Bayesian parameter estimation in chiral effective field theory using the Hamiltonian Monte Carlo method

Isak Svensson,[*] Andreas Ekström, and Christian Forssén

Department of Physics, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

The number of low-energy constants (LECs) in chiral effective field theory (χEFT) grows rapidly with increasing chiral order, necessitating the use of Markov chain Monte Carlo techniques for sampling their posterior probability density function. For this we introduce a Hamiltonian Monte Carlo (HMC) algorithm and sample the LEC posterior up to next-to-next-to-leading order (NNLO) in the two-nucleon sector of χEFT. We find that the sampling efficiency of HMC is three to six times higher compared to an affine-invariant sampling algorithm. We analyze the empirical coverage probability and validate that the NNLO model yields predictions for two-nucleon scattering data with largely reliable credible intervals, provided that one ignores the leading-order EFT expansion parameter when inferring the variance of the truncation error. We also find that the NNLO truncation error dominates the error budget.

I. INTRODUCTION

Chiral effective field theory (χEFT) descriptions [1–4] of the strong nuclear interaction depend on low-energy constants (LECs) that govern the strength of the various interaction terms. Their numerical values must be inferred from data and are best described by a posterior probability density function (PDF). Clearly, this parametric uncertainty will combine with the inherent discrepancy of χEFT, i.e., the epistemic gap between model predictions and real world observations. Thus, their joint uncertainty must be estimated and propagated forward before the credibility of the underlying theory and its predictions for nuclear observables can be assessed. Drawing samples from this posterior predictive distribution (PPD) is key to analyze the implications of physical and probabilistic modeling choices in the *ab initio* description of nuclear systems. Fortunately, operating with a χEFT endowed with a power counting offers a principal handle to estimate the relevant model discrepancy in terms of Bayesian credible intervals for the EFT truncation error [5]. Still, drawing samples from relevant posterior PDFs presents a formidable challenge, particularly for high-dimensional parameter volumes.

Here, we introduce the Hamiltonian Monte Carlo (HMC) method [6] to sample LEC posteriors and PPDs in χEFT. HMC is a Markov chain Monte Carlo (MCMC) [7,8] method that exploits the equations of Hamiltonian dynamics for drawing uncorrelated PDF samples with a high acceptance probability. Crucially, HMC performs well also in cases of high-dimensional probability distributions where most other MCMC algorithms fail to converge within a reasonable time frame. In this paper we demonstrate how to use HMC to efficiently sample the LEC posteriors at leading order (LO), next-to-leading order (NLO), and next-to-next-to-leading order (NNLO) in deltaless χEFT. We also perform model checking and model validation by drawing samples from the PPD for elastic nucleon-nucleon (NN) scattering in the neutron-proton (np) and proton-proton (pp) channels. All PDFs are conditioned on data from the recent Granada database [9,10] of measured NN scattering cross sections. We employ a statistical model for the EFT truncation error, proposed by Wesolowski *et al.* [11], to account for the model discrepancy due to excluded contributions from higher chiral orders.

In a Bayesian data analysis it is straightforward to condition on existing results. We make use of a previous Roy-Steiner analysis [12,13] to incorporate prior knowledge about the LECs that govern the strengths of subleading pion-nucleon (πN) interactions. We also place a prior in accordance with naturalness expectations on the LECs that govern contact interaction strengths. Throughout this paper, we define the necessary one- and two-pion exchanges and contact interactions according to Ref. [3] and follow the same conventions for the potential, scattering amplitudes, and NN scattering observables as in Ref. [14]. We use a nonlocal super-Gaussian momentum-space regulator with a fixed cutoff $\Lambda = 450$ MeV.

II. BAYESIAN PARAMETER ESTIMATION IN χEFT

Bayes's theorem

$$\mathrm{pr}(\vec{\alpha}|D, I) = \frac{\mathrm{pr}(D|\vec{\alpha}, I) \cdot \mathrm{pr}(\vec{\alpha}|I)}{\mathrm{pr}(D|I)} \tag{1}$$

---

*isak.svensson@chalmers.se

provides a straightforward way to express the posterior PDF $\mathrm{pr}(\vec{\alpha}|D, I)$ for the LECs $\vec{\alpha}$ in terms of a likelihood $\mathrm{pr}(D|\vec{\alpha}, I)$, prior $\mathrm{pr}(\vec{\alpha}|I)$, and marginal likelihood (or evidence) $\mathrm{pr}(D|I)$. In this paper, the LEC posterior is conditioned on NN scattering data D and additional prior information I. The marginal likelihood, which does not depend on $\vec{\alpha}$, plays no role in parameter estimation and we therefore have

$$\mathrm{pr}(\vec{\alpha}|D, I) \propto \mathrm{pr}(D|\vec{\alpha}, I) \cdot \mathrm{pr}(\vec{\alpha}|I). \tag{2}$$

A hallmark of the Bayesian approach is the transparent and straightforward inclusion of prior information I. A main goal of this paper is to incorporate the probabilistic model for the EFT truncation error from Ref. [11] and introduce an HMC algorithm to efficiently sample the posterior PDF for all LECs up to NNLO.

In the following subsections we present the NN data we use, our prior, and our likelihood, and specify the hyperparameters and details of the model for relating experimental NN scattering data to a χEFT prediction at a given chiral order.

A. Prior

Our full prior for the LECs is written as a product of two independent priors: $\mathrm{pr}(\vec{\alpha}_{NN}|I)$ for the contact LECs and $\mathrm{pr}(\vec{\alpha}_{\pi N}|I)$ for the πN LECs. This form,

$$\mathrm{pr}(\vec{\alpha}|I) = \mathrm{pr}(\vec{\alpha}_{NN}|I) \cdot \mathrm{pr}(\vec{\alpha}_{\pi N}|I), \tag{3}$$

implies no prior assumption of correlation between LECs from these sectors. Furthermore, we adopt independent and identical Gaussian PDFs for all contact LECs with zero mean and standard deviation $\bar{\alpha} = 5$. This is a rather weak prior which again makes no assumption of correlations. However, it mildly encodes the naturalness expectation of the LECs by penalizing LEC values $\gg 1$, thus safeguarding somewhat against overfitting [15]. The exact value of $\bar{\alpha}$ does not have a major impact on the outcome, as the large NN data set used in the likelihood strongly dominates over the prior. The prior at LO and NLO in χEFT—where no πN LECs appear—is thus given by

$$\mathrm{pr}(\vec{\alpha}|I) = \mathcal{N}(\vec{0}, \Sigma_{\mathrm{prior}}) \tag{4}$$

with

$$(\Sigma_{\mathrm{prior}})_{ij} = \bar{\alpha}^2 \delta_{ij}. \tag{5}$$

Here, $\mathcal{N}(\vec{\mu}, \Sigma)$ denotes a multivariate normal (Gaussian) PDF with mean vector $\vec{\mu}$ and covariance matrix Σ.

For the πN LECs c_1, c_3, and c_4, entering at NNLO, we have chosen a much more restrictive prior based on mean values and covariance matrices extracted from the maximum-likelihood fit of a Roy-Steiner analysis of the πN scattering amplitudes by Siemens *et al.* [13]. That analysis proceeds in a kinematical region of chiral perturbation theory that exhibits a stronger curvature with respect to the subthreshold parameters of the πN scattering amplitudes. Once matched to the πN LECs in χEFT, we obtain a rather informative prior for the corresponding part of the potential. We will return to a more detailed discussion of the πN LECs when analyzing the MCMC posteriors in Sec. IV B.

To be specific, the LECs that we consider at each order are

$$\vec{\alpha}_{\mathrm{LO}} = (\widetilde{C}_{1S0}, \widetilde{C}_{3S1}), \tag{6}$$

$$\vec{\alpha}_{\mathrm{NLO}} = (\widetilde{C}_{1S0}^{np}, \widetilde{C}_{1S0}^{pp}, \widetilde{C}_{3S1}, C_{1S0}, C_{3P0},$$
$$C_{1P1}, C_{3P1}, C_{3S1}, C_{3S1-3D1}, C_{3P2}), \tag{7}$$

$$\vec{\alpha}_{\mathrm{NNLO}} = (c_1, c_3, c_4, \widetilde{C}_{1S0}^{np}, \widetilde{C}_{1S0}^{pp}, \widetilde{C}_{3S1}, C_{1S0},$$
$$C_{3P0}, C_{1P1}, C_{3P1}, C_{3S1}, C_{3S1-3D1}, C_{3P2}). \tag{8}$$

We employ a conventional notation linked to the momentum partial-wave basis (see, e.g., Ref. [3]). Note also that isospin-breaking effects enter at NLO and only in the $^1S_0$ partial wave.

B. Elastic nucleon-nucleon scattering observables

We condition our LEC posterior on experimental data. We use roughly two thirds of the Granada 2013 database [9,10] to define the training data set D. We hold out all scattering data in the range $80 \leqslant T_{\mathrm{lab}} \leqslant 100$ MeV of laboratory scattering energies and assign it to a validation data set $\widetilde{D}$. We further assign to $\widetilde{D}$ all data in the energy range $290 < T_{\mathrm{lab}} \leqslant 350$ MeV, i.e., just above the pion-production threshold, plus the set of integrated np scattering cross sections from Ref. [16]. Overall, this choice of data split enables detailed model checking while leaving ample information for estimating the LECs at each chiral order up to NNLO. In all, D consists of 4366 experimental data points to be used as input in the parameter estimation process, while $\widetilde{D}$ contains 2018 validation data points. The details of the NN scattering data used for parameter estimation and validation are presented in Table I.

Given a set of numerical values for the LECs, we compute scattering amplitudes by solving the Lippmann-Schwinger equation for all partial waves with maximum total angular-momentum quantum number $J_{\mathrm{max}} \leqslant 30$. This is more than enough to converge the physical model predictions for the resulting scattering observables. Thus, we neglect all sources of numerical or computational method uncertainties going forward.

In pp scattering we include all relevant electromagnetic effects, as outlined in Ref. [14]: the static Coulomb interaction and its relativistic correction, the first-order approximation to the vacuum polarization, and relevant magnetic moment interactions. This set of long-ranged electromagnetic interactions has been demonstrated by the Nijmegen group to be sufficient for explaining the observed low-energy pp scattering data [18,19]. For this reason, we also neglect any theoretical model discrepancy due to neglected higher-order contributions to the electromagnetic interaction.

C. Likelihood and EFT truncation error

In this paper we relate an experimental measurement y_{exp} of some observable y to the true value y_{true} via a statistical model

$$y_{\mathrm{exp}} = y_{\mathrm{true}} + \delta y_{\mathrm{exp}}. \tag{9}$$

This also introduces the experimental uncertainty, δy_{exp}, as a random variable for which we employ the standard deviations provided in the Granada database. We also relate the true value

TABLE I. The distribution of observables in the training and validation data sets denoted D and $\widetilde{D}$, respectively. $\widetilde{D}$ is composed of all available data in the $80 \leqslant T_{\text{lab}} \leqslant 100$ and $290 < T_{\text{lab}} \leqslant 350$ MeV ranges, plus one set of np SGT data [16] covering the $33 \leqslant T_{\text{lab}} \leqslant 350$ MeV range. All other available data constitute D. We denote scattering observables using the SAID nomenclature [17], e.g., integrated cross section (SGT), unpolarized differential cross section (DSG), polarization (P), polarization of the beam (PB), and polarization of the target (PT). These observables make up the bulk part of the experimental NN database.

Obs.	Training data (D)			Validation data ($\widetilde{D}$)		
	np	pp	Total (%)	np	pp	Total (%)
SGT	315	0	315 (7.2)	84	0	84 (4.2)
SGTL	11	0	11 (0.3)	4	0	4 (0.2)
SGTT	16	0	16 (0.4)	3	0	3 (0.1)
DSG	1221	756	1977 (45.2)	457	159	616 (30.5)
A	5	47	52 (1.2)	0	24	24 (1.2)
AP	0	5	5 (0.1)	0	0	0 (0.0)
AT	30	0	30 (0.7)	35	0	35 (1.7)
AXX	0	143	143 (3.2)	0	120	120 (5.9)
AYY	64	151	215 (4.8)	46	151	197 (9.8)
AZX	0	137	137 (3.1)	0	120	120 (5.9)
AZZ	45	39	84 (1.9)	27	10	37 (1.8)
CKP	0	1	1 (0.0)	0	1	1 (0.0)
D	13	56	69 (1.6)	14	42	56 (2.8)
D0SK	8	0	8 (0.2)	14	0	14 (0.7)
DT	39	0	39 (0.9)	39	0	39 (1.9)
MSKN	0	8	8 (0.2)	0	8	8 (0.4)
MSSN	0	8	8 (0.2)	0	8	8 (0.4)
NNKK	8	0	8 (0.2)	0	0	0 (0.0)
NSKN	12	0	12 (0.3)	13	0	13 (0.6)
NSSN	4	0	4 (0.1)	14	0	14 (0.7)
P	0	489	489 (11.2)	0	260	260 (12.9)
PB	590	0	590 (13.5)	253	0	253 (12.5)
PT	38	0	38 (0.9)	19	0	19 (0.9)
R	5	50	55 (1.3)	0	50	50 (2.5)
RP	0	22	22 (0.5)	0	5	5 (0.2)
RPT	1	0	1 (0.0)	1	0	1 (0.0)
RT	29	0	29 (0.7)	37	0	37 (1.8)
All	2454	1912	4366 (100)	1060	958	2018 (100)

and our theory prediction y_{th} via

$$y_{\text{true}} = y_{\text{th}} + \delta y_{\text{th}}, \tag{10}$$

where δy_{th} is the model discrepancy term. We will model δy_{th} as coming from the truncation of the chiral expansion in χEFT at some finite chiral order k. When doing so we tacitly assume that the entire epistemic uncertainty of the theory can be systematically reduced by going to higher orders in χEFT.

To model the truncation error we follow Furnstahl *et al.* [5] and Wesolowski *et al.* [11] and formally write the χEFT expansion for some observable prediction y_{th} up to chiral order k as

$$y_{\text{th}}^{(k)} = y_{\text{ref}} \sum_{\nu=0}^{k} c_\nu Q^\nu, \tag{11}$$

where y_{ref} is a reference value, c_ν are dimensionless EFT expansion coefficients, and Q is a dimensionless expansion parameter that we assign as

$$Q = \frac{\max(m_\pi, p)}{\Lambda_b} \tag{12}$$

where m_π is the pion mass, p is a soft scale associated with the observable, and Λ_b is a hard scale. In this paper, we will set $\Lambda_b = 600$ MeV and p is given by the NN scattering momentum. All contributions to the potential at chiral order $\nu = 1$ vanish, i.e., $c_1 = 0$, since we employ Weinberg power counting. We refer to the different orders as LO ($\nu = 0$), NLO ($\nu = 2$), and NNLO ($\nu = 3$).

