

A NEW APPROACH TOWARDS IMPLEMENTATION OF  
POLARIZED ATOMIC ORBITALS IN Q-CHEM USING  
NEURAL NETWORKS

Ronit Sarangi

Department of Chemistry  
University of Southern California

November, 2020

CSCI 596 Final Report

# Abstract

The application of electronic structure calculations to bigger systems is limited due to their scaling with the number of basis functions. However, we need to use basis sets with several polarization and diffuse functions in order to capture the essential chemistry of the system and in variational methods a bigger basis set affords more flexibility and closer energy to complete basis set limit.

In this work we introduce basis sets called polarized atomic orbitals (PAOs) as described by Lee and Head-Gordon [J. Chem. Phys. 107, 9085 (1997)]. PAOs have the same size as traditional minimal basis sets and are formed from atom-centered linear combinations of a bigger basis set. The transformation is such that the PAOs are distorted in an optimal way for their environment. Lee and Head-Gordon scheme for PAOs is implemented in Q-Chem. We propose the implementation of the scheme introduced by Schütt and VandeVondele [J. Chem. Theory Comput. 14(8), 4168 (2018)] which can be coupled with a machine learning (ML) model like neural network (NN). In this approach the ML takes over the complex optimization part of PAO calculations and is expected to speed-up ab-initio molecular dynamics (AIMD) calculations of larger systems significantly. We believe this approach will widen the range of systems we can perform AIMD calculations on.

We propose benchmarking systems like bulk water and long chain alkanes to compare both approaches. We expect potential practical applications of our proposed approach in studying  $\text{TiO}_2$  nanoparticles, protein folding, enzyme activity studies etc.

## Contents

<b>I. Introduction</b>	4
<b>II. Proposal</b>	6
<b>III. Theoretical Background</b>	7
A. Polarized atomic orbitals	7
B. Lee and Head-Gordon approach	9
C. Schütt's Approach: Potential parametrization	11
D. Machine Learning	11
E. Neural Networks	13
<b>IV. Systems of interest</b>	15
A. Test systems	15
B. Practical applications	16
<b>V. Final Remarks</b>	17
<b>References</b>	21

## I. INTRODUCTION

Electronic structure theory is constantly evolving with new methods being introduced constantly in order to study physical phenomena. Self consistent field methods like HF and DFT have played a key role in such calculations and continue to be important to this day[1]. However, most electronic structure calculations are limited by the size of the system because of their computational cost quantified by scaling. HF calculations tend to scale as  $N^4$ , where  $N$  is the number of basis functions used for a calculation but with efficient algorithms and approximations the scaling can be brought down to  $\sim N^3$  and the same goes for DFT. The situation is worse for post-HF methods like MP2, coupled cluster etc. which have even higher scaling. This has led to significant effort being put in the development of computationally efficient methods in order to make such calculations possible for bigger systems. Eventually, linear scaling methods were developed which scale linearly with system size (scaling is  $N$ ) and with the rapid increase in computational power single point calculations for systems containing around 1,000,000 atoms[2] became possible.

With that in mind, let us turn our attention to the importance of basis sets in a quantum mechanical calculation. Even now new basis sets are being introduced to study some specific phenomena. For example, individual gauge for localized orbitals (IGLO) basis sets used for study of magnetic properties like NMR constants[3], complex basis functions introduced for study of resonances[4], tight core basis sets introduced for study of core electrons[5] and the list goes on. With increasing complexity of the system the basis set size keeps increasing as well, compounding the scaling issue. There is a need to reduce the basis set size, while keeping accuracy the same. This is contradictory for variational SCF methods like HF since bigger basis means more independent variables, which in turn means energy closer to complete basis set limit. Basis set size vs accuracy has been an issue for decades and has been addressed by Mulliken[6], Adams[7], Lee and Head-Gordon[8], Berghold[9] who proposed slightly different versions of a similar idea: Adaptive basis sets.

Traditional basis sets used in electronic structure calculations are atom-centered and isotropic but adaptive basis sets are modified to suit their chemical environment, like polarizing towards nearby atoms. Adaptive basis sets will out-perform a traditional basis set of the same size[8][10] because the former presents a chemically accurate picture. Taking into account the localized nature of chemical bonds it is expected that adaptive basis sets, if

properly optimized, can outperform ever bigger traditional basis sets. This idea was further developed by Lee and Head-Gordon when they introduced polarized atomic orbitals (PAOs), which will be our main focus in this proposal.

PAOs are defined as atom centered functions whose dimensions can be the same as that of a minimal basis set, i.e., one function per atomic orbital. They are formed from the linear combination of a bigger basis set and are optimized such that the PAOs minimize the energy of the system within some constraint. We will follow two different approaches towards calculating PAO: The traditional Lee and Head-Gordon approach[8] and the “Potential parametrization” approach introduced by Schütt and VandeVondele[10]. Regardless of the method, obtaining PAOs is a complex optimization scheme and is an additional overhead on top of SCF calculation. This is probably the reason why the developments in PAOs have been scarce and far apart. Disregarding the optimization problem for now, PAOs can be applied in several interesting ways, for example,

- Force field based MD simulations can be replaced with ab-initio MD (AIMD) with PAO minimal basis sets, without the drawbacks of traditional minimal basis sets.
- QM-MM calculations can be converted to QM-QM[11] calculations, when a higher level of theory+basis can be used for the system of interest and the rest can be managed with a low level of theory and PAO basis.
- General speedup of electronic structure calculations, keeping in mind PAOs need to be benchmarked extensively first.

With the above points as motivation we can return to the optimization issue, especially with respect to calculations involving dynamics. Large scale dynamics calculations are usually based on classical force field methods, but their disadvantages are their empirical nature and ignoring quantum effects. On the other hand using quantum mechanical methods for describing molecular dynamics (MD), while providing excellent accuracy, is computationally intractable for large systems due to their scaling issues. A combination of PAO and linear scaling methods can alleviate those issues but the overhead of optimizing PAOs for each molecular configuration in these calculations becomes the bottleneck. In this work, as described by Schütt et al., we want to offload the complex optimization problem of obtaining PAOs to a machine learning (ML) model like a neural network (NN). The proposed approach

can act as a drop in to any standard SCF calculation and takes care of the bottleneck of optimizing PAOs at every time step in a dynamics calculation. We have chosen an neural network as our ML model because of its ease of implementation and flexibility in terms of adjustable parameters.

