Proceedings of the National Conference On Undergraduate Research (NCUR) 2014 University of Kentucky, Lexington, KY April 3-5, 2014

# The Use of Natural Orbitals in Predicting Molecular Properties

Evan Jahrman Department of Chemistry Gonzaga University 502 East Boone Avenue Spokane, Washington 99258 USA

Faculty Advisor: Dr. Gergely Gidofalvi

# Abstract

Pragmatic modeling of a chemical system requires a method that will produce results of desirable accuracy while avoiding excessive computational cost. This process begins by specifying the size and type of atomic orbital basis set to be used. In general, as the number of orbitals increases so does the accuracy of the computed energies. Unfortunately, this increase in basis set size comes with severe memory and time requirements. Yet it is possible to truncate the number of orbitals and defray costs, while retaining a high level of accuracy, if an appropriate type of orbital is chosen. While it seems appropriate to choose a molecular orbital basis that is optimized with regards to energy considerations, their slow convergence when modeling the instantaneous repulsion between electrons makes them effete for truncation purposes. As an alternative, natural orbitals have enjoyed considerable popularity. Of course, before natural orbitals can be used to model a chemical system, they must first be generated. While expensive in the exact case, an approximate one-electron reduced density matrix can be quickly constructed with the aid of multi-reference Møller-Plesset perturbation theory. Once the density matrix is diagonalized, the acquired eigenvectors represent the desired natural orbitals and each associated eigenvalue corresponds to the probability of finding an electron in that orbital. These probabilities have the desirable property of providing an intuitive scheme for ordering the orbitals according to their importance in the electronic wave function, making the acquired natural orbitals a logical choice for a one-electron basis as they facilitate the truncation of orbitals. Yet traditional methods for generating natural orbitals are incompatible with the computation of global potential energy surfaces and the prediction of transition state geometries that are difficult to characterize experimentally. The procedure advocated here overcomes this shortcoming by including multiple electronic configurations as references and fixing the number of retained orbitals after considering both occupation numbers and molecular symmetries. The accuracy and robustness of the method is assessed by modeling simple bond cleavage and triatomic reactions. When compared to calculations employing standard (and large) basis sets, this selection scheme yields potential energy surfaces with high accuracy at a significantly reduced computational cost.

# Keywords: Electron Correlation, Natural Orbitals, Electronic Structure

# **1. Introduction**

The behavior of electrons governs the vast majority of chemical processes. For that reason, a method able to simulate the interactions between electrons beyond the mean-field approximation can be used to accurately predict the properties of novel systems, investigate transition states that are difficult to study in a laboratory setting, or test the efficacy of scientific models by comparing computational and experimental results. Furthermore, correlated methods that incorporate the effects of instantaneous electron repulsion can reliably analyze the various electronic states accessible to a system and provide valuable insight into physical phenomena. A commendable example is the study of the glycine-cytosine base pair by Sauri et al.<sup>1</sup> The authors used a relatively expensive correlated method to

produce a potential energy surface (PES) for the system in vacuo before complementing the calculation with quantum and molecular mechanical methods to simulate the expected biological environment. These techniques aided in the proposal, analysis, and comparison of three competing proton/hydrogen transfer mechanisms. This is one example illustrating the utility of molecular modeling, although the theories and programs implemented by Sauri et al. are applicable to a variety of other systems and subfields of chemistry. In addition, the method employed above is only one of the numerous computational techniques that electronic structure theorists have developed to determine chemical properties of molecules. These methods are designed to meet the needs of chemists; the archetypical example is the demand for greater accuracy that led to new, correlated methods with heavy computational requirements. While this increase in accuracy is desirable, it necessitates the investigation of procedures for defraying costs. For that reason, the worth of new methods is judged in terms of accuracy and efficiency, yet care should be taken in how these two criteria are evaluated.

The computational efficiency of a given method is assessed according to its time and memory requirements, whereas accuracy is often evaluated through the variational principle. For a variational method, any approximate solution will be an energetic upper bound to the exact solution<sup>2</sup>. In the context of electronic structure methods, this means that as an atomic orbital (AO) basis set approaches completeness, infinite size, and a more accurate method is employed, the energy of the approximate solution converges to the exact energy from above. As the system's energy is of interest, the electronic Schrödinger equation must be used, but it is important to note that its solutions can take the form of either wave functions or electron densities<sup>3</sup>. The latter alternative provides the rationale for density functional theory, which has become a popular choice for the modeling of otherwise prohibitively large systems. Unfortunately, even some of density functional theory's most popular functionals rely on empirical parameters<sup>4</sup> and can obscure important insights into a system's electronic structure and bonding<sup>5</sup>. Unlike density functional theory, the accuracy of wave function methods can be systematically improved by using more sophisticated models for electron correlation and larger basis sets that allow for greater flexibility in the orbitals, both of which entail an increase in computational cost. The purpose of the present work is to show how the bottle neck associated with the AO basis set size can be partially circumvented, while still acquiring accurate results, by transforming the virtual orbitals employed in a correlated calculation.

