Machine learning optimization of the collocation point set for solving the Kohn-Sham equation

Jonas Ku, Aditya Kamath, Tucker Carrington Jr, Sergei Manzhos

a Department of Mechanical Engineering, National University of Singapore, Block EA #07-08, 9 Engineering Drive 1, Singapore 117576 Singapore.
b Chemistry Department, Queen's University, Kingston, Ontario K7L 3N6, Canada
c Centre Énergie Matériaux Télécommunications, Institut National de la Recherche Scientifique, 1650, boulevard Lionel-Boulet, Varennes QC J3X1S2 Canada

Abstract
The rectangular collocation approach makes it possible to solve the Schrödinger equation with basis functions that do not have amplitude in all regions in which wavefunctions have significant amplitude. Collocation points can be restricted to a small region of space. As no integrals are computed, there are no problems due to discontinuities in the potential, and there is no need to use integrable basis functions. In this paper, we show, for the Kohn-Sham equation, that machine learning can be used to drastically reduce the size of the collocation point set. This is demonstrated by solving the Kohn-Sham equations for CO and H2O. We solve the Kohn-Sham equation on a given effective potential which is a critical part of all DFT calculations, and monitor orbital energies and orbital shapes. We use a combination of Gaussian process regression and a genetic algorithm to reduce the collocation point set size by more than an order of magnitude (from about 51,000 points to 2,000 points) while retaining mHartree accuracy.

1 Introduction
In this paper, we propose new ideas for choosing the points used with rectangular collocation to solve the electronic Schrödinger equation using the Kohn-Sham equation as a test case. Rectangular collocation has been used to solve both the vibrational and the electronic Schrödinger equations.1-13 We solve the time-independent Schrödinger equation (SE)

1 Author to whom correspondence should be addressed. E-mail: sergei.manzhos@emt.inrs.ca; Tel: +1 514 2286900.
\[ \hat{H} \psi(x) \equiv (\hat{T} + \hat{V}) \psi(x) = E \psi(x), \] (1)

where \( \hat{H} \) is the Hamiltonian operator, which is the sum of a kinetic energy operator (KEO) \( \hat{T} \) and the potential energy operator \( \hat{V} \); \( \psi(x) \) is a wavefunction; and \( E \) is an energy. It needs to be solved when computing electronic energy levels, as well as when computing (ro)vibrational spectra.\textsuperscript{14-16} In the Born-Oppenheimer approximation assumed here, one can separately solve the electronic and nuclear Schrödinger equations. In the electronic case, \( x \) is a vector of all the electronic coordinates (of which there are \( 3N_e \), where \( N_e \) is the number of electrons considered in the model) and the potential term \( \hat{V} = V(x) \) is due to Coulombic interactions among electrons and nuclei. Considering \( \psi(x) \) in Eq. 1 as an explicit function of the \( 3N_e \) (Cartesian) coordinates gives rise to various wavefunction-based \textit{ab initio} approaches.\textsuperscript{17-21} In this paper we solve a specific electronic Schrödinger equation: the single-electron Kohn Sham equation which is the basis of Kohn-Sham Density Functional Theory (DFT).\textsuperscript{22,23} In the KS case, Eq. 1 takes the form,

\[ \frac{1}{2} \Delta \psi(x) + V_{\text{eff}}(x) \psi(x) = E \psi(x), \] (2)

where

\[ V_{\text{eff}}(x) = V_{\text{nuc}}(x) + \int \frac{\rho(x')}{|x-x'|} dx' + V_{\text{XC}}(x) \]

\[ \rho(x) = \sum_{i=1}^{N_e} |\psi_i(x)|^2, \quad \int_{\infty} |\psi_i(x)|^2 dx = 1, \] (3)

where \( \rho(x) \) is the electron density, \( V_{\text{eff}}(x) \) is an effective potential, \( V_{\text{nuc}}(x) \) is the exchange-correlation potential, \( V_{\text{nuc}}(x) \) is the Coulombic potential due to the nuclei, \( \psi_i(x) \) are the orbitals, and \( i \) indexes different solutions of Eq. 2. Here, \( x \) is a vector of three Cartesian coordinates. We use atomic units (\textit{a.u.}). Because Kohn-Sham DFT is computationally much less costly than wavefunction-based methods, it dominates \textit{ab initio} materials modeling and much of molecular modeling. It is Eq. 2 that is solved in this article. We note that both \( V(x) \) in wavefunction-based approaches and \( V_{\text{eff}}(x) \) in DFT are singular due to the singular nature of the (Coulombic) ionic potential. Pseudopotentials can be used to remove the singularity at the cost of accuracy.\textsuperscript{24}
The time-independent Schrödinger equation one must solve to compute a ro-vibrational spectrum of an isolated molecule has the form of Eq. 1, but the potential is obtained from the Born-Oppenheimer approximation, and the KEO, written in terms of coordinates that specify the shape and orientation of the molecule, is complicated. Cartesian coordinates can be used when translation and rotation are frustrated.

As discussed in section 2, collocation is useful for solving the electronic Schrödinger equation because it obviates the need to choose quadrature points with which integrals can be approximated. In particular, because there are no integrals, singularities and slow potential decay in the asymptotic region pose no problems. Methods related to collocation have been previously used to solve the electronic Schrödinger equation. In some cases,\textsuperscript{12,13,25-27} analytic integrals are used for some matrix elements; this is fundamentally different from our approach in which there are no integrals and all matrix elements are determined by evaluating functions at points. In other cases,\textsuperscript{12,28-30} collocation is used, but with complicated basis functions.

In a recent work,\textsuperscript{31} we have shown that rectangular collocation can be effective for solving the electronic Schrödinger equation. We explicitly excluded points from the cusp region and concentrated collocation points around atoms without sampling the exponential tails of the wavefunction. We showed that this point placement strategy is advantageous compared to a uniform grid. Slater-type basis functions were used, which is difficult with a variational method. We solved the Schrödinger equation for the H atom and H\textsubscript{2}\textsuperscript{+} and the Kohn-Sham equation for CO and H\textsubscript{2}O. The same ideas will also work when the operator on the left in Eq. 2 is the Fock operator. When solving the KS equation, to achieve our target mHa accuracy, we used about \( M = 51,000 \) collocation points.\textsuperscript{31} A linear combination of atomic orbitals (LCAO) type basis was used with about \( N = 150 \) basis functions, i.e. we had \( M \gg N \). Optimization of collocation points and of the basis can greatly improve accuracy or decrease the calculation cost, as we have shown for the case of the vibrational Schrödinger equation.\textsuperscript{3-7,9,10} In Ref. 31, the point set was not optimized.

