Frozen natural orbitals for ionized states within equation-of-motion coupled-cluster formalism

Arie Landau, Kirill Khistyaev, Stanislav Dolgikh, and Anna I. Krylov[a)]
Department of Chemistry, University of Southern California, Los Angeles, California 90089-0482, USA

(Received 11 June 2009; accepted 2 December 2009; published online 6 January 2010)

The frozen natural orbital (FNO) approach, which has been successfully used in ground-state coupled-cluster calculations, is extended to open-shell ionized electronic states within equation-of-motion coupled-cluster (EOM-IP-CC) formalism. FNOs enable truncation of the virtual orbital space significantly reducing the computational cost with a negligible decline in accuracy. Implementation of the MP2-based FNO truncation scheme within EOM-IP-CC is presented and benchmarked using ionized states of beryllium, dihydrogen dimer, water, water dimer, nitrogen, and uracil dimer. The results show that the natural occupation threshold, i.e., percentage of the total natural occupation recovered in the truncated virtual orbital space, provides a more robust truncation criterion as compared to the fixed percentage of virtual orbitals retained. Employing 99%–99.5% natural occupation threshold, which results in the virtual space reduction by 70%–30%, yields errors below 1 kcal/mol. Moreover, the total energies exhibit linear dependence as a function of the percentage of the natural occupation retained allowing for extrapolation to the full virtual space values. The capabilities of the new method are demonstrated by the calculation of the 12 lowest vertical ionization energies (IEs) and the lowest adiabatic IE of guanine. In addition to IE calculations, we present the scans of potential energy surfaces (PESs) for ionized $(H_2O)_2$ and $(H_2)_2$. The scans demonstrate that the FNO truncation does not introduce significant nonparallelity errors and accurately describes the PESs shapes and the corresponding energy differences, e.g., dissociation energies. © *2010 American Institute of Physics*. [doi:10.1063/1.3276630]

I. INTRODUCTION

Proper account of electron correlation,[1] which is crucial for accurate calculations of chemically relevant properties, leads to computational methods that scale steeply with the number of electrons and the size of the one-electron basis set, i.e., the number of occupied (O) and unoccupied [or virtual (V)] orbitals. For example, the scaling of the second-order many-body perturbation theory is O^2V^3, the coupled-cluster method with single and double substitutions (CCSD)—as O^2V^4, and CCSDT (including triples)—as O^3V^5. This limits the applications of these reliable and accurate wave function based methods to moderate-size systems.

To reduce computational costs, the core and sometimes higher virtual orbitals are excluded (frozen) from the correlation calculations. Freezing inner-shell electrons, which can be justified by a relatively small role the core electrons play in chemical transformations, is a common practice, and many standard basis sets are optimized for correlating only the valence electrons.[1,2] Larger computational savings can, in principle, be achieved by freezing significant fraction of the virtual space, e.g., using a fraction f of the full virtual space (FVS) results in a $1/(1-f)^4$-fold speed up for each CCSD iteration. However, simply freezing higher-energy canonical virtual orbitals leads to large errors in correlation energy because of the slow convergence of the dynamical correlation energy with the virtual space size. Fortunately, the virtual orbital space can be reduced without significant loss of accuracy if the orbitals are reordered based on their importance for the total correlation energy. This can be achieved by using the natural orbital (NO) approach.

The NOs are eigenstates of the one-electron density matrix (DM),

$$\gamma_{pq} = \langle \Psi | p^+ q | \Psi \rangle, \tag{1}$$

where Ψ is an N-electron wave function and p^+ and q are creation and annihilation operators for orbitals ϕ_p and ϕ_q, respectively. The trace of the DM equals to the number of electrons and its eigenvalues (which are non-negative) can be interpreted as the occupation numbers associated with the respective NOs. For example, the DM corresponding to a Hartree–Fock (HF) wave function is a diagonal matrix and the respective natural occupations equal one for the occupied spin orbitals and zero for virtuals. A correlated DM, which is not diagonal and has nonzero elements in the OO, OV, as well as VV blocks, yields fractional natural occupation numbers. Thus, natural occupation numbers contain information about correlation, i.e., larger occupation numbers signify larger contributions of the configurations involving these orbitals to the total correlation energy. Consequently, the NOs can be used to represent a correlated wave function in a more compact form,[3] which has been exploited in a variety of approaches. The common strategy is to use a lower-level (or approximate) correlated DM (obtained by a less expensive calculation) to compute the NOs, and then use them in a higher-level correlation method.

[a)]Electronic mail: krylov@usc.edu.

The NOs were introduced more than half a century ago by Löwdin,[3] who suggested that they lead to faster convergence of the configuration interaction (CI) expansion and afford a more compact representation of the correlated CI wave function. He showed that the rapidity of the convergence can be measured by the smallness of a function that goes to zero in the limiting case in which the correlated wave function is represented by a single Slater determinant composed of the "best orbitals." The analysis of the measuring function showed that the NOs yield the smallest number of essential orbitals for CI energies to converge.

Later, Barr and Davidson[4] introduced the frozen natural orbitals (FNOs) defined as follows: the occupied orbitals are *frozen* to their HF values and only the virtual orbitals are transformed. This makes the reference and the correlation energies invariant to the transformation.

The advantages of NOs over HF orbitals were demonstrated quantitatively by Shavitt *et al.*[5] in their comprehensive investigation of the energy convergence for water. Since their introduction in 1955, the NOs have been used extensively in CI calculations.[6–12]

In many CI applications, the NOs were used to achieve more compact representation of the wave function and different energy-based criteria[5–8] were employed to select the most important configurations. Alternatively, one may use NOs to define an active space and include all the configurations (as specified by the correlation method) within this active space.[12–16]

Several procedures of generating NOs have been exploited. Some of the CI applications[5–8] employed perturbation theory criteria to construct a compact CI wave function for computing the NOs, which were used in subsequent CI calculations. Iterative variants of this approach were also explored.[6,8] An alternative strategy is to employ a less computationally demanding method, e.g., Møller–Plesset (MP2), to compute the NOs, select an active space based on their occupation numbers, and use them in a subsequent higher-level calculation. For example, Jensen *et al.*[14] used the MP2 NOs to select an initial guess for complete active space self-consistent field (CASSCF) calculations. Abrams and Sherrill[13] employed single-reference based NOs in CASCI as an alternative to the CASSCF orbitals.

Using the FNO-based truncation in perturbation theory and CC methods was pioneered by Bartlett and co-workers.[15–17] In their recent benchmark study,[15,16] they investigated errors introduced by FNO truncation in the equilibrium geometries computed by the CCSD(T) (Refs. 18 and 19) and Λ-CCSD(T) (Refs. 20–23) methods, and demonstrated that truncating orbital space by 20%–60% introduces errors in bond lengths of 0.2–0.3 pm when using basis sets of cc-pVTZ quality and above.

NOs were also used in developing basis sets for correlated calculations in the so-called atomic-NO (ANO) scheme.[24] The contracted ANO sets are obtained from a correlated atomic uncontracted calculation, and then used in molecular calculations without further modification. These basis sets have demonstrated excellent performance, and using FNOs in molecular calculations can be viewed as an extension of this idea by allowing the basis set to adjust to a specific molecule.

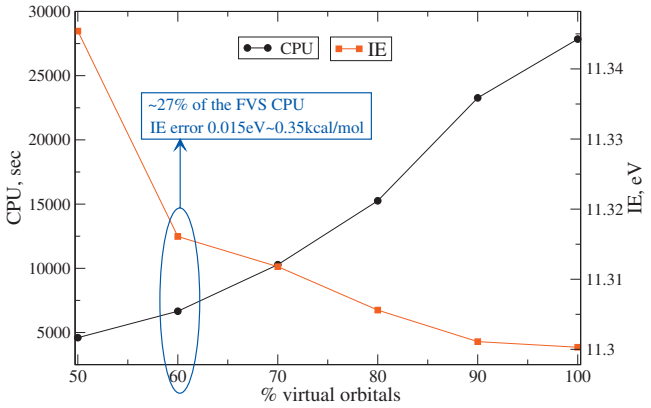

FIG. 1. CPU (black circles, values on the left *y*-axis) and the lowest IE of water dimer (red squares, right *y*-axis) as a function of the percentage of virtual NOs retained in the EOM-IP-CC(2,3)/cc-pVTZ+dff calculation. The calculations were performed using a serial version of Q-CHEM on a 2xXeon EM64T 3.2 GHz machine.

There have been several other suggestions for obtaining better orbitals for correlated calculations.[25,26] In the optimized virtual orbital space (OVOS) method,[25] a Hylleraas functional that constitutes an upper bound to the second-order energy correction is used to optimize a set of active virtual orbitals. Sosa *et al.*[17] showed that the more expensive iterative OVOS performs slightly better than FNO in some cases, while in other situations they found perfect agreement between the two approaches.

Finally, the split-localized orbitals[26] (i.e., localized NOs) were found to provide even faster convergence for CI expansions than the original NOs. The efficiency of the split-localized molecular orbitals (MOs) in generating compact CI expansions is attributed to their localized nature, which can be exploited in reduced-scaling *ab initio* methods.

