Interpolation and extrapolation in a 51-dimensional variable space:

system-agnostic construction of high-dimensional PES

Hiroki Sugisawa\textsuperscript{a,b}, Tomonori Ida\textsuperscript{b}, and R. V. Krems\textsuperscript{a}

\textsuperscript{a}Department of Chemistry, University of British Columbia, Vancouver, B.C. V6T 1Z1, Canada
\textsuperscript{b}Division of Material Chemistry, Graduate School of Natural Science and Technology, Kanazawa University, Kakuma, Kanazawa 920-1192, Japan

(Dated: February 3, 2020)

Abstract

Machine learning is becoming an increasingly powerful tool for physics research. High-dimensional physics problems are generally modelled by neural networks (NNs). However, NNs require large data and are problem-specific. An alternative approach is probabilistic modelling based on Gaussian processes (GPs), which are system-agnostic and can be fully automated. However, GPs suffer from the numerical complexity and have been limited to low-dimensional physics problems. Here, we illustrate that it is possible to build an accurate GP model of a 51-dimensional physical system based on 5000 inputs, using the fragmentation of the input space and entropy-based enhancement of the model complexity. We demonstrate the potency of these algorithms by constructing the global potential energy surface (PES) for a molecular system with 19 atoms. We illustrate that GP models thus constructed have generalization power, allowing us to extrapolate the PES from low energies ($<10,000$ cm$^{-1}$), yielding a PES at high energies ($>20,000$ cm$^{-1}$). This opens the prospect for building machine-learning models that can be used to search for new physics, such as phase transitions, in high-dimensional physics problems with unknown property landscapes by extrapolation.
Machine learning (ML) is becoming an increasingly powerful tool for applications in physics and chemistry research. At the core of these applications are models that interpolate in multi-dimensional physical spaces. These models can be used as surrogates of the solutions of physical equations [1], for optimal control applications [2], design, automation and optimization of experiments [3–6] and numerical computations [7–16]. There are two general approaches to building interpolation models. One is based on parametric models such as neural networks (NN). NNs are flexible, but building a proper NN is a problem-specific task, which requires careful design of the NN architecture, control of overfitting and a suitable choice of the functions and parameters in the NN. NNs require a large number of physical data for building accurate models. NNs generally cannot extrapolate. The second approach is probabilistic modelling, which, in most applications, is based on Gaussian processes (GP) [23]. Compared to NNs, GPs have several advantages: GPs require less information than NNs to make accurate predictions, training a GP does not require manual work, GPs have been shown to be capable of extrapolation [17–19]. Therefore, GPs are a popular choice of ML models for applications requiring automation, unbiased predictions and predictions by extrapolation. The major limitation of GP applications is the numerical difficulty of training GP models, which increases with the number of training points \( n \) as \( \mathcal{O}(n^3) \). This has restricted most applications of GP models in physics to low-dimensional problems.

Here, we demonstrate the application of GPs to interpolation and extrapolation in a 51-dimensional (51D) physical space. The error of ML models generally decreases with \( n \) as \( \propto 1/\sqrt{n} \) and several previous studies have attempted to extend the application of GPs to high-dimensional model problems by developing algorithms for training models with a large number of data (large \( n \)) [20, 21]. Some of these studies have demonstrated the construction of GP interpolation models with \( n \approx 1,000,000 \). However: (i) it has not been shown if such models provide enough accuracy to be suitable for physical applications; (ii) most physical data come from the numerical solutions of differential equation such as the Schrödinger equation or experiments so it is difficult to obtain this much informations for training GP models; (iii) the numerical evaluation of GP predictions scales as \( \mathcal{O}(n) \) so GP models trained by data with large \( n \) are not suitable for physical applications requiring a large number of evaluations; (iv) extrapolation with GP models in high-dimensional spaces has not been previously considered. Here, we follow Refs. [17–19, 22], to improve the interpolation and extrapolation accuracy of GP models in high-dimensional spaces by increasing
the complexity of models, instead of increasing $n$. We show that this allows us to build GP models capable of interpolation and extrapolation in a 51D space based on $n \approx 5,000$ inputs. This opens up the prospect for applications of GPs, and hence for taking advantage of non-parametric modelling (including automation and non-parametric extrapolation), for high-dimensional physics problems.

