PCCP

Accepted Manuscript

Volume 1 | Number 1 | Jan 2013 | Pages 1–100

PCCP

Physical Chemistry Chemical Physics

www.rsc.org/pccp

ROYAL SOCIETY OF CHEMISTRY

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

ERPA-APSG: a computationally efficient geminal-based method for accurate description of chemical systems

Ewa Pastorczak[1] and Katarzyna Pernal[2, a)]

[1] *Institute of Applied Radiation Chemistry, Lodz University of Technology,*
ul. Wroblewskiego 15, 93-590 Lodz, Poland

[2] *Institute of Physics, Lodz University of Technology, ul. Wolczanska 219,*
90-924 Lodz, Poland

(Dated: 23 February 2015)

Most computational chemistry methods cannot provide a uniformly accurate description of dynamic and static electron correlation. In this paper we present performance of the ERPA-APSG method based on the antisymmetrized product of strongly orthogonal geminal theory (APSG) and the recently proposed extended random phase approximation (ERPA) intergeminal correlation correction. We show that the ERPA-APSG approach is capable of accounting for both dynamic and static correlation, yielding excellent results when applied to describing conformational changes of molecules, twisting of the ethylene molecule, and deprotonation reactions.

---

a) Electronic mail: pernalk@gmail.com

I. THE ERPA-APSG METHOD

Computational methods in chemistry still struggle with the problem of providing good accuracy in predicting diversified electronic structures of molecules at manageable computational cost. Density functional-based approximations can be highly efficient but most of them suffer from severe inaccuracies when multireference effects come into play. As a result they are incapable of correctly describing breaking of covalent bonds - a problem of fundamental importance in chemistry, or singlet states of diradical character. On the other hand, wavefunction methods can, depending on the level of approximation, provide results of required accuracy but the computational cost is often prohibitively high. The most often used wavefunction methods are either of a single-reference character (e.g. couple cluster approximations), thus often unreliable in description of strongly correlated systems, or they are based on a multireference wavefunction (e.g. multireference configuration interaction) and miss an important portion of dynamic electron correlation.

Geminal theories have been proposed as alternatives to one-electron Hartree-Fock method and, by relying on two-electron functions called geminals, they are potentially capable of accounting for dissociation of bonds[1,2]. Imposing a strong orthogonality condition on geminals leads to a relatively simple optimization problem for the energy but this simplification has serious consequences in the performance of the resulting method, called antisymmetrized product of the strongly orthogonal geminals (APSG)[3]. Some of them originate from the lack of intergeminal electron correlation in the APSG energy.

Expansion of spatial parts of geminals $\{\psi_I\}$ in a set of natural orbitals $\{\varphi_p\}$ corresponding to a singlet-state APSG wavefunction ansatz takes diagonal form[4], namely

$$\forall_I \quad \psi_I(\mathbf{r}_1, \mathbf{r}_2) = \sum_{p \in I} c_p \varphi_p(\mathbf{r}_1) \varphi_p(\mathbf{r}_2) \quad . \tag{1}$$

Arai has shown that strong orthogonality of geminals implies that subspaces in which geminals are expanded (Arai spaces) are disjoint[5]. In other words, a given natural orbital φ_p is assigned to only one geminal of the index denoted by I_p. The expression for the APSG energy reads[4]

$$\begin{aligned} E^{APSG}[\{c_p\}, \{\varphi_p\}] = 2 \sum_p c_p^2 \, h_{pp} + \sum_{pq} \delta_{I_p I_q} \, c_p c_q \, \langle pp|qq \rangle \\ + \sum_{pq} (1 - \delta_{I_p I_q}) \, c_p^2 c_q^2 \, [2 \langle pq|pq \rangle - \langle pq|qp \rangle] \quad . \end{aligned} \tag{2}$$

