Uncertainty quantification in nuclear shell model

Sota Yoshida,1,* Noritaka Shimizu,2 Tomoaki Togashi,2 and Takaharu Otsuka3,1,2,4
1Department of Physics, the University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
2Center for Nuclear Study, the University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
3RIKEN Nishina Center, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
4Instituut voor Kern- en Stralingsfysica, KU Leuven, B-3001 Leuven, Belgium
(Dated: November 28, 2018)

The uncertainty quantifications of theoretical results are of great importance to make meaningful comparisons of those results with experimental data and to make predictions in experimentally unknown regions. By quantifying uncertainties, one can make more solid statements about, e.g., origins of discrepancy in some quantities between theory and experiment. We propose a novel method for uncertainty quantification for the effective interactions of nuclear shell-model calculations as an example. The effective interaction is specified by a set of parameters, and its probability distribution in the multi-dimensional parameter space is considered. This enables us to quantify the agreement with experimental data in a statistical manner and the resulting confidence intervals show unexpectedly large variations. Moreover, we point out that a large deviation of the confidence interval for the energy in shell-model calculations from the corresponding experimental data can be used as an indicator of some exotic property, e.g. α clustering, etc. Other possible applications and impacts are also discussed.

Introduction. Many modern theoretical models have been developed by iterative cycles of solving "forward modeling problem" and "inverse modeling problem" [1], and it has provided deep insight into phenomena of interest. In particular, such cycles have been playing key roles in nuclear physics, because the fundamental interaction, the nuclear force, is still uncertain due to its complexity including the non-perturbative character in the low-energy regime. In applications of such theoretical models, one has to solve the inverse modeling problem, i.e. the optimization problem to minimize deviations of model estimates from observations. However, if one uses a model with so-called point estimation of the optimal input parameters, there is a risk of overfitting to the given observations and of lacking the generalization ability to the data not taken into the optimization process. Therefore, it is desired that any theoretical model should be properly accompanied with uncertainty estimates, see e.g. Editorial on Physical Review A [2]. Theoretical studies in nuclear theory with uncertainty estimates are being expanded, e.g., mean-field calculations [3–10], nuclear potentials [11–14], etc.

The shell-model calculation, which is called the configuration interaction method in other fields of science, has been providing successful and systematic descriptions of a wide variety of properties of light to medium-mass nuclei. Therefore one can expect that this model well approximates wave functions of nuclei, see e.g. Refs. [15–17]. Conventional shell-model Hamiltonians have been constructed by the G-matrix theory [18, 19] with minor phenomenological corrections. Not only from phenomenological perspectives, but also from microscopic viewpoints, shell model is playing a key role with recent developments in nuclear forces from chiral effective field theory (chiral EFT) [20, 21] and many-body methods to derive shell-model effective interactions for a physically motivated model space [22–27]. The latter approaches combining ab initio methods and shell-model calculations can act as a foothold for better understanding of the nuclear potential. Under these circumstances, it is an urgent task to assess the validity of nuclear shell model through the evaluations of their uncertainties stemming from input effective interactions. This is because by performing detailed analyses of uncertainties coming from input parameters in the theory one can make more solid statements, such as on origins of the discrepancies between theoretical predictions and experimental observations, or on which observables should be reproduced by the theory within a given model space. In this study, we provide, for the first time, uncertainty quantifications of shell-model spectra in a statistical manner.

Inference on effective interactions. In what follows, we introduce a way to quantify uncertainties in shell-model results by introducing Bayesian inference on input parameters and then we consider the shell-model results with ensembles of effective interactions. It provides rich information about the validity of theoretical estimates, while in conventional shell-model studies results with only one or a few parameter sets are compared with experimental values.

We take the 0p-shell space on top of the 4He core as the model space. The 0p shell consists of two orbitals 0p1/2 and 0p3/2 and these are abbreviated as p1/2 and p3/2 in this paper. In this case, the shell-model Hamiltonian contains the 17 parameters with the isospin symmetry: 2 single-particle energies (SPEs) and 15 two-body matrix elements (TBMEs). We do inference on these parameters under a given set of experimental energy values for some p-shell nuclei and regard the parameters not as points but as probability distributions over 17-dimensional parameter space. All calculations within the 0p-shell space can

* s.yoshida@nt.phys.s.u-tokyo.ac.jp
be done by an exact diagonalization method with relatively low computational costs. In this case, since there is no uncertainty coming from the many-body calculation, i.e. the diagonalization process, we can extract a theoretical uncertainty coming merely from input parameters.

