The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Gaussian process regression adaptive density-guided approach: Toward calculations of potential energy surfaces for larger molecules

Cite as: J. Chem. Phys. 159, 024102 (2023); doi: 10.1063/5.0152367
Submitted: 29 March 2023 • Accepted: 15 June 2023 •
Published Online: 10 July 2023

Denis G. Artiukhin,[1,a] Ian H. Godtliebsen,[2,b] Gunnar Schmitz,[3,c] and Ove Christiansen[2,d]

AFFILIATIONS

[1] Institut für Chemie und Biochemie, Freie Universität Berlin, Arnimallee 22, 14195 Berlin, Germany
[2] Department of Chemistry, Aarhus Universitet, DK-8000 Aarhus, Denmark
[3] Lehrstuhl für Theoretische Chemie II, Ruhr-Universität Bochum, Universitätstraße 150, 44801 Bochum, Germany

[a] **Author to whom correspondence should be addressed:** denis.artiukhin@fu-berlin.de
[b] **E-mail:** ian@chem.au.dk
[c] **E-mail:** gunnar.schmitz@rub.de
[d] **E-mail:** ove@chem.au.dk

ABSTRACT

We present a new program implementation of the Gaussian process regression adaptive density-guided approach [Schmitz *et al.*, J. Chem. Phys. **153**, 064105 (2020)] for automatic and cost-efficient potential energy surface construction in the MidasCpp program. A number of technical and methodological improvements made allowed us to extend this approach toward calculations of larger molecular systems than those previously accessible and maintain the very high accuracy of constructed potential energy surfaces. On the methodological side, improvements were made by using a Δ-learning approach, predicting the difference against a fully harmonic potential, and employing a computationally more efficient hyperparameter optimization procedure. We demonstrate the performance of this method on a test set of molecules of growing size and show that up to 80% of single point calculations could be avoided, introducing a root mean square deviation in fundamental excitations of about 3 cm^{-1}. A much higher accuracy with errors below 1 cm^{-1} could be achieved with tighter convergence thresholds still reducing the number of single point computations by up to 68%. We further support our findings with a detailed analysis of wall times measured while employing different electronic structure methods. Our results demonstrate that GPR-ADGA is an effective tool, which could be applied for cost-efficient calculations of potential energy surfaces suitable for highly accurate vibrational spectra simulations.

I. INTRODUCTION

Constructions of potential energy surfaces (PESs) are of vast interest for many fields of chemistry as they can provide a detailed insight into dynamics and reactivity of molecules. Although the dimensionality of the PES increases linearly with the number of nuclei K_{nuc}, i.e., it is equal to the number of vibrational modes $M = 3K_{nuc} - 6(5)$, the computational cost of the PES construction scales exponentially due to the need to compute mode–mode coupling terms. As the result, calculations of accurate fully coupled PESs are prohibitively expensive and are only possible for up to four-atomic molecules. A commonly used approach to reduce the computational cost of the PES construction relies on the restriction of high-order mode couplings and is known under several names such as the *n*-mode expansion,[1–5] cluster expansion,[6] or high-dimensional model representation.[7] This method reduces the cost of the PES generation significantly and enables calculations of up to a few dozens of atoms. A large number of different strategies could be applied to further decrease the cost of *n*-mode-expanded PESs. These include, to name but a few, many-body expansions,[8,9] approximate computations of high-order coupling terms,[4,10–16] various screening techniques,[4,17–23] vibrational space dimensionality reduction,[17,24,25] and the use of molecular symmetry.[14,26]

If PESs are constructed on a grid, the corresponding computational cost could be high due to a non-optimal spatial placement of individual grid points and, as the result, a large number of them. The

design of good ways of constructing the grid of points to avoid a very large computational burden can require significant human time. The adaptive density-guided approach (ADGA)[27–29] is designed to mitigate these issues. It constructs grids within the n-mode representation in an iterative procedure being guided by one-mode vibrational densities. The ADGA ensures that a modest number of single points (SPs) are computed while maintaining a very high accuracy of the PES. Furthermore, it enables a fully automatic determination of the grid dimensions and granularity without using any prior knowledge about the molecular system. Having proved to be a reliable method for PES computations, the ADGA was further extended methodologically with various algorithms for the grid boundary extension,[29] the use of energy derivatives and molecular point groups of symmetry,[14] and different fitting functions for the analytical representation of PESs.[29] Additionally, combinations of the ADGA with multiresolution PES computations[13] and double incremental PES expansions[30] were presented.

The use of machine learning (ML) algorithms for constructing PESs and/or assisting in their computations is also gaining its momentum. Thus, neural networks (NNs) were already successfully employed for PES representation and for vibrational structure calculations.[31–37] Furthermore, in case of molecular dynamic simulations, it was demonstrated that NNs can extend the simulation time and treat large molecular systems with an accuracy similar to that of density functional theory.[38–42] Gaussian Process Regression (GPR),[43] a nonparametric Bayesian ML approach, deserves special attention as it provides uncertainties for predicted data points. This allows us to estimate the quality of the fit at regions of interest and make decisions on whether these regions should be supplied with additional training data points. Note, however, that uncertainty estimates could also be made by combining several NNs in a committee.[44] This characteristic of GPR makes it well-suited for the use in Bayesian optimization and active learning. In this regard, Jinnouchi *et al.*[45,46] demonstrated an on-the-fly force-field generation scheme. The use of GPR for the direct representation of the PES was reported in Refs. 47–52, whereas applications of GPR for positioning grid points were very recently demonstrated by Schneider *et al.*[53] Furthermore, it was employed to accelerate certain computational steps by, for example, evaluating PES matrix elements in a convenient format,[54] accelerating time-dependent dynamics,[55] and speeding up numerical structure optimization.[56–58] For a recent review on the use of GPR in the context of computational chemistry and material research, we refer to Ref. 59.

Recently, a combination of the ADGA and GPR method was presented in Ref. 60. The new approach, dubbed GPR-ADGA, employed statistical uncertainties from GPR along-side with averaged vibrational densities calculated with the ADGA as criteria for choosing whether SPs should be predicted with inexpensive GPR or calculated with a more accurate yet costly electronic structure method. The performance of GPR-ADGA was assessed by computing fundamental excitation energies from generated PESs. It was demonstrated that GPR-ADGA could reduce the number of SPs by 65%–90% while introducing a root mean square deviation (RMSD) in fundamental frequencies below 2 cm^{-1} compared to the standard ADGA. It is interesting to note that similar reduction in the number of SPs and errors in fundamental frequencies was obtained with the approach by Schneider *et al.*[53] Common to these works is that GPR, one way or the other, is involved in deciding where the final SPs are placed, and resulting from this, a significantly smaller set of SPs is sufficient. It can thus be firmly established that the statistical GPR theory can be used for significantly boosting PES computations. Unfortunately, the original GPR-ADGA algorithm was applicable only for small systems (order four atoms or so) due to a number of technical and methodological difficulties arising from the GPR algorithm itself and its combination with the ADGA (for more details, see Sec. II E). In the current work, we lift this limitation and demonstrate an improved and extended version of GPR-ADGA, which is applied for PES computations of up to 10-atomic molecules while maintaining a high accuracy in fundamental excitation energies and a large reduction in the number of SPs (when compared to the standard ADGA).

This work is organized as follows: The underlying theory of the GPR-ADGA method together with its recent technical and methodological extensions is described in Sec. II. The computational details are provided in Sec. III and followed by GPR-ADGA computations of PESs presented in Sec. IV. Subsequently, conclusions to this work are given in Sec. V.

II. THEORY

In the following, we briefly summarize GPR-ADGA components such as the n-mode expansion[1–5] in Sec. II A, the theory behind the ADGA[27–29] in Sec. II B, and GPR[43] in Sec. II C. Then, in Sec. II D, we describe the main idea of the GPR-ADGA method and focus on its recent methodological extensions, which enable calculations of larger molecules, in Sec. II E.

A. n-mode expansion

As was described above in Sec. I, constructions of full-dimensional PESs $V(\mathbf{q})$, depending on $M = 3K_{\mathrm{nuc}} - 6(5)$ number of normal vibrational coordinates $\mathbf{q} = \{q_1, q_2, \ldots, q_M\}$, are prohibitively expensive for more than about four atoms, i.e., for $K_{\mathrm{nuc}} \gtrsim 4$. In order to lift this limitation and construct PESs for larger molecular systems, additional approximations need to be invoked. To that end, we first define mode combinations (MCs) $\mathbf{m}_k$ as sets $\{m_1, m_2, \ldots, m_k\}$ containing k coordinate indices. Subsequently, mode combination ranges (MCRs) are formed as sets of MCs (for more details on MCs and MCRs, see Ref. 8). A full-dimensional PES $V(\mathbf{q})$ can then be represented as[8]

$$V(\mathbf{q}) = \sum_{\mathbf{m}_k \in \mathrm{MCR}} \tilde{V}^{\mathbf{m}_k} = \sum_{\mathbf{m}_k \in \mathrm{MCR}} \sum_{\substack{m_{k'} \subseteq m_k \\ m_{k'} \in \mathrm{MCR}}} (-1)^{k-k'} V^{\mathbf{m}_{k'}}, \qquad (1)$$

where the outer sum in the right-hand side runs over all MCs $\mathbf{m}_k$ from the MCR and the inner sum runs over all subsets of $\mathbf{m}_k$ (including $\mathbf{m}_k$ itself). In this formulation, the potential $V(\mathbf{q})$ is conveniently represented as a sum of its lower-dimensional cuts excluding overcounting of equivalent terms. In Eq. (1), the equality sign holds if the MCR contains MCs $\mathbf{m}_k$ of up to Mth order. Constructing the MCR from MCs with at most n-mode indices (where $n < M$), one neglects mode–mode couplings of higher orders and provides an approximate treatment of the PES. This approach drastically reduces the number of SPs to be computed (compared to a fully coupled PES)

and is often referred to as the n-mode approximation.[1–5] The number of SPs to be computed when constructing a PES on a grid of points within the n-mode representation is given by

$$N_{\mathrm{SPs}} = \sum_{k=1}^{n} \binom{M}{k} (g_k)^k, \tag{2}$$

where g_k is the number of SPs in the direct product grid per MC $\mathbf{m}_k$. The number g_k required for the accurate PES representation is usually unknown and could vary for different regions of the same PES. Another complication lies in the constructions of such grids of points. Static grids with predefined and equidistantly separated points offer an easy solution to this issue. However, this approach is by no means optimal and often results in a large number of SPs to be computed (for example, see Ref. 30).

