Quantifying uncertainties and correlations in the nuclear-matter equation of state

C. Drischler [1,2,*] J. A. Melendez [3,†] R. J. Furnstahl [3,‡] and D. R. Phillips [4,§]

[1]*Department of Physics, University of California, Berkeley, California 94720, USA*
[2]*Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[3]*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*
[4]*Department of Physics and Astronomy and Institute of Nuclear and Particle Physics, Ohio University, Athens, Ohio 45701, USA*

We perform statistically rigorous uncertainty quantification (UQ) for chiral effective field theory (χEFT) applied to infinite nuclear matter up to twice nuclear saturation density. The equation of state (EOS) is based on high-order many-body perturbation theory calculations with nucleon-nucleon and three-nucleon interactions up to fourth order in the χEFT expansion. From these calculations our newly developed Bayesian machine-learning approach extracts the size and smoothness properties of the correlated EFT truncation error. We then propose a novel extension that uses multitask machine learning to reveal correlations between the EOS at different proton fractions. The inferred in-medium χEFT breakdown scale in pure neutron matter and symmetric nuclear matter is consistent with that from free-space nucleon-nucleon scattering. These significant advances allow us to provide posterior distributions for the nuclear saturation point and propagate theoretical uncertainties to derived quantities: the pressure and incompressibility of symmetric nuclear matter, the nuclear symmetry energy, and its derivative. Our results, which are validated by statistical diagnostics, demonstrate that an understanding of truncation-error correlations between different densities and different observables is crucial for reliable UQ. The methods developed here are publicly available as annotated Jupyter notebooks.

I. INTRODUCTION

Calculations of observables in chiral effective field theory (χEFT) [1–4] are truncated at a finite order in the EFT expansion, leaving a residual error that should be quantified to enable robust comparisons to experiment and competing theories [5]. While χEFT is widely used to predict the nuclear-matter equation of state (EOS) [6–20] (see also Refs. [21–23] for recent reviews), a proper statistical analysis of the χEFT truncation errors for the EOS and associated observables is lacking. This work fills that gap. Implications for neutron-rich matter are elucidated in a companion paper [24].

In a recent paper, a Bayesian model for EFT truncation errors was developed that included their correlation across continuous independent variables, such as energy or scattering angle [25]. The model uses machine learning (ML) to determine the convergence and correlation pattern of the χEFT expansion by calibrating Gaussian processes (GPs) to the computed orders; this leads to statistical estimates of the omitted higher orders. Here we extend this method and apply it to many-body observables in infinite (nuclear) matter computed using χEFT nucleon-nucleon (NN) and three-nucleon

(3N) interactions up to next-to-next-to-next-to-leading order (N^{3}LO); specifically, to the EOS in the limits of pure neutron matter (PNM) and symmetric nuclear matter (SNM).

Our Bayesian model incorporates two types of correlations in the χEFT truncation error for different quantities. Given an observable $y(x)$, we diagnose and assess the impact of

(1) Type x: Correlations between $y(x)$ and $y(x')$. In infinite matter, the input points x could be Fermi momentum $k_{\rm F}$ or the density n. Type-x correlations are quantified and propagated via the ML truncation-error model proposed in Ref. [25].
(2) Type y: Correlations between discrete observables $y_i(x)$ and $y_j(x')$, e.g., between the EOS of PNM and SNM. This can include type-x correlations if the observables are considered at different input locations. Type-y correlations also include correlations between an observable and its derivatives [26–29]. We expand upon our ML framework to incorporate these novel correlations using *multitask* ML algorithms [30–33].

Both types are not only crucial to a robust uncertainty quantification (UQ) in infinite matter but also reveal physics about the system. In particular, type-x correlations tell us the persistence of information in χEFT across different densities; building them into our error model facilitates the reliable computation of derivatives of $y(x)$.

Our previous applications of Bayesian analysis to χEFT truncation errors focused on NN observables calculated using

---

*cdrischler@berkeley.edu
†melendez.27@osu.edu
‡furnstahl.1@osu.edu
§phillid1@ohio.edu

Weinberg power counting [25,34–36]. Reference [25] developed both the truncation-error model with correlations and statistical model-checking diagnostics (see also Ref. [37]) to check the validity of the model's postulated order-by-order convergence pattern and correlation structure. For some NN potentials we found a validation of basic EFT convergence expectations, with a breakdown scale Λ_b that is consistent with 600 MeV, despite the lack of renormalization-group (RG) invariance with this power-counting scheme. For other potentials our diagnostics clearly show deviations from these expectations, which were attributed to regulator artifacts. Thus, the statistical tools offer not only theoretical uncertainty bands, but also an alternative to RG diagnostics of whether an EFT is performing as advertised.

The stage is set for more wide-ranging applications and tests of these tools. There are three main reasons why infinite matter provides an attractive laboratory for studying their use and what they tell us about the convergence pattern of χEFT. First, translational invariance provides many simplifications in characterizing the system: it permits clear focus on the bulk properties of nuclear matter without confusion from surface effects that might complicate observables' convergence pattern in finite nuclei. Second, the densities of relevance for infinite matter are higher than for light nuclei, which provides a different perspective on the convergence of χEFT. Also, the controlled specification of density can help to illuminate the nature of that convergence, which can be obscured in applications to observables in finite nuclei. Third, 3N interactions drive nuclear saturation in SNM. They are quantitatively important at every proton fraction at nuclear saturation density $n_0 \approx 0.16$ fm^{-3} ($\rho_0 \approx 2.7 \times 10^{14}$ g cm^{-3}) and above.

Calculations of infinite matter with the N^{3}LO χEFT interactions we use for our study are feasible because many-body perturbation theory (MBPT) is available as a controlled and computationally efficient many-body method. The technology from Refs. [16,18] extends to high enough order in MBPT to ensure adequate many-body convergence for these χEFT NN and 3N potentials. Furthermore, the ability of MBPT to isolate contributions from different classes of diagrams can provide new insights into what determines the convergence pattern: calculations with and without 3N forces can be directly compared.

Given the marked difference in density between the NN system and infinite matter the reader may doubt that the same χEFT convergence pattern prevails in the two systems. 3N forces are present in infinite matter but, tautologically, not in the NN system. Many-body effects such as Pauli blocking may also affect the way that different χEFT orders contribute to observables. The diagnostics used to check that truncation errors were behaving as advertised in Refs. [25,34,35] can also be applied to χEFT calculations of the EOS. They allow us to determine whether the EFT breakdown scale in infinite matter is consistent with that found from analyses of few-nucleon observables.

The paper is organized as follows. The statistical model of Ref. [25] is briefly reviewed in Sec. II, including an explanation of the model's treatment of type-*x* correlations. In Sec. III we present the EOS for recent χEFT NN and 3N interactions up to N^{3}LO, analyze the corresponding order-by-order EFT

coefficients, obtain a Bayesian posterior for the EFT breakdown scale, provide error bands for the EOS in the limit of SNM and PNM, and present error ellipses for the predicted nuclear saturation point. We then study type-*y* correlations (those between observables) in Sec. IV and thereby derive uncertainty bands for the symmetry energy as a function of density. This is followed by new results for derivatives of the SNM EOS: order-by-order uncertainty estimates for the pressure and incompressibility. Section V has our summary and outlook. Further details—regarding our statistical model-checking diagnostics, the order-by-order values of the SNM and PNM EOS, and derivatives of GPs and multitask GPs—are presented in Appendices A–C. Additional figures are given in the Supplemental Material [38].

Although some results for PNM are presented here, this paper focuses on SNM. A companion paper [24] discusses the parallel analysis for PNM, and provides error estimates for the symmetry energy and its slope parameter that include the impact of both type-*x* and type-*y* correlations. The results presented in these two papers do not constitute an exhaustive study; rather we identify trends and issues and anticipate future refinements of our approach.

II. A MODEL OF EFT TRUNCATION ERRORS THAT INCLUDES CORRELATIONS

A. Previous work

In previous works we proposed a pointwise Bayesian statistical model to estimate EFT truncation errors for predicted observables y [34,35]. This model formalizes the notion of convergence of y at a single sampling point to allow one to credibly assess whether experimental data are consistent with theory. The convergence pattern should be a consequence of the EFT power counting. We incorporate this "expert knowledge" of the convergence pattern in our prior probability distributions. Those beliefs are subsequently updated using the actual order-by-order EFT results $\{y_0, y_1, y_2, \ldots, y_k\}$ [39]. Note that $y_1 \equiv 0$ in χEFT.

The authors of Ref. [40] applied the simplest version of the pointwise model from Ref. [34] to Brueckner-Hartree-Fock calculations of infinite matter based on an NN-only χEFT potential. They used $Q = k_F/\Lambda_b$ as the expansion parameter and found credibility intervals for PNM and SNM. They then followed Ref. [34] and applied a consistency check on the empirical coverage to validate their choice of Λ_b. The present work goes significantly beyond that of Ref. [40] in using a much more accurate many-body method, including 3N forces, analyzing and accounting for correlations within and between observables, applying a suite of model-checking diagnostics, and deriving posteriors for Λ_b.

B. Including correlations

To include the effects of correlations between EFT predictions at different values of independent variables, generically denoted as *x*, we extended the pointwise model to functions $y(x)$, encoding the idea of curvewise convergence for observables via GPs [25]. GPs are powerful tools for both regression and classification, and have become popular in many

fields, including statistics, physics, and applied mathematics [41–43]. The GP parameters are interpretable from an EFT convergence standpoint, and can be easily calibrated against known order-by-order predictions.

We give here a brief overview of the statistical model and refer the reader to Ref. [25] for details and examples (including a Jupyter notebook that reproduces the figures in that paper). The model for the truncation error $\delta y_k(x)$ at order k in the EFT expansion (e.g., $k = 4$ at N^{3}LO in χEFT) is based on the decomposition

$$\delta y_k(x) = y_{\text{ref}}(x) \sum_{n=k+1}^{\infty} c_n(x) Q^n(x), \qquad (1)$$

where $y_{\text{ref}}(x)$ is a dimensionful quantity that sets the reference scale of variation with x, $Q(x)$ is a dimensionless expansion parameter, and the $c_n(x)$ are dimensionless coefficients. The observable $y_k(x)$ itself at order k is decomposed as

$$y_k(x) = y_{\text{ref}}(x) \sum_{n=0}^{k} c_n(x) Q^n(x), \qquad (2)$$

where the observable coefficients $c_n(x)$ are extracted from the order-by-order calculations, given $y_{\text{ref}}(x)$ and $Q(x)$, using

$$y_0(x) \equiv y_{\text{ref}}(x) c_0(x), \qquad (3)$$

$$\Delta y_n(x) \equiv y_{\text{ref}}(x) c_n(x) Q^n(x). \qquad (4)$$

Here, $\Delta y_n(x)$ is the order-n correction to the observable. Since all scales have been factored into $y_{\text{ref}}(x)$ and $Q(x)$, the $c_n(x)$ are expected to be *natural*, or, in other words, of order 1, assuming there are no systematic cancellations (e.g., fine tuning) that would make the coefficients much smaller than the reference size.

We postulate that the properties of the unobserved $c_{n>k}(x)$ are the same as the observed $c_{n\leqslant k}(x)$. Specifically, our model assumes that all the $c_n(x)$ are independent and identically distributed (i.i.d.) random curves. We formalize the EFT convergence assumptions by modeling the coefficients $c_n(x)$ as independent draws from a single underlying GP. A brief introduction to GPs in this context is given in Ref. [25], whose notation is $\mathcal{GP}[m(x), \kappa(x, x')]$ for some mean function $m(x)$ and positive-semidefinite covariance function (also called kernel) $\kappa(x, x')$. For more in-depth discussions, see Refs. [43–45]. We adopt $m(x) = 0$ since corrections are just as likely to be positive as they are to be negative, and $\kappa(x, x'; \bar{c}, \ell) = \bar{c}^2 r(x, x'; \ell)$, so that[1]

$$c_n(x) \,|\, \bar{c}^2, \ell \overset{\text{iid}}{\sim} \mathcal{GP}[0, \bar{c}^2 r(x, x'; \ell)]. \qquad (5)$$

We choose the correlation function $r(x, x'; \ell)$ to be a radial basis function (RBF), which ensures that the $c_n(x)$ are very smooth functions (up to numerical noise that is handled by a

white-noise term),

$$r(x, x'; \ell) = \exp\left[-\frac{(x - x')^2}{2\ell^2}\right]. \qquad (6)$$

The length scale ℓ controls how quickly the $c_n(x)$ vary as a function of x; a small length scale implies that the $c_n(x)$ vary quickly, whereas a large length scale implies the opposite. Importantly, this kernel is *stationary*, meaning that it only depends on the absolute difference $|x - x'|$. Stationarity implies that the $c_n(x)$ should share similar correlation properties across all x, up to fluctuations. For example, the curves should not vary rapidly at small x and flatten out at large x.

To update the $\bar{c}^2$ based on observed $c_n(x)$ requires a prior. We choose the scaled inverse-chi-squared distribution [25] with $\nu = 10$ degrees of freedom and scale parameter $\tau^2 = (\nu - 2)/\nu$. This informative prior on $\bar{c}^2$ has a mean value of 1, which builds in our assumption that the coefficients should be naturally sized.

Thus, in summary, the $c_n(x)$ are modeled as random functions. These functions share a common variance $\bar{c}^2$, and should each look like random draws from a GP with an RBF kernel $\kappa(x, x'; \ell)$.[2] The hyperparameters $\bar{c}^2$ and ℓ can be learned from the order-by-order χEFT predictions; see Ref. [25].

Some intuition about the nature of the GPs used to model $c_n(x)$ in this work can be gleaned from Fig. 1. In each subplot are four random draws from a GP with an RBF kernel. The GPs have an arbitrary mean and marginal variance that are the same in all panels. Each row depicts a different GP, with increasing length scale ℓ from the top row to the bottom. Comparing different columns illustrates the nature of the fluctuations one should expect when working with only a few samples. The smoothness of the coefficient functions $c_n(x)$—a feature of the RBF kernel—is consistent with expectations and observations from EFTs; e.g., compare to the real $c_n(x)$s from NN scattering in Fig. 10 of Ref. [35] and from nuclear matter in Figs. 2 and 3. We contend that the modeling of $c_n(x)$ as draws from GPs is supported by the examples shown in Fig. 1. However, it is easy to be fooled by visual evidence. We therefore rely on the model checking diagnostics in Appendix A to validate the GP hypothesis.

Once we make the inductive step that the higher-order coefficients (which we do not have) also obey Eq. (5), it follows that the truncation error $\delta y_k(x)$ defined by Eq. (1) is a geometric sum over independent normally distributed variables. Its distribution is [25]

$$\delta y_k(x) \,|\, \bar{c}^2, \ell, Q \sim \mathcal{GP}[0, \bar{c}^2 R_{\delta k}(x, x'; \ell)], \qquad (7)$$

with

$$R_{\delta k}(x, x'; \ell) \equiv y_{\text{ref}}(x) y_{\text{ref}}(x') \frac{[Q(x)Q(x')]^{k+1}}{1 - Q(x)Q(x')} r(x, x'; \ell). \qquad (8)$$

The marginal variance of Eq. (7) is x dependent in general and equal to $\bar{c}^2 R_{\delta k}(x, x; \ell)$ [25].

---

[1]We use the common shorthand notation in statistics, in which $z \sim \cdots$ reads as "the variable z is distributed as $\cdots$." Some authors also use $\text{pr}(z) = \cdots$. See also Ref. [25]. The "i.i.d." above the $\sim$ indicates that the c_ns are a set of "i.i.d." random curves.