The tabulated AO basis sets employed in conventional applications are obtained by minimizing atomic energies and offer a compact representation for the electronic structure of atoms. However, their use in molecular modeling introduces redundancies and AO basis sets are not optimal (from the perspective of being compact) for the description the electronic structure of molecules. This means that the one-electron (orbital) basis is unnecessarily large. If the basis can be altered such that the majority of the electronic information is held by relatively few orbitals, then those of less consequence can be removed with impunity; nevertheless, an appropriate criterion must first be selected to order the orbitals before a truncation can take place. In this work, natural orbitals (NOs) are advocated because they are eigenvectors of the one particle reduced density matrix. The corresponding eigenvalue of an NO, referred to as its occupation number, is interpreted as the probability of that orbital being occupied. For a twoelectron system, NOs have been shown to outperform any other orbital basis<sup>6</sup>. Although no such proof has been given for many-electron systems, it is reasonable to suspect that the NOs of lowest probability provide the least contribution to the electronic wave function and their elimination allows for the bypassing of inconsequential configurations. This expectation is further motivated by the Eckart-Young theorem<sup>7</sup>, which states that approximating a matrix by reducing its rank incurs the least error when the smallest singular value of its decomposition is made zero. In the case of the density matrix, the singular values of its factorization are the occupation numbers. Therefore, a truncation of the set of unoccupied orbitals, also referred to as the virtual space, by removing the NOs of lowest occupation promises to be an effective approach and has already been implemented in single point calculations<sup>8-9</sup>. For applications that aim to describe molecular properties at the equilibrium geometry or to determine reaction energies by taking differences in the energies of products and reactants, eliminating NOs with occupation numbers below a user-specified threshold significantly reduces computational costs and provides reliable results<sup>5</sup>.

Nonetheless, there are several applications, such as the description of transition states, that necessitate the accurate modeling of the entire PES, and a simplistic truncation scheme based on occupation numbers alone is inadequate in these contexts. The reason for this is that the occupation of an orbital depends on the underlying nuclear coordinates; therefore, both the number and symmetries of retained orbitals are liable to change along a surface. Discontinuities in the basis prevent the generation of a smooth surface and suggest the presence of an unphysical force. To alleviate this problem and keep the discontinuity in the energy nearly negligible, the occupation threshold could be kept very small; however, such an approach is not desirable as it results in increased computational cost. A more pragmatic approach is to retain a set number of virtual orbitals; if point group symmetry is considered, this implies that the number of virtual orbitals for each irreducible representation (IRREP) of the point group is kept constant. While not common, previous work indicates that such a truncation scheme provides a basis for correlated calculations that is

superior to alternatives of the same size<sup>10</sup>. By supplementing this procedure with an orbital-following protocol that utilizes the overlap between NOs at adjacent molecular geometries, it is possible to keep the number and character of retained NOs uniform across a smooth PES.

In this work, the discussion of the underlying theory and equations for the construction of NOs is followed by a discussion of our specific implementation in GAMESS<sup>11</sup>. The paper is concluded by applications of the proposed truncation method to systems with challenging electronic structures. The relatively small size of the chosen examples allows for comparisons between the results with and without truncation and enables the quantitative assessment of the truncation method's utility. Nonetheless, the method presented here is quite general and offers a promising approach for decreasing the computational burdens associated with large-scale applications.

# 2. Theory

# 2.1. Essential Concepts

Although the electronic Schrodinger equation has been known for almost a hundred years, the interactions between electrons render the computation of exact solutions impractical and made necessary the development of several methods to compute approximate solutions. The simplest ab initio methods, such as Hartree-Fock theory, rely on the assumption that the wave function can be expressed as a single configuration in which the electrons occupy the lowest-energy orbitals. A consequence of this is that the single-configuration approximation yields an effective oneelectron Hamiltonian (the Fock operator) where each electron interacts with the average electric field generated by the remaining N-I electrons, where N is the total number of electrons in the molecule. This approximation is adequate for most molecules near their equilibrium geometries where there is a clear energy separation between the occupied and unoccupied orbitals. However, there are several scenarios, including transition metal compounds and bond cleavage, where the single-configuration approximation breaks down. As an example, consider the homolytic dissociation of molecular hydrogen in the gas phase. It can be shown<sup>12</sup> that the Hartree-Fock wave function, described by the  $\sigma_{g}^{2}$  configuration, introduces ionic terms into the wave function at the dissociation limit where the  $\sigma_g$  and antibonding  $\sigma_u$  orbitals become degenerate. To provide a qualitatively correct description of the wave function and eliminate the ionic contamination requires the inclusion of a second,  $\sigma_u^2$ , configuration into the wave function. The inability of a single configuration to adequately characterize the wave function arises in a variety of electronic structure contexts and is known as static correlation. To correctly capture static correlation effects, a multi-configurational approach is needed to address near degeneracies between configurations. Similar to Hartree-Fock methods, these techniques minimize the energy by varying orbital parameters according to a mean-field approximation; unlike Hartee-Fock theory, the possible configurations are also parameters in the wave function<sup>2</sup>.

