

# Bayesian, frequentist, and information geometric approaches to parametric uncertainty quantification of classical empirical interatomic potentials

Cite as: *J. Chem. Phys.* **156**, 214103 (2022); <https://doi.org/10.1063/5.0084988>

Submitted: 11 January 2022 • Accepted: 15 May 2022 • Accepted Manuscript Online: 16 May 2022 • Published Online: 01 June 2022

 [Yonatan Kurniawan](#),  [Cody L. Petrie](#), [Kinamo J. Williams](#), et al.



[View Online](#)



[Export Citation](#)



[CrossMark](#)

## ARTICLES YOU MAY BE INTERESTED IN

### [Density-functional theory vs density-functional fits](#)

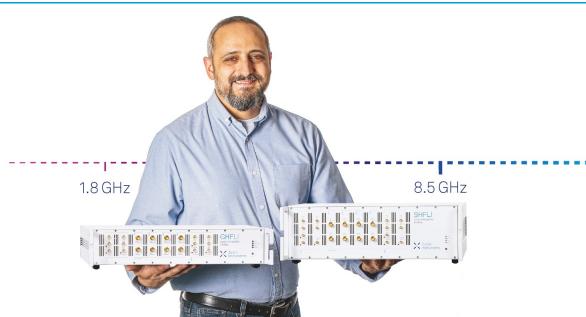
*The Journal of Chemical Physics* **156**, 214101 (2022); <https://doi.org/10.1063/5.0091198>

### [Computational materials discovery](#)

*The Journal of Chemical Physics* **156**, 210401 (2022); <https://doi.org/10.1063/5.0096008>

### [Transition orbital projection approach for excited state tracking](#)

*The Journal of Chemical Physics* **156**, 214104 (2022); <https://doi.org/10.1063/5.0081207>



**Trailblazers.** 

Meet the Lock-in Amplifiers that measure microwaves.

 [Zurich Instruments](#) [Find out more](#)

# Bayesian, frequentist, and information geometric approaches to parametric uncertainty quantification of classical empirical interatomic potentials

Cite as: J. Chem. Phys. 156, 214103 (2022); doi: 10.1063/5.0084988

Submitted: 11 January 2022 • Accepted: 15 May 2022 •

Published Online: 1 June 2022



Yonatan Kurniawan,<sup>1</sup> Cody L. Petrie,<sup>1</sup> Kinamo J. Williams Jr.,<sup>1</sup> Mark K. Transtrum,<sup>1,a)</sup> Ellad B. Tadmor,<sup>2</sup> Ryan S. Elliott,<sup>2</sup> Daniel S. Karls,<sup>2</sup> and Mingjian Wen<sup>3</sup>

## AFFILIATIONS

<sup>1</sup> Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84604, USA

<sup>2</sup> Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, Minnesota 55455, USA

<sup>3</sup> Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

<sup>a)</sup> Author to whom correspondence should be addressed: [mktranstrum@byu.edu](mailto:mktranstrum@byu.edu)

## ABSTRACT

In this paper, we consider the problem of quantifying parametric uncertainty in classical empirical interatomic potentials (IPs) using both Bayesian (Markov Chain Monte Carlo) and frequentist (profile likelihood) methods. We interface these tools with the Open Knowledgebase of Interatomic Models and study three models based on the Lennard-Jones, Morse, and Stillinger-Weber potentials. We confirm that IPs are typically sloppy, i.e., insensitive to coordinated changes in some parameter combinations. Because the inverse problem in such models is ill-conditioned, parameters are unidentifiable. This presents challenges for traditional statistical methods, as we demonstrate and interpret within both Bayesian and frequentist frameworks. We use information geometry to illuminate the underlying cause of this phenomenon and show that IPs have global properties similar to those of sloppy models from fields, such as systems biology, power systems, and critical phenomena. IPs correspond to bounded manifolds with a hierarchy of widths, leading to low effective dimensionality in the model. We show how information geometry can motivate new, natural parameterizations that improve the stability and interpretation of uncertainty quantification analysis and further suggest simplified, less-sloppy models.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0084988>

## I. INTRODUCTION

Interatomic potentials (IPs) are a foundational tool in computational materials science.<sup>1</sup> They allow modelers to make efficient predictions of materials properties without reference to the complicated sub-atomic structure. Recently, there has been considerable interest in applying methods of uncertainty quantification (UQ) to IPs.<sup>2–5</sup> UQ assesses the reliability of materials predictions, leveraging tools from statistical inference.<sup>6</sup> Statistical analysis of similar inverse problems in physics has motivated the study of *sloppy models*.<sup>7</sup> Sloppy models lead to extremely ill-conditioned inverse problems and pose several challenges for standard statistical methods.<sup>8,9</sup> This work considers the application of UQ to IPs in the context of sloppy models. We focus on the structural uncertainties of the IPs

as opposed to the random errors that come from the details of the simulation.<sup>10</sup> We find that many IPs are sloppy, which leads to challenges in interpreting UQ results, and use information geometry to mitigate some of these challenges. (We will be using the terms “IPs” and “models” largely interchangeably in this paper, with “models” used in more generic settings, and “IPs” when specifically thinking of the training and predictions of interatomic models, i.e., potentials.)

Classical IPs have been widely used in materials science to circumvent the computational cost of quantum calculations, such as density functional theory (DFT), by approximating the interaction energy between atoms without considering the electrons. However, a universal functional form that can describe all types of atomic bonding has not been discovered. Thus, IPs are often designed for specific

purposes, resulting in a plethora of models.<sup>11</sup> Efforts such as the Open Knowledgebase of Interatomic Models (OpenKIM)<sup>12</sup> aim to organize and standardize these IPs.

In the development of classical empirical IPs, the parameters are typically fit to match experimental or first principles data of some microscopic properties, such as the lattice parameters and elastic constants of single crystals or the potential energy and atomic forces associated with various atomic configurations.<sup>13</sup> They are then used in conjunction with simulation codes to predict other properties that are not used in the fitting process. Thus, UQ is relevant for assessing the reliability of these out-of-sample predictions.

In this paper, we are primarily interested in parametric uncertainty, i.e., uncertainty in the model's parameters, which is quantified through, for example, a Bayesian posterior distribution or confidence regions on the parameter space. In our formulation, the IP is used in two evaluation scenarios: one that makes predictions for the training data (e.g., energy and forces), and a second that makes predictions for other material quantities of interest. By varying the parameters, the IP traces out a set of possible predictions for each scenario. The set of predictions made by an IP for each evaluation scenario generates a corresponding *model manifold*,<sup>8</sup> which can be studied using methods of information geometry. The IP parameters act as coordinates on the model manifold and distances on the manifold measure statistical identifiability. The UQ process propagates uncertainties from training data to uncertainties in parameters via the (pseudo) inverse of the first evaluation scenario. These uncertainties then propagate forward through the second scenario to give uncertainties in the predictions for the quantities of interest. The inverse of an evaluation scenario is not given explicitly but is only accessible through iterative evaluations of the scenario. Consequently, the first uncertainty propagation process is

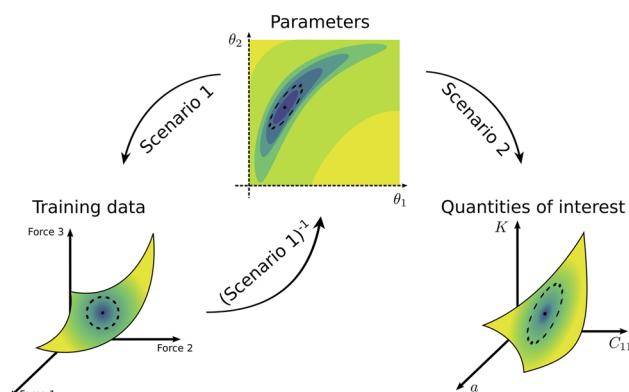
the more challenging one and the focus of this study. Given the parametric uncertainties, they can then be propagated to other quantities of interest, for example, as in Ref. 14. This entire process is illustrated in Fig. 1.

Many UQ methods have been developed to propagate uncertainties in data to uncertainties in parameters.<sup>15</sup> In materials science, Markov Chain Monte Carlo (MCMC) sampling of the Bayesian posterior is the most common approach. Being a Bayesian method, this requires a prior distribution, and several prior distributions have been used, including uniform,<sup>16–22</sup> normal,<sup>23</sup> Jeffreys prior,<sup>24</sup> and maximum entropy.<sup>25</sup> Other approaches to UQ include *F*-statistics estimations,<sup>26</sup> ANOVA-based methods,<sup>27</sup> and multi-objective optimization.<sup>28</sup> Other fields have used the profile likelihood method,<sup>29–31</sup> to the best of our knowledge, ours is their first application to IPs.

The study of inverse problems in statistical physics has identified an important property of many multi-parameter models, known as *sloppiness*. Sloppy models are characterized by predictions that are insensitive to coordinated changes in combinations of parameters.<sup>32</sup> Inverse problems for sloppy models are extremely ill-conditioned and, as we will show below, present obstacles for standard UQ methods, and there is broad interest in developing UQ methods that leverage the effective low-dimensionality of these problems to improve UQ performance.<sup>33–35</sup> Sloppiness was first systematically studied in 2003 by Brown and Sethna in the context of systems biology models.<sup>7</sup> The relevant object is the Fisher Information Matrix (FIM)<sup>36</sup> that quantifies the information that data in a given evaluation scenario carry about parameters in a model. Eigenvalues of the FIM provide a local measure of sloppiness. For sloppy models, the FIM eigenvalues span many orders of magnitude and have many small eigenvalues. These small eigenvalues correspond to sloppy combinations of parameters, i.e., those that are ill-constrained by data.<sup>8,37–39</sup>

Although sloppy models are usually identified by their characteristic FIM spectrum, the theory of sloppy models is couched in the tools of information geometry, i.e., the application of differential geometry to statistics.<sup>9,40</sup> The key object is the model manifold, which comprises the set of all predictions a model makes within a given evaluation scenario for all allowed parameter values. As we show in Sec. II E, the model manifold is embedded in data space. This space is useful because the distance corresponds to statistical identifiability. In other words, points that are distant on the model manifold are statistically distinguishable, while nearby points are not. Statistical identifiability therefore induces a Riemannian metric on the parameter space that is given by the FIM. Using the FIM, one can, in principle, measure the “information distance” across the manifold. We refer to such a distance as the manifold's width. This width depends on the direction in which the distance is measured. For sloppy models, widths in some directions can be many orders of magnitude larger than in other directions. This hierarchy of widths suggests that the model exhibits a low effective dimensionality.<sup>8,41–44</sup> When a model manifold is very thin, the parameters associated with the thin directions are unidentifiable from data.

Sloppy models are ubiquitous in many scientific fields, including critical phenomena,<sup>42</sup> systems biology,<sup>9,37,45–48</sup> power systems stability,<sup>49</sup> particle accelerators,<sup>50</sup> and others.<sup>39</sup> In molecular modeling, it has been shown that IPs typically exhibit the same characteristic sloppy FIM spectrum;<sup>2–4,51</sup> however, the techniques of information geometry have not yet found application to IPs. As



**FIG. 1.** Uncertainty propagation in information geometry. An IP is described by several parameters, collectively forming a parameter space (top center). By considering the predictions for all allowed parameter values for a given evaluation scenario, the model (IP) maps out a set of possible predictions in data space, known as the model manifold (bottom left, right). Uncertainty in training data (bottom left) is propagated through the inverse function to uncertainties in parameter space, represented here by contours of constant likelihood, i.e., cost, in the parameter space. Parametric uncertainty can then be propagated forward in another evaluation scenario to material quantities of interest, such as lattice constant, elastic constant, and bulk modulus (bottom right).

materials models increase in complexity, especially with the advent of machine learning models,<sup>5,52–54</sup> sloppiness will become increasingly relevant. This underscores the importance of understanding the effects of sloppiness on IPs and other materials science models.

This work reports on several novel results applying sloppy model analysis to the UQ of IPs. Since sloppy models are ill-conditioned, they pose unique challenges for standard UQ methods. Results are often highly sensitive to details of the problem formulation and difficult to interpret. We illustrate these challenges with both Bayesian and frequentist UQ methods, introducing the profile likelihood methods to the material modeling community. A new contribution of this work is the application of information geometry techniques to bear on this problem and discuss how they can illuminate and mitigate these challenges.

This paper is organized as follows: First, we precisely formulate the problem in Sec. II A and discuss using the FIM as a local analysis of sloppiness in Sec. II B. Then, we discuss the Bayesian methods in Sec. II C. We follow Frederiksen *et al.*<sup>4</sup> and introduce an effective sampling temperature to estimate the model accuracy. This study is the first extensive study of the role of this sampling temperature in the identifiability of the model parameters. We describe frequentist methods in Sec. II D; to the best of our knowledge, ours is the first application of profile likelihoods to IPs. Finally, we describe information geometric tools, specifically geodesics, central to our sloppy model analysis in Sec. II E. This study is the first application of information geometry to interatomic potentials. In particular, we demonstrate that the same global properties observed in other sloppy models (specifically manifold boundaries) are also present in IPs. This opens up the possibility of using novel parameter reductions techniques, such as the manifold boundary approximation method (MBAM)<sup>40</sup> for IP development. Section II F presents the models used in this study, i.e., the IPs and quantities of interest. We present the results for each method in Sec. III. Finally, we discuss the effects sloppiness has on UQ in Sec. IV and the prospect for accurate, efficient UQ in IPs generally. From this cross-sectional survey, we make several recommendations that we hope will be beneficial for both IP developers and practitioners.

## II. METHODS

In this section, we introduce the general methods that will be used later in this paper. We pay particular attention to the mathematical assumptions that different methods require and the types of calculations that they enable. Then, we introduce the IPs and datasets on which we conduct our study.

### A. Defining cost

The minimal elements for parametric UQ are (1) a collection of data,  $\{y_m\}_{m=1}^M$  (where  $M$  is the number of data points), (2) a parameterized family of models that make predictions  $\{f_m(\theta)\}_{m=1}^M$ , and (3) a metric for comparing the model predictions to data,  $\|\cdot\|$ . We assume that the model depends on  $N$  parameters  $\theta \in \mathcal{D} \subseteq \mathbb{R}^N$ . Here,  $\mathcal{D}$  is the physically allowed domain; for example, it is common for some parameters to be restricted to positive values. It is convenient to interpret both the data and model predictions as vectors in an  $M$ -dimensional *data space*:  $y_m \rightarrow \mathbf{y} \in \mathbb{R}^M$ ,  $f_m(\theta) \rightarrow \mathbf{f} : \mathcal{D} \subseteq \mathbb{R}^N \rightarrow \mathbb{R}^M$ .

The third requirement, a metric, defines a *cost function* (also known as a loss function) that quantifies how well specific parameter values fit the available data,  $C(\theta) = \|\mathbf{y} - \mathbf{f}(\theta)\|$ . The best fit parameters, denoted by  $\theta^*$ , minimize the cost. By far the most common choice for a metric is (weighted) least squares,

$$C(\theta) = \frac{1}{2} \sum_{m=1}^M r_m(\theta)^2, \quad (1)$$

where we have introduced the *residuals*

$$r_m(\theta) = \frac{y_m - f_m(\theta)}{\sigma_m} \quad (2)$$

that depends on the inverse weights  $\sigma_m$  that act as error bars for each data point.

The cost function has a probabilistic interpretation as the negative log-likelihood,

$$P(\mathbf{y}|\theta) \sim \exp(-C(\theta)). \quad (3)$$

Equation (1) corresponds to the case that residuals are independent and identically distributed Gaussian random variables:  $r_m \sim \mathcal{N}(0, 1)$  or, equivalently,  $y_m \sim \mathcal{N}(f_m(\theta), \sigma_m^2)$ . Probability acts as a *measure* on data space that induces a measure on parameters space that we use to quantify uncertainty. For stochastic processes, the stochastic variation in the data is a natural measure, in which case the inverse weights,  $\sigma_m$ , are often taken to be the standard errors estimated from repeated observations. Because they are often associated with experimental error bars, we refer to  $\sigma_m$  as the “error bars” below, even though they may not be explicitly related to any experimental error. When working with DFT data, one does not usually have an error bar that is exactly analogous to an experimental uncertainty. However, there is a growing recognition of the need to estimate the errors in DFT calculations.<sup>55,56</sup> If such error estimates are available, they can be used as weights in the cost function. Because predictions may be made for quantities that carry different physical units (e.g., energies vs forces), choosing  $\sigma_m$  is minimally necessary for Eq. (1) to be dimensionally consistent. Deliberate selection of  $\sigma_m$  is an important part of the UQ problem formulation since any eventual measure of the uncertainty in the model parameters will be derived from the choice of measure in data space.