The present algorithms can be used to model any high-dimensional physics problem that depends on a large number ($\sim 50$) of parameters. We demonstrate the potency of these algorithms by constructing the global potential energy surface (PES) for a molecular system with 19 atoms, which, to the best of our knowledge, represents the highest-dimensional PES reported to date. In this context, the variable parameters are the coordinates in the $19 \times 3 - 6$ configuration space. Beyond the demonstration of the applicability of GPs to high-dimensional physical problems, the present work demonstrates a method for constructing global PES for large molecular systems. Global PESs are required for computations of stable and metastable geometries of polyatomic systems, zero point energies, proton transfer mechanisms, densities of states and thermodynamic properties. A major thrust of recent research has been to develop ML models for PES [29–52], using approaches based on neural networks (NNs) [29–39] and kernel methods [1, 26, 53–55], including GP regression [40–52]. However, the construction of accurate global PESs for systems with more than 10 atoms remains a challenging task. NNs can be difficult to construct and generalize for systems with unknown geometrical features. GP models are non-parametric and can be applied to any system without discrimination but suffer from the numerical complexity of training and evaluating the models. This has limited previous work on GP models of PES to systems with 4 to 6 atoms, using $n \approx 1,000 – 10,000$ ab initio points. The goal of the present work is to obtain GP models of similar accuracy, using similar $n$, but for systems with many more degrees of freedom.

We begin by a brief description of the conventional algorithm for GP regression. A GP $y(x)$ can be considered as a limit of a Bayesian neural network with an infinite number of hidden neurons [1]. In this work, the inputs $x = [x_1, ..., x_N]^\top$ are the variables describing the internal coordinates of a polyatomic system. The output $y$ is the value of the potential energy. GPs produce a normal distribution $P(y)$ of values $y$ at any $x$. The goal is to condition $P(y)$ by $n$ known values of the potential energy $y = [y_1, ..., y_n]^\top$ at $n$ points $[x_1, ..., x_n]^\top$ of the $N$-dimensional variable space. The mean of this conditional distribution
at an arbitrary point $x_*$ is given by \[ \mu_* = k_*^\top K^{-1} y, \] (1)

where $k_*$ is a vector with $n$ entries $k(x_*, x_i)$ and $K$ is a square $n \times n$ matrix with entries $k(x_i, x_j)$. The quantities $k(x, x')$ are the kernels, which represent the covariance of the normal distributions of $y$ at $x$ and at $x'$ \[ k(x, x') = \mathcal{M}_v = \frac{2^{1-v}}{\Gamma(v)} \left( \sqrt{2v} r(x, x') \right)^v \mathcal{K}_v \left( \sqrt{2v} r(x, x') \right) \] (2)

where $r^2(x, x') = (x - x')^\top M (x - x')$, $M$ is a diagonal matrix with $N$ parameters, $\mathcal{K}_v$ is the modified Bessel function, $\Gamma$ is the Gamma function, and $v$ is a half-integer. The parameters of the kernel function are found by maximizing the logarithm of the marginal likelihood \[ \log \mathcal{L} = -\frac{1}{2} y^\top K^{-1} y - \frac{1}{2} \log |K| - \frac{n}{2} \log 2\pi. \] (3)

We now make three observations: (i) while Eq. (1) can interpolate any smooth function with any kernel function if $n \to \infty$, for finite $n$, the interpolation accuracy depends on the functional form of the kernel function $k(x, x')$; (ii) Eq. (3) is related to cross entropy of the model and data distributions so maximizing Eq. (3) enhances the information content in the model (1); (iii) Eq. (1) uses the training points $y = [y_1, ..., y_n]^\top$ directly, so the prediction accuracy is sensitive to the distribution of these points in the $N$-dimensional space. We exploit these observations to enhance the accuracy of the GP model without increasing $n$.

The system considered here is the protonated imidazole dimer, shown in Figure 1 (a). The potential energy of the molecule was calculated using the Gaussian program package \[56\] at the MP2/6-31++G(d,p) level of theory. To compute the global PES, we started with the known geometry of imidazole dimer in Ref. \[57\], reoptimized it with the MP2/6-31++G(d,p) calculations and used the resulting lowest-energy structure as our starting guess. The global deviation of the molecule from this geometry was described using the Cartesian XYZ coordinates for each carbon and nitrogen atom, with the sampling range $[-0.05, +0.05]$ Å for each XYZ-direction. The configurations for the terminal hydrogen
atoms were sampled so that the distance between each hydrogen and its adjacent atom is within $[-0.1, +0.2] \, \text{Å}$ and the angle is $[-20^\circ, +20^\circ]$, as illustrated in Figure 1 (a). Within these coordinate ranges, the potential energy was computed at 15,000 points, randomly generated using the Latin hypercube sampling method to avoid clustering [43]. The resulting \textit{ab initio} points cover the energy range between zero and 35,000 cm$^{-1}$. To quantify the accuracy of resulting PES, we compute the root-mean-square error (RMSE) using a large number of \textit{ab initio} points that are not used for training GP models.