Truncating the χEFT expansion at order k induces a truncation error given by

$$\delta y_{\text{th}}^{(k)} = y_{\text{ref}} \sum_{\nu=k+1}^{K} c_\nu Q^\nu \tag{13}$$

where $K \to \infty$. Assuming that *all* c_ν coefficients, including those for which $\nu \leqslant k$, are independent and identically distributed, we can use known lower-order coefficients $c_0, \ldots, c_k$ to learn about the single PDF from which the unknown higher-order coefficients should also be sampled. This, combined with a prior assumption about the form of the PDF for the c_ν coefficients, provides us with a prescription for quantitatively estimating $\delta y_{\text{th}}^{(k)}$.

We assume a Gaussian prior with variance $\bar{c}^2$ for c_ν:

$$\text{pr}(c_\nu|I) = \mathcal{N}(0, \bar{c}^2). \tag{14}$$

All physical scales reside in Q and y_{ref} and it is reasonable to assume that the c_ν coefficients are of order 1. This is an assumption we will test explicitly in Sec. II C 2 when estimating $\bar{c}$ from order-by-order shifts in the prediction of different NN scattering observables y.

Placing a normal prior for c_ν, with known $\bar{c}^2$, leads to a normal PDF with variance σ_{th}^2 for $\delta y_{\text{th}}^{(k)}$ in the limit $K \to \infty$ given by [11]

$$\text{pr}(\delta y_{\text{th}}^{(k)}|\bar{c}^2, Q, I) = \mathcal{N}(0, \sigma_{\text{th}}^2), \tag{15}$$

where we have also conditioned on Q, and

$$\sigma_{\text{th}}^2 = \bar{c}^2 y_{\text{ref}}^2 \frac{Q^{2(k+1)}}{1 - Q^2}. \tag{16}$$

In this paper we will assume that the theory errors for different observables (y_i, y_j) are completely uncorrelated. This leads to a diagonal covariance matrix Σ_{th} for the truncation error at order k that is given by

$$(\Sigma_{\text{th}})_{ij} = \sigma_{\text{th},i}^2 \delta_{ij}. \tag{17}$$

Note that the LO truncation error is proportional to $Q^4/(1 - Q^2)$ since the $\nu = 1$ chiral order vanishes such that the first nonzero term in Eq. (13) corresponds to $\nu = 2$.

The likelihood function for the entire data set D can—assuming that experimental and theoretical errors are independent—be expressed as

$$\text{pr}(D|\vec{\alpha}, I) \propto \exp\left[-\tfrac{1}{2}\vec{r}^T \cdot (\Sigma_{\text{exp}} + \Sigma_{\text{th}})^{-1} \cdot \vec{r}\right], \tag{18}$$

with the residual vector $\vec{r}$ defined as

$$\vec{r} = \vec{y}_{\text{th}}^{(k)}(\vec{\alpha}) - \vec{y}_{\text{exp}}, \qquad (19)$$

i.e., the difference between the experimental data and the corresponding set of χEFT predictions at order k given a vector $\vec{\alpha}$ of LEC values. We employ a diagonal covariance matrix Σ_{exp} also for the experimental data and use the normalization factors from the Granada database [9,10] for the joint systematic uncertainty of a group of data originating from the same experiment [20]. Note also that the likelihood (18) is implicitly conditional on $\bar{c}$ and the breakdown scale Λ_b.

Let us also mention two possible extensions of the statistical model used in this paper. First, instead of placing an implicit delta prior on $\bar{c}^2$, one could follow standard Bayesian practice and exploit the class of conjugate priors, assuming, e.g., an inverse-χ^2 PDF such that $\text{pr}(\bar{c}^2|I) = \chi^{-2}(\nu_0, \tau_0^2)$.[1] The conjugacy of this prior with respect to the normal PDF leads to a Student's t PDF for the expansion parameters c_ν and the truncation error $\delta y_{\text{th}}^{(k)}$ (see Ref. [21]). The Student's t distribution [22] is essentially a heavier-tailed variant of the normal distribution, arising here as the result of the unknown variance $\bar{c}^2$. Second, one could relax the assumption of completely uncorrelated truncation errors and model the covariance structure of the expansion coefficients c_ν using, e.g., a Gaussian process where the correlation length in Q is one of its hyperparameters [21].

1. Choice of reference values

The reference value, y_{ref}, in Eq. (13) for each observable can be assigned in different ways. Options include using an experimental value, picking some suitable theoretical prediction, or a value motivated from an order of magnitude estimate. We have chosen to let LO predictions set the scale for observables. But since LO predictions depend on the LECs $\vec{\alpha}$—which at this point in the analysis remain to be inferred—it is necessary to extract a reasonable point estimate of LO LEC values $\vec{\alpha}_{\nu=0}^\star$. We do this by maximizing the likelihood (18) at LO (i.e., setting $k = 0$) using Σ_{th} from Eq. (17) with $y_{\text{ref}} = y_{\text{exp}}$ and $\bar{c} = 1$, as motivated by naturalness. The choice of using experimental data as reference values when maximizing the likelihood was technically convenient but has little practical impact on the subsequent steps of the analysis. In this way, we let the maximum likelihood estimate (MLE) at LO, $y_{\text{th}}^{(0)}(\vec{\alpha}_{\nu=0}^\star)$, be the reference value for integrated (SGT) and differential cross section (DSG) observables. For spin polarization and correlation cross sections we use $y_{\text{ref}} = 0.15$ as motivated by the average of the LO maximum-likelihood prediction.

2. Estimating $\bar{c}$

We proceed to estimate $\bar{c}$ from the finite set of expansion coefficients c_ν obtained by rearranging Eq. (11) and setting $\vec{\alpha} = \vec{\alpha}_\nu^\star$. In general, we obtain a vector, $\vec{c}_\nu$, of expansion

---

[1]Here ν_0 and τ_0 are hyperparameters corresponding to the degrees of freedom and scale for the χ^{-2} PDF. This PDF is equivalent to a normal-inverse-gamma PDF.

TABLE II. Results from the $\bar{c}$ analysis. $\bar{c}_\nu$ denotes the rms value of the EFT expansion coefficients at a particular order ν, while $\bar{c}_{0...\nu}$ denotes the rms value up to, and including, order k.

Order	ν	$\bar{c}_\nu$	$\bar{c}_{0...\nu}$	Outliers (order ν)	Outliers (all orders)
LO	0	1.17	1.17	1/54	1/54
NLO	2	4.95	2.08	0/54	15/108
NNLO	3	2.84	2.72	1/54	8/162

coefficients at order ν:

$$c_{\nu,i} = \frac{y_{\text{th},i}^{(\nu)}(\vec{\alpha}_\nu^\star) - y_{\text{th},i}^{(\nu-1)}(\vec{\alpha}_{\nu-1}^\star)}{y_{\text{ref},i}Q_i^\nu}, \quad i = 1, \ldots, N_o, \qquad (20)$$

for a representative set of $N_o = 54$ observables at different scattering energies and angles. Specifically, we used a combination of total cross sections, differential cross sections, and spin observables in np and pp at an energy grid $T_{\text{lab}} = 20, 70, 120, 170, 220$, and 270 MeV, and scattering angles $\theta = 50°$ and $150°$, such that we cover the relevant kinematical regions while being sufficiently well separated in the energy-angle variables to reduce the influence of finite correlations in the expansion coefficients. We account for the vanishing of chiral order $\nu = 1$ when computing the order-by-order differences.

At each order we then compute a rms value, i.e.,

$$\bar{c}_\nu = \sqrt{\frac{1}{N_o} \sum_{i=1}^{N_o} c_{\nu,i}^2}. \qquad (21)$$

Our choice to set y_{ref} according to the LO MLE effectively constrains the expansion coefficient c_0, leading to $c_0 = 1.17$, accounting for the averaged spin polarizations. See the third column in Table II for results at the other two orders. We also define an estimate of $\bar{c}$ as the total rms value of the $c_{\nu,i}$ coefficients up to, and including, the order k at which we truncate the EFT expansion (see the fourth column in Table II). The highest-order estimate for $\bar{c}$ is given by $\bar{c}_{0...3}$ and therefore includes information from LO, NLO, and NNLO. Note that outlier values of $c_{\nu,i}$ were removed before computing all rms values since they would otherwise influence our estimate disproportionally. We used the following procedure to determine outliers.

(1) Compute the lower and upper quartiles $C_{25\%}$ and $C_{75\%}$ of $c_{\nu,i}$.

(2) Determine the distance ΔC between the quartiles.

(3) Discard any $c_{\nu,i}$ that falls outside the interval $[C_{25\%} - 3\Delta C, C_{75\%} + 3\Delta C]$.

The numbers of removed outliers at each order are provided in Table II.

Since we express the observable predictions as an EFT expansion in Eq. (11) we expect to observe expansion coefficients of natural size. On average we find relatively natural values that characterize the truncation error. However, the larger difference in the χEFT predictions when going from LO to NLO is a signature of an irregular convergence pattern and incurs an unexpectedly large value for $\bar{c}_{\nu=2}$. We analyze the consequences of this in Sec. V.

III. HAMILTONIAN MONTE CARLO

Hamiltonian Monte Carlo shares some key features with the canonical Metropolis-Hastings algorithm [7,8]. These MCMC methods draw samples from a PDF $\text{pr}(\vec{\alpha})$ by producing an ergodic Markov chain of states the unique stationary distribution of which is $\text{pr}(\vec{\alpha})$. Given a current state $\vec{\alpha}_n$ of the MCMC chain, the Metropolis-Hastings algorithm proposes a new state $\vec{\alpha}'_n$ from a proposal distribution $q(\vec{\alpha}'_n|\vec{\alpha}_n)$. The new state is accepted with a probability a given by

$$a = \min(1, r) \qquad (22)$$

where r is the Hastings ratio:

$$r = \frac{\text{pr}(\vec{\alpha}'_n)q(\vec{\alpha}'_n|\vec{\alpha}_n)}{\text{pr}(\vec{\alpha}_n)q(\vec{\alpha}_n|\vec{\alpha}'_n)}. \qquad (23)$$

The next state of the chain, $\vec{\alpha}_{n+1}$, will be set to $\vec{\alpha}'_n$ if the update is accepted, or to a copy of $\vec{\alpha}_n$ if it is rejected. The proposal distribution is typically a normal distribution but many other options exist. The Metropolis-Hastings algorithm is a single-particle (or single-walker) algorithm, i.e., it is only aware of a single location in the parameter space at any given time. The primary drawback of the random-walk Metropolis-Hastings algorithm, and most of its derivatives, is that the elements of the resulting MCMC chain are strongly correlated with each other in most practical applications. Such correlations decrease the amount of information contained in the MCMC chain, requiring longer chains than would otherwise be necessary. The problem is exacerbated when the dimensionality of the parameter space for the sampled distribution $\text{pr}(\vec{\alpha})$ increases.

In 1987, lattice field theorists devised an ingenious MCMC algorithm based on Hamiltonian dynamics to combat the problem of correlated samples. They dubbed the algorithm hybrid Monte Carlo (HMC) [6], but it has subsequently become known as Hamiltonian Monte Carlo [23], retaining the acronym. HMC is based on the Metropolis-Hastings algorithm, but the method for proposing a new state is radically modified. The PDF to be sampled is treated as a potential energy surface and the current state of the chain is regarded as a particle with position coordinates given by the parameters of the PDF. By endowing the HMC "particle" at iteration n with a randomly drawn momentum $\vec{p}_n$ we can associate the joint state $(\vec{\alpha}_n, \vec{p}_n)$ with a total energy and, by extension, a probability to find the system in that state. Simulating the particle's trajectory for a finite period of time subsequently yields a proposed state $(\vec{\alpha}'_n, \vec{p}'_n)$. The proposed state retains the previous total energy since this is a conserved quantity in Hamiltonian dynamics and the new state will therefore be accepted. The momentum $\vec{p}'_n$ is then discarded. In practice, the total energy will only be approximately conserved due to numerical simulation errors, and we vet the proposed state in a manner similar to Eq. (22).

The result is a completely uncorrelated sample of the target PDF, assuming that the length of the trajectory is appropriately chosen. Clearly, the length of the MCMC chain produced by HMC can then be drastically reduced compared to chains with a high degree of correlation. We will quantify this statement by extracting an effective number of samples

in Sec. IV D. The strongest advantage of using the HMC algorithm is its potential for producing uncorrelated samples even when the target PDF is high-dimensional. Unfortunately, simulating Hamiltonian dynamics for each proposed state is computationally expensive compared to the method of proposing new states by applying small random perturbations. It is therefore apparent that the reduction in overall chain length and intersample correlations must be sufficiently large to warrant the increased cost per sample.

In the following subsections we present the HMC algorithm in detail. We have opted to write a custom implementation[2] using PYTHON [25] and NUMPY [26] in lieu of using a standard package such as STAN [27]. Specific implementation choices will be presented as appropriate. Our implementation and the mathematical details presented here are largely based on Ref. [23]. A conceptual introduction to HMC can be found in Ref. [28].

A. From potential energy to posterior probability

In HMC we generate samples from the d-dimensional posterior PDF $\text{pr}(\vec{\alpha}|D, I)$ by simulating classical Hamiltonian dynamics for a particle moving under the influence of a potential proportional to the posterior itself. To see this, we let the phase space of states—specified by position and momentum coordinates $(\vec{\alpha}, \vec{p})$ with total energy H—be described by a Boltzmann distribution:

$$\text{pr}(\vec{\alpha}, \vec{p}) = \frac{1}{Z} \exp[-H(\vec{\alpha}, \vec{p})/T]. \qquad (24)$$

For a classical Hamiltonian function

$$H(\vec{\alpha}, \vec{p}) = K(\vec{p}) + U(\vec{\alpha}) \qquad (25)$$

with kinetic energy K and potential energy U, the Boltzmann distribution factorizes and the marginal Boltzmann distribution for the position vector $\vec{\alpha}$ is independent of the distribution for $\vec{p}$. For our purposes, the temperature T is some function that makes the exponent dimensionless, and we set $T = 1$. The factor Z is independent of $(\vec{\alpha}, \vec{p})$ and ensures proper normalization of the distribution in Eq. (24). From the exponential Boltzmann factor we identify the potential energy term as the negative log posterior (NLP). Disregarding the constant marginal likelihood we define

$$\begin{aligned} U(\vec{\alpha}) &= -\log[\text{pr}(\vec{\alpha}|D, I)] \\ &= -\log[\text{pr}(D|\vec{\alpha}, I)\text{pr}(\vec{\alpha}|I)]. \end{aligned} \qquad (26)$$

In the last step we also used Bayes's theorem (1) to link the potential directly to the likelihood and the prior.