\textit{In this paper, we optimize the collocation point set to decrease the cost of solving the Kohn-Sham equation.} We solve Kohn-Sham equation on a given effective potential – a critical step in any DFT calculation. This step produces orbitals and their energies. In a complete DFT calculation, to achieve self-consistency, the orbitals and energies are used to assemble a new effective potential and density for the next step and for total energy calculations. This is done with well-established procedures and is not considered here. Orbital energies themselves are also explicitly used in many types of simulations including linear-response time-dependent DFT calculations,\textsuperscript{32} estimations of ionization potentials (IP) and electron affinities (EA)
including higher IP and lower EA;\textsuperscript{33} in applications, they are also widely used to approximate redox potentials and driving forces to electron transfer.\textsuperscript{34,35} In addition, electronic coupling is often estimated based on orbital energy splitting\textsuperscript{36} and energies of core electronic orbitals are helpful in the analysis of XPS spectra.\textsuperscript{37} We therefore focus on orbital energies in order to optimize collocation points. We optimize the collocation points by choosing them to minimize the root means square error \textit{(rmse)} of the potential obtained by representing the exact potential with Gaussian processes (GP).\textsuperscript{38,39} We hypothesize that points that are the best for representing the potential are also good for solving the SE. This is strikingly demonstrated by calculations on formaldehyde, where it was shown that it is possible to compute accurate vibrational levels using only points on the slabs required to determine a multimode representation\textsuperscript{40-43} of the PES.\textsuperscript{5} It was also shown in Ref. 11 where the fitting points (used with GP and neural network PES representations) that reduced the global PES error also reduced the error in vibrational levels computed with collocation. To minimize the potential \textit{rmse}, we use a genetic algorithm (GA).\textsuperscript{44} A GA is used to avoid the problem of local minima. We feed the GA populations obtained by optimizing the parameters of a probability distribution used to select them. We show, for the solution of the KS equation for CO and H\textsubscript{2}O, that this scheme is capable of reducing the collocation point set size from about 51,000 points to only 2,000-3,000 points while maintaining mHa accuracy.

2 Collocation and its advantages

Both the electronic and the nuclear Schrödinger equations are typically solved by expanding the wavefunction in a basis:

$$\psi(x) \approx \sum_{k=1}^{N} c_k \phi_k(x),$$  \hspace{1cm} (4)

with a vector of coefficients $c$ and basis functions $\phi_k(x)$. The SE then becomes

$$\sum_{k=1}^{N} c_k \hat{H} \phi_k(x) + V(x) \sum_{k=1}^{N} c_k \phi_k(x) = E \sum_{k=1}^{N} c_k \phi_k(x).$$  \hspace{1cm} (5)

If Eq. 5 is multiplied on the left by $\phi_m(x)$ and integrated over the support of the basis functions, one obtains the matrix version of the Schrödinger equation in the variational approach:\textsuperscript{15}

$$Hc = Ec,$$  \hspace{1cm} (6)
where $H_{ij} = \langle \phi_j | \hat{H} | \phi_i \rangle$ and $S_{ij} = \langle \phi_j | \phi_i \rangle$. Eq. 6 is a square matrix equation with Hermitian matrices. Eq. 6 has real eigenvalues that are upper limits for exact energies and converge to energy levels as the basis approaches completeness. This standard variational approach requires computing many integrals because elements of $H$ and $S$ are integrals. Integrals are also necessary to compute an energy from coupled cluster theory (where $E = \langle \Psi | \hat{H} | \Psi \rangle$ with $\Psi$ the many body wavefunction).

When using Eq. 6, it is common to compute some integrals (e.g. those for the KEO and $S$) exactly and others using quadrature; this means that quadrature approximations to integrals must be accurate. Using a mix of exact integrals and quadrature causes several problems:

(i) Basis functions must have amplitude wherever wavefunctions have significant amplitude. Quadrature points must be chosen so that quadrature approximations to integrals computed in the region in which wavefunctions have significant amplitude are accurate. This may require having quadrature points close to points at which a term in the Hamiltonian is singular. In the electronic problem, the Coulombic potential supports both core, atom-centered states, which require dense grids, as well as delocalized states which require extended quadrature grids.

(ii) To ensure that quadrature approximations to integrals are accurate, it is best to choose basis functions for which integrals of $\phi_k(x)V(x)\phi_{k'}(x)$ and $\phi_k(x)\phi_{k'}(x)$ are finite. This limits the space of possible functions. The basis is deemed complete if the $L^2$ norm $\int (\psi(x) - \sum_{k=1}^{N} c_k \phi_k(x))^2 dx \approx 0$.

(iii) If the Hamiltonian has singularities, it is tricky to find basis functions with which one can represent wavefunctions everywhere, which is necessary in a variational calculation. Singularities also make quadrature approximations to integrals inaccurate. For example, it is not practical to represent a nuclear cusp with plane waves. For vibrational calculations, the singularities (if any) are in the KEO and not the PES. For electronic calculations, the singularities are in the potential.

(iv) For some problems, the quadrature grid required to compute accurate potential and overlap matrix elements is larger than the basis size. When solving the vibrational problem, if the quadrature grid is large enough, an analytic PES is necessary, as the number of points is too large to be computed ab initio.
Collocation is an alternative way to cast the Schrödinger equation into matrix form. It also uses a basis expansion of the wavefunction, but the coefficients are determined by demanding that the Schrödinger equation be satisfied at a set of points \( \{x_i\}, i = 1, \ldots, M \):

\[
\sum_{k=1}^{N} c_k \hat{\phi}_k(x_i) + V(x) \sum_{k=1}^{N} c_k \phi_k(x_i) = E \sum_{k=1}^{N} c_k \phi_k(x_i) \tag{7}
\]

or in matrix form

\[
Dc + VFc = Ec. \tag{8}
\]

where \( D_{ik} = \hat{\phi}_k(x_i) \), \( F_{ik} = \phi_k(x_i) \), \( V_{ik} = \delta_{ik}V(x_i) \). The elements of the matrix \( D \) can be computed analytically or numerically. The numerical application of the KEO allows one to use the Cartesian KEO and apply it to any basis functions in any coordinates. Equation 8 is a rectangular matrix equation because in general \( M > N \). It can be approximately solved as a rectangular matrix pencil. Alternatively, it can be converted to a square generalized eigenvalue problem, that can be solved exactly, by left multiplying by \( F^T \):

\[
F^T(D + VF)c = EF^TFc. \tag{9}
\]

The exact solutions of Eq. 9 are approximate solutions of Eq. 8. Several authors have noticed that when Eq. 9 is used, it is not necessary that matrix elements be good approximations for integrals. This “rectangular collocation” approach has the following properties and advantages:

(i) No integrals are computed. Although \( F^TVF \) has the same form as a quadrature approximation (with unit weights), accurate energies can be computed without choosing points so that elements of \( F^TVF \) are quadrature approximations to integrals. For this reason, there is no need to use points in the entire region in which wavefunctions have significant amplitude.