To build the 51D surface, we change the above algorithm for constructing the GP model as follows. First, we follow Refs. [17,19] to increase the complexity of the kernel function by defining a set of basis kernel functions and combining these basis functions into linear combinations that produce the larger value of $\mathcal{L}$ in Eq. (3). The basis functions include

FIG. 1: Schematic illustration of the protonated imidazole dimer. Panel (a) shows the coordinate displacements of each atom used to obtain the global surface. Panels (b) and (c) illustrate the fragmentation of the full GP.
the functions in Eq. (2) with \( v = \frac{3}{2}, \frac{5}{2} \) and \( \infty \) as well as the rational quadratic kernel \( \mathcal{M}_{RQ} = \left(1 + \frac{|x-x'|^2}{2\alpha^2}\right)^{-\alpha} \). Note that we use a different metric for model selection from that in Refs. [17–19, 22]. Second, we split the full configuration space into smaller parts and represent the energy of the entire molecular system as

\[
E_{\text{total}}(\mathbf{R}) = \mathcal{E}_1(\mathbf{R}_1) + \mathcal{E}_2(\mathbf{R}_2) + \mathcal{E}_{12}(\mathbf{R}_{12}),
\]

where \( \mathbf{R} \) is a 51D-vector, \( \mathcal{E}_1 \) and \( \mathcal{E}_2 \) are independent GP models depending on vectors of lower dimensionality, and \( \mathcal{E}_{12} \) is a GP model that brings the fragments 1 and 2 together into the full surface and that depends on the vector \( \mathbf{R}_{12} \) with the dimensionality to be determined. The model (4) is hereafter referred to as ‘Composite GP’. While the fragmentation (4) is general, here, we use \( \mathbf{R}_1 \) and \( \mathbf{R}_2 \) to represent the separate 21D imidazole fragments, shown in Figure 1 (b). To determine the dimensionality of \( \mathbf{R}_{12} \), we construct a series of surfaces, sampling a different number of active degrees of freedom in \( \mathbf{R}_{12} \), corresponding to the fragments shown in Figure 1 (c). Our results show that \( \mathbf{R}_{12} \) must account for all 51 dimensions in order for Eq. (4) to be accurate.

The main interpolation results of this work are summarized in Table I, illustrating

- that it is possible to construct a 51D surface based on \( n = 5,000 \) with the global error under 0.2 kcal/mol; and

- how the fragmentation (4) and increasing the complexity of the kernels improve the accuracy of the resulting surface.

Here, ‘Simple GP’ refers to the 51D model of the surface trained directly by \textit{ab initio} points in the \( \mathbf{R} \)-space. ‘Complex \( k' \) refers to the complex kernels. To identify such kernels, we use the greedy-search algorithm – as in Refs. [19, 22] – that combines different simple kernel functions in order to maximize the log-likelihood function. This algorithm determined the following complex kernels for the composite models: \( k = a\mathcal{M}_{v=5/2} + b\mathcal{M}_{v=3/2} + c\mathcal{M}_{v=\infty} \), with \( a, b \) and \( c \) being free parameters, for \( \mathcal{E}_1 \) and \( \mathcal{E}_2 \); and \( k = (a\mathcal{M}_{v=5/2} \times \mathcal{M}_{RQ} + b\mathcal{M}_{v=\infty}) \times \mathcal{M}_{v=1/2} \) for \( \mathcal{E}_{12} \). For the simple GP model with complex \( k \), this algorithm determined the kernel \( k = a\mathcal{M}_{v=5/2} + b\mathcal{M}_{v=3/2} + c\mathcal{M}_{v=\infty} + d\mathcal{M}_{v=\infty} \) to give the optimal results. The results labeled ‘Simple \( k' \) in Table I refer to GP models with \( k = \mathcal{M}_{v=5/2} \).

To illustrate the extrapolation power of high-dimensional GP models, we construct a series of surfaces using the \textit{ab initio} points at low energies and predict the global surface
TABLE I: The RMSE for the full 51D surface computed using 10,000 points in the energy range [0, 35000] cm$^{-1}$ as a function of the number of training points $n$.