A common choice for  $\sigma$  that we advocate is to choose the weights to be some fractional value of the data,<sup>57</sup> potentially including an additional cutoff to deal with near-zero data,<sup>58</sup>

$$\sigma_m = \sqrt{c_1^2 + c_2^2 \|y_m\|^2}, \quad (4)$$

where  $c_1$  is the cutoff or padding term and  $c_2$  is a constant that sets of the scale of the uncertainty from the data. In general, choosing  $\sigma_m$  as a fractional tolerance, e.g., 10% of the data, is a reasonable choice and what we use in this study. This choice corresponds to  $c_1 = 0$  and  $c_2 = 0.1$ . [The numerical values of our data are large enough that Eq. (4) is insensitive to the choice of  $c_1$ , so we use zero.]

For IPs, errors in the data are often not the primary source of errors in the inferred model. Because IPs are simplified functional forms that do not capture all of the nuances of quantum mechanics, there is an additional error due to missing physics. We decompose the model error into contributions from the bias,  $b$ , representing

missing physics, and errors in the data,  $\epsilon$ . The data and the model are then related by

$$y_m = f_m(\theta) + b_m + \epsilon_m. \quad (5)$$

Conceptually, the bias can be understood as accuracy of the model, and a fundamental problem is to estimate the bias of an IP for problem-specific applications.

To estimate model accuracy, Frederiksen *et al.*<sup>4</sup> suggested tempering the cost by an effective temperature, which modifies Eq. (3) as

$$P(y|\theta) \sim \exp(-C(\theta)/T). \quad (6)$$

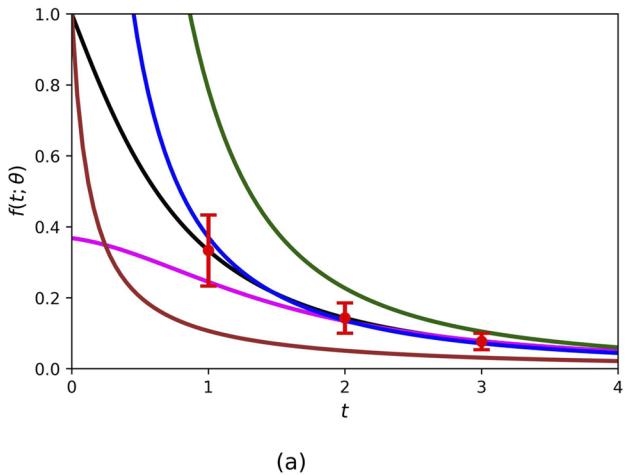
This is to account for the model inadequacy, or bias, as discussed in Ref. 15. Functionally, the temperature  $T$  uniformly scales the weights in the residuals and can be adjusted as a coarse measure of the overall accuracy of the model. In fact, the tempering technique is a common practice to improve UQ performance when the model does not match the data generating process.<sup>59</sup> A natural choice of temperature motivated by analogy with the equipartition theorem<sup>4</sup> is

$$T_0 = \frac{2C_0}{N}, \quad (7)$$

where  $C_0$  is the minimum cost of the model and  $N$  is the degrees of freedom of the model, which often is set to the number of parameters. Since this choice of temperature includes the best fit cost, it incorporates information about the accuracy of the model with respect to the training data. Using this natural temperature, we find that the bias far outweighs the errors in the data and our uncertainty estimates give a rough indication of the accuracy of an IP.

To illustrate key ideas throughout this section, we use a two-parameter toy model of the form

$$f(t; \theta) = \frac{1}{t^2 + \theta_1 t + \theta_2}. \quad (8)$$



(a)

We make predictions at times  $t = 1.0, 2.0, 3.0$ , i.e.,  $f_m(\theta) = f(t_m; \theta)$ . We restrict  $\theta_i \geq 0$  (a common physical constraint on parameter values), which suggests working with the log-transformed parameter values. We use  $\mathbf{y} = [1/3, 1/7, 1/13]^T$  with tolerances  $\sigma_m$  set to be 30% of the data  $\mathbf{y}$  for the purpose of visual clarity. These data, along with model predictions for several values of the parameters, are shown in Fig. 2(a).

It is useful to visually consider cost contours for this model, shown in Fig. 2(b). In general, we are interested in describing the regions in parameter space with low cost. The model in Eq. (8) is sloppy, as manifested by its insensitivity to coordinated variations in some parameter directions. Because of this, the cost contours in the sloppy directions are elongated and the aspect ratio of the low-cost canyon around the minimum is very large. Away from the best fit, many of the cost contours do not close; canyons stretch to the edges of parameter space and flatten into broad plateaus. Extreme values of the parameters can have finite, and in many cases, very small cost. These features are ubiquitous in sloppy models and play a fundamental role in quantifying the parametric uncertainty.

For models with many parameters, direct visualization of the cost surface is not possible. However, we have constructed this toy model to illustrate issues that are typical of high-dimensional parameter spaces. In Secs. II B–II E, we describe several tools for analyzing the cost surface of multi-parameter models.

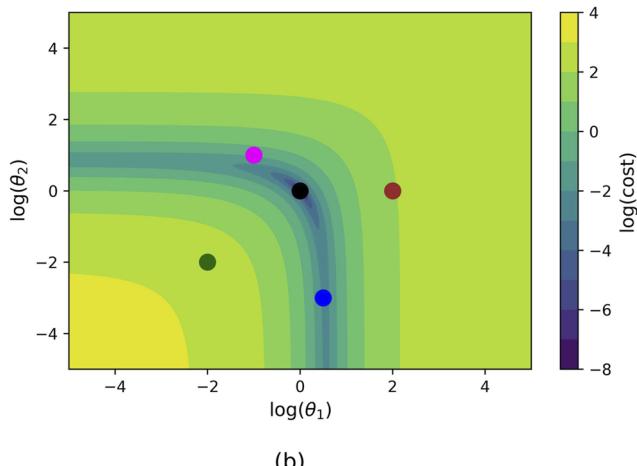
## B. Fisher information: Sloppy model analysis

To quantify the local geometry of the cost surface in a neighborhood of the best fit, we linearize the residuals about  $\theta^*$ ,

$$r_m(\theta) \approx r_m(\theta^*) + \frac{\partial r_m}{\partial \theta}(\theta - \theta^*). \quad (9)$$

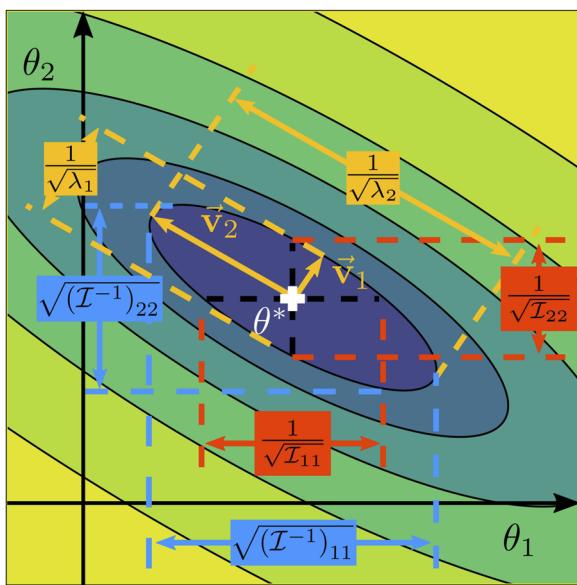
To lowest order, the cost function becomes

$$C(\theta) \approx C(\theta^*) + \frac{1}{2}(\theta - \theta^*)^T (J^T J)(\theta - \theta^*), \quad (10)$$



(b)

**FIG. 2.** Time series (a) and cost surface (b) for the toy model in Eq. (8). Data are given by the red points in (a), with the error bars set to 30% of the data. Model predictions for different parameter values and their corresponding points on the cost surface are shown in matching colors. Contours in (b) represent the errors between the model predictions and data as calculated in Eq. (1).



**FIG. 3.** The Fisher information describes the local geometry of the cost surface. Locally, cost contours are represented by ellipses where the axes are aligned with the eigenvectors of the FIM. The aspect ratio of the ellipse is given by the square root of the ratio of the eigenvalues.

where we have used the fact that  $\nabla C = 0$  at  $\theta^*$  and introduced the Jacobian of the residual function  $J_{mn} = \partial r_m / \partial \theta_n = -(1/\sigma_m) \partial f_m / \partial \theta_n$  evaluated at  $\theta^*$ . The squared Jacobian appearing in Eq. (10) is the Fisher Information Matrix (FIM),

$$\mathcal{I} = J^T J. \quad (11)$$

The FIM is an important statistical quantity; its inverse is a lower bound on the covariance of an unbiased estimator of  $\theta$ , known as the Cramér–Rao bound.<sup>60</sup>

The local geometry of the cost surface around the best fit is described by the FIM, as we illustrate in Fig. 3. Diagonals of the FIM describe the change in cost to each parameter individually, ignoring any potential correlations among parameters. Cost contours form ellipses, aligned with the eigenvectors of the FIM, whose aspect ratio is given by the square root of the ratio of the eigenvalues. Elongated directions are parallel to the eigenvectors with small eigenvalues, indicating that the data carry little information about those parameter combinations. These parameter combinations are only weakly constrained by the data and have large uncertainties in their inferred values. Projecting these ellipses onto the parameter axes estimates the uncertainty in each individual parameter, given by the diagonals of the inverse FIM.

### C. Bayesian analysis

The most common UQ methods in molecular modeling use a Bayesian framework. In Bayesian statistics, the parametric uncertainty is described by a posterior distribution given by Bayes' theorem,

$$P(\theta|y) \propto L(\theta|y) \times \pi(\theta), \quad (12)$$

where  $L(\theta|y)$  and  $\pi(\theta)$  are the likelihood and the prior distribution of the model's parameters, respectively.<sup>61,62</sup> The likelihood is functionally the same as the probability distribution of the observed data conditioned on the different values of the parameters, i.e.,  $L(\theta|y) = P(y|\theta)$ . As before, we temper the likelihood with a tunable “sampling temperature.” In a Bayesian context, this introduction is motivated by a formal analogy to the Boltzmann distribution.<sup>7,63,64</sup> The cost is analogous to the internal energy of a system, so low-temperature distributions are concentrated near the low-energy (i.e., low cost) region of parameter space. Formally, the temperature uniformly scales the tolerances  $\sigma_m$  in Eq. (2). In our case, the temperature scales the tolerances to be reflective of the model's accuracy and not just the error in the data itself. Continuing the analogy, Eq. (12) becomes

$$P(\theta|y) \sim \exp(-(C(\theta) - S(\theta)T)/T), \quad (13)$$

in which the prior is analogous to entropy:  $S = \log \pi$ .

We sample from the posterior distribution using an MCMC algorithm. There are several black-box libraries for MCMC sampling. In this work, we used the `ptemcee` Python package, which utilizes an affine invariance property of the sampler.<sup>65,66</sup> In addition, this method generates chains at different temperatures and mixes them with an appropriate acceptance probability.<sup>67</sup> Traditionally, parallel tempering is utilized as a device to improve convergence rates by allowing walkers to skip over regions in parameter space with higher cost values and possibly find different minima (if they exist).<sup>59</sup> Our use of the sampling temperature to estimate the scale of the model accuracy [see Eq. (5)]. Previous results have shown that a natural sampling temperature given by  $T_0 = 2C_0/N$  [see Eq. (7)] gives a good estimate of the systematic errors in the model.<sup>4</sup> Thus, raising the temperature from  $T = 1$  to  $T = T_0$  transitions the sampling from the target accuracy to a more realistic estimate of the actual systematic errors.

For the toy model in Eq. (8) and Fig. 2, data are generated from the model itself, i.e., there is no bias. In this case, varying the sampling temperature effectively changes the size of the error bars  $\sigma_m$  in the model. Of particular interest is how the samplers interact with the plateaus and canyons of the cost surface as the temperature is varied.

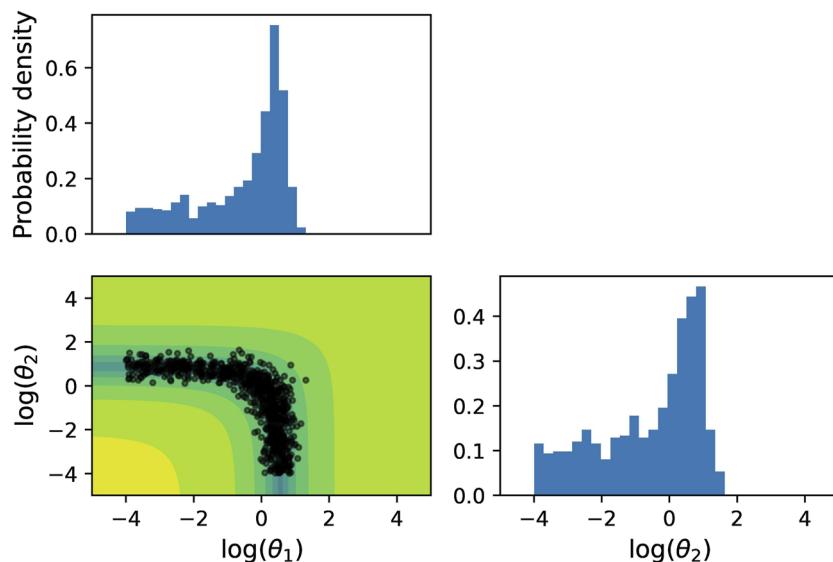
To assess the convergence, we simulate multiple chains and use the multivariate potential scale reduction factor (PSRF), denoted by  $\hat{R}^P$ .<sup>68–70</sup> The value of  $\hat{R}^P$  is related to the ratio of the covariance between and within the independent chains, given by

$$\hat{R}^P = \frac{n-1}{n} + \frac{m+1}{m} \lambda_{\max}(W^{-1}B/n), \quad (14)$$

where  $n$  and  $m$  are the numbers of iterations and chains, respectively, and  $\lambda_{\max}(A)$  denotes the largest eigenvalue of matrix  $A$ .  $B/n$  and  $W$  are the variance between and within the independent chains,  $\psi_j$ ,

$$\begin{aligned} \frac{B}{n} &= \frac{1}{m-1} \sum_{j=1}^m (\bar{\psi}_j - \bar{\psi})(\bar{\psi}_j - \bar{\psi})^T \\ W &= \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{t=1}^n (\psi_{jt} - \bar{\psi}_j)(\psi_{jt} - \bar{\psi}_j)^T. \end{aligned} \quad (15)$$

Note that as  $n \rightarrow \infty$ , the value of  $\hat{R}^P$  declines to one.<sup>68,69</sup> Thus, as the MCMC samples converge to a stationary distribution, the value



**FIG. 4.** MCMC samples of the model defined by Eq. (8). In the lower left frame, the samples are plotted on top of the cost contour. The univariate marginal distributions of the samples are given on the diagonal. We pick the sampling temperature  $T = 0.5$  to highlight the effect of sloppiness to the Bayesian sampling, i.e., parameter evaporation.

of  $\hat{R}^p$  approaches one; however, the converse is not necessarily true. Common thresholds of  $\hat{R}^p$  are in the range of 1.1 to 1.05.<sup>69</sup> In this work, we have used the more stringent requirement ( $\hat{R}^p < 1.05$ ).

To illustrate, we sample the posterior of the model in Eq. (8) with a uniform prior distribution that is non-zero over  $(-4, 4)$ . Figure 4 shows the result of the sampling as an array of plots that summarize the sample. Along the main diagonal, we plot the univariate marginal distributions, i.e., projection of the samples' distribution onto a parameter axis. In the lower left frame of the array, we show samples in two-dimensional parameter space. In higher dimensions, these plots are two-dimensional marginal distributions. In this example, we have superimposed the samples on top of the cost contours [compare this to Fig. 2(b)].

In Fig. 4, note that the MCMC samples are located along the canyon. This is expected as the posterior is large in regions of high likelihood, i.e., low cost. Thus, the samples accurately conform to the cost contours and quantify the uncertainty in the parameter estimates. However, by inspecting Fig. 4, we can anticipate a potential problem when the cost contours have flat, elongated canyons that extend to extreme parameter values. In this scenario, it will be common to “evaporate” parameters, i.e., have samples that extend over the full range of model parameters. Parameter evaporation was first observed when sampling posterior distributions for sloppy models in systems biology;<sup>50</sup> however, the phenomenon occurs in molecular models as well, as we document below. The term parameter evaporation continues the analogy to statistical mechanics. In the language of statistical inference, one would say that the posterior does not concentrate with respect to the prior and is *unidentifiable*.

Closely related to the parameter evaporation is the question of choosing the prior distribution,  $\pi(\theta)$ , on parameter space. Here, we have used a uniform prior, a common choice for an uninformative prior. However, we find the prior can strongly influence the posterior distribution in subtle ways. Even the apparently innocuous uniform prior can introduce strong biases. Note that in Fig. 4, the samples stop evaporating due to the boundaries of the prior distribution. Consequently, the marginal distribution of the samples has

a hard cutoff at this boundary. With a broader prior, the posterior will be even wider, and samples may no longer be concentrated near the best fit. The long canyons and broad plateaus of the cost surface would dominate the samples. We intuitively explain this effect in terms of a trade-off between energy and entropy in the sampling process. While the single most probable parameter value is the best fit (i.e., ground state), there are many more sub-optimal parameter values along the canyon. In other words, the broad prior introduces a large entropy in some regions of the parameter space. For broad priors, the entropic contribution dominates the sampling.