B. Adaptive density-guided approach

The ADGA[27–29] is designed for a fully automatic grid construction and has an advantage over above-mentioned static grid approaches. It calculates PESs employing the n-mode representation and an iterative procedure, which is guided by one-mode averaged vibrational densities of the form

$$\rho^{\mathrm{ave}}_{\mathrm{iter}}(q_{m_k}) = \frac{1}{N^{m_k}_{\mathrm{modal}}} \sum_{s^{m_k}=1}^{N^{m_k}_{\mathrm{modal}}} |\varphi^{m_k}_{s^{m_k}}(q_{m_k})|^2, \tag{3}$$

which are obtained in each iteration from vibrational self-consistent field (VSCF) calculations.[61–64] In Eq. (3), $\varphi^{m_k}_{s^{m_k}}(q_{m_k})$ are orthonormal one-mode wave functions (modals) used to describe a vibrational mode q_{m_k} and $N^{m_k}_{\mathrm{modal}}$ is the number of these modals. Averaged vibrational densities $\rho^{\mathrm{ave}}_{\mathrm{iter}}(q_{m_k})$ and the corresponding one-mode potentials $V^{m_k}_{\mathrm{iter}}(q_{m_k})$ are then used to calculate an energy-like quantity, which in the one-dimensional case is given by

$$\Xi^{m_k}_{\mathrm{iter}} = \int_{I_{m_k}} \rho^{\mathrm{ave}}_{\mathrm{iter}}(q_{m_k}) V^{m_k}_{\mathrm{iter}}(q_{m_k}) \mathrm{d}q_{m_k}. \tag{4}$$

$\Xi^{m_k}_{\mathrm{iter}}$ is computed for all intervals I_{m_k} defined by neighboring SPs of mode q_{m_k}. If $\Xi^{m_k}_{\mathrm{iter}}$, computed for a particular interval I_{m_k}, changes significantly between two ADGA iterations, the corresponding interval is divided at the middle by inserting a new SP. Note that required SPs are computed with external electronic structure programs. The procedure stops when no significant changes in $\Xi^{m_k}_{\mathrm{iter}}$ are detected for all modes and intervals. After that, the ADGA continues analogously for higher mode couplings until convergence at the specified MC level is achieved. Furthermore, the ADGA automatically extends the grid boundaries if a non-negligible amount of the average vibrational density $\rho^{\mathrm{ave}}_{\mathrm{iter}}(q_{m_k})$ is detected outside the current grid. The convergence of the ADGA is controlled with three criteria, ϵ_{rel}, ϵ_{abs}, and ϵ_{ρ}, where ϵ_{rel} and ϵ_{abs} assess the relative and absolute change in the integral value $\Xi^{m_k}_{\mathrm{iter}}$, respectively, between subsequent iterations and ϵ_{ρ} checks for the amount of vibrational density outside the grid boundaries. For more details on these thresholds, we refer to the original works in Refs. 27 and 29.

C. Gaussian process regression

A further speed-up of the PES construction procedure within the ADGA could be achieved by replacing some of the costly SP computations with inexpensive GPR predictions. To that end, we introduce vectors of coordinates $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})^{\mathrm{T}}$, each describing a particular ith molecular conformation. These vectors $\mathbf{x}_i$ are often referred to as *input* or *feature vectors*.[65] In our previous studies in Refs. 16, 57, and 60, $\mathbf{x}_i$ were minimal sets of internal coordinates. In the current work, we still apply internal coordinates but additionally standardize them by shifting and scaling each feature (for more details, see Secs. 2.5 and S1.1 in the supplementary material). Note, however, that normal coordinates $\mathbf{q}$ are employed in the ADGA computations of this paper although extension of the ADGA to other coordinates has been reported.[66] The known values of the potential $V(\mathbf{x})$ for a given set of molecular structures $\{\mathbf{x}_i\}_i^N$ are then collected in a vector $\mathbf{v} = (V(\mathbf{x}_1), V(\mathbf{x}_2), \ldots, V(\mathbf{x}_N))^{\mathrm{T}}$, which can be regarded as a vector of *outputs* or *labels*.[65] We assume that the elements of $\mathbf{v}$ have a multivariant Gaussian distribution, i.e.,

$$\mathbf{v} \sim \mathcal{N}(\mathbf{m}, \mathbf{K} + \sigma_N^2 \mathbf{I}), \tag{5}$$

where $\mathbf{m}$ is the prior mean vector of length N, σ_N^2 is a regularization parameter or noise, and $\mathbf{I}$ is the $N \times N$ identity matrix. $\mathbf{K}$ is the prior $N \times N$ covariance matrix with elements $(\mathbf{K})_{ij}$ being equal to the kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ evaluated for molecular structures $\mathbf{x}_i$ and $\mathbf{x}_j$. The specific kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ used in this study is defined later in the text in Sec. II E. In many practical applications of GPR, including our previous studies in Refs. 16, 57, and 60, the mean vector $\mathbf{m}$ is set to zero. However, in the current work, the components of $\mathbf{m}$ are given as potential energy functions for the quantum mechanical harmonic oscillator,

$$m(\mathbf{x}_i) = E_0 + \frac{1}{2}(\mathbf{x}_i - \mathbf{x}_0)^{\mathrm{T}} \mathbf{H} (\mathbf{x}_i - \mathbf{x}_0), \tag{6}$$

with $\mathbf{x}_0$ and E_0 being the optimized molecular structure in a minimal set of internal coordinates and the corresponding reference energy, respectively, and $\mathbf{H}$ being the matrix of second derivatives of energy at $\mathbf{x}_0$ with respect to molecular displacements (i.e., the Hessian matrix). Therefore, the GPR-based approach employed in this work predicts deviations of $V(\mathbf{x})$ from a harmonic potential given in Eq. (6) and could be regarded as a variant of Δ-learning[67] or semi-parametric GPR.[43] It is, of course, trivial to extend the procedure to other $m(\mathbf{x}_i)$.

In order to predict unknown values of the potential $\mathbf{v}^* = (V(\mathbf{x}_1^*), V(\mathbf{x}_2^*), \ldots, V(\mathbf{x}_{N^*}^*))^{\mathrm{T}}$ for a set of N^* molecular structures $\{\mathbf{x}_i^*\}_i^{N^*}$, the joint Gaussian distribution of unknown values $\mathbf{v}^*$ and observations $\mathbf{v}$ is conditioned on $\mathbf{v}$ (for more details on conditioning, see Ref. 65). The resulting conditional probability (of $\mathbf{v}^*$, given $\mathbf{v}$) has a multivariant Gaussian distribution, i.e.,

$$\mathbf{v}^* | \mathbf{v} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{7}$$

with the new posterior mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The posterior mean vector $\boldsymbol{\mu}$ has the length N^* and is defined as[43]

$$\boldsymbol{\mu} = \mathbf{m}^* + (\mathbf{K}^*)^{\mathrm{T}} (\mathbf{K} + \sigma_N^2 \mathbf{I})^{-1} (\mathbf{v} - \mathbf{m}), \tag{8}$$

where $\mathbf{K}^*$ denotes the $N \times N^*$ matrix of elements $(\mathbf{K}^*)_{ij} = k(\mathbf{x}_i, \mathbf{x}_j^*)$ and $\mathbf{m}^*$ is a vector of length N^* containing mean values from Eq. (6) evaluated for $\mathbf{x}_i^*$. The posterior covariance matrix $\boldsymbol{\Sigma}$ from Eq. (7) is given by[43]

$$\boldsymbol{\Sigma} = \mathbf{K}^{**} - (\mathbf{K}^*)^{\mathrm{T}} (\mathbf{K} + \sigma_N^2 \mathbf{I})^{-1} \mathbf{K}^*. \tag{9}$$

Here, $\mathbf{K}^{**}$ is the $N^* \times N^*$ matrix of elements $(\mathbf{K}^{**})_{ij} = k(\mathbf{x}_i^*, \mathbf{x}_j^*)$. Subsequently, the posterior mean vector $\boldsymbol{\mu}$ is used as a predictor of unknown values $\mathbf{v}^*$, i.e., we assume that $\mathbf{v}^* \approx \boldsymbol{\mu}$ while the diagonal elements of the posterior covariance matrix $\mathbb{V}[V(\mathbf{x}_i^*)] = (\boldsymbol{\Sigma})_{ii}$ are employed as statistical error estimates for the predicted values $V(\mathbf{x}_i^*)$ of the potential.