The one- and two-electron integrals, $\{h_{pq}\}$ and $\{\langle pq|rs\rangle\}$, respectively, are in the representation of the natural orbitals. The second term in Eq.(2) accounts exactly for intrageminal electron interactions. As a result APSG provides an exact description of closed-shell two-electron systems and supersystems composed of noninteracting electron pairs. The last term in the APSG expression is responsible for the intergeminal effects. It can be expressed as a sum of Coulomb and exchange intergeminal interactions, apparently missing the intergeminal correlation[6]. APSG is a good starting point for systems of multiconfigurational character, since static correlation effects in electron pairs are exactly captured and two-electron bonds dissociate correctly[2,7]. The method misses, however, a significant portion of the dynamic correlation, which is a consequence of the underlying strong orthogonality condition for geminals resulting in the lack of intergeminal correlation effects. As it will be shown, this deficiency often results in inaccurate energy differences (e.g. energy barriers), which is a serious limitation of the APSG approximation from the perspective of its usefulness for chemical systems.

We have proposed[6] an intergeminal correlation correction to APSG by considering intergeminal density fluctuation terms, which, after employing the fluctuation-dissipation equation, has led to expressing the intergeminal correlation in terms of the transition density matrix elements. The latter can be found by solving the extended random phase approximation (ERPA) equations[8,9] originally formulated as a method for obtaining excitation energies. The ERPA equations can be written as an eigenvalue problem that reads

$$\left[\left(\mathbf{A}^{+}\right)^{1/2}\mathbf{A}^{-}\left(\mathbf{A}^{+}\right)^{1/2}\right]\left(\mathbf{A}^{+}\right)^{-1/2}\mathbf{Y}_{\nu}=\omega_{\nu}^{2}\left(\mathbf{A}^{+}\right)^{-1/2}\mathbf{Y}_{\nu}\quad , \tag{3}$$

where, ω_ν^2 denotes the νth eigenvalue. If the ground state APSG wavefunction is employed, the symmetric and positive definite matrices $\mathbf{A}^-$, $\mathbf{A}^+$ are given in terms of the coefficients $\{c_p\}$ and the orbitals $\{\varphi_p\}$. Their explicit forms are provided in the Appendix. The intergeminal correlation correction has been expressed in terms of the eigenvectors $\mathbf{Y}_\nu$ and the pertinent spin-summed expression reads[6]

$$
\begin{aligned}
E_{\text{intergem}} = &\sum_{p>q}\sum_{r>s}(1-\delta_{I_pI_q}\delta_{I_rI_s}\delta_{I_pI_r}) \\
&\times\Bigg[2(c_p+c_q)(c_r+c_s)\sum_{\nu}(Y_\nu)_{pq}(Y_\nu)_{rs} \\
&-[c_p^2(1-c_q^2)+c_q^2(1-c_p^2)]\delta_{pr}\delta_{qs}\Bigg]\langle pr|qs\rangle \quad . \qquad (4)
\end{aligned}
$$

Notice that terms involving indices belonging to the same geminal are considered to be of intrageminal character and are excluded from the summation.