Following standard practice, and the classical dynamics analogy, we employ a quadratic form for $K(\vec{p})$:

$$K(\vec{p}) = \tfrac{1}{2}\vec{p}^T \mathcal{M}^{-1} \vec{p}. \qquad (27)$$

Here $\mathcal{M}$ is a positive-definite, symmetric matrix, called the mass matrix. With $K(\vec{p}) = -\log[\text{pr}(\vec{p})]$ we have

$$\text{pr}(\vec{p}) = \mathcal{N}(0, \mathcal{M}). \qquad (28)$$

---

[2]Available as free software [24].

The HMC algorithm thus samples the joint PDF $\mathrm{pr}(\vec{\alpha}, \vec{p})$. Marginalizing over the auxiliary momentum $\vec{p}$—by simply discarding these coordinates in the Markov chain—leaves us with MCMC samples of $\mathrm{pr}(\vec{\alpha}|D, I)$.

A particle governed by Hamiltonian dynamics moves through $2d$-dimensional phase space on a hypersurface of constant energy. The time evolution of the system is described by Hamilton's equations

$$\frac{d\alpha_i}{dt} = \frac{\partial H}{\partial p_i}, \tag{29}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial \alpha_i} \tag{30}$$

where $i = 1 \ldots d$. The form of the Hamiltonian, Eq. (25), allows us to rewrite Hamilton's equations as

$$\frac{d\alpha_i}{dt} = \frac{\partial K}{\partial p_i} = (\mathcal{M}^{-1}p)_i, \tag{31}$$

$$\frac{dp_i}{dt} = -\frac{\partial U}{\partial \alpha_i}. \tag{32}$$

Note that $\partial U/\partial \alpha_i$ is a partial derivative of the NLP that we must evaluate in order to simulate Hamiltonian dynamics and, by extension, use HMC sampling. Realistically, this must be done through automatic differentiation (AD) except in special cases where analytic expressions for these partial derivatives are available (our implementation readily allows for both possibilities). AD generally incurs a factor of 2 overhead compared to just evaluating the target PDF [29]. In this paper, we exploit an external AD library [30] for computing the necessary gradients and we measure the computational overhead of AD to less than 50% (see Sec. IV C 2).

B. Advancing the HMC sampler

Starting from some current state $(\vec{\alpha}_n, \vec{p}_n)$, the total energy of a particle trajectory traversing the HMC phase space is conserved. We should thus always accept the proposed state $(\vec{\alpha}'_n, \vec{p}'_n)$ at the end of the Hamiltonian trajectory, discard the auxiliary momentum $\vec{p}'_n$, and store $\vec{\alpha}'_n$ as the new parameter sample $\vec{\alpha}_{n+1}$ in the HMC chain. In practice, however, numerical errors in the solution of Eqs. (31) and (32) break energy conservation, which in turn breaks detailed balance. If ignored, we can no longer guarantee that the Markov chain converges to the sought stationary distribution. The solution is to vet the proposed state, i.e., the state at the end of the Hamiltonian trajectory, with the accept/reject step of the Metropolis-Hastings algorithm. We can also ensure a symmetric proposal distribution by negating the momentum variable $\vec{p}'_n$ at the end of the particle trajectory. However, such a negation is not necessary for a quadratic momentum distribution as in Eq. (28). The new state is thus accepted with the probability a given by Eq. (22) and a Hastings ratio for the joint probabilities given by

$$r = \frac{\mathrm{pr}(\vec{\alpha}'_n, \vec{p}'_n)}{\mathrm{pr}(\vec{\alpha}_n, \vec{p}_n)} = \exp\left[-H(\vec{\alpha}'_n, \vec{p}'_n) + H(\vec{\alpha}_n, \vec{p}_n)\right]. \tag{33}$$

The HMC chain will typically be ergodic since we draw a new momentum before integrating Hamilton's equations and thereby drastically altering the total energy.

C. Leapfrogging Hamiltonian dynamics

Upon imposing the Metropolis-Hastings accept/reject criterion we implicitly require reversibility of our chain. Fortunately, Hamiltonian dynamics is time reversible and it is necessary to integrate Hamilton's equations in a way that preserves this property. Standard methods like Euler integration or explicit Runge-Kutta methods are disqualified as they do not preserve time reversibility. From a strictly practical point of view, integration methods that do not conserve the Hamiltonian also limit the length of the particle trajectory since a large accumulated error would result in an unacceptably low mean acceptance rate $\bar{a}$. To ensure time reversibility, one should use a symplectic integrator that goes hand in hand with the volume preservation of phase space that follows from Liouville's theorem. The local discretization error of a symplectic integrator is equally likely to be positive or negative in each step of the integration as long as the step size ϵ is below some threshold value. The result is that the total energy is approximately conserved for an arbitrarily long trajectory. A nice property of HMC is that $\bar{a}$ drops precipitously if ϵ is greater than the threshold value. The absence of quiet failures makes it trivial to diagnose a too large choice for ϵ. The upper bound for ϵ is generally imposed by the most constrained parameter in the posterior.

The number of leapfrog iterations L can drastically influence the performance of the HMC sampler; it is imperative that L is set neither too high nor too low. A too small number partly defeats the purpose of using HMC in the first place, as it would result in a random-walk-like behavior with highly correlated samples. In contrast, a too large value of L would waste valuable CPU cycles without improving (and possibly even curtailing) performance. Naturally, choosing a very small step size ϵ needs to be compensated for by increasing L in order to avoid random walks.

Neither ϵ nor L is fixed in our implementation. Rather, random values are drawn from predefined probability distributions prior to each invocation of the leapfrog solver according to

$$\epsilon \sim \mathcal{U}\left(\tfrac{1}{2}\epsilon^\star, \tfrac{3}{2}\epsilon^\star\right), \tag{34}$$

$$L \sim \mathcal{U}\left\{\tfrac{1}{2}L^\star, \tfrac{3}{2}L^\star\right\}, \quad \tfrac{1}{2}L^\star \geqslant 1, \tag{35}$$

where the user specifies the nominal values $\epsilon^\star$ and $L^\star$. The reasons for randomizing these leapfrog parameters are threefold. First, variations in the trajectory length ϵL may decrease correlations between samples. Second, a fixed trajectory length can result in oscillatory behavior if ϵL happens to approximately match some periodicity of the target distribution. This type of (nearly) nonergodic behavior can severely limit the efficiency of HMC. Third, the target PDF may have regions where its gradient is very steep so that the nominal ϵ is too large to resolve features in that section.

There are extensions of HMC the purpose of which is to relieve the user from the burden of tuning the hyperparameters

ϵ and L. For instance, the choice of ϵ may be automated based on acceptance rates of small trial runs. The state-of-the-art no-U-turn sampler (NUTS) terminates trajectories based on heuristic rules for when continued simulation no longer increases the performance of the sampler. Both of these improvements are described in Ref. [31].

D. Tuning in to the target distribution

An array of tunable HMC parameters have been introduced in the previous subsections: the step size ϵ, the number of leapfrog iterations L, and the mass matrix $\mathcal{M}$. A drawback of HMC is the need to carefully tune these hyperparameters to each target distribution, or risk poor performance. However, as mentioned in Sec. III C, the tuning of ϵ and L may be automated. For now, we use a manual tuning procedure that will be outlined below. It is designed to achieve efficient sampling and also involves the important mass matrix.

The keen reader may have noticed that ϵ, L, and $\mathcal{M}$ are interlinked and changing one may force us to change one (or both) of the others. The tuning procedure is therefore, to an extent, iterative. We start with ϵ.

1. Leapfrog step size and number of iterations

For tuning ϵ we exploit that the acceptance rate $\bar{a}$ is largely independent of L (a consequence of using a symplectic integrator) and produce a very short HMC chain using a small number of leapfrog steps ($L = 3$). If $\bar{a}$ is low (in practice 0%) we decrease ϵ by an order of magnitude and try again. If $\bar{a}$ is 100% we instead increase ϵ by an order of magnitude and try again. Once the appropriate magnitude is found we make fine-grained adjustments as necessary. Using this method we quickly achieved HMC acceptance rates of 99% at all three chiral orders. This value is likely too high for optimal efficiency, but we find it adequate for our purposes.

The number of required leapfrog iterations L is intimately linked to the choice of the mass matrix $\mathcal{M}$. Careful tuning of L is rather pointless until $\mathcal{M}$ is settled. We have found that $L \approx 10$–20 yields excellent performance for approximately Gaussian distributions with around a dozen or so parameters, assuming that the mass matrix (and ϵ) is well chosen. To rapidly assess the choice of L—and the overall performance—it is useful to inspect trace plots of each individual parameter. The trace plots will reveal no apparent structures if the HMC algorithm is performing well. Any remaining structures in the trace plots may be quenched by increasing L, at the obvious expense of increased computational effort, or by improving the mass matrix. The latter alternative should always take precedence if possible. Note that it is usually necessary to revisit the tuning of ϵ after the mass matrix has been updated.

2. Mass matrix

Both ϵ and L are scalar values with no distinction for each individual parameter α_i and thus cannot be used to compensate for differences in parameter scales. Like Metropolis-Hastings, and unlike, e.g., affine-invariant ensemble samplers [32] such as the EMCEE package [33], HMC is

sensitive to such differences of scale and we need a way to account for them. This is the purpose of the mass matrix $\mathcal{M}$.

Deploying a mass matrix that captures the most important features of the target distribution is absolutely critical to the performance of HMC. An improper choice of $\mathcal{M}$ can degrade the performance by several orders of magnitude. Letting $\mathcal{M} = \mathbb{1}$, i.e., an identity matrix, does not fare well with the χEFT models analyzed in this paper. To improve, we exploit published LEC uncertainties from a previous analysis [34] and construct a diagonal mass matrix. We then draw ≈ 1000 samples, using $L = 8$ at LO and $L = 20$ at NLO and NNLO, to estimate a parameter covariance matrix $\Sigma_{\vec{\alpha}}$ and construct a mass matrix according to $\mathcal{M} = \Sigma_{\vec{\alpha}}^{-1}$. We find that this approach to learn about $\mathcal{M}$ yields high HMC performance in practice. Note that one does not have to use HMC for this tuning, as we do; indeed, it may be preferable to use a more expedient method for extracting an approximate parameter covariance matrix, e.g., the more tuning-agnostic MCMC sampler EMCEE.

Local estimates of the target covariance based on, e.g., optimization and second derivatives have been used in previous studies of LEC uncertainties (see, e.g., Ref. [14]). This method was recently used to construct a Bayesian prior for the χEFT contact LECs at NNLO when estimating the c_D and c_E LECs in the three-nucleon force sector [35]. It was found that the prior and marginal posterior were largely the same. This finding reinforces the observation that the inverse of a point-estimated covariance matrix yields a performant mass matrix.

IV. SAMPLING LEC POSTERIORS USING HAMILTONIAN MONTE CARLO

In Sec. IV A we outline the sampling strategy and in Sec. IV B we present our posterior PDFs $\text{pr}(\vec{\alpha}|D, I)$ for the LECs $\vec{\alpha}$ at LO, NLO, and NNLO in χEFT sampled using HMC. These posteriors enable all subsequent inference in this paper and constitute the main result of our paper. In Sec. IV C we discuss the convergence of the MCMC chains. In Sec. IV D we highlight some of the unique aspects of the HMC algorithm by comparing with posterior samples obtained using EMCEE. In Sec. IV E we comment on multimodality and the challenge it brings.

A. Sampling strategy

We employ the same HMC sampling strategy at all chiral orders considered in this paper.

(1) To identify a ballpark region where we expect to find the posterior mode we first optimize the data likelihood in Eq. (18). At this stage we employ $\bar{c} = 1$ and $y_{\text{ref}} = y_{\text{exp}}$ to parametrize the covariance matrix for the EFT truncation error. Every subsequent HMC sampling is then randomly initiated within an overdispersed region around the MLE.

(2) To tune the mass matrix $\mathcal{M}$, we use previously published uncertainties for the LECs [34] to define its diagonal entries and draw $n_{\text{tune}} \sim 1000$ samples using $L \approx 10$–20. The resulting sample covariance matrix is inverted to yield the final mass matrix.

TABLE III. Detailed statistics of the HMC chains during the tuning and sampling phases. The number of chains at each order is M and the total number of samples is counted across all M chains. The HMC parameters $\epsilon^\star$ and $L^\star$ denote the nominal step length and total number of steps taken with the leapfrog algorithm to integrate Hamilton's equations for each HMC step. $\bar{a}$ is the average acceptance rate.

Order	n_{tune}	M	Total number of samples	$\epsilon^\star$	$L^\star$	$\bar{a}$
LO	2000	3	50 063	0.1	8	99%
NLO	2000	10	57 134	0.09	20	99%
NNLO	591	3	10 155	0.08	20	99%

(3) We determine the optimal step size ϵ from a small set of very short HMC chains consisting of 10–20 samples and easily find a step size that yields an acceptance rate $\bar{a}$ of 99%.

(4) Equipped with a well-tuned HMC algorithm, we collect $M \geqslant 3$ independent chains at different starting values for the LECs. It is important to have two or more chains to enable canonical convergence tests based on within- and between-chain variances, e.g., the Gelman-Rubin test. We also employ a convergence criterion based on the integrated autocorrelation time as recommended in Refs. [33,36] (see Sec. IV C 2).

(5) The trace plots of the HMC chains indicate that the length of the burn-in phase is very short. This is corroborated by the autocorrelation analysis presented in Sec. IV C 2. We discard the initial $\approx$10 samples from each chain, except for three chains at NLO which require us to discard $\approx$100 samples. The situation is drastically different for the EMCEE chains, where we find it necessary to discard the initial $\approx$1000 samples at LO and $\approx$ 30 000–40 000 samples at the higher orders. These lengthy burn-ins have to be repeated for each EMCEE chain. In contrast, tuning the HMC hyperparameters is a one-time cost.