Machine learning (ML) is a field with a deep and rich history of its own, rather separate from quantum chemistry but recently with the introduction of “Big-Data”, increase in computation power, and open source programs like Tensorflow, Keras etc. machine learning methods have been increasingly applied to quantum chemistry[12]. In short, the goal of ML methods is interpolation and the most applied use of ML methods in chemistry has been studying structure property relations. There have been attempts of replacing quantum mechanical calculations with ML methods but they have not been successful, simply because the level of accuracy in high level quantum mechanical calculations (chemical accuracy  $\sim 0.01$  eV) is difficult to achieve via ML methods trying to do interpolate properties from structure. Another reason is the degree of understanding we get by performing electronic structure calculations far surpasses the ones done by ML methods because the former methods are derived from first principles, with reasonable approximations while the latter methods are predicting properties from prior training data. The best use of ML methods in quantum chemistry is in conjunction with electronic structure methods with effective partitioning of jobs[13]. In this work we will mention one such example where the time consuming optimization part of the PAO calculations is offloaded to a neural network, while the property calculation is done by Q-Chem[14].

## II. PROPOSAL

We propose to implement a new approach to obtain PAOs, hereby referred to as Schütt’s approach, using a neural network in Q-Chem. The neural network will be designed using TensorFlow[15], Figure 1 shows a graphical representation of our proposal. This approach is expected to provide significant speedup in AIMD calculations of large systems compared to the already implemented method, referred to as Lee and Head-Gordon approach from now on. Our proposed method can also act as a drop in for any standard SCF calculation, making it convenient to extend the use of PAOs in post-HF calculations.

We will compare the speedup, accuracy, and stability of both approaches using several

benchmark systems like bulk water and long chain alkanes. We will investigate whether Schütt’s approach shares the advantages offered by Lee and Head-Gordon approach using the benchmarks. A detailed description of both methods are provided in Section III. We also discuss practical applications of our method in studies of  $\text{TiO}_2$  nanoparticles, enzymatic activity, and bio-molecular systems.

The outline of the rest of this proposal is as follows. In section III we discuss both the approaches in detail, key implementation steps, and their advantages and disadvantages. This is followed by a brief description of machine learning (ML) and neural networks (NN) and how it factors into our proposal. Section IV describes the benchmarking details and the potential systems our implementation can be applied to.

### III. THEORETICAL BACKGROUND

#### A. Polarized atomic orbitals

The isotropic nature of traditional basis sets does not reflect reality, where the chemical system is always polarized due to its environment. A more appropriate selection of basis functions would be polarized basis sets using local atomic information. We know that bonding is largely localized in nature apparent from the fact that MOs from an SCF calculation can be rotated such that they can be classified as either bonding, non-bonding or anti-bonding, an adaptive minimal basis set should be able to convey most of the bond information about the system. Polarized atomic orbitals (PAOs) are an example of such adaptive basis functions and are obtained by linear combination of atomic orbitals on a single atom referred to here as primary basis. PAOs minimizes the total energy when used as a basis set.

The atomic orbitals used to form the PAOs or primary basis come from a bigger traditional basis set called the secondary basis set, i.e., the primary basis is formed from the atom-centered linear transformation of the secondary basis [8, 16, 17].

$$\tilde{\phi}_\nu = \sum_{\mu} B_{\mu\nu} \phi_\mu \quad (1)$$

where,  $\tilde{\phi}_\nu$  is a PAO basis function and  $\phi_\mu$  is a secondary basis function and  $\mu, \nu$  belong to the same atom. B is a transformation matrix which is block diagonal because of the atom

centered nature of the transformations.

As an example, say our secondary basis set is 6-31G for a carbon atom. We know that this basis has one contracted gaussian function for the core 1s orbital, two contracted gaussians for valence 2s and two for valence 2p, making the total number of basis functions 9. However, a minimal basis set for carbon atom has only 5 basis functions, 1 core and 4 valence. From 1, each function of the minimal PAO basis is formed from the linear combination of all functions of a larger, secondary basis like 6-31G such that the PAO basis minimizes the energy when used. This also means the transformation matrix needs to contain the structural information of the system. In addition to that the linear combination of 9 functions gives 9 independent functions, but we only need a few of them (5 to be exact) and hence there needs to be a selection of which combination to take.

The flexibility of the secondary basis due to more basis functions is not lost when transformed to PAOs. This is because the PAOs are formed from a linear combination which can be understood as a rotation, a fact exploited in Lee and Head-Gordon’s approach. PAOs should be constructed from a secondary basis with higher angular momentum functions to give them flexibility to adapt to the molecular environment. This is one of the ways to improve the PAO performance, another way is to increase the PAO size from a minimal basis set. Doing this would defeat our purpose of trying to reduce the number of basis functions, so it is best to avoid it.

The transformation matrix,  $B$ , plays a critical role since,

$$\begin{aligned}\tilde{P} &= BPB^\dagger \\ \tilde{H} &= B^\dagger HB \\ \tilde{S} &= B^\dagger SB,\end{aligned}\tag{2}$$

where  $P$ ,  $H$ , and  $S$  are the density matrix, hamiltonian matrix, and overlap matrix respectively in the secondary basis and the tilde indicates the same for PAO basis. The relations above are true assuming the secondary AOs are orthogonal, which they usually are not but with the following simplification of  $B$ ,

$$B = NUY = S^{-1/2}U\tag{3}$$

where  $N$  orthogonalizes the secondary basis functions, and hence contains the inverse of the atom centered overlap matrix elements.  $U$  is the unitary matrix which rotates the orthogonal



secondary basis functions to obtain the PAOs. All the matrices involved here are block diagonal and hence computation involving them is not expensive. This parametrization of  $B$  is common in both Lee and Head-Gordon approach and from equations 2 and 3 it is clear that the transformation matrix, and in turn the PAOs completely depend on  $U$  and the approaches of Lee and Head-Gordon and Scüth vary on how to obtain this unitary matrix.

