Machine Learning Adaptive Basis Sets for Efficient Large Scale Density Functional Theory Simulation

Ole Schütt[†] and Joost VandeVondele*,[†,‡]

[†]Department of Materials, ETH Zürich, 8093 Zürich, Switzerland

[‡]Swiss National Supercomputing Centre (CSCS), 6900 Lugano, Switzerland

S *Supporting Information*

ABSTRACT: It is chemically intuitive that an optimal atom centered basis set must adapt to its atomic environment, for example by polarizing toward nearby atoms. Adaptive basis sets of small size can be significantly more accurate than traditional atom centered basis sets of the same size. The small size and well conditioned nature of these basis sets leads to large saving in computational cost, in particular in a linear scaling framework. Here, it is shown that machine learning can be used to predict such adaptive basis sets using local geometrical information only. As a result, various properties of standard DFT calculations can be easily obtained at much lower costs, including nuclear gradients. In our approach, a rotationally invariant parametrization of the basis is obtained by employing a potential anchored on neighboring atoms to ultimately construct a rotation matrix that turns a traditional atom centered basis set into a suitable adaptive basis set. The method is demonstrated using MD simulations of liquid water, where it is shown that minimal basis sets yield structural properties in fair agreement with basis set converged results, while reducing the computational cost in the best case by a factor of 200 and the required flops by 4 orders of magnitude. Already a very small training set yields satisfactory results as the variational nature of the method provides robustness.
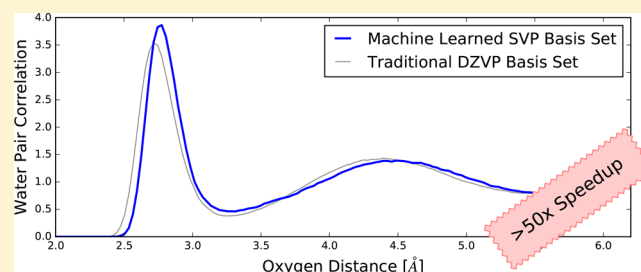
1. INTRODUCTION

The rapid increase in computational power and the development of linear scaling methods[1,2] now allow for easy single-point density functional theory (DFT) energy calculations of systems with 10,000−1,000,000 atoms.[3,4] However, the approach is computationally demanding for routine application, especially if first-principles molecular dynamics or relaxation is required. The computational cost of a DFT calculation depends critically on the size and condition number of the employed basis set. Traditional linear scaling DFT implementations employ basis sets which are atom centered, static, and isotropic. Since molecular systems are never isotropic, it is apparent that isotropic basis sets are suboptimal. Therefore, in this work a scheme is presented to define small adaptive basis sets as a function of the local chemical environment. These chemical environments are subject to change, e.g., during the aforementioned relaxations or sampling. In order to map chemical environments to basis functions in a predictable fashion, a machine learning (ML) approach is used. The analytic nature of a ML framework allows for the calculation of exact analytic forces, as required for dynamic simulations.

The idea of representing the electronic structure with adapted atomic or quasi-atomic basis functions dates back several decades. It underlays, e.g., many early tools used for the investigation of bonding order.[5−10] Also more recent methods for extracting atomic orbitals from molecular orbitals build on

this idea.[11−16] Besides using adaptive basis sets for analytic tasks, they can also be used to speed up SCF algorithms, which was pioneered by Adams.[17−19] The approach was later refined by Lee and Head-Gordon[20,21] and subsequently applied to linear scaling DFT by Berghold et al.[22] Many linear scaling DFT packages have also developed their own adaptive basis set scheme: The CONQUEST program[4] uses local support functions, derived either from plane waves[23] or pseudoatomic orbitals.[24] The ONETEP program[25] uses nonorthogonal generalized Wannier functions (NGWFs).[26] The BigDFT program[27] uses a minimal set of on-the-fly optimized contracted basis functions.[28] Other related methods include numeric atomic orbital[29−32] and localized filter diagonalization.[33−37] Recently Mao et al. used perturbation theory to correct for the error introduced into a DFT calculation by a minimal adaptive basis.[38]

Here, we focus on polarized atomic orbitals (PAOs) and build on the work of Berghold et al.[22] PAOs are linear combinations of atomic orbitals (AOs) on a single atomic center, called primary basis in the following, that minimize the total energy when used as a basis. As a result, small PAO basis sets are usually of good quality and their variational aspect is advantageous when computing properties, such as, e.g., nuclear gradients. While there is no fundamental restriction on the

PAO basis size, minimal PAO basis sets have been studied in the most detail and are also the focus of this work. Despite their qualities, the use of PAOs in simulation has been very limited, which we attribute to the difficulty of optimizing these PAOs for each molecular geometry in addition to the implied approximation. Our aim is to exploit the adaptivity of the PAO basis but to avoid this tedious optimization step by a machine learning approach.

The application of machine learning techniques to quantum chemistry is a rather young and very active field. For a recent review see Ramakrishnan and von Lilienfeld.[39] Its aim is to mitigate the high computational cost associated with quantum calculations. Initially, the research focused mostly on predicting observable properties directly from atomic positions.[40] For example, very successful recent applications include the derivation of force fields using neural network descriptions.[41−43] However, such end-to-end predictions pose a very challenging learning problem. As a consequence they require large amounts of training data with increasing system size, and the learning must be repeated for each property. Fortunately, the past decades of research have provided a wealth of quantum chemical insights. One can therefore build onto established approximations, such as DFT, and apply machine learning only to small, but expensive, subparts of the algorithms. Examples are schemes for learning the kinetic energy functional to perform orbital free DFT[44] or learning the electronic density of states at the Fermi energy.[45] Alternatively, machine learning can be used to improve the accuracy of semiempirical methods by making their parameters configuration-dependent.[46,47] In this work, machine learning is used to predict suitable PAO basis sets for a given chemical environment. The present method is thus essentially a standard DFT calculation in a geometry-dependent, optimized basis. Contrary to methods learning specific properties, including the total energy, the present method thus provides access to all properties in DFT calculations.