All the above applications of NOs were performed in a state-specific fashion, and focused on a single state (most often, the ground-state) calculation, except for Ref. 6(b) in which average NOs for the several states were employed to describe the neutral and ionized states of ethylene. This work presents the extension of FNO to a multistate method, i.e., equation-of-motion coupled-cluster for ionized states (EOM-IP-CC). The FNOs are computed for the reference (neutral) state and then are used to describe several target (ionized) states of interest. Quite surprisingly, we found that using reference-state FNOs to truncate the virtual space does not introduce imbalance in the reference CCSD and target (ionized) EOM-IP-CCSD wave functions.

Figure 1 illustrates the potential of the FNO approach using the calculation of the lowest ionization energy (IE) of the water dimer as an example. The calculation is performed using EOM-IP-CC with explicit inclusion of triple excitations in the EOM part [EOM-IP-CC(2,3)], and an augmented cc-pVTZ basis (see Sec. III for details). Figure 1 shows the IE value and the corresponding CPU as a function of the percentage of virtual orbitals (POVO) retained. As the size of virtual space increases, the CPU rises steeply and IE ap-

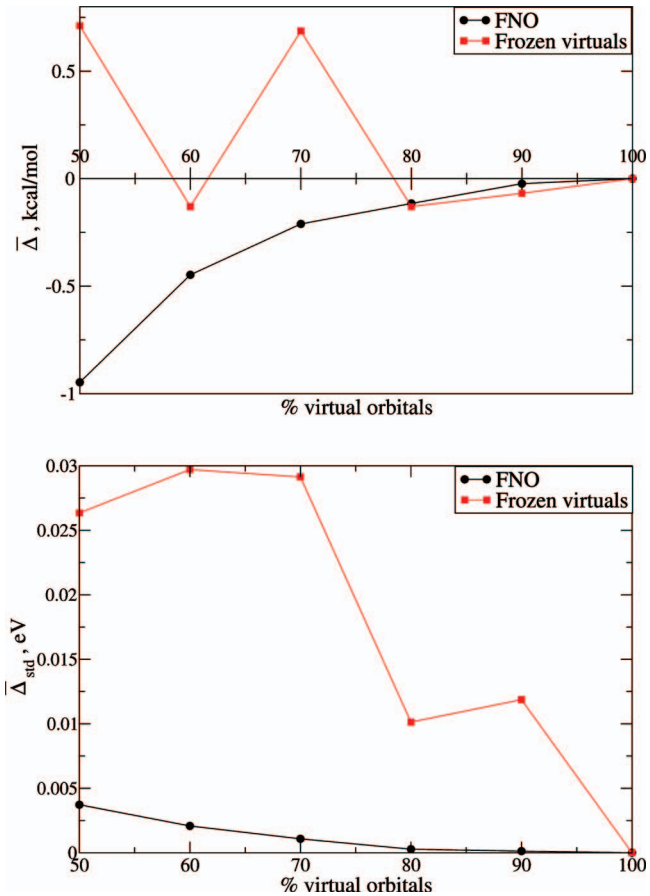

FIG. 2. MEs (upper panel) and STDs (bottom panel) for the six lowest IEs of water dimer at the EOM-IP-CC(2,3)/cc-pVTZ+dff level using FNOs vs frozen virtual canonical HF orbitals.

proaches the FVS value (shown as the 100% point). The calculation using 60% of the virtual orbitals requires only about 27% of the FVS CPU time, while the error in IE (relative to the FVS value) is less than 0.5 kcal/mol. Thus, considerable computational savings are possible, with only minor decline in accuracy. At the same time, freezing the same number of the highest-energy canonical HF orbitals introduces much larger errors, as shown in Fig. 2. The mean errors (MEs) for the six lowest IEs of the water dimer (shown in the top panel) in the FNO calculations systematically decrease as the percentage of retained orbitals increases, whereas freezing canonical virtual MOs produces nonsystematic (and larger) errors. The difference becomes even more pronounced when considering standard deviations (STDs) (shown at the bottom panel of Fig. 2).

The goal of this work is to benchmark the performance of the FNO-based truncation schemes for IEs as well as energy differences along potential energy surfaces (PESs). We consider multiple electronic states in beryllium, $(H_2)_2$, water, nitrogen, water dimer, uracil dimer, and guanine. Throughout this paper, IEs refer to vertical IEs (VIEs) unless otherwise noted. The errors introduced by the FNO truncation are computed as the difference between the full and truncated virtual space calculations. The errors are characterized by the following statistical parameters: the ME, $\bar{\Delta}$, and the STD, Δ_{STD}. All FNO/EOM-IP-CCSD MEs equal the mean absolute error, $\bar{\Delta}_{\mathrm{abs}}$, because the IEs are systematically underestimated in the truncated calculations. To characterize the quality of PESs, we analyzed the difference between the FVS and FNO curves in terms of the nonparallelity errors (NPEs),[27] which are computed as the difference between the maximum and minimum errors along a PES scan. In addition, the errors in dissociation energies (DEs) are examined.

The structure of the paper is as follows: Sec. II outlines EOM-IP-CC approach and presents the FNO formalism. Section III describes the computational details. The results are presented in Sec. IV. Our concluding remarks are given in Sec. V.

II. FORMALISM

A. The equation-of-motion coupled-cluster method for ionized states

The performance of FNO for the ground-state CC energies and PESs was benchmarked in Refs. 15 and 16. In this work, we present the results for the EOM-IP-CC methods.[28–33] We implemented the FNO approach within the EOM-IP-CC codes[33] as well as ground-state CC methods in the *Q-CHEM* electronic structure package.[34]

EOM-IP-CC describes open-shell doublet states as "ionized" states derived from a well-behaved closed-shell reference wave function. By avoiding open-shell reference, the method is free from spin contamination and is less susceptible to symmetry breaking problems. The EOM-IP-CC wave function has the following form:

$$|\Psi_R\rangle = Re^T|\Phi_0\rangle, \tag{2}$$

where the linear EOM operator R acts on the reference CCSD wave function $e^T|\Phi_0\rangle$. The operator T is an excitation operator satisfying the reference-state CC equations,

$$\langle\Phi_\mu|\bar{H}|\Phi_0\rangle = 0, \tag{3}$$

where Φ_μ are the μ-tuply excited determinants (with respect to the reference determinant Φ_0). In FNO calculations, T and R are defined within the active orbital space. The EOM wave functions are the eigenstates of the so-called similarity-transformed Hamiltonian $\bar{H}=e^{-T}He^T$,

$$\bar{H}R = ER. \tag{4}$$

The EOM amplitudes and the corresponding IEs are found by diagonalizing $\bar{H}$ in the space of target configurations defined by the choice of the operator R and the reference Φ_0 (see Ref. 35 for examples of different EOM models).

In EOM-IP-CCSD, the CC and EOM operators are truncated at double excitation level, i.e., such that only the single and double excitation operators ($1h1p$ and $2h2p$) are retained in T, and R contains $1h$ and $2h1p$ operators (h and p represent a hole and a particle, respectively).

Accuracy of the target states description can be improved by including triple excitations ($3h2p$) in R (while retaining only singles and doubles in the CC expansion) giving rise to EOM-IP-CC(2,3).[33,36]

Once the FNOs are formed and the virtual space is truncated (as described below), the correlated calculations are

performed within the active orbital space as usual. The truncation of the virtual orbital space affects the target energies by modifying the similarity-transformed Hamiltonian (via T) and by truncating the subspace in which $\bar{H}$ is diagonalized. Note that the spectrum of $\bar{H}$ does not depend on the form of T, thus, exact eigenvalues can be obtained by solving Eq. (4) in the full CI space. However, the eigenvalues of $\bar{H}$ in the space of the reference, $1h$, and $2h1p$ configurations are affected by the choice of T. One may expect that errors introduced by approximate T will be of similar magnitude for the reference and the target states, that is, that all the eigenstates of $\bar{H}$ in the full $\{0,\Phi_i,\Phi_{ij}^a\}$ (or $\{0,\Phi_i^a,\Phi_{ij}^{ab}\}$) space will be affected in a similar way. However, when the target configurational space is truncated by restricting R to the active space, one may anticipate larger errors because the CI-like expansion of the EOM states becomes shorter. In extreme cases, the truncation can exclude leading configurations of the target states (e.g., consider an excited state derived from an excitation to a target orbital which is not important for the ground-state correlation, and is therefore excluded from the active space). Since the leading EOM configurations of the principal ionized states are the $1h$ ones, EOM-IP wave functions should be less sensitive to the truncation of the target $\{0,\Phi_i,\Phi_{ij}^a\}$ space than, for example, EOM-EA (EOM for electron attachment) or EOM-EE (EOM for excitation energies) ones. However, the truncation of the target EOM space may affect the EOM-IP energies by reducing the ability of EOM-IP ansatz to account for the differential correlation and orbital relaxation effects, which are described by $2h1h$ (and $3h2p$) configurations. Our results indicate that the truncation results only in a minor accuracy loss.

B. Frozen natural orbitals

The first-order correction to the HF wave function is given by a linear combination of doubly excited determinants,

$$|\Psi^{(1)}\rangle = \frac{1}{4}\sum_{ijab}\frac{\langle ab||ij\rangle}{\epsilon_{ij}^{ab}}|\Phi_{ij}^{ab}\rangle, \tag{5}$$

where $\langle ab||ij\rangle$ are the two-electron antisymmetrized Coulomb integrals and the denominator is

$$\epsilon_{ij}^{ab} = \epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b, \tag{6}$$

where ϵ_p denotes HF orbital energy, and the occupied and virtual orbitals are denoted by $i,j,k\ldots$ and $a,b,c\ldots$, respectively.