### B. Lee and Head-Gordon approach

Since  $U$  is unitary and the operation it performs is linear transformation or rotations, Lee and Head-Gordon parametrized  $U$  as a product of  $2 \times 2$  rotation matrices, which rotate the PAO space into retained and excluded basis functions. The functions finally used as PAO basis are called retained functions and the rest are referred to as excluded functions.

$$B = S^{-1/2}U(\Theta) \quad (4)$$

The calculation to obtain  $U$  then becomes an optimization problem such that by using the PAO basis the SCF energy is minimized. Lee and Head-Gordon used a simultaneous optimization method to obtain the matrix  $U$ , which is similar to an unconstrained density matrix SCF solution. A very similar method was introduced by Lee, Venderbilt, and Nunes [18] who introduced a functional hereby referred to as LNV functional,

$$\mathbb{P} = 3PSP - 2PSPSP \quad (5)$$

given any density matrix  $P$  and overlap matrix  $S$ , we obtain a “purified” density matrix  $\mathbb{P}$  by solving the equation above. The term “purified” refers to the fact that  $\mathbb{P}$  is closer to idempotency than  $P$ , and so is closer to being a true density matrix. This condition reinforces the idempotence property of the density matrix instead of forcing it. For an arbitrary density matrix, we can introduce a Lagrangian,

$$\mathcal{L} = Tr \left( \mathbb{P}H + \frac{1}{2}\mathbb{P}\mathbb{P}\mathbb{P} \right) \quad (6)$$

used in unconstrained density matrix approach since we are not explicitly constraining the density matrix to be idempotent, but purifying it using the LNV functional. In equation

6,  $H$  is the hamiltonian matrix and  $\Pi$  is the two electron integrals. All the matrices in equation 5 and 6 are in the secondary basis.

From equation 2 we know that  $P = P(\tilde{P}, B)$  i.e, the density matrix in the secondary basis is a function of the PAO density matrix and transformation matrix and so Lee and Head-Gordon proposed a variational optimization of the Lagrangian with respect to  $\tilde{P}$  and  $B$  simultaneously, significantly reducing the number of independent variables involved. This minimization is done using a Newton-Raphson method where we need an initial guess for  $\tilde{P}$  and  $B$ . The form of the Lagrangian when the PAO density matrix is substituted is shown below only for the one electron part,

$$\begin{aligned}
\Omega(P) &= 3Tr(PSPH) - 2Tr(PSPSH) \\
&= 3Tr[(B\tilde{P}B^\dagger)S(B\tilde{P}B^\dagger)H] - 2Tr[(B\tilde{P}B^\dagger)S(B\tilde{P}B^\dagger)S(B\tilde{P}B^\dagger)H] \\
&= 3Tr[B\tilde{P}(B^\dagger SB)\tilde{P}B^\dagger H] - 2Tr[B\tilde{P}(B^\dagger SB)\tilde{P}(B^\dagger SB)\tilde{P}B^\dagger H] \\
&= 3Tr[\tilde{P}\tilde{S}\tilde{P}\tilde{H}] - 2Tr[\tilde{P}\tilde{S}\tilde{P}\tilde{S}\tilde{P}\tilde{H}]
\end{aligned} \tag{7}$$

which shows that  $\Omega(P)$  can be completely expressed using matrices in the smaller PAO basis and since these matrices are block diagonal, the computation is significantly simplified and sped-up. The steps of this method are outlined in Figure 2(a).

There are several advantages offered by this method of obtaining PAOs, the most important of which is the variational nature of this method. Because of this the total energy using PAO as the basis set will be greater than the total energy with the larger basis. This also means method to calculate nuclear gradients and Hessians remains the same as conventional SCF. Lee and Head-Gordon also found that the errors introduced by the PAOs are mostly local in nature, i.e., the deficiency lies in describing local character or a bond.

However, this approach tends to depend too much on how good the initial guess is because the LNV functional is known to give runaway solution for density matrices and hence, depending on the initial guess the convergence can be fast or slow. The overhead of constructing the hamiltonian after every optimization step is not negligible when application in dynamics is desired. We also have to keep in mind that the Newton-Raphson method needs the second derivatives to work and is an additional cost. These costs might be negligible for single point calculations but when we have calculations like AIMD, all of the calculations have to be performed at each time step and these negligible costs can add up over long times.

### C. Schütt’s Approach: Potential parametrization

The idea behind potential parametrization was the need to map the chemical environment to the PAO basis sets through machine learning[10]. In order to meet this criteria the unitary transformation operator,  $U$  needs to have additional properties which are not necessarily shared by Lee and Head-Gordon’s parametrization of  $U$ .

In order to maintain the unitary nature of  $U$ , Schütt et al proposed it be constructed using the eigenvectors of an auxiliary Hamiltonian,  $H_{aux}$  where,

$$H_{aux} = H_0 + V \quad (8)$$

$H_0$  is the isolated atom hamiltonian and  $V$  is the parametrizable “potential” due to the environment of the surrounding atoms. This is the reason this approach is called potential parametrization of  $U$ , since  $U$  is obtained from the lowest eigenvalues of this auxiliary Hamiltonian. The idea is quite similar to a mean field approach but in this case the form of the mean field potential is chosen by us instead of mathematically derived. The potential  $V$  can be further simplified as,

$$V = \sum_i X_i V_i \quad (9)$$

where the sum runs over all atoms,  $V_i$  is the form of the potential contributed by each neighboring atom and  $X_i$  is understood as the weight. The breakdown of the potential in equation 9 makes it rotationally invariant for machine learning approaches. This will be further explain in the next section. The steps involved in this method are outlined in Figures 2(b) and 3.

### D. Machine Learning

According to an introductory machine learning class offered by Stanford University, machine learning is defined as “the science of getting computers to act without being explicitly programmed”. Arthur Samuel came up with the term “machine learning” [19] and now this field is involved in many aspects of everyday life, whether we realize it or not. Its application to chemistry, however, is a recent phenomena fueled by the easy availability of commercial

quantum chemistry software and access to property-structure databases. The major application, but not the only one, is the structure to property mapping using machine learning algorithms which we will use as well. Some prominent examples are filtering molecules for use in Pharmaceuticals[20], calculation of melting points[21], study of chemical reactivity[22], frontier orbital eigenvalues[23], and XANES spectra prediction [24–26].

$$\mathcal{F}(\textit{structure}) \mapsto \textit{observables} \tag{10}$$

When probing a system for properties, we feed the structure and charge to the quantum chemistry software and it solves the Schrodinger equation using a variety of well understood approximations and gives the wavefunction/density which can be further used to calculate properties. This process is computationally expensive depending on the complexity of the system, the method requested by the user, and how well the method is implemented. From a mathematical point of view, if we consider a map between system structure and quantum mechanical observables,  $\mathcal{F}$ , then it is clear to see that this map will be very high dimensional and cannot be analytically expressed. But if we consider a small subset of systems which are similar in size and properties then we might be able to perform interpolative regression. This exact procedure is carried out a machine learning algorithms which are designed for such high dimensional interpolation.