2. METHODS

2.1. Polarized Atomic Orbitals.
The polarized atomic orbital basis is derived from a larger primary basis through linear combinations among functions centered on the same atom. In the following, the notation from Berghold et al.[22] has been adopted. Variables with a tilde denote objects in the smaller PAO basis, while undecorated variables refer to objects in the primary basis. Formally, a PAO basis function $\tilde{\varphi}_\mu$ can be written as a weighted sum of primary basis functions φ_ν, where μ and ν belong to the same atom:

$$\tilde{\varphi}_\mu = \sum_\nu B_{\mu\nu} \varphi_\nu \tag{1}$$

As a consequence of the atom-wise contractions, the transformation matrix B assumes a rectangular block-diagonal structure. Since the primary basis is nonorthogonal, the tensor property of the involved matrices has to be taken into account.[48] Covariant matrices such as the Kohn−Sham matrix H and the overlap matrix S transform differently than the contravariant density matrix P. Hence, two transformation matrices A and B are introduced.

PAO ← primary   primary ← PAO

$\tilde{P} = A^T P A$	$P = B\tilde{P}B^T$	contravariant
$\tilde{H} = B^T H B$	$H = A\tilde{H}A^T$	covariant
$\tilde{S} = B^T S B$	$S = A\tilde{S}A^T$	covariant (2)

Notice that $A^T B = B^T A = \tilde{\mathbb{I}}$ gives the identity matrix in the PAO basis, while $AB^T = BA^T$ is the projector onto the subspace spanned by the PAO basis within the primary basis. In order to treat the matrices A and B in a simple and unified fashion, they are rewritten as a product of three matrices:

$$B = NUY \qquad A = N^{-1}UY \tag{3}$$

Due to the atom-wise contractions, the matrices N, U, and Y are block-diagonal as well. The matrices $N^{\pm 1}$ transform into the orthonormal basis, in which co- and contravariance coincide and the distinction can be dropped. The unitary matrix U rotates the orthonormalized primary basis functions of each atom such that the desired PAO basis functions become the first m_I components. The selector matrix Y is a rectangular matrix, which selects for each atom the first m_I components. Each atomic block Y_I of the selector matrix is a rectangular identity matrix of dimension $n_I \cdot m_I$, where n_I denotes the size of the primary basis and m_I the PAO basis size for the given atom I:

$$Y_I^T = \begin{pmatrix} 1 & & 0 & 0 & ... & 0 \\ & \ddots & & \vdots & & \vdots \\ 0 & & 1 & 0 & ... & 0 \end{pmatrix} \tag{4}$$

In the formulation from eq 3 the PAO basis is now solely determined by the unitary diagonal blocks of matrix U, without any loss of generality. None of the matrix multiplications required in the transformation is expensive to compute, because the matrices either are block-diagonal or expressed in the small PAO basis.

2.2. Potential Parametrization.
The PAO basis is determined by the unitary matrix U. In order to ensure the unitariness of U, it is constructed from the eigenvectors of an *auxiliary atomic Hamiltonian* H_{aux}:

$$H_{aux} = N(H_0 + V)N \tag{5}$$

$$H_{aux} = U \, \text{diag}(\{\lambda_i\})U^\dagger \tag{6}$$

Effectivly, the lowest m states of the auxiliary Hamiltonian are taken as PAO basis functions. Here, the atomic Hamiltonian H_0 describes the isolated spherical atom, and V is the polarization potential that models the influence of neighboring atoms. In the absence of V the PAO basis will reproduce the isolated atom exactly.

In the context of machine learning a parametrization should also be rotationally invariant. A parametrization without rotational invariance, on the contrary, would require training data for all possible orientations and still bear the risk of introducing artificial torque forces. In this work rotational invariant parameters X are obtained by expanding the potential V into terms V_i that are anchored on the neighboring atoms:

$$V = \sum_i X_i V_i \tag{7}$$

When the system is rotated, the potential terms V_i change accordingly, while the X_i remain invariant. As a consquence, the optimal $\vec{X}$ is independent of the system's orientation.

Explicit Form of the Potential Terms. The explicit form of the potential terms V_i must be sufficiently flexible to span the relevant subspace. Yet, they must also depend smoothly on atomic positions, be independent of the atom ordering, and be sufficiently local in nature. While we expect that more advanced forms can be found, the following scheme has been employed:

$$(V_i)_{\mu\nu} = (V_{k\mu\nu})_{\mu\nu} = \int P_u \phi_\mu^*(\vec{r})\, \Phi_k(\vec{r})\, P_v \phi_\nu(\vec{r})\, d\vec{r} \tag{8}$$

where

$$\Phi_k(\vec{r}) = \sum_{J \neq I} w_J \cdot (\vec{r} - \vec{R}_J)^{2k} \exp(-\beta_J(\vec{r} - \vec{R}_J)^2),$$

$$k = 0, 1, 2, \ldots \tag{9}$$

is a potential that results from spherical Gaussian potentials centered on all neighboring atoms.

$$P_u \phi_\nu = \begin{cases} \phi_\nu, & \phi_\nu \in \text{shell } u \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

is a projector on shells of basis functions that share a common radial part, and the same angular momentum number l, but have different m quantum numbers. Specializing the terms by different l quantum number and radial part introduces the needed flexibility, while retaining the rotational invariance. Nonlocal pseudopotentials have some resemblance to this scheme. Finally, additional terms are added that just result from the central atom, these are give by

$$(V_{uv})_{\mu\nu} = \begin{cases} P_u P_v \delta_{\mu\nu}, & l_u = l_v \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

Trivially degenerated terms with $l_u = l_v = 0$ are included only once. The weights w_J and exponents β_J could be used for fine-tunning the potential terms. However, througout this work simply $w_J = 1$, $\beta_J = 2$, and $k \leq 2$ are used.