D. Combined GPR-ADGA methodology

As was mentioned above, the general idea of GPR-ADGA[60] lies in using the GPR variance for a predicted potential value $V(\mathbf{q}_i^*)$ with the corresponding averaged VSCF vibrational density as criteria for choosing whether the predicted value $V(\mathbf{q}_i^*)$ should be included in the PES as is or re-calculated with a more accurate and expensive electronic structure method. In practice, the procedure is carried out in an iterative manner and starts from training the GPR predictor on the available dataset of coordinates $\mathbf{q}_i$ and corresponding energy values $V(\mathbf{q}_i)$ (and possibly energy derivatives with respect to $\mathbf{q}_i$) computed with an electronic structure method. Then, the ADGA calculation is carried out until its full convergence using GPR-predicted SPs. For each SP of the constructed PES, the following quantity,

$$\Omega^{\mathbf{m}_n}_{\mathrm{box}_i}(\mathbf{q}_i^*) = A^{\mathbf{m}_n} \rho^{\mathbf{m}_n} \mathbb{V}[V(\mathbf{q}_i^*)], \tag{10}$$

is evaluated. In Eq. (10), $A^{\mathbf{m}_n}$ is the box size (i.e., length, area, or volume for one-, two-, or three-dimensional potential cuts, respectively) equal to $\int_{\mathrm{box}_i} \mathrm{d}q_{m_1}, \mathrm{d}q_{m_2}, \ldots, \mathrm{d}q_{m_n}$, $\rho^{\mathbf{m}_n}$ is a vibrational density computed as a product $\prod_{m_k \in \mathbf{m}_n} \rho^{\mathrm{ave}}_{\mathrm{iter}}(q_{m_k})$ of one-mode VSCF averaged vibrational densities $\rho^{\mathrm{ave}}_{\mathrm{iter}}(q_{m_k})$, and $\mathbb{V}[V(\mathbf{q}_i^*)]$ is the GPR variance for the predicted energy value $V(\mathbf{q}_i^*)$. The units of $\rho^{\mathbf{m}_n}$ are inverse of those for $A^{\mathbf{m}_n}$. Therefore, the units of $\Omega^{\mathbf{m}_n}_{\mathrm{box}_i}(\mathbf{q}_i^*)$ are those of the variance $\mathbb{V}[V(\mathbf{q}_i^*)]$ and are $[\mathrm{energy}^2]$. A measure involving the square root of the variance $(\Omega^{\mathbf{m}_n}_{\mathrm{box}_i}(\mathbf{q}_i^*) = A^{\mathbf{m}_n} \rho^{\mathbf{m}_n} \sqrt{\mathbb{V}[V(\mathbf{q}_i^*)]})$ with units of energy can also be used but was found inferior in the original work and was thus not pursued here. SPs for which the value of $\Omega^{\mathbf{m}_n}_{\mathrm{box}_i}(\mathbf{q}_i^*)$ is larger than the specified threshold T_Ω (and some other selection rules are fulfilled; see Ref. 60) are collected in a list. All SPs from this list are then re-calculated with the electronic structure method and added to the existing dataset of energy values. Subsequently, the steps including the GPR predictor training, ADGA computation, and another selection of SPs are repeated. The procedure is continued until the list of SPs for which $\Omega^{\mathbf{m}_n}_{\mathrm{box}_i}(\mathbf{q}_i^*) > T_\Omega$ is empty.

E. Extension to larger molecules

A large number of technical and methodological modifications to original GPR-ADGA[60] in the Molecular Interactions Dynamics And Simulation Chemistry Program Package (MidasCpp)[68] were performed within this work to enable computations of PESs for up to ten atoms (for examples of PES computations, see Sec. IV). The former includes (i) predicting energy values $V(\mathbf{q}_i^*)$ in batches for better stability and easier parallelization, (ii) OpenMP parallelization over the number of batches, (iii) more efficient use of memory and disk space, and (iv) implementation of initial guesses for hyperparameter optimization (see Sec. S1.2 in the supplementary material) as well as a general clean-up removing redundant steps and improving the overall performance. Methodological changes require a more detailed consideration.

As was mentioned already in Sec. II C and shown in Eq. (6), we use a variant of the Δ-learning technique by predicting the difference between the actual PES and a PES described within the harmonic approximation. This choice has two advantages. Thus, by predicting the anharmonicity correction instead of the full PES, we potentially also decrease errors in predicted values. Furthermore, this allows us using the derivative information for the reference point $\mathbf{x}_0$ in the GPR mean function as opposed to placing it directly into the training set (as was performed in our previous studies in Refs. 16 and 60). Therefore, we avoid enlarging the training set size.

Another important methodological change is related to hyperparameter optimization. In our previous studies in Refs. 16 and 60, we used kernels with one signal variance σ_f^2 and M-number of characteristic length-scale parameters l_i, each being optimized for an individual degree of freedom. This provided a better optimization flexibility and allowed GPR to adjust to the physical nature of each coordinate. However, with the number of atoms growing, hyperparameter optimization quickly becomes prohibitively expensive, whereas a large number of parameters to be optimized lead to numerical instabilities and multiple minimas being present on the hypersurface. These limitations were lifted by first standardizing coordinates, i.e., by shifting and scaling each feature and label in the original training set. In the current work, we center our dataset by removing the mean value of each feature/label and subsequently dividing them by their population standard deviations. Note that data standardization could also be applied to training sets containing derivatives (for a more detailed description of this procedure and for the full list of data standardization options implemented in MidasCpp, see Sec. S1.1 in the supplementary material). This transformation ensures that all features are within comparable-length intervals, and thereby, it supports the use of a single length-scale parameter l for all degrees of freedom. As a result, the use of simpler kernels is enabled, such as the squared exponential kernel,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2l^2}\right), \tag{11}$$

which was employed in this work. Furthermore, we implemented an additional criterion controlling the hyperparameter optimization procedure. In our setup, hyperparameter optimization is not performed during a GPR iteration if a predefined number of SPs per 2M-cut is present in the training dataset, i.e., hyperparameters are only optimized for the first few GPR iterations and kept fixed afterward. This allows for yet another reduction in the overall computational cost since the hyperparameter optimization is more demanding for larger training set sizes.

III. COMPUTATIONAL DETAILS

To validate the new variant of GPR-ADGA, we chose a test set of molecules of growing size (from three to ten atoms). This set includes water, formaldehyde, ethylene, imidazole, and pyrimidine (for Lewis structures, see Fig. 1). The corresponding molecular structures were optimized in Orca[69] using the Hartree–Fock with three corrections (HF-3c) approach.[70] Subsequently, same electronic

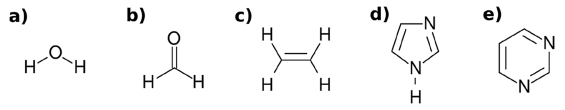

FIG. 1. Lewis structures of (a) water, (b) formaldehyde, (c) ethylene, (d) imidazole, and (e) pyrimidine used in this work.

structure method was applied to compute SPs and second derivatives of energy as presented in Sec. IV A.

Generation of normal coordinates and construction of PESs with the ADGA[27–29] and GPR-ADGA[60] were carried out in the MIDASCPP.[68] In all presented PES computations, the n-mode expansion was truncated at the second order. To find the optimal set of ADGA thresholds[27,29] in terms of accuracy of fundamental excitations and the computational cost, we carried out a small benchmark study using the above-mentioned test set of molecules and ten sets of ADGA criteria. The raw data from this benchmark are given in Sec. S2 in the supplementary material. The use of thresholds $\epsilon_{\text{rel}} = 1.0 \times 10^{-2}$, $\epsilon_{\text{abs}} = 1.0 \times 10^{-5}$, and $\epsilon_{\rho} = 1.0 \times 10^{-3}$ led to minimum numbers of SPs being computed while producing RMSDs in fundamental excitations below 1 cm^{-1}. For this reason, these thresholds were selected as optimal and applied in all GPR-ADGA and reference ADGA computations (unless stated otherwise). For the determination of initial grid boundaries, harmonic oscillator turning points with the quantum number $v = 2$ were applied. Four VSCF modals were included in the mean density given in Eq. (3). Reference ADGA computations were carried out using the dynamic extension of grid boundaries and gradient-guided basis set determination as described in Ref. 29. Same settings were found to be incompatible with GPR-ADGA and were disabled. Contrary to reference ADGA computations, GPR-ADGA was allowed to extend the potential grid boundaries already from the first iteration (by default, ADGA does so starting from iteration two and onward).