This first-order correction can be used to construct the virtual-virtual block of the second-order DM,[37]

$$D_{ab}^{(2)} = \frac{1}{2}\langle\Psi^{(1)}|a_a^\dagger a_b + a_b^\dagger a_a|\Psi^{(1)}\rangle = \frac{1}{2}\sum_{cij}\frac{\langle ac||ij\rangle\langle ij||cb\rangle}{\epsilon_{ij}^{ac}\epsilon_{ij}^{cb}}. \tag{7}$$

The FNOs are defined as eigenstates of the VV block of the DM and their coefficients and occupations (the eigenvalues) are found by diagonalizing the DM. Using two different truncation schemes described below, we retain the FNOs with the highest occupations in the correlated calculations, and treat the rest as frozen.

Finally, we block diagonalize the matrix of the Fock operator to transform the virtual NOs to a semicanonical representation.[38,39] After these steps, there are three (or four) orbital subspaces: canonical HF occupied orbitals (active and frozen core), semicanonical active, and frozen virtual NOs. The separation between the occupied and virtual subspaces is not affected by these transformations, and the OV block of the Fock matrix remains zero. In energy calculations, the inactive virtual orbitals are simply frozen; however, in orbital response equations for analytic gradient and property calculations, the inactive FNOs need to be considered,[16] similar to the frozen core and frozen virtual orbitals.[28,29,40]

C. Different truncation criteria and extrapolation to the FVS values

Since there is no clear gap in natural occupation numbers, the choice of the cutoff criterion is open ended. One can simply specify the number of virtual orbitals to retain, e.g., as the percentage of total virtual orbitals, as was done in Refs. 15 and 16. Alternatively, one can specify the percentage of total natural occupation (in the virtual space) that needs to be recovered in the truncated space. We will refer to these two criteria as the POVO and the occupation threshold (OCCT) cutoffs, respectively.

When large bases are employed, the natural population spectrum becomes very dense and successive occupation numbers may differ by less than 10^{-6}. Moreover, in systems with high symmetry, the population numbers may be exactly degenerate. To avoid splitting of the degenerate manifolds, the OCCT cutoff procedure checks for the degeneracies and expands the active space to include all nearly degenerate orbitals within a specified threshold. Proper account of degeneracies is very important in PES calculations and for preserving exact degeneracies of the target states (e.g., as in Jahn–Teller systems).

Since the OCCT criterion is based on the correlation in a specific molecule, it yields more consistent results than POVO, as demonstrated by the numerical examples below. Moreover, we found that the errors decrease linearly as a function of the total natural occupation recovered, which can be exploited by extrapolating truncated virtual space calculations to the FVS values. This extrapolation scheme will be referred to as the extrapolated FNO (XFNO) procedure. The extrapolation is performed using linear regression and the coefficient of determination R^2 is used to evaluate its quality. The linear behavior is exhibited by the total energies of the ground as well as the ionized states, as demonstrated in Fig. 3 that shows the errors in the total energies of the ground and the six lowest ionized states of water dimer as a function of the OCCT. The ground-state energy is calculated by CCSD and the energies of the ionized states—by EOM-IP-CCSD and EOM-IP-CC(2,3) (see Sec. II A). In all cases, the errors in the total energies decrease linearly as a function of the OCCT. Note that the errors in the ground and ionized states energies are of the same order. Moreover, EOM-IP-CCSD and EOM-IP-CC(2,3) yield similar magnitude of errors even though one may expect larger errors when triples are in-

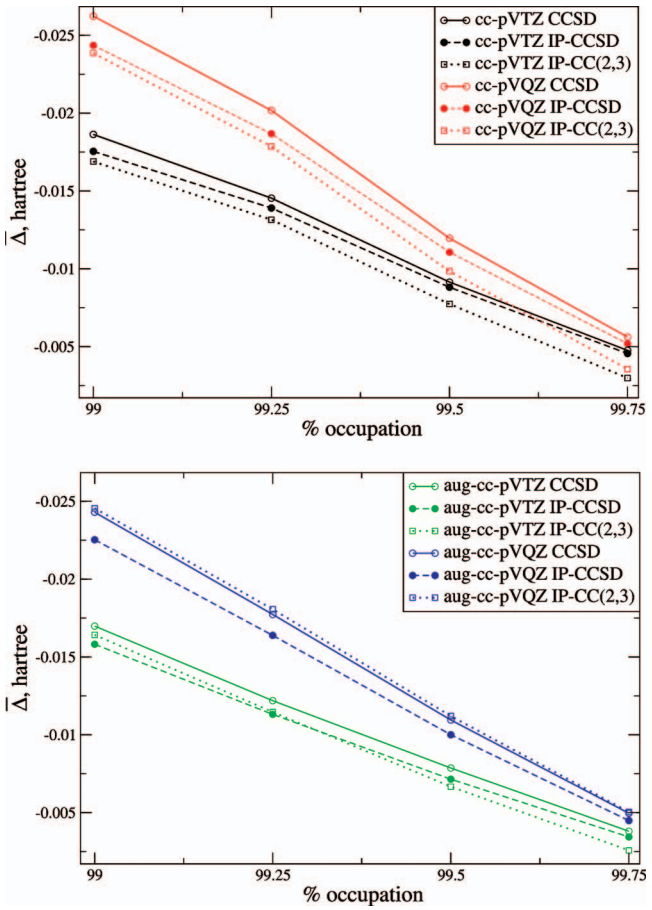

FIG. 3. Errors in total energies of the ground and the lowest six ionized states of $(H_2O)_2^+$ due to FNO truncation as a function of the total natural occupation retained.

cluded (the triple corrections reduce the IEs by about 0.1–0.2 eV in the small and large basis sets, respectively). Thus, Fig. 3 demonstrates that the FNO-based truncation performs equally well for the ground and the ionized states.

The diagonal elements of the Fock matrix in the semicanonical FNO basis of active orbitals are always shifted to higher energies relative to their original canonical energies (while the inactive orbitals are shifted to the lower energies). Figure 4 illustrates this point for beryllium. It presents the

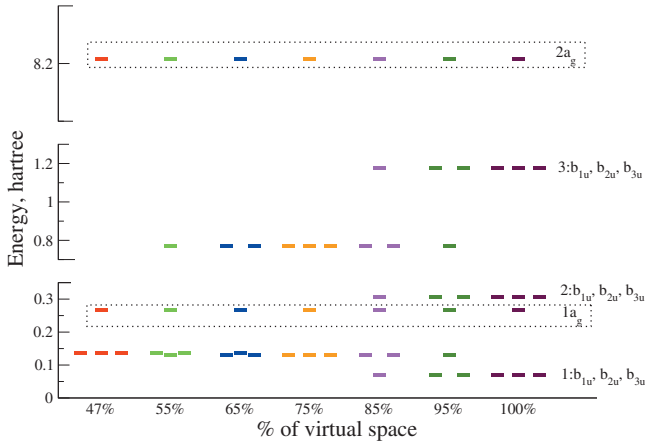

FIG. 4. Energies of the active virtual orbitals for different FNO truncations of the virtual space of beryllium using the 6-311G basis set.

diagonal Fock matrix elements for active orbitals for different FNO truncations of the virtual space. In the FVS (100% population) the 11 virtual orbitals (the 6-311G basis is used) are split into three sets of degenerate p orbitals (which have b_{1u}, b_{2u}, and b_{3u} symmetries in the D_{2h} subgroup) and two s-orbitals (of a_g symmetry). Using 47% of the virtual space yields five active NOs, i.e., two s and three degenerate p orbitals. Because all the s-like NOs are active, they preserve their canonical HF energies. Remarkably, the second s-like NO has the highest HF energy (>8 hartree), yet it is included in this heavily truncated virtual space. This explains the poor performance of the energy-based truncation of virtual space illustrated in Fig. 2. In this truncated space, only three p-like NOs are active and their energies are higher than the energies of the lowest canonical p-manifold. As we increase the threshold to 55%, one more p-orbital (of b_{1u} symmetry) enters the active space, which slightly lowers the energy of the first b_{1u} NO breaking the degeneracy of the p-manifold. Further increase in the threshold (to 75%) results in additional two p-like NOs restoring the degeneracy. The energies of the two p-manifolds are shifted to higher energy with respect to the two lowest canonical p-sets. The same pattern is observed as the third manifold of p-orbitals becomes active. This example demonstrates that the degeneracy of symmetry-equivalent orbitals needs to be properly treated by the population-based truncation procedure to preserve the degeneracy of the ionized states. The trend in orbital energies suggests that the FNOs, which describe well the correlation for the ground and ionized states, may not be optimal for EA and EE calculations, since the target virtual orbitals may not be well described in the truncated virtual space.

We found that the FNO truncation often causes slower convergence of the CCSD and EOM procedures, as was observed for CI by Cave *et al.*,[9] which is counterintuitive in view of the increased virtual orbital energies. Nevertheless, despite the larger number of iterations, the FNO-based truncation of orbital space reduces computational effort considerably, as demonstrated by the numerical examples presented in this paper.

III. COMPUTATIONAL DETAILS

The IEs of water, water dimer, and nitrogen were computed using four different Dunning basis sets:[41,42] cc-pVTZ, aug-cc-pVTZ, cc-pVQZ, and aug-cc-pVQZ. Calculations of the beryllium cation employed cc-pVQZ.[41] The PES scan in $(H_2)_2^+$ was computed with aug-cc-pVTZ.[41] All occupied orbitals were active in the correlated calculations. To estimate the effect of freezing 1s orbitals in light atoms, we computed IEs of water using the cc-pVQZ basis set. We found that freezing the core orbitals reduces the IEs by about 0.01 eV.