Despite a large number of examples of ML methods being used this way, all of these models invariably fail to generalize. This is because these models are trained on a small subset of chemical space, and have no relation to quantum mechanics or physical reality. These models are simply maps from a high dimensional space to a particular observable, which are not generalizable. We want to avoid going by such a route by introducing a smaller but computationally demanding problem for ML methods to tackle, while leaving the property determination to QM methods.

As mentioned in the previous section, we have an auxiliary Hamiltonian  $H_{aux}$  with a parametrizable potential term  $V$ , to determine the PAO unitary matrix as in equation 8 and 9. Since  $V_i$  constitutes the contribution of the surrounding atoms to the potential, the  $X_i$  are usually determined by a complex optimization scheme. We propose to substitute this procedure by a supervised ML model, a neural network to be specific.

## E. Neural Networks

The ML model we propose to take over the mapping in equation 11 is a neural network, also referred to as a multilayer perceptron[27].

$$\mathcal{F}(\text{structure}) \mapsto X \quad (11)$$

The motivation behind designing neural network based ML modes was to model the human brain. Just like a brain has neurons, fires signals from the neurons which are propagated through dendrites to other neurons, a neural network contains a neuron/perceptron which is connected to several other neurons in a layer like structure as presented in Figure 5. As shown in Figure 4, a single neuron can take multiple inputs from other neurons associated with certain weights, add the weighted inputs and pass them through a nonlinear activation function, which determines the output of the neuron.

$$y = \sigma\left(\sum_i x_i w_i + b_i\right) \quad (12)$$

In equation 12,  $y$  is the neuron output and  $b_i$  are the bias in each neuron. The activation function,  $\sigma$  decides whether a certain neuron is activated or not and is usually a sigmoid function like tanh, logistic, ReLu etc. The weights and biases are the learnable parameters which are adjusted during the training of the network by a process called Back-propagation (backprop)[28]. We will not go too deep into the detail of how a NN works, but would like to reiterate that ML models like a NN are built for these kind of complex mapping problems. Before moving on, we should mention some properties of the domain of the ML model, i.e., structure.

The process of quantifying the domain of a ML model is called featurization[29], and the vectors generated from this process are called feature vectors,  $\{q_i\}$  where  $i$  is equal to or less than the number of input neurons. The feature vectors in our case are a quantification of structure/arrangement of our system, and they need to obey several properties,

- Computation without much overhead
- Rotationally and translationally invariant
- Encodable/Standardizable

The first property is obvious since we do not want much overhead in the ML process itself, but the second property is important to discuss. Since we are getting the potential an atom experiences due to nearby atoms of our system from the ML model, we need it to be rotationally invariant because of physical nature of potentials. This means even if we rotate the system, the potential experienced by one atom due to others does not change and the same goes for translation. So, our feature vectors need to be the same if we obtain a rotated or translated structure. The final property has deals with instabilities in the ML model, which is taken care of by standardization which makes the feature vector input ready. There are several well known ways of featurization[30],

1. Radial distribution function
2. Overlap matrices
3. Coulomb matrices
4. Connectivity matrices

An example of overlap matrices/function is used in Scütt’s approach,

$$Q_{jk} = \sqrt{\frac{\beta_j}{\pi}} \sqrt{\frac{\beta_k}{\pi}} \int e^{-\sigma_i(r-R_i)^2} e^{-\beta_j(r-R_j)^2} e^{-\beta_k(r-R_k)^2} dr \quad (13)$$

where j,k are atom indexes,  $\beta$ s are used to differentiate between different types of atoms,  $\sigma_i$  acts as a screening parameter. This matrix is obeys all the rules above, and is also invariant to permutation of equivalent atoms. The eigenvalues of this matrix are used as the feature vectors. There are several other examples and there is no obvious advantage to using one or the other, beside chemical reasoning and ease of determination. We plan to test several other featurization techniques for our method.

Before building the NN, we need to decide on its activation function, how many input neurons it needs to be have, how many output neurons, how many hidden layers, and how many neurons per hidden layer. These properties are called “hyperparameters” of a network and they have an effect on how well our network behaves. There is no common wisdom on how to set them, so we need extensive testing with different hyperparameters, but there are recent studies on how to automate this procedure using another ML algorithm like gaussian processes. Building ML modes has become easier and accessible to everyone because of open source programs like TensorFlow, PyTorch, Keras etc.

## IV. SYSTEMS OF INTEREST

A general scheme of our implementation is presented in figure 1. After successfully integrating this approach in QChem we have several tests cases to judge the performance of our approach, and several interesting systems where we will apply our PAO approach. This section is dedicated to the discussion about these systems,

### A. Test systems

1. Compare benchmarks from Lee and Head-Gordon studies with Scütt’s approach without the ML part
2. Compare timings and properties for both approaches, with ML

The first test is to compare Scütt’s approach to Lee and Head-Gordon’s approach using the benchmark systems studied in the latter work. There were several advantages to Lee and Head-Gordon’s approach like their method is exactly variational in nature, the errors introduced by PAOs were local i.e., scale with the number of atoms in the system. Since our approach is not strictly variational, we want to compare benchmarks to gauge the reliability of our method. The systems used in[8] are mostly linear alkanes and polyenes. The significance of using linear systems is mostly to determine what kind of errors we are getting by using PAOs instead of standard basis sets. If the errors were mostly local, i.e., arising from the inability of the PAO basis to describe intra-atomic interactions and some bonding, then the errors ( $E_{PAO} - E_{HF}$ ) will scale almost linearly with number of carbons in the alkane. If we observe non-linearity then the errors are non-local in nature which means the method cannot describe long range interactions properly. We don’t want the ML part since we also want to compare the speedups of both approaches, and more importantly we cannot use ML methods in single point calculations, unless previously trained.

The second test will include the ML/NN part and will be a testament to our approach since we will be doing AIMD where Lee and Head-Gordon’s approach becomes computationally expensive. Such a test has not been done before, so we shall use the standard system for these kinds of calculations, a water box. Describing bulk water has been a challenge made apparent by the fact that there are still force fields being developed for water, new

properties being discovered using quantum chemical methods, etc. What makes water such an interesting system are the highly directional but relatively weak hydrogen bonds with non-directional van der-Waal interactions. Being a universal solvent, water and its interactions with solutes needs to be described properly in order to study aqueous systems. There are several well established properties we should test after performing the AIMD calculation like radial distribution function, diffusion coefficient etc.