To provide the initial training set of points for GPR during the first GPR-ADGA iteration, the static grid approach[71] was used. The constructed initial static grids contained two SPs per each one-mode combination and the reference SP energy (i.e., 7, 13, 25, 43, and 49 SPs for water, formaldehyde, ethylene, imidazole, and pyrimidine, respectively). The corresponding static grid boundaries were set using the harmonic oscillator turning points defined by quantum number $v = 10$. Second derivatives of energy with respect to molecular displacements (Hessian) at the equilibrium molecular structure were computed and used to set up the mean function as described in Sec. II C. Furthermore, one-mode grid boundaries were added to the GPR-ADGA list of points, which are to be calculated, during the first GPR iteration. The training data, both features and labels, were standardized shifting by mean values and scaling by population standard deviations (for more details on standardization options, see Sec. S1.1 in the supplementary material). Shifting and scaling factors were computed using the whole training set of data (no special treatment of outliers). Note that calculations of new standardization factors and re-standardization of data were carried out each time the training set was extended. Similar to Ref. 60, the Bunch–Kaufmann decomposition was applied to solve the GPR linear system of equations. Due to several methodological changes to GPR-ADGA, old thresholds such as values for the selection criterion T_Ω and the noise term σ_N^2, found to be optimal in Ref. 60, were not applicable to the current setup. In order to find new optimal values of T_Ω and σ_N^2, we carried out GPR-ADGA computations for the water molecule, varying both the criteria independently by a factor of ten from 1.0×10^{-7} to 1.0×10^{-14} (see Sec. S3 in the supplementary material). The best trade-off between accuracy and performance was obtained when T_Ω was equal to σ_N^2. Furthermore, both the criteria being simultaneously varied from 1.0×10^{-8} to 1.0×10^{-11} provided a series of computations with growing computational cost and accuracy and consistently converging to the reference ADGA. In the following, a series of such GPR-ADGA computations are demonstrated and discussed in Sec. IV. The hyperparameters were optimized by minimizing the negative logarithm of the marginal likelihood[43] with the iRprop algorithm.[72] To find the optimal number of SPs per two-mode cut, which can be used as the threshold for stopping the hyperparameter optimization and subsequently re-using hyperparameters (see Sec. II E), we tested several values from 0 (no optimization) to 30. The results are demonstrated in Sec. S4 in the supplementary material. The value of 15 was found optimal and was applied in all presented GPR-ADGA calculations.

To obtain analytical representations of PESs, the linear fit using up to tenth order polynomials was applied. Note that the polynomial order used for fitting depends on the number of SPs and never exceeds it. The polynomial order increases with the number of SPs up to the specified value of ten. For high-mode potentials, an additional cutoff controls that the combination of polynomials does not exceed the 10-th order (for more details, see the supplementary material of Ref. 29). The fitted PESs were used to compute fundamental excitation energies with VSCF.[61–64] The accuracy of GPR-ADGA was assessed by computing maximal, minimal, and root mean square deviations in fundamental excitations with respect to the reference ADGA. Additionally, kernel density estimation (KDE) curves were constructed for the difference between ADGA and GPR-ADGA fundamental excitations, i.e., for $\Delta\omega = \omega(\text{ADGA}) - \omega(\text{GPR} - \text{ADGA})$ using the Seaborn[73] library. For this purpose, a Gaussian-type kernel was used.

For a further demonstration of the GPR-ADGA computational cost in Sec. IV B, the molecular structure of ethylene was re-optimized in the TURBOMOLE program package V7.0.[74] For this purpose, the Hartree–Fock (HF) method from the DSCF module[75] as well as explicitly correlated versions of the Møller–Plesset perturbation theory to second order of perturbation in conjunction with the resolution-of-the-identity approximation (RI-MP2-F12)[76–78] and coupled cluster with single, double, and perturbative triple correction [CCSD(F12*) (T)] from the CCSDF12 module[79] was employed. Note that CCSD(F12*) is also known as CCSD-F12c.[80,81] As the basis set, the correlation-consistent polarized valence double-ζ cc-pVDZ-F12[82] was used in all cases. In F12 calculations, the complementary auxiliary basis set (CABS) approach[83] was adopted. The corresponding CABS threshold was set to 1.0×10^{-8}. Additionally, the frozen core approximation excluding all orbitals with energies below −3 a.u. from the correlation treatment was used. Same electronic structure methods and settings were applied for subsequent ADGA and GPR-ADGA calculations of PESs. Calculations of the prior GPR mean function were carried out using numerical Hessians. Both ADGA and GPR-ADGA computations were

performed in parallel on a single Intel Xeon E5-2680 v2 @ 2.8 GHz/128 GB compute-node with 20 cores in total.

IV. RESULTS

In the following, we demonstrate the performance of GPR-ADGA for PES construction using a test set of five molecules of growing size and compare it with the standard ADGA in Sec. IV A. For this proof-of-principle study, we carried out inexpensive PES computations employing the HF-3c electronic structure method. Further in Sec. IV B, we analyze the computational cost of GPR-ADGA in detail by performing PES computations for the ethylene molecule using a series of different electronic structure methods.

A. Calculations of potential energy surfaces

Performance of GPR-ADGA compared to the reference ADGA is demonstrated in Fig. 2 (top) and (bottom). As can be seen and as was expected, the use of tighter thresholds $T_\Omega = \sigma_N^2$ results in generally smaller deviations in fundamental excitation energies and larger numbers of SPs being computed. Thus, with $T_\Omega = \sigma_N^2 = 1.0 \times 10^{-8}$, RMSD values are always below 3.2 cm^{-1}, whereas the reduction in the number of SPs compared to the ADGA is maximal and reaches about 57%–79% (for original values, see Secs. S2 and S4 in the supplementary material). For the tightest thresholds considered here, i.e., for $T_\Omega = \sigma_N^2 = 1.0 \times 10^{-11}$, all deviations are below about 0.7 cm^{-1}, while the reduction in SPs is the smallest and varies from about 4% to 68%. Computational savings are always higher for looser thresholds, whereas the dependence found for RMSD values is not always consistent. From Fig. 2 (bottom), it can also be seen that the gain in terms of the computational cost strongly and consistently depends on the molecular size. GPR-ADGA performs the best for smaller molecules such as water, where 68%–79% of SP calculations is avoided. Unfortunately, a lower reduction of about 4%–57% is found for pyrimidine. Note, however, that for this molecule GPR-ADGA still allows us to reach the very high accuracy of ~1 cm^{-1} while computing about half the number of SPs required for the ADGA (see results for $T_\Omega = \sigma_N^2 = 1.0 \times 10^{-9}$). This behavior of GPR-ADGA is probably related to the reference ADGA computations being more efficient for larger molecules. To demonstrate this, we can calculate approximate numbers of SPs computed with standard ADGA per 2M-cut potential. Taking the total numbers of SPs in PESs constructed with the reference ADGA (Sec. S2 in the supplementary material) and calculating the number of two-mode combinations as $\binom{M}{2} = M!/[2!(M-2)!]$, we can verify that about 179, 121, 63, 81, and 58 SPs per 2M-cut function are computed for water, formaldehyde, ethylene, imidazole, and pyrimidine,

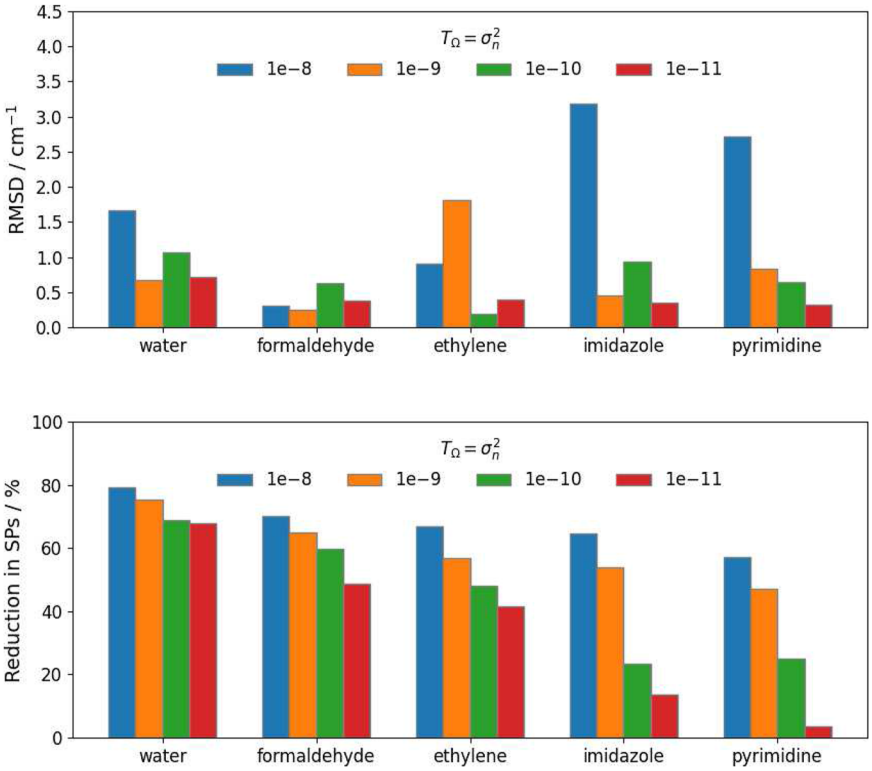

FIG. 2. Comparison of 2M PESs calculated for the chosen test set of molecules using GPR-ADGA and reference ADGA. RMSDs (in cm^{-1}) of VSCF fundamental frequencies are shown at the top. Reduction in the number of SPs (in %) is given at the bottom. The number of SPs calculated with the reference ADGA is equal to 536, 1812, 4145, 17 105, and 16 022 for water, formaldehyde, ethylene, imidazole, and pyrimidine, respectively. Results generated with GPR-ADGA criteria T_Ω and σ_N^2 being simultaneously varied in series 1.0×10^{-8}, 1.0×10^{-9}, 1.0×10^{-10}, and 1.0×10^{-11} are shown in blue, orange, green, and red colors, respectively.

respectively. Therefore, we could conclude that the ADGA computes smaller numbers of SPs per mode combination of larger molecules while still reaching the same level of accuracy. As the result, a rather modest additional reduction in the number of SPs could be achieved by means of GPR-ADGA.