Water and water dimer calculations were performed using the geometry of the neutral optimized at the MP2/6-311(+,+)G(d,p) level taken from Ref. 43. The uracil dimer calculations were conducted using the geometry of the neutral from the S22 set of Hobza and co-workers[44] and the 6-311(+)G(d,p) basis set,[45,46] as in Ref. 47.

The geometry of the neutral guanine molecule was opti-

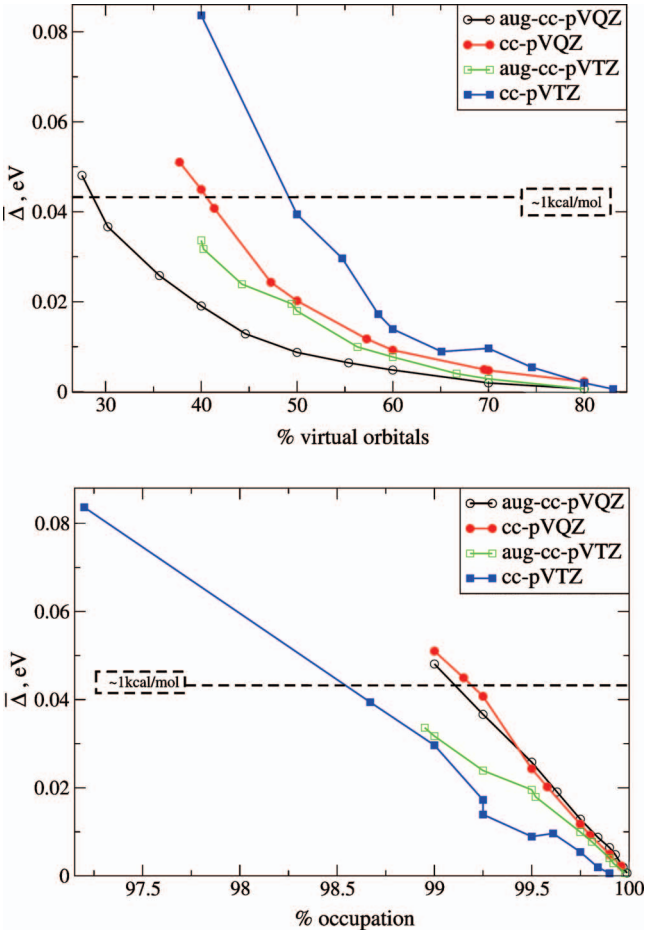

FIG. 5. MEs for the lowest six IEs of water dimer computed by EOM-IP-CCSD using two different cutoff criteria, i.e., POVO (upper panel) and OCCT (lower panel) thresholds.

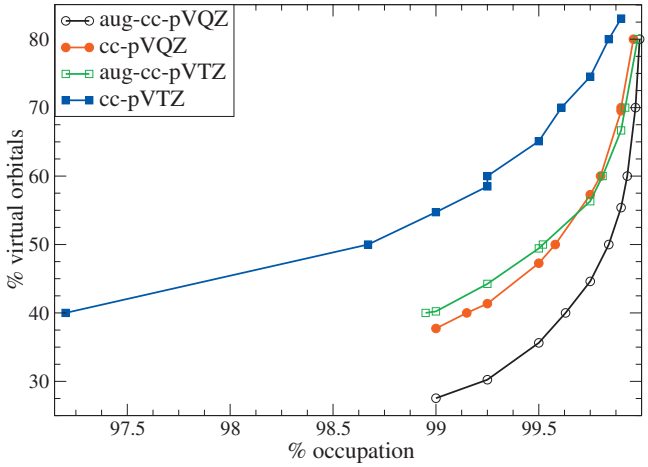

FIG. 6. The POVO as a function of the total natural occupation retained in water dimer calculations.

mized by resolution-of-identity MP2 (RI-MP2) (Refs. 48–50) with the cc-pVTZ+dff basis set (and aux-cc-pVTZ as an auxiliary basis), and is discussed in the supplementary material.[51]

The cc-pVTZ+dff basis is an extended cc-pVTZ basis set augmented by *s* and *p* diffuse functions for the second row elements, and an *s*-function for hydrogen. The exponents of the diffuse functions are taken from the 6-311(2+,+)G basis set.[45,46]

The geometry of the guanine cation was optimized using the $\omega B97X$ functional[52] with cc-pVTZ+dff. The IEs (vertical and adiabatic) of guanine were computed using the cc-pVTZ (Ref. 41) and cc-pVTZ+dff basis sets.

In uracil dimer and guanine calculations all core electrons were frozen. We also employed the RI technique[48] in integral transformation with the aux-cc-pVTZ auxiliary basis. Molecular geometries and relevant energies are provided in the supplementary material.[51]

IV. RESULTS AND DISCUSSION

A. Ionization energies of $(H_2O)_2$, H_2O, N_2, Be, and uracil dimer

Figure 5 presents the MEs for the six lowest IEs of water dimer computed by EOM-IP-CCSD with different basis sets. The top panel presents the MEs as a function of the retained POVO. The POVO threshold is varied from 30% to 83%. As expected, the errors decrease as the number of active orbitals increases; however, there is no simple functional dependence on the threshold. As noted in the previous benchmark studies,[15,16] smaller bases yield larger errors with the same retained percentage of orbitals as compared to the larger bases; however, even a relatively compact cc-pVTZ virtual space can be truncated by 50% with no significant penalty. Even for relatively small values (POVO≈30%), the MEs are around 1 kcal/mol, except for the cc-pVTZ calculation with POVO=40%, which yields ME of about 2 kcal/mol. This point appears as an outlier in Fig. 5 (top). However, when we plot these MEs versus the total natural occupation retained in the truncated space (Fig. 5, bottom panel), the trend becomes clearer: The MEs decrease linearly as a function of the total natural occupation recovered (as specified by the OCCT). This linear behavior is exploited in the XFNO scheme (see Sec. II C), which extrapolates the truncated virtual space energies to the FVS values. Note that the problematic cc-pVTZ point (with 2 kcal/mol ME) has much lower occupation (97.2%) relative to other points, and, therefore, larger error for this calculation can be predicted by considering total natural occupation retained in the truncated space.

Figure 6 presents the POVO versus the total natural occupation recovered. As one can see, the same amount of total natural population can be recovered in a smaller fraction of the virtual space for larger bases, thus confirming that more "dead wood" can be efficiently eliminated from large bases. In this plot, the problematic cc-pVTZ point stands out as having low occupation, whereas the rest of the points (even with much lower POVOs) have higher fraction of total natural occupation recovered. Consequently, the respective calculations yield smaller errors, as illustrated by Fig. 5. Thus, the errors due to the reduced orbital space depend on the percentage of the retained total natural occupation rather than on the percentage of the virtual orbitals used. When more than 99% of natural occupation is retained (which corresponds to 30%–60% of FVS in large and small basis sets, respectively), the MEs are within chemical accuracy.

Similar trends are observed in the EOM-IP-CCSD calculations of water and nitrogen [see supplementary material

(Ref. 51)]. In both cases the errors depend linearly on the retained population. The water results are very similar to the water dimer and for nitrogen, the errors become less than 1 kcal/mol when more than 99.5% of the total natural occupation is recovered, which corresponds to 40%–70% of the FVS in large and small basis sets, respectively.

To test the performance of the FNO truncation scheme in smaller systems, we considered the beryllium atom, where the difference between the neutral and the cation is more pronounced and one may expect larger errors due to the FNO truncation. The FVS EOM-IP-CCSD/cc-pVQZ value of the beryllium's lowest IE is 9.303 eV. By using only 45% of the orbital space, we recover 99.2% of the total natural population. The respective FNO error is about 0.14 kcal/mol. The errors decrease linearly as a function of the OCCT, and extrapolation using 99.2%, 99.3%, 99.4%, and 99.5% occupations reduces the error by half. The extrapolated value is 9.300 eV (with $R^2=0.983$).

The results for a larger system, the uracil dimer, are also encouraging [see supplementary material (Ref. 51)]. For the ten lowest IEs, the errors behave linearly and OCCT of 99%, which corresponds to about 60% of the FVS, brings the errors below 1 kcal/mol.

Finally, we investigated the performance of the FNO truncation of virtual space in EOM-IP-CC calculations with triple excitations using EOM-IP-CC(2,3). Figure 7 (top panel) shows that the MEs no longer exhibit linear behavior with respect to the total natural occupation; however, the errors are very small for a wide range of occupations. Even for the OCCT below 98%, the MEs are less than 0.5 kcal/mol. The STDs shown at the bottom panel of Fig. 7 decrease as a function of the OCCT and become an order of magnitude smaller than the chemical accuracy above 99% of the total natural occupation. The total energies of the neutral and ionized states do behave linearly as a function of the OCCT (Fig. 3), and separate extrapolations (for the ground and ionized state total energies) may be performed. For example, using the aug-cc-pVQZ basis, the extrapolated six IEs obtained from the calculations with 99%, 99.25%, 99.5%, and 99.75% OCCTs are in excellent agreement with the FVS values, i.e., $\bar{\Delta}=0.05$ kcal/mol and $\bar{\Delta}_{abs}=0.06$ kcal/mol. The R^2 value for the ground and six ionized states is 1.000.