2.3. Machine Learning.
Machine learning essentially means to approximate a complex, usually unknown, function from a given set of training points. The amount of required training data grows with the difficulty of the learning problem. Therefore, the learning problem should be kept as small as possible by exploiting a priori knowledge about the function's domain and codomain.

For the co-domain side this simplification is achieved through the previously described potential parametrization. It takes as input a PAO parameter vector and returns the unitary matrix that eventually determines the PAO basis: $\vec{X} \rightarrow U$.

For the domain side a so-called *descriptor* is used. It takes as input all atom positions and returns a low-dimensional feature vector that characterizes the chemical environment: $\{\vec{R}_I\} \rightarrow \vec{q}$. The search for a good general-purpose descriptor is an ongoing research effort.[49−51] For this work a variant of the descriptor proposed by Sadeghi et al.[52] and inspired by Rupp et al.[53] was chosen. For each atom I an overlap matrix of its surrounding atoms is constructed:

$$Q_{JK} = \sqrt{\frac{\beta_J}{\pi}} \sqrt{\frac{\beta_K}{\pi}} \int e^{-\sigma_I(\vec{r}-\vec{R}_I)^2}\, e^{-\beta_J(\vec{r}-\vec{R}_J)^2}\, e^{-\beta_K(\vec{r}-\vec{R}_K)^2}\, d\vec{r} \tag{12}$$

The eigenvalues of this overlap matrix are then used as descriptor. They are invariant under rotation of the system and permutation of equivalent atoms. The exponent σ_I acts as a screening parameter, while β_J allows the descriptor to distinguish between different atomic species. With these two simplifications in place, the learning machinery only has to perform the remaining mapping of feature vectors onto PAO parameter vectors: $\vec{q} \rightarrow \vec{X}$. A number of different learning methods have been proposed, including neural networks[54] and regression.[40] For this work a Gaussian process (GP)[55] was chosen as a relatively simple, but well characterized, ML procedure. As kernel served the popular squared exponential covariance function:

$$\text{cov}(\vec{q}, \vec{q}\,') = \exp\left[-\frac{(\vec{q} - \vec{q}\,')^2}{2\sigma^2} \right] \tag{13}$$

However, the PAO-ML scheme makes no assumptions about the employed ML algorithm and can be used in combination with any other machine learning method. Finally, a small number of *hyper-parameters* had to be optimized to achieve good results. While fixing the descriptor screening to $\sigma_I = 1$ and the GP noise level to $\epsilon = 10^{-4}$, the descriptor's β_J and the GP length scale σ were determined with a derivative-free optimizer as $\beta_O = 0.09$, $\beta_H = 0.23$, and $\sigma = 0.46$ au. For an overview of the entire PAO-ML scheme see Figure 1.
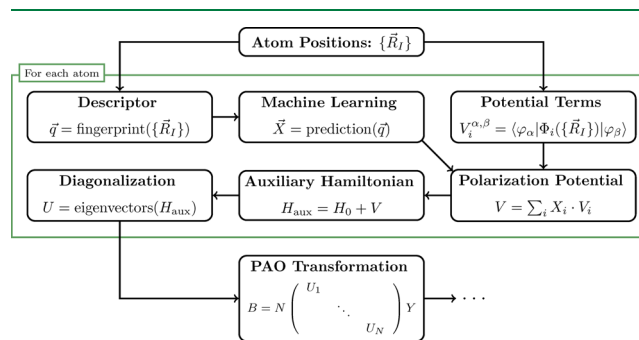
Figure 1. Overview of the PAO-ML scheme for using the potential parametrization and machine learning to calculate the PAO basis from given atomic positions.

2.4. Analytic Forces.
In order to run molecular dynamics simulations, accurate forces are essential. Forces are the derivative of the total energy with respect to atom positions. While a variationally optimized PAO basis does not contribute any additional force terms, the same does not hold for approximately optimized PAO basis sets. The advantage of using a pretrained machine learning scheme is the possibility to calculate accurate forces nevertheless.

The PAO-ML scheme contributes two force terms that have to be added to the common DFT forces $\vec{F}_{DFT}$. One term originates from the potential terms V_i from eq 8, which are anchored on neighboring atoms. The other force term arises from the descriptor, which takes atom positions as input. Both additional terms can be calculated analytically:

$$\vec{F} = -\frac{dE}{d\vec{R}} = \vec{F}_{DFT} - \frac{\partial E}{\partial V}\frac{\partial V}{\partial \vec{R}} - \frac{\partial E}{\partial \vec{X}}\frac{\partial \vec{X}}{\partial \vec{q}}\frac{\partial \vec{q}}{\partial \vec{R}} \qquad (14)$$

2.5. Training Data Acquisition. Training data are obtained by explicitly optimizing the PAO basis for a set of training structures. This poses an intricate minimization problem because the total energy must be minimal with respect to the rotation matrix U and the density matrix $\tilde{P}$. Additionally, the solution has to be self-consistent because the Kohn−Sham matrix H depends on the density matrix. Significant speedups can be obtained from temporarily relaxing the self-consistency by fixing the Kohn−Sham matrix H during an optimization cycle of $\tilde{P}$ and U.

