Quantum Gaussian process model of potential energy surface for a polyatomic molecule

J. Dai and R. V. Krems

Department of Chemistry, University of British Columbia,
Vancouver, B.C. V6T 1Z1, Canada
Stewart Blusson Quantum Matter Institute,
Vancouver, B.C. V6T 1Z4, Canada

(Dated: February 23, 2022)

Abstract

With gates of a quantum computer designed to encode multi-dimensional vectors, projections of quantum computer states onto specific qubit states can produce kernels of reproducing kernel Hilbert spaces. We show that quantum kernels obtained with a fixed ansatz implementable on current quantum computers can be used for accurate regression models of global potential energy surfaces (PES) for polyatomic molecules. To obtain accurate regression models, we apply Bayesian optimization to maximize marginal likelihood by varying the parameters of the quantum gates. This yields Gaussian process models with quantum kernels. We illustrate the effect of qubit entanglement in the quantum kernels and explore the generalization performance of quantum Gaussian processes by extrapolating global six-dimensional PES in the energy domain.
I. INTRODUCTION

Predicting properties of complex molecules from first principles is considered to be one of the most promising applications of quantum computing. A computation of molecular properties within the Born-Oppenheimer approximation requires solving the electronic structure problem, fitting the results of potential energy calculations to produce global PES and solving the nuclear dynamics problem with the PES thus obtained. Several algorithms have been recently developed for solving electronic structure [1–16] and nuclear dynamics [17–22] problems on noisy intermediate-scale quantum (NISQ) computers. However, quantum algorithms for producing global PES of polyatomic molecules have not yet been demonstrated. The present work builds a quantum regression model of a six-dimensional PES for the molecular ion \( \text{H}_3\text{O}^+ \). Our results demonstrate a comparison with the corresponding classical models and illustrate the role of entanglement of the qubits used for the quantum algorithm of constructing PES.

Recent work has demonstrated that PES of polyatomic molecules can be accurately represented by machine learning (ML) regression models, based on neural networks [23–36] or kernel methods [36–56]. Quantum computers have opened the possibility to research the quantum analogues of ML algorithms [57–73]. It has been shown that gate-based quantum devices can be used to build quantum kernels for kernel ML models [60–68, 73]. While most applications of quantum kernels have been for support vector classification of low-dimensional data [61–68], several studies have considered quantum algorithms for regression [69–73]. Particularly relevant for the present work is Ref. [73] that applied Gaussian process regression to several model applications, such as regression of the one-dimensional function \( x \sin x \). The goal of Ref. [73] was to simulate classical kernels using coherent states, or truncations of coherent states. In order to extend this work to regression problems for fitting PES, it is necessary to overcome several challenges. First, there is no general quantum circuit ansatz for building performant quantum kernels for PES interpolation. It is not known how to build a sequence of quantum gates in order to build the best quantum kernel for accurate models of PES. Second, accurate kernel regression models for complex problems with sparse data require optimization of kernel parameters. However, quantum kernel estimation is expensive, requiring many quantum measurements for each pair of training points. In addition, as will be illustrated in this work, the cost function used to train regression models
with quantum kernels can be very sensitive to quantum circuit parameters. This makes kernel parameter optimization difficult. Third, the number of quantum kernel parameters grows quickly with the number of qubits and gates in the corresponding quantum circuit. This precludes grid search of optimal kernel parameters, often used for building classical kernel ridge regression models. This also makes search of optimal quantum circuit ansatz difficult.

Here, we demonstrate that quantum kernel regression models with a fixed quantum circuit ansatz readily deployable on current gate-based quantum computers can yield comparable accuracy with classical ML models. Our focus is on building accurate models with a small number of training points, aiming to produce global PES with a small number of \textit{ab initio} potential energy calculations. To achieve this, we employ Bayesian optimization for tuning the parameters of quantum gates and optimize kernels by maximizing a modified version of log marginal likelihood. We consider two problems: interpolation of PES in a six-dimensional (6D) configuration space and extrapolation of PES in the energy domain. We show that the quantum models may exhibit better extrapolation accuracy than classical models with radial basis function kernels \cite{74}, when trained by the same number and distribution of potential energy points. We also show that the accuracy of quantum models is significantly enhanced by two-qubit gates, which illustrates a critical role of qubit entanglement in quantum kernels for regression problems. By demonstrating Bayesian regression models with quantum kernels, our work complements Ref. \cite{73} to set the stage for the quantum analogue of Bayesian optimization on quantum computing devices.

II. CLASSICAL VS QUANTUM MODELS

We use Gaussian process (GP) models to represent PES of the molecule H$_3$O$^+$. The molecular geometry is described by the six-dimensional (6D) vector $\mathbf{x}$, as in our previous work \cite{76}, where we built classical GP models of PES for H$_3$O$^+$. A GP model is trained by $n$ input-output pairs, with inputs represented by $n$ molecular geometries $\mathbf{x}_i$ and outputs by $n$ corresponding values of the potential energy, collected into a column vector $\mathbf{y}$. The prediction of potential energy at an arbitrary point $\mathbf{x}^*$ in the 6D input space is given by \cite{77}:
\[
\hat{f}(x^*) = k^\top(x^*) \left[ K + \sigma^2 I \right]^{-1} y
\]

where \( \sigma^2 \) is a hyperparameter representing variance of data noise, \( I \) is the identity matrix, \( K \) is an \( n \times n \) kernel matrix with entries \( k(x_i, x_j) \), \( k^\top(x^*) \) is the transpose of a column vector with \( n \) entries \( k(x^*, x_i) \), and \( x_i \) and \( x_j \) represent the molecular geometries for the training points \( i \) and \( j \). Because PESs are noiseless, we set \( \sigma^2 \) to zero.