It is also interesting to compare the performance of the ADGA and GPR-ADGA for the largest molecule from our test set, pyrimidine, while varying the convergence thresholds ϵ_{rel}, ϵ_{abs}, and ϵ_{ρ}. Results of this analysis are presented in Fig. 3, whereas original values are given in Sec. S5 in the supplementary material. As can be seen, RMSD values do not strongly depend on the ADGA convergence criteria ϵ_{rel}, ϵ_{abs}, and ϵ_{ρ} and change by at most ~1 cm^{-1}. Contrary to that, the reduction in the number of SPs changes from 4%–57% (for "normal") to 44%–76% (for "extra tight"). This trend can be explained by the fact that the number of SPs calculated with the reference ADGA is larger for tighter convergence thresholds, whereas GPR-ADGA shows a rather weak dependence on ϵ_{rel}, ϵ_{abs}, and ϵ_{ρ} (see Sec. S5 in the supplementary material). For example, the number of SPs computed with the reference ADGA grows by about a factor of two (from 16 022 to 33 836 SPs) when the criteria are changed from "normal" to "extra tight." At the same time, only a rather modest increase of about 18% is found for GPR-ADGA using $T_{\Omega} = \sigma_N^2 = 1.0 \times 10^{-11}$ (from 15 447 to 18 833 SPs). These results again support the previously discussed point that the relatively smaller gain of GPR-ADGA for large molecules when compared to the ADGA is related to a higher efficiency of the reference ADGA rather than drawbacks of the GPR-ADGA methodology.

To further analyze the accuracy and precision aspects of the GPR-ADGA method, we present KDE curves demonstrating distributions of errors in VSCF fundamental frequencies in Fig. 4. Results are generated for a set of deviations in fundamental excitations belonging to all five molecules from our test set. In the presented KDE curves, the position of the function's maximal value (i.e., the position of the peak) corresponds to the error $\Delta\omega$ with the largest probability. The closer the position of the peak to the zero at the $\Delta\omega$ axis the more accurate the results obtained. The precision is reflected in the KDE curve's broadness: A broader KDE curve corresponds to a larger error distribution and a lower precision, whereas, on the opposite, a narrower curve indicates at a higher precision. As can be seen from Fig. 4, both the accuracy and precision of GPR-ADGA are consistently improving for tighter thresholds $T_{\Omega} = \sigma_N^2$. Thus, a very broad KDE curve with a maximum at about 2.2 cm^{-1} is obtained for $T_{\Omega} = \sigma_N^2 = 1.0 \times 10^{-8}$. With $T_{\Omega} = \sigma_N^2 = 1.0 \times 10^{-9}$ and 1.0×10^{-10}, much more narrow KDE curves with the largest

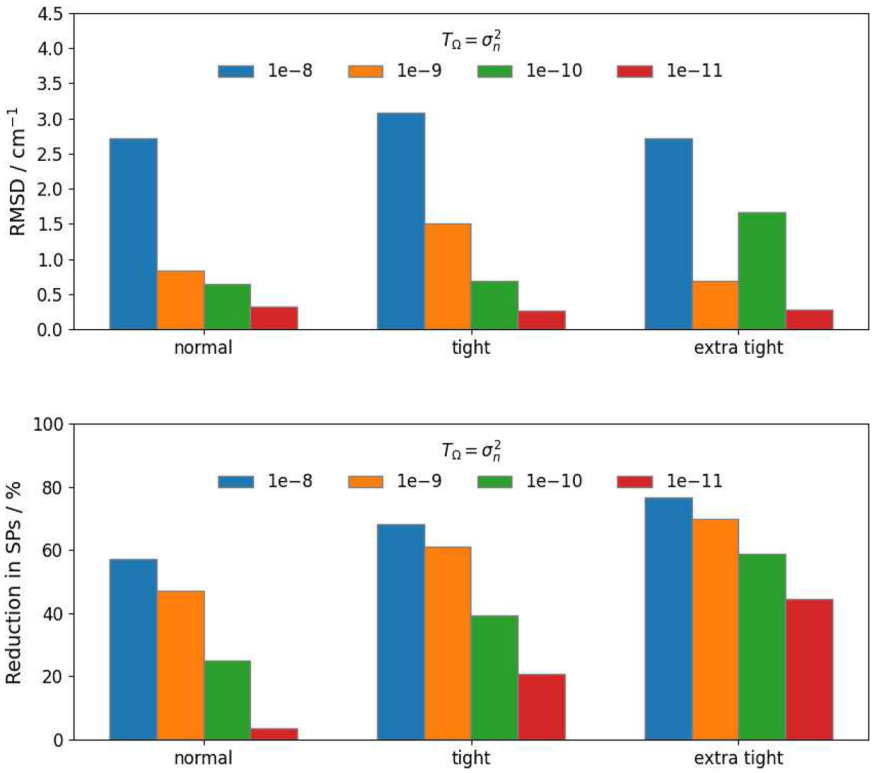

FIG. 3. Comparison of 2M PESs calculated for pyrimidine using GPR-ADGA and reference ADGA. RMSDs (in cm^{-1}) of VSCF fundamental frequencies are shown at the top. Reduction in the number of SPs (in %) is given at the bottom. The number of SPs calculated with the reference ADGA is equal to 16 022. The employed sets of ADGA convergence thresholds (for both, GPR-ADGA and reference ADGA computations) are denoted "normal" ($\epsilon_{\text{rel}} = 1.0 \times 10^{-2}$, $\epsilon_{\text{abs}} = 1.0 \times 10^{-5}$, $\epsilon_{\rho} = 1.0 \times 10^{-3}$), "tight" ($\epsilon_{\text{rel}} = 1.0 \times 10^{-3}$, $\epsilon_{\text{abs}} = 1.0 \times 10^{-5}$, $\epsilon_{\rho} = 1.0 \times 10^{-3}$), and "extra tight" ($\epsilon_{\text{rel}} = 1.0 \times 10^{-2}$, $\epsilon_{\text{abs}} = 1.0 \times 10^{-6}$, $\epsilon_{\rho} = 1.0 \times 10^{-3}$). Results generated with GPR-ADGA criteria T_{Ω} and σ_N^2 being simultaneously varied in series 1.0×10^{-8}, 1.0×10^{-9}, 1.0×10^{-10}, and 1.0×10^{-11} are shown in blue, orange, green, and red colors, respectively.

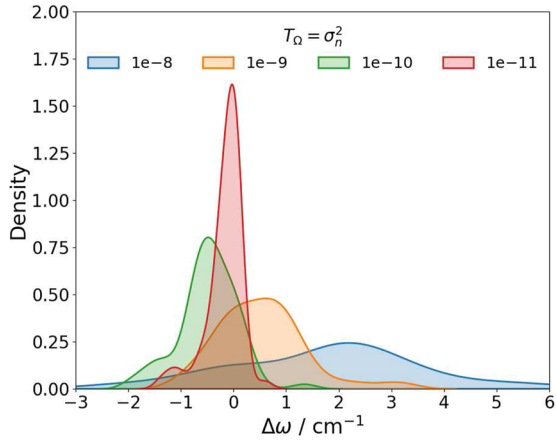

FIG. 4. KDE curves for deviations in VSCF fundamental frequencies $\Delta\omega = \omega(\text{ADGA}) - \omega(\text{GPR} - \text{ADGA})$ (in cm^{-1}) calculated for all five molecules from the test set. Results generated with GPR-ADGA criteria T_Ω and σ_N^2 being simultaneously varied in series 1.0×10^{-8}, 1.0×10^{-9}, 1.0×10^{-10}, and 1.0×10^{-11} are shown in blue, orange, green, and red colors, respectively.

probability errors of only about 0.6 and −0.5 cm^{-1}, respectively, are obtained. Finally, the best results in terms of accuracy and precision are found for $T_\Omega = \sigma_N^2 = 1.0 \times 10^{-11}$.