We should also compare the speedups we get using our approach with Lee and Head-Gordon’s approach. Since there have been no studies on how such a calculation depends on the form of the potential,  $V_i$  and the feature vector,  $q_i$ , we would like to conduct tests with various forms and compare the properties, speedup, NN stability, training efficiency etc.

### B. Practical applications

If we observe encouraging results from the test systems, the next step is studying complex, bigger systems of practical interest. We already have some in mind,

1. Systems where only semi-empirical methods have been applied due to size constraints.
2. Attempting to replace systems where QM-MM methods are necessary by QM-QM methods.
3. Moving forward from SCF methods to correlated methods like MP2, CCSD etc.

There are several studies, especially on nanoparticle systems[31, 32] done using semi-empirical methods[33]. There are several reasons for this, one of which has to do with size of the system. The usual size of the nanoparticles under consideration is  $\sim 1\text{-}5$  nm, which is too big for standard quantum mechanical methods to handle. However, with our approach using PAO minimal basis sets, we might be able to attempt such calculations, at least in the HF level.

Even bigger systems like bio-molecular systems[34, 35], metal surfaces etc. where the “reaction” is localized but influenced by its surroundings the method of choice is QM-MM, where the immediate surrounding of the system is treated with a high level quantum mechanical method while the rest of the system is treated with molecular mechanics using a force field. There are several disadvantages to this approach like if we have a dynamic system,



then atoms frequently move in and out of the QM region which leads to abrupt change in potential energy surfaces introducing unphysical effects, depending on the embedding scheme, the QM region may or may not experience the full effect of its surroundings. But using our approach we can try using a lower level quantum mechanical method on the surroundings which hopefully alleviates some issues, such methods are referred to as QM-QM methods[36]. We still would need to figure out the embedding scheme, but this is one of the applications we would like to explore.

Finally, throughout this proposal we have discussed the application of PAOs in non-correlated methods like HF. However, introducing correlation effects is important if we want to achieve chemical accuracy in our calculations. High level methods like MP2, CCSD etc. introduce correlation by using excited state determinants[37]. However, since PAO basis sets are minimal, we will have very few virtual orbitals to form such determinants. This issue might be addressed by increasing the size of the PAO basis but as discussed previously, it will defeat the purpose of introducing them. This dilemma needs a closer look and we would like to explore it further.

## V. FINAL REMARKS

In this proposal we introduce the implementation scheme of a new approach towards polarized atomic orbitals (PAOs) in Q-Chem. We outline this approach introduced by Scütt et al and introduce a machine learning model which can couple with it.

If successful, this method will significantly broaden the type and size of systems that can be treated quantum mechanically. Extensive benchmarking studies would be needed before we can actually put this method to practical use and we outline a few potential systems like linear chain alkanes and bulk water system.

This work also shows how we, as computational chemists can take advantage of machine learning like neural networks to offload some computational expensive part of a calculation while keeping the advantages of first principle methods.

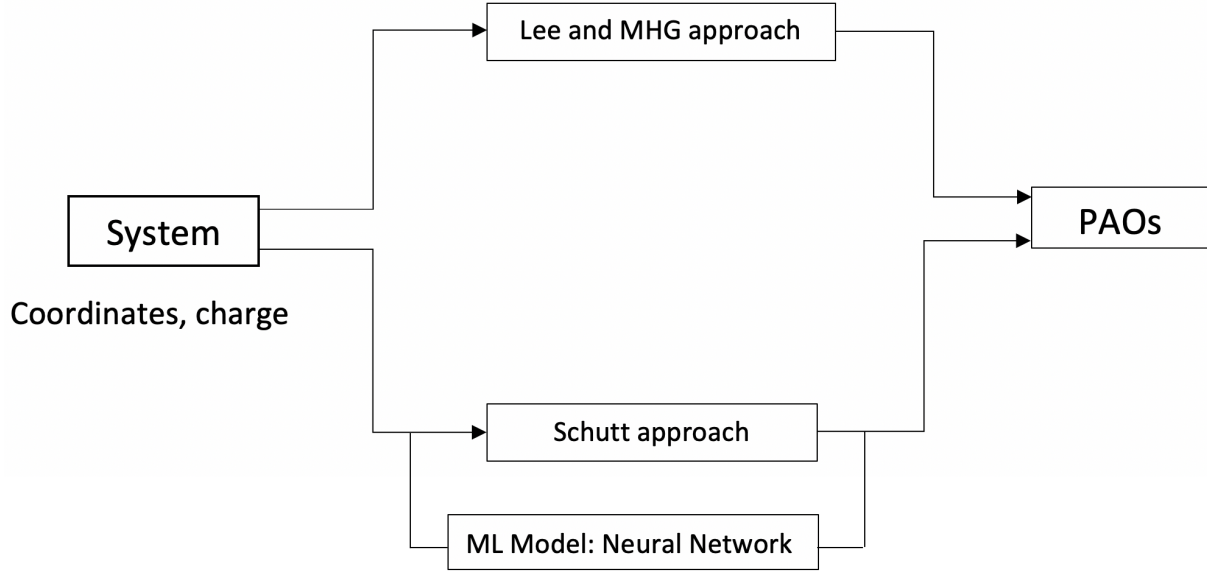


FIG. 1: Skeleton of our proposal - Implementation of Schütt's approach for PAO calculation in Q-Chem and comparing with existing Lee and MHG approach