The function \( k(x, x') \) yielding the elements of the kernel matrix is the covariance function of the GP \[77\]. It must satisfy the properties of a kernel function of a reproducing kernel Hilbert space (RKHS). Specifically, \( k(x, x') \) must be positive-definite and symmetric to interchange of \( x \) and \( x' \). In the present work, we build GP models with classical and quantum kernels. In both cases, the prediction of the model is given by Eq. (1). The difference is in the kernel matrix \( K \). For classical models, we use radial basis functions (RBF) as the kernel function,

\[
k(x, x') = \exp \left( -\theta ||x - x'||^2 \right) .
\]

RBF kernels are known to be universal \[74\] and provide benchmark results for quantum models developed in this work.

For quantum kernels, we consider a quantum computer with \( m \) qubits, initially in state \( |0^m\rangle \). A sequence of gates operating on these qubits produces a quantum state \( \mathcal{U}(x)|0^m\rangle \). The measurable square of the inner product

\[
k(x, x') = |\langle 0^m | \mathcal{U}(x')^\dagger \mathcal{U}(x) |0^m\rangle|^2
\]

satisfies all the properties of a kernel of an RKHS. In order to build such quantum kernels, one must encode information about input vectors into parameters of the quantum gates of a quantum computer.

In the present work, we use the quantum circuit depicted in Figure 1 to build quantum kernels. This quantum circuit was introduced in Ref. [64] for classification problems. We use one qubit to represent one dimension of the input space, resulting in a 6-qubit quantum circuit for the present problem. Each qubit is initialized in state \( |0\rangle \). Following the initialization, quantum states are created by a sequence of gate operations \( \mathcal{U}(x')^\dagger \mathcal{U}(x) \), as depicted in the upper panel of Figure 1. The values of the kernels are obtained by projecting the resulting quantum states onto the state \( |0^m\rangle \).
As shown in Figure 1, the unitary transformation $U$ includes a sequence of three types of quantum gates: the Hadamard gates ($H$),

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

which put the individual qubits into coherent superposition states, the single-qubit rotation gates $R_Z$,

$$R_Z(\phi_i) = \begin{pmatrix} e^{-i\phi_i} & 0 \\ 0 & e^{i\phi_i} \end{pmatrix}$$

and the two-qubit rotation gates $R_{ZZ}$,

$$R_{ZZ}(\phi_{ij}) = \begin{pmatrix} e^{-i\phi_{ij}} & 0 & 0 & 0 \\ 0 & e^{i\phi_{ij}} & 0 & 0 \\ 0 & 0 & e^{i\phi_{ij}} & 0 \\ 0 & 0 & 0 & e^{-i\phi_{ij}} \end{pmatrix}.$$  

The two-qubit gates introduce entanglement.

The input vectors $\mathbf{x}$ are encoded into the quantum gates as follows:

$$\phi_i = \frac{x^i}{\theta_i}$$

$$\phi_{ij} = \exp\left(-\frac{(x^i - x^j)}{\theta_{ij}}\right),$$

where the superscripts in $x^i$ and $x^j$ denote the $i$-th and $j$-th components of the 6D vector $\mathbf{x}^T = [x^1, \ldots, x^6]$, and $\theta_i$ and $\theta_{ij}$ are parameters of the quantum circuit to be optimized. As shown in Figure 1, the unitary transformation $U$ is built as

$$U = U H^\otimes n U H^\otimes n,$$

with

$$U = \exp \left[ -i \left( \sum_i \phi_i(\mathbf{x}, \theta_i) \sigma_{Z,i} + \sum_{i,j>i} \phi_{ij}(\mathbf{x}, \theta_{ij}) \sigma_{Z,i} \sigma_{Z,j} \right) \right]$$

where $\sigma_{Z,i}$ is the Pauli $Z$-gate acting on qubit $i$, and the second term in the exponent correspond to the $R_{ZZ}$ gate. This ansatz includes a sequence of two-qubit rotations, entangling each pair of the qubits in the circuit. The order of the individual $R_{ZZ}$ gates in Eq. (10) is arbitrary, because the $\sigma_{Z,i}$ operators commute. The parameters $\theta_i$ are independent for each
one-qubit rotation gate in $U$. In order to simplify the optimization of the kernel parameters, we require that the parameters of all two-qubit gates $\theta_{ij}$ be the same and set them equal to a single variable parameter $\theta_{12}$. The number of free parameters in the quantum kernel is thus equal to the number of $R_Z$ gates plus one, for a total of 7 parameters in $U$.