In Bayesian analysis the parameter distribution is described by the posterior distribution, i.e. the conditional probability distribution under observation of data. This posterior is obtained through Bayes’ theorem,

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \propto P(D|\theta)P(\theta),$$

where \( \theta \) is a set of multidimensional parameters. In this study, \( \theta \) corresponds to the 17-dimensional parameters for the \( p \)-shell effective interactions. The \( D \) in the equation above denotes the data set taken into account for the parameter inference. We take the uniform distribution as the prior, \( P(\theta) \propto 1 \), and the ordinary likelihood function,

$$P(D|\theta) = \exp (-\chi^2(\theta)/2),$$

with the squared errors,

$$\chi^2(\theta) \equiv \sum_{n=1}^{N_D} \frac{(O_n^{\text{exp}} - O_n^{\text{th}}[\theta])^2}{\Delta O_n}.$$  

Here \( N_D \) denotes the number of data, \( O_n^{\text{exp}} \) is an experimental value for the observable labeled by \( n \), and \( O_n^{\text{th}}[\theta] \) is the corresponding theoretical results with \( \theta \). We note that since we now assume the isospin symmetry in effective interactions, the Coulomb corrections to energy values for \{\( O_n^{\text{th}}[\theta] \}\} are calculated in accordance with Ref. [28]. The \( \Delta O \) in the Eq. (3) is the typical error of the observables and this should contain both theoretical and experimental errors in general. However, the experimental error is negligible in the present work. For \{\( O_n^{\text{exp}} \}\}, we use a fixed 33 data for the energy values of ground and excited states of the \( p \)-shell nuclei throughout this work. Those are essentially the same set as used in the work by Cohen and Kurath [28] with the exception of a few updated data. The interest reader is referred to the Supplemental Material [29]. If one intends to change the data set in the fitting procedure, one can modify the parameter distributions relatively easily by regarding the previously obtained posterior as the prior and by doing the similar inference for the data added. Furthermore, if one makes pseudo data for unknown states, one can estimate their impacts on parameters in the same manner. This flexibility is a benefit of introducing Bayesian inference.

We typically have two classes to evaluate the posterior \( P(\theta|D) \). One is the asymptotically exact method, Markov Chain Monte Carlo (MCMC) [30]. If one could achieve an infinite number of iterations for the MCMC, one would obtain the samples, i.e. the ensemble of effective interactions that are exactly obeying posterior distributions. However, in the case of highly multimodal or too steep true posterior, it is hard to obtain fully converged results by the MCMC with currently available computer resources. The other class is the approximation schemes, which are literally approximation methods for posteriors; one can write down a posterior in a simple and closed form like a Gaussian distribution. In this class, computational costs are in general much less than those for the MCMC. The advantages and disadvantages for these two classes are complementary. In this work, we employ the Laplace approximation (LA), which belongs to the latter class. In the problems of interest, the LA tends to search a wider region in the parameter space than the MCMC, because the \( \chi^2 \) potential is considerably steep around the global minimum. This means that uncertainties in the parameters evaluated by LA samples are larger than those obtained by the MCMC. We adopt the LA because it mitigates the risk of too definitely estimating the uncertainty.

When applying the LA, the posterior is approximated by multivariate Gaussian around the so-called Maximum A Posteriori (MAP) estimate:

$$\theta_{\text{MAP}} = \arg \max_{\theta} P(D|\theta)P(\theta) = \arg \min_{\theta} \chi^2(\theta),$$

$$P(D|\theta) \approx \frac{|A|}{\sqrt{(2\pi)^n}} \exp \left( -\frac{1}{2} \bar{\theta}^T A \bar{\theta} \right),$$

with \( \bar{\theta} = \theta - \theta_{\text{MAP}} \) and the Hessian matrix \( A \)

$$A = -\nabla \nabla \log P(D|\theta)_{\theta=\theta_{\text{MAP}}}. \quad (6)$$

In Eq. (4), we used Eqs. (1-2) and \( P(\theta) \propto 1 \). Each element of the Hessian matrix \( A \) is written as

$$A_{ij} = \sum_{n=1}^{N_D} \frac{1}{(\Delta O_n)^2} \frac{\partial O_n^{\text{th}}[\theta]}{\partial \theta_i} \frac{\partial O_n^{\text{th}}[\theta]}{\partial \theta_j} + \sum_{n=1}^{N_D} \frac{(O_n^{\text{th}}[\theta] - O_n^{\text{exp}}) \partial^2 O_n^{\text{th}}[\theta]}{(\Delta O_n)^2} \frac{\partial O_n^{\text{th}}[\theta]}{\partial \theta_i} \frac{\partial O_n^{\text{th}}[\theta]}{\partial \theta_j}. \quad (7)$$