[2]Generalizations of this model are discussed in Ref. [25]; Appendix A considers relaxing the assumption of a single GP in favor of separate GPs for NN and 3N contributions.
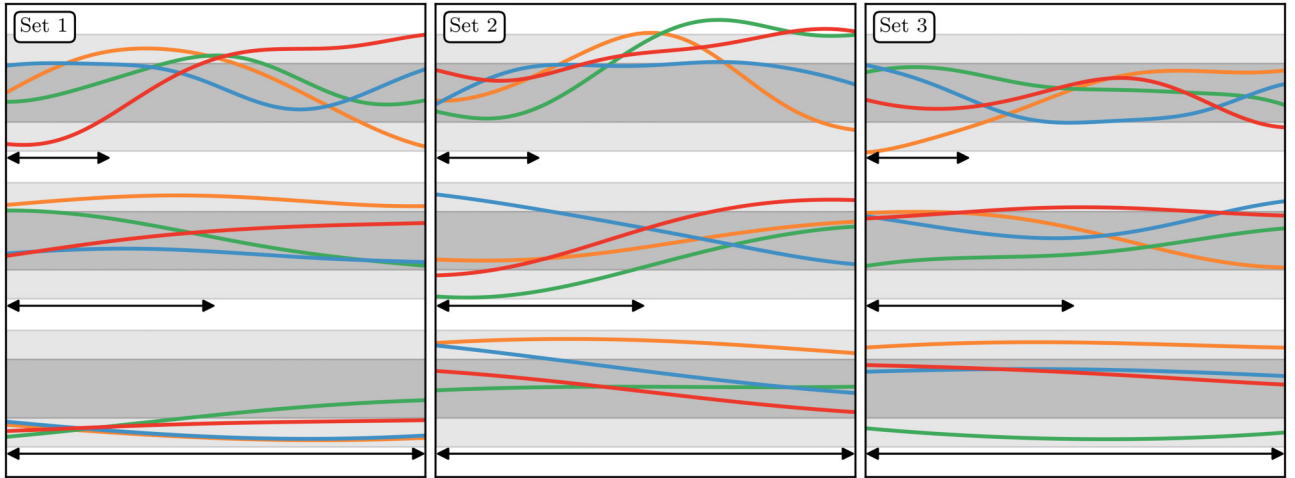
FIG. 1. The three columns show three sets of random functions. For each column the three rows show the output of GPs that have the same (arbitrary) mean and marginal variance, but differ in their radial basis function (RBF) kernel: a different length scale is used in each of the three rows. The length scales are indicated by the double-headed arrow at the bottom of each panel Each panel then contains four functions, i.e., four draws from the GP. The x-axis could represent an independent variable such as energy, angle, or density and the y-axis can be thought of as the order-by-order EFT coefficients for an observable of interest. The dark (light) shaded bands represent one (two) standard deviations from the mean.

Note that one can embed our truncation error model within a Bayesian parameter estimation framework to find posteriors for the low-energy constants (LECs) of χEFT interactions [36,46]. Here we will take the LECs as given for our analysis of infinite matter. An important subject for future work is a complete Bayesian analysis that consistently combines uncertainties on LECs from fitting nuclear interactions to data with truncation errors in order to find the full uncertainty in χEFT predictions.

III. RESULTS FOR PNM AND SNM

This section describes how we incorporate type-x correlations in our model of EFT convergence, i.e., account for the fact that the truncation error varies smoothly with
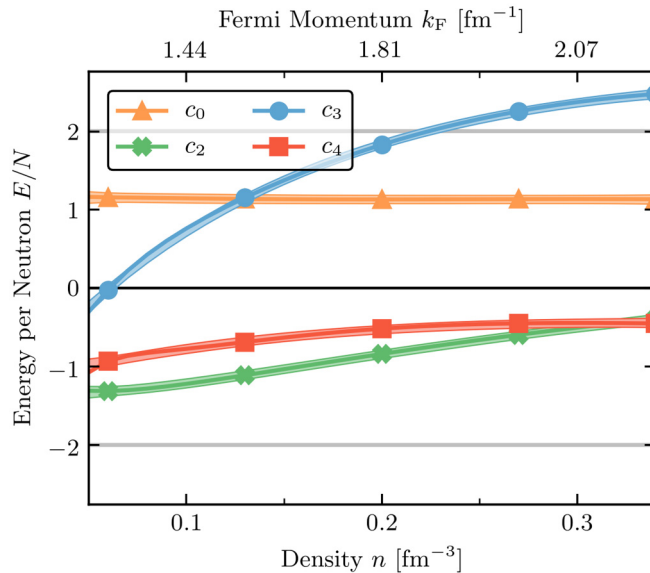

FIG. 2. Observable coefficients, c_n, for $E/N(n)$ up to N³LO, as a function of density n, obtained using the $\Lambda = 500$ MeV interactions in Table I. Markers indicate training points, gray lines indicate $2\bar{c}$, and colored bands are 68% credible intervals of the interpolating GPs. The estimated hyperparameters are given by $\bar{c} = 1.0$ and $\ell = 0.97$ fm^{-1}. Note that the secondary x axis (at the top of the figure) is not linear in k_F^{PNM}.
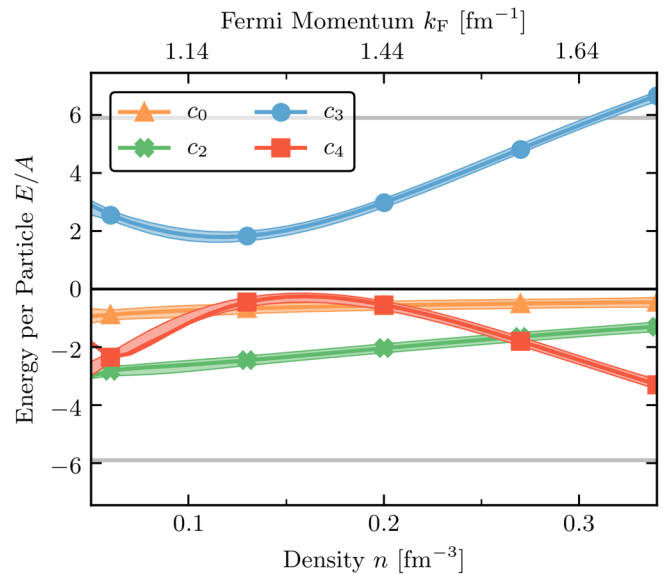

FIG. 3. Observable coefficients c_n for $E/A(n)$ up to N³LO, as a function of density n, using the $\Lambda = 500$ MeV interactions in Table I. See Fig. 2 for the notation. The estimated hyperparameters are given by $\bar{c} = 2.9$ and $\ell = 0.48$ fm^{-1}. Note that the secondary x axis (at the top of the figure) is not linear in k_F^{SNM}.

TABLE I. NN and 3N interactions considered in this work [18]. The interactions are based on the order-by-order NN potentials by Entem, Machleidt, and Nosyk [48] (EMN) with momentum cutoffs $\Lambda = 450$ and 500 MeV up to N^{3}LO and 3N forces at the same order and cutoff. Reference [18] fit the two 3N LECs c_D and c_E to the triton binding energy and the empirical saturation point of SNM. Chiral 3N forces up to N^{3}LO also depend on the NN LECs C_S and C_T as well as the πN LECs c_1, c_3, and c_4. Their values were taken from the associated NN potential. The applied LO and NLO potentials are NN only and therefore omitted. More details can be found in Ref. [18] and its Supplemental Material.

Chiral order	NN potential	c_D	c_E
N^{2}LO	EMN 450 MeV	2.25	0.07
N^{2}LO	EMN 500 MeV	-1.75	-0.64
N^{3}LO	EMN 450 MeV	0.00	-1.32
N^{3}LO	EMN 500 MeV	-3.00	-2.22

density. With the Bayesian framework described in Sec. II, it is straightforward to analyze the convergence patterns of the energy per particle in PNM (E/N) and SNM (E/A) and obtain the size and correlation structure of the truncation error. We then extract first posterior distributions for the predicted saturation point of SNM.

A. Nuclear-matter equation of state

Our analysis is based on the MBPT calculations of $E/N(n)$ and $E/A(n)$ up to $2n_0$ in Refs. [18,24] and Refs. [18,47], respectively, equidistantly sampled at $n = 0.05, 0.06, \ldots, 0.34$ fm^{-3}. These high-order MBPT calculations are driven by the novel Monte Carlo framework introduced in Ref. [18]. In this framework, arbitrary interaction and many-body diagrams can be efficiently evaluated using automatic code generation, which enables calculations with controlled many-body uncertainties for the employed NN and 3N interactions (see the references for details).

The authors of Ref. [18] combined NN potentials developed by Entem, Machleidt, and Nosyk [48] with 3N forces at the same order and cutoff to construct a set of order-by-order NN and 3N interactions up to N^{3}LO. The two 3N LECs c_D and c_E that govern, respectively, the intermediate- and short-range 3N contributions at N^{2}LO were adjusted to the triton binding energy and the empirical saturation point of SNM. For the momentum cutoffs $\Lambda = 450$ and 500 MeV, three 3N forces with different combinations of c_D and c_E and reasonable saturation properties were obtained at N^{2}LO and N^{3}LO. However, the terms of the 3N forces proportional to c_D and c_E do not contribute to PNM with nonlocal regulators [49], so there is only one neutron-matter EOS determined for each momentum cutoff and chiral order. And even our results for SNM at a given cutoff do not differ significantly for the different 3N fits. We therefore restrict the discussion here to one Hamiltonian for each cutoff as summarized in Table I. Additional figures focusing on the 450 MeV potentials are given in the Supplemental Material [38].

B. Extracting observable coefficients

The observable coefficients $c_n(x)$ form the backbone of the convergence model and the training of the GP hyperparameters. To extract the coefficients, we need to assign values to x, $y_{\text{ref}}(x)$, and $Q(x)$ based on the system under consideration.

For the independent variable x we have two clear choices: k_F or the density $n = g k_F^3 / (6\pi^2)$ with the spin-isospin degeneracy $g = 2$ for PNM and $g = 4$ for SNM. The choice of x is important because we assume a GP kernel for $c_n(x)$ that is *stationary*, i.e., the variance and correlation length are the same across the entire space of the independent variable selected. In other words, we want the $c_n(x)$ to be approximately as curvy at low density as at high density, and this will (presumably) be better satisfied for one of these two choices than for the other. Since we do not have strong theoretical arguments regarding stationarity of the EFT truncation error in either k_F or n, we instead rely on empirical evidence. When plotting the observable coefficients as functions of n, they are more compressed at low density and stretched out at high density; when plotting versus k_F, the coefficients appear to have a slightly more homogeneous correlation structure. Therefore, although the evidence is only slight, we choose our GP input space to be k_F. Because readers usually want to know the density dependence of results, we still display graphs in which n varies linearly and is the main independent variable. But it should be borne in mind that the type-x correlations are actually formulated in k_F. Predictions across a larger range of k_F would be needed to provide more conclusive evidence for this choice of x.

Although we have chosen k_F as the GP input space, we could, in fact, have defined the input space with k_F replaced by γk_F for some constant $\gamma > 0$. However, we emphasize that our analysis is actually independent of the choice of γ. This is because the RBF kernel (6) is stationary and we use a scale invariant prior on ℓ for each system. Therefore, the posterior for $\bar{c}$ and ℓ contain the same information regardless of how we scale k_F. The only effect of the choice of γ is a cosmetic one: the posterior in ℓ will be scaled by that factor too. We return to this point when discussing the nuclear symmetry energy in Sec. IV A.

We choose the reference scale for nuclear-matter EOS to be

$$y_{\text{ref}}(k_F) = 16 \text{ MeV} \times \left(\frac{k_F}{k_{F,0}} \right)^2, \tag{9}$$

where $k_{F,0}$ is the Fermi momentum associated with $n_0 = 0.16$ fm^{-3}; i.e., $k_{F,0}^{\text{PNM}} = 1.680$ fm^{-1} and $k_{F,0}^{\text{SNM}} = 1.333$ fm^{-1}. Our findings indicate that this is a good approximation to the size of the LO predictions of $E/N(n)$ and $E/A(n)$, and sets a reasonable scale for the convergence for higher χEFT orders.

A natural choice for the expansion parameter based on experience with free-space NN scattering is $Q \propto k_F/\Lambda_b$. Conceptually, we might want to consider $Q = \gamma k_F/\Lambda_b$, where the constant prefactor $\gamma > 0$ arises, e.g., from an average of momenta over the Fermi sea ($\gamma = \sqrt{3/5}$) [18]. In fact, because our statistical model only constrains $Q(k_F)$, different choices of γ are implicitly considered if we take the simplest choice:

$$Q(k_F) = \frac{k_F}{\Lambda_b}. \tag{10}$$

A soft scale γk_F with $\gamma \neq 1$ will give the same results we find with Eq. (10) as long as the inferred value of Λ_b is also rescaled to $\gamma \Lambda_b$. This means we cannot make a connection to NN scattering results, where $\Lambda_b \approx 600$ MeV is favored, solely based on the results of the statistical analysis. If $\Lambda_b = 600$ MeV were asserted to be the breakdown scale we could infer γ. Conversely, if theoretical arguments for a particular γ were adduced we could determine Λ_b. For this first study we take the simplest choice, Eq. (10), let Λ_b adopt a value learned from the data, and reserve speculation over the broader meaning of the result found.

However, the choice of the soft scale in the numerator of Eq. (10) is not clear when computing a quantity, such as the symmetry energy, that is obtained from both $E/N(n)$ and $E/A(n)$ at a specific density. Then $k_F^{\rm PNM}$ and $k_F^{\rm SNM}$ differ, and this would have to be accounted for when defining the Q used to extract the coefficients. Ultimately this issue does not affect our results though: we deal with it through the use of *multitask* GPs (see Sec. IV A).

The EFT expectation is then that each successive order $\Delta y_n(x)$ should decrease by about a factor of $Q = k_F/\Lambda_b \gtrsim 1/3$ (depending on density), except for Δy_2, where the change from LO to NLO leads to Q^2 improvement. Tables III and IV of Appendix B show that the predicted energy per particle for PNM and SNM is consistent with this convergence pattern, as long as we allow for statistical variations of the coefficients $c_2(k_F)$, $c_3(k_F)$, and $c_4(k_F)$ in Eq. (4).

Reference [18] showed that the residual MBPT uncertainty is much smaller than the estimated χEFT truncation error for the interactions considered here. Nevertheless, to be conservative, we assign an (uncorrelated) uncertainty of 0.1% (or $\geqslant 20$ keV, whichever is greater) to the total energy per particle in order to account for this residual uncertainty. That is, when predicting the EOS with theory uncertainties given by Eq. (7), we add a white noise term to the truncation error kernel, $\bar{c}^2 R_{\delta k}$, of Eq. (8).

But, before making those predictions, we need to obtain the GP hyperparameters from the convergence pattern of χEFT. Because each $c_n(k_F)$ is extracted using a different power of $Q(k_F)$, even this small noise in the many-body calculations gets magnified as the EFT order n grows. Our GP approach can straightforwardly account for such uncertainties in the training data. We smooth the EOS by fitting a GP at each order before computing the observable coefficients to obtain a noise level that is approximately constant in EFT order. This has no noticeable effect on the total energy per particle. We then include a white noise term (called a nugget in this context) in the RBF kernel [Eq. (6)] when training the GP hyperparameters. This regularizes the matrix inversion of our GP framework. The variance of the white noise is chosen to be $\sigma^2 = 5 \times 10^{-4}$.