Computing the energy within the ERPA-APSG method involves two steps. First, the optimal APSG ground state energy is found by minimizing the function given in Eq.(2) with respect to orbitals under the orthogonality constraint and the expansion coefficients constrained by a normalization condition for each geminal, i.e. $\forall_I \ \sum_{p\in I} c_p^2 = 1$. In addition, full optimization of the APSG energy involves finding optimal Arai spaces. Overall, the APSG optimization scales as $M^4 N_{gem}$ with the size of the basis set M and with the number of geminals N_{gem}[10]. The second step consists in solving the ERPA equations (3) and computing the intergeminal correction (4) that is added to the APSG energy. Notice that indices of the $(\mathbf{Y}_\nu)_{pq}$ elements run through all distinct pairs (pq), where $p > q$. Thus, the dimensionality of the ERPA eigenproblem, Eq.(3), for M basis set functions amounts to $M(M-1)/2$ and is much greater than that of the RPA problem in which the size of the $\mathbf{Y}_\nu$ vector is only $N_{occ}N_{virt}$, where N_{occ} and N_{vir} are, respectively, numbers of the occupied and virtual orbitals. In general, all of the APSG natural orbitals are fractionally occupied, i.e. each natural occupation number n_p is different from 0 and 1 (notice that the values of the natural occupation numbers, $\{n_p\}$, directly related to the expansion coefficients by the relation $\forall_p \ n_p = c_p^2$, are bound by 0 and 1). Each Arai space includes one "strongly" occupied natural orbital of the occupation number greater than 1/2, the rest of orbitals being "weakly" occupied, i.e. their occupancies are smaller than 1/2. If the APSG wavefunction describes a molecule close to its equilibrium geometry (or, more generally, bonds in the molecule are not stretched) the occupancies of the "strongly" occupied orbitals are close to 1, whereas those for the "weakly" occupied ones are small and close to 0. The occupation pattern is different in the case of a molecule with a stretched covalent two-electron bond. Then the occupation number of one of the "strongly" occupied orbitals and of one of the "weakly" occupied ones deviates strongly from 1 and 0. In fact, in the dissociation limit the occupancies of two natural orbitals, the so-called frontier orbitals, achieve a value 1/2. Frontier (bonding and antibonding) orbitals belong to the same geminal, which is localized on the breaking bond. The ERPA equations treat all pairs of orbitals on equal footing. However, as it has been mentioned in Ref.[8] not all pairs are equally important for excitation energies obtained from the ERPA method. We have also investigated that reducing the dimension of the main matrix in the eigenproblem (3) to $N_{strong}N_{weak}$, where N_{strong} and N_{weak} denote the number of "strongly" and "weakly" occupied orbitals, respectively, influences the value of the intergeminal correlation correction by a few mHartrees for investigated molecules. In particular, truncated implementation of the intergeminal correction does not influence the method's performance for systems in which multiple bonds are broken (e.g. the description of the symmetric stretching of O-H bond in a water molecule, shown in Ref.[6] is faithfully reproduced by ERPA-APSG). In addition, we have found out that obtaining smooth potential energy curves is much more demanding if solution of the full (instead of reduced) ERPA equations is employed in Eq.(4), which is due to the fact that definitions of the $\mathbf{A}^{-}$, $\mathbf{A}^{+}$ matrices involve dividing by sums and differences of two coefficients, $c_p \pm c_q$, that may be of the values close to 0.

Since truncating the ERPA matrices dramatically reduces the computational cost of solving the ERPA equations and it significantly improves the smoothness of the potential energy curves, we propose that the ERPA-APSG method employs the vectors $\mathbf{Y}_\nu$ resulting from solving the truncated ERPA equations. The assumed truncation involves partitioning of the set of the natural orbitals into "strongly"-occupied, "weakly"-occupied and frontier orbital subsets and removing from the main ERPA matrix rows and columns corresponding to the "strongly"-"strongly" and "weakly"-"weakly" occupied pairs. Notice that if the ERPA-APSG method is applied to atoms or molecules without stretched bonds there is no need to select frontier orbitals and the respective set is empty. The truncated ERPA eigenproblem has a similar (or the same if there are no frontier orbitals) scaling as that of the RPA method. In the following section, we will show that such defined ERPA-APSG method leads to obtaining excellent energy differences for processes governed by both dynamic and static electron correlation.