The first derivative of the term is obtained by the Hellman-Feynmann theorem [31] and the second derivative term above is numerically evaluated using the finite differences of the first derivative terms,

$$\frac{\partial^2 O_n^{\text{th}}[\theta]}{\partial \theta_i \partial \theta_j} = \frac{1}{2\epsilon} \left( \frac{\partial O_n^{\text{th}}[\theta_i^+]}{\partial \theta_j} - \frac{\partial O_n^{\text{th}}[\theta_i^-]}{\partial \theta_j} \right) + \mathcal{O}(\epsilon^2), \quad (8)$$

where \( O_n[\theta_j^\pm] \) denotes the \( O_n \) value evaluated by the parameters whose \( j \)-th components are slightly shifted by a small value \( \pm \epsilon \) from the \( \theta_{\text{MAP}} \).

While we use the uniform prior and the LA for the posterior, (this particular choice reduces the parameter inference to the problem of chi-square minimization and covariance matrix analysis as in Refs. [3–5, 7–14]), the previously mentioned benefit of Bayesian inference could be achieved in future works by a fully Bayesian treatment of posterior distributions.
The concrete procedures to quantify the uncertainties in shell-model calculations by the LA are the following: (i) to search the optimal interaction, i.e. the $\theta_{\text{MAP}}$ in Eq. (4), with respect to the given data set by means of optimization methods, (ii) to calculate the Hessian matrix according to Eqs. (6)-(8), and (iii) to generate a sufficiently large number of the parameter set $\{\theta\}$ obeying the posterior defined in Eq. (5) and to evaluate statistical quantities for observables of interest. For (i), we prefix the optimal interaction by Stochastic Gradient Descent, see e.g. [32]. The root mean square errors for the given data set are summarized in Tab. I in comparison with the results for three sets of the interaction proposed by Cohen and Kurath (CK) [28]. We note that the mass-dependence of the form $(A/6)^{-0.3}$ is introduced for TBMEs while three CK interactions are mass-independent. Three CK interactions, denoted CKpot, CKtb1, and CKtb2, correspond, respectively, to (8-16)POT, (8-16)2BME, and (6-16)2BME, in the original work. Here, the numbers in the parentheses denote the range of the mass number used in fit. In what follows, we use the total root mean square error for the optimal interaction as the typical error in Eq. (3), $\Delta \mathcal{O} = 0.35$ MeV. In the procedure (iii), we perform shell-model calculations with $50,000$ LA samples. This number is large enough to suppress the error coming from stochastic choices of samples; The typical error in the mean values of energy eigenvalues is less than 0.1%.

Results. We show the marginal distributions of the 17 parameters for the p-shell interaction in Fig. 1. Red curves with the shaded area and blue bars respectively denote $3\sigma$ and $1\sigma$ confidence intervals for the parameter distributions. The $1\sigma$ deviations of parameters are given by the square root of diagonal components of $A^{-1}$. The white dots plotted at the medians of Gaussian distributions correspond to the parameter values of the MAP estimation. Some TBMEs, whose total angular momentum $J$ and total isospin $T$ are $(J, T) = (1, 0)$, show relatively large uncertainties. If we take Gamow-Teller transition strengths or electromagnetic observables into fit, which are sensitive to those parameters, we expect that the uncertainties would become smaller. However, this brings additional issues on the quenching factor or the effective charges. In this work, we do not enter into the detail of those quantities for the sake of simplicity.

In Fig. 2, we show some selected results of energy spectra. Theoretical results by the LA samples are shown by so-called violin plots in comparison with (i) experimental data [33], (ii) the results by one of the CK interactions [28], and (iii) those by the microscopically derived interaction by the valence-space in-medium similarity renormalization group (VS-IMSRG) [24]. The height and width of violins show, respectively, $3\sigma$ confidence intervals and appearance frequencies of the quantities with respect to all the $50,000$ LA samples. Here we use the notation $\bar{\sigma}$ for standard deviations of the results with all the LA samples in order to distinguish from the ordinary statistical term $\sigma$ for Gaussian distributions. Their mean values (horizontal solid lines) and $1\bar{\sigma}$ confidence intervals (error bars) are shown in violin plots. Experimental data and theoretical results are classified with the angular momentum $J$, the parity $\pi$, and the isospin $T$. We note that the total isospin is not determined in some data, in which cases most plausible values are taken from the corresponding theoretical results. The CKtb1 was determined from the almost same data as the present study, and gives the minimum total root mean square errors among the three CK interactions (see Tab. I). Excitation spectra like Fig. 2 (a, c) are convenient for comparisons with experimental studies with gamma-ray measurements, e.g. to determine which states are relevant to newly measured gamma rays. One can also work with energy eigenvalues, as shown in Fig. 2 (b, d). These are convenient for purposes to see systematic deviations (e.g. with respect to the mass numbers) of shell-model results from experimental data.

