. Mod. Phys.* **81**, 1773–1825 (2009).
3. Hammer, H.-W., König, S. & van Kolck, U. Nuclear effective field theory: status and perspectives. *Rev. Mod. Phys.* **92**, 025004 (2020).
4. Epelbaum, E. et al. Three nucleon forces from chiral effective field theory. *Phys. Rev. C* **66**, 064001 (2002).
5. Hoferichter, M., Ruiz de Elvira, J., Kubis, B. & Meißen, U.-G. Roy-Steiner equation analysis of pion-nucleon scattering. *Phys. Rept.* **625**, 1–88 (2016).
6. Siemens, D. et al. Reconciling threshold and subthreshold expansions for pion-nucleon scattering. *Phys. Lett. B* **770**, 27–34 (2017).
7. Hergert, H. A guided tour of ab initio nuclear many-body theory. *Front. Phys.* **8**, 379 (2020).
8. Hebeler, K. Three-nucleon forces: Implementation and applications to atomic nuclei and dense matter. *Phys. Rept.* **890**, 1–116 (2021).
9. Hagen, G. et al. Neutron and weak-charge distributions of the ^{48}Ca nucleus. *Nat. Phys.* **12**, 186–190 (2015).
10. Gysbers, P. et al. Discrepancy between experimental and theoretical β -decay rates resolved from first principles. *Nat. Phys.* **15**, 428–431 (2019).
11. Stroberg, S. R., Holt, J. D., Schwenk, A. & Simonis, J. Ab initio limits of atomic nuclei. *Phys. Rev. Lett.* **126**, 022501 (2021).
12. Hu, B. et al. Ab initio predictions link the neutron skin of ^{208}Pb to nuclear forces. *Nat. Phys.* **18**, 1196–1200 (2022).
13. Hebeler, K., Lattimer, J. M., Pethick, C. J. & Schwenk, A. Equation of state and neutron star properties constrained by nuclear physics and observation. *Astrophys. J.* **773**, 11 (2013).
14. Lynn, J. E. et al. Chiral three-nucleon interactions in light nuclei, neutron- α scattering, and neutron matter. *Phys. Rev. Lett.* **116**, 062501 (2016).
15. Tews, I., Margueron, J. & Reddy, S. Critical examination of constraints on the equation of state of dense matter obtained from GW170817. *Phys. Rev. C* **98**, 045804 (2018).
16. Drischler, C., Hebeler, K. & Schwenk, A. Chiral interactions up to next-to-next-to-next-to-leading order and nuclear saturation. *Phys. Rev. Lett.* **122**, 042501 (2019).
17. Drischler, C., Holt, J. W. & Wellenhofer, C. Chiral effective field theory and the high-density nuclear equation of state. *Ann. Rev. Nucl. Part. Sci.* **71**, 403–432 (2021).
18. Keller, J., Hebeler, K. & Schwenk, A. Nuclear equation of state for arbitrary proton fraction and temperature based on chiral effective

field theory and a Gaussian process emulator. *Phys. Rev. Lett.* **130**, 072701 (2023).