The covariance functions of the GP models are thus parametrized by $\theta$ in the classical models and $\theta_{i=1,\ldots,6}$ and $\theta_{12}$, hereafter represented collectively by $\theta$, in the quantum models. GP models are trained by maximizing the logarithm of marginal likelihood (LML), which yields optimal parameters of the kernels [77]. For GPs, LML can be written in closed form in terms of the kernel matrix $K$ and its determinant as follows [77]:

$$\log L(\theta) = -\frac{1}{2} y^\top (K + \sigma^2 I)^{-1} y - \frac{1}{2} \log |K + \sigma^2 I| - \frac{n}{2} \log 2\pi,$$

where the dependence on $\theta$ is through the elements of the kernel matrix. While it is straightforward to train classical GP models with the RBF kernel by maximizing LML, it will be illustrated in the next section that LML for quantum models is extremely sensitive to $\theta$ in some parts of the parameter space, leading to rapid variation of LML and lack of convergence of LML optimization. In order to overcome this problem, we show that quantum models can be trained by optimizing the following objective function instead of the LML:

$$\mathcal{O}(\theta) = \log[L(\theta) + a]$$

where $a$ is a hyperparameter, set to 1 in the present work. It will be shown that the constant $a$ stabilizes optimization of LML and improves convergence.

To build quantum kernels, we use simulated qubits as implemented in the IBM qiskit package, using Statevector [78]. Quantum states are generated by the operation of gate sequences on qubits initially all in state $|0\rangle$. The gate operations are noiseless. The kernels as defined in Eq. (3) are computed from the corresponding probability amplitudes in the quantum states of $m$ qubits after the sequence of gate operations. In order to examine the role of qubit entanglement, we consider two types of kernels for the quantum models: (a) kernels constructed as described above with 7 parameters $\theta_i$ and $\theta_{12}$; (b) kernels constructed as described above, but with all two-qubit gates $R_{ZZ}$ replaced with identity matrices, yielding quantum circuits with 6 free parameters and no entanglement between qubits. We will refer to these kernels as entangled and unentangled kernels, respectively.
III. RESULTS

Although Ref. [64] illustrated that the quantum circuit ansatz described in the previous section can be used to build kernels for classification problems, this ansatz has not been used for regression models. Therefore, our first goal is to explore the possibility of using the quantum circuit depicted in Figure 1 for regression problems. Specifically, we aim to build accurate interpolation and extrapolation models with a limited number of training points (200 to 1500 for a 6D problem). In this limit, and especially for extrapolation problems, kernel regression models must be sensitive to kernels. We use a comparison with the models based on optimized RBF kernels to benchmark the performance of the quantum kernels. It should be noted that RBF kernels do not always represent the best classical kernels for kernel models of PES. As we demonstrated previously, classical GP models of PES can be improved by increasing the kernel complexity by combining different simple mathematical forms of kernels into composite kernels [76, 79]. However, RBF kernels are proven to be universal [74] and represent one of the most frequently used type of kernels. Our goal is not to illustrate that quantum kernels can outperform classical kernels for small data regression problems. Rather, we aim to show that quantum kernels can produce regression models of similar accuracy as classical kernels.

Specifically, the present section illustrates:

- how to optimize quantum circuits to build accurate quantum GP models;
- the feasibility of building accurate GP regression models with quantum kernels using a fixed quantum circuit ansatz depicted in Figure 1;
- comparison of quantum GP models for interpolation and extrapolation (in the energy domain) of PES with the classical models with optimized RBF kernels;
- comparison of quantum GP models of PES with and without entanglement between qubits.

The ab initio results for the PES of H$_3$O$^+$ are taken from Ref. [75]. There are a total of 31124 potential energy points, spanning the energy range $[0, 21000]$ cm$^{-1}$. We construct global 6D PES by training GP models using $n$ ab initio points in a specific energy interval. The value of $n$ and the energy range for the training points is specified in the caption for
each figure. The accuracy of the resulting models is quantified by computing the root mean squared error (RMSE)

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{f}(x_i))^2},
\]

(13)

where \(\hat{f}\) are the GP model predictions given by Eq. (1), \(y_i\) represent the \textit{ab initio} potential energy points from Ref [75], and the sum extends over all \textit{ab initio} points that are not used for training the models. For models trained by potential energy points from a limited range of energies (e.g. at energies \(\leq 10,000 \text{ cm}^{-1}\)), these RMSEs covering the entire energy range up to 21,000 cm\(^{-1}\) quantify the ability of GP models to extrapolate in the energy domain.

A. Quantum kernel optimization

For classical GP models with simple analytical kernel functions, LML optimization is usually performed with a gradient-based optimization method, quickly converging to desired estimates of kernel parameters. As follows from the above description of quantum kernels, LML maximization for quantum models requires optimization of a large number of parameters \(\theta\), with kernels given by the probability amplitudes in a quantum state. When implemented on a quantum computer, the present algorithm will yield kernels as quantum measurement outcomes, instead of analytical functions. This makes optimization of LML, or equivalently \(O(\theta)\), much more challenging. In this section, we illustrate that accurate quantum GP models can be obtained by optimizing \(O(\theta)\) in Eq. (12) with Bayesian optimization (BO).