The effect is clearly demonstrated with a simple example. Consider a one-dimensional cost function given by

$$C(\theta) = \begin{cases} C_0, & \theta < l, \\ C_0 + \Delta, & \theta > l, \end{cases} \quad (16)$$

where  $\theta$  is a non-negative parameter and  $C_0$ ,  $\Delta$ , and  $l$  are non-negative constants. We take a prior  $\pi(\theta) = \mathcal{U}(0, L)$ , i.e., a uniform, flat prior from zero to a positive value of  $L$ . After calculating the posterior distribution, we find the average cost for this scenario to be

$$\langle C \rangle = \begin{cases} C_0, & L < l, \\ C_0 + \Delta \left( \frac{L - l}{L + l(e^{\Delta/T} - 1)} \right), & L > l. \end{cases} \quad (17)$$

Note that for very large  $L$  (i.e., very broad prior),  $\langle C \rangle \rightarrow C_0 + \Delta$ . That is to say, for a sufficiently broad prior, the posterior distribution is dominated by bad fits (large cost) because of their large entropic contribution. This result holds for any non-zero sampling temperature and regardless of how bad the fit is (i.e., the size of  $\Delta$ ).

As we will see in Sec. III, the trade-off between entropy and energy is even more nuanced for “sloppy” cost landscapes with many dimensions. Rarely is there an objectively “correct” prior, and any choice is almost certain to introduce artifacts into the statistics of the posterior. In these cases, it is unclear to what extent the posterior

accurately reflects the target uncertainty. One solution is to sample with multiple priors and temperatures, a computationally expensive task, and try to assess the effect of prior and sampling temperature on the results. Because this practice generates multiple posteriors, it potentially undermines the Bayesian paradigm in which a single posterior summarizes all the information one has about the parameters of a model. However, we believe that these extra steps are an important intermediate analysis in understanding the effect of the prior on the posterior and necessary for constructing a reliable posterior. Alternatively, one could use a formalism that does not require an *a priori* measure on parameter space. This is the domain of frequentist statistics, which we discuss next.

#### D. Frequentist analysis

Although much has been said about the philosophical differences between Bayesians and frequentists,<sup>71,72</sup> here we use a functional distinction. The choice of prior in Sec. II C was a central question. The prior acts as a measure on parameter space, i.e., a weight function whose integral generalizes the concept of lengths and volumes.<sup>73</sup> In the frequentist approach, no such measure exists. Without a measure on parameter space, we lose the machinery of a posterior distribution, but we also need fewer mathematical assumptions. With no prior, the goal is to describe the set of parameter values that have small cost (i.e., below some statistically defined threshold) without attaching any (probabilistic) weight to regions of parameter space. While there are many frequentist tools available, we use the profile likelihood.<sup>74</sup> To the best of our knowledge, profile likelihoods have not been previously applied to the study of interatomic potentials. Importantly, profile likelihoods generate paths through the parameter space that we will compare with the geodesic curves generated by the information geometric analysis in Sec. II E.

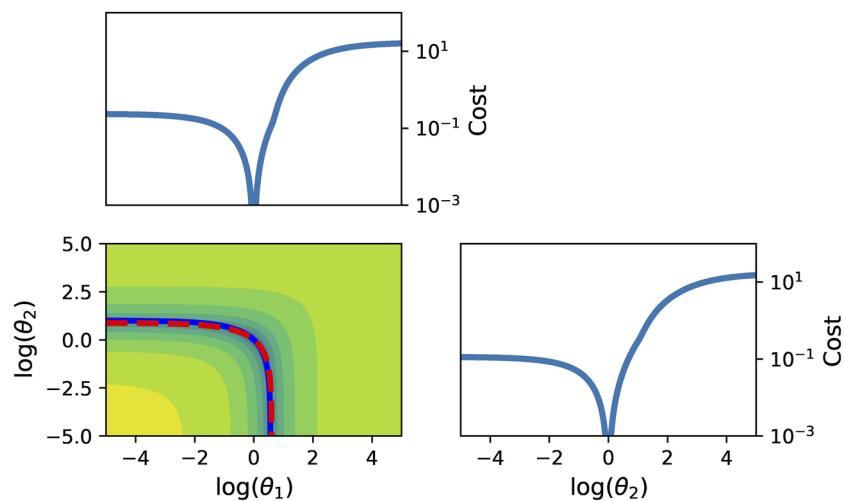
The basic idea is to select one parameter, fix it to a constant value, and globally optimize the likelihood function (i.e., minimize the cost) over the remaining  $N - 1$  model parameters.<sup>74,75</sup> By varying the value to which the parameter is fixed, we trace out a “profile” of

how the cost depends on this parameter in the context of the rest of the model. The procedure is best understood through example, as we now demonstrate on the toy model from Eq. (8) in Fig. 5. As before, we summarize results with a two-by-two array of figures. Consider the cost contours in the lower left frame of the plot array. The red curve is the set of points obtained by fixing  $\theta_1$  to a constant value and optimizing the cost over  $\theta_2$ . The optimization searched over vertical slices of the parameter space for each value of  $\theta_1$ . Similarly, the blue curve is the set of points found by fixing  $\theta_2$  and optimizing over horizontal slices, i.e., over  $\theta_1$ . Along the main diagonal, we plot the cost along each of these profile likelihood paths.

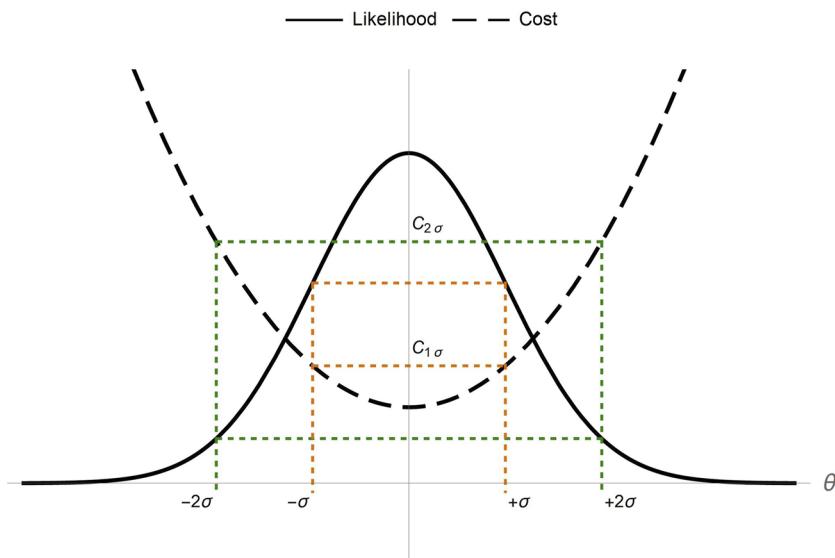
By construction, the profile likelihood paths trace out the canyon on the cost contour. By comparing the paths and the cost along the paths, we extract information about how variation of the parameters affects the variation in the predictions. The paths also tell us how the parameters correlate with each other. Statistical confidence levels correspond to an allowed error or cost threshold. For a given confidence level, the uncertainty of an individual parameter is given by the width of the profile likelihood that has cost values lower than this cost threshold. This idea is illustrated in Fig. 6 for a simple Gaussian likelihood function.

To calculate profile likelihoods, we developed a Python package, `profile_likelihood`,<sup>76</sup> that additionally interfaces with IPs from the OpenKIM repository at <https://openkim.org>. We use the Levenberg–Marquardt algorithm with geodesic acceleration in the optimization process.<sup>77</sup>

The profile likelihood analysis method has its own challenges and limitations. Optimizing multi-dimensional cost functions can be challenging;<sup>8</sup> however, by using the result of the previous optimization as the starting point of each iteration, convergence is relatively fast and stable. Additionally, the profiling process effectively projects parameter curves onto the parameter axes. As we will see in Sec. III, if the cost canyon curves or bends over on itself, it will be missed by the profile likelihood. More broadly, the construction is not invariant to reparameterization. Just as priors may introduce artifacts in the Bayesian framework, the parameterization can introduce artifacts into the profile likelihood. To avoid these



**FIG. 5.** Profile likelihood for the model in Eq. (8). In the lower left frame, the red and blue curves show the paths traced from the profile likelihood computation for the parameter on the horizontal and vertical axes, respectively. The cost profiles, i.e., the cost along these paths, are given on the diagonal.



**FIG. 6.** Confidence intervals for a Gaussian likelihood at several confidence levels.  $C_{1\sigma}$  and  $C_{2\sigma}$  correspond to the cost threshold at 68% and 95% confidence levels, respectively. The confidence interval of  $\theta$  given 68% confidence level spans from  $-\sigma$  to  $+\sigma$ .

issues, we next use information geometry to study the uncertainty in a parameterization-independent way.

### E. Information geometry

Information geometry is an approach to statistics in which a multi-parameter model is interpreted as a high-dimensional manifold. We study this manifold using computational differential geometry that allows us to extract the key geometric and topological features of the model manifold. These features shed light on issues related to UQ.

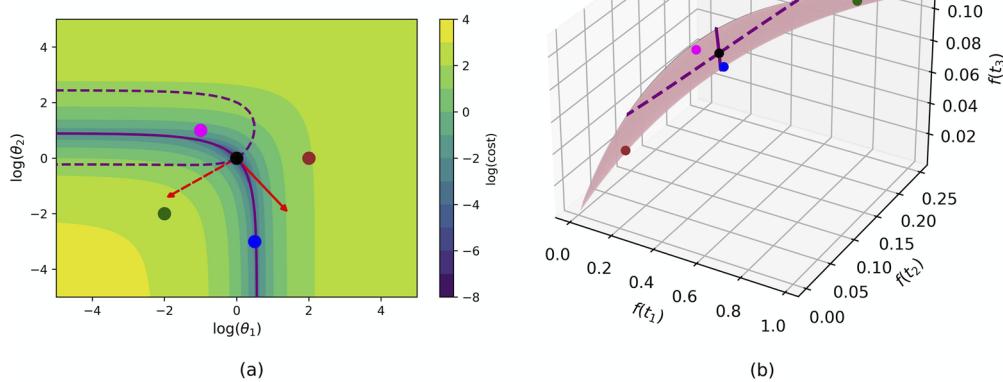
As we have seen in Sec. II A, a multi-parameter model makes a set of predictions for a given evaluation scenario,  $f_m(\theta)$ , that we

interpret as a vector in data space. That is to say, the model is a mapping between parameter space and data space,

$$f : \mathcal{D} \subseteq \mathbb{R}^N \rightarrow \mathbb{R}^M. \quad (18)$$

Conceptually, the model manifold is constructed by mapping all possible parameter values to their corresponding predictions in data space, i.e., the model manifold is the image of parameter space under the model map, illustrated for the toy model [Eq. (8)] in Fig. 7.

Critically, note that the manifold is bounded by two one-dimensional segments. We focus on the boundaries of the model manifold as they are the geometric feature most relevant to parameter uncertainty. In parameter space, we have seen that there can



**FIG. 7.** Model manifold of the toy model in Eq. (8). Each point in parameter space (a) corresponds to a set of predictions in data space (b). The set of all possible predictions trace out the model manifold. For reference, colored points are the same as those in Fig. 2. The red arrows in parameter space show the eigenvectors of the FIM. Geodesics (solid and dashed curves) relate manifold structures, such as manifold boundaries, to parameter space.

be large or infinite uncertainties when cost contours do not close, i.e., confidence regions extend to the limits of the parameter domain. Infinite, high-entropy regions of parameter space are mapped to finite regions near the boundary of the model manifold, and thus, these contours, i.e., the contours that do not close, are generic when the model manifold is bounded. Since the cost corresponds to the distance in data space, the images of cost contours are approximately concentric circles on the model manifold. Each segment of the boundary is a manifold of co-dimension one. Associated with each boundary segment is a parameter or combination of parameters that are *practically unidentifiable* at some level of statistical confidence.

We use geodesics, i.e., distance minimizing curves, on the model manifold to find the unidentifiable parameter combination associated with each boundary segment.<sup>40</sup> Geodesics are a natural extension of the local FIM analysis. As such, the boundaries and parameter combinations they find identify the global sloppiness of the model. Geodesics traverse the low cost regions in parameter space to find unidentifiable parameter combinations—those whose uncertainty will be large—which can be removed to reduce the complexity and sloppiness of the model.

We approximate geodesics curves along the model manifold by numerically solving the geodesic equation,

$$\frac{\partial^2 \theta^i}{\partial \tau^2} = - \sum_{j,k} \Gamma_{jk}^i \frac{\partial \theta^j}{\partial \tau} \frac{\partial \theta^k}{\partial \tau}, \quad (19)$$

where

$$\Gamma_{jk}^i = \sum_{l,m} (\mathcal{I}^{-1})^{il} \frac{\partial y_m}{\partial \theta^l} \frac{\partial^2 y_m}{\partial \theta^j \partial \theta^k} \quad (20)$$

are the so-called Christoffel symbols,  $\mathcal{I}$  is the FIM,  $\theta$  are the parameters, and  $\tau$  is the arc length of the geodesic along the model manifold. We numerically solve for the path of the geodesic by treating this equation as an initial value problem where the initial position is given by the nominal parameter values. The initial direction is given by a few of the sloppiest eigenvectors. We provide a simple example script for calculating geodesics on GitHub.<sup>78</sup> To illustrate this, we again turn to the toy model in Eq. (8). Figure 7 shows two geodesics that pass through the best fit point on the model manifold, the solid and dashed curves in both parameter space (a) and on the model manifold (b). Note how the geodesics rotate in parameter space to naturally follow the cost contours and align with the unidentifiable parameters. As the geodesic curves approach the edge of the model manifold, we see that either  $\theta_1 \rightarrow 0$  or  $\theta_2 \rightarrow 0$ . From the correspondence between the parameter space picture and the data space picture, we deduce that the upper boundary segment on the model manifold corresponds to  $\theta_1 \rightarrow 0$ , while the lower segment corresponds to  $\theta_2 \rightarrow 0$ . These limiting values indicate which combinations of parameters are unidentifiable and have unbounded uncertainties. We often calculate several geodesics emanating from the nominal parameter values in several of the sloppiest eigenvectors of the FIM. These geodesics will help us identify the least identifiable parameter (combinations) in the model.

In this simple example, the two boundary segments are already aligned with the parameters of the model. In more realistic models,

boundary segments often correspond to the coordinated combinations of bare parameters. In these cases, we will use geodesics to identify the correlation and find a more natural, identifiable reparameterization.

## F. Interatomic potentials and tests

In this study, we apply the methods described in Secs. II B–II E to empirical IPs taken from the OpenKIM repository.<sup>12,79</sup> The OpenKIM framework has a standardized collection of models, data, and tests for computing materials properties that make the UQ process reproducible and transferable. For this study, we chose the Lennard-Jones (LJ) and Morse potentials parameterized for silicon and nickel, respectively, to validate methods and demonstrate general principles on low-dimensional models.<sup>80</sup> We then extend the investigation to the molybdenum disulfide (MoS<sub>2</sub>) system using the more complex Stillinger–Weber (SW) potential.<sup>2</sup>

These potentials are categorized as cluster potentials. Given a system with  $N$  atoms, the total potential energy,  $\mathcal{V}$ , is

$$\mathcal{V} = \sum_{\substack{i, j=1 \\ i < j}}^N \phi_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{\substack{i, j, k=1 \\ i < j < k}}^N \phi_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots, \quad (21)$$

where  $\phi_n$  denotes the  $n$ -body potential function and  $\mathbf{r}_i$  is the position of atom  $i$ .