| Number of training points | RMSE [kcal/mol] |
|---------------------------|-----------------|
|                           | Simple GP       | Simple GP       | Composite GP | Composite GP |
|                           | Simple $k$      | Complex $k$     | Simple $k$   | Complex $k$  |
| 1000                      | 3.285           | 2.480           | 0.9875       | 0.7837       |
| 2000                      | 2.353           | 1.545           | 0.7416       | 0.5373       |
| 3000                      | 1.883           | 0.8569          | 0.6203       | 0.4123       |
| 4000                      | 1.537           | 0.8666          | 0.5315       | 0.2642       |
| 5000                      | 1.286           | 0.7709          | 0.4776       | 0.1815       |

at high energies. Table II summarizes the results. The errors reported in Table II are computed using 7,092 ab initio points in the energy range [20,000 – 35,000] cm$^{-1}$. Models A, B and C are trained by $n$ potential energy points in the energy ranges [0 – 35,000]; [0 – 20,000] cm$^{-1}$, and [0 – 10,000] cm$^{-1}$, respectively. It is impressive to see that model C with $n = 5,000$ points, all at energy below 10,000 cm$^{-1}$, produces a 51D-surface in the energy range [20,000 – 35,000] cm$^{-1}$ with the global error $\approx 0.6$ kcal/mol. This represents the relative average accuracy of better than 1% in this energy range.

Models B and C use no information about the PES at energies above 20,000 cm$^{-1}$. The largest deviation of these model predictions from the ab initio results in the energy range [20,000 – 35,000] cm$^{-1}$ is 514 cm$^{-1}$ for model A and 2740 cm$^{-1}$ for model C (both for the composite, complex kernel case with $n = 5000$). This represents the relative error for that single worst point of < 2.6% (model A) and < 13.7% (model C). Figure 2 illustrates the accuracy of the interpolation and energy extrapolation of the surface represented by model B with complex kernels.

To illustrate the global performance of the 51D GP PES in the computation of observables, we calculate the vibrational frequencies for the 51 normal modes of the molecule. To do that, we diagonalize the Hessian matrix constructed directly from the ab initio results (hereafter referred to as ‘Exact’) and from the GP models. Table III compares the GP model results with the exact results. Table III shows that all normal modes with the frequencies
TABLE II: RMSEs for 7,092 testing points of the energy range [20,000 − 35,000] cm$^{-1}$ computed for three kinds of models \([4]\) trained by \(n\) ab initio points in the energy range [0 − 35,000] cm$^{-1}$ (models A); [0 − 20,000] cm$^{-1}$ (models B), and [0 − 10,000] cm$^{-1}$ (models C).

| \(n\) | Models A | Models B | Models C |
|-------|----------|----------|----------|
|       | [cm$^{-1}$] | kcal/mol | [cm$^{-1}$] | kcal/mol | [cm$^{-1}$] | kcal/mol |
| 1000  | 304.4    | 0.8702   | 332.7     | 0.9513   | 622.8     | 1.781    |
| 2000  | 208.7    | 0.5967   | 220.3     | 0.6299   | 430.0     | 1.230    |
| 3000  | 161.1    | 0.4605   | 168.2     | 0.4810   | 371.3     | 1.062    |
| 4000  | 103.3    | 0.2954   | 115.1     | 0.3291   | 287.9     | 0.8233   |
| 5000  | 71.01    | 0.2030   | 86.53     | 0.2474   | 222.1     | 0.6350   |

> 100 cm$^{-1}$ are well described by the composite GP PES with \(n = 5000\). Moreover, the GP PES constructed with \(n = 1000\) ab initio points captures qualitatively 48 out of 51 normal modes. This illustrates that a qualitatively correct 51D PES can be constructed with 1000 ab initio points.

To show that the GP PESs are smooth and physical, we compute the potential energy profile describing proton transfer between the imidazole molecules. Figure 3 shows that the potential energy predicted by the composite GP model \([4]\) trained with \(n = 5,000\) ab initio points is in perfect agreement with the ab initio results for this minimum energy proton transfer path. Note that the curves shown in Figure 3 represent the minimum of a 51D surface for fixed imidazole - H$^+$ separations.

In conclusion, this work has three important results. First, it shows that it is possible to build a GP model of a 51-dimensional physical system based on only 5000 inputs. Note that the GP model uses a random distribution of training points in the 51D configuration space and no information about the underlying evolution in the physical space is used. This opens up the possibility of taking advantage of GPs for high-dimensional systems that have until now been fitted with NNs. GPs – being non-parametric models – have several advantages over NNs. In particular, GPs are system-agnostic and their training can be easily automated, which makes feasible the construction of a large number of models for a large number of different systems. GPs offer not only the model prediction but also the Bayesian
FIG. 2: The 51D GP model (circles) in comparison with ab initio results (squares). The size of the square represents the energy interval 200 cm\(^{-1}\). The results are shown for the GP model C trained with 5000 ab initio points (not shown) at energies below 20,000 cm\(^{-1}\) (shaded region). \(D\) is the Euclidean distance from the equilibrium geometry of the 51D molecule.