With choices for x, $y_{\rm ref}(x)$, and $Q(x)$ in hand, we can make the observations of the previous paragraph rigorous through a statistical analysis of the convergence pattern of $E/N(n)$ and $E/A(n)$. The corresponding observable coefficients $c_n(k_F)$ have been extracted in Figs. 2 and 3 using $\Lambda_b = 600$ MeV. (We return to the choice of Λ_b shortly.) Each curve appears to be naturally sized and relatively smooth across this range of density. Below N^{2}LO, where only NN forces are present in
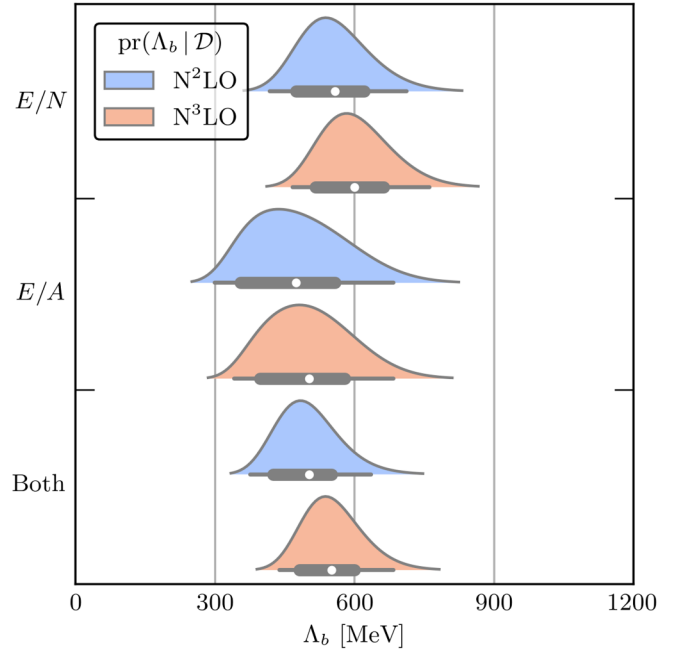

FIG. 4. The posteriors for the EFT breakdown scale Λ_b using orders through N^{2}LO (blue bands) and N^{3}LO (red bands) corresponding to Figs. 2 and 3. The upper pair of posteriors comes from analyzing E/N, the middle pair from E/A, and the bottom from a combined analysis. In all cases a Gaussian prior centered at $\Lambda_b = 600 \pm 150$ MeV is used. The combined N^{3}LO posterior is consistent with the $\Lambda_b \approx 600$ MeV found when considering free-space NN scattering observables [35].

χEFT, the coefficients have particularly large length scales, whereas the higher orders exhibit greater curvature because 3N contributions affect their density dependence. The GP hyperparameters, $\bar{c}$ and ℓ, are trained on all coefficients except c_0 (LO), and hence find values that represent the features of all $c_n(k_F)$ simultaneously. [The leading-order term is often disregarded when we perform this induction on the $c_n(k_F)$ because it informed the selection of $y_{\rm ref}$.] As a test that these hyperparameters are appropriate, we plot in Figs. 2 and 3 also the GP interpolants (colored 1σ bands) with these $\bar{c}$ and ℓ values. These smoothly interpolate the density-dependence of each coefficient, including the untrained data.

The hyperparameter $\bar{c}$ is easily updated due to its conjugate prior, whereas ℓ is determined by optimizing the log-likelihood. To support the use of $\Lambda_b = 600$ MeV as a point estimate for the EFT breakdown scale, along with the fit values of ℓ, we provide posterior distributions $\text{pr}(\Lambda_b, \ell \,|\, \mathcal{D})$ trained on each $E/N(n)$, $E/A(n)$, and both simultaneously (assuming that they are independent data). The marginal posteriors $\text{pr}(\Lambda_b \,|\, \mathcal{D})$ and $\text{pr}(\ell \,|\, \mathcal{D})$ can then be obtained by integrating over ℓ and Λ_b, respectively. We use a Gaussian prior $\text{pr}(\Lambda_b) = 600 \pm 150$ MeV for the breakdown scale, and a scale invariant prior $\text{pr}(\ell) = 1/\ell$ for the each length scale of $E/N(n)$ and $E/A(n)$.

Figure 4 shows the posteriors for Λ_b. We compute the posteriors using observable coefficients up to N^{2}LO and N^{3}LO (neglecting LO) as a check of the robustness of our analysis.
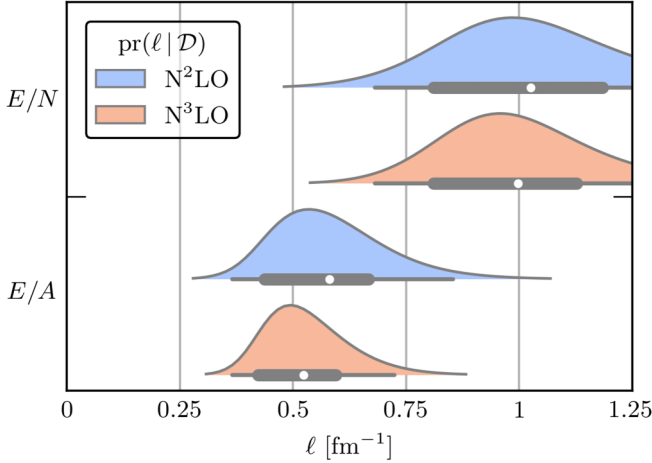
FIG. 5. Length-scale posteriors organized similarly to Fig. 4. A scale-invariant prior proportional to $1/\ell$ is used, and each length scale is relative to the k_F of each system. If we were to use a single k_F prescription, the length scales in PNM and SNM, ℓ_{PNM} and ℓ_{SNM}, respectively, would transform just as k_F, making the posteriors shift towards agreement. See the discussion of the input space in the main text.

It is clear that the results behave consistently between orders and between observables, with $E/A(n)$ possibly preferring a smaller breakdown scale than $E/N(n)$. When combined, the posterior has a median around $\Lambda_b \approx 560$ MeV with a 1σ spread of over 50 MeV. This is consistent with $\Lambda_b \approx 600$ MeV as extracted from np cross sections and angular observables of free-space NN scattering. Thus, we choose $\Lambda_b = 600$ MeV here for simplicity. We reiterate that, in reality, only the ratio $Q(k_F) = k_F/\Lambda_b$ is determined by this analysis, and hence extracting Λ_b is contingent on our choice of the numerator, k_F. If, for example, k_F^{PNM} were used for $E/A(n)$, this would bring the Λ_b posteriors to even better agreement, with a median value $\Lambda \gtrsim 600$ MeV. Because $k_F^{PNM} > k_F^{SNM}$ for a given density, and with the estimate $\Lambda_b \simeq 600$ MeV, the truncation error of $E/A(n)$ would then grow accordingly.

Figure 5 depicts the length scale posteriors for the observable coefficients in PNM and SNM. If the length scales were put on a common scale then the posteriors would become more aligned. The fact that the $c_n(k_F)$ from PNM and SNM could share a common ℓ in the same k_F scale could prove useful when modeling the correlations of the convergence patterns in $E/N(n)$ and $E/A(n)$. We return to this in Sec. IV A. Both length scales are relatively large given the range of k_F used in this work. This implies that the truncation error is a *highly* correlated quantity, and hence that an estimate of this correlation will prove crucial to a robust UQ in infinite matter.

We have provided one reasonable implementation of a GP-based EFT convergence model in this subsection. Other choices for x, $y_{ref}(x)$, and $Q(x)$ could be made. We provide an example of an alternative $y_{ref}(x)$ in Appendix A, which also provides model checking diagnostics for the interested reader to verify our convergence model for these systems. We find evidence that $c_3(k_F)$ may be an outlier in terms of the large

effect of the 3N contributions that enter χEFT at N^{2}LO. If one does not believe that such large corrections will continue, it may prove useful to leave $c_3(k_F)$ (N^{2}LO) out of our inductive model for higher-order terms.

Additionally, the diagnostics point to the possibility that the NN-only coefficients $c_0(k_F)$ (LO) and $c_2(k_F)$ (NLO) may have a different correlation structure than higher orders. As noted above, this is suggested by a visual inspection of Figs. 2 and 3, where $c_0(k_F)$ and $c_2(k_F)$ appear much flatter than $c_3(k_F)$ (N^{2}LO) and $c_4(k_F)$ (N^{3}LO). An investigation in this direction is presented in Appendix A. There we have attempted to isolate the strongly repulsive 3N contributions that change the correlation structure by splitting the coefficients into NN-only and residual 3N coefficients with each having different k_F dependence in $y_{ref}(x)$. This succeeds in making the coefficients more uniform and improves the diagnostics for PNM, but does not improve SNM significantly. Crucially, the order-by-order uncertainty bands for PNM and SNM presented in the next section are almost unchanged when this alternative model is used; the saturation ellipses do become slightly larger though. We provide these details, along with annotated Jupyter notebooks [50] that generate them, to promote further investigation, possibly with other EFT implementations, into the systematic convergence of infinite matter.

C. Quantified uncertainties for PNM and SNM

The GP truncation error model described in Sec. II combined with the hyperparameter estimates now permit the first statistically rigorous χEFT uncertainty bands in infinite matter. Figures 6 and 7 depict predictions of $E/N(n)$ and $E/A(n)$, respectively, up to N^{3}LO with 68% credible intervals (colored bands). From left to right, top to bottom, the panels show the order-by-order progression of EFT uncertainties as the χEFT order increases. The gray box in Fig. 7 represents the empirical saturation point, $n_0 = 0.164 \pm 0.007$ fm^{-3} with $E/A(n_0) = -15.86 \pm 0.57$ MeV, obtained from a set of energy-density functionals in Ref. [18,51]. We stress, however, that the quoted uncertainty in the empirical saturation point does not permit a statistical interpretation (e.g., 1σ credibility interval), in contrast to the results discussed this work.

In these figures, one might be tempted to count how frequently 1σ bands from one order contain the prediction of the subsequent orders, and to compare the frequency to the nominal value of 68%. If the bands are too conservative (aggressive), more (less) than 68% of the points will lie within the bands. But the highly correlated nature of the truncation error renders such an assessment difficult: if any point along a given curve is contained (not contained) within an uncertainty band, nearby points will also likely be inside (outside) the band. Long correlation lengths indicate that the effective sample size, N_{eff}, is much smaller then the number of data points. $\sqrt{N_{eff}}$ fluctuations mean that the bands, within which statistically consistent truncation-error prescriptions fall, are wider.

This is not a failure of our convergence model; these correlations are real and must be dealt with one way or another. On the contrary, we can provide estimates of exactly how correlated the truncation error is, and use that to inform us
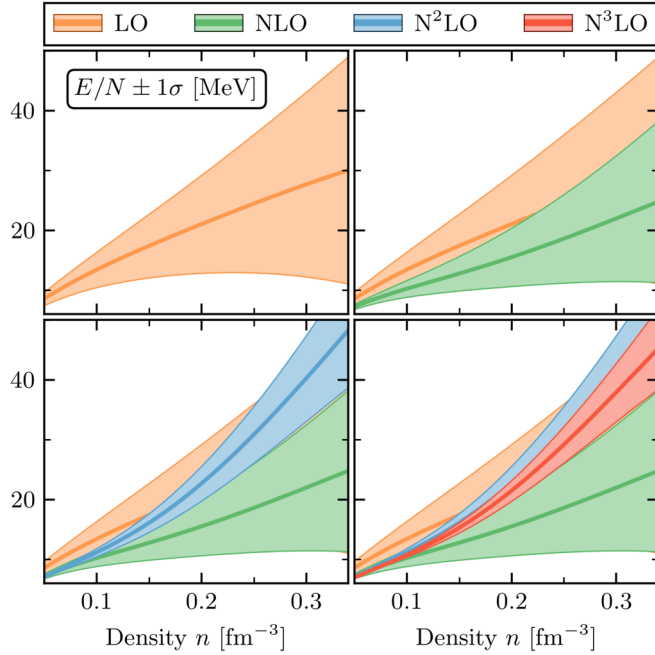
FIG. 6. Energy per particle in PNM with truncation errors using the $\Lambda = 500$ MeV interactions in Table I. From left to right, top to bottom, the panels show the order-by-order progression of EFT uncertainties as the χEFT order increases. The bands indicate 68% credible intervals.
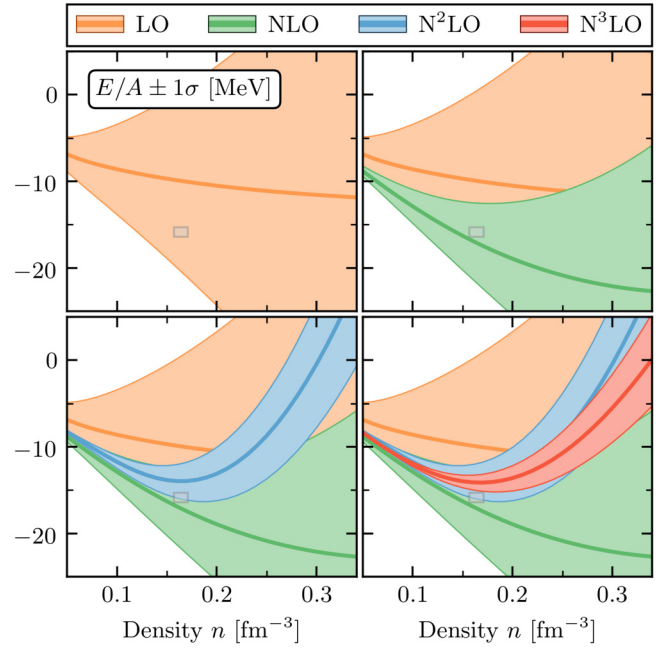

FIG. 7. Similar to Fig. 6 but for SNM. The gray box depicts the empirical saturation point, $n_0 = 0.164 \pm 0.007$ fm^{-3} with $E/A(n_0) = -15.86 \pm 0.57$ MeV, obtained from a set of energy-density functionals [18,51] (see the main text for details).

how perturbed we should be by any perceived "failure" of the uncertainty bands to align with our intuitions. We show a useful diagnostic tool in Fig. 8, which plots the empirical coverage of the credible intervals versus the choice of credible interval [25,35]. Importantly, we provide gray bands which account for random fluctuations due to the finite sample size $N_{\rm eff}$. As anticipated, the bands are very large due to the presence of correlations, much larger than the binomial bands one would obtain by naïve counting (compare to Figs. 11 to 20 in Ref. [35]). Correctly accounting for these correlations shows that the empirical coverage is statistically consistent for all credible intervals.

Next, we determine the location of the nuclear saturation point [i.e., the minimum of $E/A(n)$]. Ours is the first analysis to do this with fully correlated truncation errors. To assess saturation properties of nuclear interactions, one could follow a "Coester plot" approach (see, e.g., Refs. [16,18]) and ask whether the uncertainty bands in Fig. 7 overlap with the empirical saturation point (gray box) at a certain credibility level and how close the mean values are to that region. For example, while the N^{2}LO band completely overlaps, the N^{3}LO does so only partially.