Individual selection of the chemically significant configurations requires insight and experience for even the simplest cases, and, as systems become larger, quickly becomes untenable. Alternatively, the complete active-space self-consistent field (CASSCF) method provides a means to substitute knowledge of the important configurations for that of important orbitals. This is achieved by partitioning the orbitals into three classes (Figure 1). The first class contains the low-energy doubly-occupied orbitals, while the third set of orbitals, denoted virtual orbitals, consists of the high-energy orbitals that are unoccupied in the CASSCF wave function. The active space contains those orbitals that are expected to be important in describing the multi-configurational character of the wave function, and is denoted ( $N_{\alpha}$ ,  $n_a$ ), where  $N_a$  and  $n_a$  are the number of electrons and orbitals, respectively. For bond cleavage, the active space must contain the bonding and antibonding orbitals corresponding to the bonds being broken/formed, however, certain cases require the expansion of the active space size beyond this minimum. Larger active spaces yield greater accuracy, but also lead to significant increases in computational cost as every permissible permutation of the active electrons among the active orbitals is included in the CASSCF wave function. Thus, to improve upon the CASSCF results requires alternative strategies.

To include the dynamic correlation effects associated with the instantaneous repulsion between electrons, correlated multi-reference methods augment the CASSCF wave function by including excited configurations obtained by promoting electrons from active to virtual orbitals. The configurations may be classified according to their excitation level relative to the CASSCF wave function, which specifies the number of promoted electrons; for example, the configuration obtained by promoting three electrons from active to virtual orbitals would be classified as a triply-excited configuration. Note that the number of configurations with *m* electrons promoted to virtual orbitals scales as  $O(n_v^m)$ , where  $n_v$  is the number of virtual orbitals. Since the Hamiltonian is a two-electron operator, only singly- and doubly-excited configurations interact directly with the reference CASSCF wave function, and

most correlated multi-reference methods, such as multi-reference configuration interaction with single and double excitations (MRCISD), only include these configurations<sup>2</sup>. These multi-reference correlated methods produce highly accurate PESs<sup>13</sup> but are computationally expensive. For a given active space size, the scaling of the memory requirements and computational cost of MRCISD ( $O(n_v^2)$ , and  $O(n_v^4)$ , respectively) limit its applicability to medium-sized AO basis sets. Nevertheless, the applicability of MRCISD and other multi-reference methods can be extended by reducing the dimensionality of the virtual space. In molecular applications, the underlying AO basis is transformed into a molecular orbital (MO) basis before attempting truncations. These MOs are also often referred to as canonical orbitals since they are eigenvectors of the Fock operator and are ordered according to increasing orbital energy. Based on arguments from perturbation theory, a simplistic approach is to discard the high-energy CASSCF MOs prior to a correlated calculation; however, as will be shown in section 3.1, this approach is generally unsuccessful. On the other hand, an improved truncation criterion arises from transforming the AO basis to NOs.

# 2.2. Approximate Natural Orbitals

The transformation of virtual orbitals from atomic to canonical or natural orbitals is validated by the fact that any set of vectors related to an existing basis set by a unitary transformation corresponds to an equivalent choice of basis. However, obtaining this transformation matrix for constructing the NOs requires the one-electron reduced density matrix, which in the exact case requires foreknowledge of the exact electronic wave function. The challenge, then, is to construct an approximate wave function with a method that captures the important electronic features of the system, has low computational cost, and is sufficiently robust for use in a range of chemical systems. The long history of NOs in configurational interaction methods<sup>6</sup> suggests MRCISD could be used for that purpose, however, prior work in our group demonstrates that the construction of such an accurate density matrix is unwarranted. Instead, the method presented in this paper uses a significantly cheaper method based on second-order perturbation theory to construct an approximate density matrix. Using a Møller-Plesset (MP) type expression, the amplitudes  $(t_{ii}^{ab})$  of doubly-excited configurations<sup>14</sup> are given by

$$t_{ij}^{ab} = -\frac{(ai|bj)}{\varepsilon_a - \varepsilon_i + \varepsilon_b - \varepsilon_j},\tag{1}$$

where indices *i* and *j* denote occupied orbitals, and virtual orbitals are denoted with indices *a* and *b*. An example of such a doubly-excited configuration is shown in Figure 1.  $\varepsilon_k$  is the orbital energy for orbital *k*, and (ai|bj) represents a two-electron integral<sup>2</sup> related to the repulsion between electrons. Given the MP-type amplitude expression in Eq. (1), an element of the virtual-virtual block of the variational density matrix is given by

$$D_{ab}^{(2)} = \sum_{ijc} t_{ij}^{ac} (2t_{ij}^{cb} - t_{ji}^{cb}).$$
<sup>(2)</sup>