error of the model. GPs can be used for Bayesian optimization, a very efficient, gradient-free optimization technique that is at the core of many applications of ML for physics and chemistry experiments [3, 5, 6, 58]. Second, we have constructed the global PES for the largest molecular system to date. Such PES can be used for a variety of applications, including the study of the role of different degrees of freedom in proton exchange processes or accurate calculations of molecular densities of states. We note that the accuracy of the global PES obtained here (< 0.2 kcal/mol) is significantly better than the accuracy of < 1 kcal/mol typically targeted in parametric fits for large molecular systems with unknown landscapes. We have illustrated an algorithm that produced this PES with a much smaller
FIG. 3: The minimum energy path for the proton transfer in the protonated imidazole dimer: solid curve - \textit{ab initio} calculations; broken curves and green symbols – the results from the 51D GP models as indicated in the legend box. The green symbols representing the 51D composite GP model with the complex kernel is in excellent agreement with the \textit{ab initio} calculations.

Number of \textit{ab initio} points than typically needed for other approaches. This not only makes the training and evaluation of GPs fast, but also shows that high-dimensional PES for large polyatomic molecules can be computed with expensive high-level \textit{ab initio} methods. Third, we have illustrated that 51D GP models have generalization power that allowed us to extrapolate the PES from low energies ($< 10,000$ cm$^{-1}$) to high energies ($> 20,000$ cm$^{-1}$). This opens up the possibility to extend the Bayesian methods for searching new physics, such as the approach in Ref. \[19\] to identify phase transitions, to high-dimensional physics problems with unknown property landscapes.
TABLE III: Vibrational frequencies (cm$^{-1}$) for the normal modes of the protonated imidazole dimer computed from the global PES given by Eq. (4) with different $n$.

| $n$ | 1000 | 4000 | 5000 | Exact |
|-----|------|------|------|-------|
| 1   | 3602.3 | 3635.5 | 3686.5 | 3698.6 |
| 2   | 3587.0 | 3624.6 | 3680.9 | 3688.2 |
| 3   | 3553.7 | 3435.0 | 3379.4 | 3379.6 |
| 4   | 3452.8 | 3430.2 | 3367.3 | 3370.3 |
| 5   | 3412.9 | 3380.1 | 3359.9 | 3367.3 |
| 6   | 3410.3 | 3367.2 | 3350.4 | 3361.6 |
| 7   | 3368.5 | 3329.4 | 3337.5 | 3352.0 |
| 8   | 3363.9 | 3323.9 | 3324.8 | 3345.4 |
| 9   | 2031.0 | 2002.5 | 2031.0 | 2018.9 |
| 10  | 1700.1 | 1687.4 | 1689.4 | 1676.8 |
| 11  | 1641.3 | 1584.1 | 1587.0 | 1584.4 |
| 12  | 1597.2 | 1579.3 | 1574.8 | 1583.6 |
| 13  | 1580.6 | 1564.1 | 1563.9 | 1564.0 |
| 14  | 1570.5 | 1551.7 | 1553.9 | 1552.2 |
| 15  | 1519.2 | 1504.6 | 1500.8 | 1498.8 |
| 16  | 1518.7 | 1502.7 | 1495.5 | 1497.3 |
| 17  | 1441.9 | 1426.7 | 1423.7 | 1422.8 |
| 18  | 1396.6 | 1403.0 | 1401.9 | 1400.3 |
| 19  | 1347.2 | 1355.6 | 1360.2 | 1360.6 |
| 20  | 1287.5 | 1277.2 | 1284.0 | 1281.0 |
| 21  | 1276.5 | 1272.7 | 1281.4 | 1272.3 |
| 22  | 1244.3 | 1217.2 | 1212.3 | 1207.6 |
| 23  | 1237.5 | 1203.6 | 1198.2 | 1198.1 |
| 24  | 1190.6 | 1199.3 | 1192.7 | 1185.8 |
| 25  | 1176.9 | 1190.3 | 1185.6 | 1182.7 |
| 26  | 1150.2 | 1131.9 | 1144.0 | 1134.6 |
| 27  | 1139.9 | 1126.4 | 1140.7 | 1132.9 |
| 28  | 1106.2 | 1110.5 | 1112.9 | 1110.3 |
| 29  | 1087.8 | 1088.8 | 1096.3 | 1093.7 |
| 30  | 982.3  | 958.5  | 955.7  | 957.9  |
| 31  | 964.6  | 936.4  | 936.2  | 936.1  |
| 32  | 955.8  | 932.5  | 929.7  | 932.0  |
| 33  | 920.1  | 843.8  | 851.5  | 851.2  |
| 34  | 862.0  | 795.9  | 794.6  | 789.8  |
| 35  | 806.4  | 790.3  | 779.3  | 780.5  |
| 36  | 790.6  | 773.3  | 768.7  | 777.3  |
| 37  | 750.1  | 761.9  | 756.6  | 760.4  |
| 38  | 746.5  | 723.0  | 710.0  | 721.3  |
| 39  | 730.3  | 714.7  | 708.2  | 709.7  |
| 40  | 700.1  | 687.6  | 682.0  | 685.9  |
| 41  | 683.8  | 673.6  | 671.5  | 673.2  |
| 42  | 655.3  | 606.8  | 620.2  | 616.2  |
| 43  | 570.7  | 601.5  | 603.1  | 613.8  |
| 44  | 490.2  | 508.2  | 579.0  | 581.8  |
| 45  | 458.9  | 477.6  | 560.3  | 561.0  |
| 46  | 213.0  | 178.6  | 168.8  | 159.7  |
| 47  | 209.2  | 164.2  | 141.4  | 143.8  |
| 48  | 145.8  | 112.7  | 106.2  | 123.5  |
| 49  | 87.5   | 76.4   | 91.2   | 48.7   |
| 50  | 17.8   | 62.3   | 82.5   | 44.1   |
| 51  | -8.4   | 26.7   | 78.2   | 30.3   |
Acknowledgments