Instead, we create here the foundation for a statistical analysis of nuclear saturation properties and obtain the joint posterior $\mathrm{pr}(E/A(n_0), n_0 \mid \mathcal{D})$ given $\mathcal{D}$, the order-by-order predictions of $E/A(n)$ up to $2n_0$. We compute this distribution by sampling thousands of curves from the GP interpolant of $E/A(n)$ and extracting for each of them the minimum. The resulting posteriors at N^{2}LO (blue bands) and N^{3}LO (red bands) are depicted in Fig. 9 with 2σ ellipses. They are

well-approximated at N^{2}LO and N^{3}LO by a two-dimensional Gaussian. The mean and covariance of the highest-order prediction, is given by

$$
\begin{bmatrix} n_0 \\ \frac{E}{A}(n_0) \end{bmatrix} \approx \begin{bmatrix} 0.170 \\ -14.3 \end{bmatrix} \quad \text{and} \quad \Sigma \approx \begin{bmatrix} 0.016^2 & -0.015 \\ -0.015 & 1.0^2 \end{bmatrix}
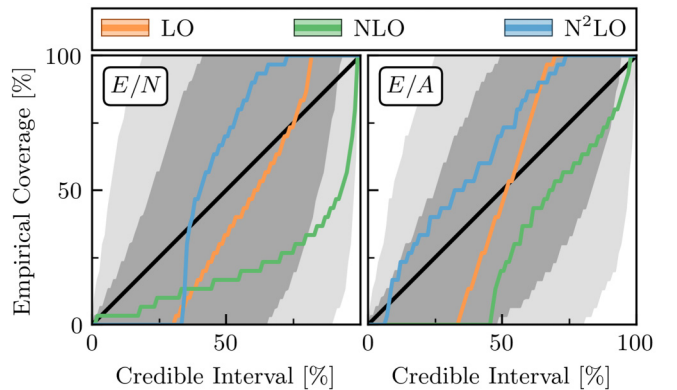$$

(11)

FIG. 8. Credible-interval diagnostics for the $E/N(n)$ (left-hand side) and $E/A(n)$ uncertainty bands (right-hand side) for the $\Lambda = 500$ MeV interactions in Table I; for details see Ref. [25]. At each order we construct an uncertainty band for the upcoming correction (not the full truncation error) and test whether the next order is contained within it at a specific credible interval. The expected size of fluctuations due to the finite effective sample size of the curves is depicted using dark (light) gray bands for the 68% (95%) interval. Both bands are quite large, which shows that correlations are crucial to assess whether truncation errors have been properly assigned.
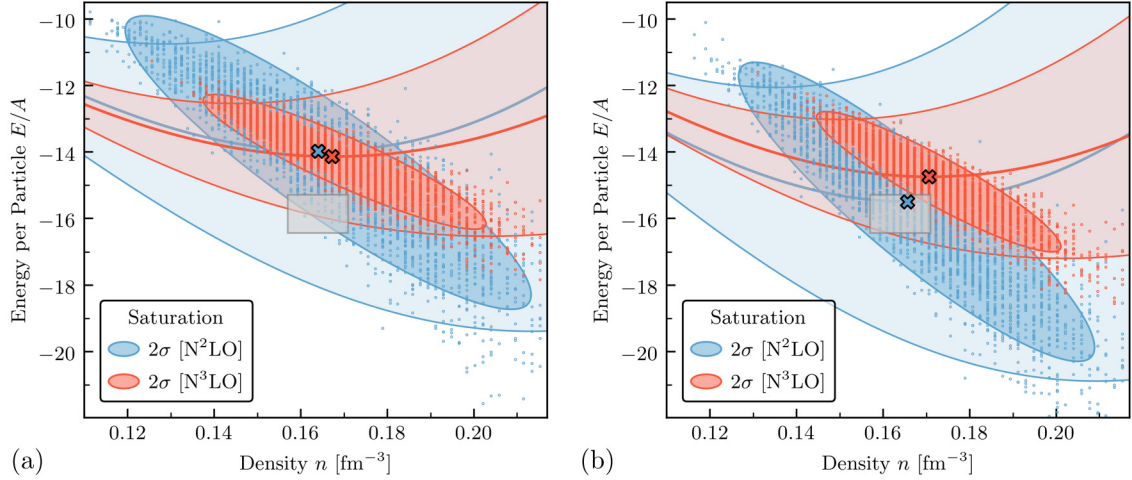
FIG. 9. Predicted nuclear saturation point of SNM at N^{2}LO (blue) and N^{3}LO (red) with 95% (2σ) credible interval ellipses including correlated truncation errors. Panels (a) and (b) show the $\Lambda = 500$ MeV and $\Lambda = 450$ MeV interactions described in Table I, respectively. The colored crosses depict the minimum of each EOS obtained at that order without accounting for truncation errors. In contrast, each dot is the minimum of a curve sampled from the corresponding $E/A(n)$ GP in Fig. 7, which are used to fit the ellipses. These should be compared to the 2σ uncertainty bands in the same color, which are estimated using $\Lambda_b = 600$ MeV. The gray box again depicts the empirical saturation point.

for the $\Lambda = 500$ MeV potentials [Fig. 9(a)] and by

$$
\begin{bmatrix} n_0 \\ \frac{E}{A}(n_0) \end{bmatrix} \approx \begin{bmatrix} 0.173 \\ -14.9 \end{bmatrix} \quad \text{and} \quad \Sigma \approx \begin{bmatrix} 0.014^2 & -0.014 \\ -0.014 & 1.1^2 \end{bmatrix}
$$

(12)

for the $\Lambda = 450$ MeV potentials [Fig. 9(b)] in Table I. The off-diagonal terms in the covariance matrices render the posteriors elliptical, rotated at an angle such that the N^{3}LO bands barely (if at all) overlap with the empirical saturation point. These findings are consistent with the conclusions in Ref. [18], albeit that work did not employ the statistical tools presented here.

Reference [52] studied binding energies and charge radii of medium-mass to heavy nuclei based on these χEFT NN and 3N interaction constrained by empirical saturation properties (see Figs. 6 and 7 in that reference). While the selected closed-shell oxygen, calcium, and nickel isotopes are under-bound (as expected from the findings in infinite matter), the charge radii are too large, opposite to the expectation from infinite matter. Furthermore, the sensitivity of the observables to the 3N low-energy coupling c_D is significantly less than that to the infinite-matter properties is (see Figs. 8 and 9 of Ref. [52]). The link between finite nuclei in this mass range and infinite matter thus seems to be more intricate than one might naïvely expect [52].

IV. RESULTS FOR DERIVED QUANTITIES

This section describes the second set of correlations addressed in this work: "type-y" correlations between observables. Section IV A proposes a novel correlation structure between $E/N(n)$ and $E/A(n)$, prescribing how they combine to yield the symmetry energy, whereas Sec. IV B discusses how the energy per particle is correlated with its derivatives and related quantities.

A. Nuclear symmetry energy

The nuclear-matter EOS at zero temperature as a function of the total nucleon density $n = n_n + n_p$ and isospin asymmetry $\beta = (n_n - n_p)/n$ can be expanded about SNM ($\beta = 0$),

$$
\frac{E}{A}(n, \beta) \approx \frac{E}{A}(n, \beta = 0) + \beta^2 S_2(n),
$$

(13)

with the neutron (proton) density given by n_n (n_p). Microscopic calculations based on χEFT NN and 3N interactions up to $\lesssim n_0$ have shown that the standard (quadratic) expansion (13) works reasonably well [9,51] (see also Refs. [53,54]). The density-dependent nuclear symmetry energy $S_2(n)$ is then given by the difference,

$$
S_2(n) \approx \frac{E}{N}(n) - \frac{E}{A}(n).
$$

(14)

That is, $S_2(n)$ is determined by the energy per particle in PNM ($\beta = 1$) and SNM. Hence, treating the truncation error of $S_2(n)$ completely uncorrelated with the errors of $E/N(n)$ and $E/A(n)$ is questionable. But how can the correlations between the convergence pattern of $E/N(n)$ and $E/A(n)$ be estimated and incorporated into our UQ model?

The contribution to the energy per particle from each χEFT order is dictated by observable coefficients $c_n(k_F)$. We take a first step towards quantifying the extent to which the $c_n(k_F)$ in PNM are correlated with those of SNM. This enables a diagnosis of the correlation between their EFT truncation errors. To this end, we calculate the empirical Pearson correlation coefficient ρ between the coefficients evaluated at the same density, assuming they have mean zero. For the $\Lambda = 500$ MeV interactions, we find a *very strong* [55,56] correlation $\rho \approx 0.94$ that is visualized in Fig. 10(a). For the $\Lambda = 450$ MeV interactions, there exists a *strong* [55,56] correlation $\rho \approx 0.75$; see Fig. 10(c). Note that ρ is not an observable, since it is the correlation of the E/A and E/N EFT coefficients at the EFT orders which have been computed. As such it depends
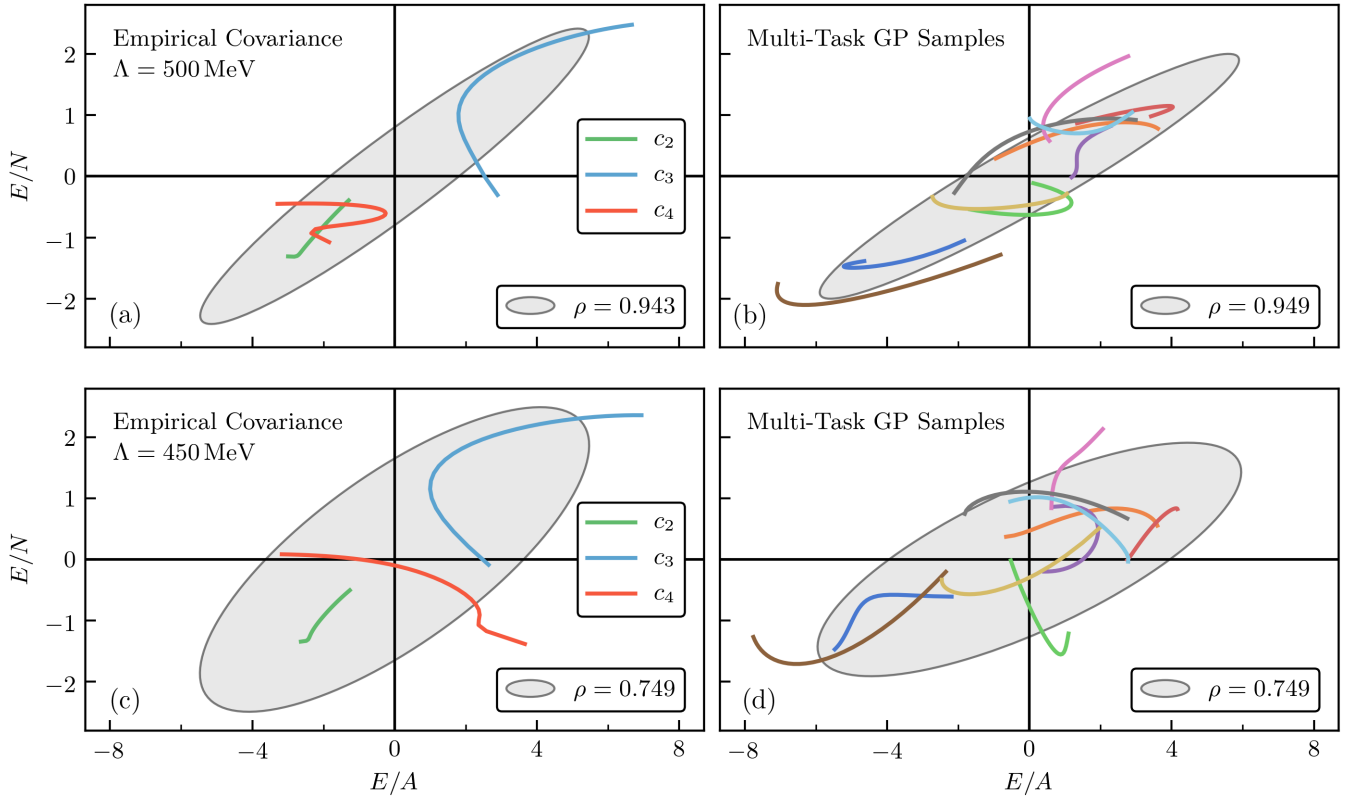
FIG. 10. Correlations between the observable coefficients $c_n(k_F)$ from $E/N(n)$ and $E/A(n)$. The points on each coefficient curve are at the same density. (a) The coefficients from the $\Lambda = 500$ MeV interactions. (b) Toy coefficients created using a correlated multioutput GP trained on the coefficients shown in (a). Each dimension of the multi-output GP ellipse takes into account the associated $\bar{c}$ from Figs. 2 and 3, which need not be the same as the empirical variances of the ellipse in (a). For the $\Lambda = 450$ MeV interactions, the corresponding plots are shown in (c) and (d).

on the choice of y_{ref} and Q. ρ will also be a function of the EFT cutoff, since RG running shuffles contributions between different orders.

But we cannot simply propose an arbitrary correlation structure between $E/N(n)$ and $E/A(n')$ at any given n and n' because the covariance must remain positive semidefinite. As described in Secs. II and III, the truncation errors of each EOS are treated as an output from a GP. Correlations between discrete outputs of various GPs have been addressed in the literature by *multitask* GPs, also known as cokriging [30,31].

Since we use the RBF kernel (6) for each individual observable, the correlations between $E/N(n)$ and $E/A(n')$ can be directly modeled. Furthermore, if the correlation lengths in PNM and SNM are permitted to differ between the observables, we can model this dependence using an RBF cross covariance kernel with a length scale and correlation coefficient ρ that are determined by each individual correlation length [31]; see Eq. (C16). We choose to employ this model for the $\Lambda = 500$ MeV interactions, where the predicted ρ accurately matches the empirical correlation. Appendix C 2 describes this multitask model and also proposes an alternative where ρ can be tuned to the data. This modified model is applied to the $\Lambda = 450$ MeV interactions, since it can better fit the empirical correlation in that case.

We make use of the scale independence of the GP input in k_F, as discussed in Sec. III B, so that we can define a

space where $E/N(n)$ and $E/A(n)$ at the same k_F are also at the same density. Concretely, we simply use k_F^{PNM} as the input space for $E/A(n)$. This only involves scaling the $c_n(k_F)$ length scale to leave the RBF kernel invariant, but the Q and y_{ref} defined in Sec. III B for $E/A(n)$ still use k_F^{SNM}, and hence remain the same. This is a necessary step in correlating $E/N(n)$ with $E/A(n')$ such that points with $n = n'$ are the most highly correlated, while the correlations drop as distance $|n - n'|$ grows.