(ii) Any basis functions can be used. There is no requirement that \( \phi_k(x)V(x)\phi_{k'}(x) \) or \( \phi_k(x)\phi_{k'}(x) \) be integrable. There is no requirement that basis functions have amplitude where wavefunctions have amplitude. We have recently shown that collocation works well even when the basis functions do not satisfy the integrability condition.
(iii) Because energy levels can be computed with basis functions that have no amplitude at singularities (even when wavefunctions have significant amplitude at the singularities) and it is not necessary to approximate integrals, singularities cause no problems.\textsuperscript{31}

(iv) The basis is deemed complete if, at the $M>N$ collocation points,  
$$
\psi(x_i) - \sum_{k=1}^{N} c_k \phi_k(x_i) \approx 0, \text{ i.e. the series converges pointwise.}
$$

(v) Any point distribution can be used. Important parts of space can be emphasized. In fact, accurate energies can be computed using only points in a small region of space. For example, we have shown that levels are accurate even when the collocation set does not extend to the classical turning points or when the points are placed on low-dimensional slabs.\textsuperscript{5,31}

(vi) As no integrals need to be converged, the required collocation point set is, in some cases, small enough that the potential can be directly evaluated at the points, obviating the need to fit a PES.\textsuperscript{4,6-10}

(vii) When using Eq. 9, it is not necessary to store matrices of size $M \times N$, it is possible to store matrices only of dimension $N \times N$, see Ref. 1.

These are significant advantages which should be utilized in applications.

One disadvantage of collocation is the fact that the matrix $F^T (D + VF)$ is not symmetric, which increases the cost of computing the eigenvalues. Eigenvalues that are not converged may be complex. Although collocation has the advantage that accurate spectra can be computed with fewer points than would be required with quadrature, when the point set is not dense, the wavefunction between the points might be poor, as the method only focuses on satisfying the Schrödinger equation at the points. Although many collocation point sets work well, for some points sets, the condition number of $F^TF$ is large and it is not possible to calculate accurate eigenvalues. Rectangular collocation is useful because there is a large middle ground between a point set with which quadrature approximations for elements of the overlap matrix are accurate and point sets for which the condition number of $F^TF$ is a problem. Rectangular collocation has been successfully applied to compute anharmonic molecular spectra of molecules with as many as 15 fully coupled degrees of freedom.\textsuperscript{1-11} The relatively small size of the collocation point set and ability to compute the potential matrix without an analytic PES make it possible to compute anharmonic spectra for molecules on surfaces\textsuperscript{6-8,10} which is useful in applications such as heterogeneous catalysis, photocatalysis, and electrochemical power...
Collocation has recently been used in multi-configuration time dependent Hartree\textsuperscript{62} calculations.\textsuperscript{63,64}

3 Methods

The Kohn-Sham Eq. 2 for CO and H\textsubscript{2}O was solved using Eqs. 8, 9 with \(\mathbf{x} = \{x, y, z\}\). The C-O bond length was set to 1.128 Å; the H-O bond length to 0.965 Å, and the HOH angle to 104 degrees. We solve the Kohn-Sham equation on a given effective potential computed with Gaussian 09.\textsuperscript{65} Electrostatic potentials (ESP) and the density were output from single-point Gaussian 09 calculations as cube files with a resolution of 0.0945 Bohr and 200 points in each direction. The 6-311++g(2d,2p) basis was used. We used the X\(\alpha\) functional\textsuperscript{66} to simplify constructing the exchange-correlation potential (which in the case of X\(\alpha\) is exchange-only):

\[
V_x(x) = 0.7 \frac{3}{2} \left(\frac{3}{\pi}\right)^{\frac{3}{2}} \rho(x)^{\frac{3}{2}}.
\]

Any exchange-correlation functional could be used; we chose X\(\alpha\) to avoid possible inaccuracies when extracting \(V_{\text{eff}}\) from a DFT code and for direct comparison with the results of Ref. 31. The \(V_x\) of Eq. (10) computed on the grid was added to the ESP to form \(V_{\text{eff}}\).

Elements of the matrix \(D\) are computed with a five-point finite-difference (FD) stencil, with a FD step of \(1 \times 10^{-6}\) a.u. The calculations were performed in Octave;\textsuperscript{67} the generalized eigenvalue problem was solved using the \(\text{eig}\) function. The quality of the solution was evaluated vs. the reference levels and by monitoring the residual\textsuperscript{31}

\[
\text{Res} \equiv \left| \frac{\langle \psi \mid [H \psi - E \psi] \rangle}{\langle \psi \mid \psi \rangle} \right| \approx \frac{|c^T F T (D + (V - E) F) c|}{e^T F T F c}
\]

for relevant levels. In Eq. 11, “\(\mid \)\)” implies that the absolute values of the elements of vectors are used; it is not a norm. We use the residual of Eq. 11 because it is a rectangular residual vector, \((D + (V - E) F) c\), weighted with the absolute value of the wavefunction. Energies with errors close to a mHa are achieved when \(\text{Res}\) is about 0.02 a.u.

The basis functions have the form \(\phi_{k\ell m}(x) = \phi_\ell(R_k) Y_{\ell m}(x - x_k)\), where \(Y_{\ell m}\) is a spherical harmonic, \(R_k = \|x - x_k\|\), and \(x_k\) is the ‘center’ of the \(k^{th}\) set of basis function. The functions are centered on the ions. Multiple basis functions are positioned at each \(x_k\), and their number is controlled by a parameter \(l_{\text{max}}\) such that \(l = 0, 1, \ldots, l_{\text{max}}, \) and \(m = -l, -l+1, \ldots, l.\)
Simple exponential functions $\phi_\zeta(R) = e^{-\epsilon_\zeta R}$ were used for radial components. The exponential form has the advantage that it has the correct form close to the nuclear cusp. The parameter $\epsilon_\zeta$ determines the basis function width. At each center, we use several $\epsilon_\zeta$ which we denote, $\zeta = 1, \ldots, \zeta_{\text{max}}$. That is, we use a multi-zeta type basis. The values of $\epsilon_\zeta$ were chosen to approximately minimize $Res$.