We thank Rodrigo Vargas-Hernandez for useful discussions and technical assistance. HS would like to thank Prof. Mark J. MacLachlan from the University of British Columbia for assistance with the research visit to Canada, during which this work was carried out. This work was supported by TOBITATE! Young Ambassador Program (No. S191N133010001; Japan), JSPS KAKENHI (Grant No. 19K05371; Tokyo, Japan) and NSERC of Canada.
[1] R. V. Krems, Bayesian Machine Learning for Quantum Molecular Dynamics, *Phys. Chem. Chem. Phys.* **21**, 13392 (2019).

[2] M. Benning, E. Celledoni, M. J. Ehrhardt, B. Owren, C.-B. Schönlieb, arXiv: 1904.05657 (2019).

[3] R. Gómez-Bombarelli *et al.*, Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach, *Nat. Mat.* **15**, 1120 (2016).

[4] J.N. Wei, D Duvenaud, and A. Aspuru-Guzik, Neural networks for the prediction of organic chemistry reactions, *ACS Cent. Scie.* **2**, 725 (2016).

[5] L. M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L. P.-E. Yunker, J. E Hein, and A. Aspuru-Guzik, ChemOS: Orchestrating autonomous experimentation, *Science Robotics* **3**, 19 (2018).

[6] F. Häse, L. M. Roch, C. Kreisbeck, and A. Aspuru-Guzik, Phoenics: A Bayesian optimizer for chemistry *ACS Cent. Sci.* **4**, 1134 (2018).

[7] P. L. A. Popelier, QCTFF: On the construction of a novel protein force field, *Int. J. Quant. Chem.* **115**, 1005 (2015).

[8] V. Botu and R. Ramprasad, Adaptive machine learning framework to accelerate ab initio molecular dynamics, *Int. J. Quant. Chem.* **115**, 1074 (2015).

[9] M. Caccin, Z. Li, J. R. Kermode, and A. De Vita, A framework for machine-learning-augmented multiscale atomistic simulations on parallel supercomputers, *Int. J. Quant. Chem.* **115**, 1129 (2015).

[10] J. Wu, Y. Zhou, and X. Xu, The X1 family of methods that combines B3LYP with neural network corrections for an accurate yet efficient prediction of thermochemistry , *Int. J. Quant. Chem.* **115**, 1021 (2015).

[11] K. Vu, J. C. Snyder, L. Li, M. Rupp, B. F. Chen, T. Khelif, K.-R. Müller, and K. Burke, Understanding kernel ridge regression: Common behaviors from simple functions to density functionals , *Int. J. Quant. Chem.* **115**, 1115 (2015).

[12] J. J. Mortensen, K. Kaasbjerg, S. L. Frederiksen, J. K. Nørskov, J. P. Sethna, and K. W. Jacobsen, Bayesian Error Estimation in Density-Functional Theory, *Phys. Rev. Lett.* **95**, 216401 (2005).
[13] A. J. Medford, J. Wellendorff, A. Vojvodic, F. Studt, F. Abild-Pedersen, K. W. Jacobsen, T. Bligaard, and J. K. Nørskov, Catalysis. Assessing the reliability of calculated catalytic ammonia synthesis rates, Science 345, 197 (2014).