We take the point estimates of $\bar{c}_i$ and ℓ_i in PNM and SNM, as determined in Sec. III B, and use them to construct the total covariance matrix within and between PNM and SNM. Remarkably, the predicted correlation coefficient for the $\Lambda = 500$ MeV interactions is $\rho = 0.949$, which agrees with the empirical value to two digits.[3] We can then emulate $c_n(k_F)$ from our correlated PNM–SNM system as shown in Fig. 10(b). The emulated coefficients appear quite similar to the actual $c_n(k_F)$, in that we could not tell them apart if they were not already distinguished. This gives us confi-

_____

[3]This should likely not be taken as anything other than coincidence, as ρ was not tuned to the empirical correlation. Rather, ρ is a prediction based on the individual length scales of each marginal process. Tuning of ρ can be done if the length scales of PNM and SNM are the same; see Appendix C 1.
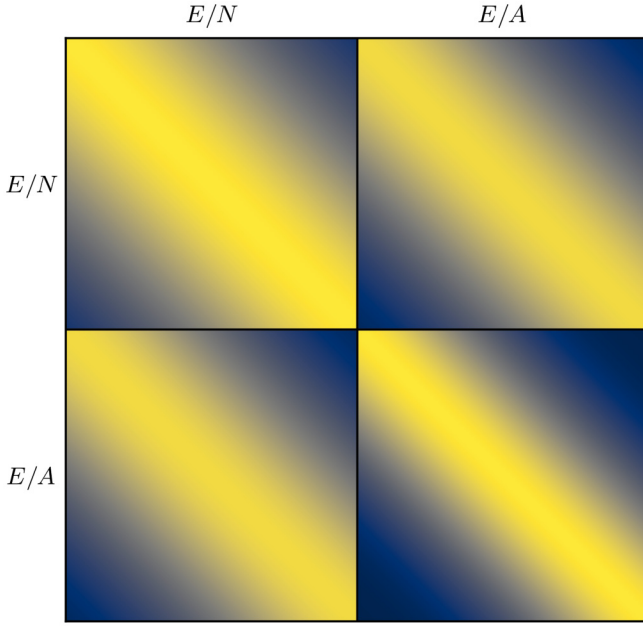
FIG. 11. Total correlation matrix of $E/N(n)$ and $E/A(n)$ assuming a multitask GP model that was trained to order-by-order results of the $\Lambda = 500$ MeV interactions. Each submatrix uses the same grid spaced linearly in k_F. The diagonal blocks show the autocorrelation, and the off-diagonal block is known as the cross correlation. All use RBF kernels [see Eqs. (6), (C12), and (C16)]. The length scale of the $E/A(n)$ blocks has been transformed to k_F^{PNM} as discussed in the text. Hence, points of equal density between $E/N(n)$ and $E/A(n)$ lie on the diagonal band of the off-diagonal blocks, making them the most highly correlated, but the $E/A(n)$ autocorrelation is unchanged. The length scales were determined by fitting to the $E/N(n)$ and $E/A(n)$ coefficients independently as shown in Sec. III B. The correlation $\rho = 0.95$ is a prediction (which agrees with the empirical correlation) given these length scales.
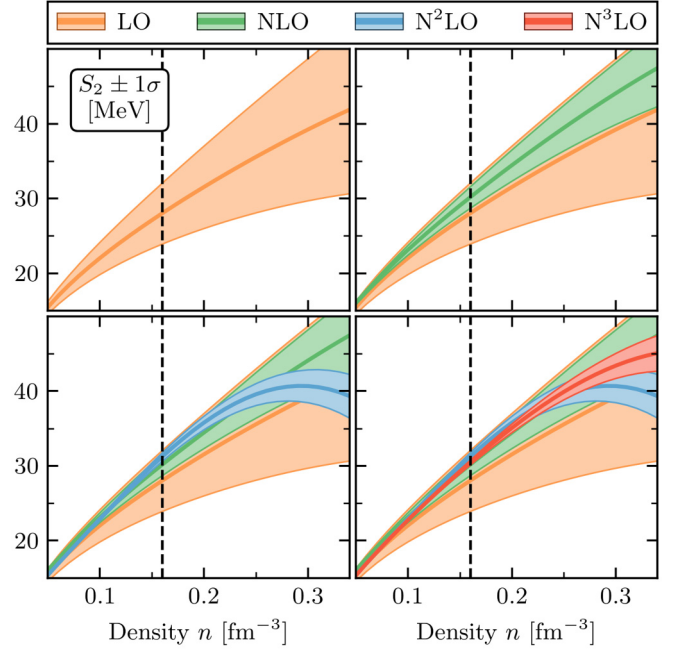

FIG. 12. Similar to Fig. 6, but these are the order-by-order predictions of the symmetry energy $S_2(n)$ for the $\Lambda = 500$ MeV interactions.

density dependence,

$$L(n) = 3n \frac{d}{dn} S_2(n), \quad (15)$$

as discussed in the companion paper [24]. Computing $L(n)$ with full uncertainty propagation requires a discussion of how to take derivatives of GPs, the topic of the next subsection and Appendix C 1.

B. Derivatives and related observables

An important feature of GP interpolants is that they allow straightforward computation of derivatives that are smooth and have theoretical uncertainties that are fully propagated from their antiderivatives. Finite differencing and parametric fits do not achieve this. In the following, we discuss how to evaluate first- and second-order derivatives in SNM using GPs and refer to the companion publication [24] for a detailed discussion of PNM. Appendix C 1 gives details on how to compute the GP for derivative quantities.

Specifically, we consider here the pressure,

$$P(n) = n^2 \frac{d}{dn} \frac{E}{A}(n), \quad (16)$$

and the incompressibility,

$$K = 9n_0^2 \frac{d^2}{dn^2} \frac{E}{A}(n)\bigg|_{n=n_0}. \quad (17)$$

Notice that K is evaluated at the predicted (not at the empirical) saturation density, n_0. We thus need to account for the uncertainties in n_0 in addition to the ones in $E/A(n)$ and its derivatives. The χEFT truncation errors in $E/A(n)$ lead to a distribution for n_0. In Sec. III C we derived the posterior

dence in our approach. The total correlation matrix describing correlations within and between PNM and SNM is given in Fig. 11.

The correlation between PNM and SNM is somewhat weaker (but still strong) for the $\Lambda = 450$ MeV interactions; see Fig. 10(c). This cross correlation is not well predicted by Eq. (C16), so we present a different approach to modeling ρ in this case. One can tune ρ, which we take to be equal to the empirical correlation $\rho = 0.75$, so long as we make the reasonable approximation $\ell_{PNM} \approx \ell_{SNM}$ (see Appendix C 2 for details). The samples from this tuned multitask GP look similar to the actual $c_n(k_F)$; see Fig. 10(d). This shows that our multitask kernels are a flexible way to introduce correlations between observables.

With the full correlation structure of PNM and SNM in hand, we can compute $S_2(n)$ with truncation errors. The results for the $\Lambda = 500$ MeV interactions are shown in Fig. 12. If correlations between PNM and SNM had instead been neglected, i.e., truncation errors simply added in quadrature, the size of the truncation uncertainty would be >5 times larger (≈ 2 times larger for $\Lambda = 450$ MeV). This factor is particularly important for constraining $S_2(n)$ and its (rescaled)
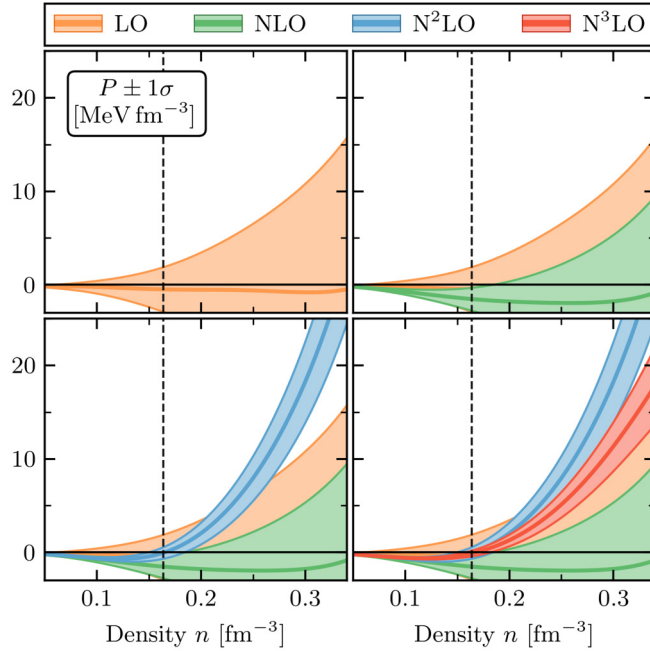
FIG. 13. Order-by-order predictions of the pressure $P(n)$ of SNM, including differentiation and truncation uncertainty, for the $\Lambda = 500$ MeV interactions. See the main text for details.

$\mathrm{pr}(n_0 \,|\, \mathcal{D})$, which is approximately Gaussian. Hence, we can estimate K and its full uncertainty at any given n_0 by computing $\mathrm{pr}(K \,|\, \mathcal{D}, n_0)$, and subsequently sum over all plausible n_0 values via

$$\mathrm{pr}(K \,|\, \mathcal{D}) = \int \mathrm{pr}(K \,|\, \mathcal{D}, n_0) \mathrm{pr}(n_0 \,|\, \mathcal{D}) dn_0. \qquad (18)$$

We perform a similar summation in Ref. [24] to compute the posterior for the symmetry energy $[S_v = S_2(n_0)]$ and its slope parameter $[L = L(n_0)]$ evaluated at the predicted range for n_0.

GPs have advantages over parametrizations of the EOS (e.g., series expansions) when derivatives up to high densities are desired. In particular, second-order (and higher) derivatives tend to magnify numerical instabilities in these parametrizations. We have verified this instability by performing a global Bayesian fit of a power series in k_F [51],

$$\frac{E}{A}\big(n; \{d_\nu\}_{\nu=2}^M\big) = \sum_{\nu=2}^{M} d_\nu \left(\frac{n}{n_0}\right)^{\frac{\nu}{3}}, \qquad (19)$$

to $E/A(n)$, with $M \geqslant 2$ assigned by maximizing the Bayesian evidence in an attempt to prevent overfitting. Despite using the evidence as a safeguard, we obtained fit coefficients d_ν that are unnaturally large in magnitude and that alternated in sign, a classic symptom of overfitting. (Similar symptoms are seen in the fit coefficients presented in Table II of Ref. [51].) Consequently, our parametric fits of Eq. (19) were unable to predict, e.g., K reliably. In contrast, GPs are like splines [44] in being more sensitive to local information than a global parametric fit like Eq. (19). Our GP model therefore yields more reliable derivatives. Its locality does mean, though, that there are possible edge effects when computing derivatives—
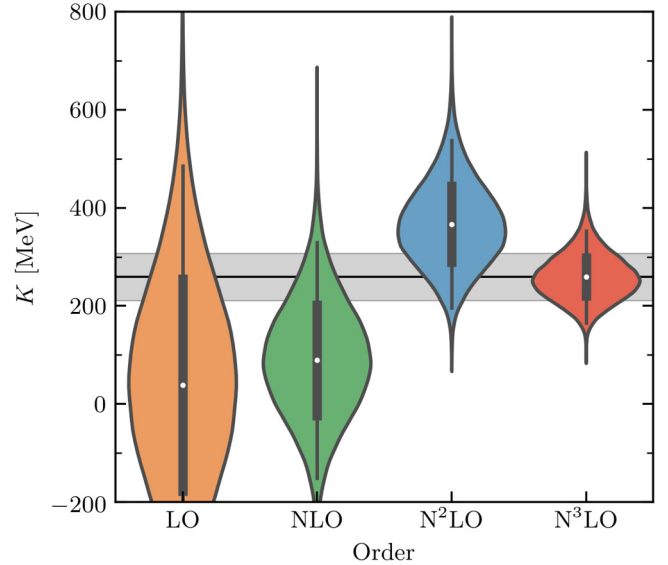

FIG. 14. Violin plots of the incompressibility K of SNM, shown order by order for the $\Lambda = 500$ MeV interactions. Curves show the entire smoothed posterior (and its reflection). Each posterior includes differentiation and truncation uncertainty, and are marginalized over all plausible saturation densities $n_0 = 0.17 \pm 0.01$ fm^{-3}; see the main text. Dots and bars indicate the mean value, along with the 1σ and 2σ uncertainties. The black line and gray band extends the N^{3}LO mean and 1σ uncertainty to more easily compare χEFT orders.

especially higher derivatives—near the edge of the region where there is data.

Honest UQ for these observables is only possible with a correlated model of uncertainty. If correlations are neglected, as in the "standard EFT" error prescription [57,58], then derivatives can be arbitrarily uncertain and thus unrealistic.

Figure 13 shows the pressure of SNM with 68% credible intervals. Although LO and NLO have negative mean values across all shown densities, nuclear saturation (i.e., $P = 0$) near n_0 could be achieved within the large uncertainties, even for these NN-only interactions. Indeed, the wide range of densities at which the pressure *could* cross zero at LO and NLO suggests that the nuclear saturation point is somewhat fine tuned. Perhaps this is not surprising: in EFT, if an observable that is not zero at LO has a zero crossing, the position of that crossing is, by definition, sensitive to higher-order corrections because lower orders must cancel there. From a χEFT consistency point of view it is reassuring that N^{2}LO and N^{3}LO are consistently within the bands of the previous orders at $\lesssim n_0$. They do, however, begin to diverge at higher densities. As with $E/N(n)$ and $E/A(n)$ we stress that the consistency of the uncertainty bands is difficult to gauge due to the long correlation length of the truncation error, and hence small effective sample size of data.

Figure 14 shows our order-by-order results for K based on the $\Lambda = 500$ MeV interactions. These predictions use our best estimate for n_0 in the integral (18): the Gaussian posterior $n_0 = 0.17 \pm 0.01$ fm^{-3} at N^{3}LO determined in Sec. III C. At LO and NLO, where the empirical saturation point is typically

not well reproduced, this choice leads to wide-spread distributions whose 1σ regions reach $K < 0$, even though nuclear saturation requires $K > 0$. In general, the uncertainty bands are consistent across χEFT orders and settle at 260 ± 54 MeV (292 ± 54 MeV) for the N^{3}LO interaction with $\Lambda = 500$ MeV (450 MeV).

V. SUMMARY AND OUTLOOK

Order-by-order predictions of a well-behaved EFT should converge regularly towards the all-orders value. We have formalized this idea into a falsifiable EFT convergence model using Bayesian statistics and presented the first application to infinite matter up to $2n_0$. The EOS is based on order-by-order calculations in MBPT with NN and 3N interactions up to N^{3}LO [18,24,47]. While this work focuses on key properties of SNM, our companion publication [24] is dedicated to PNM and its astrophysical applications. Together, they set a new standard [50] for UQ in infinite matter calculations.

Section III provides the first truncation error bands for infinite matter that account for correlations in density. Our findings indicate that the truncation errors are highly correlated, rendering the qualitative judgment of credible intervals more difficult. A full understanding, therefore, requires the diagnostic tools discussed in Ref. [25]. Specifically, we have verified the importance of truncation error correlations between

(i) the EOS at different densities,
(ii) different observables, such as $E/N(n)$ and $E/A(n)$ to determine $S_2(n)$, and
(iii) the EOS and its derivatives.

Our truncation error model also allows us to do the first efficient and accurate propagation of EOS theoretical uncertainties (including χEFT truncation errors) to derived quantities. We have pointed out the advantages of our approach over global parametrizations of the EOS such as fitted series expansions: even maximizing the Bayesian evidence could not prevent overfitting of the series expansion (19). Such numerical instabilities are magnified when computing derivatives: especially second and higher derivatives. Our nonparametric GP approach does not suffer from these instabilities.

We then studied nuclear saturation properties, including the incompressibility and pressure of SNM, and symmetry energy, with theoretical uncertainties fully quantified. We have also provided the first probability distributions for the predicted saturation region by sampling from our correlated error model. Table II summarizes our constraints for the two momentum cutoffs at the 1σ level. The results all agree at the level of 2σ, and for the saturation properties the agreement is better than that. This indicates only a mild cutoff dependence for these observables. Our findings indicate that taking into account the correlations in density is necessary to quantify these properties. The methods developed here are publicly available as annotated Jupyter notebooks. [50].

In Sec. I we raised several points about the in-medium convergence of χEFT. Here are some of our conclusions:

TABLE II. Our 1σ-level constraints on the density (in fm^{-3}), energy per particle, and incompressibility of SNM at saturation as well as the symmetry energy and its derivative (all in MeV). These are given for two different N^{3}LO χEFT Hamiltonians. We emphasize that n_0 and $\frac{E}{A}(n_0)$ as well as $S_2(n_0)$ and $L(n_0)$ are correlated (i.e., the covariance matrix is not diagonal). We also remind the reader that the uncertainties quoted here are solely due to truncation of the χEFT expansion: they do not account for the uncertainty in the χEFT LECs. A full assessment of the uncertainties in χEFT's predictions for nuclear matter will require more work on χEFT NN and 3N interactions up to N^{3}LO, as discussed in the main text.

	$\Lambda = 450$ MeV	$\Lambda = 500$ MeV	See also
n_0	0.170(16)	0.173(14)	Fig. 9[a]
$\frac{E}{A}(n_0)$	$-14.3(1.0)$	$-14.9(1.1)$	Fig. 9[a]
K	260.0(54.0)	292.0(54.0)	Fig. 14
$S_2(n_0)$	33.5(1.3)	31.7(1.1)	Fig. 12[b]
$L(n_0)$	67.8(4.0)	59.8(4.1)	Ref. [24][b]

[a]Mean and covariance matrix are given in Eqs. (11) and (12).
[b]Figure 2 of Ref. [24] shows the constraints in the $S_2(n_0)$-$L(n_0)$ plane for the $\Lambda = 500$ MeV interaction; for the $\Lambda = 450$ MeV interaction see the Supplemental Material of Ref. [24].