The Lennard-Jones (LJ) potential is a pair potential, i.e., Eq. (21) only consists of the two-body (pair-wise) interaction term, and the higher order potential functions are set to zero. The pair-wise interaction has two parameters, given by

$$\begin{aligned} \phi_{\text{LJ}}(r_{ij}) &= 4\epsilon \left( \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right) + \Delta, \\ \Delta &= -4\epsilon \left( \left( \frac{\sigma}{r_{\text{cut}}} \right)^{12} - \left( \frac{\sigma}{r_{\text{cut}}} \right)^6 \right), \end{aligned} \quad (22)$$

where  $r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$  is the distance between atoms  $i$  and  $j$ . The potential is only non-zero when  $r_{ij} < r_{\text{cut}} = 7.9118 \text{ \AA}$ . The parameter  $\epsilon$  is an energy scaling factor in the potential, while  $\sigma$  is related to the equilibrium distance of the pair interaction. The shifting factor,  $\Delta$ , is chosen so that the potential is continuous at  $r_{\text{cut}}$ .<sup>81–84</sup>

The Morse potential is also a pair potential, similar to the LJ potential. The pair-wise interaction with three parameters is given by

$$\begin{aligned} \phi_{\text{M}}(r_{ij}) &= \epsilon \left( -e^{-2C(r_{ij} - r_0)} + 2e^{-C(r_{ij} - r_0)} \right) + \Delta, \\ \Delta &= -\epsilon \left( -e^{-2C(r_{\text{cut}} - r_0)} + 2e^{-C(r_{\text{cut}} - r_0)} \right), \end{aligned} \quad (23)$$

where  $\epsilon$  is an energy scaling factor,  $r_0$  is the equilibrium distance, and  $C$  controls the width of the potential well. Again, the potential is only non-zero when  $r_{ij} < r_{\text{cut}} = 9.75476 \text{ \AA}$  and  $\Delta$  is chosen such that the potential is continuous at  $r_{\text{cut}}$ .<sup>85,86</sup>

We use these pair potentials to predict the unrelaxed energy and forces of silicon (LJ potential) and nickel (Morse potential) atoms in

an evaluation scenario comprised of a randomly perturbed body-centered triclinic configuration with periodic boundary conditions. The lattice parameters are given as follows:<sup>87,88</sup>

$$\begin{aligned} a &= 3.128\,752\,5 \text{ \AA}, & \alpha &= 87.253\,180\,549\,684\,44^\circ, \\ b &= 3.151\,46 \text{ \AA}, & \beta &= 93.340\,747\,774\,135\,02^\circ, \\ c &= 3.131\,210\,44 \text{ \AA}, & \gamma &= 91.231\,344\,620\,111\,88^\circ. \end{aligned}$$

We also use a random triclinic configuration with 64 silicon atoms<sup>89</sup> with several other cluster potentials in a broader survey of the Fisher information. We generate artificial data for both of these models using the default parameter values reported in OpenKIM. We calculate the energy and forces evaluated at the default parameters and add 10% Gaussian noise (for forces, we use 10% of the magnitude of the force vector). The introduction of the noise in the data results in new sets of best fit parameters, which for the LJ potential are  $\epsilon = 3.345\,459\,22$  eV and  $\sigma = 1.981\,715\,08$  \AA, while for the Morse potential are  $\epsilon = -0.243\,660\,83$  eV,  $C = 0.655\,085\,52$  \AA<sup>-1</sup>, and  $r_0 = 4.487\,043\,15$  \AA. Note that the default parameter values were found by fitting to experimental data.<sup>90,91</sup> Because these original training data were not generated by the model, the parameter values reflect some model inadequacy. We will later explore how the parameter uncertainty responds to the scale of model inadequacy by tempering this likelihood for a range of sampling temperatures.

Note that these structures are not the ground states for silicon or nickel, but they are tests available in OpenKIM that are convenient for validating methods on low-dimensional models. We will see that they clearly illustrate the problems that the sloppiness of the model brings to the standard UQ methods and the phenomena that we will discuss later are generic, at some confidence level, to any atomic configuration used.

We extend the analysis to the Stillinger–Weber (SW) potential for monolayer MoS<sub>2</sub>,<sup>2,92,93</sup> which contains both two-body and three-body interactions. The two-body interaction takes the form

$$\phi_2^{IJ}(r_{ij}) = A_{IJ} \left( B_{IJ} \left( \frac{\sigma_{IJ}}{r_{ij}} \right)^{p_{IJ}} - \left( \frac{\sigma_{IJ}}{r_{ij}} \right)^{q_{IJ}} \right) \exp \left( \frac{\sigma_{IJ}}{r_{ij} - r_{IJ}^{\text{cut}}} \right), \quad (24)$$

where uppercase subscripts denote the types of atoms, e.g.,  $A_{IJ}$  is the parameter  $A$  corresponding to interaction between atoms of type  $I$  and type  $J$ . The three-body term is given by

$$\begin{aligned} \phi_3^{IJK}(r_{ij}, r_{ik}, \beta_{jik}) &= \lambda_{IJK} (\cos \beta_{jik} - \cos \beta_{jik}^0)^2 \\ &\times \exp \left( \frac{\gamma_{IJ}}{r_{ij} - r_{IJ}^{\text{cut}}} + \frac{\gamma_{IK}}{r_{ik} - r_{IK}^{\text{cut}}} \right), \end{aligned} \quad (25)$$

with  $\beta_{jik}$  being the angle between the  $i-j$  and  $i-k$  bonds.

We calibrate this potential to DFT data of the atomic forces in configurations near the equilibrium state at 750 K, as described in Ref. 2. Since the data are not generated from the IP, they contain bias and we will temper the likelihood over a range of temperatures to account for it. Our formulation follows closely that of the original paper (e.g., we set  $q_{IJ} = 0$ , fix  $\gamma$  to be the same for all types of interaction, and use the same training set); however, we make a few

TABLE I. Fitted parameters of the two-body term in the SW potential for MoS<sub>2</sub>.

| Parameter      | Interaction  |              |              |
|----------------|--------------|--------------|--------------|
|                | Mo–Mo        | Mo–S         | S–S          |
| $A$ (eV)       | 18.431 006 0 | 8.838 613 05 | 0.374 633 96 |
| $B$            | 0.006 417 86 | 1.047 936 03 | 561.429 270  |
| $p$            | 4.737 178 13 | 8.266 217 44 | 2.661 969 13 |
| $\sigma$ (\AA) | 6.169 404 54 | 1.929 679 91 | 0.419 048 14 |

TABLE II. Fitted parameters of the three-body term in the SW potential for MoS<sub>2</sub>.

| Parameter                              | Value        |
|--|--------------|
| $\lambda_{S\text{-}Mo\text{-}S}$ (eV)  | 4.287 840 76 |
| $\lambda_{Mo\text{-}S\text{-}Mo}$ (eV) | 14.428 502 6 |
| $\gamma$ (\AA)                         | 1.538 005 00 |

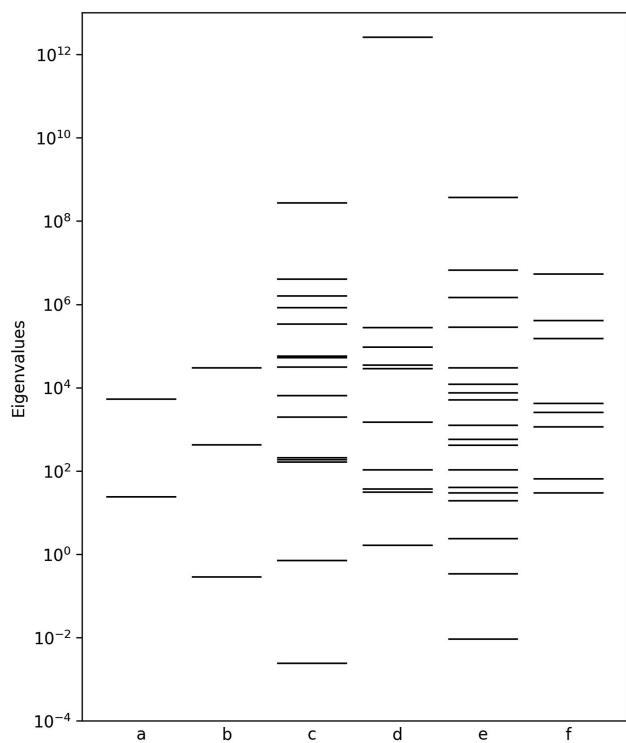
changes. First, we allow parameters  $p_{IJ}$  to take any positive real value and remove the relation between  $\sigma_{IJ}$  and the equilibrium lattice constants of the system. We also do not require  $d\phi_2/dr|_{r=d} = 0$  at the equilibrium bond length  $d$ , which removes the constraint on  $B_{IJ}$ . The remaining free parameters are  $A_{IJ}$ ,  $B_{IJ}$ ,  $p_{IJ}$ , and  $\sigma_{IJ}$  for each type of pair-wise interaction (Mo–Mo, Mo–S and S–S interactions),  $\lambda_{IJK}$  for S–Mo–S and Mo–S–Mo interactions, and  $\gamma$ .

We again choose error tolerances to be 10% of predicted values. Note that this leads to non-uniform weighting factors in our cost function, unlike Ref. 2. Fitting this model leads to a new set of optimal parameter values listed in Table I for the two-body interaction term and Table II for the three-body interaction term.<sup>94</sup> Other parameters that are not listed in these tables take the same values as listed in Ref. 2, such as the cutoff radii and the reference bond angle. The cost at the best fit is  $1.390 \times 10^6$ . Because the fitting data are forces near equilibrium, the error bars are very small (leading to a large cost) with larger weight on configurations near equilibrium. However, in our Bayesian analysis, we sample the posterior at many temperatures, effectively scaling these small error bars up to something more reasonable. The process provides a systematic study of the role of error bars in quantifying parametric uncertainty in sloppy, molecular models.

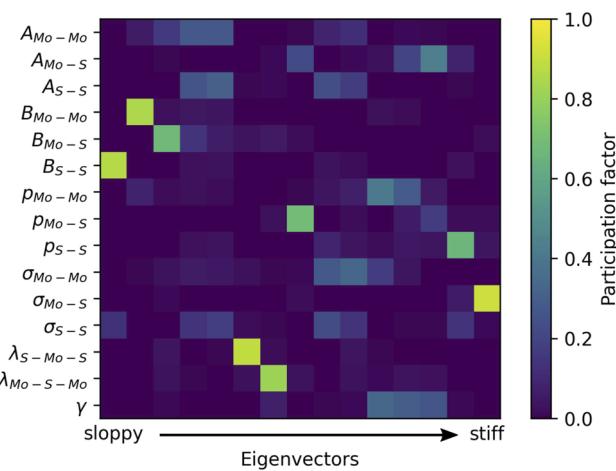
After the calibration process, we propagate the parametric uncertainty of this potential in a second evaluation scenario to predict the uncertainty of the change in energy as a response to the lattice stretching and compression. This calculation is done by creating MoS<sub>2</sub> unit cells with various in-plane lattice constants  $a$  and then relaxing the atoms in the perpendicular, out-of-plane, direction. We probe the calculation in the range  $(a - a_0) \in [-0.5, 0.5]$  \AA, where  $a_0$  is the equilibrium lattice constant. Then, we compare the uncertainty to the result in Ref. 2 qualitatively.

### III. RESULTS

Figure 8 shows the eigenvalues of the FIM for the models in Sec. II F, evaluated at the nominal values of the parameters. We use these models to perform energy and forces calculations, as the

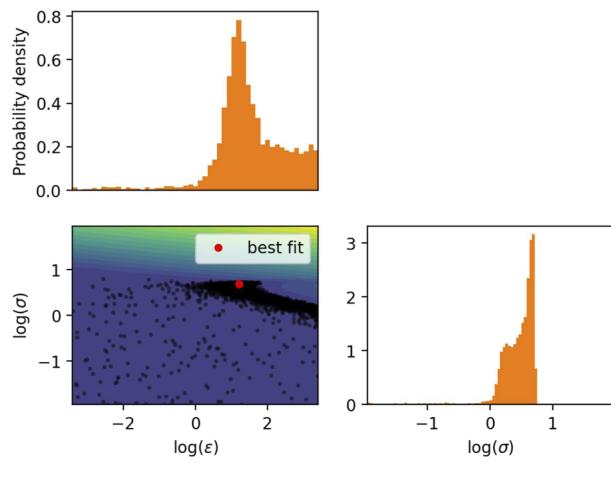
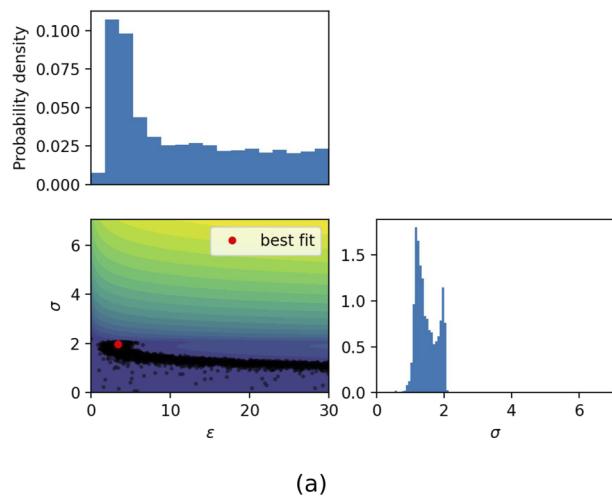


**FIG. 8.** Eigenvalues of the FIM for various IPs: (a) LJ for Si, (b) Morse for Ni, (c) SW for  $\text{MoS}_2$ , and (d) Khor-Das Sarma, (e) EDIP, and (f) SW, each for Si. For each model, the larger (smaller) eigenvalues represent stiff (sloppy) parameter combinations in the direction of their respective eigenvectors.

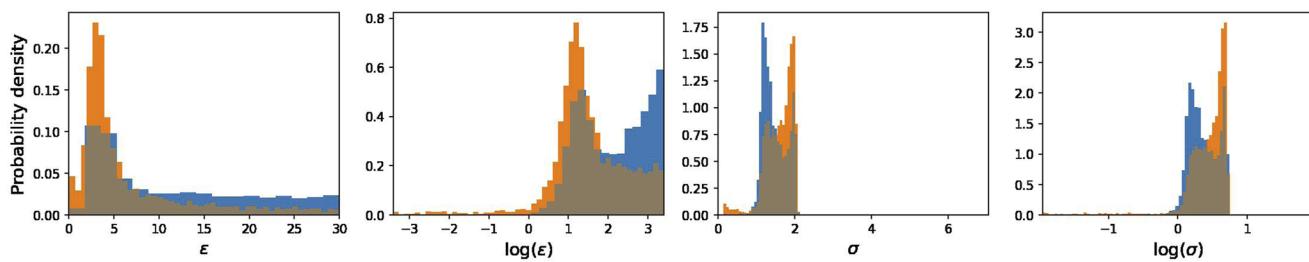


**FIG. 9.** Participation factor of the SW potential for  $\text{MoS}_2$ , calculated as the element-wise square of the eigenvectors of the FIM. Each column corresponds to an eigenvector, increasing in stiffness from left to right. The participation factor shows how much each parameter contributes to each eigenvector. The parameter direction is given by the logarithm of the labels on the vertical axis. The sloppiest eigenvector (the left most column) is mostly in the  $\log(B_{S-S})$  direction.

first evaluation scenario, as explained in Sec. II F. Note that they are sloppy; the eigenvalues cover many orders of magnitude, indicating that many parameters are unidentifiable from the data. To illustrate that sloppiness is a general property of IPs, we also include the eigenvalues of the FIM, evaluated at the fitted parameters, for the Khor-Das Sarma potential<sup>95,96</sup> (three-body potential), Environment Dependent Interatomic Potential (EDIP)<sup>97–99</sup> (bond-order potential), and the original SW potential for silicon<sup>100–103</sup> in predicting the energy and forces of the atoms in a random triclinic silicon



**FIG. 10.** MCMC samples for the LJ potential, sampled in (a) linear and (b) log parameter scales, at sampling temperature  $T = 21.5$ . The original parameterization of the potential is given by the red dot. The samples are plotted against the cost contour on the lower left frame on each figure, with samples condensed around the low cost canyon. The marginal distributions are shown on the diagonal for each figure.