Hereafter, we consider 82 low-lying states of the p-shell nuclei in order to confirm that the parameter distributions are not overfitted to the 33 states. We mainly focus on the energy spectra for $^{12,14}$C in Fig. 2 as a representative example. By taking a close look at these results, we can make some important remarks. Firstly, we can see significant variances in the height of confidence intervals. This enables us to quantify the relative reliabilities of the theoretical estimates. The overall scale of the

| $\theta_{\text{MAP}}$ | CKpot | CKtb1 | CKtb2 |
|-----------------------|-------|-------|-------|
| total RMS             | 0.35  | 0.57  | 0.47  | 0.54  |

FIG. 1. The marginal distributions of 17 parameters for the p-shell, in MeV. Red curves and blue bars denote the $3\sigma$ and $1\sigma$ confidence intervals, respectively. Each plot is scaled to the same height for aesthetic purposes. White dots denote the mean of the distribution and, namely, the $\theta_{\text{MAP}}$. The TBMEs are abbreviated as $v(abcd, JT)$, and the orbits are labeled by $1 = p_{1/2}$ and $3 = p_{3/2}$.
whose mean energy values for the 82 states are reproduced by the LA samples (not in fit) and/or many-body structure. Secondly, most of the states are described with the LA samples within 1σ deviations, as shown in Fig. 2. Experimental data on ENSDF [33] are shown by horizontal dotted lines in red. Corresponding theoretical results with the LA samples are summarized in violin plots in green or pink. Here green and pink ones corresponds to the states in fit and not in fit, respectively. In each violin plot, mean value and 1σ error are shown by horizontal line and error bar in white, respectively, where σ means standard deviations of all results. Results by a CK interaction and the VS-IMSRG are also shown.

FIG. 2. Excitation energies (left column) and energies (right column) of 12C (panels (a) and (b)) and 14C ((c) and (d)). Each state is classified by its (J^T, N_JT), where (J^T, N_JT) stands for the N-th lowest states with the total angular momentum J^T and the total isospin T. Experimental data on ENSDF [33] are shown by horizontal dotted lines in red. Corresponding theoretical results with the LA samples are summarized in violin plots in green or pink. Here green and pink ones corresponds to the states in fit and not in fit, respectively. In each violin plot, mean value and 1σ error are shown by horizontal line and error bar in white, respectively, where σ means standard deviations of all results. Results by a CK interaction and the VS-IMSRG are also shown.

can be interpreted as an indicator of exotic structures, e.g. α clustering, intruder configurations, core excitations, etc. Although we cannot give a criteria for exotic structures, we can deduce which states are likely to have exotic structures by looking at the relative size of the discrepancies between confidence intervals and observations. This is an important outcome of the evaluation of uncertainties in shell-model calculations.

Even if there is no preliminary knowledge of which states have exotic structures, one can quantify how the states are to be (or not to be) taken into account through fitting procedures. For example, if one adds the Hoyle state into the data set in fit, the total root mean square error becomes worse by ~ 39% for the 74 states which are reproduced within 2σ error by the already obtained LA samples with the 33 data. In this way, one can identify certain states which should not be included in the fitting procedure, and then update the optimal interaction to a more utilitarian one. Furthermore, we can assess the validity of the LA by looking into the shapes of violin plots. If the LA failed to capture the global structure of χ^2, i.e. in the case that there are many local minima or plateaus in the χ^2-potential over the parameter...
space, it is expected that the mean values of observables calculated with the LA samples will deviate from the evaluation with the MAP estimate and that the shape of violin plots will become asymmetric. However, the deviation is not significant and the shapes of violin are almost symmetric. Regarding all the 82 states, the typical discrepancy in energy values between a result by the $\theta_{\text{MAP}}$ and a mean value of the results with the LA samples is about 0.1% (see [29]).

Finally, the methodology discussed in this study would be useful for microscopic understandings too. In the current microscopic calculations like the VS-IMSRG and the coupled-cluster method, the input nuclear potential is regarded as the dominant source of uncertainty [43]. The capability of shell-model calculations is important to know which states are likely to be described by future microscopically derived effective interactions. By visualizing the confidence intervals of shell-model calculations, there is also a possibility to figure out some missing contribution or higher order terms which are not to be omitted in modern nuclear forces and many-body methods. For example, if the considered states have large deformations and are located in the confidence intervals of the shell-model calculation with phenomenologically constructed effective interactions, one may need to take account of higher order couplings between a valence space and an outer space.