BO is a gradient-free optimization method that uses a balance between the prediction of a GP and the Bayesian uncertainty of the prediction to determine how to sample the function under optimization [80]. Here, we apply BO to find the parameters of the quantum circuit \(\theta\) that maximize \(O(\theta)\). BO begins with the evaluation of \(O(\theta)\) at a small number of randomly selected values of \(\theta\). The results of these evaluations are used to train a (classical) GP model \(\mathcal{F}(\theta)\) characterized by the mean of the GP \(\mathcal{F}\) denoted as \(\mu(\theta)\) and by the uncertainty of the GP \(\mathcal{F}\) denoted as \(\sigma(\theta)\). The subsequent evaluation of \(O(\theta)\) is performed at the maximum of the acquisition function \(\alpha(\theta)\) defined as

\[
\alpha(\theta) = \mu(\theta) + \kappa \sigma(\theta),
\]

(14)
where $\kappa$ is a hyperparameter that determines the balance between exploration and exploitation. The result of the new evaluation of $O(\theta)$ is added to the set of the previous evaluations and the new set of $O$ values is used to train a new GP model $\mathcal{F}$. The procedure is iterated until convergence is reached.

We use RBF kernels for the GP models $\mathcal{F}$, initialize BO with 20 randomly chosen points and typically reach optimal results with $\sim 30 - 100$ iterations sampling 6 or 7 dimensions of the $\theta$ parameter space. We use the value of $\kappa$ in Eq. (14) set to 1. We have repeated calculations with multiple values of $\kappa$ and found that this choice of $\kappa$ leads to optimal convergence of BO for the present problems.

Figure 2 illustrates the results of optimization of LML using the objective functions defined in Eqs. (11) and (12) for unentangled (left panel) and entangled (right panel) kernels. These optimization problems vary 6 and 7 parameters, respectively. LML exhibits sharp variation with $\theta$, with characteristic drops (c.f., right panel of Figure 2), suggesting the presence of singularities for some values of the quantum circuit parameters. Qubit entanglement makes the optimization of LML more challenging. However, introducing a constant under the logarithm of the objective function as in Eq. (12) stabilizes optimization and improves convergence for GP models with both unentangled and entangled kernels. We have repeated optimization with several different values of $a \in [0, 10]$ in Eq. (12) and found that the results are not sensitive to the value of $a$. All of the calculations reported in this work use the value $a = 1$.

Figure 3 illustrates the effect of qubit entanglement on the results of LML optimization and the accuracy of the corresponding quantum GP models quantified by the RMSE over the entire data set. The results illustrate that including qubit entanglement enhances the accuracy of the quantum models. The right panel of Figure 3 shows that accurate models of 6D PES based on entangled kernels can be obtained with as few as 20 iterations of kernel optimization. This illustrates both the feasibility of obtaining accurate regression models with the fixed ansatz in Figure 1, and the efficiency of BO for optimizing quantum circuits for quantum regression problems.

Figure 4 illustrates convergence of BO of LML for quantum models with entangled kernels based on different numbers of training points. As expected, LML increases and RMSE decreases with the number of training points. The optimization of LML converges with less than 30 iterations of BO for all three models. Figure 4 illustrates that quantum GPs
produce reasonable models of 6D PES, when trained with as few as 200 potential energy points randomly sampled from the 6D configuration space. It can be observed that the optimization of the quantum circuit parameters reduces the RMSE for models with 1000 potential energy points by a factor of 3. These results illustrate that the quantum circuit ansatz introduced in Ref. [64] for classification problems is also effective for regression problems and that it is flexible enough to allow learning of complex functions by optimization of quantum gate parameters.

B. Quantum vs classical GP models of PES – interpolation

Figure 5 illustrates the interpolation performance of the optimized quantum regression model of the 6D PES of H₃O⁺ built with 1000 potential energy points. The line represents the quantum model predictions and the symbols – the potential energy points randomly sampled as functions of the separation $R$ between the centers of mass of H₂⁺ and OH fragments. At each value of $R$, we locate the energy point in the original set of ab initio points by varying the angles and/or the interatomic distances within the fragments. This energy point is then compared with the GP predictions. The training data for this model are sampled from the entire energy range of the PES. The quantum model is based on the entangled kernel and is obtained with 72 iterations of BO. The RMSE of the model is 82.30 cm⁻¹. While this is a remarkable performance of the quantum kernel, we note that the accuracy of the model can be further increased by increasing the number of training points (c.f., Figure 4).

It is instructive to compare the performance of this quantum model with that of the quantum model based on unentangled qubits and of the classical GP model. Figure 6 shows that the quantum model with entangled qubits is significantly more accurate than the quantum model with unentangled qubits. This illustrates the importance of two-qubit gates in the quantum circuit ansatz. Figure 6 also illustrates that the accuracy of the GP model with the optimized RBF kernel is very close to the accuracy of the model with the entangled kernel, except for $n = 100$. Both models approach the RMSE of about 37 cm⁻¹ as the number of training points increases.
C. **Extrapolation in the energy domain**

Several recent studies have explored the application of GP models for extrapolation problems. It was shown that the generalization accuracy of GP models increases if the complexity of GP kernels is increased by combining different simple kernels into composite kernels through an algorithm using Bayesian Information Criterion as the model selection metric [81–83]. It was shown that GP models thus constructed can extrapolate the properties of complex quantum systems across quantum phase transition lines [84]. The same approach was used to enhance the accuracy of GP models of PES for polyatomic molecules [76, 79, 85]. Since quantum circuits offer a conceptually different approach to building kernels for GP models, it is instructive to examine the potential of quantum kernels to extrapolate.