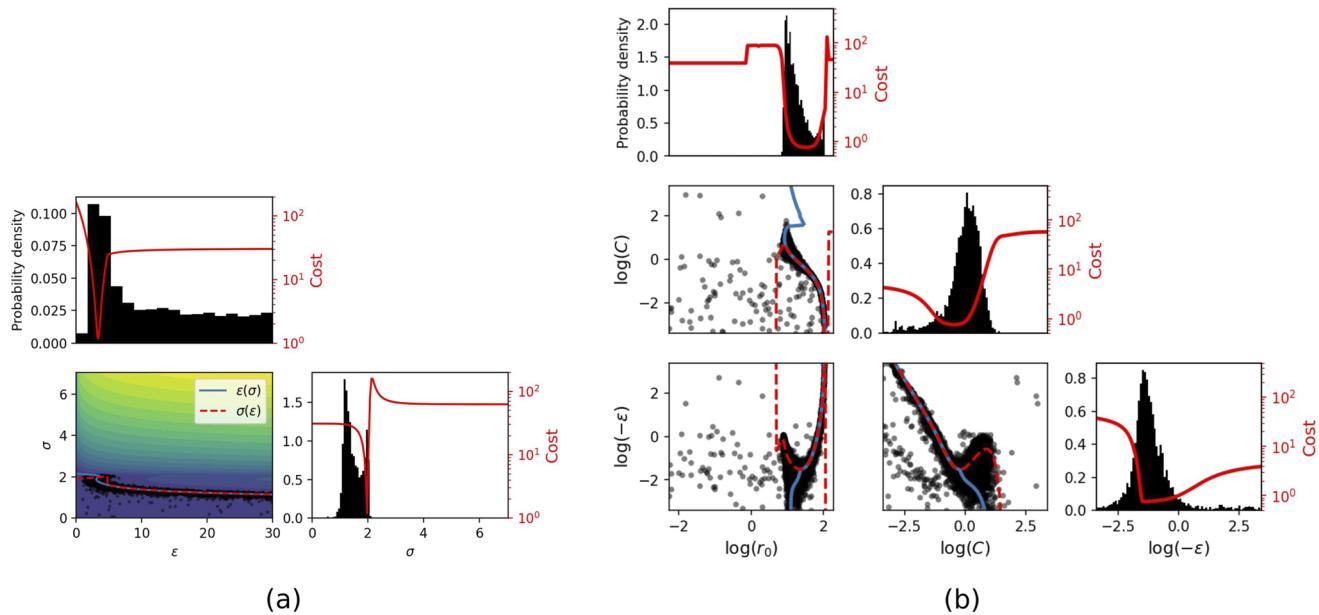
**FIG. 11.** Comparison of the marginal distributions for the MCMC simulations sampled in linear (blue) and log (orange) parameter scales. A uniform prior was used in both cases. Differences in the posterior distributions reflect the role of parameter scaling and choice of prior.

configuration<sup>88,89</sup> (the original parameters for these potentials can be found in OpenKIM<sup>104–106</sup>). Figure 9 shows the participation factor<sup>107,108</sup> for the SW MoS<sub>2</sub> model, i.e., how much each parameter contributes to each eigenvector. Participation factors are calculated as the element-wise square of the eigenvectors of the FIM. We conclude that the sloppiest direction, indicated by the eigenvector with the smallest eigenvalue, is dominated by the parameter  $B_{S-S}$ . Similarly, we can read off the participation factors of each parameter in the other eigendirections.

The FIM is a local calculation and computationally inexpensive compared to other methods discussed here. As a result, we recommend using the FIM as an initial step to UQ. We will revisit the

results from the FIM when we extend the analysis and compare the results to more global methods.

We now consider the results of the Bayesian analysis for the LJ potential. Bayesian analysis requires a prior and a common choice by molecular modelers is the uniform prior. Figure 10 shows these results sampled for uniform on both the linear [Fig. 10(a)] and log scales [Fig. 10(b)]. In both cases, we use a uniform prior in their respective parameter space, bounded by a rectangular region defined by  $0 < \epsilon < 30$  and  $0 < \sigma < 2^{-1/6}r_{\text{cut}}$  in linear parameter space and  $|\log(\epsilon)| < \log(30)$  and  $|\log(\sigma)| < \log(2^{-1/6}r_{\text{cut}})$  in log parameter space. The upper bound of  $\sigma$  is chosen so that the pair-wise equilibrium length is less than the cutoff distance. Although both priors



**FIG. 12.** MCMC samples and profile likelihood for the (a) LJ and (b) Morse potentials at sampling temperatures  $T = 21.5$  and  $T = 10.0$ , respectively. We plot the cost surface for LJ because it only contains two parameters; in general, it is not possible to plot the cost surface, e.g., for Morse. On the lower triangle frames, the MCMC samples are plotted as the black points, while the red and blue curves show the profile likelihood paths for the parameters on the horizontal and vertical axes, respectively. On the diagonal, we superimpose the cost profiles (red curves) on top of the marginal distribution of the MCMC samples. These plots show qualitative agreement between the two methods for low-dimensional models. MCMC samples are concentrated around the profile likelihood paths, indicating that the sampling is energy-dominated, and there are no significant artifacts from the choice of prior. However, there are signs of large entropy regions that could dominate the sampling at higher temperatures, e.g., evaporation at large negative values of  $\log(r_0)$ .

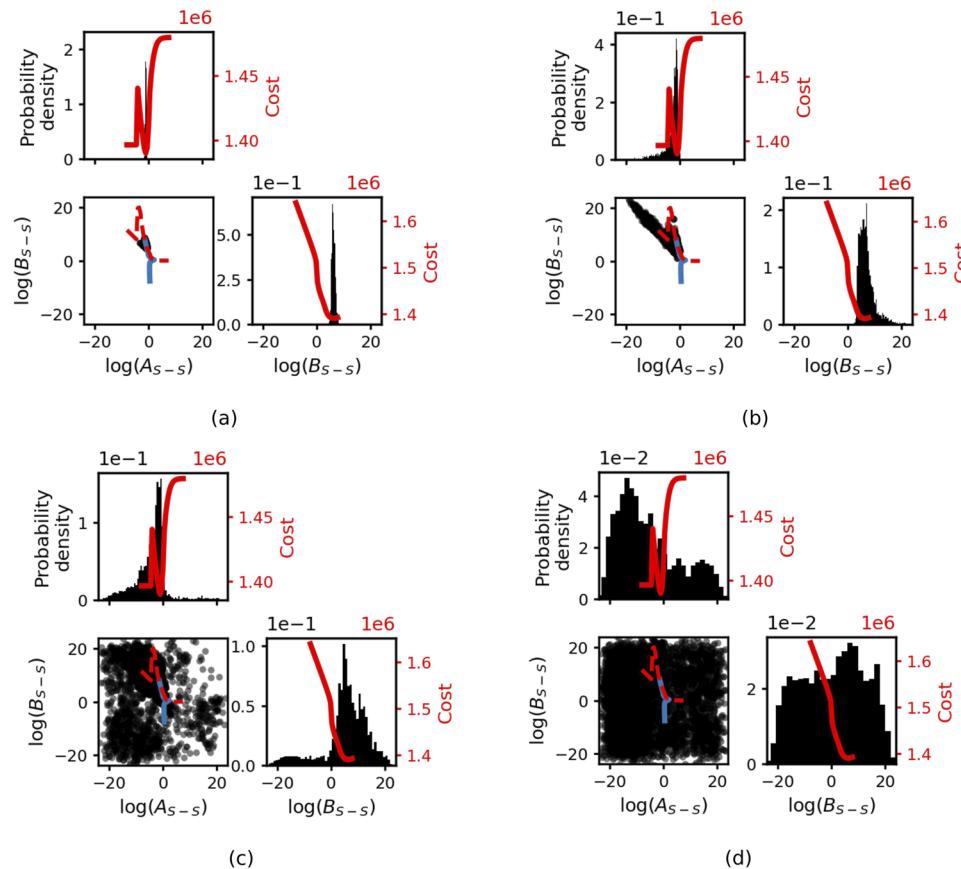
are uniform and correspond to approximately the same range, they distribute prior probability weight differently and lead to different posterior distributions. At first glance, the sampling is in good agreement with what is expected from the cost surface. Samples dominate the regions of low cost and give a visual validation that the samples are converged. Figure 11 compares the marginal distributions for each parameter on both linear and log scales. Note how the parameter scaling and, by extension, the choice of prior can have a strong impact on the posterior distribution. On a log scale, there is a broad, flat plateau for large, negative values of  $\log(\sigma)$  and  $\log(\epsilon)$ . These choices affect how the uncertainties are interpreted and eventually propagated to new predictions. This example illustrates the nuanced relationship between parameterization and prior selection. These effects can be much more severe and harder to tease out in high-dimensional parameter spaces.

In more general terms, recall that the prior defines how one measures volume in parameter space. Regions of the parameter space with large volume may dominate samples, in analogy to statistical mechanics in which high-entropy configurations can dominate an ensemble. It is also related to (though not exactly the same as) the phenomenon known as Lindley's paradox in which Bayesian and frequentist approaches can give different results in a hypothesis test when a broad prior is used.<sup>109</sup> This issue can become especially subtle for sloppy models in high dimensions. These models are

insensitive to coordinated changes in many parameters, indicating that there are large regions of parameter space with nearly identical fits, i.e., fits with high-entropy contributions to the posterior. In these cases, large entropic contributions may dominate their relative frequency in the posterior. The high dimensionality makes it difficult to quantify the role of the energy vs entropy in the final sample and, by extension, justify the choice of prior.

For high-dimensional sloppy models, it is instructive to compare the results of the Bayesian and frequentist techniques, as done in Fig. 12 for the LJ and Morse potentials. For the Morse potential, we also use a uniform prior, bounded by  $|\log(r_0)| < \log(r_{\text{cut}})$ ,  $|\log(C)| < \log(30)$ , and  $|\log(-\epsilon)| < \log(30)$ . These cases illustrate how the two methods agree in low-dimensional examples that are well-understood. Note that the samples are very dense in regions around the paths of the profile likelihood, indicating that the posterior is energy-dominated and that there are not significant artifacts from the prior. Furthermore, the marginal distributions of each parameter are congruous with the profile likelihoods (main diagonal). However, there are hints of large-entropy regions that could become significant at higher sampling temperatures, for example, samples evaporating on the sub-optimal region at large negative values of  $\log(r_0)$ .

We now turn to the SW model in Fig. 13. We set the prior distribution to be uniform in a rectangular region, defined by

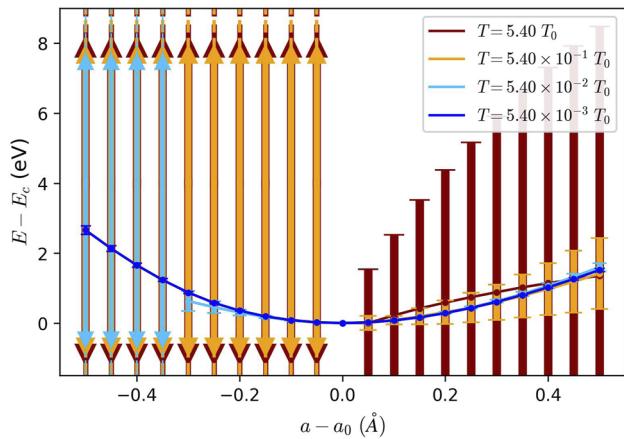


**FIG. 13.** MCMC samples and profile likelihood for parameters  $A$  and  $B$  of the S-S interaction in the SW MoS<sub>2</sub> potential with a uniform prior in log parameter space at sampling temperatures: (a)  $T = 5.40 \times 10^{-3} T_0$ , (b)  $T = 5.40 \times 10^{-2} T_0$ , (c)  $T = 5.40 \times 10^{-1} T_0$ , and (d)  $T = 5.40 T_0$ , where  $T_0$  is a natural temperature. Note that different parameter combinations evaporate at different temperatures, e.g.,  $A_{\text{S-S}}$  and  $B_{\text{S-S}}$  evaporate at lower temperature in a coordinated way (b), while they evaporate in all directions at higher temperature (c) and (d).

$|\log(\theta_i)| < 24$ , where  $\theta_i$  are the parameters in this potential. Figure 13 summarizes a Bayesian sampling at four temperatures for parameters  $A_{S-S}$  and  $B_{S-S}$  ( $T = 5.40 \times 10^{-3} T_0, 5.40 \times 10^{-2} T_0, 5.40 \times 10^{-1} T_0, 5.40 T_0$ , where  $T_0 \approx 1.85 \times 10^5$  is the natural temperature). The sampling results at other sampling temperatures for the other parameters can be found in the [supplementary material](#).

At low temperatures, the profile likelihoods again agree with the Bayesian sampling. Next, we increase the sampling temperature. Recall that the temperature uniformly scales the error bars in Eq. (2). As the temperature rises, the uncertainty estimates in the parameters also increase; however, it does not increase uniformly in each of the parameters. At some critical temperatures, the uncertainty in a particular parameter abruptly transitions to infinity. For example, note that, from the spread of the samples, the uncertainties in the parameters  $A_{S-S}$  and  $B_{S-S}$  are relatively small at  $T = 5.40 \times 10^{-3} T_0$  but become effectively infinite at  $T = 5.40 \times 10^{-1} T_0$ . The posterior has transitioned from a distribution of low-temperature, energy-dominated samples to high-temperature, entropy-dominated samples. The higher sampling temperature has “evaporated” the parameter. We discuss this further in Sec. IV.

For the next step of UQ for this model, we propagate the parametric uncertainty and calculate the uncertainty of the change in energy as a response to lattice stretching and compression.<sup>110</sup> Figure 14 shows the uncertainty of this quantity of interest, calculated at several different sampling temperatures using the ensembles in Fig. 13. Note that at lower sampling temperatures, such as at  $T = 5.40 \times 10^{-3} T_0$  (blue), the uncertainties of the predicted quantities are finite. Moreover, the uncertainty in the tension domain ( $a > a_0$ ) matches the distribution of predicted quantities from various models in Ref. 2. However, at higher temperatures, the uncertainties diverge as the MoS<sub>2</sub> sheet is compressed.

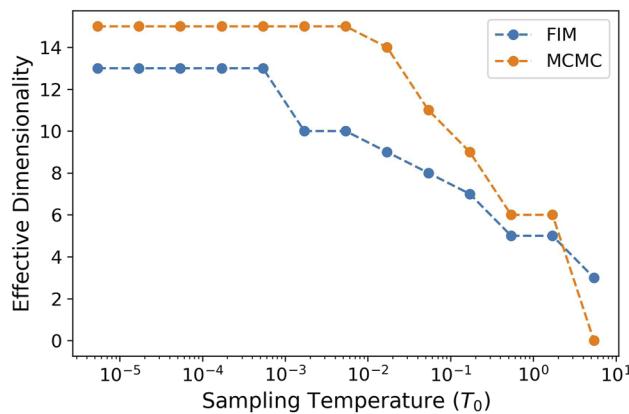


**FIG. 14.** Propagated uncertainty of changes in energy as a function of lattice stretching and compression. The parameters  $a_0$  and  $E_c$  are the equilibrium lattice constant and cohesive energy, respectively. The uncertainty of this quantity is calculated at several different temperatures from the ensembles in Fig. 13. Note that the uncertainty of the energy at higher temperature diverges to infinity. This is a result of parameter evaporation, where the evaporated parameters predict infinite energy.

At higher temperatures, some of the MCMC walkers sample regions with extreme values of parameters, near the edge of the support of the prior. These evaporated samples represent interactions with a very strong repulsive force in the compression domain. The magnitude of the energy grows very fast as the lattice is compressed. Consequently, the uncertainty of the energy in this domain diverges. Furthermore, we are unable to propagate the uncertainty from the  $T = 5.40 T_0$  samples. The ensemble at this temperature contains many samples representing extreme potentials, e.g., a semi-infinite square-well potential.

The phenomenon of parameter evaporation illustrated in Fig. 13 has been observed previously.<sup>50</sup> When a parameter evaporates, its marginal posterior distribution approaches its prior. Evaporated parameters do not participate in the statistics of the model; they do not encode any information in the data and do not constrain future predictions. In other words, the effective dimensionality of the model is reduced by the number of evaporated parameters. However, evaporated parameters affect statistical methods, slowing down convergence in both MCMC sampling and profile likelihood optimization. They also obscure interpretation since the evaporated parameters are often combinations of the bare parameters. Performing UQ with these “nuisance” parameters is challenging.

Parameter evaporation is a global manifestation of the “sloppiness” phenomenon. Sloppiness was first recognized as the exponential distribution of FIM eigenvalues, as in Fig. 8, as a local property. However, it was later shown using information geometry that sloppiness is a global property of the model and evaluation scenario. For sloppy models, the entire model manifold is systematically compressed into an object of low effective dimensionality, and in many practical cases, the eigenvalues of the FIM (local property) are a good estimate for the widths of the model manifold (global property).<sup>8,41–44</sup> We check this correspondence for the case of the SW IP by comparing the eigenvalues in Fig. 8 with the number of effective (non-evaporated) parameters in the model at each sampling temperature. We consider a parameter “evaporated” if the samples approach a boundary of the prior corresponding to this parameter.



**FIG. 15.** Relation between sampling temperature and the effective dimensionality, obtained from the local (FIM) and global (MCMC) estimates. The effective dimensionality from the MCMC samples is the number of non-evaporated parameters at a given temperature. The local estimate is the number of eigenvalues of the Fisher information larger than a given temperature.

Although, in general, there is a subtle difference between the evaporated and the non-evaporated parameters, the temperature ladder we use is sparse enough that there is a clear distinction between the two. We show the comparison between the local estimate and the result deduced from MCMC in Fig. 15, and we find good agreement.

Sloppiness in high dimensions leads to cost contours that do not close and complicates the question of prior selection and the role of parameter-space entropy in the Bayesian posteriors. We now use information geometry to better understand this phenomenon, first using the LJ model as a motivating example.