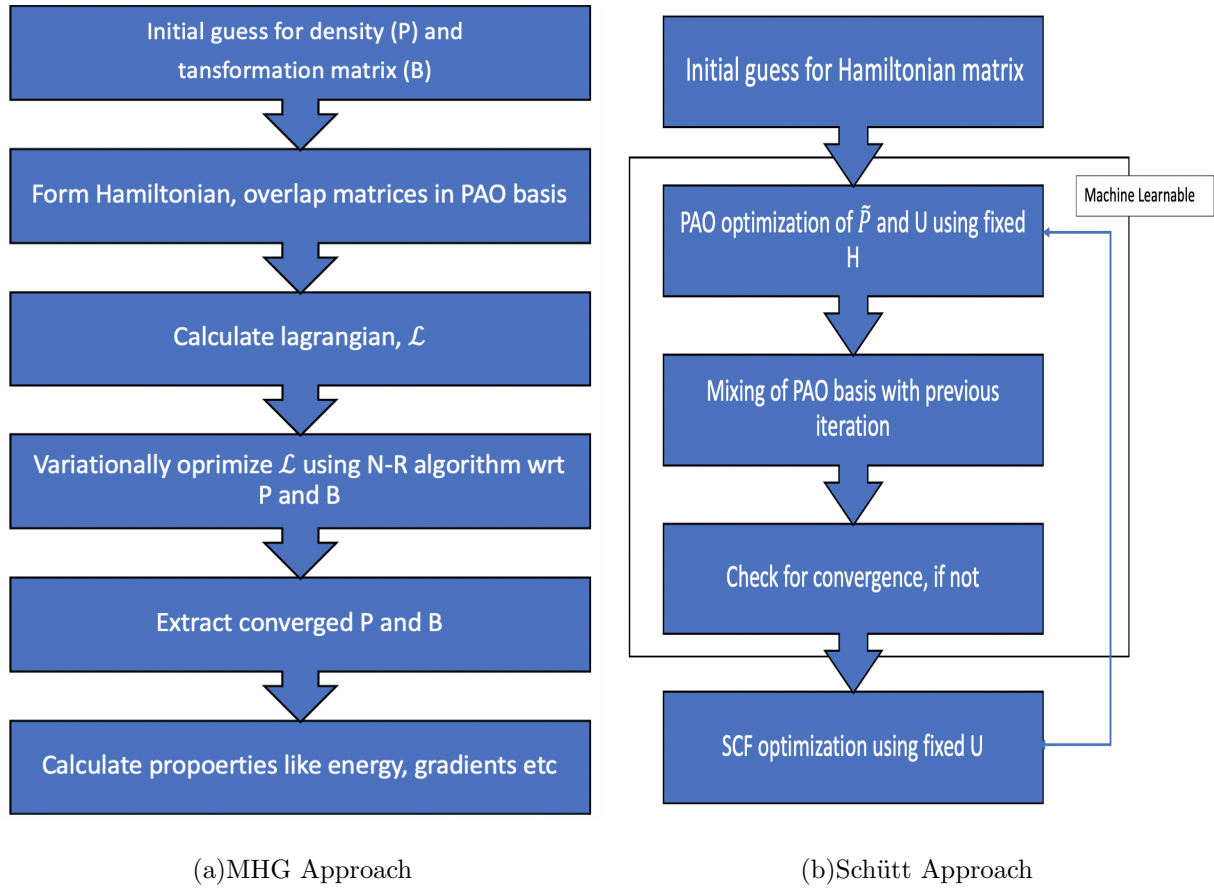


FIG. 2: Comparing both the approaches

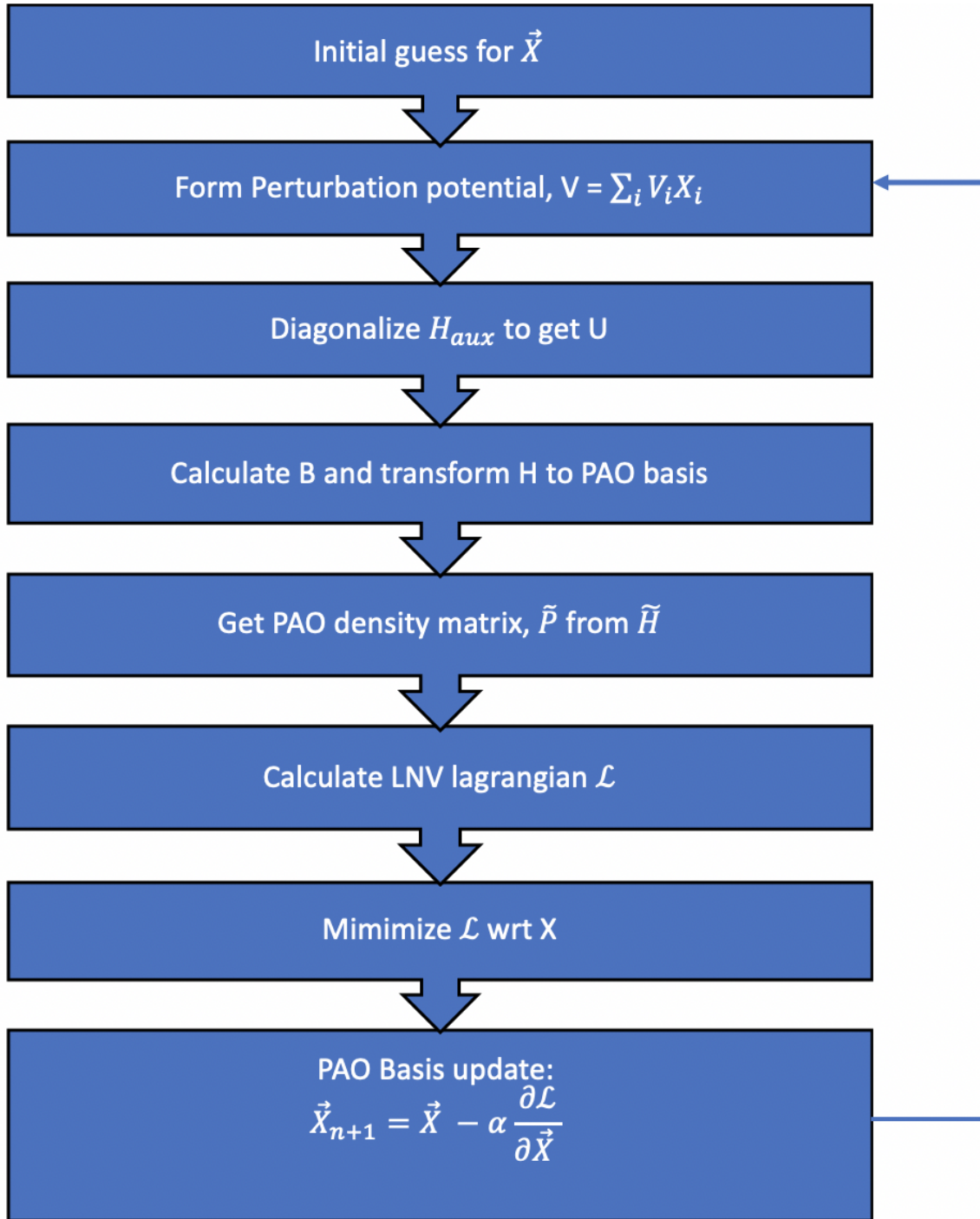


FIG. 3: Potential parametrization - The step which gets taken over by the neural network