(1) The convergence plots for PNM and SNM show regular convergence with increasing order, as seen for NN observables.
(2) The statistical model formulated with $Q(k_F) = k_F/\Lambda_b$ and $y_{ref}(k_F) = 16$ MeV $\times (k_F/k_{F0})^2$ provides a reasonable characterization of the convergence pattern in infinite matter. This yields consistent uncertainty bands for both $E/N(n)$ and $E/A(n)$, within statistical fluctuations.
(3) With our model checking diagnostics given in Appendix A, we have found evidence that 3N interactions show somewhat different characteristics (e.g., different length scales), and the N^{2}LO coefficient $c_3(k_F)$ may be an outlier due to the first nonvanishing 3N contributions at this order. This points to the possibility that one should not use $c_3(k_F)$ to infer truncation error properties, though more work along these lines with different interactions is needed.
(4) The posterior for Λ_b calculated using $Q(k_F) = k_F/\Lambda_b$ is consistent with determinations from NN observables.

Calculations of finite nuclei with χEFT potentials have often been found to predict too-small radii [59–61]. Since the link between infinite matter, heavy to medium-mass nuclei, and few-body systems has yet to be fully understood [52,62], examining the statistical correlation of radius systematics with the predicted saturation density in SNM could clarify the origin of these deficiencies. The methods developed here could facilitate the statistically consistent inclusion of empirical saturation properties in fits of χEFT potentials (cf. Refs. [18,63]).

Future studies will extend our analysis to asymmetric matter with arbitrary proton fractions (e.g., neutron-star matter) and finite temperature. In this case, the discrete correlations between $E/N(n)$ and $E/A(n)$ found here could be naturally

handled by $c_n(k_{\rm F})$ that are correlated in both density and proton fraction. To elucidate the full dependence of the EOS on the nuclear interactions, however, improved order-by-order NN and 3N interactions need to be developed up to N^{3}LO [52,61,62]. Our physically motivated GP model can also be applied to efficiently compute nuclear saturation properties using Bayesian optimization frameworks [64,65] and to account for uncertainties in the fits of the LECs using Monte Carlo sampling [36,46,66]. This last task is particularly important. It will presumably expand the error bars presented here, which only account for truncation error, and not for uncertainties in the LECs.

APPENDIX A: MODEL-CHECKING DIAGNOSTICS

In this appendix we provide more details on our model-checking diagnostics.

There are three underlying assumptions in our statistical model for EFT truncation errors that we seek to validate [25]. First, the coefficients $c_n(x)$ at each order are well characterized as independent draws (i.i.d. realizations) from a single GP. Second, the parametrized mean and covariance functions that characterize this GP have been correctly estimated (i.e., we have found a consistent mean, variance, and correlation length). Third, the GP learned from the known coefficients predicts a statistically meaningful distribution for the truncation error.

For this validation we adopt here the preferred menu of model-checking diagnostics advocated in Ref. [25]. These are

  (i) Distribution of Mahalanobis distance (MD) for the order-by-order coefficients. To check if pointwise data from a coefficient $\{c_n(x_i)\}_i$ followed an uncorrelated normal distribution, we would calculate the sum of squares of the scaled residuals and compare

to the χ^2 distribution with the appropriate number of degrees of freedom. The generalization for a multivariate correlated normal distribution—that is, when $c_n(x)$ is correlated in x, as in our case—is to calculate the MD for the extracted $c_n(k_{\rm F})$ at a specific order n at M validation $k_{\rm F}$'s and compare to a reference distribution. For a GP that is a χ^2 distribution with M degrees of freedom.

 (ii) Pivoted Cholesky (PC) decomposition of the MD plotted graphically against the index and compared to a standard normal distribution. This provides specific information about misestimated variance (too large or small values across all indices) or correlation structure (failing distribution at large index).

(iii) Credible interval diagnostic (CID). A plot of the CID for truncation error shows whether a $100\alpha\%$ credible interval learned up to a given order contains approximately $100\alpha\%$ of a set of validation points representing the next order result. This test can be carried out at $k - 1$ orders, where k is the number of χEFT coefficients in hand, since it requires knowledge of the result one order beyond that at which validation is being carried out.

Examples and associated Python code for carrying out these diagnostics are given in Refs. [25,67].

The observable coefficients for the $\Lambda = 450$ MeV interactions of Table I are shown in Fig. 15. The two panels are the analog of those shown for the $\Lambda = 500$ MeV interactions in Figs. 2 and 3. The MD and PC diagnostics are applied to the $\Lambda = 450$ MeV and $\Lambda = 500$ MeV observable coefficients in Figs. 16 and 17 respectively. The reference scale (9) and expansion parameter (10) are used in all of these figures.

The CID for $\Lambda = 450$ MeV is very similar to $\Lambda = 500$ MeV, which was shown in Fig. 8. The MD diagnostics for the two PNM cases are also similar. Both show that the $c_3(k_{\rm F})$ coefficient may be an outlier. This supports the qualitative observation, apparent in Fig. 2 and the right-hand panel of Fig. 15, and discussed in Sec. III B, that $c_3(k_{\rm F})$ has a different shape to $c_0(k_{\rm F})$ and $c_2(k_{\rm F})$. We reiterate that this presumably happens because of the 3N contributions that enter χEFT at N^{2}LO. The PC diagnostic, while behaving well at small index, shows a decreased range of points at the highest index for both PNM and SNM. This happens because of the size of the white noise term $\sigma^2 = 5 \times 10^{-4}$ that was used for numerical stability (see Sec. III B). The MD diagnostics for the two SNM cases show larger discrepancies with the reference distributions, with no points lying within the 50% credible intervals.

As pointed out in Sec. III B, this problem reflects a mismatch of the assumed NN-only and 3N correlation structures. To explore this further, we consider an alternative χEFT convergence model. In particular, we split the coefficients into NN-only and residual 3N coefficients, $c_n^{(2)}(k_{\rm F})$ and $c_n^{(3)}(k_{\rm F})$, and assign different $k_{\rm F}$ dependencies to the $y_{\rm ref}(k_{\rm F})$ associated with each. In this variant of the truncation-error model we use a constant $y_{\rm ref}(k_{\rm F}) = 16$ MeV for the $c_n^{(2)}(k_{\rm F})$ coefficients,
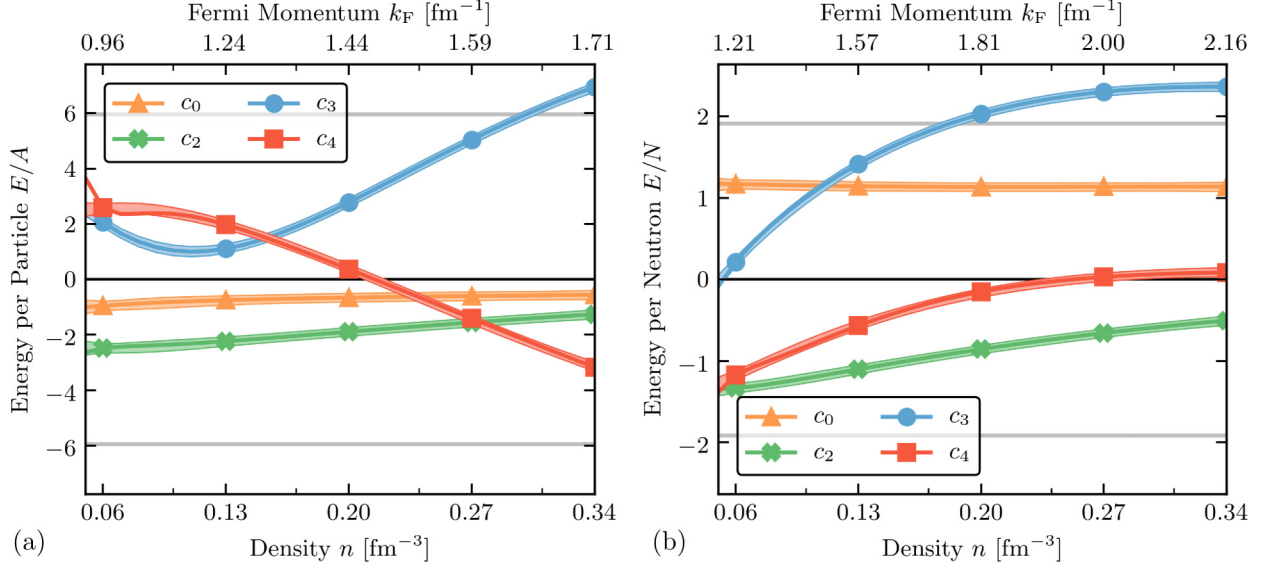
FIG. 15. Observable coefficients for (a) $E/A(n)$ (SNM) and (b) $E/N(n)$ (PNM) up to N³LO using the $\Lambda = 450$ MeV interactions in Table I. Markers indicate training points, gray bands indicate $2\bar{c}$ and colored bands are 68% credible intervals of the interpolating GPs. The estimated hyperparameters are given by $\bar{c} = 3.0$ and $\ell = 0.50$ fm^{-1} for SNM and $\bar{c} = 0.96$ and $\ell = 0.81$ fm^{-1} for PNM.

while for the $c_n^{(3)}(k_F)$ coefficients we use

$$y_{\text{ref}}(k_F) = 16 \text{ MeV} \times \left(\frac{k_F}{k_{F,0}} \right)^3. \qquad (A1)$$

Here, as in Eq. (9), $k_{F,0}$ is the Fermi momentum associated with $n_0 = 0.16$ fm^{-3}, namely $k_{F,0}^{\text{PNM}} = 1.680$ fm^{-1} and $k_{F,0}^{\text{SNM}} = 1.333$ fm^{-1}. This form is chosen to roughly capture the extra k_F dependence of 3N contributions relative to leading NN contributions.

The observable coefficients for $E/N(n)$ (PNM) for the alternative model are given in Fig. 18 along with the MD and PC diagnostics. Comparing to Fig. 2, we see that the observable coefficients have less variability with the alternative reference scale. This is verified by the diagnostic plots, which show reasonable distributions for MD and the PC at each index, with the $c_3^{(3)}(k_F)$ coefficient being less of an outlier than the combined $c_3(k_F)$ coefficient in Fig. 17. The order-by-order credible intervals for $E/N(n)$ obtained with this alternative truncation-error model are not shown, but are very close
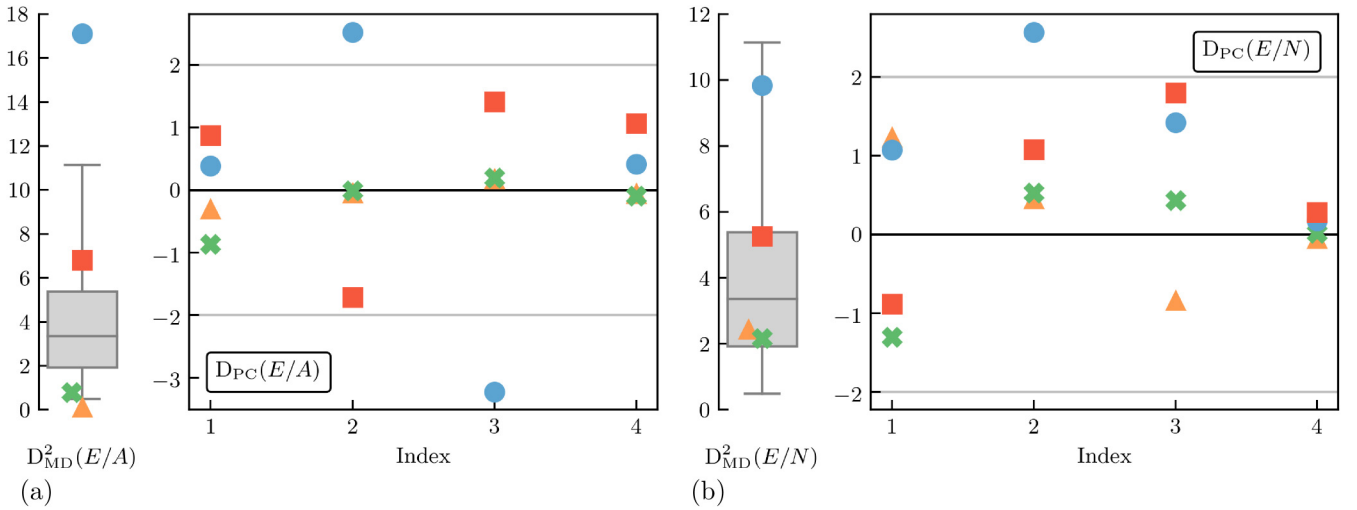

FIG. 16. Model checking diagnostics applied to (a) SNM and (b) PNM coefficients from the $\Lambda = 450$ MeV interactions in Fig. 15. The MD computed against the underlying process is shown in the left panel of each subplot. The interior line, box end caps, and whiskers on the box plot show the median, 50% credible intervals, and 95% credible intervals, respectively. The right panel shows the PC diagnostic $\mathbf{D}_{\text{PC}}$ vs index, with gray lines that represent its 2σ error bands. Both diagnostics point to the $c_3(k_F)$ coefficient as a possible outlier. See Ref. [25] for more details about analyzing these plots.
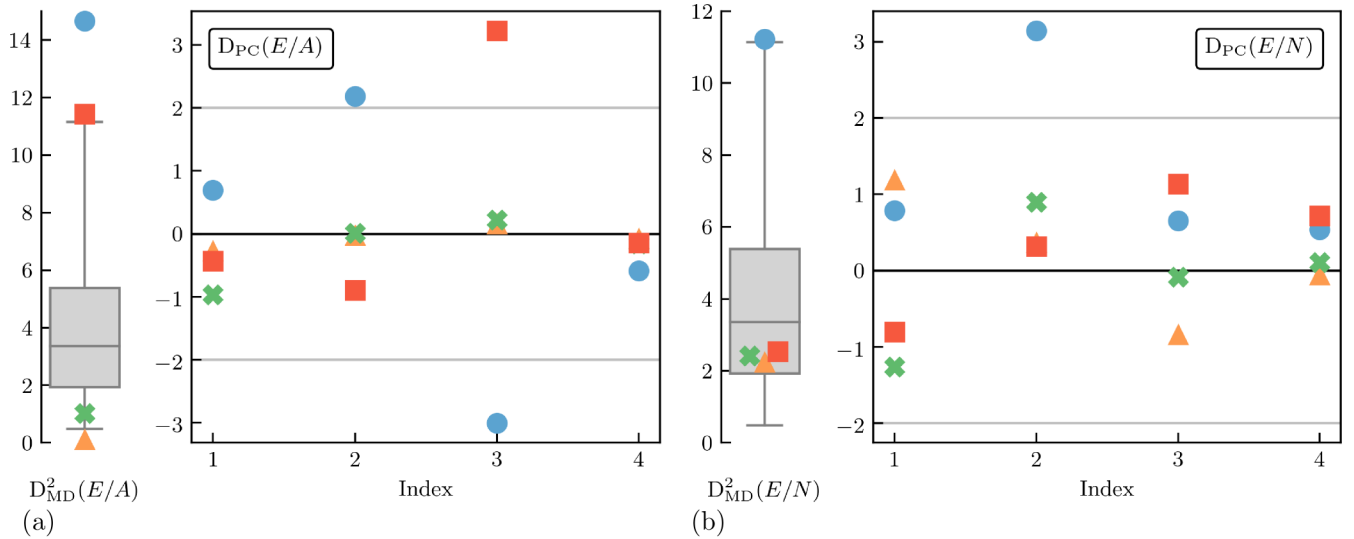
FIG. 17. Similar to Fig. 16 but for the $\Lambda = 500$ MeV interactions in Figs. 2 and 3.

to those from using Eq. (9). The Λ_b posterior (Fig. 19) is compatible with that shown in the main text (Fig. 4) although the maximum *a posteriori* (MAP) value is somewhat larger here. The posterior for ℓ, Fig. 20, has a smaller MAP value than the one shown in Fig. 5, but is consistent with Fig. 18(a).