TABLE I. Inversion and rotational barriers. Absolute errors of the barriers with respect to either CCSD(T) or TCSCF-CISD values are provided in parenthesis. Geometries: $^a R_{OH} = 1.808\,\text{a.u.}$, $\alpha_{HOH} = 104.5°$, $^b R_{NH} = 1.912\,\text{a.u.}$, $\alpha = 67°$ (see Fig.2), $^c R_{OH} = 1.827\,\text{a.u.}$, $R_{OO} = 2.748\,\text{a.u.}$, $\alpha_{HOO} = 99.5°$, α=115.9° (see Fig.3), $^d R_{CC} = 1.330\,\text{a.u.}$, $R_{CH} = 1.076\,\text{a.u.}$, $\alpha_{HCC} = 121.7°$, $\alpha_{\text{torsional}}$=0°. Basis sets used: H_2O, NH_3 - cc-pVTZ[11]; H_2O_2 - cc-pVDZ[11]; C_2H_4 - DZP[12]. ^{e}Results taken from Ref.[12]

Molecule	Method	Total energy [Ha]			Barrier [kcal/mol]	
		Opt. geom.^{a}	Linear			
H_2O	CCSD(T)	-76.3339	-76.2785		34.8 (0.0)	
	MP2	-76.3205	-76.2663		34.0 (0.8)	
	APSG	-76.1569	-76.0886		42.9 (8.1)	
	ERPA-APSG	-76.3297	-76.2751		34.2 (0.5)	
		Opt. geom.^{b}	Planar			
NH_3	CCSD(T)	-56.4746	-56.4637		6.8 (0.0)	
	MP2	-56.4546	-56.4442		6.5 (0.3)	
	APSG	-56.3258	-56.3095		10.2 (3.4)	
	ERPA-APSG	-56.4726	-56.4618		6.8 (0.0)	
		Opt. geom.^{c}	*cis*	*trans*	*cis*	*trans*
H_2O_2	CCSD(T)	-151.1939	-151.1797	-151.1925	8.9 (0.0)	0.9 (0.0)
	MP2	-151.1705	-151.1562	-151.1692	9.0 (0.1)	0.9 (0.0)
	APSG	-150.9338	-150.9188	-150.9329	9.4 (0.5)	0.5 (0.4)
	ERPA-APSG	-151.1624	-151.1483	-151.1611	8.9 (0.0)	0.8 (0.1)
		Opt. geom.^{d}	Twisted (90°)			
C_2H_4	TCSCF-CISD^e	-78.3659	-78.2457		75.4 (0.0)	
	CASSCF^e	-78.1895	-78.0646		78.4 (3.0)	
	MP2	-78.3529	-78.1907		101.8 (26.4)	
	APSG	-78.1920	-78.0646		80.0 (4.6)	
	ERPA-APSG	-78.3700	-78.2445		78.8 (3.4)	

II. RESULTS

First, we investigate inversion of H_2O, umbrella inversion of NH_3, and *cis* and *trans* torsional barriers of H_2O_2 molecule. The results are collected in Table I and Figs. 1-3. As a reference the CCSD(T) results obtained by the molpro package[13] are employed. For comparison we have also included MP2 results. The values of the total ERPA-APSG energies are in agreement with their CCSD(T) counterparts within 2-4 [mHa] for NH_3 and H_2O and around 30 [mHa] for H_2O_2 not only at equilibrium geometries but along the whole energy curves. This is quite remarkable, taking into account that the APSG energies deviate from the reference CCSD(T) values by a few hundreds of [mHa], cf. Table I. Shapes of inversion curves are in excellent agreement with those obtained by the CCSD(T) method which is reflected in highly accurate energy barriers for the considered inversions. The absolute errors with respect to CCSD(T) values amount to only 0.0 [kcal/mol], 0.5 [kcal/mol], 0.0 [kcal/mol] and 0.1 [kcal/mol] for, respectively, inversion of NH_3, H_2O, *cis*-H_2O_2, and *trans*-H_2O_2 and they are significantly lower than those of the APSG method. Notice that the inversion energy barriers predicted by the ERPA-APSG method are comparable and equally close to the reference CCSD(T) values, as these obtained by employing the MP2 approach.