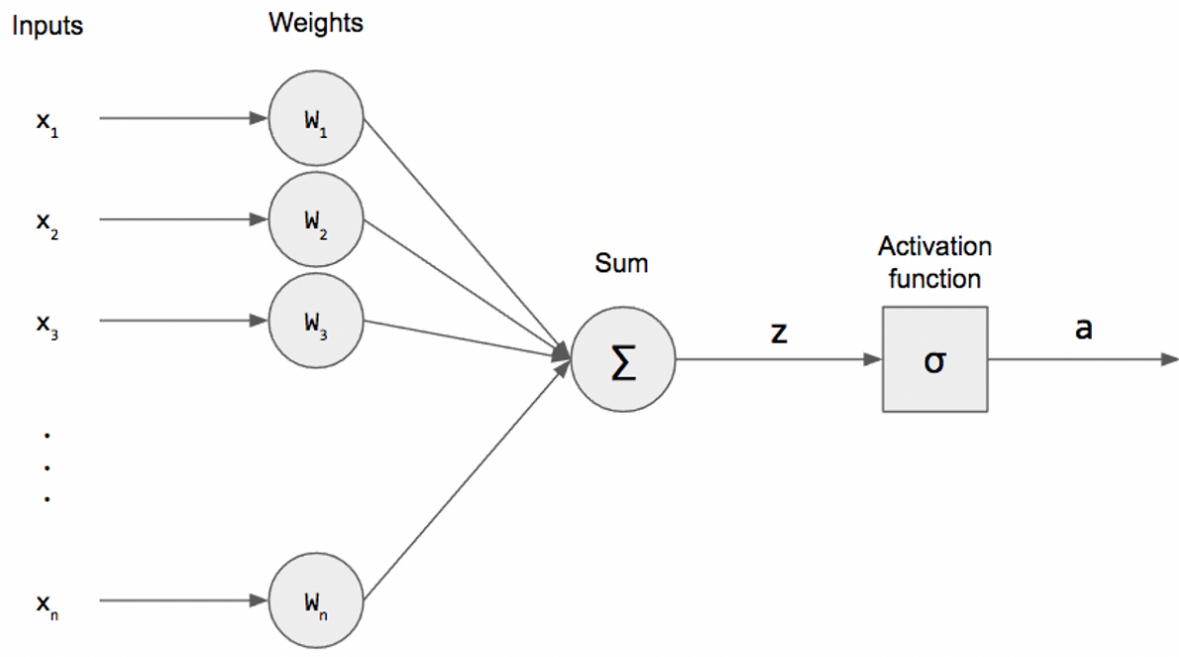


FIG. 4: Outline of a basic perceptron, also referred to as a single neuron

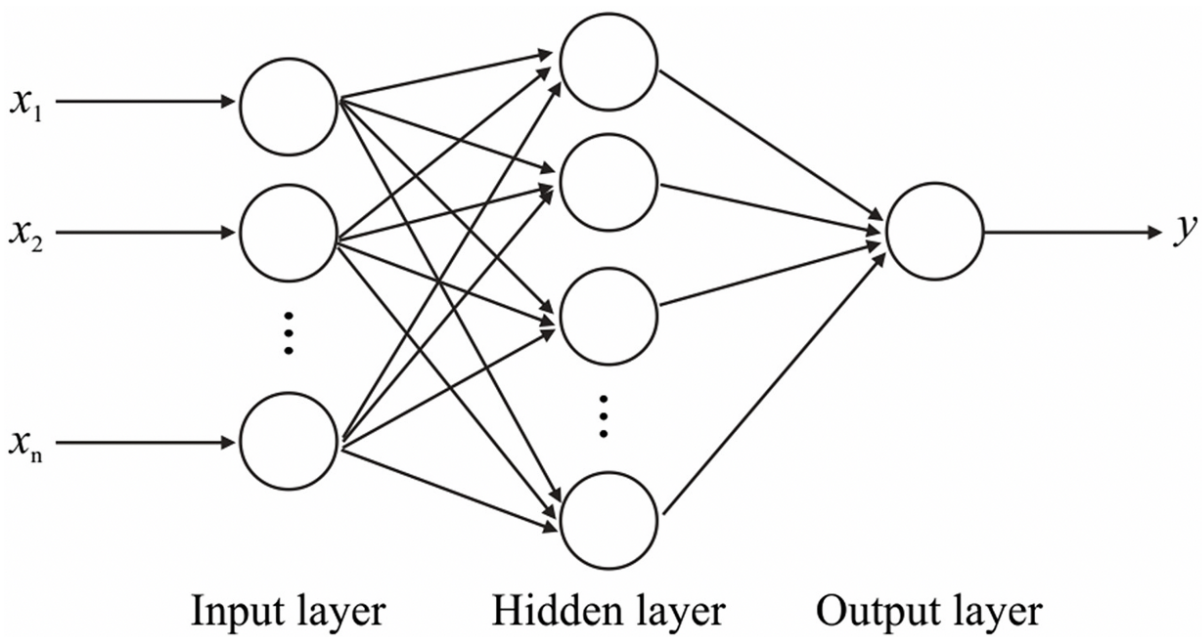


FIG. 5: Multilayer perceptron, i.e., a standard neural network with a single hidden layer