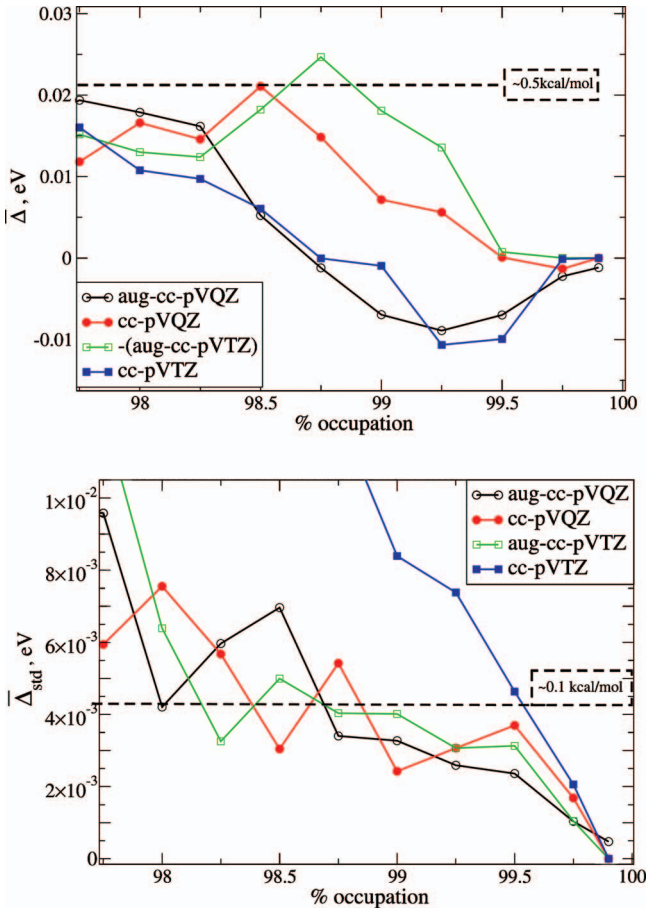

FIG. 7. MEs (upper panel) and STD (lower panel) for the lowest six IEs with respect to the total natural occupation retained in the EOM-IP-CC(2,3) calculations of water dimer. The aug-cc-pVTZ values are plotted with an opposite sign for convenience.

B. Guanine

Guanine, the largest nucleobase present in DNA and RNA, has the lowest gas-phase IE. Guanine has several tautomers, and here we consider only the biologically relevant (keto-N9H) form. Guanine exhibits the most pronounced nonplanarity of the amino group, as discussed in several recent works;[53–55] however, the cation is planar. We performed structural optimization [see supplementary material (Ref. 51) for discussion] followed by IE calculations. We computed the 12 lowest VIEs and the lowest adiabatic IE (AIE) and compare the results with the experimental[56,57] and other theoretical works.[58–64]

Figure 8 shows the MEs for the lowest 12 IEs of guanine versus the OCCT. As in the previous examples, the errors exhibit linear behavior, and the occupation above 99.3% (which corresponds to about 62% of the FVS) yields errors below 1 kcal/mol. The computational savings due to the truncation are more significant for this large system, i.e., the FVS calculation took more than 10 days, while the FNO calculations were completed in 2–5 days only.

Table I and Fig. 9 present the results for the twelve lowest ionized states. The leading MOs and their weights in the

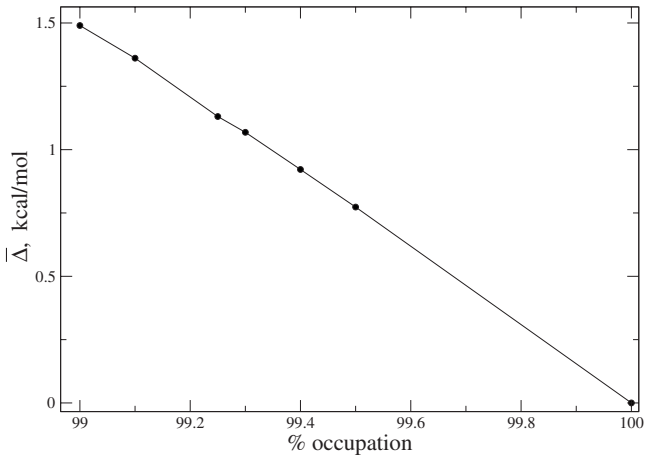

FIG. 8. ME for the lowest 12 IEs with respect to the OCCT in the EOM-IP-CCSD/cc-pVTZ calculation of guanine.

TABLE I. 12 lowest IEs (eV) of guanine: Koopmans' energies (the respective MO number is given in parenthesis) and the EOM-IP-CCSD IEs computed in the FVS and by XFNO (the weight of the leading MO is shown in parenthesis). The results of the ultraviolet photoelectron spectroscopy measurements and the synchrotron radiation valence-shell photoionization spectra are also shown (band assignment is given in parentheses).

Koopmans[a]	FVS[a]	XFNO[a]	XFNO[b]	Experiment[c]	Experiment[d]
8.109 (1)	8.049 (0.96)	8.053	8.148	8.28[e] (A)	8.26 (A)
11.646 (3)	9.766 (0.90)	9.761	9.859	9.9 (B)	9.81 (B)
12.490 (5)	10.039 (0.89)	10.035	10.127	10.4 (C)	10.36 (C)
11.129 (2)	10.201 (0.94)	10.199	10.288	9.9 (B)	9.81 (B)
11.701 (4)	10.502 (0.89)	10.498	10.584	10.4 (C)	10.36 (C)
12.490 (6)	11.298 (0.87)	11.296	11.382	11.2 (D)	11.14 (D)
13.497 (7)	11.485 (0.91)	11.482	11.573	11.2 (D)	11.14 (D)
14.912 (8)	13.462 (0.94)	13.468	13.550	13.0 (E)	13.05 (E)
16.681 (11)	14.357 (0.80)	14.356	14.447		14.39 (F)
16.218 (9)	14.502 (0.86)	14.501	14.584		14.39 (F)
16.327 (10)	14.687 (0.75)	14.679	14.769		15.06 (G)
17.198 (12)	15.391 (0.94)	15.382	15.474		15.06 (G)

[a]cc-pVTZ.
[b]cc-pVTZ+dff.
[c]Reference 57.
[d]Reference 65.
[e]Reference 56 reports 8.24 eV.

EOM-IP-CCSD wave functions of the ionized states are shown in Fig. 9 along with their Koopmans and EOM-IP-CCSD ionization energies. Orbital numbers (given in parenthesis) correspond to the MO numbers in Table I, i.e., the MO No. 1 is the HOMO. Except for the lowest IE, the Koopmans energies deviate from the EOM-IP-CC values by 1 eV and even more than 2 eV for the states of a mixed character. Correlation and mixing of configurations change the order of the states. For example, the states derived from ionization of the mixed third and fifth MOs (the weight of the second dominant MO is ≈0.3 in both cases) appear as the second and third ionized states that are below the ionized state derived by ionization of the second MO. In this nonplanar C_1 molecule, σ-like orbitals can mix with π-like out-of-plane orbitals, e.g., the 11th ionized state is a mixture of the 9th (π), 10th (σ/π), and 11th (σ) MOs, with the respective weights of 0.3, 0.75, and 0.49. The ninth and tenth MOs are also of a mixed π/σ character.

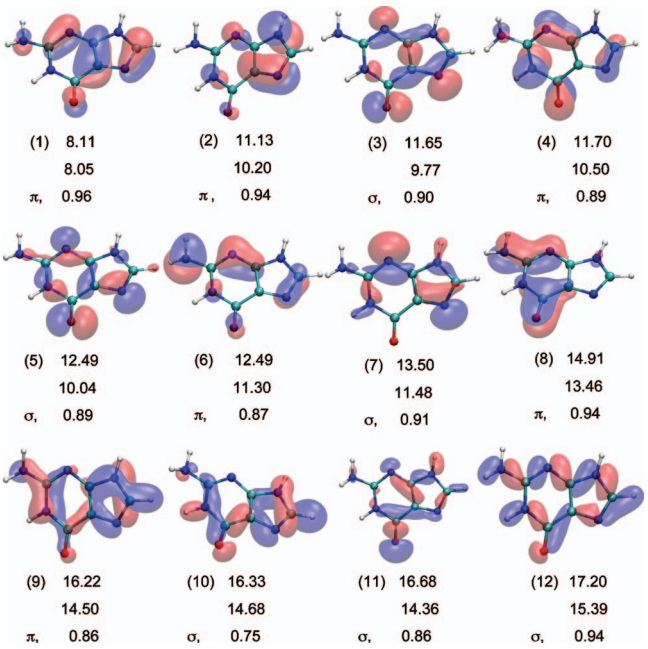

FIG. 9. The 12 highest occupied MOs of guanine and the respective Koopmans' (HF/cc-pVTZ) and correlated (EOM-IP-CC/cc-pVTZ) ionization energies (eV). Orbital type and weights of the leading EOM-IP-CC (R_1) amplitudes are also shown.

Table I summarizes the IEs and compares the FVS and the XFNO values. It also contains the experimental maxima of the peaks in the ultraviolet photoelectron spectra[57] and the results of the recent synchrotron radiation valence-shell photoionization measurements.[65] The assignment of the experimental peaks is ambiguous due to the dense character of the spectrum and overlapping bands. We followed the assignment given in Ref. 57 and use the band labels from Ref. 65.