Figure 7 compares the extrapolation accuracy of quantum models with both entangled and untangled kernels and the classical model with the RBF kernel. The results shown in Figure 7 are obtained with models trained by random samples of *ab initio* potential energy points from the energy interval below the energy threshold indicated on the horizontal axis. The RMSEs shown are calculated for the entire energy range of the PES extending to 21,000 cm\(^{-1}\). Figure 7 illustrates two important results. First, including the entanglement between qubits into the quantum circuit enhances the extrapolation accuracy to a great extent. Second, models with entangled kernels appear to outperform models with the RBF kernels for low thresholds of the training data range, corresponding to a larger extrapolation interval.

To illustrate the comparison between the model predictions and the original *ab initio* energies, we show in Figure 8 the results of several models corresponding to different energy ranges of the training samples (shown by the shaded intervals). All models illustrated in Figure 8 are trained by 1500 *ab initio* points. The lines represent the GP model predictions and the symbols – the potential energy points sampled as functions of the separation between \(\text{H}_2^+\) and OH fragments. As in Figure 4, at each value of \(R\), we locate the energy point in the original set of *ab initio* points by varying the angles and/or the interatomic distances within the fragments. This energy point is then compared with the GP predictions. The functional form of PES at high energies is qualitatively different from that at low energies. Figure 8 shows that optimized quantum kernels can produce GP models that generalize predictions to different function distributions.
IV. CONCLUSION

We have demonstrated that quantum circuits of gate-based quantum computers can be used to build kernels for regression models of global PES for polyatomic molecules. Such kernels can be obtained by measuring the individual qubit states. We have shown that such kernels can be constructed with a fixed quantum circuit ansatz, previously used for classification problems, provided the quantum gate parameters are optimized to maximize \( \log[\mathcal{L} + 1] \), where \( \mathcal{L} \) is marginal likelihood. This yields Gaussian process models of PES with quantum kernels. While the standard procedure for training Gaussian process models is to maximize \( \log \mathcal{L} \), our results illustrate that \( \log \mathcal{L} \) is very sensitive to variation of the circuit parameters, making the optimization challenging. However, we have shown that maximization of \( \log[\mathcal{L} + 1] \) can be performed with Bayesian optimization, yielding stable results that correspond to accurate regression models with quantum kernels.

We have compared the accuracy of Gaussian process models of PES with quantum kernels based on entangled qubits, quantum kernels with unentangled kernels and classical Gaussian process models with RBF kernels. In all cases considered, the accuracy of quantum models including two-qubit rotation gates is comparable with the accuracy of classical models with RBF kernels. The quantum models with entangled kernels outperform the classical models with optimized RBF kernels for the class of problems aiming to construct the 6D PES at high energies based on 1500 \textit{ab initio} points at low energies. At the same time, the accuracy of all quantum models drops significantly, when the entangling two-qubit gates are omitted from the quantum circuits. This illustrates the critical role of qubit entanglement in the quantum kernel computation algorithm.

Our work demonstrates that quantum kernels obtained with a small number of qubits and quantum gates can be used for accurate regression models. This is important because finite fidelity of current NISQ devices is a major obstacle to increasing the size of quantum circuits. The quantum circuit used in the present work can be readily implemented on the current IBM quantum computer. Moreover, we have built quantum kernels for Gaussian process models, which themselves could be used as surrogate models underlying Bayesian optimization. Thus, our work complements Ref. \cite{ref} to pave the way for the development of the quantum analogue of Bayesian optimization. If quantum kernels prove to offer better inference for supervised learning tasks with a small number of training points than classical
kernels, Bayesian optimization with quantum GPs may offer a useful application of quantum computing to optimization of functions that are exceedingly expensive to evaluate.

ACKNOWLEDGMENT

This work was supported by NSERC of Canada
[1] J. D. Whitfield, J. Biamonte, and A. Aspuru-Guzik, Simulation of electronic structure Hamiltonians using quantum computers, *Mol. Phys.* **109**, 5 (2011).

[2] I. Kassal, J. Whitfield, A. Perdomo-Ortiz, M.-H. Yung, and A. Aspuru-Guzik, Simulating chemistry using quantum computers, *Annu. Rev. Phys. Chem.* **62**, 185 (2011).

[3] I. D. Kivlichan, J. McClean, N. Wiebe, C. Gidney, A. Aspuru-Guzik, G. K. Chan, and R. Babbush, Quantum simulation of electronic structure with linear depth and connectivity, *Phys. Rev. Lett.* **120**, 110501 (2018).

[4] M. B. Hastings, D. Wecker, B. Bauer, and M. Troyer, Improving quantum algorithms for quantum chemistry, arXiv:1811.11184.

[5] I. G. Ryabinkin, T. C. Yen, S. N. Genin, and A. F. Izmaylov, Qubit coupled cluster method: a systematic approach to quantum chemistry on a quantum computer, *J. Chem. Theor. Comp.* **14**, 12 (2018).