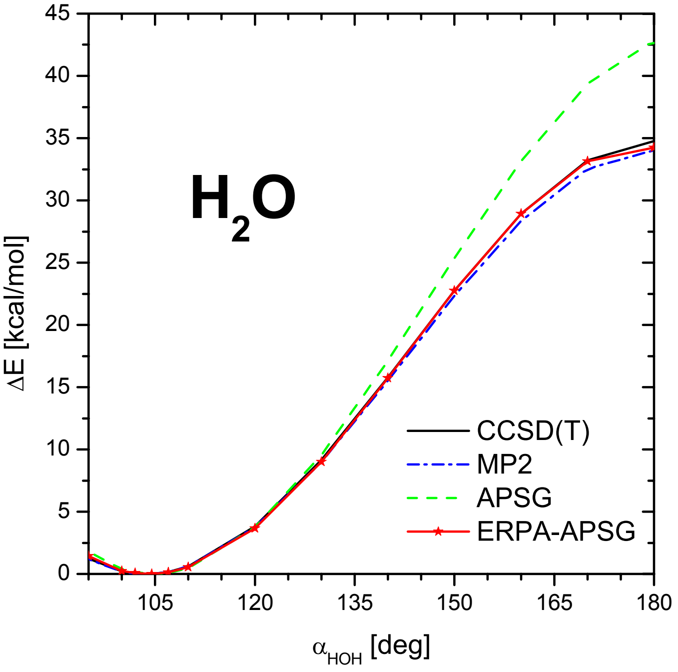

FIG. 1. Inversion of the H_2O molecule in cc-pVTZ[11] basis set.

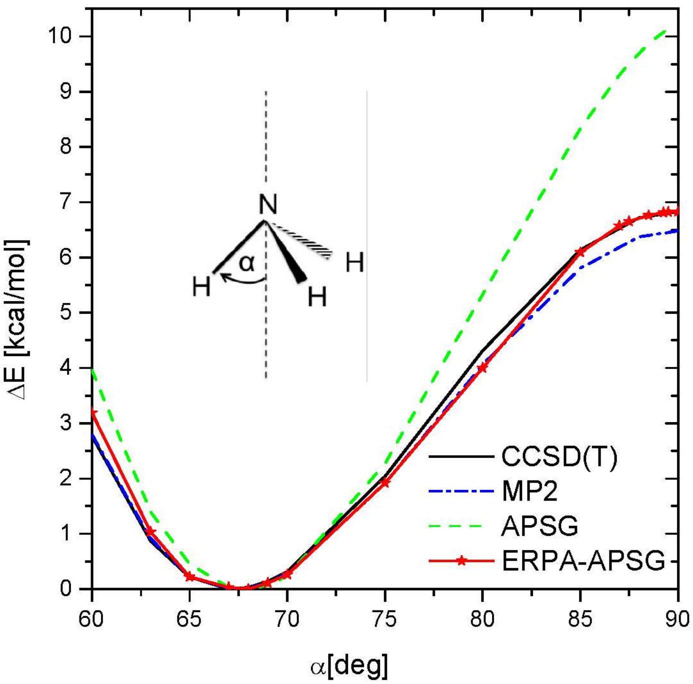

FIG. 2. Umbrella inversion of the NH_3 molecule in cc-pVTZ[11] basis set.