B. Computational cost

As was demonstrated in Sec. IV A, GPR-ADGA allows us to reduce the number of computed SPs by up to 68% (compared to the reference ADGA) while keeping the RMSD in fundamental excitation energies below 0.7 cm^{-1}. If a lower accuracy of about 3 cm^{-1} is considered sufficient, even larger reduction in the number of SPs of up to about 80% could be reached. Although SP calculations are embarrassingly parallelizable and could be performed on many dedicated computational nodes, they often remain the main computational bottleneck of the PES construction. Therefore, the reduction by 80% reported in this work is rather considerable. One might argue, however, that a smaller number of SPs being computed does not necessarily mean a faster PES construction procedure as GPR-ADGA has a much larger computational overhead than the ADGA. This argumentation is correct, and the GPR-ADGA method could indeed be more expensive than the standard ADGA. Thus, GPR-ADGA executes several ADGA computations until their full convergence using GPR as a provider of new SPs. This means that VSCF calculations and fitting of constructed PESs are repeated multiple times on each GPR iteration. The number of these iterations could be considerable and reach up to 40 for very tight convergence criteria $T_\Omega = \sigma_N^2 = 1.0 \times 10^{-11}$. Note, however, that VSCF calculations within the ADGA are performed for one-dimensional mode cuts and therefore formally scale linearly with the number of modes M. Furthermore, additional computational cost is introduced with the use of the GPR algorithm. In this regard, the most expensive steps are the inversion of the covariance matrix, i.e., calculations of the term $(\mathbf{K} + \sigma_N^2\mathbf{I})^{-1}$, and the solution of an equivalent system of linear equations to find weights $\boldsymbol{\omega} = (\mathbf{K} + \sigma_N^2\mathbf{I})^{-1}(\mathbf{v} - \mathbf{m})$ as seen from Eqs. (8) and (9). Both steps scale as $\mathcal{O}(N^3)$ with the number of training points N. For computing predictions $\boldsymbol{\mu}$ and uncertainties $(\boldsymbol{\Sigma})_{ii}$, it is sufficient to calculate weights $\boldsymbol{\omega}$ only once per GPR iteration. However, the hyperparameter optimization procedure updates hyperparameters and therefore requires re-computing the inverse on each optimization cycle. Therefore, the computational cost of hyperparameter optimization is dominated by the inversion of $(\mathbf{K} + \sigma_N^2\mathbf{I})$ and scales as $\mathcal{O}(N^3)$. Subsequent calculations of the gradient of the marginal likelihood with respect to the hyperparameters would scale only as $\mathcal{O}(N^2)$ per hyperparameter[43] as the inversion of $(\mathbf{K} + \sigma_N^2\mathbf{I})$ could be re-used for the same set of hyperparameters. However, as the hyperparameter optimization process continuously updates the set of hyperparameters, the inverse of $(\mathbf{K} + \sigma_N^2\mathbf{I})$ has to be re-calculated at each step. Since the training set size N grows from iteration to iteration, the overhead of using GPR increases as well. Although in our setup the hyperparameter optimization is not carried out for latter GPR iterations featuring the largest training sets, as was described in Sec. II E, it still affects the total computational cost of GPR-ADGA. Finally, the matrix–matrix multiplication $(\mathbf{K} + \sigma_N^2\mathbf{I})^{-1}\mathbf{K}^*$ from Eq. (9) scales as $\mathcal{O}(N^2N^*)$, where N^* is often much larger than N in practical applications of GPR-ADGA. Performing GPR predictions in batches with an OpenMP parallelization over the number of these batches, as already mentioned in Sec. II E, offers a way to mitigate this step. Note that the second matrix–matrix multiplication in Eq. (9) involving $(\mathbf{K}^*)^{\mathrm{T}}$ formally scales as $\mathcal{O}(NN^{*2})$. However, because only diagonal elements of the posterior covariance matrix $\boldsymbol{\Sigma}$ are required, the actual computational scaling is reduced to $\mathcal{O}(NN^*)$.

Despite the described above overhead, GPR-ADGA could still considerably reduce the cost of the overall PES construction procedure when expensive electronic structure methods are used. To demonstrate this, we calculated wall times of PES generation using a series of electronic structure methods with increasing computational costs: HF, RI-MP2-F12, and CCSD(F12*) (T). The results are presented in Fig. 5, whereas the original values of the wall time as well as central processing unit (CPU) time are provided in Sec. S6 in the supplementary material. As can be seen from Fig. 5 (top), with HF being used, GPR-ADGA reduces the time spent on calculating SPs by about a factor of two (i.e., by ~44%–56%) compared to the reference ADGA. This, however, does not lead to a decreased total computational cost due to a considerable overhead of running multiple VSCF calculations and hyperparameter optimizations. These two types of computations amount in about 18%–24% and 20%–22%, respectively, of the GPR-ADGA wall time. As the result, depending on the thresholds being used, GPR-ADGA is comparable or more expensive than the standard ADGA. The situation changes when RI-MP2-F12 is used for calculating SPs as seen in Fig. 5 (middle). The reduction in the SP computational cost remains about the same, whereas 15%–20% and 10%–15% of the GPR-ADGA total computational time are spent on VSCF and hyperparameter optimization. This leads to GPR-ADGA being 22%–32% faster than the ADGA. Finally, for the most expensive electronic structure method CCSD(F12*) (T), from those applied in this work, the cost of SPs becomes dominant in GPR-ADGA with all other computational steps amounting in only about 5%–9% of the total wall time. As the result, GPR-ADGA computation employing CCSD(F12*) (T) is about twice as fast as the reference ADGA using the same electronic structure method. Due to a very steep increase in the computational cost of PES construction with the number of atoms, one can expect

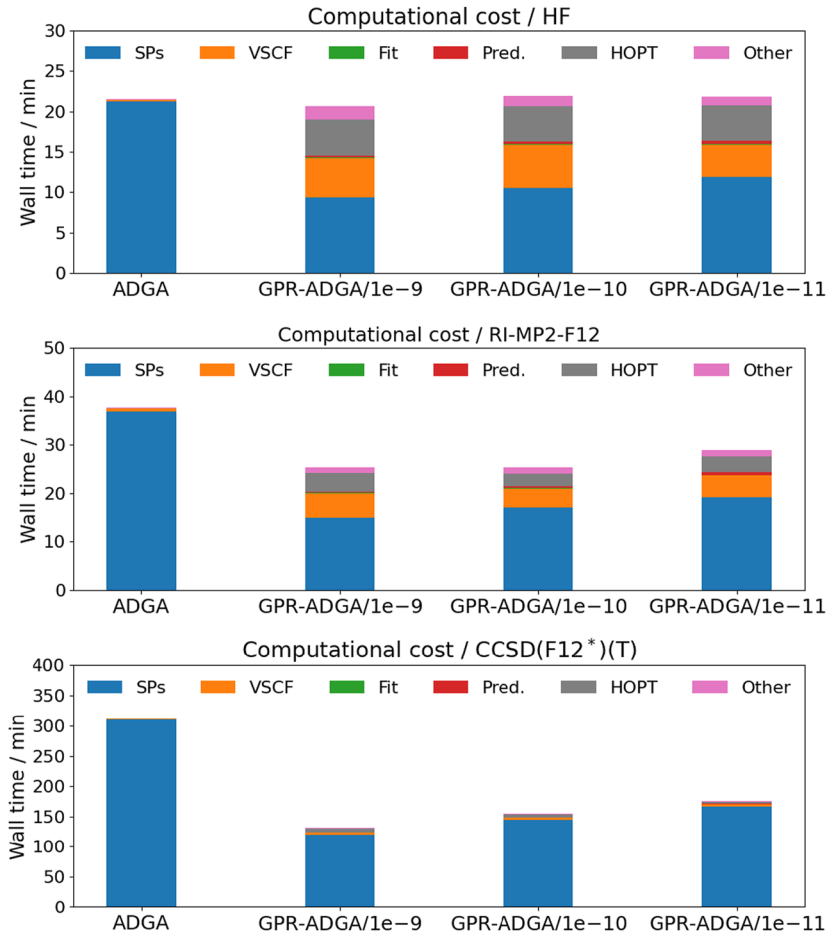

FIG. 5. Wall times required for GPR-ADGA and reference ADGA computations of ethylene PESs. Results are provided for the HF (top), RI-MP2-F12 (middle), and CCSD(F12*) (T) (bottom) methods. GPR-ADGA criteria T_Ω and σ_N^2 are simultaneously varied in series 1.0×10^{-9}, 1.0×10^{-10}, and 1.0×10^{-11}. Times spent on SPs, VSCF, polynomial fit of the PES, energy predictions, hyperparameter optimization (denoted as HOPT), and other operations are given in blue, orange, green, red, gray, and pink colors, respectively.

that the GPR-ADGA overhead becomes negligibly small (relative to the cost of SPs) for larger molecules and more expensive electronic structure methods.

In full analogy to the results presented in Sec. IV A, we also assessed the accuracy of the constructed GPR-ADGA 2M PESs of ethylene by computing RMSDs in fundamental excitations and using ADGA as the reference. To that end, vibrational coupled cluster with up to two-mode excitations (VCC[2])[84–88] was applied. The results are demonstrated in Sec. S7 in the supplementary material and show very similar trends to those from Sec. IV A. Thus, the RMSD consistently decreases for tighter convergence thresholds T_Ω and σ_N^2 and reaches values below 1 cm^{-1} for $T_\Omega = \sigma_N^2 = 1.0 \times 10^{-11}$.

V. CONCLUSIONS

In this work, we presented a new and improved program implementation of the GPR-ADGA method[60] in MidasCpp.[68] A number of technical and methodological extensions were introduced, enabling GPR-ADGA calculations of PESs for larger molecules than those accessible previously while maintaining a very high accuracy in fundamental excitation energies and a considerable reduction in the number of SPs compared to the reference ADGA. The performance of GPR-ADGA was assessed on a test set of five molecules of increasing size from three to ten atoms. Convergence thresholds were introduced, allowing one to reach a desired balance between the accuracy and the efficiency of the PES construction. Thus, if the RMSD in fundamental excitation energies of about 3 cm^{-1} is considered sufficient, calculations of up to 80% of SPs could be avoided by using GPR-ADGA. A higher accuracy of about 0.7 cm^{-1} or better could be reached with tighter GPR-ADGA convergence thresholds while reducing the number of SPs by up to 68%. The reduction in the number of SPs was found to be smaller for larger molecules. This, however, was explained by a higher efficiency of the reference ADGA for large molecular systems rather than by drawbacks of the GPR-ADGA methodology.

Additionally, we analyzed the computational cost of the PES construction with GPR-ADGA by carrying out calculations of ethylene with a series of electronic structure methods such as HF, RI-MP2-F12, and CCSD(F12*) (T) and presenting total wall times. We showed that due to an increased overhead of GPR-ADGA compared to the reference ADGA and despite the number of SPs being considerably reduced, no computational gain could be reached while using HF. For more accurate and expensive RI-MP2-F12, GPR-ADGA is by 22%–32% faster than the ADGA employing the same electronic structure method. For CCSD(F12*) (T), the overhead of running GPR-ADGA becomes negligibly small compared to the wall time spent on SPs. As the result, the computational gain is about the same as the reduction in the number of SPs and reaches 44%–58%. Therefore, we conclude that it is the most advantageous to use GPR-ADGA in conjunction with very accurate and costly electronic structure methods.