[6] K. Setia and J. D. Whitfield, Bravyi-Kitaev superfast simulation of electronic structure on a quantum computer, *J. Chem. Phys.* **148**, 164104 (2018).

[7] R. Xia, T. Bian, and S. Kais, Electronic structure calculations and the Ising Hamiltonian, *J. Phys. Chem. B* **122**, 113 (2018).

[8] K. Sugisaki, S. Yamamoto, S. Nakazawa, K. Toyota, K. Sato, D. Shiomi, and T. Takui, Quantum chemistry on quantum computers: a polynomial-time quantum algorithm for constructing the wave functions of open-shell molecules, *J. Phys. Chem. A* **120**, 32 (2016).

[9] S. Wei, H. Li, and G. Long, A full quantum eigen solver for quantum chemistry simulations, *Research* **2020**, 1486935 (2020).

[10] T. Bian, D. Murphy, R. Xia, A. Daskin, and S. Kais, Quantum computing methods for electronic states of the water molecule, *Mol. Phys.* **117**, 15 (2019).

[11] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven, and G. K. Chan, Low-depth quantum simulation of materials, *Phys. Rev. X* **8**, 011044 (2018).

[12] N. C. Rubin, A hybrid classical/quantum approach for large-scale studies of quantum systems with density matrix embedding theory, arXiv:1610.06910.

[13] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets,
Nature 549, 242 (2017).

[14] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, and D. A. Buell et al., Hartree-Fock on a superconducting qubit quantum computer, Science 369, 6507 (2020).

[15] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. 92, 015003 (2020).

[16] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, Quantum chemistry in the age of quantum computing, Chem. Rev. 119, 19 (2019).

[17] P. J. Ollitrault, A. Miessen, and I. Tavernelli, Molecular quantum dynamics: a quantum computing perspective, Acc. Chem. Res. 54, 23 (2021).

[18] P. J. Ollitrault, G. Mazzola, and I. Tavernelli, Nonadiabatic molecular quantum dynamics with quantum computers, Phys. Rev. Lett. 125, 260511 (2020).

[19] R. J. MacDonell, C. E. Dickerson, C. J. T. Birch, A. Kumar, C. L. Edmunds, M. J. Biercuk, C. Hempel and I. Kassal, Analog quantum simulation of chemical dynamics, Chem. Sci. 12, 9794 (2021).

[20] I. Kassal, S. P. Jordan, P. J. Love, M. Mohseni, and A. Aspuru-Guzik, Polynomial-time quantum algorithm for the simulation of chemical dynamics, Proc. Natl. Acad. Sci. U.S.A. 105, 18681 (2008).

[21] A. Roggero, C. Gu, A. Baroni, and T. Papenbrock, Preparation of excited states for nuclear dynamics on a quantum computer, Phys. Rev. C 102, 064624 (2020).

[22] E. T. Holland, K. A. Wendt, K. Kravvaris, X. Wu, W. E. Ormand, J. L DuBois, S. Quaglioni, and F. Pederiva, Optimal control for the quantum simulation of nuclear dynamics, Phys. Rev. A 101, 062307 (2020).

[23] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, Quantum-chemical insights from deep tensor neural networks, Nat. Commun. 8, 13890 (2017).

[24] O. T. Unke and M. Meuwly, PhysNet: a neural network for predicting energies, forces, dipole moments and partial charges, J. Chem. Theor. Comp. 15, 3678 (2019).

[25] S. Manzhos and T. Jr. Carrington, A random-sampling high dimensional model representation neural network for building potential energy surfaces, J. Chem. Phys. 125, 084109 (2006).
[26] S. Manzhos, X. Wang, R. Dawes, and T. Jr. Carrington, A nested molecule-independent neural network approach for high-quality potential fits, *J. Phys. Chem. A* **110**, 5295 (2006).

[27] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Phys. Rev. Lett.* **98**, 146401 (2007).

[28] J. Behler, Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations, *Phys. Chem. Chem. Phys.* **13**, 17930 (2011).

[29] J. Behler, Constructing high-dimensional neural network potentials: A tutorial review, *Int. J. Quant. Chem.* **115**, 1032 (2015).

[30] E. Pradhan and A. Brown, A ground state potential energy surface for HONO based on a neural network with exponential fitting functions, *Phys. Chem. Chem. Phys.* **19**, 22272 (2017).

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

[32] S. Manzhos, R. Dawes, and T. Jr. Carrington, Neural network-based approaches for building high dimensional and quantum dynamics-friendly potential energy surfaces, *Int. J. Quant. Chem.* **115**, 1012 (2015).

[33] 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, *J. Chem. Phys.* **138**, 154301 (2013).

[34] 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, *J. Phys. Chem. C* **122**, 1761 (2018).

[35] S. Manzhos and T. Jr. Carrington, Neural network potential energy surfaces for small molecules and reactions, *Chem. Rev.* **121**, 16 (2021).

[36] M. Meuwly, Machine learning for chemical reactions, *Chem. Rev.* **121**, 16 (2021).

[37] 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).

[38] 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).
[39] A. P. Bartók and G. Csányi, Gaussian approximation potentials: A brief tutorial introduction, *Int. J. Quant. Chem.* **115**, 1051 (2015).