[14] M. Fritz, M. Fernández-Serra, and J. M. Soler, Optimization of an exchange-correlation density functional for water, J. Chem. Phys. 144, 224101 (2016).

[15] R. A. Vargas-Hernández, Bayesian optimization for tuning and selecting hybrid-density functionals, arXiv:1903.10678

[16] J. Proppe and M. Reiher, Reliable Estimation of Prediction Uncertainty for Physicochemical Property Models, J. Chem. Theor. Comp. 13, 3297 (2017).

[17] D. K. Duvenaud, H. Nickisch, and C. E. Rasmussen, Additive Gaussian Processes, Adv. Neur. Inf. Proc. Sys. 24, 226 (2011).

[18] D. K. Duvenaud, J. Lloyd, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani, Structure Discovery in Nonparametric Regression through Compositional Kernel Search, Proceedings of the 30th International Conference on Machine Learning Research 28, 1166 (2013).

[19] R. Vargas-Hernandez, J. Sous, M. Berciu, and R. V. Krems, Extrapolating quantum observables with machine learning: Inferring multiple phase transitions from properties of a single phase, Phys. Rev. Lett. 121, 255702 (2018).

[20] Hensman, J.; Fusi, N.; Lawrence, N. D. Gaussian processes for big data, Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI2013), 2013, Report number: UAI-P-2013-PG-282-290.

[21] Liu, H.; Ong, Y.-S.; Shen, X.; Cai, J.; When Gaussian Process Meets Big Data: A Review of Scalable GPs, arXiv:1807.01065.

[22] J. Dai and R. V. Krems, Interpolation and extrapolation of global potential energy surfaces for polyatomic systems by Gaussian processes with composite kernels, arXiv:1907.08717 (2019).

[23] C. E. Rasmussen, and C. K. I. Williams, Gaussian Processes for Machine Learning (The MIT Press, Cambridge, 2006).

[24] B. J. Braams and J. M. Bowman, Permutationally invariant potential energy surfaces in high dimensionality, Int. Rev. Phys. Chem. 28, 577 (2009).

[25] J. N. Murrell, S. Carter, S. C. Farantos, P. Huxley, and A. J. C. Varandas, Molecular Potential Energy Functions (Wiley, Chichester, England, 1984).

[26] T. Hollebeek, T.-S. Ho, and H. Rabitz, Constructing multidimensional molecular potential
energy surfaces from \textit{ab initio} data, \textit{Annu. Rev. Phys. Chem.} \textbf{50}, 537 (1999).

[27] M. A. Collins, Molecular potential-energy surfaces for chemical reaction dynamics, \textit{Theor. Chem. Acc.} \textbf{108}, 313 (2002).

[28] C. M. Handley and P. L. A. Popelier, Potential Energy Surfaces Fitted by Artificial Neural Networks, \textit{J. Phys. Chem.} A \textbf{114}, 3371 (2010).

[29] J. Behler, Perspective: Machine learning potentials for atomistic simulations, \textit{J. Chem. Phys.} \textbf{145}, 170901 (2016)

[30] S. Manzhos and T. Carrington, Jr., A random-sampling high dimensional model representation neural network for building potential energy surfaces \textit{J. Chem. Phys.} \textbf{125}, 084109 (2006).

[31] S. Manzhos, X. Wang, R. Dawes, and T. Carrington, Jr., A nested molecule-independent neural network approach for high-quality potential fits, \textit{J. Phys. Chem.} A \textbf{110}, 5295 (2006).

[32] J Behler and M Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, \textit{Phys. Rev. Lett.} \textbf{98}, 146401 (2007)

[33] J. Behler, Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations, \textit{Phys. Chem. Chem. Phys.} \textbf{13}, 17930 (2011).

[34] J. Behler, Constructing high-dimensional neural network potentials: A tutorial review, \textit{Int. J. Quant. Chem.} \textbf{115}, 1032 (2015).

[35] E. Pradesh and A. Brown, A ground state potential energy surface for HONO based on a neural network with exponential fitting functions, \textit{Phys. Chem. Chem. Phys.} \textbf{19}, 22272 (2017).

[36] A. Leclerc and T. Carrington, Jr., Calculating vibrational spectra with sum of product basis functions without storing full-dimensional vectors or matrices, \textit{J. Chem. Phys.} \textbf{140}, 174111 (2014).

[37] S. Manzhos, R. Dawes, and T. Carrington, Neural network-based approaches for building high dimensional and quantum dynamics-friendly potential energy surfaces, \textit{Int. J. Quant. Chem.} \textbf{115}, 1012 (2015).