Our results demonstrate that GPR-ADGA could be used for highly accurate and cost-efficient PES calculations and encourage applications to various molecular systems for subsequent reliable vibrational spectra simulations. Although all PESs calculated in this work included only up to second-order mode-coupling terms, GPR-ADGA including the current implementation can easily be used for higher-level mode couplings. In fact, computations of three-mode potentials were already presented in the previous work in Ref. 60. The approach could further be improved by combining it with double incremental PES expansions[8,30] and flexible adaptation of local coordinates of nuclei.[89] This could allow to incorporate fragmentation ideas into GPR-ADGA and handle even larger molecular systems. In this case, the computational gain could be reached by using GPR-ADGA for calculating individual subsystem potentials and/or for enabling learning between subsystems of the total molecular system. The work in both of these directions is currently in progress.

SUPPLEMENTARY MATERIAL

See the supplementary material for (S1) additional theory aspects, (S2) reference ADGA-2M computations, (S3) optimal GPR-ADGA thresholds, (S4) parameters controlling hyperparameter optimization, (S5) influence of ADGA thresholds on GPR-ADGA computations of pyrimidine, (S6) GPR-ADGA computational cost, and (S7) vibrational couple cluster computations.

ACKNOWLEDGMENTS

D.G.A. acknowledges funding from the Rising Star Fellowship of the Department of Biology, Chemistry, Pharmacy of Freie Universität Berlin. O.C. acknowledges support from the Independent Research Fund Denmark through Grant No. 1026-00122B. We gratefully acknowledge funding from the Novo Nordisk Foundation through the Exploratory Interdisciplinary Synergy Programme Project No. NNF19OC0057790. The authors thank Professor Dr. Guntram Rauhut for helpful discussions. Computations were performed at the Centre for Scientific Computing Aarhus (CSCAA).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Denis G. Artiukhin: Conceptualization (equal); Data curation (lead); Formal analysis (lead); Funding acquisition (equal); Investigation (lead); Methodology (lead); Software (equal); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **Ian H. Godtliebsen**: Conceptualization (equal); Data curation (equal); Investigation (equal); Methodology (equal); Software (equal); Writing – review & editing (equal). **Gunnar Schmitz**: Conceptualization (supporting); Methodology (supporting); Software (supporting); Writing – review & editing (equal). **Ove Christiansen**: Conceptualization (lead); Data curation (equal); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (lead); Resources (lead); Software (supporting); Supervision (lead); Writing – review & editing (equal).

DATA AVAILABILITY

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

REFERENCES

[1] J. O. Jung and R. B. Gerber, "Vibrational wave functions and spectroscopy of $(H_2O)_n$, n=2,3,4,5: Vibrational self-consistent field with correlation corrections," J. Chem. Phys. **105**, 10332 (1996).
[2] S. Carter, S. J. Culik, and J. M. Bowman, "Vibrational self-consistent field method for many-mode systems: A new approach and application to the vibrations of CO adsorbed on Cu(100)," J. Chem. Phys. **107**, 10458 (1997).
[3] J. M. Bowman, S. Carter, and X. Huang, "MULTIMODE: A code to calculate rovibrational energies of polyatomic molecules," Int. Rev. Phys. Chem. **22**, 533–549 (2003).
[4] G. Rauhut, "Efficient calculation of potential energy surfaces for the generation of vibrational wave functions," J. Chem. Phys. **121**, 9313–9322 (2004).
[5] J. Kongsted and O. Christiansen, "Automatic generation of force fields and property surfaces for use in variational vibrational calculations of anharmonic vibrational energies and zero-point vibrational averaged properties," J. Chem. Phys. **125**, 124108 (2006).
[6] H.-D. Meyer, "Studying molecular quantum dynamics with the multiconfiguration time-dependent Hartree method," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 351–374 (2012).
[7] H. Rabitz and Ö. F. Aliş, "General foundations of high-dimensional model representations," J. Math. Chem. **25**, 197–233 (1999).
[8] C. König and O. Christiansen, "Linear-scaling generation of potential energy surfaces using a double incremental expansion," J. Chem. Phys. **145**, 064105 (2016).
[9] D. Madsen, O. Christiansen, and C. König, "Anharmonic vibrational spectra from double incremental potential energy and dipole surfaces," Phys. Chem. Chem. Phys. **20**, 3445–3456 (2018).
[10] K. Yagi, S. Hirata, and K. Hirao, "Multiresolution potential energy surfaces for vibrational state calculations," Theor. Chem. Acc. **118**, 681–691 (2007).
[11] G. Rauhut and T. Hrenar, "A combined variational and perturbational study on the vibrational spectrum of P_2F_4," Chem. Phys. **346**, 160–166 (2008).
[12] G. Rauhut and B. Hartke, "Modeling of high-order many-mode terms in the expansion of multidimensional potential energy surfaces: Application to vibrational spectra," J. Chem. Phys. **131**(1), 014108 (2009).

[13]M. Sparta, I.-M. Høyvik, D. Toffoli, and O. Christiansen, "Potential energy surfaces for vibrational structure calculations from a multiresolution adaptive density-guided approach: Implementation and test calculations," J. Phys. Chem. A **113**, 8712–8723 (2009).
[14]M. Sparta, M. B. Hansen, E. Matito, D. Toffoli, and O. Christiansen, "Using electronic energy derivative information in automated potential energy surface construction for vibrational calculations," J. Chem. Theory Comput. **6**, 3162–3175 (2010).
[15]P. Meier, G. Bellchambers, J. Klepp, F. R. Manby, and G. Rauhut, "Modeling of high-order terms in potential energy surface expansions using the reference-geometry Harris–Foulkes method," Phys. Chem. Chem. Phys. **15**, 10233–10240 (2013).
[16]G. Schmitz, D. G. Artiukhin, and O. Christiansen, "Approximate high mode coupling potentials using Gaussian process regression and adaptive density guided sampling," J. Chem. Phys. **150**, 131102 (2019).
[17]D. M. Benoit, "Fast vibrational self-consistent field calculations through a reduced mode–mode coupling scheme," J. Chem. Phys. **120**, 562–573 (2004).
[18]D. M. Benoit, "Efficient correlation-corrected vibrational self-consistent field computation of OH-stretch frequencies using a low-scaling algorithm," J. Chem. Phys. **125**, 244110 (2006).
[19]L. Pele and R. B. Gerber, "On the number of significant mode–mode anharmonic couplings in vibrational calculations: Correlation-corrected vibrational self-consistent field treatment of di-, tri-, and tetrapeptides," J. Chem. Phys. **128**, 165105 (2008).
[20]D. M. Benoit, "Fast vibrational calculation of anharmonic OH-stretch frequencies for two low-energy noradrenaline conformers," J. Chem. Phys. **129**, 234304 (2008).
[21]P. Seidler, T. Kaga, K. Yagi, O. Christiansen, and K. Hirao, "On the coupling strength in potential energy surfaces for vibrational calculations," Chem. Phys. Lett. **483**, 138–142 (2009).
[22]X. Cheng and R. P. Steele, "Efficient anharmonic vibrational spectroscopy for large molecules using local-mode coordinates," J. Chem. Phys. **141**, 104105 (2014).
[23]E. L. Klinting, O. Christiansen, and C. König, "Toward accurate theoretical vibrational spectra: A case study for maleimide," J. Phys. Chem. A **124**, 2616 (2020).
[24]K. Mackeprang, V. Hänninen, L. Halonen, and H. G. Kjaergaard, "The effect of large amplitude motions on the vibrational intensities in hydrogen bonded complexes," J. Chem. Phys. **142**, 094304 (2015).
[25]K. Yagi, K. Yamada, C. Kobayashi, and Y. Sugita, "Anharmonic vibrational analysis of biomolecules and solvated molecules using hybrid QM/MM computations," J. Chem. Theory Comput. **15**, 1924–1938 (2019).
[26]B. Ziegler and G. Rauhut, "Rigorous use of symmetry within the construction of multidimensional potential energy surfaces," J. Chem. Phys. **149**(16), 164110 (2018).
[27]M. Sparta, D. Toffoli, and O. Christiansen, "An adaptive density-guided approach for the generation of potential energy surfaces of polyatomic molecules," Theor. Chem. Acc. **123**, 413–429 (2009).
[28]D. Toffoli, M. Sparta, and O. Christiansen, "Accurate multimode vibrational calculations using a B-spline basis: Theory, tests and application to dioxirane and diazirinone," Mol. Phys. **109**, 673–685 (2011).
[29]E. L. Klinting, B. Thomsen, I. H. Godtliebsen, and O. Christiansen, "Employing general fit-bases for construction of potential energy surfaces with an adaptive density-guided approach," J. Chem. Phys. **148**, 064113 (2018).
[30]D. G. Artiukhin, E. L. Klinting, C. König, and O. Christiansen, "Adaptive density-guided approach to double incremental potential energy surface construction," J. Chem. Phys. **152**, 194105 (2020).
[31]S. Manzhos and T. Carrington, "Using neural networks to represent potential surfaces as sums of products," J. Chem. Phys. **125**(19), 194105 (2006).
[32]S. Manzhos and T. Carrington, "Using neural networks, optimized coordinates, and high-dimensional model representations to obtain a vinyl bromide potential surface," J. Chem. Phys. **129**(22), 224104 (2008).
[33]A. Brown and E. Pradhan, "Fitting potential energy surfaces to sum-of-products form with neural networks using exponential neurons," J. Theor. Comput. Chem. **16**(05), 1730001 (2017).
[34]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**, 22272–22281 (2017).
[35]E. Pradhan and A. Brown, "Vibrational energies for HFCO using a neural network sum of exponentials potential energy surface," J. Chem. Phys. **144**(17), 174305 (2016).
[36]E. Pradhan and A. Brown, "Neural network exponential fitting of a potential energy surface with multiple minima: Application to HFCO," J. Mol. Spectrosc. **330**, 158–164 (2016).
[37]A. M. Cooper, P. P. Hallmen, and J. Kästner, "Potential energy surface interpolation with neural networks for instanton rate calculations," J. Chem. Phys. **148**, 094106 (2018).
[38]J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," Phys. Rev. Lett. **98**, 146401 (2007).
[39]J. Behler, "Neural network potential-energy surfaces in chemistry: A tool for large-scale simulations," Phys. Chem. Chem. Phys. **13**, 17930–17955 (2011).
[40]J. Behler, "First principles neural network potentials for reactive simulations of large molecular and condensed systems," Angew. Chem., Int. Ed. **56**(42), 12828–12840 (2017).
[41]J. Behler, "Perspective: Machine learning potentials for atomistic simulations," J. Chem. Phys. **145**(17), 170901 (2016).
[42]T. W. Ko, J. A. Finkler, S. Goedecker, and J. Behler, "A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer," Nat. Commun. **12**(1), 398 (2021).
[43]C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, *Adaptive Computation and Machine Learning* (MIT Press, 2005).
[44]C. Schran, K. Brezina, and O. Marsalek, "Committee neural network potentials control generalization errors and enable active learning," J. Chem. Phys. **153**(10), 104105 (2020).
[45]R. Jinnouchi, F. Karsai, and G. Kresse, "On-the-fly machine learning force field generation: Application to melting points," Phys. Rev. B **100**, 014105 (2019).
[46]R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, and M. Bokdam, "Phase transitions of hybrid perovskites simulated by machine-learning force fields trained on the fly with Bayesian inference," Phys. Rev. Lett. **122**, 225701 (2019).
[47]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).
[48]A. P. Bartók and G. Csányi, "Gaussian approximation potentials: A brief tutorial introduction," Int. J. Quantum Chem. **115**(16), 1051–1057 (2015).
[49]L. Mones, N. Bernstein, and G. Csányi, "Exploration, sampling, and reconstruction of free energy surfaces with Gaussian process regression," J. Chem. Theory Comput. **12**(10), 5100–5110 (2016).
[50]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**(22), 224001 (2016).
[51]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**(13), 2552–2557 (2017).
[52]N. Yang, S. Hill, S. Manzhos, and T. Carrington, "A local Gaussian processes method for fitting potential surfaces that obviates the need to invert large matrices," J. Mol. Spectrosc. **393**, 111774 (2023).
[53]M. Schneider, D. Born, J. Kästner, and G. Rauhut, "Positioning of grid points for spanning potential energy surfaces—How much effort is really needed?," J. Chem. Phys. **158**, 144118 (2023).
[54]J. P. Alborzpour, D. P. Tew, and S. Habershon, "Efficient and accurate evaluation of potential energy matrix elements for quantum dynamics using Gaussian process regression," J. Chem. Phys. **145**(17), 174112 (2016).
[55]G. W. Richings and S. Habershon, "MCTDH on-the-fly: Efficient grid-based quantum dynamics without pre-computed potential energy surfaces," J. Chem. Phys. **148**(13), 134116 (2018).
[56]A. Denzel and J. Kästner, "Gaussian process regression for geometry optimization," J. Chem. Phys. **148**, 094114 (2018).
[57]G. Schmitz and O. Christiansen, "Gaussian process regression to accelerate geometry optimizations relying on numerical differentiation," J. Chem. Phys. **148**, 241704 (2018).