- 
- [1] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. McGraw-Hill, New York, 1989.
- [2] D R Bowler and T Miyazaki. Calculations for millions of atoms with density functional theory: linear scaling shows its potential. *Journal of Physics: Condensed Matter*, 22(7):074207, feb 2010.
- [3] Cynthia J. Jameson, Angel C. de Dios, and A. Keith Jameson. Nuclear magnetic shielding of nitrogen in ammonia. *The Journal of Chemical Physics*, 95(2):1069–1079, 1991.
- [4] Nimrod Moiseyev. *Non-Hermitian Quantum Mechanics*. 2011.
- [5] Brian Prascher, David Woon, Kirk Peterson, Thom Dunning, and Angela Wilson. Gaussian basis sets for use in correlated molecular calculations. vii. valence, core-valence, and scalar relativistic basis sets for li, be, na, and mg. *Theor. Chem. Acc.*, 128:69–82, 01 2011.
- [6] R. S. Mulliken. Criteria for the construction of good self-consistent-field molecular orbital wave functions, and the significance of lcao-mo population analysis. *The Journal of Chemical Physics*, 36(12):3428–3439, 1962.
- [7] William H. Adams. On the solution of the hartree-fock equation in terms of localized orbitals. *The Journal of Chemical Physics*, 34(1):89–102, 1961.
- [8] Michael S. Lee and Martin Head-Gordon. Polarized atomic orbitals for self-consistent field electronic structure calculations. *The Journal of Chemical Physics*, 107(21):9085–9095, 1997.
- [9] Gerd Berghold, Michele Parrinello, and Jürg Hutter. Polarized atomic orbitals for linear scaling methods. *The Journal of Chemical Physics*, 116(5):1800–1810, 2002.
- [10] Ole Schütt and Joost VandeVondele. Machine learning adaptive basis sets for efficient large scale density functional theory simulation. *Journal of Chemical Theory and Computation*, 14(8):4168–4175, 2018. PMID: 29957943.
- [11] Lung Wa Chung, W. M. C. Sameera, Romain Ramozzi, Alister J. Page, Miho Hatanaka, Galina P. Petrova, Travis V. Harris, Xin Li, Zhuofeng Ke, Fengyi Liu, Hai-Bei Li, Lina Ding, and Keiji Morokuma. The oniom method and its applications. *Chemical Reviews*, 115(12):5678–5796, 2015. PMID: 25853797.
- [12] Raghunathan Ramakrishnan and O. Anatole von Lilienfeld. *Machine Learning, Quantum Chemistry, and Chemical Space*, chapter 5, pages 225–256. John Wiley and Sons, Ltd, 2017.

- [13] Pavlo O. Dral. Quantum chemistry in the age of machine learning. *The Journal of Physical Chemistry Letters*, 11(6):2336–2347, 2020. PMID: 32125858.
- [14] Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A.T.B.; Wormit, M.; Kussmann, J.; Lange, A.W.; Behn, A.; Deng, J.; Feng, X., et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. 113:184–215, 2015.
- [15] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [16] Michael Lee and Martin Head-Gordon. 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. *Computers and chemistry*, 24:295–301, 06 2000.
- [17] Michael S. Lee and Martin Head-Gordon. Extracting polarized atomic orbitals from molecular orbital calculations. *International Journal of Quantum Chemistry*, 76(2):169–184, 2000.
- [18] X.-P. Li, R. W. Nunes, and David Vanderbilt. Density-matrix electronic-structure method with linear system-size scaling. *Phys. Rev. B*, 47:10891–10894, Apr 1993.
- [19] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, 1959.
- [20] Jean-Loup Faulon, Donald P. Visco, and Ramdas S. Pophale. The signature molecular descriptor. 1. using extended valence sequences in qsar and qspr studies. *Journal of Chemical Information and Computer Sciences*, 43(3):707–720, 2003. PMID: 12767129.
- [21] M. Karthikeyan, Robert C. Glen, and Andreas Bender. General melting point prediction based on a diverse compound data set and artificial neural networks. *Journal of Chemical Information and Modeling*, 45(3):581–590, 2005. PMID: 15921448.
- [22] Matthew A. Kayala and Pierre Baldi. Reactionpredictor: Prediction of complex chemical reactions at the mechanistic level using machine learning. *Journal of Chemical Information*

- and Modeling*, 52(10):2526–2540, 2012. PMID: 22978639.
- [23] Edward O. Pyzer-Knapp, Kewei Li, and Alan Aspuru-Guzik. Learning from the harvard clean energy project: The use of neural networks to accelerate materials discovery. *Advanced Functional Materials*, 25(41):6495–6502, 2015.
  - [24] C. D. Rankine, M. M. M. Madkhali, and T. J. Penfold. A deep neural network for the rapid prediction of x-ray absorption spectra. *The Journal of Physical Chemistry A*, 124(21):4263–4270, 2020. PMID: 32369378.
  - [25] Janis Timoshenko and Anatoly I. Frenkel. “inverting” x-ray absorption spectra of catalysts by machine learning in search for activity descriptors. *ACS Catalysis*, 9(11):10192–10211, 2019.
  - [26] Matthew R. Carbone, Mehmet Topsakal, Deyu Lu, and Shinjae Yoo. Machine-learning x-ray absorption spectra to quantitative accuracy. *Phys. Rev. Lett.*, 124:156401, Apr 2020.
  - [27] C. Van Der Malsburg. Frank rosenblatt: Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. Berlin, Heidelberg, 1986. Springer Berlin Heidelberg.
  - [28] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
  - [29] Jörg Behler. Constructing high-dimensional neural network potentials: A tutorial review. *International Journal of Quantum Chemistry*, 115(16):1032–1050, 2015.
  - [30] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.
  - [31] N.S. Phala and E. van Steen. Intrinsic reactivity of gold nanoparticles: Classical, semi-empirical and dft studies. *Gold Bull*, 40:150–153, 2007.
  - [32] Alexander V. Vorontsov and Panagiotis G. Smirniotis. Semiempirical computational study of oxygen vacancies in a decahedral anatase nanoparticle. *International Journal of Quantum Chemistry*, 119(5):e25806, 2019.
  - [33] Anders S. Christensen, Tomáš Kubař, Qiang Cui, and Marcus Elstner. Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. *Chemical Reviews*, 116(9):5301–5337, 2016.
  - [34] Fernanda Duarte, Beat A. Amrein, David Blaha-Nelson, and Shina C.L. Kamerlin. Recent advances in qm/mm free energy calculations using reference potentials. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1850(5):954 – 965, 2015.
  - [35] Hao Hu and Weitao Yang. Development and application of ab initio qm/mm methods for

- mechanistic simulation of reactions in solution and in enzymes. *Journal of Molecular Structure: THEOCHEM*, 898(1):17 – 30, 2009.
- [36] Maristella Alessio, Denis Usvyat, and Joachim Sauer. Chemically accurate adsorption energies: Co and h<sub>2</sub>o on the mgo(001) surface. *Journal of Chemical Theory and Computation*, 15(2):1329–1344, 2019.
- [37] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular electronic structure theory*. Wiley & Sons, 2000.