19. Arthuis, P., Hebeler, K. & Schwenk, A. Neutron-rich nuclei and neutron skins from chiral low-resolution interactions. <https://arxiv.org/abs/2401.06675> (2024).
20. Capano, C. D. et al. Stringent constraints on neutron-star radii from multimessenger observations and nuclear theory. *Nat. Astron.* **4**, 625–632 (2020).
21. Dietrich, T. et al. Multimessenger constraints on the neutron-star equation of state and the Hubble constant. *Science* **370**, 1450–1453 (2020).
22. Al-Mamun, M. et al. Combining electromagnetic and gravitational-wave constraints on neutron-star masses and radii. *Phys. Rev. Lett.* **126**, 061101 (2021).
23. Essick, R., Tews, I., Landry, P. & Schwenk, A. Astrophysical constraints on the symmetry energy and the neutron skin of Pb208 with minimal modeling assumptions. *Phys. Rev. Lett.* **127**, 192701 (2021).
24. Huth, S. et al. Constraining neutron-star matter with microscopic and macroscopic collisions. *Nature* **606**, 276–280 (2022).
25. Annala, E. et al. Strongly interacting matter exhibits deconfined behavior in massive neutron stars. *Nat. Commun.* **14**, 8451 (2023).
26. Raaijmakers, G. et al. Constraints on the dense matter equation of state and neutron star properties from NICER's mass–radius estimate of PSR J0740+6620 and multimessenger observations. *Astrophys. J. Lett.* **918**, L29 (2021).
27. Rutherford, N. et al. Constraining the dense matter equation of state with new NICER mass–radius measurements and new chiral effective field theory inputs. *Astrophys. J. Lett.* **971**, L19 (2024).
28. Maselli, A., Sabatucci, A. & Benhar, O. Constraining three-nucleon forces with multimessenger data. *Phys. Rev. C* **103**, 065804 (2021).
29. Sabatucci, A., Benhar, O., Maselli, A. & Pacilio, C. Sensitivity of neutron star observations to three-nucleon forces. *Phys. Rev. D* **106**, 083010 (2022).
30. Rose, H. et al. Revealing the strength of three-nucleon interactions with the proposed Einstein Telescope. *Phys. Rev. C* **108**, 025811 (2023).
31. Flanagan, E. E. & Hinderer, T. Constraining neutron star tidal Love numbers with gravitational wave detectors. *Phys. Rev. D* **77**, 021502 (2008).
32. Hinderer, T., Lackey, B. D., Lang, R. N. & Read, J. S. Tidal deformability of neutron stars with realistic equations of state and their gravitational wave signatures in binary inspiral. *Phys. Rev. D* **81**, 123016 (2010).
33. Cook, P., Jammooa, D., Hjorth-Jensen, M., Lee, D. D. & Lee, D. Parametric matrix models. *Nat. Commun.* **16**, 5929 (2025).
34. Somasundaram, R. et al. Emulators for scarce and noisy data: application to auxiliary field diffusion Monte Carlo for the deuteron. *Phys. Lett. B* **866**, 139558 (2025).
35. Tews, I., Carlson, J., Gandolfi, S. & Reddy, S. Constraining the speed of sound inside neutron stars with chiral effective field theory interactions and observations. *Astrophys. J.* **860**, 149 (2018).
36. Drischler, C., Furnstahl, R. J., Melendez, J. A. & Phillips, D. R. How well do we know the neutron-matter equation of state at the densities inside neutron stars? A Bayesian approach with correlated uncertainties. *Phys. Rev. Lett.* **125**, 202702 (2020).
37. Somasundaram, R., Tews, I. & Margueron, J. Investigating signatures of phase transitions in neutron-star cores. *Phys. Rev. C* **107**, 025801 (2023).
38. Reed, B. T. et al. Toward accelerated nuclear-physics parameter estimation from binary neutron star mergers: emulators for the Tolman–Oppenheimer–Volkoff Equations. *Astrophys. J.* **974**, 285 (2024).
39. Abbott, B. P. et al. GW170817: observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.* **119**, 161101 (2017).
40. Abbott, B. P. et al. Properties of the binary neutron star merger GW170817. *Phys. Rev. X* **9**, 011001 (2019).
41. Riley, T. E. et al. A NICER view of PSR J0030+0451: millisecond pulsar parameter estimation. *Astrophys. J. Lett.* **887**, L21 (2019).
42. Miller, M. C. et al. PSR J0030+0451 mass and radius from NICER data and implications for the properties of neutron star matter. *Astrophys. J. Lett.* **887**, L24 (2019).
43. Riley, T. E. et al. A NICER view of the massive pulsar PSR J0740+6620 informed by radio timing and XMM-Newton spectroscopy. *Astrophys. J. Lett.* **918**, L27 (2021).
44. Miller, M. C. et al. The radius of PSR J0740+6620 from NICER and XMM-Newton data. *Astrophys. J. Lett.* **918**, L28 (2021).
45. Choudhury, D. et al. A NICER view of the nearest and brightest millisecond pulsar: PSR J0437–4715. *Astrophys. J. Lett.* **971**, L20 (2024).
46. Salmi, T. et al. The radius of PSR J0740+6620 from NICER with NICER background estimates. *Astrophys. J.* **941**, 150 (2022).
47. Hebeler, K. & Schwenk, A. Chiral three-nucleon forces and neutron matter. *Phys. Rev. C* **82**, 014314 (2010).
48. Punturo, M. et al. The Einstein Telescope: a third-generation gravitational wave observatory. *Class. Quant. Grav.* **27**, 194002 (2010).
49. Maggiore, M. et al. Science case for the Einstein Telescope. *JCAP* **03**, 050 (2020).
50. Reitze, D. et al. Cosmic Explorer: the U.S. contribution to gravitational-wave astronomy beyond LIGO. *Bull. Am. Astron. Soc.* **51**, 035 (2019).
51. Evans, M. et al. A horizon study for Cosmic Explorer: science, observatories, and community. <https://arxiv.org/abs/2109.09882> (2021).
52. Borhanian, S. & Sathyaprakash, B. S. Listening to the universe with next generation ground-based gravitational-wave detectors. *Phys. Rev. D* **110**, 083040 (2024).
53. Evans, M. et al. Cosmic explorer: a submission to the NSF MPSAC ngGW subcommittee. <https://arxiv.org/abs/2306.13745> (2023).
54. Branchesi, M. et al. Science with the Einstein Telescope: a comparison of different designs. *JCAP* **2023**, 068 (2023).
55. Gupta, I. et al. Characterizing gravitational wave detector networks: from A[#] to Cosmic Explorer. *Class. Quant. Grav.* **41**, 245001 (2024).
56. Smith, R. et al. Bayesian inference for gravitational waves from binary neutron star mergers in third generation observatories. *Phys. Rev. Lett.* **127**, 081102 (2021).
57. Finstad, D., White, L. V. & Brown, D. A. Prospects for a precise equation of state measurement from advanced LIGO and Cosmic Explorer. *Astrophys. J.* **955**, 45 (2023).
58. Abbott, R. et al. Population of merging compact binaries inferred using gravitational waves through GWTC-3. *Phys. Rev. X* **13**, 011048 (2023).
59. Landry, P., Essick, R. & Chatzioannou, K. Nonparametric constraints on neutron star matter with existing and upcoming gravitational wave and pulsar observations. *Phys. Rev. D* **101**, 123007 (2020).
60. Landry, P. & Chakravarti, K. Prospects for constraining twin stars with next-generation gravitational-wave detectors. <https://arxiv.org/abs/2212.09733> (2022).
61. Cutler, C. & Flanagan, E. E. Gravitational waves from merging compact binaries: how accurately can one extract the binary's parameters from the inspiral wave form? *Phys. Rev. D* **49**, 2658 (1994).
62. Poisson, E. & Will, C. M. Gravitational waves from inspiraling compact binaries: parameter estimation using second post-Newtonian wave forms. *Phys. Rev. D* **52**, 848 (1995).
63. Veitch, J. et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D* **91**, 042003 (2015).