Detailed information about the tuning and sampling phases at LO–NNLO is summarized in Table III. Note that the number of tuning samples n_{tune} is larger than necessary at LO and NLO. The step sizes ϵ and acceptance rates $\bar{a}$, which are closely linked, are remarkably similar across the three chiral orders. The reason is that the step size is generally limited by the most constrained parameter(s) which in all three cases are the LO LECs $\widetilde{C}$.

B. LEC posteriors

The posterior PDFs for the LECs are multivariate but will be presented using univariate and bivariate projections. These so-called corner plots of the LO, NLO, and NNLO posteriors are shown in Figs. 1–3, respectively.

At all orders, the locations of the maximum *a posteriori* (MAP) probability and widths of the posterior PDFs are similar to the corresponding measures we obtained using frequentist parameter estimation in Ref. [14]. This is largely due to the fact that we are using nearly the same database of thousands of NN scattering cross sections in both analyses and that the inference is likelihood dominated. The main modification to the database comes from setting aside part of the data for validation in this paper. We also employ identically regulated χEFT interactions, and closely related diagonal covariance matrices for estimating uncorrelated EFT truncation errors. Despite several apparent similarities it is very important to realize that we are comparing results from two fundamentally

different approaches. The use of Bayesian inference methods allows us to assign a probability (density) measure to LEC values themselves. In the frequentist approach we are estimating covariances from the gradients at the maximum likelihood estimator of the *data*.

1. LO

At LO we consider the two NN contact LECs present at this order: $\widetilde{C}_{1S0}$ and $\widetilde{C}_{3S1}$, acting in the S waves. The corner plot in Fig. 1 reveals that they are both very well constrained by the NN scattering data D and appear to be uncorrelated with each other. We note that $\widetilde{C}_{1S0}$ is considerably more constrained than $\widetilde{C}_{3S1}$. This is likely due to (i) the isovector (isoscalar) character of $\widetilde{C}_{1S0}$ ($\widetilde{C}_{3S1}$) and (ii) the fact that pp data are more abundant and more precise than np data at low scattering energies where the truncation error is relatively small.
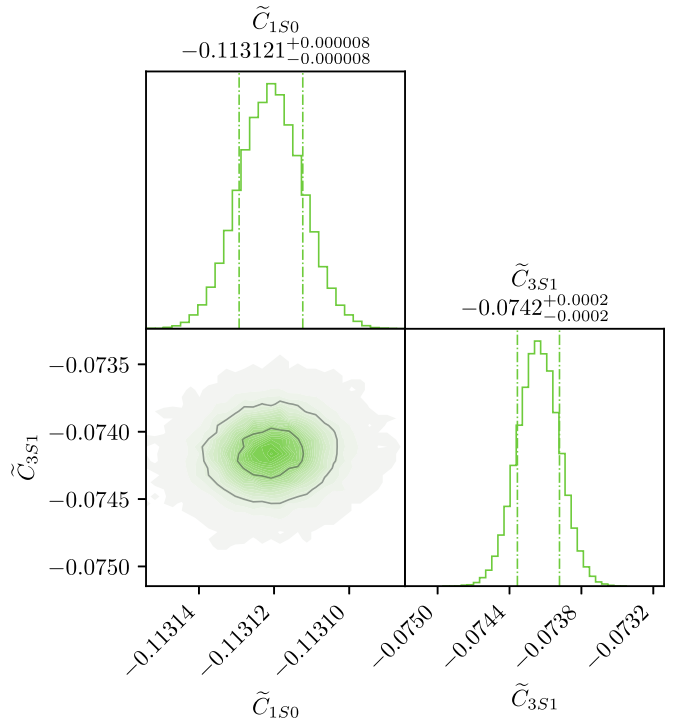

FIG. 1. LO posterior sampled with HMC. The LECs are shown in units of $10^4\,\text{GeV}^{-2}$. The inner (outer) gray contour line encloses 39% (86%) of the probability mass. The dot-dashed vertical lines indicate a 68% credibility interval in the univariate marginals. White areas indicate zero counts of samples.
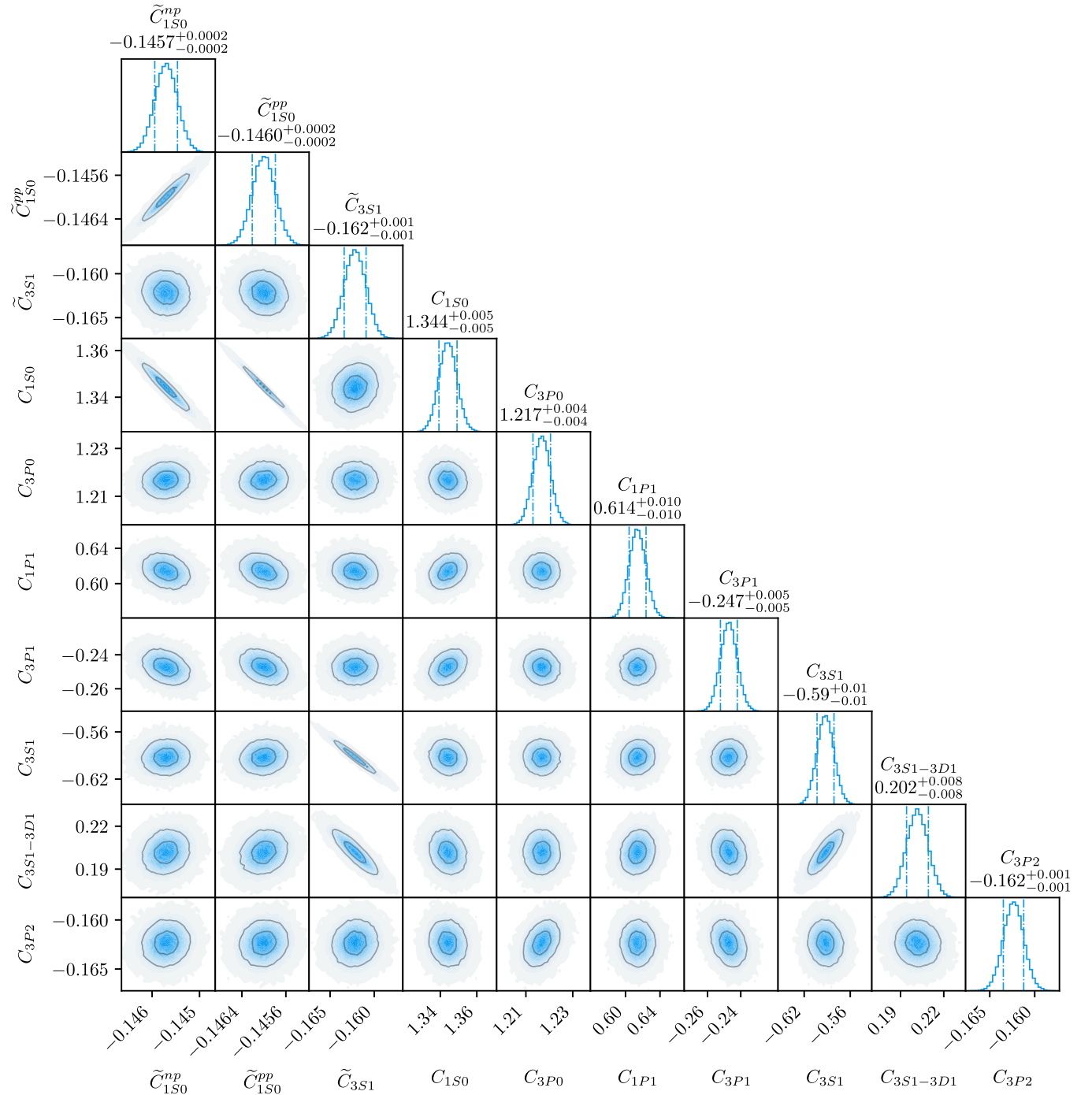
FIG. 2. NLO posterior sampled with HMC. The LECs are shown in units of $10^4\,\mathrm{GeV}^{-2}$ for the LO LECs and $10^4\,\mathrm{GeV}^{-4}$ for the NLO LECs. The inner (outer) gray contour line encloses 39% (86%) of the probability mass. The dot-dashed vertical lines indicate a 68% credibility interval in the univariate marginals.

2. NLO

Several contact LECs are introduced at NLO and we find that the LEC posterior exhibits noticeable correlations in certain directions (see Fig. 2). The presence of such correlations indicates a level of parameter redundancy in the model. From a statistical perspective there exist methods, e.g., singular value decomposition, to identify and retain only the most important parameters (or linear combinations of parameters) of a model to explain data, so-called stiff directions in the parameter space. However, before doing so it is worthwhile to inspect the model structure from a physics perspective. In the present case we identify a strong correlation between the LECs $\widetilde{C}_{1S0}^{pp}$ and $\widetilde{C}_{1S0}^{np}$. Following conventional counting of the isospin-breaking effects in χEFT we encounter the leading isospin-dependent $^1S_0$ contacts at NLO. This is also
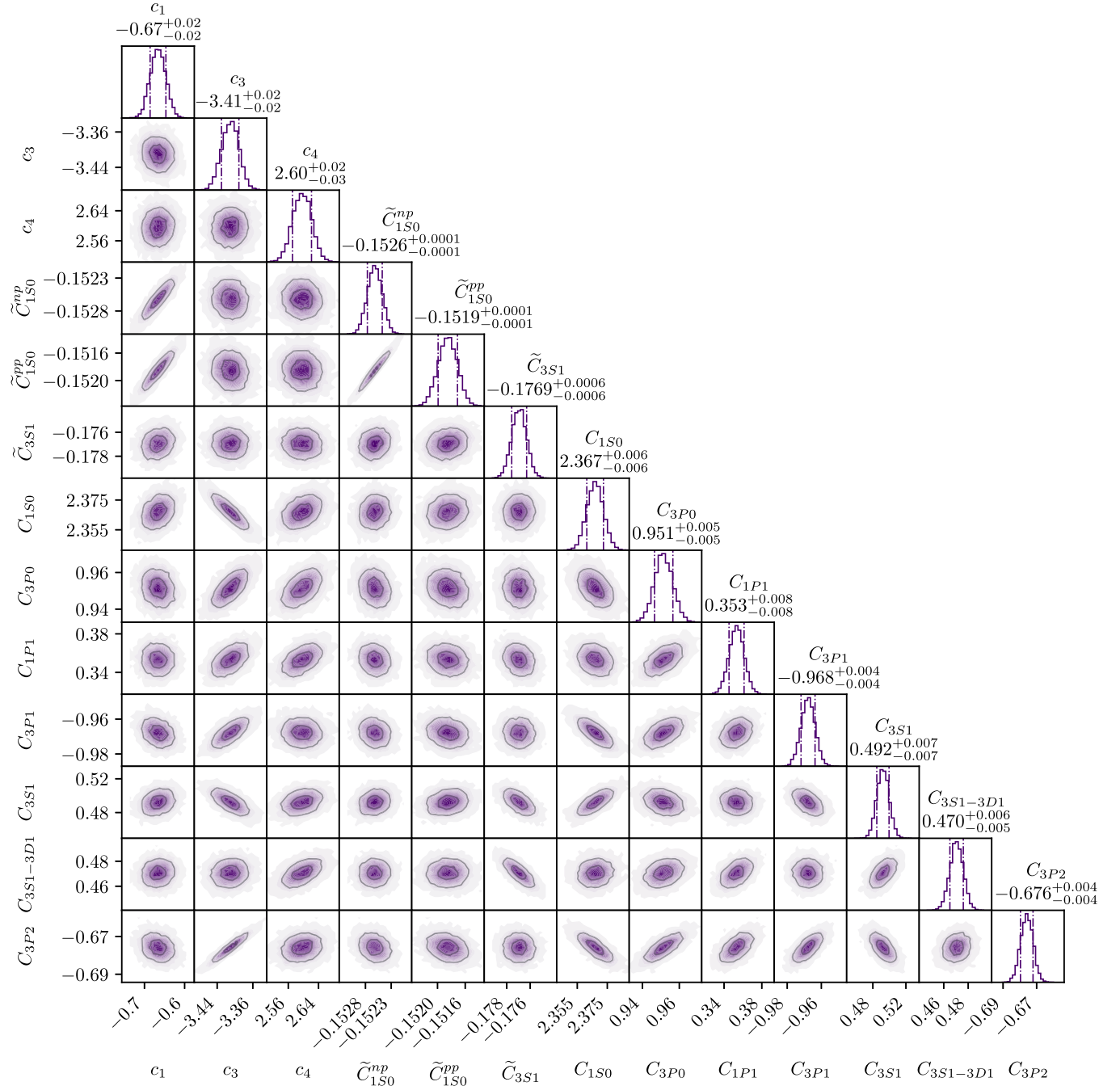
FIG. 3. NNLO posterior sampled with HMC. The LECs are shown in units of $10^4\,\mathrm{GeV}^{-2}$ for the LO contact LECs, $10^4\,\mathrm{GeV}^{-4}$ for the NLO contact LECs, and GeV^{-1} for the πN LECs. The inner (outer) gray contour line encloses 39% (86%) of the probability mass. The dot-dashed vertical lines indicate a 68% credibility interval in the univariate marginals.

in line with the results from a high-precision data analysis by the Nijmegen group [18] demonstrating that strong and electromagnetic interactions break charge independence and most prominently in the $^1S_0$ channel. However, since isospin breaking is a comparatively small effect, a non-negligible EFT truncation error at NLO is likely to dilute isospin sensitivity with respect to the NN data being used.