Conclusions. We have proposed a novel method to quantify uncertainties in shell-model calculations. By introducing violin plots, we have visualized the unexpectedly large variations of confidence intervals which differ by states. These plots can be used to quantify the agreement with experimental data in a statistical manner. We have also proposed that a large deviation of theoretical confidence intervals from the corresponding experimental data can be regarded as an indicator of some exotic structure like clustering, intruder configurations, core excitations, etc.

The methodology discussed in this work may have various possible future applications. By implementing Bayesian inference on the parameters, we can sequentially construct shell-model effective interactions while sorting out data. This would be useful especially for heavier regions where even one iteration for the optimization process is computationally demanding. It would be also interesting to extend this analysis to the valence-space Hamiltonians based on chiral EFT potentials for a valence space [44]. One can also extend this study to include electromagnetic observables or Gamow-Teller transition strengths into the data set. That is one of the future perspectives of this work and of great importance for astrophysical applications.

If these uncertainty estimates could be done for any other phenomenological models in which theoretical uncertainties have not been discussed in detail, it may provide new insight into a wide variety of phenomena in nuclear physics.

Acknowledgments. We thank Achim Schwenk and Yutaka Utsuno for fruitful discussions and comments. We also thank Patrick Copinger for carefully reading the manuscript. This work was supported in part by JSPS KAKENHI (Grant Numbers 17J06775 and 17K05433), JSPS Overseas Challenge Program for Young Researchers (No. 20188031), the priority issue 9 to be tackled by using Post-K computer, and the CNS-RIKEN joint project for large-scale nuclear structure calculations.

[1] A. Tarantola, Nature Physics 2, 492 (2006).
[2] The Editors, Phys. Rev. A 83, 040001 (2011).
[3] P.-G. Reinhard and W. Nazarewicz, Phys. Rev. C 81, 051303 (2010).
[4] F. J. Fattoyev and J. Piekarewicz, Phys. Rev. C 84, 064302 (2011).
[5] M. Kortelainen, J. Erler, W. Nazarewicz, N. Birge, Y. Gao, and E. Olsen, Phys. Rev. C 88, 031305 (2013).
[6] Y. Gao, J. Dobaczewski, M. Kortelainen, J. Toivanen, and D. Tarpanov, Phys. Rev. C 87, 034324 (2013).
[7] J. Dobaczewski, W. Nazarewicz, and P.-G. Reinhard, Journal of Physics G: Nuclear and Particle Physics 41, 074001 (2014).
[8] S. Goriely and R. Capote, Phys. Rev. C 89, 054318 (2014).
[9] S. E. Aghemava, A. V. Afanasjev, D. Ray, and P. Ring, Phys. Rev. C 89, 054320 (2014).
[10] J. D. McDonnell, N. Schunck, D. Higdon, J. Sarich, S. M. Wild, and W. Nazarewicz, Phys. Rev. Lett. 114, 122501 (2015).
[11] R. N. Pérez, J. E. Amaro, and E. R. Arriola, Phys. Rev. C 88, 064002 (2013); R. Navarro Pérez, J. E. Amaro, and E. R. Arriola, Phys. Rev. C 91, 029901(E) (2015).
[12] R. Navarro Pérez, J. E. Amaro, and E. Ruiz Arriola, Phys. Rev. C 89, 064006 (2014).
[13] A. Ekström, B. D. Carlsson, K. A. Wendt, C. Forssén, M. H. Jensen, R. Machleidt, and S. M. Wild, Journal of Physics G: Nuclear and Particle Physics 42, 034003 (2015).
[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] B. A. Brown, Progress in Particle and Nuclear Physics 47, 517 (2001).
[16] E. Caurier, G. Martinez-Pinedo, F. Nowacki, A. Poves, and A. P. Zuker, Rev. Mod. Phys. 77, 427 (2005).
[17] T. Otsuka, A. Gade, O. Sorlin, T. Suzuki, and Y. Utsuno, “Evolution of nuclear structure in exotic nuclei and nuclear forces,” (2018), arXiv:1805.06501.
[18] T. T. S. Kuo and G. E. Brown, Nuclear Physics 85, 40 (1966).
[19] M. Hjorth-Jensen, T. T. S. Kuo, and E. Osnes, Physics Reports 261, 125 (1995).
[20] R. Machleidt and D. Entem, Phys. Rept. 503, 1 (2011).