Regularization. For high-quality training data the optimal parameters $\vec{X}$ should be unique and vary smoothly with atomic positions. To this end, two carefully designed regularization terms were introduced. The first term is inspired by Tikhonov regularization[56] and penalizes expansion on linearly dependent potential terms in eq 7. The second term is a L_2 regularization for the excess degrees of freedom in the potential V. Together both regularizations can be expressed via the overlap matrix of the potential terms:

$$\mathcal{V}_{ij} = \text{tr}[V_i^T V_j] \qquad (15)$$

as

$$R = \beta \min\left[1, \frac{\alpha}{\mathcal{V}}\right] + c\mathcal{V}$$

$$E_{reg} = \sum_{ij} X_i R_{ij} X_j \qquad (16)$$

Througout this work the values $\alpha = 10^{-6}$ and $\beta = c = 1$ mHa are used.

3. RESULTS

In this section, the performance of the method for bulk liquid water is explored. This system has a long tradition within the first-principles MD community, as it is both important and difficult to describe.[57] From an energetic point of view, the challenge arises from the delicate balance between directional hydrogen bonding and nondirectional interactions such as van der Waals interactions.[58] The relatively weak interaction can furthermore be influenced by technical aspects, such as basis set quality. Additionally, the liquid is a disordered state, which requires sampling of configurations for a proper description. The disorder makes it also an interesting test case for the ML approach, as the variability of the environment of each molecule can be large.

3.1. Learning Curve. In order to validate the PAO-ML method a learning curve is recorded. To do this, 71 frames containing 64 water molecules, spaced 100 fs apart, are taken from an earlier MP2 MD simulation at ambient temperature and pressure.[59] The first 30 frames are used as training data while the last 30 frames serve as a test set. For each training frame the optimal PAO basis is determined via explicit optimization using DZVP-MOLOPT-GTH as the primary basis. The PAO-ML method is then used to predict basis sets for all test frames based on an increasing number of training frames. The learning curve in Figure 2 shows the standard deviation of the energy difference with respect to the primary basis taken across all 64 water molecules in all 30 test frames. It shows that already a single frame, i.e., 64 molecular geometries,
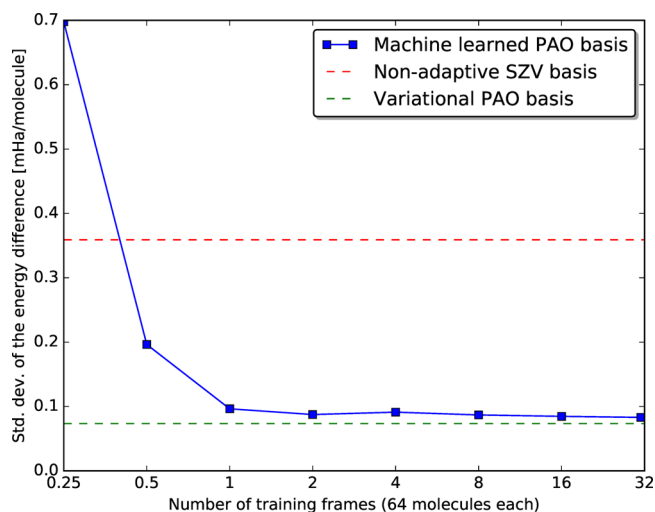

Figure 2. Learning curve showing the decreasing error of PAO-ML (blue) with increased training set size. For comparison the error of a variationally optimized PAO basis (green) and a traditional minimal SZV-MOLOPT-GTH (red) basis set are shown. With very little training data, the variational limit is approached by the ML method.

is sufficient training data to yield an error below 0.1 mHa per water molecule. The curve furthermore shows good resilience against overfitting as the error continues to decrease even for large training sets, eventually reaching 0.083 mHa per molecule. In comparison, a traditional minimal (SZV-MOLOPT-GTH) basis set exhibits an error of 0.360 mHa. The learning can at best reach the accuracy of the underlying PAO approximation (0.074 mHa). It is unlikely that the current descriptor would be sufficient to attain that bound.

3.2. Consistency of Energy and Forces. In order to validate that the forces provided by the PAO-ML implementation are consistent with its energies, a series of short molecular dynamics simulations with different time steps was performed on a water dimer. For the integration of Newton's law of motion the velocity−Verlet algorithm[60] has been employed, which has an integration error that is of second order in the time step. Figure 3 shows the fluctuations obtained with time steps of 0.4, 0.2, and 0.1 fs. The standard variations extracted from these fluctuation curves are 5.00, 1.23, and 0.31 μHa. This matches nicely the 4-fold decrease expected for a time step halving and confirms the consistency of the PAO-ML implementation.