Due to the expression of the amplitudes in Eq. (1), an important consequence is that both orbital energies and electron-electron repulsions are incorporated in the construction of the density matrix according to Eq. (2). Thus, its eigenvectors, the NOs, offer a more sophisticated parameterization of the virtual space than conventional (canonical) MOs. From Eq. (2) it is apparent that constructing the virtual block of the density matrix scales as  $O(n_o^2 n_v^3)$ , where  $n_o$  is the number of occupied orbitals. As the scaling is less severe than that of computing the MRCISD wave function, construction of  $\mathbf{D}^{(2)}$  is not expected to significantly increase the computational effort. Nonetheless, for large molecules with many occupied orbitals, Eq. (2) can become demanding. To alleviate this, we decompose the summation over *i* and *j* into two sums, symbolically denoted as  $\Sigma_{ij} \rightarrow \Sigma_i + \Sigma_{i\neq j}$ . The first, restricted sum (*i=j*), only includes those terms that correspond to double excitations from the same occupied orbital, whereas the second sum only considers those amplitudes that result from promoting electrons from different occupied orbitals. Neglecting the second summation yields an approximation for computing the density matrix

$$\widetilde{D}_{ab}^{(2)} = \sum_{ic} t_{ii}^{ac} t_{ii}^{cb}.$$
(3)

Since Eq. (3) only incorporates a restricted class of configurations, we will refer to Eq. (3) as the restricted approximation. Note that, in addition to scaling linearly with the number of occupied orbitals, computing  $\tilde{\mathbf{D}}^{(2)}$  requires the evaluation of a significantly smaller number of two-electron integrals in Eq. (1). Clearly the form in Eq. (3) is preferable and has been implemented in computational suites such as MOLCAS<sup>5</sup>. Despite its utility, to the best of our knowledge, the accuracy of the restricted approximation has not been investigated in the literature but will be tested here.



Figure 1. definition of orbital classes in the CASSCF and MRCISD wave functions. The doubly-excited configuration formed by an excitation from occupied orbitals i/j to virtual orbitals a/b is shown by grey lines.

#### 2.3. Details Of Implementation

All relevant subroutines were written in the Fortran90 programming language and interfaced with GAMESS. The NOs are computed by diagonalizing the density matrix using LAPACK<sup>15</sup> subroutines. In the current implementation, the use of NOs is supported for the MRCISD and second-order multi-reference Møller-Plesset perturbation (MRMP2) subroutines of GAMESS, although it can be interfaced with other correlated methods as long as canonical orbitals are available. Abelian point-group symmetry is enforced throughout the construction of the density matrix. Therefore, the density matrix is block-diagonal and only those elements for which virtual orbitals *a* and *b* belong to the same IRREP need to be calculated. Enforcing point-group symmetry has several advantages. The block-diagonal form prevents the mixing of orbitals with different symmetries and reduces the time associated with diagonalizing the density matrix. Once the NOs are constructed, the virtual space is truncated according to user-specifications. Although the selection scheme based on occupation threshold is supported in our implementation, most of the results shown in Section 3 utilize the truncation scheme based on the number of orbitals. For molecules that belong to a point group with more than one IRREP, this requires the number of orbitals per IRREP to be specified. One and two-electron integrals in the NO basis are computed from the AO integrals with the internal transformation subroutines of GAMESS.

At the first nuclear geometry of the PES, usually the equilibrium geometry, the set of NOs to be retained is decided solely by the user's input and the employed selection scheme. In order to ensure the character of the NOs varies smoothly along the PES, an orbital-following scheme based on the overlap between NOs  $\phi_{\alpha}$  and  $\phi_{\beta}(S_{\alpha,\beta})$ 

$$S_{\alpha,\beta} = \sum_{ij} c^*_{\alpha,i} S_{i,j} c_{j,\beta} \tag{4}$$

has been implemented.  $S_{i,j}$  is the overlap between AOs *i* and *j* at the different molecular geometries. Assume that, at some reference geometry of the PES, the set of orbitals,  $\phi_{\alpha}$ , included in the correlated calculation is available. At the next nuclear conformation, the full set of NOs { $\phi_{\beta}$ } is computed by diagonalizing the density matrix. For each  $\phi_{\alpha}$ , its

overlap with all NOs from the previous geometry is computed, and the NO to be included is obtained by selecting  $\phi_{\beta}$  for which  $S_{\alpha,\beta}$  is largest. This process maximizes the similarity between the set of NOs, and repeating this selection scheme at each point along the PES ensures that the character of the NOs included in the correlated calculation is uniform across the PES.