The observable coefficients and diagnostics for $E/A(n)$ (SNM) for the alternative model are given in Fig. 21. Here we see the $c_3^{(3)}(k_F)$ coefficient has become even more of an outlier than it was in the approach used in the main text, so the model with split $y_{ref}(k_F)$'s has not succeeded.

We invite the reader to take advantage of the freely available Jupyter notebooks [50] to further investigate these issues.

APPENDIX B: TABULATED VALUES FOR THE EOS

Tables III and IV give numerical values for the EOS in the limit of SNM (left-hand side) and PNM (right-hand side) up to N³LO. The NN and 3N interactions in Table I are used. See the captions for more details.

APPENDIX C: MULTITASK GAUSSIAN PROCESSES

In this appendix, we provide more technical details on modeling with *multitask* GPs, with particular emphasis on two cases: (1) a function and its derivatives and (2) multiple generic functions. Multitask GPs, also known as multioutput GPs, are used to model multiple curves $y_i(x)$ simultaneously,
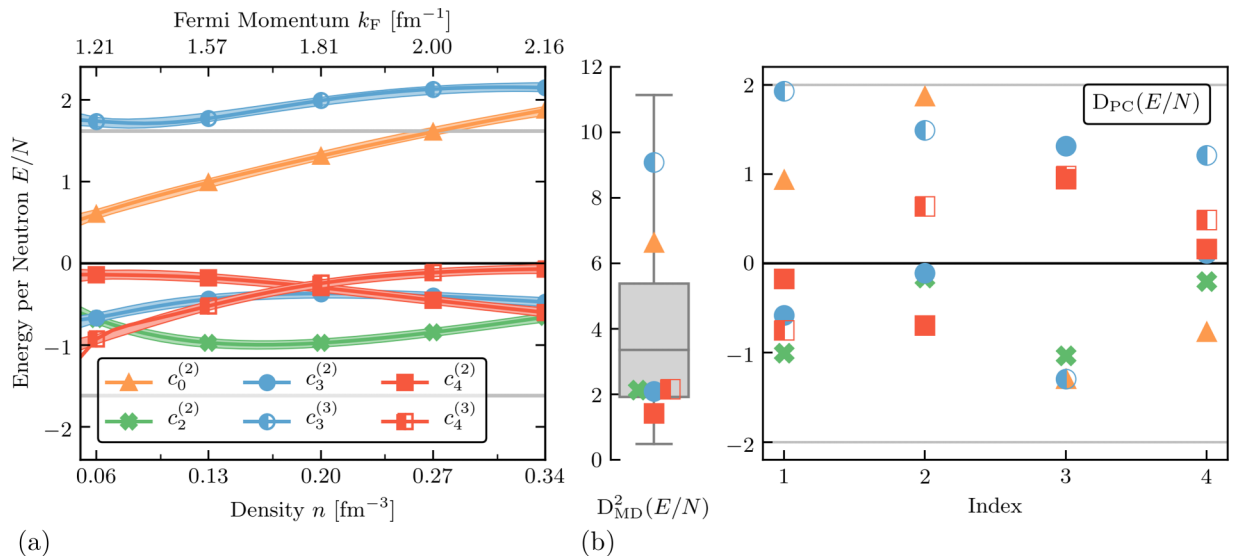

FIG. 18. Observables coefficients and model diagnostics for $E/N(n)$ (PNM) up to N³LO for the $\Lambda = 500$ MeV interactions in Table I using separate reference scales for NN-only (denoted with a superscript "(2)") and 3N contributions (denoted with a superscript "(3)"). The model for $y_{ref}(k_F)$ is explained in the text [see Eq. (A1)].
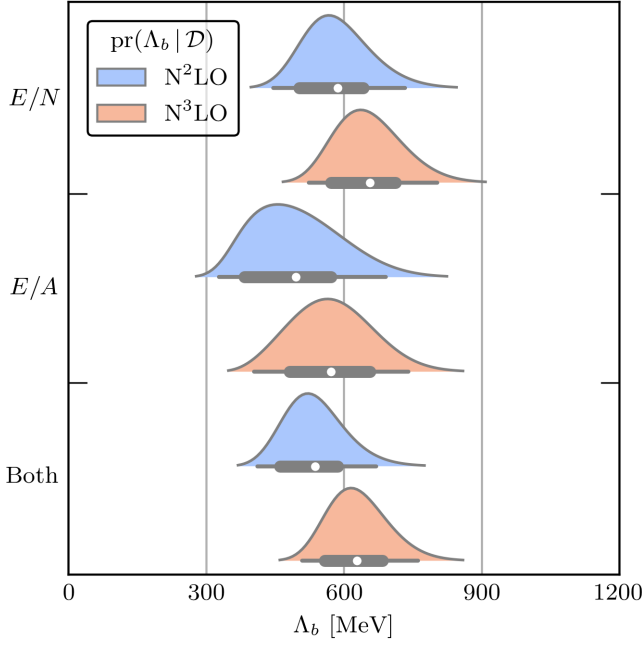
FIG. 19. The posteriors for the EFT breakdown scale Λ_b as in Fig. 4 but using the alternative model for $y_{\mathrm{ref}}(k_{\mathrm{F}})$ as explained in the text [see Eq. (A1)].

FIG. 20. Length-scale posteriors as in Fig. 5 but using the alternative model for $y_{\mathrm{ref}}(k_{\mathrm{F}})$ as explained in the text [see Eq. (A1)].

while possibly learning about their interdependencies to improve predictions.

1. A function and its derivatives

The derivative is a linear operator, so a Gaussian random variable remains closed under this operation. Assume that $f(x)$ is distributed as $\mathcal{GP}[m(x), \kappa(x, x')]$. The joint distribution of a function $f(x)$ and its derivative $\partial_x f(x)$

is then

$$\begin{bmatrix} f(x) \\ \partial_x f(x) \end{bmatrix} \sim \mathcal{GP}[m^\nabla(x), \, \kappa^\nabla(x, x')], \tag{C1}$$

$$\text{with} \quad m^\nabla(x) = \begin{bmatrix} m(x) \\ \partial_x m(x) \end{bmatrix}, \quad \text{and} \tag{C2}$$

$$\kappa^\nabla(x, x') = \begin{bmatrix} \kappa(x, x') & \partial_{x'}^\mathsf{T} \kappa(x, x') \\ \partial_x \kappa(x, x') & \partial_x \partial_{x'}^\mathsf{T} \kappa(x, x') \end{bmatrix}, \tag{C3}$$

where ∂_x is a d-dimensional vector if $x \in \mathbb{R}^d$, making $m^\nabla(x)$ and $\kappa^\nabla(x, x')$ then $(d+1)$- and $(d+1) \times (d+1)$-dimensional, respectively. For example, if $f(x)$ has a prior mean of 0 and has been estimated by fitting to a set of training points $(\mathbf{x}, \mathbf{y})$, then the conditional mean and variance are

FIG. 21. Same as Fig. 18 but for $E/A(n)$ (SNM).

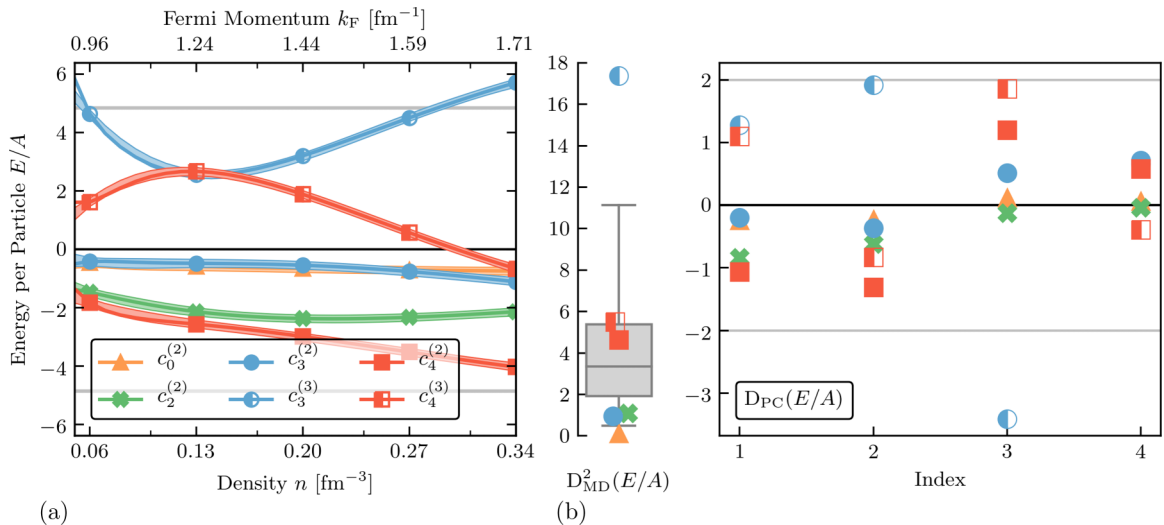TABLE III. Energy per particle in MeV for SNM (left-hand side) and PNM (right-hand side) at four orders in the χ EFT expansion for the interactions with $\Lambda = 450$ MeV in Table I. The density n is given in units of fm^{-3} and the Fermi momentum k_F in fm^{-1}. Notice that k_F^{SNM} and k_F^{PNM} at same density are different.

			Symmetric nuclear matter (SNM)		
n	k_F^{SNM}	LO	NLO	N^{2}LO	N^{3}LO
0.05	0.90	−7.40	−9.12	−8.60	−8.41
0.06	0.96	−7.93	−9.99	−9.46	−9.22
0.07	1.01	−8.39	−10.87	−10.30	−10.01
0.08	1.06	−8.79	−11.71	−11.13	−10.78
0.09	1.10	−9.16	−12.55	−11.94	−11.50
0.10	1.14	−9.49	−13.36	−12.70	−12.20
0.11	1.18	−9.79	−14.14	−13.41	−12.80
0.12	1.21	−10.12	−14.90	−14.04	−13.36
0.13	1.24	−10.43	−15.62	−14.58	−13.82
0.14	1.27	−10.71	−16.34	−14.98	−14.19
0.15	1.30	−10.99	−17.01	−15.29	−14.47
0.16	1.33	−11.27	−17.69	−15.46	−14.66
0.17	1.36	−11.50	−18.32	−15.51	−14.72
0.18	1.39	−11.79	−18.92	−15.39	−14.65
0.19	1.41	−11.99	−19.51	−15.13	−14.50
0.20	1.44	−12.29	−20.08	−14.69	−14.23
0.21	1.46	−12.48	−20.61	−14.14	−13.83

			Pure neutron matter (PNM)		
n	k_F^{PNM}	LO	NLO	N^{2}LO	N^{3}LO
0.05	1.14	8.64	7.25	7.22	7.02
0.06	1.21	9.72	7.96	8.07	7.81
0.07	1.27	10.73	8.61	8.91	8.60
0.08	1.33	11.69	9.21	9.76	9.40
0.09	1.39	12.60	9.77	10.64	10.22
0.10	1.44	13.48	10.32	11.56	11.09
0.11	1.48	14.32	10.84	12.51	12.01
0.12	1.53	15.14	11.35	13.51	12.97
0.13	1.57	15.94	11.86	14.56	14.01
0.14	1.61	16.71	12.36	15.66	15.10
0.15	1.64	17.46	12.86	16.81	16.26
0.16	1.68	18.20	13.36	18.02	17.48
0.17	1.71	18.93	13.85	19.27	18.75
0.18	1.75	19.62	14.35	20.56	20.10
0.19	1.78	20.32	14.86	21.90	21.49
0.20	1.81	20.99	15.36	23.27	22.94
0.21	1.84	21.67	15.87	24.70	24.41

given by

$$\tilde{m}(x) = \kappa(x, \mathbf{x})K^{-1}\mathbf{y}, \tag{C4}$$

$$\tilde{\kappa}(x, x') = \kappa(x, x') - \kappa(x, \mathbf{x})K^{-1}\kappa(\mathbf{x}, x'), \tag{C5}$$

where $K = \kappa(\mathbf{x}, \mathbf{x})$ (see also Refs. [25,43]). Then the distribution of $f(x)$ with any of its derivatives would involve differentiating $\tilde{m}$ and $\tilde{\kappa}$, which includes at least two derivatives of the kernel κ. The generalization to higher derivatives follows straightforwardly.

For the squared exponential kernel (RBF) employed here, an analytic expression for an arbitrary number of derivatives

exists. Consider the nth order (scalar) mixed partial derivative,

$$\partial_{n_1, n_2, \ldots, n_d}^n \equiv \frac{\partial^n}{\partial x_1^{n_1} \partial x_2^{n_2} \cdots \partial x_d^{n_d}}, \tag{C6}$$

with $\partial_{n_1', n_2', \ldots, n_d'}^{n'}$ defined similarly for x'. Stationary kernels like the RBF kernel obey $\partial_{x'}\kappa(x, x') = -\partial_x\kappa(x, x')$, which implies

$$\partial_{n_1, \ldots, n_d}^n \partial_{n_1', \ldots, n_d'}^{n'}\kappa(x, x') = (-1)^{n'} \partial_{N_1, \ldots, N_d}^N \kappa(x, x'), \tag{C7}$$

where $N_i = n_i + n_i'$ and all derivatives act on x. To compute this Nth order derivative, we make use of the (physicists')

TABLE IV. Same as Table III but for the interactions with $\Lambda = 500$ MeV in Table I.

Symmetric nuclear matter (SNM)					
n	k_F^{SNM}	LO	NLO	N^{2}LO	N^{3}LO
0.05	0.90	−6.86	−8.82	−8.26	−8.37
0.06	0.96	−7.34	−9.68	−9.00	−9.20
0.07	1.01	−7.72	−10.52	−9.73	−9.97
0.08	1.06	−8.05	−11.33	−10.43	−10.71
0.09	1.10	−8.34	−12.11	−11.10	−11.40
0.10	1.14	−8.60	−12.89	−11.74	−12.01
0.11	1.18	−8.84	−13.63	−12.32	−12.57
0.12	1.21	−9.06	−14.33	−12.83	−13.05
0.13	1.24	−9.28	−15.00	−13.25	−13.46
0.14	1.27	−9.46	−15.65	−13.60	−13.76
0.15	1.30	−9.67	−16.28	−13.84	−13.97
0.16	1.33	−9.88	−16.89	−13.95	−14.10
0.17	1.36	−10.06	−17.46	−13.96	−14.14
0.18	1.39	−10.24	−17.99	−13.86	−14.04
0.19	1.41	−10.46	−18.52	−13.62	−13.88
0.20	1.44	−10.60	−19.00	−13.24	−13.59
0.21	1.46	−10.74	−19.48	−12.67	−13.23

Pure neutron matter (PNM)					
n	k_F^{PNM}	LO	NLO	N^{2}LO	N^{3}LO
0.05	1.14	8.59	7.23	7.11	6.96
0.06	1.21	9.66	7.92	7.91	7.71
0.07	1.27	10.66	8.55	8.70	8.45
0.08	1.33	11.62	9.14	9.50	9.19
0.09	1.39	12.54	9.69	10.31	9.94
0.10	1.44	13.42	10.22	11.16	10.72
0.11	1.48	14.26	10.75	12.05	11.53
0.12	1.53	15.09	11.26	12.99	12.39
0.13	1.57	15.88	11.77	13.97	13.30
0.14	1.61	16.66	12.28	15.02	14.27
0.15	1.64	17.41	12.78	16.12	15.29
0.16	1.68	18.15	13.29	17.29	16.38
0.17	1.71	18.88	13.82	18.52	17.53
0.18	1.75	19.58	14.35	19.81	18.74
0.19	1.78	20.27	14.89	21.17	20.01
0.20	1.81	20.97	15.43	22.59	21.37
0.21	1.84	21.63	16.00	24.06	22.78

Hermite polynomial relation

$$H_n(z) = (-1)^n e^{z^2} \frac{d^n}{dz^n} e^{-z^2}. \tag{C8}$$

The squared exponential covariance function is given by

$$\kappa(x, x') = \bar{c}^2 e^{-\frac{1}{2}(x-x')^\intercal L^{-1}(x-x')}, \tag{C9}$$

where we assume a diagonal correlation length matrix $L = \text{diag}(\ell_1^2, \ell_2^2, \ldots, \ell_d^2)$. Now it is useful to make a change of variables $z = L^{-1/2}(x - x')/\sqrt{2}$, from which it follows that

$$\frac{d}{dx_i} = \frac{1}{\sqrt{2}\ell_i} \frac{d}{dz_i}. \tag{C10}$$

This transformation allows the kernel to be separable in z, i.e., $\kappa(x, x') = \bar{c}^2 \prod_{i=1}^d e^{-z_i^2}$. Using Eq. (C8) the desired derivative

follows as

$$\partial_{n_1, n_2, \ldots, n_d}^n \partial_{n_1', n_2', \ldots, n_d'}^{n'} \kappa(x, x')$$

$$= (-1)^{n'} \bar{c}^2 \left[\prod_{i=1}^d \left(\frac{1}{\sqrt{2}\ell_i} \right)^{N_i} \frac{\partial^{N_i}}{\partial z_i^{N_i}} e^{-z_i^2} \right]$$

$$= (-1)^n \left[\prod_{i=1}^d \left(\frac{1}{\sqrt{2}\ell_i} \right)^{N_i} H_{N_i}(z_i) \right] \kappa(x, x'). \tag{C11}$$

Certain observables require the sum of a function with one or more of its derivatives. The distribution of such a sum follows straightforwardly as the sum of correlated Gaussians.