Figure 16 shows four geodesics paths in parameter space radiating from the best fit. Although difficult to discern from this figure, the geodesics that found boundaries took the same asymptotic form:  $\log(\epsilon) \rightarrow \infty$ ,  $\log(\sigma) \rightarrow -\infty$ , with  $\log(\epsilon)$  diverging six times faster than  $\log(\sigma)$ . The significance of this result is more apparent when expressed in the so-called *AB* form with  $A = 4\epsilon\sigma^{12}$  and  $B = 4\epsilon\sigma^6$ . As the original parameters,  $\epsilon$  and  $\sigma$ , approach the boundary, they are correlated such that  $A \rightarrow 0$  while  $B$  remains finite. This suggests that the *AB* parameterization is a more natural parameterization, which from Eq. (22) gives

$$\begin{aligned} \phi_{\text{LJ}}(r_{ij}) &= 4\epsilon \left( \sigma^{12} \left( \frac{1}{r_{ij}^{12}} - \frac{1}{r_{\text{cut}}^{12}} \right) - \sigma^6 \left( \frac{1}{r_{ij}^6} - \frac{1}{r_{\text{cut}}^6} \right) \right) \\ &= A \left( \frac{1}{r_{ij}^{12}} - \frac{1}{r_{\text{cut}}^{12}} \right) - B \left( \frac{1}{r_{ij}^6} - \frac{1}{r_{\text{cut}}^6} \right) \\ &\rightarrow -B \left( \frac{1}{r_{ij}^6} - \frac{1}{r_{\text{cut}}^6} \right). \end{aligned} \quad (26)$$

For this fitting problem,  $B$  is the identifiable parameter combination. There is a natural limit that removes the unidentifiable parameter,  $A \rightarrow 0$ , that leads to a physically interpretable reduced model, i.e., a purely attractive potential. Although this reduced model loses the physics of the repulsive part of the potential, the data to which the model was fit included atomic configurations

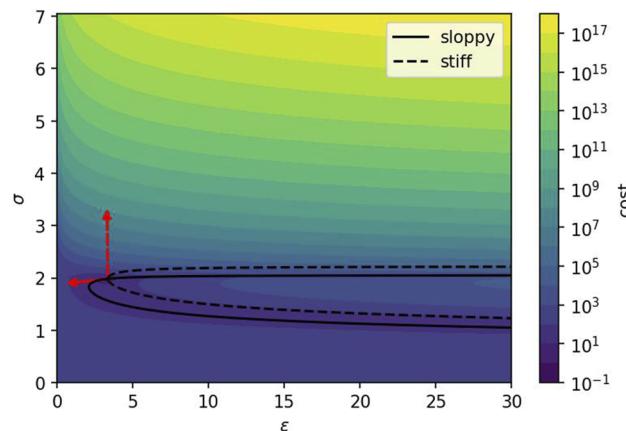


FIG. 16. Geodesics for the LJ potential in the sloppy and stiff directions. Each geodesic starts at the best fit parameters and moves in the forward and backward directions of the sloppy and stiff eigenvectors of the FIM, shown by the corresponding arrows. By considering the difference in the scaling of the parameter axes, the eigenvectors are orthogonal to each other.

that only probed the attractive regime. In the case when the data only probe the repulsive regime,  $A$  becomes the identifiable parameter, and removing the unidentifiable parameter ( $B \rightarrow 0$ ) will lead to a purely repulsive potential, instead. This behavior is illustrated in Fig. 17. Thus, the geometry (1) reflects the information content of the data, (2) explains the correlations among the inferred parameters, (3) isolates unidentifiable combinations of parameters, and (4) suggests reduced models for simplifying the statistics.

An analogous calculation on the Morse potential reveals many of the same themes. We find the geodesic initially aligned with the sloppiest eigenvector of the FIM eventually approaches a boundary in which  $C \rightarrow 0$  and  $\epsilon \rightarrow -\infty$ , as seen in Fig. 18. Note that this geodesic aligns with the MCMC results, i.e., the low-cost canyon. As

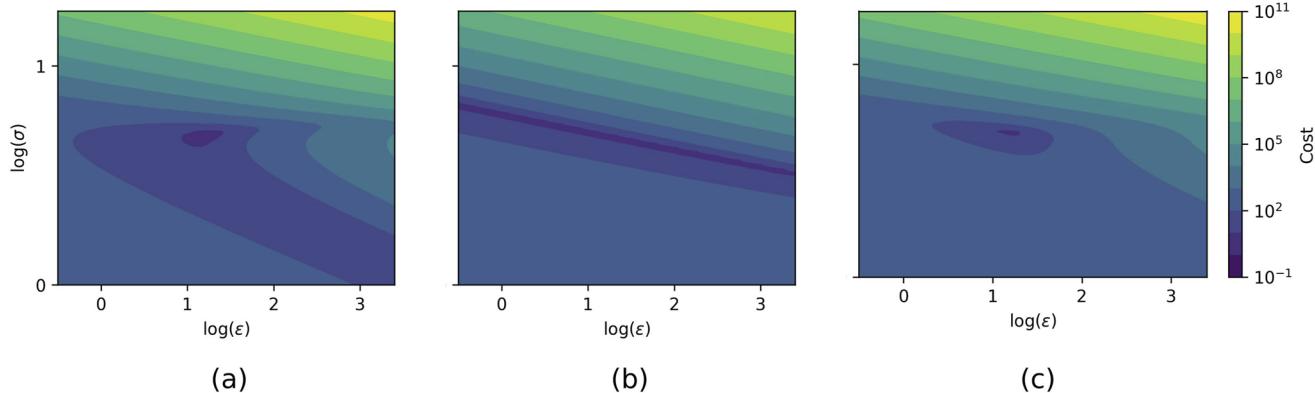
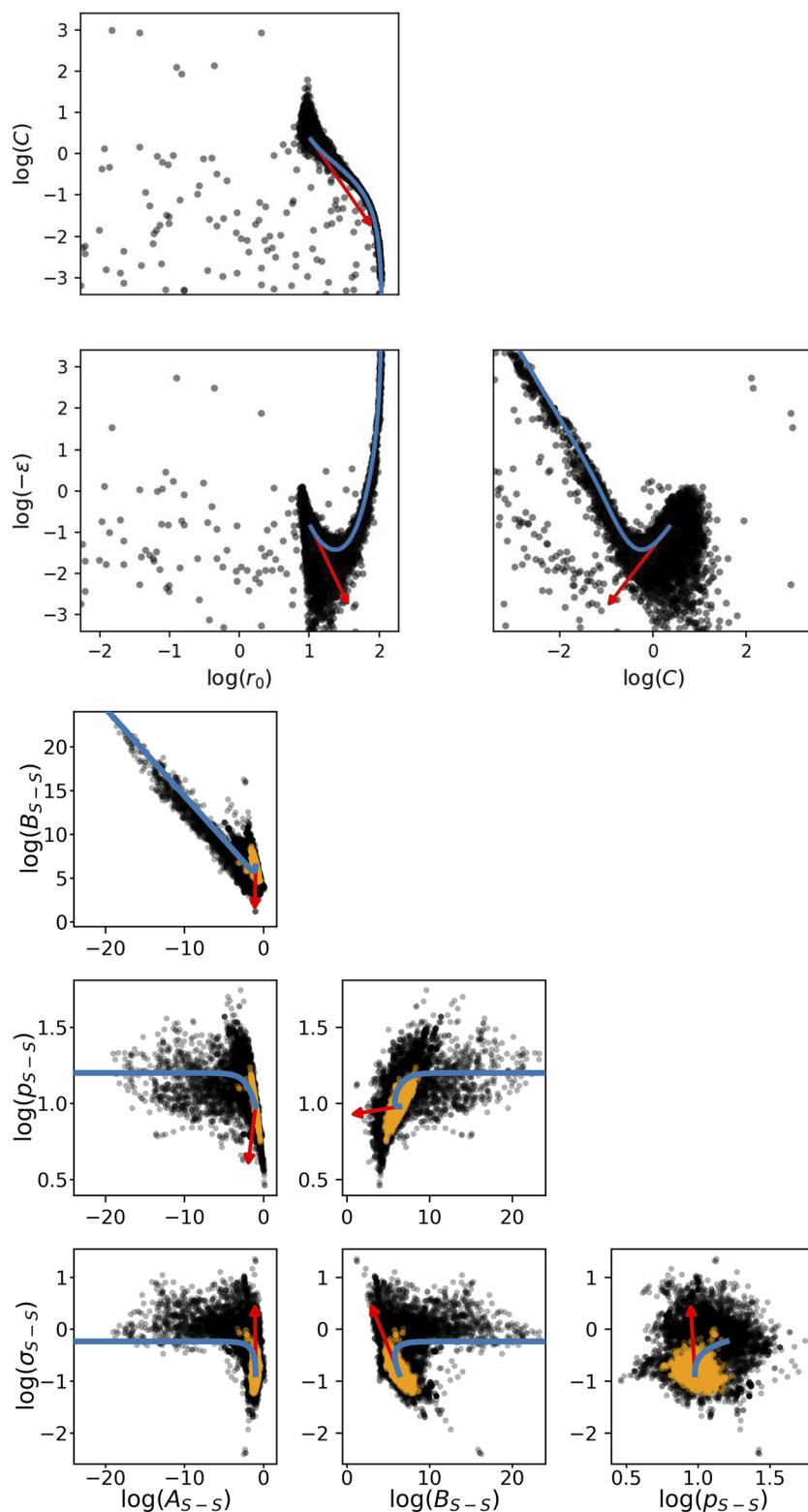
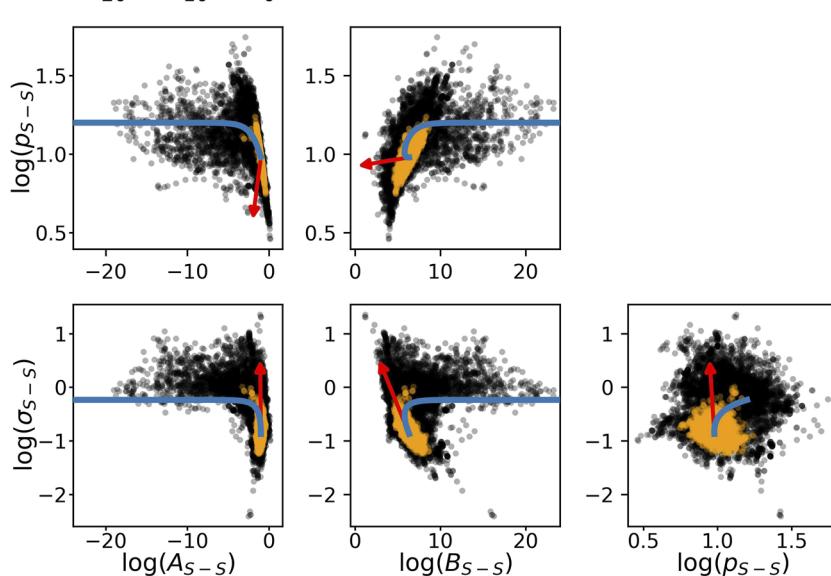


FIG. 17. Cost surface of the LJ potentials with data that probe the (a) attractive, (b) repulsive, and (c) both regimes. When only the data in the attractive (repulsive) regime are used, parameter  $A$  ( $B$ ), which corresponds to repulsive (attractive) interaction in the LJ potential, is unidentifiable. This is shown by the direction of the low cost canyon in (a) and (b). When the data probe both regimes, these canyons are raised and parameters  $A$  and  $B$  become more identifiable. However, note that the plateau extending to  $\{\epsilon, \sigma\} \rightarrow 0$  direction exists in all cases, which shows that there is a fundamental limitation in the LJ potential in predicting energy and forces.



**FIG. 18.** Geodesic in the sloppiest direction for the Morse potential. Geodesics (blue curves) are shown with the MCMC samples (black points) to illustrate that the geodesics follow the same low cost canyons as MCMC. These geodesics reveal specific parameter limits leading to boundaries of the model manifold, e.g.,  $C \rightarrow 0$  as  $\epsilon \rightarrow -\infty$ .



**FIG. 19.** Geodesics in the sloppiest direction for the SW model. Geodesics (blue curves) are plotted with MCMC results at sampling temperatures of  $5.40 \times 10^{-3} T_0$  (orange points) and  $5.40 \times 10^{-2} T_0$  (black points). Axis scaling is set to show detail, not to reflect the boundaries of the uniform prior. The geodesic started in the local sloppy direction defined by the FIM,  $B \rightarrow 0$  as  $\sigma_{S-S} \rightarrow \infty$ . Eventually, the geodesic turned to find the boundary given by the limit  $A_{S-S} \rightarrow 0$  as  $B_{S-S} \rightarrow \infty$ .

with LJ, this geodesic suggests a natural reparameterization of the model,

$$k = -2\epsilon C^2. \quad (27)$$

As  $\epsilon$  and  $C$  approach extreme values, the specific combination  $k$  remains finite. In this parameterization,  $k$  and  $r_0$  are the identifiable parameter combinations, while  $C$  is unidentifiable since the geodesic approaches the boundary given by  $C \rightarrow 0$ . Evaluating the limit  $C \rightarrow 0$  at constant  $k$  and  $r_0$  leads to the simplified model,

$$\tilde{\phi}_M(r_{ij}) = \frac{1}{2}k(r_{ij} - r_0)^2 - \frac{1}{2}k(r_{\text{cut}} - r_0)^2. \quad (28)$$

In this limit, the model is a simple harmonic potential parameterized by an equilibrium position,  $r_0$ , and a stiffness,  $k$ . This indicates that the configurations do not carry enough information about the IP's anharmonicity to constrain those parameter combinations. Once again, the geometry reflects the information content of the data, explains observed correlations, isolates the unidentifiable combinations, and suggests alternative parameterizations and simplified models.

We now consider the SW potential. As before, we calculate a geodesic in the sloppiest direction and find that it encounters the boundary  $A_{S-S} \rightarrow 0$  and  $B_{S-S} \rightarrow \infty$ , as shown in Fig. 19. Again, we reparameterize the model,

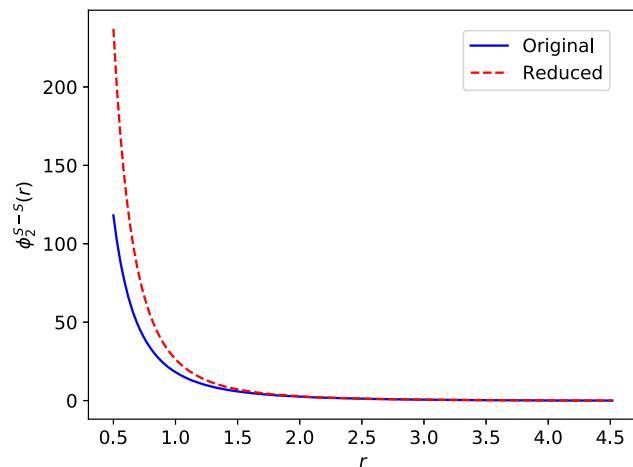
$$\theta = A_{S-S}B_{S-S}. \quad (29)$$

Both  $A_{S-S}$  and  $B_{S-S}$  are unidentifiable parameters, but there is an identifiable combination given by  $\theta = A_{S-S}B_{S-S}$ . Note that as  $A_{S-S} \rightarrow 0$  at constant  $\theta$ ,  $B_{S-S} \rightarrow \infty$ , consistent with the evaporation in Fig. 19. Furthermore, considering  $A_{S-S}, 1/B_{S-S} \rightarrow 0$  at constant  $\theta$  leads to the reduced form

$$\tilde{\phi}_2^{S-S}(r_{ij}) = \theta \left( \frac{\sigma}{r_{ij}} \right)^p \exp \left( \frac{\sigma}{r_{ij} - r_{\text{cut}}} \right). \quad (30)$$

Figure 20 shows the plots of the two-body S-S interaction term for both the original and reduced models. In addition to having fewer identifiable parameters, reduced models include the physics that is informed by the data. In this case, details about the repulsive core were removed at the boundary, resulting in a stronger repulsion at short distances that were not constrained by the fit. This example illustrates that reduced order models change the nature of the physics that is encoded in the potential. While the simple model may be just as accurate as the original model on the training data, it may be less accurate for downstream applications. Additional work is needed to develop these information geometric insights into concrete tools for selecting and developing interatomic potentials for target applications. Alternatively, by identifying the short-range interactions as the missing information in the training set, information geometry can potentially guide the extension of training data to constrain relevant parameter combinations.

It is interesting to compare the geodesic to the purely local analysis of the Fisher information. The initial direction of the geodesic is given by the sloppiest eigendirection of the Fisher information. That eigendirection, as can be seen in Fig. 9, is dominated by the parameters  $B_{S-S}$  and  $\sigma_{S-S}$ . Initially, the geodesic decreases  $B_{S-S}$  and increases  $\sigma_{S-S}$ ; however, as  $\sigma_{S-S}$  increases, it plays a more important



**FIG. 20.** Reduced models on the boundaries have fewer identifiable parameters and abstract away irrelevant aspects of the physics. Geodesics identified the boundary defined by the coordinated limit  $A_{S-S} \rightarrow 0, B_{S-S} \rightarrow \infty$ . This figure compares the forms of the original and reduced two-body SW potentials for the S-S interaction. This reduction abstracts away details about the repulsive core of the potential.

role in the model. It is no longer part of the sloppy combination, so the geodesic rotates to align with  $A_{S-S}$  and  $B_{S-S}$ , the parameters that eventually participate in the boundary (see Fig. 19). This simple comparison illustrates how the geodesic naturally extends the local analysis.