[58]D. Born and J. Kästner, "Geometry optimization in internal coordinates based on Gaussian process regression: Comparison of two approaches," J. Chem. Theory Comput. **17**, 5955–5967 (2021).
[59]V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, "Gaussian process regression for materials and molecules," Chem. Rev. **121**(16), 10073–10141 (2021).
[60]G. Schmitz, E. L. Klinting, and O. Christiansen, "A Gaussian process regression adaptive density guided approach for potential energy surface construction," J. Chem. Phys. **153**, 064105 (2020).
[61]J. M. Bowman, "Self-consistent field energies and wavefunctions for coupled oscillators," J. Chem. Phys. **68**, 608–610 (1978).
[62]R. B. Gerber and M. A. Ratner, "A semiclassical self-consistent field (SC SCF) approximation for eigenvalues of coupled-vibration systems," Chem. Phys. Lett. **68**, 195–198 (1979).
[63]O. Christiansen, "A second quantization formulation of multimode dynamics," J. Chem. Phys. **120**, 2140–2148 (2004).
[64]M. B. Hansen, M. Sparta, P. Seidler, D. Toffoli, and O. Christiansen, "New formulation and implementation of vibrational self-consistent field theory," J. Chem. Theory Comput. **6**, 235–248 (2010).
[65]K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, 2012).
[66]E. L. Klinting, D. Lauvergnat, and O. Christiansen, "Vibrational coupled cluster computations in polyspherical coordinates with the exact analytical kinetic energy operator," J. Chem. Theory Comput. **16**(7), 4505–4520 (2020).
[67]R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, "Big data meets quantum chemistry approximations: The Δ-machine learning approach," J. Chem. Theory Comput. **11**(5), 2087–2096 (2015).
[68]O. Christiansen, D. G. Artiukhin, I. H. Godtliebsen, E. M. Gras, W. Győrffy, M. B. Hansen, M. B. Hansen, E. L. Klinting, J. Kongsted, C. König, D. Madsen, N. K. Madsen, K. Monrad, G. Schmitz, P. Seidler, K. Sneskov, M. Sparta, B. Thomsen, D. Toffoli, and A. Zoccante, MidasCpp, version 2022.10.0, https://midascpp.gitlab.io/.
[69]F. Neese, "The ORCA program system," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 73–78 (2012).
[70]R. Sure and S. Grimme, "Corrected small basis set Hartree–Fock method for large systems," J. Comput. Chem. **34**, 1672–1685 (2013).
[71]D. Toffoli, J. Kongsted, and O. Christiansen, "Automatic generation of potential energy and property surfaces of polyatomic molecules in normal coordinates," J. Chem. Phys. **127**, 204106 (2007).
[72]C. Igel and M. Hüsken, "Improving the Rprop learning algorithm," in *Proceedings of the Second International Symposium on Neural Computation* (ICSC Academic Press, 2000), pp. 115–121.
[73]M. L. Waskom, "seaborn: Statistical data visualization," J. Open Source Software **6**(60), 3021 (2021).
[74]TURBOMOLE V7.0 2015, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.
[75]R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, "Electronic structure calculations on workstation computers: The program system turbomole," Chem. Phys. Lett. **162**(3), 165–169 (1989).
[76]J. L. Whitten, "Coulombic potential energy integrals and approximations," J. Chem. Phys. **58**(10), 4496–4501 (1973).
[77]B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, "On some approximations in applications of $X\alpha$ theory," J. Chem. Phys. **71**(8), 3396–3402 (1979).
[78]O. Vahtras, J. Almlöf, and M. W. Feyereisen, "Integral approximations for LCAO-SCF calculations," Chem. Phys. Lett. **213**, 514–518 (1993).
[79]C. Hättig and F. Weigend, "CC2 excitation energy calculations on large molecules using the resolution of the identity approximation," J. Chem. Phys. **113**, 5154 (2000).
[80]H.-J. Werner, G. Knizia, and F. R. Manby, "Explicitly correlated coupled cluster methods with pair-specific geminals," Mol. Phys. **109**(3), 407–417 (2011).
[81]H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, "Molpro: A general-purpose quantum chemistry program package," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 242–253 (2012).
[82]K. A. Peterson, T. B. Adler, and H.-J. Werner, "Systematically convergent basis sets for explicitly correlated wavefunctions: The atoms H, He, B–Ne, and Al–Ar," J. Chem. Phys. **128**(8), 084102 (2008).
[83]E. F. Valeev, "Improving on the resolution of the identity in linear R12 ab initio theories," Chem. Phys. Lett. **395**, 190–195 (2004).
[84]O. Christiansen, "Vibrational coupled cluster theory," J. Chem. Phys. **120**, 2149 (2004).
[85]O. Christiansen, "Vibrational structure theory: New vibrational wave function methods for calculation of anharmonic vibrational energies and vibrational contributions to molecular properties," Phys. Chem. Chem. Phys. **9**, 2942–2953 (2007).
[86]P. Seidler and O. Christiansen, "Vibrational excitation energies from vibrational coupled cluster response theory," J. Chem. Phys. **126**, 204101 (2007).
[87]P. Seidler, M. Sparta, and O. Christiansen, "Vibrational coupled cluster response theory: A general implementation," J. Chem. Phys. **134**, 054119 (2011).
[88]N. K. Madsen, I. H. Godtliebsen, S. A. Losilla, and O. Christiansen, "Tensor-decomposed vibrational coupled-cluster theory: Enabling large-scale, highly accurate vibrational-structure calculations," J. Chem. Phys. **148**, 024103 (2018).
[89]C. König, M. B. Hansen, I. H. Godtliebsen, and O. Christiansen, "FALCON: A method for flexible adaptation of local coordinates of nuclei," J. Chem. Phys. **144**, 074108 (2016).