We computed the lowest energy levels starting from core $1s$ levels of O and C and up to the molecular valence states (HOMO, highest occupied molecular orbital, and LUMO, lowest unoccupied molecular orbital). As identified in Ref. 31, to get millihartree accuracy for CO, seven $\zeta$-components were required as well as $l_{\text{max}} = 2$ (called an $spd$ basis). The $\epsilon_\zeta$ values were $0.15, 0.195, 0.270, 0.315, 0.480, 0.810$, and $1.350$ (all in $a.u.$), multiplied by the value of $Z$, where $Z$ is nuclear charge. For H$_2$O, six $\zeta$-components were required and $l_{\text{max}} = 2$ (an $spd$ basis). The $\epsilon_\zeta$ values for O were $0.30, 0.39, 0.54, 0.63, 0.96$, and $1.62$ (all in $a.u.$), all multiplied by the value of $Z$, and $0.60, 0.78, 1.08, 1.26, 1.92, 3.24$, for H. These parameters were manually selected to approximately minimize $Res$ and were found to be relatively insensitive to collocation point distribution. The basis size was 126 for CO and 162 for H$_2$O.

Collocation points $x_i$ were selected from the uniform grid of 200×200×200 output by Gaussian 09. In general, collocation points can be distributed as one wishes, e.g. there is no need for them to be on a grid. They could be selected from any large pool of points, but the uniform grid is convenient because we are interfacing with $ab$ initio calculations. First, point sets similar to those used in Ref. 31 were constructed by filtering all the points on the uniform grid. A grid point $x_i$ is accepted into a collocation point set if

$$\frac{V_{\text{max}} - V(x_i) + \delta}{V_{\text{max}} - V_{\text{min}} + \kappa} > \text{rand},$$

where $\text{rand}$ is a random number in the range [0, 1]. Because there is a random element in Eq 12, the point set depends on the specific random sequence. This point selection scheme is similar to the simple potential-weighted scheme used for point selection in vibrational calculations\textsuperscript{1,11} except that we introduce $\kappa$ to flatten the distribution between $V_{\text{min}}$ and $V' = V_{\text{min}} - \kappa$ and thereby ensure that the point set contains enough points between $V_{\text{min}}$ and $V'$. This flattening is helpful because the potential for a molecule has (see Figure 2) several deep minima separated by high barriers. Application of Eq. 12 with $V' = -35$ a.u., $\delta = 0.2$ a.u. (for both molecules) resulted in a collocation set of about 51,000 points, which was large enough to
achieve millihartree accuracy in Ref. 31. In this paper, the parameters \( \kappa \) and \( \delta \) are partially optimized. \( \kappa \) is selected to ensure that all minima are sampled and that the point set obtained from Eq. 12 includes \( M' = 1000, 2000, \) and \( 3000 \) points. \( \delta \) is chosen by making plots and choosing \( \delta \) values that minimize the mean absolute error (mae), on the full point set (about 51,000), of a potential represented with Gaussian Processes (GP) and obtained from a fitting set with \( M' \) points. GP was programmed using Python’s scikit-learn library.\(^7\) The rational quadratic kernel was used. Near-optimal kernel parameters were determined by doing several trial regressions and kept fixed.

Once good values of \( \kappa \) and \( \delta \) have been identified, we use them and Eq. 12 to generate 48 sets with \( M' = 1000 \) points, 48 sets with \( M' = 2000 \) points and 48 sets with \( M' = 3000 \) points. A genetic algorithm, home-coded in Python, is used to determine an optimal point set for each of these three populations. For each \( M' \), the points undergo crossover and mutation for a number of generations. Crossover is done as follows: 1) select the best 24 datasets; 2) randomly select 24 sets of 2 parents from the best 24; 3) from each set of parents, make a new dataset whose points have a 50% probability of coming from each of the two parents. Each of these 24 new datasets then undergoes mutation by replacing some of its \( M' \) points with points randomly selected from the uniform grid of 200×200×200. The “best” 24 are the sets for which the \( \text{rmse} \), on the point set with about 51,000 points, of the GP potential is the smallest. Each of the \( M' \) points has a mutation rate of 4% (probability to be mutated). Together, the best 24 sets and the 24 offspring generated from them constitute a new set of 48 points and a new “generation”. The process of making generations is repeated until convergence is achieved. We find that evolution for about 30 generations is sufficient to optimize the point set. The set with the best fitness value is then used as the collocation point set to compute energy levels.

4 Results

4.1 Optimization of parameters \( V' \) and \( \delta \) of the probability density function

Here we describe the procedure we use to choose a \( \delta \) value for CO; a similar procedure is used for H\(_2\)O. We aim to achieve mHa accuracy with about 2,000 collocation points. We therefore select 2,000 points by applying Eq. 12 to all 8,000,000 cube points for different \( V' \) and \( \delta \) pairs. When a \( V' \) and \( \delta \) pair results in more than 2,000 points after application of Eq. 12, we select randomly 2,000 points from the resulting set. If fewer than 2,000 points are selected by Eq. 12, all those points are used. For each \( V', \delta \) pair and its set of \( \leq 2,000 \) points, we compute the \( \text{rmse} \), the \( \text{mae} \) (mean absolute error) of the GP effective potential at all the
51,000 points used in Ref. 31 as well as the residual Res of the KS equation for levels up to and including the LUMO. Figure 1 shows the dependence of the test-set rmse, mae, and of Res on \( V' \) and \( \delta \). The mae and Res plots indicate that optimal values of \( \delta \) around 0. We choose the values \( \delta = 0 \) and \( V' = -108 \) a.u. See Figure 2 for a one-dimensional slice of the effective potential along the CO bond. The value of \( V' \) is chosen to be above -110 a.u. to make sure that the minimum of the effective potential well at the C atom (this minimum is finite due to discretization) is within the range of energies in which the probability of selection is unity, as well as to make sure that about 2,000 points are be obtained with Eq. 12 (lower / more negative values of \( \kappa \) and \( \delta \) result in fewer points). With \( \delta = 0 \) and \( V' = -108 \) a.u., about 2,000 points are selected with Eq. 12. Similarly, the best 1,000 and 3,000 point sets were obtained by adjusting \( V' \) so that applying Eq. 12 to the 8,000,000 cube points gives, respectively, about 1,000 or 3,000 points.

The resulting point distribution, with \( \delta = 0 \) and \( V' = -108 \) a.u. and 2,000 points is shown in Figure 3. It is superimposed on the point distribution used in Ref. 31, with about 51,000 points. With the point set of Ref. 31, all the energy errors are about a mHa. They are about \( 10^{-2} \) Ha with the \( \delta = 0 \) and \( V' = -108 \) a.u. point set, although it contains many fewer points. The GA-optimized point set of the next subsection has a similar distribution, but better energy errors.