We observe a remarkable agreement between the XFNO (using the calculations with 99%, 99.1%, 99.25%, 99.3%, 99.4%, and 99.5% OCCTs) and the FVS calculations with the cc-pVTZ basis: $\bar{\Delta}=0.078$ kcal/mol and $\bar{\Delta}_{abs}=0.095$ kcal/mol. In addition, we present the XFNO/cc-pVTZ+dff values (the FVS calculations with this large basis were prohibitively expensive). The extrapolation was done using 99%, 99.15%, 99.3%, and 99.5% occupation points, and the averaged R^2 value for the 12 ionized states is 0.999. The additional diffuse functions increase the VIEs by about 0.1 eV. The comparison with the experiment is more complicated because of the relatively broad peaks and possible presence of multiple tautomers. The synchrotron radiation study[65] also includes theoretical IEs up to 30 eV computed with algebraic-diagrammatic construction [ADC(3)] approximation and outer valence Green's function (OVGF) schemes. The ADC(3) calculations, which are approximately of IP-CCSD quality, were performed at the planar geometry and with a small basis set (6-31G). The OVGF/6-311++G** values are in a qualitatively agreement with our results

(the differences vary from −0.17 to 0.17 eV), although the relative order of some closely lying states is different.

The theoretical estimates of the lowest VIE of guanine reported by several groups vary by more than 0.3 eV: 8.27 eV at the MP2/6-311++G(d,p)//HF/6-31G(d,p) level;[58] 8.09 eV using IP/EA-corrected CASPT2 with the ANO-L 431/21 basis (optimization at the CASSCF/ANO-L 431/21 level), as well as at the CCSD(T)/aug-cc-pVDZ//CCSD/aug-cc-pVDZ level;[59] 7.91 eV using B3LYP/TZVP//B3LYP/TZVP;[60] 8.02 eV at the B3LYP/6-311++G(d,p)//HF/6-311G(d,p) level[61] [same as in the B3LYP/6-311++G(d,p) calculation[62]]; 8.01 eV at the OVGF/6-311++G**//MP2/cc-pVTZ level.[65] Our best estimate of the first VIE is 8.15 eV.

Recent photoionization experiment reported the refined onset for the thermal vaporization photoionization efficiency (PIE) curve as 7.75 ± 0.05 eV (Ref. 62) [an early value was reported as 8.1 ± 0.1 eV (Ref. 64)]. The PIE appearance energy corresponds to the AIE. Other experimental values for the AIE are in agreement with 7.77 eV (Ref. 63) and 7.8 eV.[57]

We calculated the AIE as the difference between the cation EOM-IP-CC total energy (at the cation geometry) and the neutral ground-state CCSD energy (at the neutral geometry). Both ground and ionized states' energies are obtained using the XFNO scheme. The XFNO calculations were carried out with the same total natural occupation as in the VIE calculations above. The calculated AIEs are 7.671 and 7.747 eV with the cc-pVTZ and the cc-pVTZ+dff bases, respectively [see supplementary material (Ref. 51)]. Relative to the experimental results,[62] the error in AIE computed with the extended basis is 0.05 eV. By assuming similar errors in the VIEs, we estimate the lowest VIE and AIE of the keto-N9H tautomer of guanine as 8.15 ± 0.05 and 7.75 ± 0.05 eV, respectively.

C. Potential energy surfaces

Chemical applications require accurate PESs and reliable energy differences. Along the PES scans the correlation and the character of correlated wave functions change, which presents a challenge for approximate methods relying on error cancellation. For instance, relative contributions to the correlation energy of orbitals of a different character (e.g., σ versus π) may depend strongly on molecular geometry, which can result in nonparallelity of the FNO and the FVS curves due to the changes in the character of the truncated space. This may also yield larger errors for energy differences along the PES relative to vertical energy difference calculations.

We assess the errors due to the FNO truncation along PESs for several test cases. First, we consider two different scans for ionized water dimer along the interfragment distance and along the dihedral angle. Second, we analyze the PES scan along interfragment distance for $(H_2)_2^+$. The geometries of the fragments are frozen in these calculations. For details see supplementary material.[51]

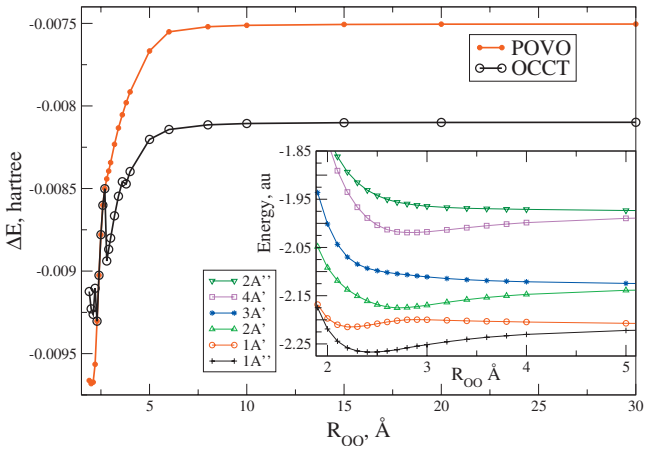

FIG. 10. Energy difference (ΔE, hartree) between the FVS and FNO calculations of the 1A′ state of water dimer using POVO=66.67% and OCCT =99.5% (EOM-IP-CC/cc-pVTZ). Inset: the PESs of the six lowest states using OCCT (energy is shifted by 150 hartree).

1. The PES scans of water dimer along interfragment and torsional coordinates

We consider the PES scans of the six lowest ionized states, of which four are bound and two are dissociative [see supplementary material (Ref. 51)]. The FVS using cc-pVTZ contains 106 orbitals. The number of FNOs retained with OCCT=99.5% changes from 71 at short O–O distances to 69 at large interfragment separations. The POVO cutoff produces a fixed number of FNOs. In this example, we use POVO=66.67%, which corresponds to 70 active orbitals. Figure 10 shows the energy differences (ΔE) between the FVS and the FNO curves of the 1A′ state for the two cutoff schemes. ΔE depends strongly on the number of orbitals retained. At short distances, 71 active orbitals yield smaller errors than 70 FNOs in the POVO scheme, then the same number of FNOs in both schemes yields the same ΔE's, and finally, at large distances, 69 FNOs give larger ΔE's than the POVO cutoff. Overall, the OCCT curve is more *parallel* to the FVS one than the POVO curve. Although the OCCT energy difference curve does not appear smooth, the actual PESs are smooth, as shown in the inset of Fig. 10.

The parallelity of two curves can be quantified using the NPE. Since the number of active FNOs in the OCCT =99.5% scan changes from 71 to 69, it is appropriate to compare it with the scans obtained using POVO=67%, 66.67%, and 65.5%, which correspond to 71, 70, and 69 FNOs, respectively. The average NPE (for the six water dimer surfaces) using the four thresholds are presented in Table II. Table II illustrates the advantage of the occupation-

TABLE II. Average NPEs (kcal/mol) of the water dimer cation for different FNO scans (the number of active NOs is given in parenthesis), EOM-IP-CC/cc-pVTZ.

Cutoff	OCCT	POVO		
Threshold	99.5%(71–69)	67%(71)	66.67%(70)	65.5%(69)
Average NPE	1.01	1.50	1.59	1.64

TABLE III. The effect of diffuse functions and triple excitations on the DEs (mhartree) of the water dimer cation using FVS (unrelaxed estimates, see text). At the bottom part, the values for EOM-IP-CC/cc-pVTZ plus additive corrections from diffuse functions and triple excitations are shown.

EOM-IP-CC/cc-pVTZ			
State	R_{min} (Å)	DE	
1A″	2.4	55.37	
1A′	2.2	3.61	
2A′	2.7	46.09	
4A′	2.8	40.20	
EOM-IP-CC/cc-pVTZ+dff			
State	R_{min} (Å)	DE	Δ diffuse
1A″	2.4	52.33	−3.04
1A′	2.2	0.54	−3.07
2A′	2.7	45.20	−0.89
4A′	2.8	38.13	−2.07
EOM-IP-CC(2,3)/cc-pVTZ			
State	R_{min} (Å)	DE	Δ triples
1A″	2.4	56.99	1.62
1A′	2.2	5.71	2.10
2A′	2.7	48.04	1.95
4A′	2.8	42.00	1.80
Additive corrections			
State	R_{min} (Å)	DE	
1A″	2.4	53.95	
1A′	2.2	2.64	
2A′	2.7	47.15	
4A′	2.8	39.93	

based truncation over using a fixed percentage of the virtual orbitals. The average NPE for the OCCT is smaller than for the three POVO thresholds.

In addition to the EOM-IP-CCSD/cc-pVTZ calculations, we also analyzed the errors introduced by the FNO truncation at higher levels of theory, i.e., in EOM-IP-CCSD/cc-pVTZ+dff and EOM-IP-CC(2,3)/cc-pVTZ. To quantify the errors in energy differences, we consider crude estimates of the equilibrium interfragment distances and the DEs obtained from the one-dimensional PES scans with unrelaxed fragments. These values for the four bound states computed using FVS are summarized in Table III.

It is clear that diffuse functions and triple excitations are important for converged DEs. The respective average effects for the four states are −2.27 mhartree (−1.42 kcal/mol) and 1.87 mhartree (1.17 kcal/mol). Interestingly, basis set and triple excitation effects almost exactly cancel each other out.