#### IV. DISCUSSION

Sloppy models are often identified by their characteristic FIM spectra with eigenvalues spanning many orders of magnitude.<sup>8,9,38,39</sup> Previous work has noted sloppiness in many contexts,<sup>9,37,42,45–48</sup> including IPs,<sup>3,4,51</sup> and our results corroborate this conclusion (see Fig. 8). Subsequent work using information geometry showed sloppiness to be a consequence of global properties of the model, specifically that the model manifold is bounded with a hierarchy of widths.<sup>3</sup> In this work, we have extended the local sloppy model analysis using (Bayesian) MCMC and (frequentist) profile likelihoods. Each of these methods gives a unique perspective on the “global sloppiness” of the model. For example, MCMC samples evaporate sloppy parameters, and cost profiles have flat plateaus resulting in diverging uncertainties. We connect these traditional statistical tools to sloppy model analysis using information geometry and geodesics. We show that the problems associated with both of these methods are features of the same underlying phenomena, sloppiness in the form of bounded model manifolds. We now discuss each of these observations in more detail.

Figure 10 illustrates that cost contours of sloppy models have broad plateaus in parameter space. These regions can be thought of as high-entropy states, i.e., large volumes of parameter space with approximately equal cost. Although these regions may have sub-optimal cost, they can dominate the Bayesian posterior because of their high entropy. Previous work in sloppy models has noted this “parameter evaporation” phenomenon when the posterior becomes

dominated by sub-optimal, high-entropy samples.<sup>50</sup> Geometrically, the model maps these high-entropy regions to compressed areas near the boundaries of the model manifold. Parameter evaporation is thus the Bayesian manifestation of sloppiness that is a consequence of model boundaries. As such, it is a nonlinear effect and distinct from other well-documented statistical phenomena related to high dimension.

Model boundaries have different distances from the data. This implies that the height of the cost plateau varies in each parameter direction. Consequently, specific parameters evaporate at different sampling temperatures. This is analogous to particles in a classical finite potential well; only those particles (MCMC walkers) with high enough energy can escape the well (cost surface). Figure 13 shows that different parameter combinations evaporate at different temperatures. Indeed, previous studies in IPs have lowered the sampling temperature specifically to avoid parameter evaporation.<sup>3,4</sup> However, for a sufficiently broad prior, parameters evaporate at any non-zero temperature, although the evaporation time may be very long for large cost barriers, making it difficult to assess convergence.

Figure 15 reports the number of evaporated parameters for different sampling temperatures for our SW potential. Note that the number of identifiable parameter combinations, i.e., non-evaporated at a given temperature, correlates with the number of eigenvalues above that temperature. Elsewhere, it has been shown that the eigenvalues of the FIM are a good approximation for the widths of the model manifold.<sup>8,41</sup> Since the sampling temperature corresponds to a distance in data space, Fig. 15 confirms that the eigenvalues are a reasonable (local) approximation for the widths of the model manifold of IPs.

Most of the challenges associated with formulating and performing a Bayesian analysis of a sloppy IP result from the interplay of entropy and energy in the posterior. Recall that the Bayesian prior acts as a measure on the parameter space (see Sec. II D), i.e., it quantifies the entropy associated with volumes of parameter space. The ambiguity in the choice of prior leads to the issues we report here. We advocate comparing MCMC with a frequentist method to assess the effect the prior has on the posterior, as we have done using profile likelihoods.

The profile likelihood analysis, being a frequentist method, does not use a prior. As such, global sloppiness manifests itself differently. The cost profiles exhibit plateaus that asymptotically approach constant values, as seen in Figs. 12 and 13. Uncertainty in a parameter is set by selecting a level of statistical significance, e.g., 95% confidence interval, and identifying those parameter values with cost less than the corresponding cost threshold. As the cost threshold approaches that of the plateau, the uncertainty diverges. This leads to uncertainty metrics that are very sensitive to the level of statistical significance, making it difficult to draw conclusions from the UQ analysis.

Another complication due to sloppiness is related to parameter correlations. Sloppy canyons and plateaus do not naturally align with the parameter axes due to correlations in the parameters. These correlations can bend sloppy canyons, as seen in Fig. 12(a) for LJ. As the profile likelihood projects bending canyons onto parameter axes, correlation is lost and the results are misleading. Using a more natural parameterization, motivated by information geometry, weakens

parameter correlations and unwinds the canyons asymptotically aligning them with the parameter axes.

Geodesics extend the local FIM analysis to a global regime. For example, consider Fig. 19. The geodesic initially pointed in the steepest direction, as indicated by the FIM, but changed directions to follow the global sloppiness as it approached the manifold boundary. This behavior is due to non-linearity in the model and is known as parameter-effects curvature.<sup>11</sup>

The global nature of geodesics is used to find boundaries of the model manifold, revealing the cost plateaus and suggesting natural parameterizations of the model. Again, consider Fig. 19 where the geodesics found the manifold boundary represented by the parameter limits  $A_{S-S} \rightarrow 0$  and  $B_{S-S} \rightarrow \infty$ . This limit demonstrates a more natural parameterization of the model with parameters  $\epsilon = 1/B_{S-S}$  and  $\theta = A_{S-S}B_{S-S}$ , where  $\epsilon$  is strictly non-negative. In this parameterization, only one parameter participates in the boundary, i.e.,  $\epsilon \rightarrow 0$  while  $\theta \sim \mathcal{O}(1)$ . The sloppy direction aligns with a single parameter  $\epsilon$  that is (mostly) uncorrelated from  $\theta$ . Furthermore, rather than a diverging confidence interval for  $B_{S-S}$ , the confidence intervals for  $\epsilon$  and  $\theta$  remain finite.

This new parameterization suggests a simplified model in which  $\epsilon$  has been removed. The Manifold Boundary Approximation Method (MBAM) is a tool that utilizes information geometry to find these boundary limits and construct reduced models. In this specific example, the reduced model would correspond to  $\epsilon \rightarrow 0$  while holding  $\theta \sim \mathcal{O}(1)$ . This limit removes the sloppy parameter from the model, leaving the identifiable combination,  $\theta = A_{S-S}B_{S-S}$ . Performing MBAM before UQ, i.e., finding less-sloppy models, would remove the challenges we have discussed. In this paper, we have performed the first step of MBAM by using geodesics to find manifold boundaries.

Reduced models often do not transfer well to new predictive regimes. However, they make new predictions with higher levels of certainty. When the large parametric uncertainties of sloppy models are propagated to new predictions, the resulting uncertainties can be large or infinite. Reducing the sloppiness of models decreases parametric uncertainty as well as the propagated uncertainty in new predictive regimes. Future work will continue this process and perform UQ on reduced models.

We have shown that sloppy models lead to ill-posed UQ problems. For Bayesians, the challenge is how to unambiguously select a prior that does not lead to large-entropy contributions in the posterior. For frequentists, the challenge is sensitivity to the confidence level and plateaus that do not naturally align with the bare parameters (i.e., occur due to correlations among parameters). By identifying the root cause of these problems, we hope this work will lead to more transparency in the future UQ studies for IPs. In particular, information geometry suggests solutions to these issues by identifying natural parameterizations near boundaries that provide simplified, less-sloppy models.

In conclusion, we provide suggestions both for model developers and UQ practitioners alike. For developers of empirical potentials, we recommend using the FIM to assess how parameters locally affect calculated quantities. To extend this analysis to a global regime, we recommend using geodesics to identify more natural parameterizations as well as additional training data that are needed to identify model parameters. For example, geodesic calculations for

the LJ potential above identified that the training data only contained information about the attractive part of the potential. To fully identify the LJ parameters, additional data that probe the repulsive regime are needed. Alternatively, unidentifiable parameters may be removed. Using these geodesics to reduce the model with an MBAM step will decrease model sloppiness and improve future UQ. This process can then be iterated starting from the reduced model until a simple, yet accurate model is attained. However, caution must be used as these parameters may be relevant for downstream applications. Future work may focus on how to selectively remove parameters for target applications.

For performing UQ of IPs, we recommend starting with the FIM analysis to assess the sloppiness of the model. This analysis also provides a local estimate of which parameters evaporate at a given sampling temperature. If performing UQ with MCMC, we recommend using several different sampling temperatures, including the natural temperature,<sup>7</sup> and some alternative priors. For example, we have sampled the SW at multiple temperatures (see Fig. 13) and for different priors (see the [supplementary material](#)). Then, we advocate comparing the sampling results to geodesics, a frequentist method, to assess the effect of the Bayesian prior on parameter uncertainty. Additionally, researchers can perform other frequentist analysis, e.g., profile likelihood. Finally, for an extended UQ analysis, researchers can apply MBAM to perform model reduction and/or collect additional training data, in which case we recommend iterating the steps in the previous paragraph. We are working on implementing these analysis tools within the KIM-based Learning-Integrated Fitting Framework (KLIFF) Python package.<sup>112</sup>

## SUPPLEMENTARY MATERIAL

In the [supplementary material](#), we first show plots of the profile likelihood and the MCMC samples at several sampling temperatures for all parameters of the SW potential for MoS<sub>2</sub> system. The sampling temperatures are given with respect to the natural temperature  $T_0 \approx 1.85 \times 10^5$ . We then present the MCMC samples for SW potential, where the sampling is done using linear parameterization with a corresponding uniform prior in parameter space.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Award Nos. DMR-1834251 and DMR-1834332. Some of the calculations were done on computational facilities provided by the Brigham Young University Office of Research Computing.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## REFERENCES

- 1 E. B. Tadmor and R. E. Miller, *Modeling Materials: Continuum, Atomistic and Multiscale Techniques* (Cambridge University Press, 2011).
- 2 M. Wen, S. N. Shirodkar, P. Plecháč, E. Kaxiras, R. S. Elliott, and E. B. Tadmor, “A force-matching Stillinger-Weber potential for MoS<sub>2</sub>: Parameterization and Fisher information theory based sensitivity analysis,” *J. Appl. Phys.* **122**, 244301 (2017).
- 3 S. Longbottom and P. Brommer, “Uncertainty quantification for classical effective potentials: An extension to *potfit*,” *Modell. Simul. Mater. Sci. Eng.* **27**, 044001 (2019).
- 4 S. L. Frederiksen, K. W. Jacobsen, K. S. Brown, and J. P. Sethna, “Bayesian ensemble approach to error estimation of interatomic potentials,” *Phys. Rev. Lett.* **93**, 165501 (2004).
- 5 M. Wen and E. B. Tadmor, “Uncertainty quantification in molecular simulations with dropout neural network potentials,” *npj Comput. Mater.* **6**, 124 (2020).
- 6 A. Chernatynskiy, S. R. Phillipot, and R. LeSar, “Uncertainty quantification in multiscale simulation of materials: A prospective,” *Annu. Rev. Mater. Res.* **43**, 157–182 (2013).
- 7 K. S. Brown and J. P. Sethna, “Statistical mechanical approaches to models with many poorly known parameters,” *Phys. Rev. E* **68**, 021904 (2003).
- 8 M. K. Transtrum, B. B. Machta, and J. P. Sethna, “Why are nonlinear fits so challenging?,” *Phys. Rev. Lett.* **104**, 060201 (2010); [arXiv:0909.3884](#).
- 9 M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, “Perspective: Sloppiness and emergent theories in physics, biology, and beyond,” *J. Chem. Phys.* **143**, 010901 (2015).
- 10 S. Wan, R. C. Sinclair, and P. V. Coveney, “Uncertainty quantification in classical molecular dynamics,” *Philos. Trans. R. Soc. London, Ser. A* **379**, 20200082 (2021).
- 11 D. W. Brenner, “The art and science of an analytic potential,” *Phys. Status Solidi B* **217**, 23–40 (2000).
- 12 E. B. Tadmor, R. S. Elliott, J. P. Sethna, R. E. Miller, and C. A. Becker, “The potential of atomistic simulations and the knowledgebase of interatomic models,” *JOM* **63**, 17 (2011).
- 13 F. Ercolelli and J. B. Adams, “Interatomic potentials from first-principles calculations: The force-matching method,” *Europhys. Lett.* **26**, 583 (1994).
- 14 J. M. Hughes, M. F. Horstemeyer, R. Carino, N. Sukhija, W. B. Lawrimore, S. Kim, and M. I. Baskes, “Hierarchical bridging between ab initio and atomistic level computations: Sensitivity and uncertainty analysis for the modified embedded-atom method (MEAM) potential (part b),” *JOM* **67**, 148–153 (2015).
- 15 M. C. Kennedy and A. O’Hagan, “Bayesian calibration of computer models,” *J. R. Stat. Soc. B* **63**, 425–464 (2001).
- 16 P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos, “Bayesian uncertainty quantification and propagation in molecular dynamics simulations: A high performance computing framework,” *J. Chem. Phys.* **137**, 144103 (2012).
- 17 F. Rizzi, R. E. Jones, B. J. Debusschere, and O. M. Knio, “Uncertainty quantification in MD simulations of concentration driven ionic flow through a silica nanopore. II. Uncertain potential parameters,” *J. Chem. Phys.* **138**, 194105 (2013).
- 18 R. A. Messerly, M. R. Shirts, and A. F. Kazakov, “Uncertainty quantification confirms unreliable extrapolation toward high pressures for united-atom Mie  $\lambda$ -6 force field,” *J. Chem. Phys.* **149**, 114109 (2018).
- 19 M. Vohra, A. Y. Nobakht, S. Shin, and S. Mahadevan, “Uncertainty quantification in non-equilibrium molecular dynamics simulations of thermal transport,” *Int. J. Heat Mass Transfer* **127**, 297–307 (2018).
- 20 M. Vohra and S. Mahadevan, “Discovering the active subspace for efficient UQ of molecular dynamics simulations of phonon transport in silicon,” *Int. J. Heat Mass Transfer* **132**, 577–586 (2019).
- 21 G. Dhaliwal, P. B. Nair, and C. V. Singh, “Uncertainty analysis and estimation of robust AIREBO parameters for graphene,” *Carbon* **142**, 300–310 (2019).
- 22 G. Dhaliwal, P. B. Nair, and C. V. Singh, “Uncertainty and sensitivity analysis of mechanical and thermal properties computed through embedded atom method potential,” *Comput. Mater. Sci.* **166**, 30–41 (2019).
- 23 L. De Simon, M. Iglesias, B. Jones, and C. Wood, “Quantifying uncertainty in thermophysical properties of walls by means of Bayesian inversion,” *Energy Build.* **177**, 220–245 (2018).
- 24 F. Rizzi, H. N. Najm, B. J. Debusschere, K. Sargsyan, M. Salloum, H. Adalsteinsson, and O. M. Knio, “Uncertainty quantification in MD simulations.

Part II: Bayesian inference of force-field parameters," *Multiscale Model. Simul.* **10**, 1460–1492 (2012).

<sup>25</sup> M. Cools-Ceuppens and T. Verstraelen, "Uncertainty prediction in molecular simulations using ab initio derived force fields," Ph.D. thesis, Ghent University, 2017.

<sup>26</sup> R. A. Messerly, T. A. Knotts, and W. V. Wilding, "Uncertainty quantification and propagation of errors of the Lennard-Jones 12-6 parameters for *n*-alkanes," *J. Chem. Phys.* **146**, 194110 (2017).

<sup>27</sup> M. A. Tschopp, B. Chris Rinderspacher, S. Nouranian, M. I. Baskes, S. R. Gwaltney, and M. F. Horstemeyer, "Quantifying parameter sensitivity and uncertainty for interatomic potential design: Application to saturated hydrocarbons," *ASME J. Risk Uncertainty Eng. Syst., Part B* **4**, 011004 (2018).

<sup>28</sup> A. Mishra, S. Hong, P. Rajak, C. Sheng, K.-i. Nomura, R. K. Kalia, A. Nakano, and P. Vashishta, "Multiobjective genetic training and uncertainty quantification of reactive force fields," *npj Comput. Mater.* **4**, 42 (2018).

<sup>29</sup> F. Lu, H. Wang, D. Yan, D. Zhang, and W. Xiao, "Application of profile likelihood function to the uncertainty analysis of hydrometeorological extreme inference," *Sci. China: Technol. Sci.* **56**, 3151–3160 (2013).

<sup>30</sup> S. Chen, Y. Li, J. Shin, and T. Kim, "Constructing confidence intervals of extreme rainfall quantiles using Bayesian, bootstrap, and profile likelihood approaches," *Sci. China: Technol. Sci.* **59**, 573–585 (2016).