#### **3. Results and Discussion**

### 3.1. Accuracy Of Approximate Natural Orbitals

For a truncation criterion to be effective, it must be capable of transforming the orbital basis set such that a substantial percentage of the virtual space can be removed with little loss of accuracy. One way of assessing this attribute is to run a series of calculations in which the underlying chemical system,  $N_2$  for the example shown in Figure 1, remains unchanged, but the number of virtual orbitals in the correlated calculation is gradually reduced. If the resultant energy converges to the full basis limit with only a small percentage of the virtual space, then the criterion demonstrates promise for general implementation. As stated previously, truncating the virtual space in canonical form is not ideal, but a comparison between energy-based and occupation-based truncations is insightful.



Figure 2. MRMP2 energies for N<sub>2</sub> relative to the full basis are shown. Diamonds correspond to points using the restricted approximation and circles to the unrestricted. The canonical results are represented by triangles.

Using molecular nitrogen as an example, the MRMP2 energy was calculated at the equilibrium bond distance<sup>16</sup> using the cc-pVQZ<sup>17-18</sup> AO basis set. The (6,6) active space consists of the bonding and antibonding orbitals formed from the 2p nitrogen AOs, and all valence electrons are correlated in the MRMP2 calculations. The NO results (restricted and unrestricted) are obtained by systematically increasing the occupation threshold to reduce the size of the virtual space. In contrast, calculations performed in the canonical basis employ the lowest-energy orbitals such that the number of virtual orbitals in each IRREP is the same as in the unrestricted NO calculations. It should be noted that, regardless of whether natural or canonical orbitals are used in the virtual space, the doubly-occupied and active orbitals are the same in all calculations.

Both sets of orbitals, natural and canonical, approach the same energy in the limit that all virtual orbitals are retained in the correlated calculation (Figure 2). Since the MRMP2 energy is invariant to rotations between active orbitals, and the NOs are obtained from the canonical ones by such rotations, the limiting behavior of the curves is as expected. The important distinction between the orbital bases is that the occupation criterion provides a much more rapid and smoother convergence than the energy criterion. For instance, with only fifty percent of the virtual space retained, the error introduced by the NOs has dropped to roughly a quarter of the error incurred by the canonical alternative. The behavior at the fifty percent mark is of special significance as the number of orbitals retained corresponds to roughly the same number of orbitals as the cc-pVTZ<sup>17-18</sup> AO basis. In addition, there appears to be little difference in the convergence provided by the NOs constructed using the unrestricted **D**<sup>(2)</sup> from Eq. (2) and the restricted **D**<sup>(2)</sup> from Eq. (3). The same trend was observed in other systems (results not shown here),

indicating that the use of  $\tilde{\mathbf{D}}^{(2)}$  to compute approximate NOs yields accurate results. Thus, the remaining applications presented in this work utilize approximate NOs.

## 3.2. Accuracy Of Ground-State PESs

There are several NO-based truncation schemes to ensure that the dimension of the virtual space is constant along the PES. For example, one could determine the number of virtual orbitals with occupation numbers above some threshold for the *entire* PES, and then select the dimension of the virtual space such that no NO with an occupation number above the threshold is excluded from the correlated calculation of the PES. A major shortcoming of such an approach is that it presumes knowledge of the *entire* set of molecular geometries that need to be considered in the correlated treatment. This is often not the case, such as in *ab-initio* on-the-fly dynamics simulations where the relevant geometries are not known ahead of time. The systematic truncation scheme presented here alleviates this by selecting the dimension of the virtual space without using a threshold based on occupation. In this scheme, the NOs are computed in a large standard AO basis set (cc-pVQZ in this work), and the dimension of the virtual space is reduced to that of a smaller AO basis set (cc-pVTZ in this work). To clarify the discussion, the convention X/Y denotes a calculation that employs the standard AO basis set X and restricts the dimension of the virtual space to a size equivalent to that of a smaller AO basis set denoted Y. Thus, a calculation that employs the cc-pVQZ basis set and does not truncate the virtual space prior to a correlated calculation is denoted QZ/QZ, while the truncation scheme adopted in this work is denoted QZ/TZ.