The average errors in DEs for the four states and NPEs for the six states due to FNO truncation as well as the respective threshold values are summarized in Table IV. Table IV demonstrates that the occupation-based cutoff criterion consistently yields better results than the percentage-based threshold, i.e., the NPEs are smaller for OCCT than for POVO in all three calculations. The DEs show similar trend: the MEs are well within chemical accuracy for the occupation-based cutoff and their values are much smaller than the values obtained with the fixed-percentage criterion. As expected, the energy difference errors along the PES scan are larger than the vertical energy difference errors. For example, at the EOM-IP-CCSD/cc-pVTZ level with OCCT =99.5% the ME in DEs is 0.62 kcal/mol (see Table IV), while the ME in IE is only 0.24 kcal/mol (see Fig. 5).

The numbers of the retained FNOs when using the cc-pVTZ+dff basis (in which the FVS contains 126 virtual orbitals) with OCCT=97% are 81 and 80 for short and large distances, respectively. The POVO=65% corresponds to 81 FNOs. For the EOM-IP-CC(2,3)/cc-pVTZ calculation, the number of FNOs is the same as in the EOM-IP-CCSD/cc-pVTZ scan.

The ME introduced by using the FNO truncation with cc-pVTZ at the EOM-IP-CC and EOM-IP-CC(2,3) levels are of the same order; however, the MEs in cc-pVTZ+dff are much smaller since the occupation used with the extended basis was considerably larger (97%) than 95% used in the other sets. 95% OCCT with the cc-pVTZ+dff retains only 58% of the FVS (relative to 66.67% with the cc-pVTZ basis). Smaller cutoffs with the extended basis result in large fluctuations of the FNO errors along the scan. This demonstrates that the PES scans are more sensitive than the IE calculations, and therefore require tighter FNO cutoffs, i.e., above 95% occupation.

Along the torsional scan [see supplementary material (Ref. 51)], the changes in electronic structure are smaller, and the occupation-based cutoff gives the same number of FNOs at all angles. Figure 11 shows the lowest PES of the water dimer cation using the full valence and the FNO-truncated (OCCT=99.5%) orbital spaces. The two surfaces are almost parallel and similar behavior is observed for other electronic states. The NPEs for the six lowest surfaces [see supplementary material (Ref. 51)] are below 0.06 mhartree, and their average is about −0.0396 mhartree (−0.025 kcal/

TABLE IV. Average NPEs and MEs for the DEs (ME) for $(H_2O)_2^+$, kcal/mol.

Method	IP-CCSD/cc-pVTZ		IP-CCSD/cc-pVTZ+dff		IP-CC(2,3)/cc-pVTZ	
Cutoff	OCCT	POVO	OCCT	POVO	OCCT	POVO
Threshold	99.5%	66.67%	99.7%	65%	99.5%	66.67%
Average NPE	1.01	1.59	0.47	0.66	1.27	1.95
DEs ME	0.62	0.91	0.19	0.39	0.68	1.06

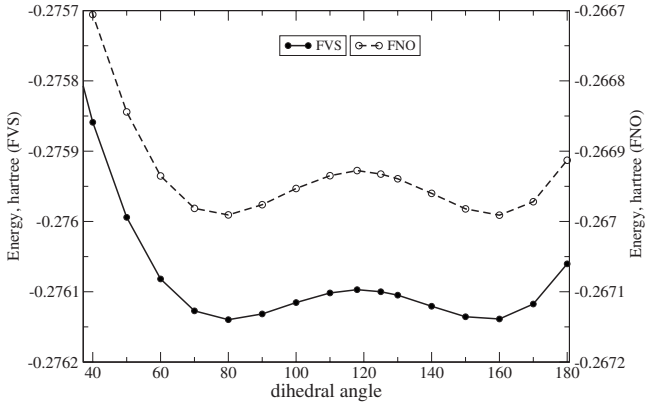

FIG. 11. The PES scan along the dihedral angle of the lowest ionized state of water dimer calculated using FVS (full line and circles, left *y*-axis) and FNO with OCCT=99.5% (dashed line and empty circles, right *y*-axis). The energies are shifted by 152 hartree.

mol). Thus, the FNO-based truncation is capable of reproducing such fine details in the shape of the PES correctly.

2. The PES scan for $(H_2)_2$

Finally, we consider the EOM-IP-CC/aug-cc-pVTZ calculations of the ground electronic state of rectangular $(H_2)_2^+$ at different interfragment separations [see supplementary material (Ref. 51)]. Table V demonstrates the importance of proper treatment of the degenerate NOs (see Sec. II C). Table V compares the NPEs for the FNO curves using the OCCT cutoff with and without including entire manifolds of degenerate orbitals in the active space (the degeneracy threshold used is 10^{-6}). In addition, we present the POVO cutoff results without an adjustment for degeneracy. The number of active orbitals used in each calculation is given in parenthesis.

The FNO yields accurate PES with all the cutoffs considered, i.e., NPEs are about 0.2 kcal/mol and less when using slightly more than half of the virtual space. Including the degenerate NOs within the OCCT scans noticeably improves the NPEs. The POVO cutoff also yields reasonable PESs. As the threshold increases, the NPEs obtained using the OCCT cutoff and proper treatment of degenerate manifolds are systematically improved, in contrast to the POVO calculations. The OCCT and POVO NPEs are converged at 99.75% and 58% thresholds, respectively.

TABLE V. The EOM-IP-CCSD/aug-cc-pVTZ NPEs (kcal/mol) along the $(H_2)_2^+$ interfragment separation scan for different OCCT thresholds with and without account of orbital degeneracy, as well as the POVO results. The number of active FNOs along the scan is shown in parenthesis.

Threshold (%)	OCCT[a]	OCCT[b]	Threshold (%)	POVO
99.70	−0.19(46–49)	−0.21(47–52)	53	−0.18(47)
99.71	−0.19(47–50)	−0.18(47–52)	54	−0.19(48)
99.72	−0.18(47–50)	−0.15(47–54)	55	−0.19(49)
99.73	−0.18(47–50)	−0.10(48–54)	56	−0.16(50)
99.74	−0.18(48–51)	−0.10(48–54)	57	−0.05(51)
99.75	−0.18(49–51)	−0.06(49–54)	58	−0.06(52)
99.76	−0.12(49–52)	−0.06(49–54)	59	−0.04(53)
99.77	−0.06(49–52)	−0.06(49–60)	60	−0.04(54)

[a]Without, degeneracy thresholds equal to 0.
[b]With, degeneracy thresholds equal to 10^{-6}.

Orbital analysis shows that the order and character of the FNOs change along the PES scan, e.g., orbitals that are significant for correlation in one region of the scan become insignificant in others. Consequently, the active space character is different at different geometries. Relatively large errors may arise due to omission of a significant orbital giving rise to large NPEs. For example, the 12th a_g orbital (using the D_{2h} symmetry labels) is a significant orbital between 1.6 and 2.4 Å. The a_g symmetry orbitals are bonding orbitals that are essential in the description of the potential well. This orbital is the 51st NO in this region but appears higher at other regions. Inclusion of this orbital results in a significant decrease in the NPE when going from POVO=56% to 57% [see supplementary material (Ref. 51)]. This orbital is included in the OCCT when degenerate manifolds are correctly treated when going from 99.70% up to 99.74% corresponding to the systematic improvement of the NPE in this threshold range [see supplementary material (Ref. 51)].

V. CONCLUSIONS

We present the implementation and benchmarking of the FNO-based truncation of the virtual orbital space within the EOM-IP family of methods. By sorting the orbitals by their relative importance for the total correlation energy, NOs enable significant truncations of the virtual space with only moderate accuracy loss.

The virtual space truncation scheme based on the MP2 NOs has been implemented and benchmarked for the ground-state CC methods by Bartlett and co-workers,[15,16] who reported excellent performance and significant computational savings. We extended the FNO approach to the EOM methods. Remarkably, virtual space truncation based on the FNOs computed for the reference-state wave function works very well for the target EOM-IP wave functions, which have rather different characters. Although the dominant amplitudes of the target ionized states are in the occupied space, which is not modified in the FNO calculations, the differential correlation and orbital relaxation described by the $2h1p$ (and $3h2p$) EOM operators could, in principle, be adversely affected by the truncation. Our results suggest that the reference-state correlation wrapped into the similarity-transformed Hamiltonian is more important for the description of the target ionized states than the differential correlation effects recovered via CI-like expansion of the EOM states.

Our benchmarks demonstrate that the FNO-based truncation scheme allows one to reduce virtual orbital space by 30%–70% (which results in 4–120-fold speedup), without introducing significant errors. Larger savings are observed in larger basis sets. The FNO-based truncation works consistently well for a variety of basis sets (including augmented bases), as well as different levels of correlation treatment, e.g., EOM-IP-CCSD and EOM-IP-CC(2,3).

For IEs, we found that a cutoff criterion based on total natural population recovered in a truncated space is more

robust than simply utilizing a certain percentage of the virtual space. Proper treatment of orbital degeneracies is important for preserving the exact degeneracies of the ionized states and smooth PESs. Because the population-based criterion takes into account correlation effects in a specific molecule (and in a given basis set), it results in more consistent error bars for a variety of systems. We found that 99%–99.5% natural population threshold (which corresponds to the 30%–70% reduction in the FVS) yields errors below 1 kcal/mol. Moreover, the errors in total energies behave linearly as a function of the population recovered, which enables extrapolation to the FVS limit giving rise to the XFNO scheme. The XFNO scheme can be employed even when the two states are not calculated on the same footing, e.g., in AIEs and EOM-IP-CC(2,3) calculations.