Table 1 shows the Gaussian 09 reference KS energies as well as energies and errors computed with collocation using point distributions obtained with Eq. 12 and different \( \kappa \) and \( \delta \) parameters. In column three of Table 1, we report energies computed with the \( \kappa \) and \( \delta \) parameters of Ref. 31, but using only 2,000 points. The errors are huge. Column 4 has energies computed with the “good” \( \kappa \) and \( \delta \) values, determined as explained in this subsection. Although the errors are much smaller, they are still unacceptably large. Even worse, the wavefunctions computed from such collocation point sets have extraneous features (see below). Much smaller errors (the GAGP columns) can be obtained, as presented in the next subsection, by using the probability density function determined by the best \( \kappa \) and \( \delta \) values as the starting point for an optimization with a GA.
Table 1. Kohn-Sham energies of CO computed with Gaussian 09 (reference) and with collocation on the same effective potential using different collocation point sets, constructed using the probability density function of Eq. 12 with particular $V$, $\delta$ parameters without or with (“GPGA”) subsequent application of machine learning (a combination of Gaussian process regression and a genetic algorithm). HOMO and LUMO energies are in bold. The sum of the eigenvalues of the occupied states and its error as well as the rmse of the GP-interpolated $V_{eff}$ are given at the bottom. The errors in the KS levels and in the sum vs. the reference are also given (“error”). All values are in a.u. The values in the columns labelled by “2,000 (Eq 12)” have poor reproducibility and should be viewed as representative.

| no. of points | 51,000 (Eq 12) | 2,000 (Eq 12) | 2,000 (Eq 12) | 1,000 (GAGP) | 2,000 (GAGP) | 3,000 (GAGP) |
|---------------|----------------|----------------|----------------|---------------|---------------|---------------|
| $k \delta$    | $k \delta$    | $k \delta$    | $k \delta$    | $k \delta$    | $k \delta$    | $k \delta$    |
| G09 energies  | energy error  | energy error  | energy error  | energy error  | energy error  | energy error  |
| -18.7401      | -18.7384 0.0017 | -36.3609 -17.6208 | -18.7270 0.0131 | -18.7550 -0.0146 | -18.8100 -0.0696 | -18.7430 -0.0032 |
| -9.9072       | -9.9037 0.0035 | -24.2358 -14.3286 | -9.8568 0.0504 | -9.9171 -0.0099 | -9.9156 -0.0884 | -9.9075 -0.0003 |
| -1.0480       | -1.0410 0.0070 | -23.6071 -22.5590 | 1.0491 -0.0011 | 1.0522 -0.0042 | 1.0624 -0.0144 | 1.0467 0.0014 |
| -0.4912       | -0.4922 -0.0010 | -23.6071 -23.1158 | -0.4917 -0.0005 | -0.4915 -0.0002 | -0.4926 -0.0014 | -0.4903 0.0009 |
| -0.4133       | -0.4162 -0.0029 | -20.4418 -20.0286 | -0.4163 -0.0030 | -0.4167 -0.0034 | -0.4279 -0.0146 | -0.4149 -0.0017 |
| -0.4133       | -0.4142 -0.0009 | -20.4418 -20.0286 | -0.4143 -0.0010 | -0.4144 -0.0011 | 0.4100 0.0033 | 0.4116 0.0017 |
| -0.3043       | -0.3043 -0.0008 | -17.8687 -17.5644 | -0.3049 -0.0006 | -0.3056 -0.0013 | -0.3055 -0.0012 | -0.3044 -0.0001 |
| -0.0518       | -0.0529 -0.0011 | -17.8687 -18.1690 | -0.0520 -0.0001 | -0.0527 -0.0009 | -0.0540 -0.0022 | -0.0526 -0.0008 |
| -0.0518       | -0.0519 -0.0001 | -17.7264 -16.7475 | -0.0517 -0.0001 | -0.0509 -0.0009 | -0.0488 0.0030 | -0.0513 0.0005 |
| 0.0384        | 0.0619 0.0235 | -17.7264 -17.7648 | 0.0893 0.0509 | 0.0864 0.0480 | 0.0976 0.0592 | 0.0822 0.0438 |

| rmse of $V_{eff}$ | 0.0019 | 0.0019 | 0.0009 | 0.0007 | 0.0018 | 0.0013 |
Table 2. Kohn-Sham energies of H₂O computed with Gaussian 09 (reference) and with collocation on the same effective potential using different collocation point sets, constructed using the probability density function of Eq. 12 with particular \( V' , \delta \) parameters without or with ("GPGA") subsequent application of machine learning (a combination of Gaussian process regression and a genetic algorithm). HOMO and LUMO energies are in bold. The sum of the eigenvalues of the occupied states and its error as well as the \( \text{rmse} \) of the GP-interpolated \( V_{\text{eff}} \) are given at the bottom. The errors in the KS levels and in the sum vs. the reference are also given ("error"). All values are in \( a.u. \). The values in columns labelled by "2,000 (Eq. 12)" have poor reproducibility and should be viewed as representative.

| no. of points | 51,000 (Eq. 12) | 2,000 (Eq. 12) | 2,000 (Eq. 12) | 1,000 (GAGP) | 2,000 (GAGP) | 3,000 (GAGP) |
|---------------|-----------------|-----------------|-----------------|--------------|--------------|--------------|
| G09 energies  | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) | \( \kappa \) \( \delta \) |
| \( 20 \) \( 20 \) | \( -35.00 \) \( 0.20 \) | \( -35.00 \) \( 0.20 \) | \( -108.00 \) \( 0.00 \) | \( -208.00 \) \( 0.00 \) | \( -208.00 \) \( 0.00 \) | \( -108.00 \) \( 0.00 \) | \( -108.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) | \( -70.00 \) \( 0.00 \) |

| \( \text{rmse of } V_{\text{eff}} \) | \( 0.0036 \) | \( 0.0026 \) | \( 0.0028 \) | \( 0.0025 \) | \( 0.0043 \) | \( 0.0016 \) |
4.2 Machine learning optimization of point distribution

CO energies computed with the collocation point set refined with the GA using the accuracy of the GP approximation to the potential are also shown in Table 1. Using only 2,000 points, errors of about one mHa are obtained. For comparison, energies with 1,000 and 3,000 collocation points are also given. They were computed from the GP+ GA optimization, starting with points obtained from Eq. 12 with $\kappa$ and $\delta$ values determined as explained in the previous section. With only 1,000 points, it was not possible to achieve mHa accuracy; with 3,000 points errors are about $10^{-3}$ a.u. for both the levels and the sum of eigenvalues for the occupied levels (the bandstructure part of the total energy). We observe that the accuracy of the energies is not a monotonic function of the accuracy of the potential. For example, better energies are obtained from intermediate GA generations (see e.g. generation 22 in Table 1 in the 2,000 point case) than in the final generation that has a slightly smaller potential error. This highlights the limits of the strategy of minimizing potential error when the density of the data is low. Nevertheless, mHa accuracy of the CO KS spectrum is maintained through final generations.