3.3. PAO-ML Molecular Dynamics of Liquid Water. So far, we have tested the performance of the method based on frames sampled with a traditional approach. More challenging for a ML method is sampling configurations based on predicted energies, in particular, to verify that instabilities and unphysical behavior are absent when the method is given the freedom to explore phase space. To test and verify the performance, molecular dynamics simulations have been performed for 64 molecules of water at experimental density and 300 K, producing trajectories between 20 and 40 ps depending on the method. Besides PAO-ML, a traditional minimal (SZV-MOLOPT-GTH) basis set, a standard basis sets of triple-ζ quality (TZV2P-MOLOPT-GTH), and density functional tight binding (DFTB)[61,62] were used. TZV2P serves as a reference converged result, while SZV and DFTB provide insight in the performance of methods with a basis set size identical to PAO-ML. The oxygen−oxygen pair correlation
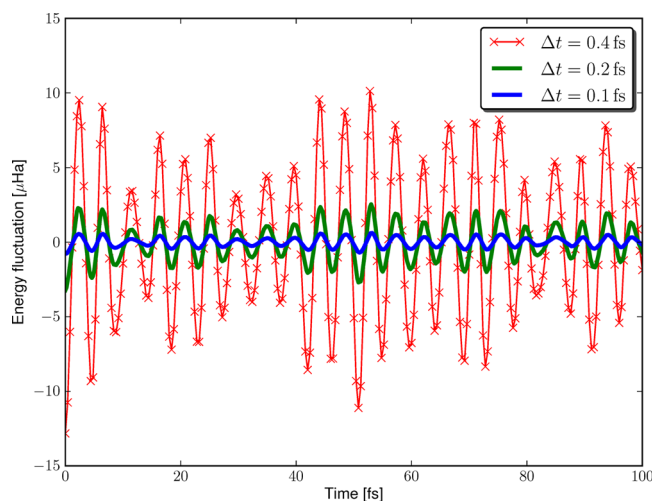
Figure 3. Energy fluctuation during a series of MD simulation of a water dimer using the PAO-ML scheme. The simulations were conducted in the *NVE* ensemble using different time steps Δt to demonstrate the consistency of the forces and thus the controllability of the integration error.

functions of liquid water are shown in Figure 4. First, these results show that the PAO-ML simulation is similar to the
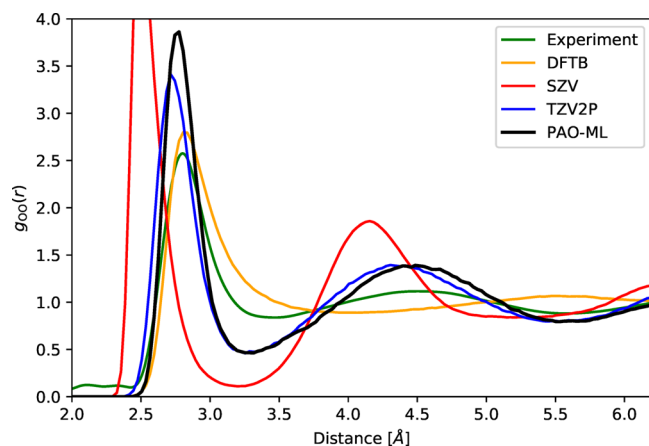
Figure 4. Shown are oxygen−oxygen pair correlation functions for liquid water at 300 K. As reference the experimental (green, ref 63) and TZV2P-MOLOPT-GTH basis sets (blue) results are shown. The SZV-MOLOPT-GTH curve (red) and DFTB (orange) are examples of results typically obtained from a minimal basis sets. The adaptive basis set PAO-ML (black) reproduces the reference (TZV2P) better than any of the alternative minimal basis set methods.

reference TZV2P-MOLOPT-GTH. The position of the first peak in the O−O pair correlation function matches well the one of the experimental reference, which is a significant improvement over the result obtained with a SZV-MOLOPT-GTH basis. Compared to the experiment, overstructuring of the first peak can be mostly attributed to the employed PBE exchange and correlation functional, as it also shows up with the triple-ζ basis set. Comparing to the DFTB results, the difference is most significant near the second solvation shell, which is mostly absent or strongly shifted to larger distances with DFTB, whereas the PAO-ML reproduces the reference results rather accurately.

3.4. Check for Unphysical Minima. We checked that the PAO-ML potential energy surface is free from unphysical

minima. To this end, the 30 test frames of bulk liquid water employed in section 3.1 were geometry optimized using PAO-ML. During this optimization the energy dropped on average by 3.14 mHa per water molecule and each atom moved on average 0.212 Å. Afterward, starting from the PAO-ML minima configuration, a second geometry optimization was performed using the DZVP-MOLOPT-GTH basis. Confirming the physical nature of the PAO-ML minima, the average energy difference between the configurations optimized with PAO-ML and DZVP is a neglibile 0.028 mHa per molecule and the positions changed on average by only 0.014 Å per atom. This confirms the quality of the PAO-ML basis.

3.5. PAO-ML Speedup. The speedup obtained with PAO-ML in the context of linear scaling calculations will be quantified. As a test system, a cubic unit cell containing 6912 water molecules (∼20000 atoms) is employed. The simulations were run on a Cray XC40 using between 64 and 400 nodes each with two CPUs. Table 1 shows the timings for

Table 1. Timings (seconds) for the Complete CP2K Energy Calculation (Full) and the Matrix Multiplication Part (mult) on a System Consisting of ∼20000 Atoms, As Described in the Text[a]

nodes	64	100	169	256	400
			PAO-ML		
full	87	58	41	33	24
mult	23	17	13	11	8
		DZVP-MOLOPT-GTH			
full	5215	2765	1996	1840	1201
mult	5036	2655	1922	1779	1165

[a]The PAO-ML method outperforms a standard DFT run with a DZVP-MOLOPT-GTH basis by a factor of at least 50×.