The capabilities of the new approach are demonstrated by calculation of vertical and AIEs of guanine. The FNO truncation can also be employed in PES calculations; however, tighter thresholds are required, i.e., OCCT above 99.5% introduces errors for energy differences (such as, DEs and NPEs) that are well within chemical accuracy.

ACKNOWLEDGMENTS

Support from the National Science Foundation (Grant No. CHE-0616271) is gratefully acknowledged. This work is conducted under auspices of the *iOpenShell* Center for Computational Studies of Electronic Structure and Spectroscopy of Open-Shell and Electronically Excited Species supported by the National Science Foundation through the CRIF:CRF Grant No. CHE-0625419+0624602+0625237. We are also grateful to the Institute of Mathematics and its Applications in Minnesota for its hospitality, computer resources, and productive environment during our stay at IMA as visiting scholars.

[1] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic Structure Theory* (Wiley, New York, 2000).
[2] For the processes when the core effects are important, basis sets that are sufficiently flexible to describe core polarization and core-valence correlation should be employed.
[3] P.-O. Löwdin, Phys. Rev. **97**, 1474 (1955).
[4] T. L. Barr and E. R. Davidson, Phys. Rev. A **1**, 644 (1970).
[5] I. Shavitt, B. J. Rosenberg, and S. Palakit, Int. J. Quantum Chem., Symp. **10**, 33 (1976).
[6] C. F. Bender and E. R. Davidson, J. Chem. Phys. **47**, 4972 (1967); S. J. Desjardins, A. D. O. Bawagan, Z. F. Liu, K. H. Tan, Y. Wang, and E. R. Davidson, J. Chem. Phys. **102**, 6385 (1995).
[7] R. J. Buenker and S. D. Peyerimhoff, Theor. Chim. Acta **35**, 33 (1974).
[8] K. H. Thunemann, J. Römelt, S. D. Peyerimhoff, and R. J. Buenker, Int. J. Quantum Chem. **11**, 743 (1977).
[9] R. J. Cave, S. S. Xantheas, and D. Feller, Theor. Chim. Acta **83**, 31 (1992).
[10] C. D. Sherrill and H. F. Schaefer III, Adv. Quantum Chem. **34**, 143 (1999).
[11] J. Ivanic and K. Ruedenberg, Theor. Chim. Acta **107**, 220 (2002).
[12] M. L. Abrams and C. D. Sherrill, J. Chem. Phys. **118**, 1604 (2003).
[13] M. L. Abrams and C. D. Sherrill, Chem. Phys. Lett. **395**, 227 (2004).
[14] H. J. Aa. Jensen, P. Jørgensen, H. Ågren, and J. Olsen, J. Chem. Phys. **88**, 3834 (1988).
[15] A. G. Taube and R. J. Bartlett, Collect. Czech. Chem. Commun. **70**, 837 (2005).
[16] A. G. Taube and R. J. Bartlett, J. Chem. Phys. **128**, 164101 (2008).
[17] C. Sosa, J. Geertsen, G. W. Trucks, and R. J. Bartlett, Chem. Phys. Lett. **159**, 148 (1989).
[18] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).
[19] J. D. Watts, J. F. Stanton, and R. J. Bartlett, Chem. Phys. Lett. **178**, 471 (1991).
[20] T. D. Crawford and J. F. Stanton, Int. J. Quantum Chem. **70**, 601 (1998).
[21] S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **108**, 5243 (1998).
[22] A. G. Taube and R. J. Bartlett, J. Chem. Phys. **128**, 044110 (2008).
[23] A. G. Taube and R. J. Bartlett, J. Chem. Phys. **128**, 044111 (2008).
[24] J. Almlöf and P. R. Taylor, J. Chem. Phys. **86**, 4070 (1987).
[25] L. Adamowicz and R. J. Bartlett, J. Chem. Phys. **86**, 6314 (1987).
[26] L. Bytautas, J. Ivanic, and K. Ruedenberg, J. Chem. Phys. **119**, 8217 (2003).
[27] M. L. Abrams and C. D. Sherrill, Chem. Phys. Lett. **412**, 121 (2005).
[28] J. F. Stanton and J. Gauss, J. Chem. Phys. **111**, 8785 (1999).
[29] P. A. Pieniazek, S. A. Arnstein, S. E. Bradforth, A. I. Krylov, and C. D. Sherrill, J. Chem. Phys. **127**, 164110 (2007).
[30] D. Sinha, D. Mukhopadhyay, and D. Mukherjee, Chem. Phys. Lett. **129**, 369 (1986).
[31] D. Sinha, D. Mukhopadhya, R. Chaudhuri, and D. Mukherjee, Chem. Phys. Lett. **154**, 544 (1989).
[32] R. Chaudhuri, D. Mukhopadhyay, and D. Mukherjee, Chem. Phys. Lett. **162**, 393 (1989).
[33] P. A. Pieniazek, S. E. Bradforth, and A. I. Krylov, J. Chem. Phys. **129**, 074104 (2008).
[34] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neil, R. A. Distasio, Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Bird, H. Daschel, R. J. Doerksen, A. Drew, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. S. Kedziora, R. Z. Khalliulin, P. Klunziger, A. M. Lee, W. Z. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Herhe, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, Phys. Chem. Chem. Phys. **8**, 3172 (2006).
[35] A. I. Krylov, Annu. Rev. Phys. Chem. **59**, 433 (2008).
[36] S. Hirata, M. Nooijen, and R. J. Bartlett, Chem. Phys. Lett. **326**, 255 (2000).
[37] The second-order DM constructed from the first-order correction to the wave function is not identical to the DM from the MP2 gradient calculations, which includes additional orbital response terms.
[38] N. C. Handy, J. A. Pople, M. Head-Gordon, K. Raghavachari, and G. W. Trucks, Chem. Phys. Lett. **164**, 185 (1989).
[39] J. D. Watts, J. Gauss, and R. J. Bartlett, J. Chem. Phys. **98**, 8718 (1993).
[40] S. V. Levchenko, T. Wang, and A. I. Krylov, J. Chem. Phys. **122**, 224106 (2005).
[41] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[42] R. A. Kendall, Jr., T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[43] P. A. Pieniazek, J. VandeVondele, P. Jungwirth, A. I. Krylov, and S. E. Bradforth, J. Phys. Chem. A **112**, 6159 (2008).
[44] P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[45] W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).
[46] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).
[47] A. A. Golubeva and A. I. Krylov, Phys. Chem. Chem. Phys. **11**, 1303 (2009).
[48] M. Feyereisen, G. Fitzgerald, and A. Komornicki, Chem. Phys. Lett. **208**, 359 (1993).
[49] F. Weigend and M. Haser, Theor. Chim. Acta **208**, 359 (1997).
[50] R. A. Distasio, Jr., R. P. Steele, Y. M. Rhee, Y. Shao, and M. Head-Gordon, J. Comput. Chem. **28**, 839 (2007).
[51] See supplementary material at http://dx.doi.org/10.1063/1.3276630 for molecular geometries and relevant energies.
[52] J.-D. Chai and M. Head-Gordon, J. Chem. Phys. **128**, 084106 (2008).
[53] T. Uchimaru, Th. Zeegers-Huyskens, A. K. Chandra, and M. T. Nguyen, J. Mol. Struct.: THEOCHEM **555**, 61 (2000).
[54] V. B. Delchev and H. Mikosch, Monatsch. Chem. **135**, 1373 (2004).
[55] M. K. Shukla and J. Leszczynski, Chem. Phys. Lett. **414**, 92 (2005).
[56] N. S. Hush and A. S. Cheung, Chem. Phys. Lett. **34**, 11 (1975).

[57] J. Lin, C. Yu, S. Peng, I. Akiyama, K. Li, L. Kao Lee, and P. R. LeBreton, J. Phys. Chem. **84**, 1006 (1980).
[58] E. Cauët, D. Dehareng, and J. Livin, J. Phys. Chem. A **110**, 9200 (2006).
[59] D. Roca-Sanjuán, M. Rubio, M. Merchán, and L. Serrano-Andrés, J. Chem. Phys. **125**, 084302 (2006).
[60] C. M. Marian, J. Phys. Chem. A **111**, 1545 (2007).
[61] M. K. Shukla and J. Leszczynski, Chem. Phys. Lett. **429**, 261 (2006).
[62] J. Zhou, O. Kostko, C. Nicolas, X. Tang, L. Belau, M. S. de Vries, and M. Ahmed, J. Phys. Chem. A **113**, 4829 (2009).
[63] V. M. Orlov, A. N. Smirnov, and Ya. M. Varshavsky, Tetrahedron Lett. **17**, 4377 (1976).
[64] L. Belau, K. R. Wilson, S. R. Leone, and M. Ahmed, J. Phys. Chem. A **111**, 7562 (2007).
[65] I. L. Zaytseva, A. B. Trofimov, J. Schirmer, O. Plekan, V. Feyer, R. Richter, M. Coreno, and K. C. Prince, "Theoretical and experimental study of valence-shell ionization spectra of guanine," J. Phys. Chem. A (in press).