We used a similar strategy to optimize the collocation point set for H$_2$O. Similar values of $\kappa/V'$ and $\delta$ were needed to select 1,000 / 2,000 / 3,000 points from the total set of 8,000,000, see Table 2. For H$_2$O, we do not choose $\kappa$ so that all points close to a H nucleus are accepted by Eq. 12 because H 1s levels do not preserve their character in the molecule, contrary to 1s levels of C and O atoms in CO. Similar to the CO case, with 2,000 points, energy errors are about $10^{-2}$ Ha, before the point set is refined with the GA. After the GA optimization, and with the same number of points, the errors are about one mHa. Similar to the CO case, we observe that intermediate generations give slightly better energies, even though in the final generation the rmse of the effective potential is slightly lower. The sum of occupied eigenvalues is accurate to $10^{-3}$ Ha in the best generations. Contrary to the case of CO, the mHa accuracy is not maintained in the final generations, and this issue persists when increasing the number of points to 3,000. The limitations of the strategy of minimizing the rmse of the potential to improve the levels appear to be more severe in the case of H$_2$O. In the future, alternative optimization criteria could be explored.

Several molecular orbitals of CO and H$_2$O, starting with the lowest-energy orbitals and through the LUMO (LUMO+1 in the case of CO as LUMO and LUMO+1 are degenerate) computed by solving the Kohn-Sham equation using 2,000 GA-GP optimized collocation points are shown in Figure 4 and Figure 5, respectively. They are visually indistinguishable from those computed by Gaussian and by the collocation method using the large set of 51,000
points, see Fig. 6 of Ref. 31. The data for Figure 4 and Figure 5 were computed with Eq. 4 on the same 200$^3$ cube points originally generated with Gaussian. Although there is no guarantee that wavefunctions computed with collocation will be accurate at points that are not collocation points, with a sufficiently dense collocation point set this is the case, as for example with the 51,000 point set used in Ref. 31. It is also the case with the small GA-optimized point sets, but not necessarily the case with the non-optimized sets. For example in Figure 6, the orbital plots of H$_2$O computed from a set of 2,000 points with the same $\delta = 0$ and $V' = -108$ a.u. but without machine learning optimization, have easily visible deficiencies. The corresponding residuals are large (on the order of 3 a.u.) although the energies have errors only on the order of $10^{-2}$ a.u. (Table 2).

5 Conclusions

We have used a genetic algorithm to optimize the collocation point set used to solve the electronic Schrödinger equation with the rectangular collocation approach. Other ideas have been used before to identify the most important region of a potential surface when studying nuclear dynamics. Here, we used machine learning for point optimization.$^{71,72}$ The optimization criterion is the accuracy of a Gaussian Processes regression potential made from the points. Using an optimized set of 2,000 points to solve the Kohn-Sham equations for CO and H$_2$O with a reference effective potential yields orbital energies and a sum of energies with errors of about 0.001 mHa. This point set is more than an order of magnitude smaller than the point set employed in Ref. 31, indicating that points that are good for representing the potential are also good for computing energies. This is useful because it means that the potential can be used to refine the point set without solving the Schrödinger equation. We fed the GA populations generated by choosing points using a probability distribution with parameters optimized with the GP interpolated potential and large set of test points. We observe that although it is helpful to start with a good probability density function, optimizing its parameters is not by itself sufficient to achieve mHa accuracy: it is necessary to also do the GA refinement.

A key advantage of rectangular collocation is the possibility to compute accurate energies using small basis sets and basis functions that have no amplitude at singularities and do not satisfy integrability conditions. Another advantage is that one may use a small collocation point set and that many point distributions work well. All of these advantages stem from the fact that no integrals are computed. In this paper, we show that machine learning can be used to significantly reduce the size of the collocation point set, without loss of accuracy. It is possible
that by using a small basis and an optimized point set with rectangular collocation it will be less costly to solve the Schrödinger equation than with standard methods.

The idea of choosing points for solving the Schrödinger equation by optimizing the accuracy of an interpolant for the potential is interesting. For the molecules studied in this paper, it works, but is not perfect. We observe that slightly better energies are obtained with point sets having some slightly larger potential rmse values. For CO, mHa accuracy was then maintained through later generations of the GA optimization, but the accuracy did not improve as the potential rmse slightly improved. For the water molecule, mHa accuracy was achieved in intermediate generations, but lost in later generations, with a couple levels having errors as large as $10^{-2}$ a.u., although the potential rmse did slightly improve. This is indicative of the limits of using potential errors to guide the choice of collocation points when the points are not dense enough and highlights the need to explore other optimization criteria. Overall, however, potential-driven GA optimization is effective. For both CO and H$_2$O, GA-optimized collocation point sets resulted in correct orbital shapes and accurate energies and with no GA optimization, although energies are relatively accurate, points drawn from the same distribution can yield bad orbitals (incorrect behavior between collocation points) and high rectangular residuals.

6 Acknowledgements
This work was supported by the Natural Science and Engineering Research Council of Canada. A. K. thanks Strategic India Initiative by the National University of Singapore.