We also detect correlations between S-wave LECs acting within the same spin channel. In general, the spin-singlet and spin-triplet partial-wave contact LECs do not exhibit any

significant correlation with each other at any of the chiral orders we examine. This is somewhat different from the frequentist analysis in Ref. [14] where correlations ($|\rho| \gtrsim 0.7$) were found between all S-wave LECs and C_{3P2}. We speculate that this difference in correlation structure could be rooted in the difference between the models of the truncation error used in the frequentist and Bayesian analyses. Upon inspection, we find that the truncation errors at NLO and NNLO employed in this paper are more than twice as large compared to the corresponding error magnitudes used in Ref. [14]. We have
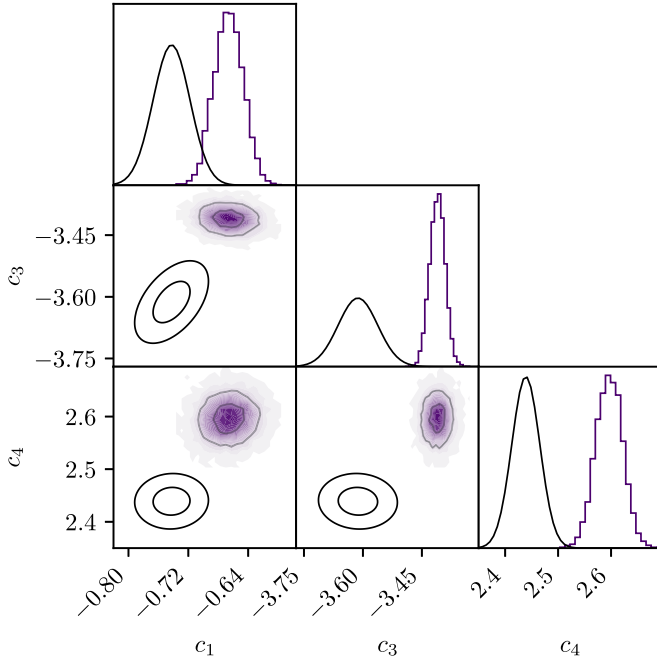
FIG. 4. Prior and posterior PDFs for the πN LECs c_1, c_3, and c_4, in units of GeV^{-1}, indicated with black-line ellipses and colored (purple jagged) regions, respectively. The inner (outer) black ellipses enclose 39% (86%) of the prior probability mass, and the jagged gray lines do the same for the posterior probability mass. The posteriors were obtained using HMC. See text and Fig. 3 for details.

not performed a systematic analysis to compare the two error models, but a smaller truncation error can certainly shift weights of, e.g., spin-polarization and spin-averaged data at different energies which in turn could induce stronger correlations between the LECs of the interaction model.

3. NNLO

At NNLO we also find strong correlations between certain LECs (see Fig. 3). Similarly to the NLO posterior, there is a very strong correlation between the charge-independence breaking contacts. However, we do not identify any unexpected correlation structures, like the ones found at N3LO in Ref. [11], to reveal a physics parameter redundancy in the model.

At this order we encounter the subleading πN LECs c_1, c_3, and c_4 for the first time. As opposed to contact LECs they act in all angular momentum channels. Certain combinations of πN and NN LECs also show significant correlations; in particular, c_1 appears positively correlated with the $\widetilde{C}_{1S0}$ LECs, as does c_3 with, e.g., C_{3P2}.

Recall that we assign a multivariate normal prior PDF for c_1, c_3, and c_4 using maximum-likelihood results from a Roy-Steiner analysis of the Δ-less πN scattering amplitudes [13]. This prior strongly regulates the values of c_1, c_3, and c_4 compared to the NN LECs which are assigned a less informative prior based on naturalness.

In Fig. 4 we show a corner plot where we focus on the differences between the prior and posterior in the πN sector. Note that we are comparing two PDFs that have very different

origin from both a statistical and methodological perspective. We can still draw the following conclusions.

(1) The NN data induce an overall 5–10% shift, in the positive direction, of the πN LECs, i.e., the NN data appear to reduce the πN subleading attraction slightly. However, the overall effect of this shift on the binding of atomic nuclei remains to be analyzed.

(2) The difference in MAP values of the πN prior and posterior is significant compared to the extent of the credible intervals. For c_1 we observe a slight overlap of the marginal PDFs. It is interesting to note that the πN vertex corresponding to c_1 does not contain any contribution from the Δ isobar.

(3) The link between low-energy πN and NN scattering processes is a hallmark of χEFT. If *all* uncertainties are accurately modeled, and we are operating with an EFT, then we expect to find an overlap between the prior and the posterior. However, we do not observe this. One possible explanation is that we estimate the truncation error in the NN sector via the uncorrelated theory covariance matrix in Eq. (17). The Roy-Steiner analysis is based on an MLE with uncorrelated data and method uncertainties [12,13] and it is not clear how to propagate an EFT error to the πN LECs matched at this order.

C. Convergence towards a stationary distribution

How many samples do we need to reach an accurate representation of the stationary target distribution with small sampling error? In connection with this, one should note that the N samples collected during some finite time period will not be independent.

It is unfortunately not possible to determine the level of convergence of a finite chain; we can only attempt to detect convergence failures. As such, all convergence diagnostics in the MCMC literature merely provide necessary but not sufficient conditions. Multiple diagnostics have been devised, of which we employ two of the most common ones: the standard Gelman-Rubin statistic ($\hat{R}$) [37,38], discussed in Sec. IV C 1, and the integrated autocorrelation time (τ) discussed in Sec. IV C 2. Both diagnostics are applied to each LEC α_i individually.

Although MCMC algorithms are ergodic and eventually explore the entire state space, we clearly face the challenge of pseudoconvergence due to multimodality when working with finite chains. This problem is discussed further in Sec. IV E.

1. The Gelman-Rubin statistic $\hat{R}$

With the Gelman-Rubin statistic $\hat{R}$ we compare the variance of the samples within a single chain to the variance between $M \geqslant 3$ chains initialized at different starting positions. Following Ref. [37], we assume that each chain contains N samples after we have discarded initial samples to reduce the memory of the starting position. We discussed the removal of the burn-in samples in Sec. IV A. Based on the $N \times M$ samples one defines a joint mean

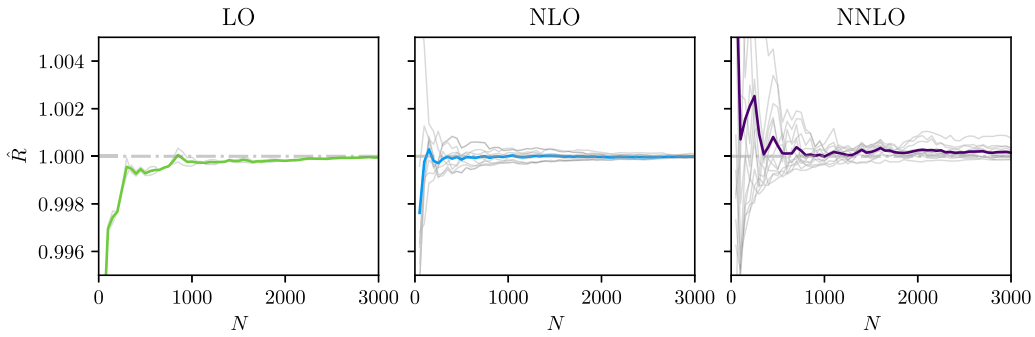$$\bar{\alpha}_i = \frac{1}{M}\sum_{m=1}^{M}\bar{\alpha}_i^{(m)} \tag{36}$$

FIG. 5. The Gelman-Rubin convergence diagnostics $\hat{R}$ for the HMC sampled chains at LO, NLO, and NNLO as a function of the number of samples N. $\hat{R}$ for each individual parameter is shown in gray, while the mean $\hat{R}$ is shown in green, blue, and purple, respectively.

for each LEC α_i based on all chains where $\bar{\alpha}_i^{(m)}$ denotes the within-chain mean for the mth chain and is given by

$$\bar{\alpha}_i^{(m)} = \frac{1}{N} \sum_{n=1}^{N} \alpha_i^{(nm)}. \tag{37}$$

In this notation, $\alpha_i^{(nm)}$ corresponds to the nth sample of the ith LEC in the mth MCMC chain. One can express the between-chain and within-chain variances for the ith LEC in terms of the corresponding means as

$$B_i = \frac{N}{M-1} \sum_{m=1}^{M} \left(\bar{\alpha}_i^{(m)} - \bar{\alpha}_i \right)^2 \tag{38}$$

and

$$W = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{N-1} \sum_{n=1}^{N} \left(\alpha_i^{(nm)} - \bar{\alpha}_i^{(m)} \right)^2, \tag{39}$$

respectively.

A weighted average of the above variances can be used to estimate the variance of the marginal posterior for the LEC α_i:

$$\mathrm{Var}^+[\alpha_i] = \frac{N-1}{N} W + \frac{1}{N} B \tag{40}$$

where the $+$ sign indicates that this quantity overestimates the posterior variance provided that the M chains are initialized at locations with greater variability compared to the true posterior. Indeed, for finite N we have that W will underestimate the marginal variance since the chains have not explored the posterior while B will overestimate the marginal variance if the initial sampling distribution is overdispersed. In the limit of $N \to \infty$ we will have that W approaches the variance of the marginal posterior. Incorporating these finite-N corrections leads to a Student's t PDF for α_i with variance (scale) estimated by

$$V = \mathrm{Var}^+[\alpha_i] + \frac{B}{MN}. \tag{41}$$

The Gelman-Rubin measure expresses the potential scale reduction by forming the ratio

$$\hat{R} = \sqrt{\frac{V}{W}}, \tag{42}$$

which approaches 1 as $N \to \infty$. A widely used threshold for declaring convergence is $\hat{R} < 1.01$ [39]. This, somewhat arbitrary, threshold simply states that the sample variance is 2% larger than the within-chain variance, and that one should expect a corresponding potential scale reduction if continuing the sampling process. In Fig. 5 we show the evolution of $\hat{R}$ with the number of HMC samples at LO, NLO, and NNLO. Clearly, our HMC chains fulfill $\hat{R} < 1.01$ as well as an even stricter threshold $\hat{R} < 1.001$. The chains obtained with EMCEE also pass the same $\hat{R}$ thresholds after a similar amount of MCMC samples.

There exist updated Gelman-Rubin measures. One can employ so-called split $\hat{R}$ [39] and rank-normalized $\hat{R}$ [40] to better handle nonstationary chains and chains distributed with a heavier tail that conspire to yield a good $\hat{R}$. We have not detected the need for using such updated $\hat{R}$ measures to analyze the χEFT posteriors in this paper.

2. The integrated autocorrelation time τ

We use the MCMC chains to compute (statistical) expectation values and we should therefore also analyze the sampling variance of such estimates. For example, one can straightforwardly estimate the mean value $\bar{\alpha}_i$ of the ith LEC within a single MCMC chain[3] [see Eq. (37)]. The sampling variance of the estimated mean value for a particular LEC, based on N uncorrelated samples, scales with $1/N$ according to

$$\mathrm{Var}[\bar{\alpha}_i] \equiv \mathrm{E}[\bar{\alpha}_i - \mathrm{E}(\alpha_i)] = \frac{\mathrm{Var}[\alpha_i]}{N}, \tag{43}$$

where $\mathrm{Var}[\alpha_i]$ is the variance of the samples with respect to the posterior $\mathrm{pr}(\alpha_i|D, I)$. This estimate only holds for uncorrelated samples, and we will demonstrate that such samples can be obtained with the HMC algorithm. In contrast, most random-walk based MCMC algorithms generate highly correlated samples. When the samples are correlated, the variance of the mean is modified according to

$$\mathrm{Var}[\bar{\alpha}_i] = \tau_i \frac{\mathrm{Var}[\alpha_i]}{N} \tag{44}$$

---

[3]In this section we discuss quantities pertaining to a single chain and therefore omit the superscript (m).
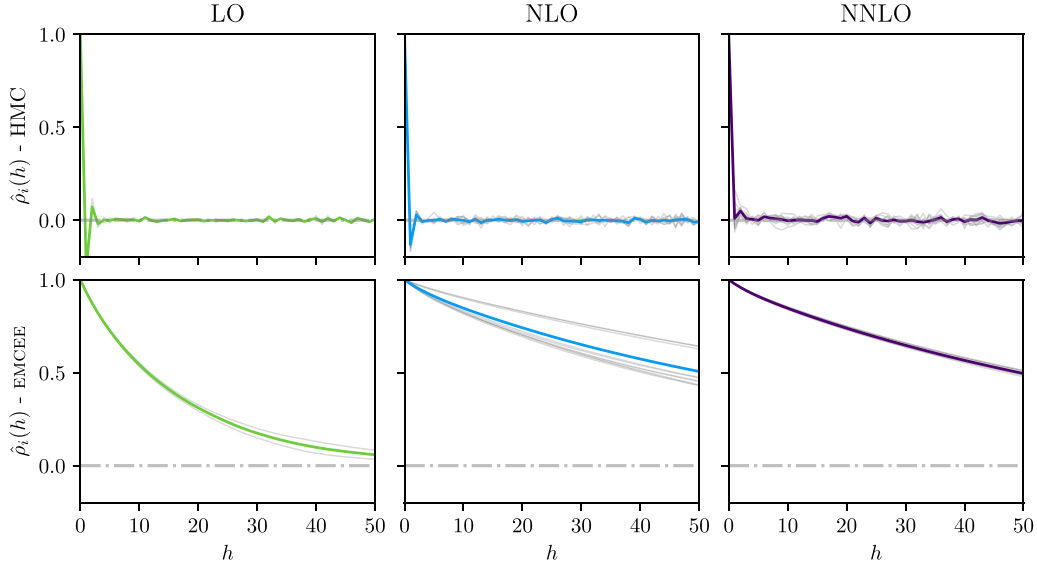
FIG. 6. Autocorrelation functions of the MCMC chains at LO-NNLO. The chains produced with HMC are shown in the first row and chains produced with EMCEE are shown in the second row. The gray lines show autocorrelations in individual parameters whereas the colored lines show the average over all parameters. The results are averaged over all walkers in the plots showing the EMCEE autocorrelations. See Table III for details about each sampling.

where τ_i is referred to as the integrated autocorrelation time for the chain of sample values of the ith LEC. It is given by

$$\tau_i = \lim_{N \to \infty} \left(1 + 2 \sum_{h=1}^{N} \rho_i(h)\right). \tag{45}$$

The autocorrelation function $\rho_i(h)$ measures the correlation between (stationary) samples separated by h MCMC steps. In the literature, h is referred to as the *lag*. Note that the integrated autocorrelation time will be different for each expectation value. Here, we limit ourselves to inspect τ_i for the mean values of the LECs and use this below to assess the convergence of an MCMC chain. In Fig. 6 we present the estimated autocorrelation functions $\hat{\rho}_i(h)$ of all LECs at LO, NLO, and NNLO, and their averages, as obtained by HMC and EMCEE. As expected, a well-tuned HMC algorithm generates virtually uncorrelated samples whereas the EMCEE chains exhibit a correlation structure that is typical for most MCMC algorithms. The HMC algorithm generates uncorrelated samples even as the dimensionality of the parameter space is increased. We see this advantageous dimensionality scaling of HMC when going from LO to NNLO. The corresponding correlation lengths of the EMCEE chains markedly increase. This is one of the primary advantages of using HMC.