both the full energy calculation and the sparse matrix multiplication part alone. Linear scaling calculations are typically dominated by matrix multiplication, which made it the target of the PAO-ML method. The largest speedup for this part is observable on a few nodes, in which case the PAO-ML scheme yields a 200× wall time reduction. The number of flops actually executed decreases by 4 orders of magnitude from 61.63×10^{15} flops for DZVP-MOLOPT-GTH to only 4.07×10^{12} flops for PAO-ML. This speedup can only be partially attributed to the smaller basis set, as the reduction in flops in the dense case would be only 56× (6 vs 23 basis functions per water molecule). This demonstrates the importance of the condition number of the overlap matrix in sparse linear algebra, because the PAO basis exhibits a condition number around 6, which is more than 2 orders of magnitude lower than for the primary DZVP basis set. Due to the large speedup of the matrix multiplication, the Kohn−Sham matrix construction becomes a major contribution to the timings. Nevertheless, on 64 nodes the PAO-ML method speedup the full calculation by 60×. Running on 400 nodes allows one to perform an SCF step in just 3.3 s.

3.6. Computational Setup. All the calculations were performed using the CP2K software.[64−66] CP2K combines a primary contracted Gaussian basis with an auxiliary plane-wave (PW) basis. This Gaussian and plane-wave (GPW)[67] scheme allows for an efficient linear-scaling calculation of the Kohn−Sham matrix. The auxiliary PW basis is used to calculate the Hartree (Coulomb) energy in linear-scaling time using fast Fourier transforms. The transformation between the Gaussian

and PW basis can be computed rapidly. The cutoff for the PW basis set was chosen to be at least 400 Ry in all simulations. While the PW basis is efficient for the Hartree energy, the primary Gaussian basis set is local in nature and allows for a sparse representation of the Kohn−Sham matrix. For the simulations, the Perdew−Burke−Ernzerhof[68] (PBE) exchange and correlation (XC) functional and Goedecker−Teter−Hutter (GTH) pseudopotentials[69] were used. The linear-scaling calculations were performed with the implementation as described in ref 3, which in particular allows for variable sparsity patterns of the matrices. All SCF optimization used the TRS4[70] algorithm. The SCF optimization was converged to a threshold (EPS_SCF) of 10^{-8} or tighter; the filtering threshold EPS_FILTER was to 10^{-7} or tighter. The default accuracy (EPS_DEFAULT) was set to 10^{-10} or tighter. All simulations were run in double precision. CP2K input files are available in the Supporting Information.

4. DISCUSSION AND CONCLUSIONS

In this work, the PAO-ML scheme has been presented and tested. PAO-ML employs machine learning techniques to infer geometry adapted atom centered basis sets from training data in a general way. The scheme can serve as an almost drop-in replacement for conventional basis sets to speedup otherwise standard DFT calculations. The method is similar to semiempirical models based on minimal basis sets but offers improved accuracy and quasi-automatic parametrization.

The PAO-ML approach has the interesting property that the optimal prediction of the parameters makes the energy minimal with respect to these parameters. During the actual simulation, this implies a certain stability of the simulation, as regions with poorly predicted parameters will be avoided due to their higher energy. Ultimately, the whole PAO-ML method provides basis sets that depend in an analytical way on the atomic coordinates. As such, analytic nuclear forces are available, making the method suitable for geometry optimization and energy conserving molecular dynamics simulations.

The performance of the method was demonstrated using MD simulations of liquid water, where it was shown that small basis sets yield structural properties that outperform those of other minimal basis set approaches. Interestingly, very small samples of training data yielded satisfactory results. Compared to the standard approach, the number of flops needed in matrix−matrix multiplications decreased by over 4 orders of magnitude, leading to an effective 60-fold run-time speedup.

Finally, it is clear that the approach presented in this work can be further refined and extended. Some early results have been published in a Ph.D. thesis.[71] Possible directions for improvements include the following: (a) systematic storage and extension of reference data to yield a general purpose machine learned framework for large scale simulation, including a more rigorous quantification of the expected error, which will improve usability; (b) refined parametrization of the PAO basis sets, reducing the number of parameters needed and the enhancing the robustness of the method; (c) nonminimal PAO basis sets; (d) extensions of the method to yield basis sets for fragments or molecules rather than atoms, which will increase accuracy and efficiency; (e) more advanced machine learning techniques and alternative descriptors, which will allow for larger training sets and improved transferability of reference results. These directions should be explored in future work.

ASSOCIATED CONTENT

⑤ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00378.

Representative input files for most simulations (ZIP)

AUTHOR INFORMATION

Corresponding Author

*E-mail: joost.vandevondele@cscs.ch.

ORCID ⓘ

Joost VandeVondele: 0000-0002-0902-5111

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, 1085.

(2) Bowler, D. R.; Miyazaki, T. $O(N)$ methods in electronic structure calculations. *Rep. Prog. Phys.* **2012**, *75*, 036503.

(3) VandeVondele, J.; Borštnik, U.; Hutter, J. Linear Scaling Self-Consistent Field Calculations with Millions of Atoms in the Condensed Phase. *J. Chem. Theory Comput.* **2012**, *8*, 3565−3573.

(4) Bowler, D. R.; Miyazaki, T. Calculations for millions of atoms with density functional theory: linear scaling shows its potential. *J. Phys.: Condens. Matter* **2010**, *22*, 074207.

(5) Mulliken, R. S. Criteria for the Construction of Good Self-Consistent-Field Molecular Orbital Wave Functions, and the Significance of LCAO-MO Population Analysis. *J. Chem. Phys.* **1962**, *36*, 3428−3439.

(6) Davidson, E. R. Electronic Population Analysis of Molecular Wavefunctions. *J. Chem. Phys.* **1967**, *46*, 3320−3324.