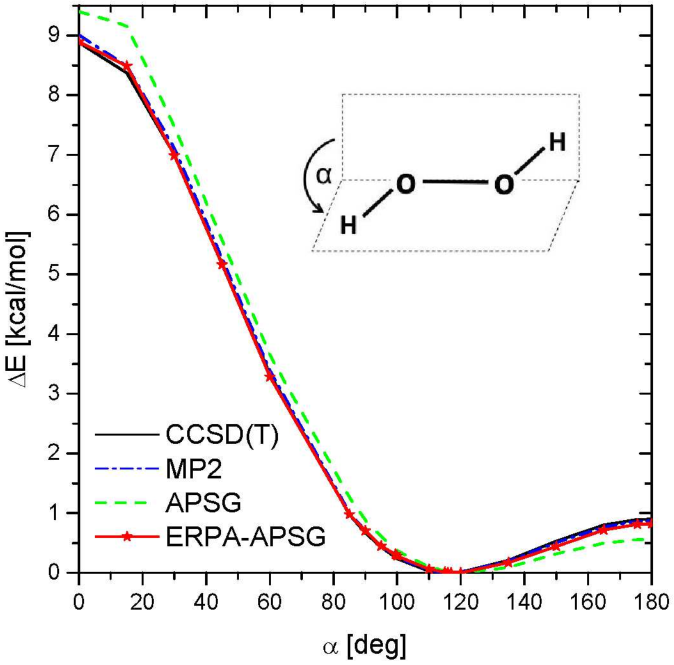

FIG. 3. Twisting of O-O bond of the H_2O_2 molecule in cc-pVDZ[11] basis set.

Rotation of molecular fragments around a single bond is a process that does not involve breaking of covalent bonds and its accurate description requires accounting for mainly dynamic electron correlation. One can wonder, however, how the ERPA-APSG method performs if multireference effects come into play. It has been shown[6] that adding the intergeminal correlation correction to the APSG energy preserves correct shapes of dissociation energy curves. To further investigate the accuracy of the ERPA-APSG method for quasi-degenerate molecules we have applied it to describing twisting of the ethylene molecule. Twisted ethylene is a challenging example of a diradical transition state. A correct wavefunction should account for degenerate $(\pi)^2$ and $(\pi^*)^2$ configurations. Spin-restricted single-reference methods overestimate the torsional energy barrier and produce an unphysical cusp (cf. the MP2 energy curve in Fig. 4). Fig.4 shows that the ERPA-APSG method inherits a proper description of the bond-breaking from the APSG method. Notice that employing the ERPA-APSG method in description of ethylene twisting involves dividing the natural orbitals into "strongly"-, "weakly"-occupied, and frontier orbitals. The lowest (in terms of occupancy) orbital of $n_p > 1/2$ and the highest orbital of the occupation number $< 1/2$ belong to the set of frontier orbitals. Their occupancies achieve a value 1/2 at torsional angle equal 90°.

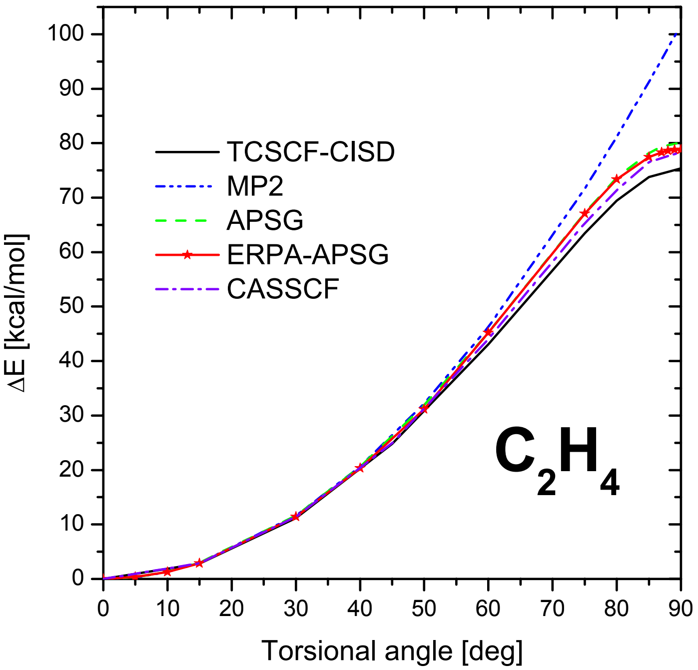

FIG. 4. Twisting of C-C bond of C_2H_4 molecule in DZP[12] basis set.