Some care is needed in the numerical computation of the integrated autocorrelation time τ_i. For large values of N in Eq. (45), the estimated autocorrelation $\hat{\rho}$ suffers from a signal-to-noise problem. Indeed, although the correlation decreases towards zero with increasing lag, the variance of the correlation does not. Following Ref. [36] we therefore truncate the sum at the smallest integer $N^\star$ such that $N^\star \geqslant c\tau_i(N^\star)$ for $c = 5$. With this in place we can monitor the evolution and convergence of τ as a function of the number of collected samples N in the chain.

As mentioned, the computation of τ is not only useful for quantifying the sampling variance but also provides a handle on the convergence of the MCMC chain. While the Gelman-Rubin statistic compares several identically prepared chains, diagnosing convergence based on the evolution of τ can be applied to a single chain. Using this method, convergence is declared when the estimation of τ has stabilized and $N \gg \tau$, where N is the length of the chain. In this paper we apply the condition $N \geqslant 50\tau$, and in Fig. 7 we show the evolution of
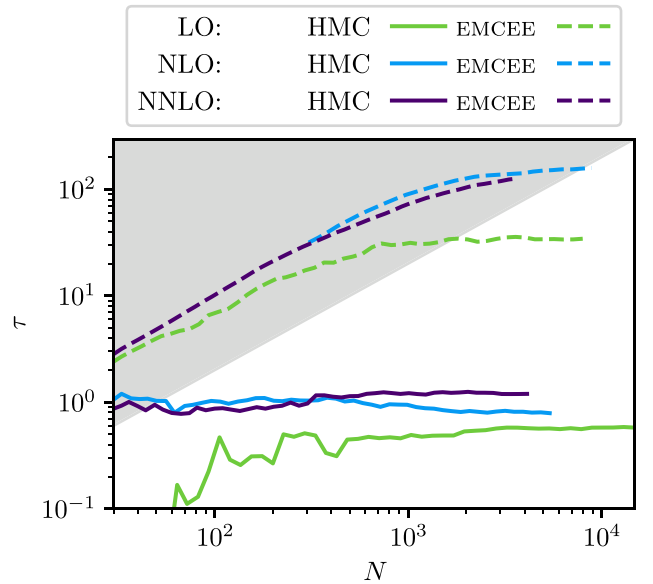

FIG. 7. Integrated autocorrelation time τ vs number of samples. The gray area indicates the zone of nonconvergence, i.e., $N < 50\tau$. The EMCEE results are averaged over all walkers, hence the seemingly low number of samples compared to Table IV.

TABLE IV. Comparison between the performance of EMCEE and HMC applied to sample the LEC posteriors at LO, NLO, and NNLO. τ is the integrated autocorrelation time. ESS is the effective sample size and N is the number of collected MCMC samples. Consequently, the ESS/N column shows how many effective samples one nominal sample is worth. The $N_{\mathcal{L}}/N$ column, where $N_{\mathcal{L}}$ is the total number of likelihood calls, shows the average number of likelihood evaluations (including tuning and burn-in) necessary to collect one nominal sample. $\mathcal{S}$ is the average real-world speedup of HMC compared to EMCEE as defined in Eq. (47).

Algorithm	Order	τ	ESS/N	$N_{\mathcal{L}}/N$	$N_{\mathcal{L}}$/ESS	AD-cost	$\mathcal{S}$
EMCEE	LO	34	0.029	1.01	35		
HMC	LO	0.58	1.7	9.07	5.3	1.10	6.0
EMCEE	NLO	158	0.0063	1.07	169		
HMC	NLO	0.78	1.3	27.5	21	1.24	6.4
EMCEE	NNLO	>126	<0.007	1.16	>146		
HMC	NNLO	1.2	0.85	23.9	28	1.43	>3.6

τ as N increases. We have also indicated the nonconvergence zone $N < 50\tau$. At LO and NLO, we fulfill τ convergence using both HMC and EMCEE. At NNLO, we fulfill τ convergence only with HMC while EMCEE falls just short of the imposed tolerance.

D. Effective sample size and efficiency of HMC

The reduction in sample quality due to correlation is often quantified with the *effective sample size*

$$\text{ESS} = \frac{N}{\tau}, \tag{46}$$

where we also introduce τ defined as the average of all τ_i for each LEC α_i. We use the ESS value to quantitatively compare the efficiencies of HMC and EMCEE. It is obvious from Eqs. (44) and (46) that the correlation structure directly impacts the ESS and hence the total computational effort required to reach a tolerable variance. Since there is a significant computational overhead involved in integrating Hamilton's equations, it is key to tune the HMC algorithm to reach a very small τ. The intrinsic benefit of using HMC also increases with the dimensionality of the parameter space.

In Table IV we present τ values and relevant related quantities for EMCEE and HMC at LO, NLO, and NNLO. The results presented in this table are based on a single chain for each order and choice of algorithm. We have, however, verified that the results are similar regardless of which one of the parallel chains is used. We note that we have $\tau \approx 1$ for the HMC sampled chains, and even $\tau < 1$ in two cases (LO and NLO). With EMCEE, τ is roughly two orders of magnitude greater than the HMC equivalents. This is solely due to the correlation structure of the respective MCMC chains. Also shown in Table IV (as ESS/N) is the average number of effective samples that one MCMC sample provides. By introducing $N_{\mathcal{L}}$—the total number of likelihood calls during sampling, tuning, and burn-in—we also show (as $N_{\mathcal{L}}/N$) the average number of calls to the likelihood function required to generate one MCMC sample. We obtain $N_{\mathcal{L}}/N > 1$ also for EMCEE since the quoted results also include burn-in. We can use the number of likelihood calls per effective sample to compare the average efficiency of HMC and EMCEE, as evaluating the likelihood constitutes nearly all of the computational effort. We therefore

define an HMC speedup factor $\mathcal{S}$ with respect to EMCEE as

$$\mathcal{S} = \frac{[N_{\mathcal{L}}/\text{ESS}]_{\text{EMCEE}}}{[N_{\mathcal{L}}/\text{ESS}]_{\text{HMC}} \times [\text{AD-cost}]}, \tag{47}$$

where we also account for the computational overhead induced by the use of AD which we employ to generate the gradients necessary for HMC. In our implementation we measure the AD overheads to 10% at LO, 24% at NLO, and 43% at NNLO, and we use these figures when quantifying the HMC speedup factors in Table IV. In summary we find that the real-world speedup factor is more than 6 at LO and NLO, and 3.6 at NNLO. The smaller speedup at NNLO is primarily due to less ideal tuning, and could be improved further. Another contributing factor is that the estimate of τ for EMCEE at NNLO has not stabilized, and the value reported in Table IV is therefore a lower bound (see Fig. 7).

We obtain $\tau < 1$ at LO and NLO when using HMC, a result which comes from drawing anticorrelated samples from the posterior PDF, as can be seen clearly in the corresponding autocorrelation functions in Fig. 6. This so-called antithetic sampling [41] leads to ESS > N, i.e., an effective number of samples greater than the number of MCMC samples with a corresponding reduction of the variance in Eq. (44). Therefore, drawing completely independent MCMC samples is not necessarily the optimal strategy. In this paper we encounter antithetic sampling for sufficiently large values of the HMC step length L once we have constructed a mass matrix $\mathcal{M}$ that we believe suits the target distribution. This is obviously an advantageous sampling strategy and antithetic sampling is one of several known variance reduction techniques in Monte Carlo sampling [23]. While we do not achieve antithetic sampling at NNLO in this paper, we believe that it can be reached with further tuning.

1. Consequences of improper tuning of the HMC hyperparameters

HMC sampling is challenging in practice, primarily due to the need for careful tuning of the mass matrix $\mathcal{M}$ to achieve high performance. It is therefore instructive to show what a failure looks like. Figure 8 demonstrates the strong autocorrelation that results when $\mathcal{M}$ is not well tuned for two chains at NLO—one where $\mathcal{M}$ is set up using a χEFT naturalness argument, and the other where it is based on previously published LEC uncertainties [34]. To construct the mass
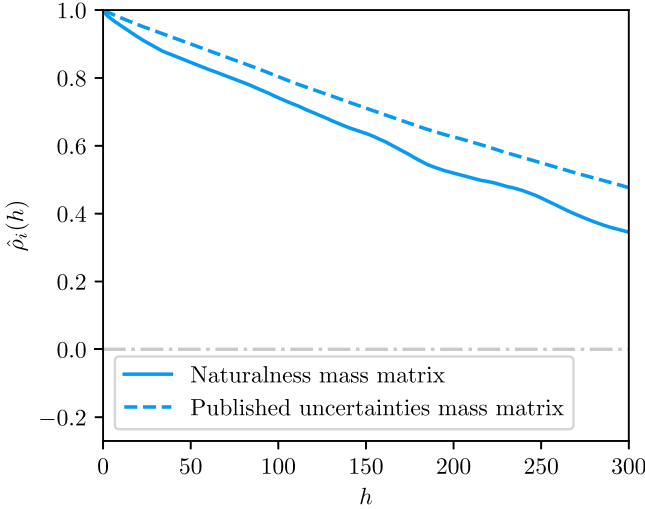
FIG. 8. Averaged autocorrelations at NLO achieved with the "naturalness" and "published uncertainties" mass matrices.

matrix using the naturalness argument we employed ratios of the contact LECs at LO–NNLO according to

$$|\widetilde{C}_i| \sim \frac{4\pi}{F_\pi^2}, \quad |C_i| \sim \frac{4\pi}{F_\pi^2 \Lambda_b^2} \tag{48}$$

where $F_\pi \simeq 92$ MeV is the pion decay constant. In both cases, the integrated autocorrelation time is very large, which, combined with the high per-sample computational cost of HMC, results in very poor performance. Autocorrelation structures like those shown in Fig. 8 are unacceptable of course, but typically seen before tuning the hyperparameters.

E. Multimodality

Although the Markov chain is ergodic, one would have to run the MCMC algorithm for an infinitely long time to visit all states. Clearly, multimodal distributions make the sampling process considerably more complicated. The Markov chain has to cross a valley of low probability to explore more than one mode. Such a crossing is naturally a low probability event and will generally occur infrequently. Different modes may also have different shapes, causing poor performance if the tuning of the MCMC algorithm is unsuitable for the other mode even if the algorithm successfully moves between modes. Standard HMC has no advantage over alternative sampling algorithms in this regard due to its single-walker nature combined with the necessity of careful problem-specific tuning. Although the situation can be improved with, e.g., tempering methods [23] it is probably better to explore other MCMC algorithms, such as MULTINEST [42], if multimodality is expected.

In our analysis we encountered one clear case of multimodality: the LEC posterior at LO. This PDF is straightforward to explore in detail because it is only two-dimensional. We performed a scan on a 500×500 grid of the LO PDF. The result is shown in Fig. 9. The main mode is located around $(\widetilde{C}_{1S0}, \widetilde{C}_{3S1}) \approx (-0.11, -0.07) \times 10^4 \, \text{GeV}^{-2}$, and marked with a cross in the lower panel of Fig. 9. A
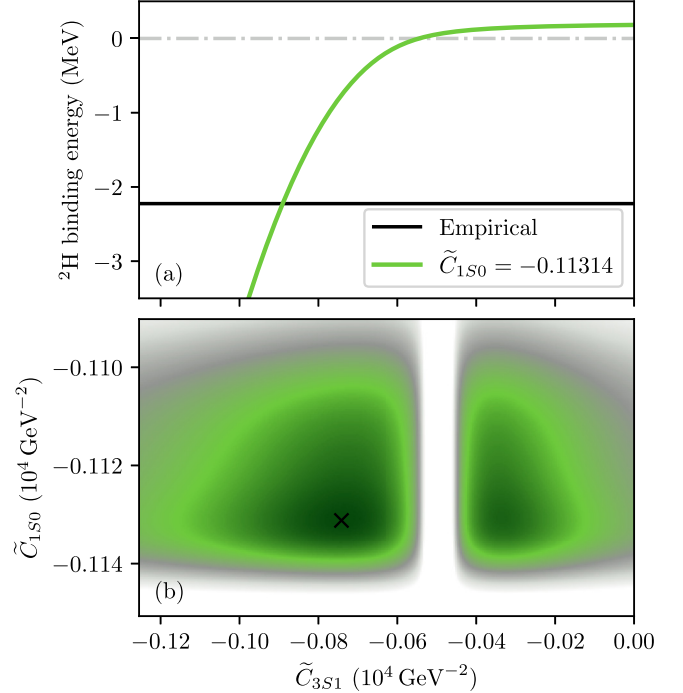

FIG. 9. (a) ^{2}H binding energy as a function of $\widetilde{C}_{3S1}$ with $\widetilde{C}_{1S0} = -0.113\,14 \times 10^4 \, \text{GeV}^{-2}$. (b) LO posterior evaluated on a lattice. The black cross marks the MAP point indicated on the diagonal in Fig. 1, i.e., $(\widetilde{C}_{1S0}, \widetilde{C}_{3S1}) = (-0.113\,12, -0.074\,16) \times 10^4 \, \text{GeV}^{-2}$.

second mode is found at $(\widetilde{C}_{1S0}, \widetilde{C}_{3S1}) \approx (-0.11, -0.03) \times 10^4 \, \text{GeV}^{-2}$. We computed the marginal likelihood $\text{pr}(D|I)$ of both modes using the Laplace approximation, and found that the second mode contains a negligible probability mass. A deep valley—many orders of magnitude lower in probability—separates the two modes and presents an effectively impenetrable barrier to the HMC sampler. The valley is caused by the breakup of the deuteron bound state, at $\widetilde{C}_{3S1} \approx -0.05 \times 10^4 \, \text{GeV}^{-2}$, that suppresses the likelihood for np SGT data at very low energies. As far as we can tell the vast majority of the probability mass for the LEC posteriors at NLO and NNLO is located in a single dominant mode. In the current paper we therefore proceed under the assumption that all PDFs are unimodal.

V. MODEL CHECKING

A probabilistic model of a physical system can only be upheld if it provides an acceptable representation of data. Therefore, we should always check to what extent the model fits the data. To that end we sample the PPDs and inspect the empirical coverages.

A. The posterior predictive distribution

As presented in Sec. II B, we have reserved roughly one third of the Granada database of experimentally measured scattering cross sections for model validation and refer to this as data set $\widetilde{D}$ (see Table I).