[40] 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).

[41] 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).

[42] B. Kolb, P. Marshall, B. Zhao, B. Jiang, and H. Guo, Representing global reactive potential energy surfaces using Gaussian processes, *J. Phys. Chem. A* **121**, 2552 (2017).

[43] A. Kamath, R. A. Vargas-Hernandez, R. V. Krems, T. Jr. Carrington, 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).

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

[45] 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).

[46] 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).

[47] 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).

[48] A. E. Wiens, A. V. Copan, and H. F. Schaefer, Multi-fidelity Gaussian process modeling for chemical energy surfaces, *Chem. Phys. Lett.* **X 3**, 100022 (2019).

[49] C. Qu, Q. Yu, B. L. Jr Van Hoozen, J. M. Bowman, and Vargas-Hernández, R. A. Assessing Gaussian process regression and permutationally invariant polynomial approaches to represent high-dimensional potential energy surfaces, *J. Chem. Theor. Comp.* **14**, 3381 (2018).

[50] Q. Song, Q. Zhang, and Q. Meng, Revisiting the Gaussian process regression for fitting high-dimensional potential energy surface and its application to the OH + HO₂ → O₂ + H₂O reaction, *J. Chem. Phys.* **152**, 134309 (2020).
[51] C. Qu, R. Conte, P. L. Houston, and J. M. Bowman, Full-dimensional potential energy surface for acetylacetone and tunneling splittings, Phys. Chem. Chem. Phys. 23, 7758 (2021).

[52] O. T. Unke and M. Meuwly, Toolkit for the construction of reproducing kernel-based representations of data: Application to multidimensional potential energy surfaces, J. Chem. Inf. Model. 57, 1923 (2017).

[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] U. T. Unke, Potential energy surfaces: from force fields to neural networks, Doctoral dissertation, University of Basel, 2019.

[57] Y. Liu, S. Arunachalam, and K. Temme, A rigorous and robust quantum speed-up in supervised machine learning, Nat. Phys. 17, 1013 (2021).

[58] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Circuit-centric quantum classifiers, Phys. Rev. A 101, 032308 (2020).

[59] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Parameterized quantum circuits as machine learning models, Quantum Sci. Technol. 4, 043001 (2019).

[60] M. Schuld, I. Sinayskiy, and F. Petruccione, An introduction to quantum machine learning, Contemp. Phys. 56, 2 (2015).

[61] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature 549, 195 (2017).

[62] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum support vector machine for big data classification, Phys. Rev. Lett. 113, 130503 (2014).

[63] S. Maria and N. Killoran, Quantum machine learning in feature hilbert spaces, Phys. Rev. Lett. 122, 040504 (2019).

[64] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature 567, 209 (2019).
[65] Y. Suzuki, H. Yano, Q. Gao, S. Uno, T. Tanaka, M. Akiyama, and N. Yamamoto, Analysis and synthesis of feature map for kernel-based quantum classifier, *Quantum Mach. Intell.* 2, 9 (2020).

[66] J. Park, B. Quanz, S. Wood, H. Higgins, and R. Harishankar, Practical application improvement to Quantum SVM: theory to practice, [arXiv:2012.07725](https://arxiv.org/abs/2012.07725).

[67] R. Chatterjee and T. Yu, Generalized coherent states, reproducing kernels, and quantum support vector machines, [arXiv:1612.03713](https://arxiv.org/abs/1612.03713).

[68] J. R. Glick, T. P. Gujarati, A. D. Córcoles, Y. Kim, A. Kandala, J. M. Gambetta, and K. Temme, Covariant quantum kernels for data with group structure, [arXiv:2105.03406](https://arxiv.org/abs/2105.03406).

[69] M. Schuld, I. Sinayskiy, and F. Petruccione, Prediction by linear regression on a quantum computer, *Phys. Rev. A* 94, 022342 (2016).

[70] G. Wang, Quantum algorithm for linear regression, *Phys. Rev. A* 96, 012335 (2017).

[71] P. Date and T. Potok, Adiabatic quantum linear regression, *Scientific Reports* 11, 21905 (2021).

[72] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, Continuous-variable quantum neural networks, *Phys. Rev. Research* 1, 033063 (2019).

[73] M. Otten, I. R. Goumiri, B. W. Priest, G. F. Chapline, and M. D. Schneider, Quantum machine learning using Gaussian processes with performant quantum kernels, [arXiv:2004.11280](https://arxiv.org/abs/2004.11280).

[74] J. Wang, Q. Chen, and Y. Chen, *RBF kernel based support vector machine with universal approximation and its application - Advances in neural networks*, edited by F. Yin, J. Wang, C. Guo, (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004), pp.512-517.

[75] Q. Yu and J. M. Bowman, *Ab initio* potential for H$_5$O$^+$ → H$^+$ + H$_2$O: A step to a many-body representation of the hydrated proton?, *J. Chem. Theor. Comp.* 12, 5284 (2016)

[76] J. Dai and R. V. Krems, Interpolation and extrapolation of global potential energy surfaces for polyatomic systems by Gaussian processes with composite kernels, *J. Chem. Theor. Comp.* 16, 3 (2020).