Table I shows that the torsional energy barrier height predicted by the ERPA-APSG approach amounts to 78.8 [kcal/mol], in a good agreement with the reference value, 75.4

[kcal/mol], predicted by the TCSCF-CISD method [a multireference configuration interaction method which includes all single and double excitations from the $(\pi)^2$ and $(\pi^*)^2$ configurations] of Krylov et al.[12]. Remarkably, the ERPA-APSG total energy values differ by only a few mHartrees from their TCSCF-CISD counterparts (see Table I). For the APSG values the deviation is more than 170 [mHa]. Nevertheless, by confronting Fig.4 one can see that shifted APSG and ERPA-APSG potential energy curves are of very similar shapes, which results in close energy barrier heights predicted by two methods. Still, the ERPA-APSG energy barrier is in a better agreement with the reference TCSCF-CISD barrier height than that of the APSG approach. It is also interesting to notice that the ERPA-APSG torsional energy barrier for ethylene is of the same quality as that obtained with a full valence CASSCF method.

Finally, we apply the ERPA-APSG approach to predict deprotonation energies of water, methanol, ethanol, and propanol. The results are collected in Table II. One immediately notices that the deprotonation energies predicted by the APSG method are in error, with respect to the CCSD(T) data, which ranges from 9.6 [kcal/mol] for H_2O to 12.8 [kcal/mol] for propanol molecule. The errors are within c.a. 4 [kcal/mol] for the tested systems if the ERPA-APSG is employed, so the reduction of errors provided by the intergeminal correlation correction is striking. Noticeably, the error does not seem to increase significantly with the size of the system.

The results presented in the paper show that the ERPA-APSG method may serve as a useful and versatile tool in predicting electronic structure of molecules or in describing chemical reactions. From the examples presented, it is evident that adding the intergeminal correlation energy to the APSG energy not only improves total energies but, more importantly, also energy differences. It offers good accuracy at modest computational cost compared with wavefunction methods like CCSD(T) or CASSCF. The greatest virtue of the ERPA-APSG approach is a balanced treatment of dynamic and static correlation effects that makes the method applicable to systems around and far from equilibrium geometries and when bonds are stretched or breaking.

TABLE II. Deprotonation energies of alcohols, ΔE. Absolute errors with respect to CCSD(T) values are provided in parenthesis. Aug-cc-pVDZ[11] basis set used in all calculations. Geometries: $^{a}R_{OH} = 1.808\,\text{a.u.}$, $\alpha_{HOH} = 104.5°$, ^{b} B3LYP/6-31G* optimized[14], ^{c}MP2/6-31G* optimized.

Molecule	Method	ΔE [kcal/mol]
H_2O^a	CCSD(T)	395.5 (0.0)
	MP2	391.6 (3.9)
	APSG	405.1 (9.6)
	ERPA-APSG	397.9 (2.4)
CH_3OH^b	CCSD(T)	393.5 (0.0)
	MP2	390.4 (3.1)
	APSG	404.1 (10.6)
	ERPA-APSG	396.5 (3.0)
$C_2H_5OH^b$	CCSD(T)	389.3 (0.0)
	MP2	386.6 (2.7)
	APSG	401.9 (12.6)
	ERPA-APSG	393.3 (4.0)
$C_3H_7OH^c$	CCSD(T)	390.9 (0.0)
	MP2	388.2 (2.7)
	APSG	403.7 (12.8)
	ERPA-APSG	395.0 (4.1)