(7) Roby, K. R. Quantum theory of chemical valence concepts. *Mol. Phys.* **1974**, *27*, 81−104.

(8) Heinzmann, R.; Ahlrichs, R. Population analysis based on occupation numbers of modified atomic orbitals (MAOs). *Theor. Chim. Acta.* **1976**, *42*, 33−45.

(9) Ehrhardt, C.; Ahlrichs, R. Population analysis based on occupation numbers II. Relationship between shared electron numbers and bond energies and characterization of hypervalent contributions. *Theor. Chim. Acta.* **1985**, *68*, 231−245.

(10) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **1988**, *88*, 899−926.

(11) Lee, M. S.; Head-Gordon, M. Extracting polarized atomic orbitals from molecular orbital calculations. *Int. J. Quantum Chem.* **2000**, *76*, 169−184.

(12) Mayer, I. Orthogonal effective atomic orbitals in the topological theory of atoms. *Can. J. Chem.* **1996**, *74*, 939−942.

(13) Cioslowski, J.; Liashenko, A. Atomic orbitals in molecules. *J. Chem. Phys.* **1998**, *108*, 4405−4412.

(14) Lu, W. C.; Wang, C. Z.; Schmidt, M. W.; Bytautas, L.; Ho, K. M.; Ruedenberg, K. Molecule intrinsic minimal basis sets. I. Exact resolution of ab initio optimized molecular orbitals in terms of deformed atomic minimal-basis orbitals. *J. Chem. Phys.* **2004**, *120*, 2629−2637.

(15) Laikov, D. N. Intrinsic minimal atomic basis representation of molecular electronic wavefunctions. *Int. J. Quantum Chem.* **2011**, *111*, 2851−2867.

(16) Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts. *J. Chem. Theory Comput.* **2013**, *9*, 4834–4843.

(17) Adams, W. H. On the Solution of the Hartree-Fock Equation in Terms of Localized Orbitals. *J. Chem. Phys.* **1961**, *34*, 89–102.

(18) Adams, W. H. Orbital Theories of Electronic Structure. *J. Chem. Phys.* **1962**, *37*, 2009–2018.

(19) Adams, W. Distortion of interacting atoms and ions. *Chem. Phys. Lett.* **1971**, *12*, 295–298.

(20) Lee, M. S.; Head-Gordon, M. Polarized atomic orbitals for self-consistent field electronic structure calculations. *J. Chem. Phys.* **1997**, *107*, 9085–9095.

(21) Lee, M. S.; Head-Gordon, M. Absolute and relative energies from polarized atomic orbital self-consistent field calculations and a second order correction.: Convergence with size and composition of the secondary basis. *Comput. Chem.* **2000**, *24*, 295–301.

(22) Berghold, G.; Parrinello, M.; Hutter, J. Polarized atomic orbitals for linear scaling methods. *J. Chem. Phys.* **2002**, *116*, 1800–1810.

(23) Bowler, D. R.; Miyazaki, T.; Gillan, M. J. Recent progress in linear scaling ab initio electronic structure techniques. *J. Phys.: Condens. Matter* **2002**, *14*, 2781.

(24) Torralba, A. S.; Todorović, M.; Brázdová, V.; Choudhury, R.; Miyazaki, T.; Gillan, M. J.; Bowler, D. R. Pseudo-atomic orbitals as basis sets for the $O(N)$ DFT code CONQUEST. *J. Phys.: Condens. Matter* **2008**, *20* (29), 294206.

(25) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. *J. Chem. Phys.* **2005**, *122*, 084119.

(26) Skylaris, C.-K.; Mostofi, A. A.; Haynes, P. D.; Diéguez, O.; Payne, M. C. Nonorthogonal generalized Wannier function pseudopotential plane-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2002**, *66*, 035119.

(27) Mohr, S.; Ratcliff, L. E.; Genovese, L.; Caliste, D.; Boulanger, P.; Goedecker, S.; Deutsch, T. Accurate and efficient linear scaling DFT calculations with universal applicability. *Phys. Chem. Chem. Phys.* **2015**, *17*, 31360–31370.

(28) Mohr, S.; Ratcliff, L. E.; Boulanger, P.; Genovese, L.; Caliste, D.; Deutsch, T.; Goedecker, S. Daubechies wavelets for linear scaling density functional theory. *J. Chem. Phys.* **2014**, *140*, 204110.

(29) Ozaki, T. Variationally optimized atomic orbitals for large-scale electronic structures. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2003**, *67*, 155108.

(30) Ozaki, T.; Kino, H. Numerical atomic basis orbitals from H to Kr. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2004**, *69*, 195113.

(31) Junquera, J.; Paz, O.; Sánchez-Portal, D.; Artacho, E. Numerical atomic orbitals for linear-scaling calculations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2001**, *64*, 235111.

(32) Basanta, M.; Dappe, Y.; Jelínek, P.; Ortega, J. Optimized atomic-like orbitals for first-principles tight-binding molecular dynamics. *Comput. Mater. Sci.* **2007**, *39*, 759–766.

(33) Rayson, M. J.; Briddon, P. R. Highly efficient method for Kohn-Sham density functional calculations of 500–10000 atom systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2009**, *80*, 205104.

(34) Rayson, M. Rapid filtration algorithm to construct a minimal basis on the fly from a primitive Gaussian basis. *Comput. Phys. Commun.* **2010**, *181*, 1051–1056.

(35) Nakata, A.; Bowler, D. R.; Miyazaki, T. Efficient Calculations with Multisite Local Orbitals in a Large-Scale DFT Code CONQUEST. *J. Chem. Theory Comput.* **2014**, *10*, 4813–4822.