Following Eq. (10) we can model the true value of an NN scattering observable as the independent sum of the predicted

value up to chiral order k and the truncation error, i.e.,

$$y_{\text{true}} = y_{\text{th}}^{(k)} + \delta y_{\text{th}}^{(k)}. \tag{49}$$

Neither term on the right hand side is known with certainty. Indeed, they are stochastic variables described by PDFs. We consider the prediction $y_{\text{th}}^{(k)}$ uncertain due to the uncertainty of the LECs, and we consider $\delta y_{\text{th}}^{(k)}$ uncertain due to the unknown EFT expansion coefficients c_ν for $\nu > k$. Additional uncertainties regarding, e.g., the breakdown scale Λ_b, the expansion parameter Q, and the correlation structure of the model predictions $y_{\text{th}}^{(k)}$ can be accounted for as well [35]. In this paper, however, we focus on quantifying the LEC uncertainty and combining this with the EFT truncation error.

Equipped with an HMC chain of samples from the LEC posterior $\text{pr}(\vec{\alpha}|D, I)$ at chiral order k, it is straightforward to evaluate the PPD $\text{pr}(y_{\text{th}}^{(k)}|D, I)$ by observing that

$$\text{pr}(y_{\text{th}}^{(k)}|D, I) = \int \text{pr}(y_{\text{th}}^{(k)}, \vec{\alpha}|D, I) \, d\vec{\alpha}$$

$$= \int \text{pr}(y_{\text{th}}^{(k)}|\vec{\alpha}, D, I) \text{pr}(\vec{\alpha}|D, I) \, d\vec{\alpha}$$

$$= \int \text{pr}(y_{\text{th}}^{(k)}|\vec{\alpha}, I) \text{pr}(\vec{\alpha}|D, I) \, d\vec{\alpha}. \tag{50}$$

The last step is a consequence of the conditional independence between $y_{\text{th}}^{(k)}$ and D given $\vec{\alpha}$. Drawing random samples from this PPD amounts to evaluating $y_{\text{th}}^{(k)}$ for each of the N samples $(\vec{a}^{(1)}, \vec{a}^{(2)}, \ldots, \vec{a}^{(N)})$ in the HMC chain. Fortunately, the HMC chains are comparatively short and the necessary computation of NN scattering observables does not pose any significant challenge. Should this become an issue one could try emulating the observable response [43–45] or use hardware acceleration [46]. Opting for emulation or acceleration will add an error term to Eq. (49) quantifying the corresponding additional uncertainty.

Next, we sample the PDF for the truncation error in Eq. (15). This is trivial for a normally distributed EFT error where the parameters $\bar{c}^2$ and y_{ref} characterize this PDF entirely.

Since we can evaluate the two terms in Eq. (49) we can also draw samples from $\text{pr}(y_{\text{true}}|D, I)$. Panels (a) and (b) in Fig. 10 show such predictions for the true value of np SGT for $0 < T_{\text{lab}} \leqslant 350$ MeV, while panels (c) and (d) show predictions of the true value for the pp spin correlation parameter (AYY) at $T_{\text{lab}} = 294.4$ MeV. These observables were not included in the training data set D. Also shown in the figure are experimental measurements of the same observables, gathered from Ref. [16] for SGT and Ref. [47] for AYY. The panels in the left column show 100 individual predictions at each order, while the panels in the right column show 68 and 95% highest density intervals (HDIs)[4] computed from 1000 predictions. We only find unimodal PPDs and they appear to be rather symmetric. Indeed, as will be discussed below, the PPDs are

---

[4] For a unimodal PDF, a p% HDI will be the smallest interval around the mode of the PDF that comprises p% of its probability mass.

dominated by the normally distributed EFT truncation error. The individual predictions were generated in an uncorrelated fashion to reflect our model for the EFT truncation error. See the Supplemental Material [48] for PPDs at LO, NLO, and NNLO of all observables present in the Granada database, i.e., a model check with respect to both the training data D and the validation data $\widetilde{D}$.

The predictions in Fig. 10 appear to converge toward the experimental results with increasing chiral order, and a visual inspection of the 68 and 95% HDIs of the PPDs indicates that they work as advertised for these observables. We note that the employed model for the truncation error does not incorporate the known symmetry constraints of the spin-scattering matrix [49], e.g., that the vector analyzing power P goes to zero at extreme scattering angles. This type of information can be straightforwardly incorporated as a boundary condition on a Gaussian process model for the EFT truncation error [50].

We find that the truncation error $\delta y_{\text{th}}^{(k)}$ is the dominating source of uncertainty in Eq. (49). Indeed, the propagated error due to the variability in the LECs is quite small in comparison since the LEC posteriors are conditioned on a very large and informative data set. In fact, we find that the truncation error dominates at all orders up to NNLO for all energies in the Granada database. Even at low energies ($T_{\text{lab}} \approx 1$ MeV) the truncation error is still five to ten times larger than the uncertainty stemming from the LECs. Of course, at low energies both errors are very small on an absolute scale. In accordance with EFT principles we would expect the HDIs to narrow with each order and widen with increasing scattering energy. Instead, we find that the widths of the HDIs at all orders are comparable, and particularly so at the larger T_{lab} values. This is a consequence of underestimating $\bar{c}$ at LO and NLO (see Table II). We remind the reader that we estimate the EFT truncation at order k by exploiting information only up to this order since this reflects what one can do in a real situation without any higher orders available. Of course, at LO the available information is particularly scarce. In Sec. V B we will explore other choices for estimating $\bar{c}$.

B. Empirical coverage probability

We compute the empirical coverage probability to assess whether our HDIs are accurate as advertised. The empirical coverage quantifies how well we meet the expectation that if we compute a $p \times 100$% HDI for the PPD of a true value for an observable, we should find that this HDI covers the measurement of said observable with probability p on average. To that end, we perform a binary test for each datum in our validation data set $\widetilde{D}$ and count the number of times the validation data fall within the specified HDI, i.e., we count the number of "hits," and compare this with the total number of data in $\widetilde{D}$. In this procedure we neglect the (generally) very small uncertainties in the experimental data. Repeating the binary coverage test for a range of values of p, i.e., for different HDI intervals, yields a summary for the empirical coverage probability that is also convenient to inspect graphically.

Figure 11 shows the resulting empirical coverage plots for the validation data set $\widetilde{D}$ using three different strategies for
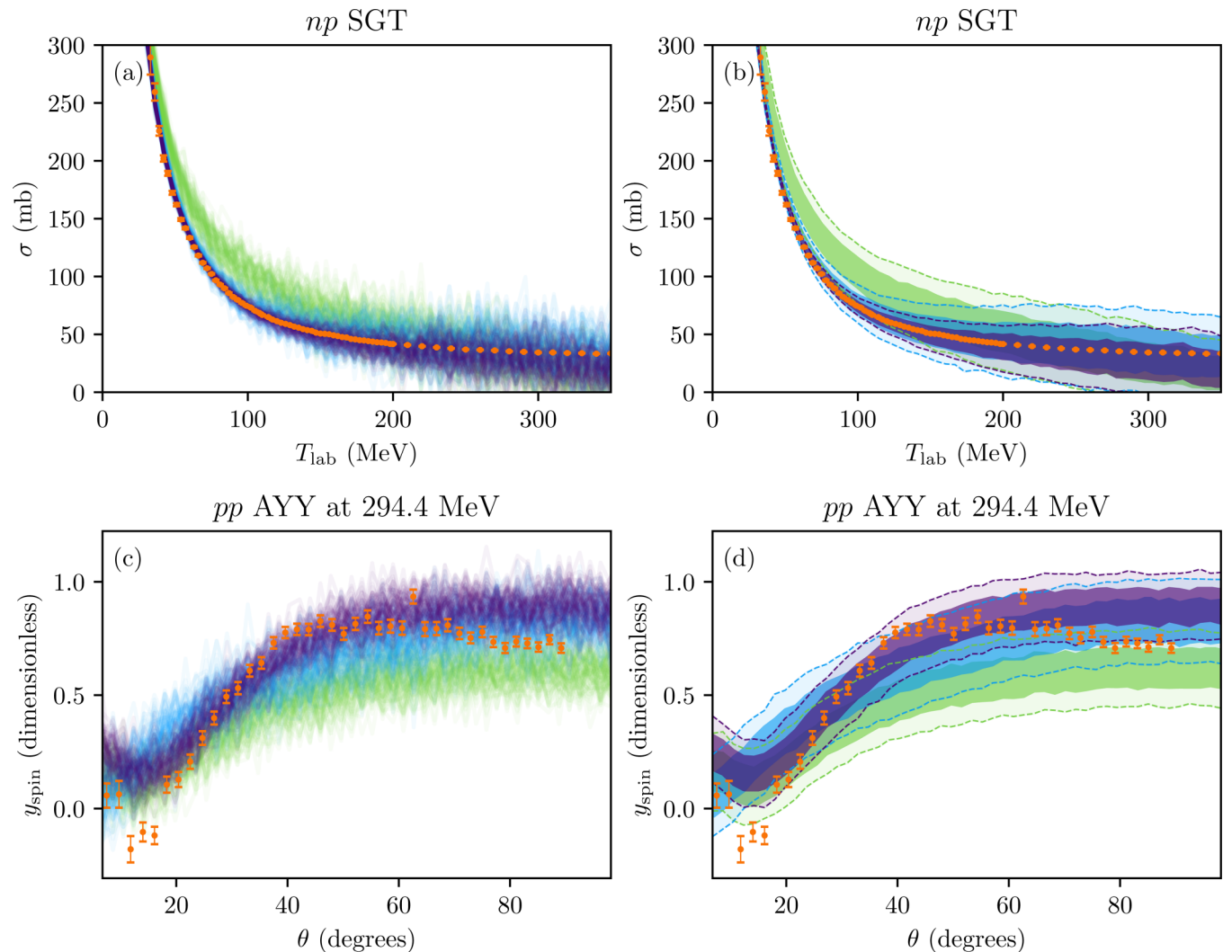
FIG. 10. Posterior predictive distributions for the true value of two *NN* scattering observables at LO (green), NLO (blue), and NNLO (purple). (a) Uncorrelated samples of $\mathrm{pr}(y_{\mathrm{true}}|D, I)$ for the *np* total cross section. Empirical data with error bars from Ref. [16] are shown in orange. (b) 68% (dark shaded regions bounded by solid lines) and 95% (light shaded regions bounded by dashed lines) HDIs of the PPDs in (a). (c) Uncorrelated samples of $\mathrm{pr}(y_{\mathrm{true}}|D, I)$ for *pp* AYY at $T_{\mathrm{lab}} = 294.4$ MeV. Empirical data with error bars from Ref. [47]. (d) 68 and 95% HDIs of the PPDs in (c).

evaluating an rms value for $\bar{c}$. In coverage plots of this kind, a $p \times 100\%$ HDI should yield a coverage such that it ends up on the 45° positive diagonal if it is working as expected. The diagonal line is indicated with a dash-dotted line in all our coverage plots. A coverage probability larger than this nominal value, i.e., a coverage probability above the diagonal, corresponds to an overly wide HDI, implying a too conservative error. The opposite situation corresponds to an overly narrow (liberal) HDI, implying a too small assigned error. Following a precautionary principle, a conservative error is preferable to the underestimated one.

The observed number of hits should also follow a binomial distribution under the assumption that the observables are uncorrelated (see, e.g., Ref. [5]). Our validation set, like the training set, contains hundreds of largely independent data groups that contain data recorded at different experimental facilities over several decades. The continuous version of the

binomial distribution is the β distribution, so we use this to compute 95% confidence intervals, indicated as a gray filled region along the diagonal in all coverage plots.

The truncation error is intimately linked to the inferred value for $\bar{c}$. In Fig. 11(a) we estimate $\bar{c}$ by the rms value $\bar{c}_k$ according to Table II. This method entails that we use information from all orders up to the one we are working at, i.e., $\bar{c}_k = \mathrm{rms}(\bar{c}_{0...k})$. Clearly, this yields liberal HDIs for the PPDs on average, in particular at LO and NNLO. Note that the LO HDI is based on an EFT truncation error essentially governed by the choice of reference values. At NLO and NNLO, the HDIs represent credible intervals conditioned on information from one and two order-by-order differences, respectively, and the corresponding coverage probabilities perform slightly better. However, we clearly underestimate $\bar{c}$.

As a second approach to determine $\bar{c}$, which we denote $\mathrm{rms}(\bar{c}_{2...k})$, we ignore the $\bar{c}_{\nu=0}$ contribution. In this case
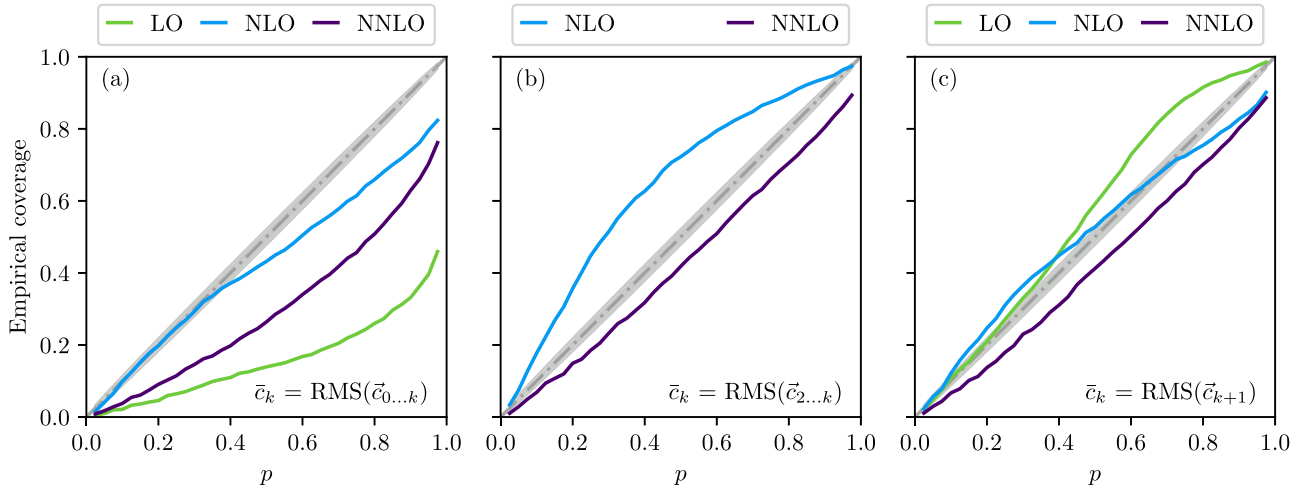
FIG. 11. Empirical coverage probabilities for $p \times 100\%$ HDI credible intervals with respect to the validation data set $\widetilde{D}$. Each panel corresponds to a different method of computing $\bar{c}$. The gray, filled regions indicate 95% confidence intervals that an observed empirical coverage is consistent with a true success rate p. (a) $\bar{c}_k = \mathrm{rms}(\vec{c}_{0...k})$. (b) $\bar{c}_k = \mathrm{rms}(\vec{c}_{2...k})$. (c) $\bar{c}_k = \mathrm{rms}(\vec{c}_{k+1})$. The computed $\bar{c}$ values are (a) $\bar{c}_{k=0} = 1.17$, $\bar{c}_{k=2} = 2.08$, $\bar{c}_{k=3} = 2.72$; (b) $\bar{c}_{k=2} = 4.95$, $\bar{c}_{k=3} = 4.19$; and (c) $\bar{c}_{k=0} = 4.95$, $\bar{c}_{k=2} = 2.84$, $\bar{c}_{k=3} = 4.12$.

the empirical coverage probability improves drastically [see Fig. 11(b)]. Motivated by the dominance of the EFT truncation error compared to the uncertainty originating from LEC variability, we do not resample the LEC posteriors when varying the hyperparameters of the truncation error. The LO predictions are absent in Fig. 11(b) to emphasize that this method for computing $\bar{c}_k$ is not relevant at that order.