III. APPENDIX

The ERPA-APSG method involves solving the ERPA equations, Eq.(3), where the elements of the matrices $\mathbf{A}^+$ and $\mathbf{A}^-$ are given in terms of the APSG natural orbitals $\{\varphi_p\}$ and the expansion coefficients $\{c_p\}$ [cf. Eq.(2)], namely[9]

$$\forall_{\substack{p>q\\r>s}} \quad A^+_{pq,rs} = (c_p + c_q)^{-1}(A_{pq,rs} + B_{pq,rs})(c_r + c_s)^{-1} \quad ,$$

$$\forall_{\substack{p>q\\r>s}} \quad A^-_{rs,pq} = (c_p - c_q)^{-1}(A_{pq,rs} - B_{pq,rs})(c_r - c_s)^{-1} \quad ,$$

where

$$
\begin{gathered}
B_{pq,rs} = A_{pq,sr} = (n_p - n_q)(\delta_{ps} h_{qr} - \delta_{qr} h_{ps}) \\
+(c_p c_s \delta_{I_p I_s} + c_q c_r \delta_{I_q I_r})(\langle pq|rs\rangle + \langle pq|sr\rangle) \\
+[n_p n_s (1-\delta_{I_p I_s}) + n_q n_r (1-\delta_{I_q I_r}) - n_p n_r (1-\delta_{I_p I_r}) - n_q n_s (1-\delta_{I_q I_s})](2\langle pr|qs\rangle - \langle pr|sq\rangle) \\
-\delta_{qr} c_p \sum_t \delta_{I_p I_t} c_t \langle ps|tt\rangle - \delta_{ps} c_q \sum_t \delta_{I_q I_t} c_t \langle qr|tt\rangle \\
-\delta_{pr} c_r \sum_t \delta_{I_r I_t} c_t \langle qs|tt\rangle - \delta_{qs} c_s \sum_t \delta_{I_s I_t} c_t \langle pr|tt\rangle \\
+\delta_{ps} \sum_t n_t [n_p (1-\delta_{I_p I_t}) - n_q (1-\delta_{I_q I_t})](2\langle qt|rt\rangle - \langle qt|tr\rangle) \\
-\delta_{qr} \sum_t n_t [n_p (1-\delta_{I_p I_t}) - n_q (1-\delta_{I_q I_t})](2\langle pt|st\rangle - \langle pt|ts\rangle)\ .
\end{gathered}
$$

The APSG natural occupation numbers $\{n_p\}$ are simply squares of the pertinent expansion coefficients, i.e. $\forall_p\ n_p = c_p^2$. Two-electron integrals are assumed to be given in terms of the APSG natural orbitals.

ACKNOWLEDGMENTS

This work was supported by the National Science Centre of Poland under grant DEC-2012/07/E/ST4/03023.

REFERENCES

[1] A. C. Hurley, J. Lennard-Jones, and J. A. Pople, Proc. Roy. Soc. (London) **A220**, 446 (1953).

[2] P. R. Surján, "An introduction to the theory of geminals," in *Topics in Current Chemistry*, Vol. 203 (Springer-Verlag, Berlin, 1999) pp. 63–88.

[3] P. R. Surján, A. Szabados, P. Jeszenszki, and T. Zoboki, J. Math.Chem. **50**, 534 (2012), and references cited therein.

[4] W. Kutzelnigg, J. Chem. Phys. **40**, 3640 (1964).

[5] T. Arai, J. Chem. Phys. **33**, 95 (1960).

[6] K. Pernal, J. Chem. Theory Comput. **10**, 4332–4341 (2014).

[7] W. Kutzelnigg, Chem. Phys. **401**, 119 (2012).

[8] K. Chatterjee and K. Pernal, J. Chem. Phys. **137**, 204109 (2012).

[9] K. Pernal, K. Chatterjee, and P. H. Kowalski, J. Chem. Phys. **140**, 014101 (2014), see also *Erratum* ibid, p.189901.

[10] V. A. Rassolov, J. Chem. Phys. **117**, 5978 (2002).

[11] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).

[12] A. I. Krylov and D. C. Sherrill, J. Chem. Phys. **116**, 3194 (2002).

[13] MOLPRO, version 2008.2, a package of ab initio programs, H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, and others, see http://www.molpro.net.

[14] A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, Chem. Phys. Lett. **270**, 419 (1997).