[77] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning* (The MIT Press, Cambridge, 2006).

[78] G. Aleksandrowicz, T. Alexander, P. Barkoutsos, L. Bello, Y. Ben-Haim, D. Bucher, F. J. Cabrera-Hernández, J. Carballo-Franquis, A. Chen, and C. Chen *et al.*, Qiskit: An open-source framework for quantum computing, doi=10.5281/zenodo.2573505.
[79] K. Asnaashari and R. V. Krems, Gradient domain machine learning with composite kernels: improving the accuracy of PES and force fields for large molecules, *Mach. Learn.: Sci. & Technol.* **3**, 015005 (2022).

[80] R. A. Vargas-Hernández, 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).

[81] G. Schwarz, Estimating the dimension of a model, *The Annals of Statistics* **2**, 461 (1978).

[82] D. K. Duvenaud, H. Nickisch, and C. E. Rasmussen, Additive gaussian processes, *Adv. Neur. Inf. Proc. Sys.* **24**, 226 (2011).

[83] 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).

[84] R. A. Vargas-Hernández, 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).

[85] H. Sugisawa, T Ida, and R. V. Krems, Gaussian process model of 51-dimensional potential energy surface for protonated imidazole dimer, *J. Chem. Phys.* **153**, 11 (2020).
FIG. 1. Quantum circuit used in the present work to build quantum kernels of Gaussian process models. The sequence of gates in $\mathbf{U}$ is determined by Eq. (10). $H$ denotes Hadamard gates and $R_Z$ – single qubit rotation gates. See text for more details.

FIG. 2. LML of quantum GP models with unentangled kernels (left panel) and entangled kernels (right panel) as functions of the number of BO iterations. Upper curves (red): LML obtained by maximization of $\mathcal{O}$ as defined by Eq. (12). Lower curves (blue): LML obtained by maximization of $\log \mathcal{L}$. All GPs are trained by the same set of $n = 1000$ energy points randomly selected from the entire energy range $[0, 21000]$ cm$^{-1}$. 
FIG. 3. Left panel: Maximum value of LML (left panel) and RMSE (right panel) of quantum GP models with entangled and unentangled kernels as functions of the number of BO iterations. All GPs are trained by the same set of $n = 1000$ energy points randomly selected from the entire energy range $[0, 21000]$ cm$^{-1}$.

FIG. 4. Maximum value of LML (left panel) and RMSE (right panel) of quantum GP models with entangled kernels as functions of the number of BO iterations for different numbers of training points: circles – $n = 200$; squares – $n = 400$; triangles – $n = 1000$. The models are trained by $ab\ initio$ points randomly sampled from the energy interval $[0, 21000]$ cm$^{-1}$.
FIG. 5. Comparison of quantum GP model predictions (solid curve) with the original potential energy points (symbols) for H$_3$O$^+$ as functions of the separation between the H$_2^+$ and OH fragments. The variable R specifies the distance between the O atom and one of the H atoms in the H$_2^+$ fragment. At each value of R, we locate the energy point in the original set of ab initio points by varying the angles and/or the interatomic distances within the fragments. This energy point is then compared with the GP predictions. The 6D GP model is trained by 1000 ab initio points randomly selected from the entire energy range and uses the entangled kernel.
FIG. 6. Dependence of the RMSE for GP models with quantum kernels based on quantum circuits with unentangled qubits (triangles), entangled qubits (circles) and classical RBF kernel (stars) on the number of training energy points. The models are trained by *ab initio* points randomly sampled from the energy interval [0, 21000] cm$^{-1}$. The RMSEs are calculated using all remaining energy points in the same energy interval that are not used for training.
FIG. 7. Extrapolation in the energy domain: RMSE for GP models with quantum kernels based on quantum circuits with unentangled qubits (triangles), entangled qubits (circles) and classical RBF kernel (stars) as functions of the training energy threshold. All models are trained by 1500 randomly selected \textit{ab initio} points from the energy interval below the indicated energy threshold. The RMSEs are calculated using all remaining energy points that are not used for training and that cover the energy interval $[0, 21000]$ cm$^{-1}$. 

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig7.png}
\caption{Extrapolation in the energy domain: RMSE for GP models with quantum kernels based on quantum circuits with unentangled qubits (triangles), entangled qubits (circles) and classical RBF kernel (stars) as functions of the training energy threshold. All models are trained by 1500 randomly selected \textit{ab initio} points from the energy interval below the indicated energy threshold. The RMSEs are calculated using all remaining energy points that are not used for training and that cover the energy interval $[0, 21000]$ cm$^{-1}$.}
\end{figure}
FIG. 8. Comparison of GP model predictions (solid curve) with the original potential energy points (symbols) for $\text{H}_3\text{O}^+$ as functions of the separation between the $\text{H}_2^+$ and OH fragments. The GP models are trained by 1500 $\text{ab initio}$ points randomly selected from the energy interval shown by the blue shaded region. The variable $R$ specifies the distance between the O atom and one of the H atoms in the $\text{H}_2^+$ fragment. At each value of $R$, we locate the energy point in the original set of $\text{ab initio}$ points by varying the angles and/or the interatomic distances within the fragments. This energy point is then compared with the GP predictions.