7 Conflicts of interest
There are no conflicts of interest to declare.

8 References
1. S. Manzhos and T. Carrington, Using an internal coordinate Gaussian basis and a space-fixed Cartesian coordinate kinetic energy operator to compute a vibrational spectrum with rectangular collocation, J. Chem. Phys., 2016, 145, 224110.
2. S. Manzhos, K. Yamashita and T. Carrington, Using a neural network based method to solve the vibrational Schrödinger equation for H$_2$O, Chem. Phys. Lett., 2009, 474, 217-221.
3. S. Manzhos, K. Yamashita and T. Carrington, On the advantages of a rectangular matrix collocation equation for computing vibrational spectra from small basis sets, Chem. Phys. Lett., 2011, 511, 434-439.
4. S. Manzhos, T. Carrington, L. Laverdure and N. Mosey, Computing the anharmonic vibrational spectrum of UF$_6$ in 15 dimensions with an optimized basis set and rectangular collocation, J. Phys. Chem. A, 2015, 119, 9557-9567.
5. S. Manzhos, X. Wang and T. Carrington, A multimode-like scheme for selecting the centers of Gaussian basis functions when computing vibrational spectra, Chem. Phys., 2018, 509, 139-144.
6. S. Manzhos, M. Chan and T. Carrington, Favorable dimensionality scaling of rectangular collocation with adaptable basis functions up to 7 dimensions, J. Chem. Phys., 2013, 139, 051101.
7. M. Chan, T. Carrington and S. Manzhos, Anharmonic vibrations of the carboxyl group in acetic acid on TiO$_2$: implications for adsorption mode assignment in dye-sensitized solar cells, Phys. Chem. Chem. Phys., 2013, 15, 10028-10034.
8. M. Chan, K. Yamashita, T. Carrington and S. Manzhos, Towards accurate spectroscopic identification of species at catalytic surfaces: anharmonic vibrations of formate on AuPt, MRS Proc., 2012, 1484, imrc12-1484-7a-0016.
9. M. Chan, S. Manzhos, T. Carrington and K. Yamashita, Parameterized bases for calculating vibrational spectra directly from ab initio data using rectangular collocation, J. Chem. Theory Comput., 2012, 8, 2053-2061.
10. S. Manzhos, T. Carrington and K. Yamashita, Calculating anharmonic vibrational frequencies of molecules adsorbed on surfaces directly from ab initio energies with a molecule-independent method: H$_2$O on Pt(111), Surf. Sci., 2011, 605, 616-622.
11. A. Kamath, R. A. Vargas-Hernández, R. V. Krems, T. 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., 2018, 148, 241702.
12. H. Nakatsuji, Discovery of a general method of solving the Schrödinger and Dirac equations that opens a way to accurately predictive quantum chemistry, Acc. Chem. Res., 2012, 45, 1480-1490.
13. H. Nakatsuji and H. Nakashima, Free-complement local-Schrödinger-equation method for solving the Schrödinger equation of atoms and molecules: basic theories and features, J. Chem. Phys., 2015, 142, 084117.
14. F. Jensen, Introduction to Computational Chemistry, Wiley, Chichester, UK, 1999.
15. A. Szabo and N. S. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, Dover, New York, USA, 1996.
16. T. Carrington, Computing (ro-)vibrational spectra of molecules with more than four atoms, *J. Chem. Phys.*, 2017, **146**, 120902.

17. J. P. Martinez, The Hartree-Fock method: from self-consistency to correct symmetry, *Ann. Phys.*, 2017, **529**, 1600328.

18. C. Møller and M. S. Plesset, Note on an approximation treatment for many-electron systems, *Phys. Rev.*, 1934, **46**, 618-622.

19. M. Head-Gordon, J. A. Pople and M. J. Frisch, MP2 energy evaluation by direct methods, *Chem. Phys. Lett.*, 1988, **153**, 503-506.

20. R. J. Bartlett and G. D. Purvis III, Many-body perturbation-theory, coupled-pair many-electron theory, and importance of quadruple excitations for correlation problem, *Int. J. Quantum Chem.*, 1978, **14**, 561-581.

21. R. J. Bartlett and M. Musiał, Coupled-cluster theory in quantum chemistry, *Rev. Mod. Phys.*, 2007, **79**, 291-352.

22. P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Phys. Rev.*, 1964, **136**, B864-B871.

23. W. Kohn and L.J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.*, 1965, **140**, A1133-A1138.

24. K. F. Garrity, J. W. Bennett, K. M. Rabe and D. Vanderbilt, Pseudopotentials for high-throughput DFT calculations, *Comput. Mater. Sci.*, 2014, **81**, 446-452.

25. R. A. Friesner, New methods for electronic structure calculations on large molecules, *Annu. Rev. Phys. Chem.*, 1991, **42**, 341-367.

26. R. A. Friesner, Solution of self-consistent field electronic structure equations by a pseudospectral method, *Chem. Phys. Lett.*, 1985, **116**, 39-43.

27. R. A. Friesner, Solution of the Hartree–Fock equations by a pseudospectral method: Application to diatomic molecules, *J. Chem. Phys. Lett.*, 1986, **85**, 1462-1468.

28. H. Nakatsuji, Structure of the exact wave function, *J. Chem. Phys.*, 2000, **113**, 2949-2956.

29. H. Nakatsuji, Scaled Schrödinger equation and the exact wave function, *Phys. Rev. Lett.*, 2004, **93**, 030403.

30. H. Nakatsuji, H. Nakashima, Y. Kurokawa and A. Ishikawa, Solving the Schrödinger Equation of Atoms and Molecules without Analytical Integration Based on the Free Iterative-Complement-Interaction Wave Function, *Phys. Rev. Lett.*, 2007, **99**, 240402.

31. S. Manzhos and T. Carrington, Using rectangular collocation with finite difference derivatives to solve electronic Schrödinger equation, *J. Chem. Phys.*, 2018, **149**, 204105.
32. M. E. Casida, Time-dependent density-functional theory for molecules and molecular solids, *J. Mol. Struct.: THEOCHEM*, 2009, **914**, 3-18.

33. Y. Mei, C. Li, N. Q. Su and W. Yang, Approximating quasiparticle and excitation energies from ground state generalized Kohn–Sham calculations, *J. Phys. Chem. A*, 2019, **123**, 666-673.

34. A. Furube, R. Katoh and K. Hara, Electron injection dynamics in dye-sensitized semiconductor nanocrystalline films, *Surf. Sci. Rep.*, 2014, **69**, 389-441.

35. W. Cao and J. Xue, Recent progress in organic photovoltaics: device architecture and optical design, *Energy Environ. Sci.*, 2014, **7**, 2123-2144.

36. L. Wang, G. Nan, X. Yang, Q. Peng, Q. Lia and Z. Shuai, Computational methods for design of organic materials with high charge mobility, *Chem. Soc. Rev.*, 2010, **39**, 423–434.

37. O. Grånäs, I. Di Marco, P. Thunström, L. Nordström, O. Eriksson, T. Björkman and J.M. Wills, Charge self-consistent dynamical mean-field theory based on the full-potential linear muffin-tin orbital method: Methodology and applications, *Comput. Mater. Sci.*, 2012, **55**, 295-302.

38. B. Kolb, P. Marshall, B. Zhao, B. Jiang and H. Guo, Representing global reactive potential energy surfaces using Gaussian processes, *J. Phys. Chem. A*, 2017, **121**, 2552-2557.

39. 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.*, 2016, **49**, 224001.

40. 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.*, 1997, **107**, 10458-10469.

41. H. Rabitz and O. F. Ali, General foundations of high-dimensional model representations, *J. Math. Chem.*, 1999, **25**, 197-233.

42. G. Rauhut, Efficient calculation of potential energy surfaces for the generation of vibrational wave functions, *J. Chem. Phys.*, 2004, **121**, 9313-9322.

43. Y. Scribano and D. Benoit, Iterative active-space selection for vibrational configuration interaction calculations using a reduced-coupling VSCF basis, *Chem. Phys. Lett.*, 2008, **458**, 384–387.