64. Biwer, C. M. et al. PyCBC inference: a python-based parameter estimation toolkit for compact binary coalescence signals. *Publ. Astron. Soc. Pac.* **131**, 024503 (2019).

65. Vallisneri, M. Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. *Phys. Rev. D* **77**, 042001 (2008).

66. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

67. Jiang, W. G. et al. Accurate bulk properties of nuclei from $\tau_{\text{fm}}=2$ to ∞ from potentials with Δ isobars. *Phys. Rev. C* **102**, 054301 (2020).

68. Petschauer, S., Haidenbauer, J., Kaiser, N., Meißner, U.-G. & Weise, W. Hyperon-nuclear interactions from SU(3) chiral effective field theory. *Front. Phys.* **8**, 12 (2020).

69. Entem, D. R., Machleidt, R. & Nosyk, Y. High-quality two-nucleon potentials up to fifth order of the chiral expansion. *Phys. Rev. C* **96**, 024004 (2017).

70. Margueron, J., Hoffmann Casali, R. & Gulminelli, F. Equation of state for dense nucleonic matter from metamodeling. I. foundational aspects. *Phys. Rev. C* **97**, 025805 (2018).

71. Margueron, J., Hoffmann Casali, R. & Gulminelli, F. Equation of state for dense nucleonic matter from metamodeling. II. predictions for neutron star properties. *Phys. Rev. C* **97**, 025806 (2018).

72. Douchin, F. & Haensel, P. A unified equation of state of dense matter and neutron star structure. *Astron. Astrophys.* **380**, 151 (2001).

73. Koehn, H. et al. From existing and new nuclear and astrophysical constraints to stringent limits on the equation of state of neutron-rich dense matter. *Phys. Rev. X* **15**, 021014 (2025).

74. Abbott, B. P. et al. GW190425: observation of a compact binary coalescence with total mass $\sim 3M_{\odot}$. *Astrophys. J. Lett.* **892**, L3 (2020).

75. Antoniadis, J. et al. A massive pulsar in a compact relativistic binary. *Science* **340**, 6131 (2013).

76. Cromartie, H. T. et al. Relativistic Shapiro delay measurements of an extremely massive millisecond pulsar. *Nat. Astron.* **4**, 72–76 (2019).

77. Fonseca, E. et al. Refined mass and geometric measurements of the high-mass PSR J0740+6620. *Astrophys. J. Lett.* **915**, L12 (2021).

78. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: the MCMC Hammer. *PASP* **125**, 306 (2013).

79. Kashyap, R., Gupta, I., Dhani, A., Bapna, M. & Sathyaprakash, B. Optimizing Bayesian model selection for equation of state of cold neutron stars. <https://arxiv.org/abs/2502.03831> (2025).

80. Vinciguerra, S. et al. An updated mass-radius analysis of the 2017–2018 NICER data set of PSR J0030+0451. *Astrophys. J.* **961**, 62 (2024).

81. Borhanian, S. GWBENCH: a novel Fisher information package for gravitational-wave benchmarking. *Class. Quant. Grav.* **38**, 175014 (2021).