If X and Y are distributed jointly as

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} K_{XX} & K_{XY} \\ K_{YX} & K_{YY} \end{bmatrix}\right), \quad (\text{C12})$$

then

$$AX + BY \sim \mathcal{N}(\mu, \Sigma), \quad (\text{C13})$$

$$\mu = A\mu_X + B\mu_Y, \text{ and} \quad (\text{C14})$$

$$\Sigma = K_{XX} + K_{YY} + BK_{YX}A^\mathsf{T} + AK_{XY}B^\mathsf{T}. \quad (\text{C15})$$

2. Two generic functions

Equations (C13)–(C15) appear simple enough, and for derivatives the cross-covariance $K_{XY} = K_{YX}^\mathsf{T}$ is given by Eq. (C3). However, the form this cross-covariance takes is less clear in the case of generic *multitask* Gaussian processes. One of the main difficulties with using GPs in the generic multitask setting is coming up with valid covariance functions that accurately model the relationships in the data.

In this work, we are interested in a multitask GP that describes $E/N(n)$ and $E/A(n)$. This means we have two processes, each individually distributed as a GP with an RBF kernel, and whose outputs are correlated with one another. Reference [31] showed that if each autocovariance, K_{XX} and K_{YY}, is generated from RBF kernels $\kappa_1(x, x'; \sigma_1, \ell_1)$ and $\kappa_2(x, x'; \sigma_2, \ell_2)$, then their cross-covariance K_{XY} can be written as another RBF kernel with a correlation length $\ell = \sqrt{(\ell_1^2 + \ell_2^2)/2}$ and correlation coefficient ρ:

$$\kappa(x, x'; \sigma_1, \sigma_2, \ell_1, \ell_2) = \sigma_1\sigma_2\rho \exp\left[-\frac{(x-x')^2}{\ell_1^2 + \ell_2^2}\right], \quad (\text{C16})$$

with

$$\rho = \sqrt{\frac{2\ell_1\ell_2}{\ell_1^2 + \ell_2^2}}. \quad (\text{C17})$$

Note that ρ is uniquely determined by ℓ_1 and ℓ_2, and, as $\ell_1 \to \ell_2$, the two outputs become 100% correlated. For the

$\Lambda = 500$ MeV interactions this model accurately reproduces the correlations found between the observable coefficients of SNM and PNM.

If one instead makes the constraint $\ell_1 = \ell_2$, or, more generally, that the same correlation kernel is used for each output, then Eq. (C17) need not be enforced. Rather, in this case, one can use the intrinsic coregionalization model [30]

$$\kappa_{\text{joint}}(x, x') = C \otimes \kappa(x, x'), \quad (\text{C18})$$

where C is a positive sem-definite matrix called the coregionalization matrix and $\otimes$ denotes the Kronecker product. C imposes the correlation structure *between* curves, while $\kappa(x, x')$ imposes a correlation structure *within* each curve. The off-diagonal components of C can then be tuned to data if desired, as long as C remains positive semidefinite.

We use the intrinsic coregionalization model, Eq. (C18), for the $\Lambda = 450$ MeV interactions. We choose the diagonal components to be $\bar{c}^2$ of $E/N(n)$ and $E/A(n)$, and the off-diagonal component of C to be $\bar{c}_{\text{PNM}}\bar{c}_{\text{SNM}}\rho$, where ρ is the empirical correlation of the c_n. For Eq. (C18), which is combined with Eqs. (C12)–(C15) to compute $S_2(n)$, we must choose a common length scale for κ. We take this to be $(\ell_{\text{PNM}} + \ell_{\text{SNM}})/2$, with ℓ_{SNM} on the k_F^{PNM} scale (i.e., it has the appropriate factor of $\sqrt[3]{2}$). Note, however, that the individual length scales are still used for all predictions requiring only PNM or SNM.

The total covariance of the truncation error model then follows directly from the total covariance of the coefficients κ:

$$\Sigma_{ij}(x, x') \equiv y_{\text{ref}\,i}(x)y_{\text{ref}\,j}(x')\frac{[Q_i(x)Q_j(x')]^{k+1}}{1 - Q_i(x)Q_j(x')}\kappa_{ij}(x, x'), \quad (\text{C19})$$

where i and j denote the observable, here $E/N(n)$ or $E/A(n)$. For $i = j$, this reduces to Eq. (8), but this extension describes correlations between the SNM and PNM truncation errors when $i \neq j$.

[1] E. Epelbaum, H.-W. Hammer, and Ulf-G. Meißner, Rev. Mod. Phys. **81**, 1773 (2009).

[2] R. Machleidt and D. R. Entem, Phys. Rep. **503**, 1 (2011).

[3] H.-W. Hammer, S. König, and U. van Kolck, Rev. Mod. Phys. **92**, 025004 (2020).

[4] I. Tews, Z. Davoudi, A. Ekström, J. D. Holt, and J. E. Lynn, J. Phys. G **47**, 103001 (2020).

[5] The Editors, Phys. Rev. A **83**, 040001 (2011).

[6] I. Tews, T. Krüger, K. Hebeler, and A. Schwenk, Phys. Rev. Lett. **110**, 032504 (2013).

[7] K. Hebeler, J. M. Lattimer, C. J. Pethick, and A. Schwenk, Astrophys. J. **773**, 11 (2013).

[8] G. Baardsen, A. Ekström, G. Hagen, and M. Hjorth-Jensen, Phys. Rev. C **88**, 054312 (2013).

[9] C. Drischler, V. Somà, and A. Schwenk, Phys. Rev. C **89**, 025806 (2014).

[10] G. Hagen, T. Papenbrock, A. Ekström, K. A. Wendt, G. Baardsen, S. Gandolfi, M. Hjorth-Jensen, and C. J. Horowitz, Phys. Rev. C **89**, 014319 (2014).

[11] A. Carbone, A. Polls, and A. Rios, Phys. Rev. C **88**, 044302 (2013).

[12] L. Coraggio, J. W. Holt, N. Itaco, R. Machleidt, L. E. Marcucci, and F. Sammarruca, Phys. Rev. C **89**, 044321 (2014).

[13] C. Wellenhofer, J. W. Holt, N. Kaiser, and W. Weise, Phys. Rev. C **89**, 064009 (2014).

[14] A. Roggero, A. Mukherjee, and F. Pederiva, Phys. Rev. Lett. **112**, 221103 (2014).

[15] J. W. Holt and N. Kaiser, Phys. Rev. C **95**, 034326 (2017).

[16] C. Drischler, A. Carbone, K. Hebeler, and A. Schwenk, Phys. Rev. C **94**, 054307 (2016).

[17] A. Ekström, G. Hagen, T. D. Morris, T. Papenbrock, and P. D. Schwartz, Phys. Rev. C **97**, 024332 (2018).

[18] C. Drischler, K. Hebeler, and A. Schwenk, Phys. Rev. Lett. **122**, 042501 (2019).

[19] D. Lonardoni, I. Tews, S. Gandolfi, and J. Carlson, Phys. Rev. Research **2**, 022033(R) (2020).

[20] M. Piarulli, I. Bombaci, D. Logoteta, A. Lovato, and R. B. Wiringa, Phys. Rev. C **101**, 045801 (2020).

[21] K. Hebeler, J. D. Holt, J. Menéndez, and A. Schwenk, Annu. Rev. Nucl. Part. Sci. **65**, 457 (2015).

[22] C. Drischler, W. Haxton, K. McElvain, E. Mereghetti, A. Nicholson, P. Vranas, and A. Walker-Loud, arXiv:1910.07961.

[23] F. Sammarruca and R. Millerson, Front. Phys. **7**, 213 (2019).

[24] C. Drischler, R. J. Furnstahl, J. A. Melendez, and D. R. Phillips, Phys. Rev. Lett. **125**, 202702 (2020).

[25] J. A. Melendez, R. J. Furnstahl, D. R. Phillips, M. T. Pratola, and S. Wesolowski, Phys. Rev. C **100**, 044001 (2019).

[26] C. E. Rasmussen, Bayesian Stat. **7**, 651 (2003).

[27] E. Solak, R. Murray-smith, W. E. Leithead, D. J. Leith, and C. E. Rasmussen, in *Advances in Neural Information Processing Systems 15*, edited by S. Becker, S. Thrun, and K. Obermayer (MIT Press, Cambridge, MA, 2003), pp. 1057–1064.

[28] D. Eriksson, K. Dong, E. Lee, D. Bindel, and A. G. Wilson, in *Advances in Neural Information Processing Systems 31*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., Red Hook, NY, 2018), pp. 6867–6877.

[29] M. Chilenski, M. Greenwald, Y. Marzouk, N. Howard, A. White, J. Rice, and J. Walk, Nucl. Fusion **55**, 023012 (2015).

[30] M. Álvarez, L. Rosasco, and N. Lawrence, *Kernels for Vector-Valued Functions: A Review*, Foundations and Trends in Machine Learning Series (Now Publishers Incorporated, Norwell, MA, 2012).

[31] A. Melkumyan and F. Ramos, in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Two, IJCAI'11* (AAAI Press, Palo Alto, 2011), pp. 1408–1413.

[32] R. Caruana, Mach. Learn. **28**, 41 (1997).

[33] Y. Zhang and Q. Yang, Natl. Sci. Rev. **5**, 30 (2017).

[34] R. J. Furnstahl, N. Klco, D. R. Phillips, and S. Wesolowski, Phys. Rev. C **92**, 024005 (2015).

[35] J. A. Melendez, S. Wesolowski, and R. J. Furnstahl, Phys. Rev. C **96**, 024003 (2017).

[36] S. Wesolowski, R. J. Furnstahl, J. A. Melendez, and D. R. Phillips, J. Phys. G **46**, 045102 (2019).

[37] L. S. Bastos and A. O'Hagan, Technometrics **51**, 425 (2009).

[38] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevC.102.054315 for the corresponding figures obtained using the $\Lambda = 450$ MeV potentials.

[39] M. Cacciari and N. Houdeau, J. High Energy Phys. 09 (2011) 039.

[40] J. Hu, P. Wei, and Y. Zhang, Phys. Lett. B **798**, 134982 (2019).

[41] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, Stat. Sci. **4**, 409 (1989).

[42] N. Cressie, Terra Nova **4**, 613 (1992).

[43] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, Adaptive computation and machine learning series (University Press Group Limited, Cambridge, MA, 2006).

[44] D. J. C. MacKay, in *Neural Networks and Machine Learning*, NATO ASI Series, Vol. 168, edited by C. M. Bishop (Springer, Berlin, 1998), pp. 133–166.

[45] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, UK, 2003).

[46] S. Wesolowski, N. Klco, R. J. Furnstahl, D. R. Phillips, and A. Thapaliya, J. Phys. G **43**, 074001 (2016).

[47] M. Leonhardt, M. Pospiech, B. Schallmo, J. Braun, C. Drischler, K. Hebeler, and A. Schwenk, Phys. Rev. Lett. **125**, 142502 (2020).

[48] D. R. Entem, R. Machleidt, and Y. Nosyk, Phys. Rev. C **96**, 024004 (2017).

[49] K. Hebeler, S. K. Bogner, R. J. Furnstahl, A. Nogga, and A. Schwenk, Phys. Rev. C **83**, 031301(R) (2011).

[50] BUQEYE Collaboration, 2020, https://buqeye.github.io/software/.

[51] C. Drischler, K. Hebeler, and A. Schwenk, Phys. Rev. C **93**, 054314 (2016).

[52] J. Hoppe, C. Drischler, K. Hebeler, A. Schwenk, and J. Simonis, Phys. Rev. C **100**, 024318 (2019).

[53] N. Kaiser, Phys. Rev. C **91**, 065201 (2015).

[54] C. Wellenhofer, J. W. Holt, and N. Kaiser, Phys. Rev. C **93**, 055802 (2016).

[55] J. D. Evans, *Straightforward Statistics for the Behavioral Sciences* (Brooks/Cole, Pacific Grove, CA, 1996).

[56] A. G. Asuero, A. Sayago, and A. G. González, Crit. Rev. Anal. Chem. **36**, 41 (2006).

[57] E. Epelbaum, H. Krebs, and U. G. Meißner, Phys. Rev. Lett. **115**, 122301 (2015).

[58] E. Epelbaum, H. Krebs, and U. G. Meißner, Eur. Phys. J. A **51**, 53 (2015).

[59] S. Binder, J. Langhammer, A. Calci, and R. Roth, Phys. Lett. B **736**, 119 (2014).

[60] V. Lapoux, V. Somà, C. Barbieri, H. Hergert, J. D. Holt, and S. R. Stroberg, Phys. Rev. Lett. **117**, 052501 (2016).

[61] E. Epelbaum, H. Krebs, and P. Reinert, Front. Phys. **8**, 98 (2020).

[62] T. Hüther, K. Vobig, K. Hebeler, R. Machleidt, and R. Roth, Phys. Lett. B **808**, 135651 (2020).

[63] A. Ekström, G. R. Jansen, K. A. Wendt, G. Hagen, T. Papenbrock, B. D. Carlsson, C. Forssén, M. Hjorth-Jensen, P. Navrátil, and W. Nazarewicz, Phys. Rev. C **91**, 051301(R) (2015).

[64] P. I. Frazier, arXiv:1807.02811.

[65] V. Picheny, D. Ginsbourger, Y. Richet, and G. Caplin, Technometrics **55**, 2 (2013).

[66] B. D. Carlsson, A. Ekström, C. Forssén, D. F. Strömberg, G. R. Jansen, O. Lilja, M. Lindby, B. A. Mattsson, and K. A. Wendt, Phys. Rev. X **6**, 011019 (2016).

[67] J. Melendez, gsum, 2020, Python package available for download from Github, github.com/buqeye/gsum.