In Fig. 11(c) we compute $\bar{c}$ as the rms value of the expansion coefficients extracted with respect to the first omitted order only, which we denote $\mathrm{rms}(\vec{c}_{k+1})$. At NNLO we exploit an MLE at N3LO. As expected, this method improves the performance of our error model, in particular at LO and NLO. There is not much difference between the coverage probabilities for the NNLO HDI in panels (b) and (c).

In Fig. 12 we inspect the coverage probability with respect to three different subsets of the validation data set $\widetilde{D}$. In panel (a), we only look at integrated np cross-section data (np SGT). In (b), we look at all data with $T_{\mathrm{lab}} \leqslant 100$ MeV, and in (c) we look at all data with $T_{\mathrm{lab}} > 290$ MeV. Here we use $\bar{c}_k = \mathrm{rms}(\vec{c}_{2...k})$. The coverage with respect to SGT data is consistently too high. The data points in this set are very correlated with each other, which makes this comparison less meaningful. In panel (b), where we retain validation data with $T_{\mathrm{lab}} \leqslant 100$ MeV, the error model at NLO performs rather well, whereas the NNLO error is too liberal. The coverage with respect to data with $T_{\mathrm{lab}} > 290$ MeV, panel (c), exhibits a similar pattern. Although the HDIs improve with information from higher orders, as expected, this is not an entirely
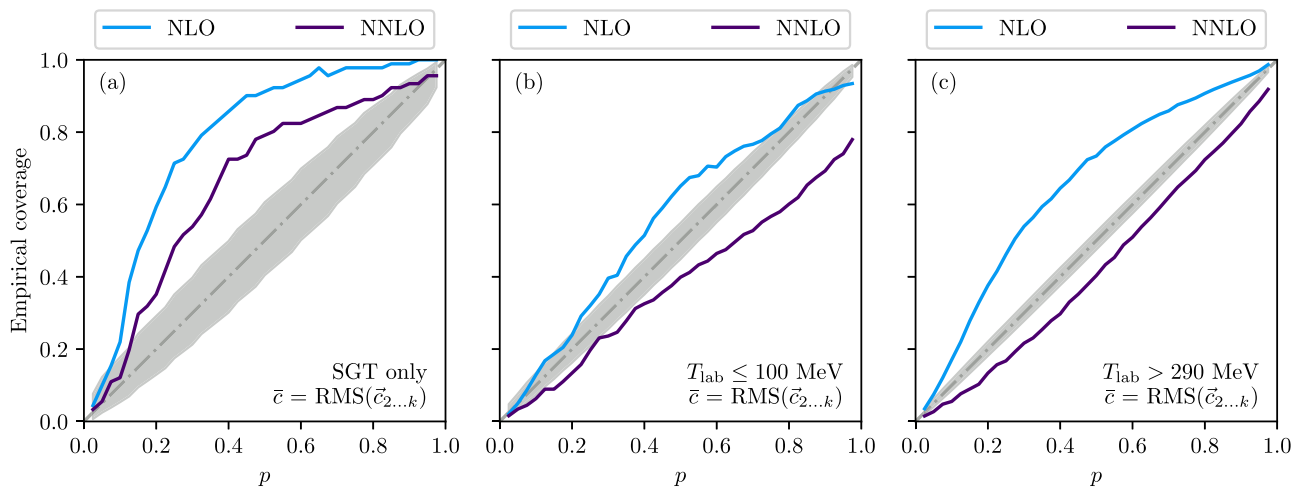

FIG. 12. Empirical coverage probabilities for $p \times 100\%$ HDI credible intervals with respect to subsets of the validation data set $\widetilde{D}$. We compute $\bar{c}$ as $\bar{c}_k = \mathrm{rms}(c_{2...k})$ as in Fig. 11(b) and use $\Lambda_b = 600$ MeV. The gray, filled regions indicate 95% confidence intervals that an observed empirical coverage is consistent with a true success rate p. (a) Only SGT observables. (b) All validation data with $T_{\mathrm{lab}} \leqslant 100$ MeV. (c) All validation data with $T_{\mathrm{lab}} > 290$ MeV.

satisfactory situation when one is making predictions at low orders of χEFT.

Finally, we estimate the consequences of modifying the employed value for the χEFT breakdown scale Λ_b. Specifically, we change Λ_b from 600 to 500 MeV in the expression for the EFT truncation error (15) without resampling the LEC posteriors. We find that decreasing Λ_b results in more natural $\bar{c}$ values while simultaneously yielding improved coverage probabilities for the EFT error. The model also becomes somewhat less sensitive to the method of computing $\bar{c}$. Conversely, raising Λ_b increases the model's sensitivity to $\bar{c}$ as well as resulting in less naturally sized expansion coefficients. The results presented here are in line with the findings in Ref. [50]. Further studies of the χEFT breakdown scale, along the lines of Ref. [35], are clearly warranted.

VI. CONCLUSIONS AND OUTLOOK

In this paper we implemented the HMC MCMC algorithm for sampling the LEC posterior PDF $\text{pr}(\vec{\alpha}|D, I)$ at LO, NLO, and NNLO in the NN sector of χEFT. We accounted for uncorrelated EFT truncation errors [5] in the sampling. Our prior was based on both reasonable assumptions and information from previous studies. For example, we assume natural EFT expansion coefficients and NN contact LECs while the πN sector was further informed by the results from a Roy-Steiner analysis [13] of πN scattering data.

We conditioned the LEC posteriors on thousands of scattering data, leading to a likelihood-dominated posterior and consequently a probability mass of the LEC posterior that is concentrated to a very small region in parameter space. At all orders, the MAP and typical widths of the posteriors are very close to the corresponding measures found using frequentist parameter estimation in, e.g., Refs. [14,34]. This is largely due to the fact that we employed nearly the same database of NN scattering cross sections in both analyses.

An analysis of the coverage probability of the HDIs for predictions indicates that the credible intervals perform largely as advertised if one excludes the LO expansion parameter when estimating $\bar{c}$ and if one has access to predictions at NNLO or beyond. Indeed, the large shift in predictions when going from LO to NLO does not provide representative information about the EFT convergence pattern. This disturbance in the order-by-order description of the nuclear interaction was also identified in a recent Bayesian analysis of three- and four-nucleon states [35]. It is hence desirable and timely to develop improved models for the EFT truncation error that explicitly account for such irregularities in the convergence pattern. At the same time it is equally important to address deficiencies in the LO description of the nucleon-nucleon interaction, in particular for making reliable low-order EFT predictions [51].

Apart from a second mode in the LO posterior, for which the deuteron is unbound, we found no clear evidence of multimodality in the LEC posteriors at NLO and NNLO. We also found that the posterior for the πN LECs at NNLO does not overlap with any significance with the narrow prior inferred from the Roy-Steiner analysis in Ref. [13]. Provided that we are operating with a low-energy EFT for the nuclear interaction, it is reasonable to expect this discrepancy to vanish if all uncertainties are accurately modeled.

We found that the HMC algorithm provides virtually uncorrelated samples of the LEC posterior at all three chiral orders. We confirm a very low level of autocorrelation which is a hallmark of the HMC algorithm. When analyzing the correlation structure in detail we found evidence of an antithetic sampling pattern in the HMC chains at chiral orders LO and NLO. This yields an integrated autocorrelation time $\tau < 1$ and an ESS greater than the number of gathered MCMC samples. If harnessed, antithetic sampling could serve as a valuable method for reducing the sample variance of LEC posteriors at higher chiral orders.

We also compared the HMC chains at the three considered orders to corresponding chains obtained using the EMCEE algorithm and found a three- to sixfold increase of the sampling efficiency. With HMC, we achieved near-instant convergence, as measured by τ, at LO, NLO, and NNLO. Using EMCEE, in contrast, we just managed to cross the convergence threshold at NLO and failed to reach a converged NNLO result. Although HMC sampling relies on access to gradients of the posterior with respect to the LECs, and requires more tuning than other MCMC algorithms, we find that the rewards justify the extra effort. It is certainly possible to devise algorithms for self-tuning of the hyperparameters with NUTS-HMC [31] being one such strategy. Our results provide a promising outlook for parameter estimation at higher chiral orders, where the increasing number of LECs most likely presents a formidable challenge to other MCMC algorithms.

[1] P. F. Bedaque and U. van Kolck, Ann. Rev. Nucl. Part. Sci. **52**, 339 (2002).

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

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

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

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

[6] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).

[7] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[8] W. K. Hastings, Biometrika **57**, 97 (1970).

[9] R. Navarro Pérez, J. E. Amaro, and E. Ruiz Arriola, Phys. Rev. C **88**, 024002 (2013).

[10] R. N. Pérez, J. E. Amaro, and E. R. Arriola, Phys. Rev. C **88**, 064002 (2013).

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

[12] M. Hoferichter, J. Ruiz de Elvira, B. Kubis, and U.-G. Meißner, Phys. Rept. **625**, 1 (2016).

[13] D. Siemens, J. Ruiz de Elvira, E. Epelbaum, M. Hoferichter, H. Krebs, B. Kubis, and U.-G. Meißner, Phys. Lett. B **770**, 27 (2017).

[14] 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).

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

[16] P. W. Lisowski, R. E. Shamu, G. F. Auchampaugh, N. S. P. King, M. S. Moore, G. L. Morgan, and T. S. Singleton, Phys. Rev. Lett. **49**, 255 (1982).

[17] INS data analysis center: SAID, http://gwdac.phys.gwu.edu, accessed: 2021-03-29.

[18] J. R. Bergervoet, P. C. van Campen, W. A. van der Sanden, and J. J. de Swart, Phys. Rev. C **38**, 15 (1988).

[19] J. R. Bergervoet, P. C. van Campen, R. A. M. Klomp, J. L. de Kok, T. A. Rijken, V. G. J. Stoks, and J. J. de Swart, Phys. Rev. C **41**, 1435 (1990).

[20] D. Stump, J. Pumplin, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung, Phys. Rev. D **65**, 014012 (2001).

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

[22] Student, Biometrika **6**, 1 (1908).

[23] R. M. Neal, in *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng (CRC, Boca Raton, FL, 2011), pp. 113–162.

[24] https://github.com/svisak/montepython.git

[25] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).

[26] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant *et al.*, Nature (London) **585**, 357 (2020).

[27] Stan Development Team, Stan modeling language users guide and reference manual, version 2.26 (2021), https://mc-stan.org/.

[28] M. Betancourt, arXiv:1701.02434.

[29] A. Griewank, Acta Numer. **12**, 321 (2003).

[30] I. Charpentier and J. Utke, Optim. Method. Softw. **24**, 1 (2009).

[31] M. D. Hoffman and A. Gelman, J. Mach. Learn. Res. **15**, 1593 (2014).

[32] J. Goodman and J. Weare, Comm. App. Math. Com. Sc. **5**, 65 (2010).

[33] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, PASP **125**, 306 (2013).

[34] P. Reinert, H. Krebs, and E. Epelbaum, Eur. Phys. J. A **54**, 86 (2018).

[35] S. Wesolowski, I. Svensson, A. Ekström, C. Forssén, R. J. Furnstahl, J. A. Melendez, and D. R. Phillips, Phys. Rev. C **104**, 064001 (2021).

[36] A. Sokal, in *Functional Integration: Basics and Applications*, edited by C. DeWitt-Morette, P. Cartier, and A. Folacci (Springer, New York, 1997), pp. 131–192.

[37] A. Gelman and D. B. Rubin, Stat. Sci. **7**, 457 (1992).

[38] S. P. Brooks and A. Gelman, J. Comput. Graphi. Stat. **7**, 434 (1998).

[39] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, A. Vehtari, and D. B. Dunson, *Bayesian Data Analysis*, Texts in Statistical Science Series (Chapman and Hall, London, 2014).

[40] A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, and P.-C. Bürkner, Bayesian Anal. **16**, 667 (2021).

[41] J. M. Hammersley and K. W. Morton, Math. Proc. Camb. Philos. Soc. **52**, 449 (1956).

[42] F. Feroz, M. P. Hobson, and M. Bridges, Mon. Not. R. Astron. Soc. **398**, 1601 (2009).

[43] S. König, A. Ekström, K. Hebeler, D. Lee, and A. Schwenk, Phys. Lett. B **810**, 135814 (2020).

[44] A. Ekström and G. Hagen, Phys. Rev. Lett. **123**, 252501 (2019).

[45] J. A. Melendez, C. Drischler, A. J. Garcia, R. J. Furnstahl, and X. Zhang, Phys. Lett. B **821**, 136608 (2021).

[46] S. B. S. Miller, A. Ekström, and C. Forssén, arXiv:2106.00454.

[47] B. von Przewoski, F. Rathmann, W. A. Dezarn, J. Doskow, M. Dzemidzic, W. Haeberli, J. G. Hardie, B. Lorentz, H. O. Meyer, P. V. Pancella, R. E. Pollock, T. Rinckel, F. Sperisen, and T. Wise, Phys. Rev. C **58**, 1897 (1998).

[48] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevC.105.014004 for posterior predictive distributions at LO, NLO, and NNLO for the nucleon-nucleon scattering cross sections in the Granada database.

[49] N. Hoshizaki, Prog. Theor. Phys. Suppl. **42**, 107 (1969).

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

[51] C. J. Yang, A. Ekström, C. Forssén, and G. Hagen, Phys. Rev. C **103**, 054304 (2021).