<sup>31</sup> Y. Wu, L. Xue, Y. Liu, and L. Ren, "Uncertainty assessment of extreme flood estimation in the Dongting Lake basin, China," *Hydrol. Res.* **50**, 1162–1176 (2019).

<sup>32</sup> It is important to point out that models may manifest sloppiness in some evaluation scenarios, but not others. Thus, "sloppiness" is a property of both the model and the specific set of predictions being considered.<sup>42</sup>

<sup>33</sup> P. G. Constantine, E. Dow, and Q. Wang, "Active subspace methods in theory and practice: Applications to kriging surfaces," *SIAM J. Sci. Comput.* **36**, A1500–A1524 (2014); [arXiv:1304.2070](https://arxiv.org/abs/1304.2070).

<sup>34</sup> T. Cui, J. Martin, Y. M. Marzouk, A. Solonen, and A. Spantini, "Inverse Probl. **30**, 114015 (2014).

<sup>35</sup> T. Cui and X. T. Tong, "A unified performance analysis of likelihood-informed subspace methods," [arXiv:2101.02417](https://arxiv.org/abs/2101.02417) [math, stat] (2021).

<sup>36</sup> A. F. Brouwer and M. C. Eisenberg, "The underlying connections between identifiability, active subspaces, and parameter space dimension reduction," [arXiv:1802.05641](https://arxiv.org/abs/1802.05641) (2018).

<sup>37</sup> A. White, M. Tolman, H. D. Thamess, H. R. Withers, K. A. Mason, and M. K. Transtrum, "The limitations of model-based experimental design and parameter estimation in sloppy systems," *PLoS Comput. Biol.* **12**, e1005227 (2016).

<sup>38</sup> R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, "Universally sloppy parameter sensitivities in systems biology models," *PLoS Comput. Biol.* **3**, e189 (2007).

<sup>39</sup> J. J. Waterfall, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and J. P. Sethna, "Sloppy-model universality class and the vandermonde matrix," *Phys. Rev. Lett.* **97**, 150601 (2006).

<sup>40</sup> M. K. Transtrum and P. Qiu, "Model reduction by manifold boundaries," *Phys. Rev. Lett.* **113**, 098701 (2014).

<sup>41</sup> M. K. Transtrum, B. B. Machta, and J. P. Sethna, "Geometry of nonlinear least squares with applications to sloppy models and optimization," *Phys. Rev. E* **83**, 036701 (2011).

<sup>42</sup> B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, "Parameter space compression underlies emergent theories and predictive models," *Science* **342**, 604–607 (2013).

<sup>43</sup> K. N. Quinn, C. B. Clement, F. De Bernardis, M. D. Niemack, and J. P. Sethna, "Visualizing probabilistic models and data with Intensive Principal Component Analysis," *Proc. Natl. Acad. Sci. U. S. A.* **116**, 13762–13767 (2019).

<sup>44</sup> K. N. Quinn, H. Wilber, A. Townsend, and J. P. Sethna, "Chebyshev approximation and the global geometry of model predictions," *Phys. Rev. Lett.* **122**, 158302 (2019).

<sup>45</sup> J. E. Jeong, Q. Zhuang, M. K. Transtrum, E. Zhou, and P. Qiu, "Experimental design and model reduction in systems biology," *Quant. Biol.* **6**, 287–306 (2018).

<sup>46</sup> M. K. Transtrum and P. Qiu, "Bridging mechanistic and phenomenological models of complex biological systems," *PLoS Comput. Biol.* **12**, e1004915 (2016).

<sup>47</sup> B. K. Mannakee, A. P. Ragsdale, M. K. Transtrum, and R. N. Gutenkunst, "Sloppiness and the geometry of parameter space," in *Uncertainty in Biology, Studies in Mechanobiology, Tissue Engineering and Biomaterials* Vol. 17 (Springer International Publishing, 2016), pp. 271–299.

<sup>48</sup> M. K. Transtrum and P. Qiu, "Optimal experiment selection for parameter estimation in biological differential equation models," *BMC Bioinf.* **13**, 181 (2012).

<sup>49</sup> M. K. Transtrum, A. T. Sarić, and A. M. Stanković, "Information geometry approach to verification of dynamic models in power systems," *IEEE Trans. Power Syst.* **33**, 440–450 (2017).

<sup>50</sup> R. Gutenkunst, "Sloppiness, modeling, and evolution in biochemical networks," Ph.D. thesis, Cornell University, 2007.

<sup>51</sup> M. Wen, J. Li, P. Brommer, R. S. Elliott, J. P. Sethna, and E. B. Tadmor, "A KIM-compliant *potfit* for fitting sloppy interatomic potentials: Application to the EDIP model for silicon," *Modell. Simul. Mater. Sci. Eng.* **25**, 014001 (2016).

<sup>52</sup> Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood *et al.*, "Performance and cost assessment of machine learning interatomic potentials," *J. Phys. Chem. A* **124**, 731–745 (2020).

<sup>53</sup> V. L. Deringer, M. A. Caro, and G. Csányi, "Machine learning interatomic potentials as emerging tools for materials science," *Adv. Mater.* **31**, 1902765 (2019).

<sup>54</sup> M. Wen and E. B. Tadmor, "Hybrid neural network potential for multilayer graphene," *Phys. Rev. B* **100**, 195419 (2019).

<sup>55</sup> J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli, "Towards an assessment of the accuracy of density functional theory for first principles simulations of water," *J. Chem. Phys.* **120**, 300–311 (2004).

<sup>56</sup> N. Schunck, J. D. McDonnell, J. Sarich, S. M. Wild, and D. Higdon, "Error analysis in nuclear density functional theory," *J. Phys. G: Nucl. Part. Phys.* **42**, 034024 (2015).

<sup>57</sup> M. R. Fellinger, H. Park, and J. W. Wilkins, "Force-matched embedded-atom method potential for niobium," *Phys. Rev. B* **81**, 144119 (2010).

<sup>58</sup> T. J. Lenosky, J. D. Kress, I. Kwon, A. F. Voter, B. Edwards, D. F. Richards, S. Yang, and J. B. Adams, "Highly optimized tight-binding model of silicon," *Phys. Rev. B* **55**, 1528–1544 (1997).

<sup>59</sup> J. W. Miller and D. B. Dunson, "Robust Bayesian inference via coarsening," *J. Am. Stat. Assoc.* **114**, 1113–1125 (2019).

<sup>60</sup> A. van den Bos, "4. Precision and accuracy," in *Parameter Estimation for Scientists and Engineers* (John Wiley & Sons, Ltd., 2007), pp. 45–97.

<sup>61</sup> W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, 2007).

<sup>62</sup> W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman & Hall/CRC Interdisciplinary Statistics; Taylor & Francis, 1995).

<sup>63</sup> C. H. LaMont and P. A. Wiggins, "Correspondence between thermodynamics and inference," *Phys. Rev. E* **99**, 052140 (2019).

<sup>64</sup> E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.* **106**, 620 (1957).

<sup>65</sup> D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "emcee: The MCMC hammer," *Publ. Astron. Soc. Pac.* **125**, 306–312 (2013); [arXiv:1202.3665](https://arxiv.org/abs/1202.3665).

<sup>66</sup> J. Goodman and J. Weare, "Ensemble samplers with affine invariance," *Commun. Appl. Math. Comput. Sci.* **5**, 65–80 (2010).

<sup>67</sup> W. D. Vosden, W. M. Farr, and I. Mandel, "Dynamic temperature selection for parallel-tempering in Markov chain Monte Carlo simulations," *Mon. Not. R. Astron. Soc.* **455**, 1919–1937 (2016); [arXiv:1501.05823](https://arxiv.org/abs/1501.05823).

<sup>68</sup> A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Stat. Sci.* **7**, 457–472 (1992).

<sup>69</sup> S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," *J. Comput. Graphical Stat.* **7**, 434–455 (1998).

<sup>70</sup> D. Vats and C. Knudson, "Revisiting the Gelman-Rubin diagnostic," *Stat. Sci.* **35**(4), 518–529 (2021).

<sup>71</sup> M. J. Bayarri and J. O. Berger, "The interplay of Bayesian and frequentist analysis," *Stat. Sci.* **19**, 58–80 (2004).

<sup>72</sup> E.-J. Wagenaers, M. Lee, T. Lodewyckx, and G. J. Iverson, "Bayesian versus frequentist inference," in *Bayesian Evaluation of Informative Hypotheses*, Statistics for Social and Behavioral Sciences (Springer, 2008), pp. 181–207.

<sup>73</sup>K. B. Athreya and S. N. Lahiri, *Measure Theory and Probability Theory* (Springer Science & Business Media, 2006).

<sup>74</sup>S. R. Cole, H. Chu, and S. Greenland, “Maximum likelihood, profile likelihood, and penalized likelihood: A primer,” *Am. J. Epidemiol.* **179**, 252–260 (2014).

<sup>75</sup>W. A. Rolke, A. M. López, and J. Conrad, “Limits and confidence intervals in the presence of nuisance parameters,” *Nucl. Instrum. Methods Phys. Res., Sect. A* **551**, 493–503 (2005).

<sup>76</sup>Y. Kurniawan, *profile\_likelihood*, [https://gitlab.com/yonatank93/profile\\_likelihood](https://gitlab.com/yonatank93/profile_likelihood), 2021.

<sup>77</sup>M. K. Transtrum and J. P. Sethna, “Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization,” *arXiv:1201.5885* [physics] (2012).

<sup>78</sup>M. K. Transtrum, MBAM GitHub repository, <https://github.com/mktranstrum/MBAM>, 2018.

<sup>79</sup>R. S. Elliott and E. B. Tadmor, Knowledgebase of Interatomic Models (KIM) application programming interface (API), <https://openkim.org/kim-api>, 2011.

<sup>80</sup>We note that the LJ and Morse IPs do not provide good representations for silicon or nickel. They are simply chosen for their simplicity to help understand the UQ methodology.

<sup>81</sup>R. S. Elliott, Efficient multi-species Lennard-Jones model with truncated or shifted cutoff v003, OpenKIM, <https://doi.org/10.25950/962b4967>, 2018.

<sup>82</sup>J. E. Jones, “On the determination of molecular fields. I. From the variation of the viscosity of a gas with temperature,” *Proc. R. Soc. London, Ser. A* **106**, 441–462 (1924).

<sup>83</sup>J. E. Jones, “On the determination of molecular fields. II. From the equation of state of a gas,” *Proc. R. Soc. London, Ser. A* **106**, 463–477 (1924).

<sup>84</sup>J. E. Lennard-Jones, “On the forces between atoms and ions,” *Proc. R. Soc. London, Ser. A* **109**, 584–497 (1925).

<sup>85</sup>R. S. Elliott and Y. Afshar, Morse pair potential shifted to zero energy at cutoff separation v004, OpenKIM, <https://doi.org/10.25950/fc55a3bb>, 2020.

<sup>86</sup>L. A. Girifalco and V. G. Weizer, “Application of the Morse potential function to cubic metals,” *Phys. Rev.* **114**, 687–690 (1959).

<sup>87</sup>D. S. Karls, Potential energy and atomic forces of periodic, non-orthogonal cell of Si atoms in a perturbed bcc structure v003, OpenKIM, [https://openkim.org/cite/TE\\_006970922000\\_003](https://openkim.org/cite/TE_006970922000_003), 2019.

<sup>88</sup>D. S. Karls, Potential energy and atomic forces of periodic, non-orthogonal cell of atoms v003, OpenKIM, <https://doi.org/10.25950/c3dca28e>, 2019.

<sup>89</sup>D. S. Karls, Potential energy and atomic forces of periodic, non-orthogonal cell of Si atoms in a perturbed random structure v003, OpenKIM, [https://openkim.org/cite/TE\\_047848074253\\_003](https://openkim.org/cite/TE_047848074253_003), 2019.

<sup>90</sup>R. S. Elliott, Efficient “universal” shifted Lennard-Jones model for all KIM API supported species developed by Elliott and Akerson (2015) v003, OpenKIM, <https://doi.org/10.25950/962b4967>, 2018.

<sup>91</sup>R. S. Elliott, Morse potential (shifted) for Ni by Girifalco and Weizer (1959) using a high-accuracy cutoff distance v004, OpenKIM, <https://doi.org/10.25950/fc55a3bb>, 2020.

<sup>92</sup>M. Wen, Modified Stillinger-Weber potential (MX2) for monolayer MoS<sub>2</sub> developed by Wen *et al.* (2017) v001, OpenKIM, <https://doi.org/10.25950/eedbdc4>, 2018.

<sup>93</sup>M. Wen, Stillinger-Weber Model Driver for Monolayer MX2 systems v001, OpenKIM, <https://doi.org/10.25950/eedbdc4>, 2018.

<sup>94</sup>Y. Kurniawan, C. Petrie, K. Williams, M. K. Transtrum, R. S. Elliott, E. B. Tadmor, D. S. Karls, and M. Wen, Modified Stillinger-Weber potential (MX2) for monolayer MoS<sub>2</sub> by Kurniawan *et al.* (2022) v000, OpenKIM, <https://doi.org/10.25950/328bfabb>, 2022.

<sup>95</sup>A. Chawla, Three-body bond-order potential by Khor and Das Sarma (1988) v000, OpenKIM, <https://doi.org/10.25950/8e5d84c2>, 2019.

<sup>96</sup>K. E. Khor and S. Das Sarma, “Proposed universal interatomic potential for elemental tetrahedrally bonded semiconductors,” *Phys. Rev. B* **38**, 3318–3322 (1988).

<sup>97</sup>D. S. Karls, Environment-Dependent Interatomic Potential (EDIP) model driver v002, OpenKIM, <https://doi.org/10.25950/545ca247>, 2018.

<sup>98</sup>M. Z. Bazant and E. Kaxiras, “Modeling of covalent bonding in solids by inversion of cohesive energy curves,” *Phys. Rev. Lett.* **77**, 4370–4373 (1996).

<sup>99</sup>J. F. Justo, M. Z. Bazant, E. Kaxiras, V. V. Bulatov, and S. Yip, “Interatomic potential for silicon defects and disordered phases,” *Phys. Rev. B* **58**, 2539–2550 (1998).

<sup>100</sup>M. Wen, Stillinger-Weber (SW) Model Driver v004, OpenKIM, <https://doi.org/10.25950/e9adf627>, 2018.

<sup>101</sup>F. H. Stillinger and T. A. Weber, “Computer simulation of local order in condensed phases of silicon,” *Phys. Rev. B* **31**, 5262–5271 (1985).

<sup>102</sup>H. Balamane, T. Halicioglu, and W. A. Tiller, “Comparative study of silicon empirical interatomic potentials,” *Phys. Rev. B* **46**, 2250–2279 (1992).

<sup>103</sup>H. Balamane, T. Halicioglu, and W. A. Tiller, “Vacancy- and adatom-induced  $\sqrt{3} \times \sqrt{3}$  reconstructions of the Si(111) surface,” *Phys. Rev. B* **40**, 9999–10001 (1989).

<sup>104</sup>A. Chawla, Three-body cluster potential for Si by Khor and Das Sarma (1988) v000, OpenKIM, <https://doi.org/10.25950/8e5d84c2>, 2019.

<sup>105</sup>D. S. Karls, EDIP model for Si developed by Justo *et al.* (1998) v002, OpenKIM, <https://doi.org/10.25950/545ca247>, 2018.

<sup>106</sup>A. K. Singh, Stillinger-Weber potential for Si developed by Balamane, Halicioglu and Tiller (1992) v004, OpenKIM, <https://doi.org/10.25950/e9adf627>, 2018.

<sup>107</sup>I. J. Perez-arriaga, G. C. Verghese, and F. C. Schweppe, “Selective modal analysis with applications to electric power systems, Part I: Heuristic introduction,” *IEEE Trans. Power Appar. Syst. PAS-101*, 3117–3125 (1982).

<sup>108</sup>F. Garofalo, L. Iannelli, and F. Vasca, “Participation factors and their connections to residues and relative gain array,” *IFAC Proc. Vol.* **35**, 125–130 (2002).

<sup>109</sup>C. Robert, “On the Jeffreys-Lindley’s paradox,” *Philos. Sci.* **81**(2), 216–232 (2014).

<sup>110</sup>The MoS<sub>2</sub> layer is constrained to remain flat so that out-of-plane wrinkling is not possible.

<sup>111</sup>D. M. Bates and D. G. Watts, “Relative curvature measures of nonlinearity,” *J. R. Stat. Soc. B* **42**, 1–16 (1980).

<sup>112</sup>M. Wen, Y. Afshar, R. S. Elliott, and E. B. Tadmor, “KLIFF: A framework to develop physics-based and machine learning interatomic potentials,” *Comput. Phys. Commun.* **272**, 108218 (2022).