To demonstrate the utility of the proposed truncation scheme, the ground-state PES of hydrogen fluoride at the MRCISD level of theory was computed. Starting from  $R_{HF}$ =0.8Å, the PES was computed on the interval  $0.6 \le R_{HF} \le 3.0$ Å in 0.1Å increments. The (6,6) active space consists of the 3-4a<sub>1</sub>, 1-2b<sub>1</sub>, and 1-2b<sub>2</sub> orbitals and all valence electrons are correlated at the MRCISD level. The small size of the system allows the quantitative comparison of the truncated calculations and calculations in the full virtual space. As expected for a variational method, the TZ/TZ (36 virtual orbitals) energies are above the QZ/QZ (77 virtual orbitals) energies, and the QZ/TZ (36 virtual orbitals) are in significantly better agreement with the full QZ/QZ results (Figure 3a). Relative to the QZ/QZ PES, the maximum errors for the TZ/TZ and QZ/TZ PESs are 34.2 mH and 12.7 mH, respectively.



Figure 3. MRCISD PESs for hydrogen fluoride (a) calculated with the QZ/QZ (dash), TZ/TZ (dot), and QZ/TZ (solid) bases. The deviation of the QZ/TZ results from the full reference are shown (b) with (triangle) and without (diamond) orbital following.

Although the QZ/TZ PES in Figure 3a appears to be smooth to a cursory investigation, a close inspection (Figure 3b) indicates the presence of unphysical discontinuities. These discontinuities are undesirable as they can adversely affect molecular properties, such as vibrational frequencies, that depend on derivatives of the energy with respect to nuclear coordinates. Although the dimension of the virtual space is restricted to be constant in the present scheme, the character of the NOs in the virtual space may change since the NOs are ordered according to descending

occupation numbers. Thus, the character of orbitals in the virtual space may change if the occupation numbers of retained and eliminated NOs switch between two nearby geometries. Indeed, the drops in the error of the QZ/TZ energies at 1.4 and 2.6 Å can be traced back to such artificial orbital switches. To eliminate these artifacts, the orbital-following procedure outlined in Section 2.3 was applied with the initial set of NOs determined near the equilibrium bond length (R=0.9 Å). Although the maximum error in the QZ/TZ is not affected by the orbital-following protocol, the corrected surface is smoother and (Figure 3b) the non-parallelity error, defined as the difference between the maximum and minimum error relative to the QZ/QZ PES, is decreased by nearly a factor of two from 4.7 mH to 2.5 mH.

The data in Table 1 demonstrate the computational improvements offered by the proposed truncation scheme. Since the dimension of the virtual space is the same for the TZ/TZ and QZ/TZ calculations, the nearly identical times for computing the PES indicates that construction of the NOs incurs marginal additional computational cost. At the same time, due to the more flexible AO basis set, the QZ/TZ results are more accurate as evidenced by the smaller non-parallelity errors. Relative to the QZ/QZ calculations, both the TZ/TZ and QZ/TZ computations require an order of magnitude less time due to the relatively steep scaling ( $O(n_v^4)$ ) of the computational effort of MRCISD with the number of virtual orbitals. Additionally, because the memory requirement of MRCISD is on the order of  $O(n_v^2)$ , the truncated calculations also offer the possibility of treating larger molecular systems.

Table 1. the time requirements and accuracy of hydrogen fluoride surfaces calculated with different bases at the MRCISD(6,6) level of theory.

	Time	NPE
Basis	(hours)	(mH)
TZ/TZ	3.4	5.1
QZ/TZ	4.0	2.5
QZ/QZ	41.9	-

#### 3.3. Accuracy Of Excited-State PESs

The formation of beryllium hydride by means of a  $C_{2v}$  insertion is dependent upon two parameters (Figure 4a), the distance from beryllium to the center of mass of the hydrogens ( $R_{Be-H_2}$ ) and the distance between the hydrogen atoms ( $R_{H-H}$ ). However, the generation and analysis of a PES is greatly simplified by introducing an implicit dependence between the two parameters as in the work of Purvis et al.<sup>19</sup> To establish the dependence of  $R_{H-H}$  on  $R_{Be-H_2}$ , the equilibrium geometry for collinear BeH<sub>2</sub> and H<sub>2</sub>, as well as the transition state are optimized at the CASSCF level of theory with a (4,6) active space that includes the 2-4a<sub>1</sub>, 1b<sub>1</sub>, and 1-2b<sub>2</sub> orbitals. The reaction pathway is then defined as two straight line segments in terms of  $R_{H-H}$  and  $R_{Be-H_2}$ (Figure 5a). The first line segment passes through the collinear equilibrium geometry and the transition state geometry. For  $R_{Be-H_2}>1.647$  Å, the second line segment is a vertical line. Although this PES does not correspond to the minimum energy  $C_{2v}$  path which would require repeated geometry optimizations, it does capture the complicated electronic structure near the transition state geometry.