[38] J. Chen, X. Xu, X. Xu, and D. H. Zhang, A global potential energy surface for the H$_2$ + OH $\leftrightarrow$ H$_2$O + H reaction using neural networks, \textit{J. Chem. Phys.} \textbf{138}, 154301 (2013).

[39] Q. Liu, X. Zhou, L. Zhou, Y. Zhang, X. Luo, H. Guo, and B. Jiang, Constructing High-Dimensional Neural Network Potential Energy Surfaces for Gas-Surface Scattering and Reactions, \textit{J. Phys. Chem.} C \textbf{122}, 1761 (2018).

[40] C. M. Handley, G. I. Hawe, D. B. Kellab and P. L. A. Popelier, Optimal construction of a
fast and accurate polarisable water potential based on multipole moments trained by machine learning, *Phys. Chem. Chem. Phys.* **11**, 6365 (2009).

[41] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons, *Phys. Rev. Lett.* **104**, 136403 (2010).

[42] A. P. Bartók and G. Csányi, Gaussian approximation potentials: A brief tutorial introduction, *Int. J. Quant. Chem.* **115**, 1051 (2015).

[43] J. Cui and R. V. Krems, Efficient non-parametric fitting of potential energy surfaces for polyatomic molecules with Gaussian processes, *J. Phys. B: At. Mol. Opt. Phys.* **49**, 224001 (2016).

[44] P. O. Dral, A. Owens, S. N. Yurchenko, and W. Thiel, Structure-based sampling and self-correcting machine learning for accurate calculations of potential energy surfaces and vibrational levels, *J. Chem. Phys.* **146**, 244108 (2017).

[45] B. Kolb, P. Marshall, B. Zhao, B. Jiang, and Hua Guo, Representing Global Reactive Potential Energy Surfaces Using Gaussian Processes, *J. Phys. Chem. A* **121**, 2552 (2017).

[46] A. Kamath, R. A. Vargas-Hernandez, R. V. Krems, T. Carrington Jr., and S. Manzhos, Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy, *J. Chem. Phys.* **148**, 241702 (2018).

[47] G. Schmitz and O. Christiansen, Gaussian process regression to accelerate geometry optimizations relying on numerical differentiation *J. Chem. Phys.* **148**, 241704 (2018).

[48] Y. Guan, S. Yang, and D. H. Zhang, Construction of reactive potential energy surfaces with Gaussian process regression: active data selection, *Mol. Phys.* **116**, 823 (2018).

[49] G. Laude, D. Calderini, D. P. Tew, and J. O. Richardson, *ab initio* instanton rate theory made efficient using Gaussian process regression, *Faraday Discuss.* **212**, 237 (2018).

[50] Y. Guan, S. Yang, and D. H. Zhang, Application of Clustering Algorithms to Partitioning Configuration Space in Fitting Reactive Potential Energy Surfaces, *J. Phys. Chem. A* **122**, 3140 (2018).

[51] A. E. Wiens, A. V. Copan, H. F. Schaefer, Multi-Fidelity Gaussian Process Modeling for Chemical Energy Surfaces, *Chem. Phys. Lett.* X, in press (2019).

[52] C. Qu, Q. Yu, B. L. Van Hoozen Jr, J. M. Bowman, and R. A. Vargas-Hernandez, Assessing Gaussian Process Regression and Permutationally Invariant Polynomial Approaches To Rep-
resent High-Dimensional Potential Energy Surfaces, *J. Chem. Theor. Comp.* 14, 3381 (2018).

[53] T. S. Ho and H. Rabitz, A general method for constructing multidimensional molecular potential energy surfaces from *ab initio* calculations, *J. Chem. Phys.* 104, 2584 (1996).

[54] T. Hollebeek, T. S. Ho, and H. Rabitz, A fast algorithm for evaluating multidimensional potential energy surfaces, *J. Chem. Phys.* 106, 7223 (1997).

[55] T. S. Ho and H. Rabitz, Reproducing kernel Hilbert space interpolation methods as a paradigm of high dimensional model representations: Application to multidimensional potential energy surface construction, *J. Chem. Phys.* 119, 6433 (2003).

[56] Gaussian 16, Revision A.03, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

[57] G. F. Mangiatordi, J. Hermet, C. Adamo, *J. Phys. Chem. A* 115, 2627 (2011).

[58] R. Vargas-Hernandez, Y. Guan, D. H. Zhang, and R. V. Krems, Bayesian optimization for the inverse scattering problem in quantum reaction dynamics, *New J. Phys.* (Fast Track Communication) 21, 022001 (2019).