(36) Lin, L.; Lu, J.; Ying, L.; E, W. Adaptive local basis set for Kohn-Sham density functional theory in a discontinuous Galerkin framework I: Total energy calculation. *J. Comput. Phys.* **2012**, *231*, 2140–2154.

(37) Lin, L.; Lu, J.; Ying, L.; E, W. Optimized local basis set for Kohn-Sham density functional theory. *J. Comput. Phys.* **2012**, *231*, 4515–4529.

(38) Mao, Y.; Horn, P. R.; Mardirossian, N.; Head-Gordon, T.; Skylaris, C.-K.; Head-Gordon, M. Approaching the basis set limit for DFT calculations using an environment-adapted minimal basis with

perturbation theory: Formulation, proof of concept, and a pilot implementation. *J. Chem. Phys.* **2016**, *145*, 044109.

(39) Ramakrishnan, R.; von Lilienfeld, O. A. *Rev. Comput. Chem.*; John Wiley & Sons, Inc., 2017; pp 225–256.

(40) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419.

(41) Handley, C. M.; Popelier, P. L. A. Potential Energy Surfaces Fitted by Artificial Neural Networks. *J. Phys. Chem. A* **2010**, *114*, 3371–3383.

(42) Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 17930–17955.

(43) Morawietz, T.; Singraber, A.; Dellago, C.; Behler, J. How van der Waals interactions determine the unique properties of water. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 8368–8373.

(44) Snyder, J. C.; Rupp, M.; Hansen, K.; Blooston, L.; Mueller, K.-R.; Burke, K. Orbital-free bond breaking via machine learning. *J. Chem. Phys.* **2013**, *139*, 224104.

(45) Schütt, K. T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K. R.; Gross, E. K. U. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 205118.

(46) Dral, P. O.; von Lilienfeld, O. A.; Thiel, W. Machine Learning of Parameters for Accurate Semiempirical Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2015**, *11*, 2120–2125.

(47) Kranz, J. J.; Kubillus, M.; Ramakrishnan, R.; von Lilienfeld, O. A.; Elstner, M. Generalized Density-Functional Tight-Binding Repulsive Potentials from Unsupervised Machine Learning. *J. Chem. Theory Comput.* **2018**, *14*, 2341–2352.

(48) White, C. A.; Maslen, P.; Lee, M. S.; Head-Gordon, M. The tensor properties of energy gradients within a non-orthogonal basis. *Chem. Phys. Lett.* **1997**, *276*, 133–138.

(49) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 184115.

(50) De, S.; Bartok, A. P.; Csanyi, G.; Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754–13769.

(51) Zhu, L.; Amsler, M.; Fuhrer, T.; Schäfer, B.; Faraji, S.; Rostami, S.; Ghasemi, S. A.; Sadeghi, A.; Grauzinyte, M.; Wolverton, C.; Goedecker, S. A fingerprint based metric for measuring similarities of crystalline structures. *J. Chem. Phys.* **2016**, *144*, 034203.

(52) Sadeghi, A.; Ghasemi, S. A.; Schäfer, B.; Mohr, S.; Lill, M. A.; Goedecker, S. Metrics for measuring distances in configuration spaces. *J. Chem. Phys.* **2013**, *139*, 184118.

(53) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(54) Behler, J. Constructing high-dimensional neural network potentials: A tutorial review. *Int. J. Quantum Chem.* **2015**, *115*, 1032–1050.

(55) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*; The MIT Press, 2006; http://www.gaussianprocess.org/gpml/ (accessed Apr. 20, 2018).

(56) Tikhonov, A. N.; Goncharsky, A. V.; Stepanov, V. V.; Yagola, A. G. *Numerical Methods for the Solution of Ill-Posed Problems*; Kluwer Academic, 1995.

(57) Gillan, M. J.; Alfè, D.; Michaelides, A. Perspective: How good is DFT for water? *J. Chem. Phys.* **2016**, *144*, 130901.

(58) Del Ben, M.; Schönherr, M.; Hutter, J.; VandeVondele, J. Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory. *J. Phys. Chem. Lett.* **2013**, *4*, 3753–3759.

(59) Del Ben, M.; Hutter, J.; VandeVondele, J. Probing the structural and dynamical properties of liquid water with models including non-local electron correlation. *J. Chem. Phys.* **2015**, *143*, 054506.

(60) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98−103.

(61) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *58*, 7260−7268.

(62) Hu, H.; Lu, Z.; Elstner, M.; Hermans, J.; Yang, W. Simulating Water with the Self-Consistent-Charge Density Functional Tight Binding Method: From Molecular Clusters to the Liquid State. *J. Phys. Chem. A* **2007**, *111*, 5685−5691.

(63) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range. *J. Chem. Phys.* **2013**, *138*, 074506.

(64) The CP2K developers group. *CP2K*, 2018; https://www.cp2k.org, (accessed Apr. 20, 2018).

(65) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 15−25.

(66) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103−128.

(67) Lippert, G.; Hutter, J.; Parrinello, M. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **1997**, *92*, 477−488.

(68) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(69) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 1703−1710.

(70) Niklasson, A. M.; Tymczak, C.; Challacombe, M. Trace resetting density matrix purification in $O(N)$ self-consistent-field theory. *J. Chem. Phys.* **2003**, *118*, 8611−8620.

(71) Schütt, O. Enabling Large Scale DFT Simulation with GPU Acceleration and Machine Learning. Ph.D. Thesis, ETH Zürich, 2017.