Figure 4. In the reaction coordinate diagram (a) for  $BeH_2$ , the dark lines represent the reaction pathway, while the dashed lines are extensions of the vertical line segment and the line segment connecting the transition state to the product. PESs of the ground (diamonds) and first excited state (squares) of  $A_1$  symmetry are shown (b) where the QZ/TZ results are represented by symbols and the QZ/QZ by lines.

To assess the proposed truncation scheme's utility in simultaneously modeling multiple electronic states, stateaveraged CASSCF calculations, with equal weights for the three lowest-energy  ${}^{1}A_{1}$  states, were performed to construct the NOs. The active space is the same as mentioned previously, and all electrons are correlated in the MRCISD calculations. The avoided crossing near the transition state geometry between the two lowest-energy  ${}^{1}A_{1}$ states is apparent in Figure 4b. Despite the complex nature of the electronic structure problem, the QZ/TZ truncation scheme accurately captures the qualitative features of the PES for both the ground and excited states. The quantitative agreement with the QZ/QZ results even without orbital-following is encouraging; the NPE for the ground- and excited-state PES is 3.8 and 3.4 mH, respectively.

# 4. Conclusion

The ultimate goal of electronic structure theory is to accurately model chemical systems and predict their properties. In many cases, this requires the computation of a global PES. Although multi-reference methods accurately capture the features a molecular PES, the steep computational scaling of these methods with the number of virtual orbitals limits their applicability to medium-sized molecules. Thus, to expand the applicability of these methods to larger molecules and establish electronic structure theory as a viable predictive tool, the development of novel methods with improved scaling properties or the implementation of approaches that employ a small fraction of the full virtual orbital space with minimal loss of accuracy is necessary.

For the latter purpose, NOs have shown considerable promise. Indeed, calculations with a limited number of NOs show that the computed energies converge to the exact results much faster than do calculations employing canonical orbitals. Thus, the use of NOs offers a method for eliminating a possibly large fraction of the virtual space while introducing only marginal errors. Furthermore, since computing the MP-type density matrix and the NOs does not require a significant amount of time, these truncation schemes allow for the computation of accurate energies at a significantly reduced cost. Inspired by ideas related to the hierarchy of standard AO basis sets in the literature, we propose a relatively aggressive truncation scheme. The essential idea is to compute the virtual orbitals in a large AO basis set. For the basis sets employed in this study, the truncation eliminates approximately fifty percent of the virtual orbitals.

The results presented here indicate that NOs are amenable to the generation of smooth global PESs by fixing the number of virtual orbitals and using an orbital-following protocol to ensure that the character of the virtual orbitals is continuous along the PES. Timings indicate that the construction of the NOs does not require significant additional computational effort. In the case of hydrogen fluoride, as measured by the NPE, the QZ/TZ surface's NPE is almost a factor of two better than the PES computed using a standard AO basis set with the same number of virtual orbitals. Applications to model the  $C_{2v}$  insertion path in BeH<sub>2</sub> indicate that truncation scheme reliably

predicts changes in the electronic structure of several electronic states, including the avoided crossing between the two lowest-energy states of  $A_1$  symmetry. Despite the aggressive truncations employed, the PES for the ground and first excited state are accurately reproduced, as shown by the small NPEs. Additionally, as indicated by the nearly indistinguishable results from computations employing the restricted ( $\tilde{\mathbf{D}}^{(2)}$ ) and unrestricted ( $\mathbf{D}^{(2)}$ ) density matrices, the use of approximate NOs is justified.

Taken together, these results are very encouraging; nonetheless, the development of approaches to estimate the energy errors introduced by the truncation is desirable. The error may be established by energy extrapolations similar to those employed in methods for estimating the complete basis set energy. Alternatively, akin to composite methods, the error may be estimated using methods for which computing the energy in the full and truncated virtual space is relatively inexpensive. At this point, it is not clear which estimation method is superior and this remains a topic for ongoing research in our group. Additionally, the current implementation requires the computation of the CASSCF wave function, and as such, requires several iterations. Depending on the size of the active space, the time for these iterations can be a sizable fraction for the total computation time. Since the exact form of the NOs is not crucial, the use of methods that approximate the CASSCF wave function could lead to significant savings in terms of time. The use of improved virtual orbitals of Freed and coworkers<sup>20</sup> is quite promising in this regard since it is not an iterative process, and the cost of computing these orbitals is equivalent to a single iteration of the CASSCF method. Future research will consider their implementation in conjunction with NOs.

# 5. Acknowledgments

This work was supported by an award from the Research Corporation for Science Advancement and a grant to Gonzaga University from the Howard Hughes Medical Institute through the Undergraduate Science Education Program. The author would also like to acknowledge the Chemistry Department of Gonzaga University and Dr. Gidofalvi.