44. M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge MA, USA, 1998.

45. J. K. L. McDonald, Successive approximations by the Rayleigh-Ritz variation method, *Phys. Rev.*, 1933, **43**, 830-833.
46. G. Avila, T. Carrington, Using nonproduct quadrature grids to solve the vibrational Schrödinger equation in 12D, *J. Chem. Phys.*, 2011, **134**, 054126.

47. G. Avila, T. Carrington, Solving the vibrational Schrödinger equation using bases pruned to include strongly coupled functions and compatible quadratures, *J. Chem. Phys.*, 2012 **137**, 174108.

48. W. Yang and A. C. Peet, The collocation method for bound solutions of the Schrödinger equation, *Chem. Phys. Lett.*, 1988, **153**, 98-104.

49. D. Gottlieb and S. Orszag, *Numerical analysis of spectral methods*, SIAM, Philadelphia, USA, 1977.

50. B. W. Shore, Solving the radial Schrödinger equation by using cubic-spline basis functions, *J. Chem. Phys.*, 1973, **58**, 3855-3866.

51. S. Manzhos, T. Carrington and K. Yamashita, Non-spectral methods for solving the Schrödinger equation for electronic and vibrational problems, *J. Phys. Chem. Lett.*, 2011, **2**, 2193-2199.

52. A. C. Peet and W. Yang, The collocation method for calculating vibrational bound states of molecular systems—with application to Ar–HCl, *J. Chem. Phys.*, 1989, **90**, 1746-1751.

53. A. C. Peet and W. Yang, An adapted form of the collocation method for calculating energy levels of rotating atom–diatom complexes, *J. Chem. Phys.*, 1989, **91**, 6598-6612.

54. G. Avila and T. Carrington, A multi-dimensional Smolyak collocation method in curvilinear coordinates for computing vibrational spectra, *J. Chem. Phys.*, 2015, **143**, 214108.

55. G. Boutry, M. Elad, G. H. Golub and P. Milanfar, The generalized eigenvalue problem for non-square pencils using a minimal perturbation approach, *SIAM J. Matrix Anal. Appl.*, 2005, **27**, 582-601.

56. S. F. Boys, Some bilinear convergence characteristics of the solutions of dissymmetric secular equations, *Proc. R. Soc. London, Ser. A*, 1969, **309**, 195-208.

57. C. M. Carlson, Use of linear nonsymmetric eigenvalue equations for a sum-over-points approach: application to H$_2^+$, *J. Chem. Phys.*, 1967, **47**, 862-863.

58. D. E. Ellis and G. S. Paintert, Discrete variational method for the energy-band problem with general crystal potentials, *Phys. Rev. B*, 1970, **2**, 2887-2898.

59. E. J. Baerends, D. E. Ellis and P. Ros, Self-consistent molecular Hartree-Fock-Slater calculations I. The computational procedure, *Chem. Phys.*, 1973, **2**, 41-51.
60. A. Kamath and S. Manzhos, Inverse multiquadratic functions as basis for the rectangular collocation method to solve the vibrational Schrödinger equation, *Mathematics*, 2018, 6, 253.

61. J.-T. Li, Z.-Y. Zhou, I. Broadwell and S.-G. Sun, In-situ infrared spectroscopic studies of electrochemical energy conversion and storage, *Acc. Chem. Res.*, 2012, 45, 485–494.

62. M.H. Beck, A. Jäckle, G.A. Worth, H.-D. Meyer, The multi-configuration time-dependent Hartree (MCTDH) method: a highly efficient algorithm for propagating wavepackets, *Phys. Rep.*, 2000, 324, 1-106.

63. R. Wodraszka and T. Carrington, A new collocation-based multi-configuration time-dependent Hartree (MCTDH) approach for solving the Schrödinger equation with a general potential energy surface, *J. Chem. Phys.*, 2018, 148, 044115.

64. R. Wodraszka and T. Carrington, A pruned collocation-based multi-configuration time-dependent Hartree approach using a Smolyak grid for solving the Schrödinger equation with a general potential energy surface, *J. Chem. Phys.*, 2019, 150, 154108.

65. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian 09, Gaussian, Inc., Wallingford CT, 2009.

66. J. C. Slater, *Quantum theory of molecules and solids, Vol. 4: The self-consistent field for molecules and solids*, McGraw-Hill, New York, USA, 1974.

67. J. W. Eaton, D. Bateman, S. Hauberg and R. Wehbring, GNU Octave Version 4.2.2 Manual: A High-level Interactive Language for Numerical Computations, 2018.

68. T. Kato, On the eigenfunctions of many-particle systems in quantum mechanics, *Commun. Pure Appl. Math.*, 1957, 10, 151-177.

69. S. Garashchuk and J. C. Light, Quasirandom distributed Gaussian bases for bound problems, *J. Chem. Phys.*, 2001, 114, 3929-3939.
70. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

71. N. DeGregorio, S. S. Iyengar, Efficient and adaptive methods for computing accurate potential surfaces for quantum nuclear effects: applications to hydrogen-transfer reactions, *J. Chem. Theory Comput.*, 2018, **14**, 14130-14147.

72. 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.*, 2009, **123**, 413-429.
9 Figures

Figure 1 Top to bottom: the interpolated potential test set \( rmse \) (top panel), \( mae \) (middle panel), and the collocation equation residual (Eq. 11, bottom panel) for CO when randomly selecting at most 2,000 points with Eq. 12.
Figure 2 A scan of CO (top panel) and H₂O (bottom panel) effective potential along the CO and OH bonds, respectively. An offset of 0.1 a.u. is applied to plot on the logarithmic scale. Finite well depths are due to the finite cube resolution.
Figure 3 Distributions of collocation points for CO (left panel) and H$_2$O (right panel): black dots – the 51,000 point distribution used in Ref. 31, blue circles – the 2,000 point distribution used here. Random subsets of points (rather than all 51,000 and 2,000 points) are shown for better visibility. C and O atoms are shown as green and red spheres, respectively, for CO. H and O atoms are shown as green and red spheres, respectively, for H$_2$O.
Figure 4 The lowest-energy orbitals of the CO molecule plotted from the solution of the Kohn-Sham equation using 2,000 GA-GP optimized collocation points. Orbitals 7 and 8 are HOMO and LUMO, respectively.
Figure 5 The lowest-energy orbitals of the H$_2$O molecule plotted from the solution of the Kohn-Sham equation using 2,000 GA-GP optimized collocation points. Orbitals 5 and 6 are HOMO and LUMO, respectively.
Figure 6 The lowest-energy orbitals of H$_2$O plotted from solutions of the Kohn-Sham equation using 2,000 non-optimized collocation points.