82. Buonanno, A., Iyer, B. R., Ochsner, E., Pan, Y. & Sathyaprakash, B. S. Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors. *Phys. Rev. D* **80**, 084043 (2009).

83. Wade, L. et al. 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* **89**, 103012 (2014).

84. Finstad, D. & Brown, D. A. Fast parameter estimation of binary mergers for multimessenger follow-up. *Astrophys. J. Lett.* **905**, L9 (2020).

85. Kunert, N., Pang, P. T. H., Tews, I., Coughlin, M. W. & Dietrich, T. Quantifying modeling uncertainties when combining multiple gravitational-wave detections from binary neutron star sources. *Phys. Rev. D* **105**, L061301 (2022).

86. Jülich Supercomputing Centre. JUWELS cluster and booster: exascale pathfinder with modular supercomputing architecture at Jülich Supercomputing Centre. *J. Large-scale Res. Facilities* **7**, A183 (2021).

87. Abbott, B. P. et al. Properties of the binary neutron star merger gw170817. *Phys. Rev. X* **9**, 011001 (2019).

88. Riley, T. E. et al. A NICER view of PSR J0030+0451: nested samples for millisecond pulsar parameter estimation *Astrophys. J. Lett.* **887**, L21 (2019).

89. Vinciguerra, S. et al. An updated mass-radius analysis of the 2017–2018 NICER data set of PSR J0030+0451. *Astrophys. J. Lett.* **961**, 62 (2023).

90. Salmi, T. et al. The radius of PSR J0740+6620 from NICER with NICER background estimates. *Astrophys. J. Lett.* **941**, 150 (2022).

91. Somasundaram, R. et al. Inferring three-nucleon couplings from multi-messenger neutron-star observations (supplemental dataset). <https://doi.org/10.5281/zenodo.1687820> (2025).

92. Somasundaram, R. et al. sv3ak/multimessenger-3n-constraints. <https://doi.org/10.5281/zenodo.16890550> (2025).

Acknowledgements

We thank C.L. Armstrong, K. Godbey, and P. Giuliani for feedback about the implementation of the PMM. We further thank D. Brown, C. Capano, C. Forssén, K. Hebeler, and W.G. Jiang for useful discussions. We also thank the Institute for Nuclear Theory at the University of Washington for its kind hospitality and stimulating research environment. R.S. acknowledges support from the Nuclear Physics from Multi-Messenger Mergers (NP3M) Focused Research Hub, which is funded by the National Science Foundation under Grant Number 21-16686, and from the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project number 20220541ECR. I.S. and A.S. were supported in part by the European Union’s Horizon 2020 research and innovation program (Grant Agreement No. 101020842). S.D., A.E.D., and I.T. were supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project number 20230315ER. A.E.D. was also supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). S.D. was also supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory, project 20250750ECR. I.T. was also supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under contract No. DE-AC52-06NA25396, by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) NUCLEI program, and by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project numbers 20220541ECR. Y.D. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-Project-ID 279384907–SFB 1245. P.L. is supported by the Natural Sciences & Engineering Research Council of Canada (NSERC). Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Colleges and Universities. Computational resources have been provided by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the U.S. Department of Energy National Nuclear Security Administration under Contract No. 89233218CNA000001, and by the National Energy Research Scientific Computing Center (NERSC), which is supported by the U.S. Department of Energy, Office of Science, under contract No. DE-AC02-05CH11231, using NERSC awards NP-ERCAP-m3319 and NP-ERCAP-m4338. I.S. and A.S. gratefully acknowledge the computing time provided on the high-performance computer Lichtenberg II at the TU Darmstadt. This is funded by the German Federal Ministry of Education and Research (BMBF) and the State of Hesse. I.S. and A.S. also gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer JUWELS⁸⁶ at Jülich Supercomputing Centre (JSC). This document has been approved for unlimited release and was assigned LA-UR-24-29187.

Author contributions

Conceptualization and project administration: R.S., I.S., P.L., A.S., I.T.; methodology and resources: R.S., I.S., S.D., P.L., A.S., I.T. Data curation: R.S., I.S., S.D., P.L. Software, validation, and writing-review and editing: R.S., I.S., S.D., A.E.D., Y.D., P.L., A.S., I.T. Formal analysis: R.S., I.S., S.D., A.E.D., P.L., A.S., I.T. Funding acquisition: S.D., P.L., A.S., I.T. Supervision: R.S., A.S., I.T. Visualization: R.S., I.S., A.E.D., P.L. Writing-original draft: R.S., I.S., Y.D., P.L., A.S., I.T.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-64756-6>.

Correspondence and requests for materials should be addressed to Rahul Somasundaram or Isak Svensson.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025