# **6. References**

1. Sauri, V.; Gobbo, J.; Serrano-Perez, J.; Lundberg, M.; Coto, P.; Serrano-Andres, L.; Borin, A.; Lindh, R.; Merchan, M.; Roca-Sanjuań, D. Proton/Hydrogen Transfer Mechanisms in the Guanine–Cytosine Base Pair: Photostability and Tautomerism. *J. Chem. Theory Comput.* **2012**, *1021*. 481-496.

2. Jensen, F. Introduction to Computational Chemistry; Wiley: 2007

3. Townsend, J. S. A Modern Approach to Quantum Mechanics; University Science Books: 2012

4. Daku, M.; Aquilante, F.; Robinson, T.; Hauser, A. Accurate Spin-State Energetics of Transition Metal Complexes. 1. CCSD(T), CASPT2, and DFT Study of [M(NCH)6]2+ (M = Fe, Co) *J. Chem. Theory Comput.* **2012**, *1021*, 4216-4231.

5. Aquilante, F.; Pedersen, T.; Veryazov, V.; Lindh, R. MOLCAS—A Software for Multiconfigurational Quantum Chemistry Calculations. *WIREs Comput. Mol. Sci.* **2013**, *3*, 143-149.

6. Lowdin, P. Quantum Theory of Many-Particle Systems. I. Physical Interpretations by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction . *Phys. Rev.* **1955**, *97*, 1474-1489.

7. Eckart, C.; Young, G. The Approximation of One Matrix by Another of Lower Rank. *Psychometrika* **1936**, *1*, 211.

8. La Macchia, G.; † Li Manni, G.; Todorova, T.; Brynda, M.; Aquilante, F.; Roos, B.; Gagliardi, L. On the Analysis of the Cr-Cr Multiple Bond in Several Classes of Dichromium Compounds. *Inorg. Chem.* **2010**, *49*, 5216 -5222.

 Roca-Sanjuan, D.; Aquilante, F.; Lindh, R. Multiconfiguration Second-Order Perturbation Theory Approach to Strong Electron Correlation in Chemistry and Photochemistry. *WIREs Comput. Mol. Sci.* 2012, *1002*. 585-603.
 Abrams, M.; Sherrill, C. A Comparison of Polarized Double-Zeta Basis Sets and Natural Orbitals for Full Configuration Interaction Benchmarks. *J. Chem. Phys.* 2003, *118*, 1603-1609.

11. Schmidt, W.; Baldridge, K.; Boatz, J.; Elbert, S.; Gordon, M.; Jensen, J.; Koseki, S.; Matsunaga, N.; Nguyen, K.; Su, S.; Windus, T.; Dupuis, M.; Montgomery, J. GAMESS. J. Comput. Chem.

12. Knowles, P. J.; Schutz, M.; Werner, H. In *AB INITIO METHODS FOR ELECTRON CORRELATION IN MOLECULES*; Modern Methods and Algorithms in Quantum Chemistry; John von Neumann Institute for

Computing: 2000; Vol. 3.

13. Sherrill, C.; Piecuch, P. The X, B, and B' States of C2: A Comparison of Coupled-Cluster and Multireference Methods with Full Configuration Interaction Benchmarks. *J. Chem. Phys.* **2005**, *122*, 1063-1081.

14. Aquilante, F.; Todorova, T.; Gagliardi, L.; Pedersen, T.; Roos, B. Systematic Truncation of the Virtual Space in Multiconfigurational Perturbation Theory. J. Chem. Phys. **2009**, 131. 1063-1069.

Anderson, E.; Bai, Z.; Bischof, C.; Blackford, S.; Demmel, J.; Dongarra, J.; Du Croz, J.; Greenbaum, A.; Hammarling, S.; McKenney, A.; Sorensen, D. LAPACK. Society for Industrial and Applied Mathematics.
 NIST Listing of Experimental Data for N2. <u>http://cccbdb.nist.gov/exp2.asp?casno=7727379</u> (8/9/2013).

17. Dunning, T. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen *J. Chem. Phys.*, **1989**, *90*, 1007-1023.

18. Woon, D.; Dunning, T. Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum Through Argon," J. Chem. Phys., **1993**, 98, 1358-1371.

19. Purvis, G.; Shepard, R.; Brown, F.; Bartlett, R. (1983), C<sub>2</sub>V Insertion pathway for BeH<sub>2</sub>: A Test

Problem for the Coupled-Cluster Single and Double Excitation Model Int. J. Quantum Chem. 1983, 23, 835–845.

20. Chaudhuri, R.; Freed, K. Improved Virtual Orbital Multireference Møller–Plesset Study of the Ground and Excited Electronic States of Protonated Acetylene, C2H3+ . *J. Chem. Phys.* **2008**, *129*, 1063-1069.