Multireference Quantum Monte Carlo Study of the O_4 Molecule

Michel Caffarel,[1] Ramón Hernández-Lamoneda,[2] Anthony Scemama,[1] and Alejandro Ramírez-Solís[3,*]
[1]Laboratoire de Chimie et Physique Quantiques, CNRS-IRSAMC Université de Toulouse, France
[2]Centro de Investigación en Química, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos 62209, México
[3]Department of Chemistry and Biochemistry, University of California at Santa Barbara, Santa Barbara, California 93106-9510, USA

(Received 27 April 2007; published 10 October 2007)

Fixed-node diffusion quantum Monte Carlo (FN-DMC) calculations are performed to obtain the most accurate dissociation barrier and heat of formation with respect to dissociation into molecular oxygen for the chemically bound tetraoxygen molecule. Multireference trial wave functions were used and built from truncated CASSCF(16,12) through a weight-consistent scheme allowing to control the fixed-node error. Results are compared with the previous *ab initio* benchmark Complete Active Space SCF Averaged Coupled Pair Functional/aug-cc-pVQZ (CASSCF-ACPF/AVQZ) results. The FN-DMC barriers to dissociation and heat of formation obtained are 11.6 ± 1.6 kcal/mol and 98.5 ± 1.9 kcal/mol, respectively. These thermochemical energies should be taken as the theoretical references when discussing the relevance of tetraoxygen in a variety of experiments and atmospheric chemical processes.

DOI: 10.1103/PhysRevLett.99.153001 PACS numbers: 31.15.−p, 02.70.Uu, 31.10.+z

Molecular oxygen is unique in its electronic structure: it is characterized by two unpaired electrons occupying degenerate π_g orbitals which lead to a ground state with a total spin of 1 and two singlet low-lying excited electronic states. Most stable molecules have zero spin and much higher lying excited electronic states. It is therefore not surprising that oxygen in a variety of forms exhibits unique properties. For example, while studying the magnetic properties of liquid oxygen, Lewis [1] proposed the formation of an O_4 species to explain the temperature dependence of the magnetic susceptibility emphasizing the failure of Curie's law to describe oxygen, even though the law generally holds for other paramagnetic substances. It is natural to expect the possible formation of an O_4 complex given the radical character of O_2. We now know that the interaction of two oxygen molecules leads to asymptotically degenerate singlet, triplet, and quintet states with stable complexes being bound by weak van der Waals forces [2–8]. In the solid state, oxygen is the only antiferromagnetic insulating phase among the elemental solids. It is the first light element for which metallization was confirmed experimentally. Several solid phases exist at room temperature and high pressures which exhibit a dramatic change of color as the pressure is increased due to changes in the nature of the intermolecular forces, the detailed explanation still being an area of active research [9–11]. One of the most interesting open problems is the determination and explanation of the structural and optical properties of the ϵ phase which is stable in a broad range of temperatures and high pressures. Some years ago Bini and co-workers [9] measured the infrared spectrum as a function of pressure and noticed the appearance of a new absorption band in the 300–600 cm^{-1} region in addition to the expected molecular absorption at $\sim$1500 cm^{-1}. The new absorption peak was explained through the formation of dimer complexes, O_4, with D_{2h} symmetry analogous to the gas phase van der Waals complex, but the much shorter (30%) intermolecular distances were taken as evidence of a new form of bonding. An alternative theoretical explanation was given by Neaton and Ashcroft [10] who proposed, on the basis of density functional theory, a structure based on linear herringbone-type chains of O_2 molecules consistent with the observed infrared spectra. Finally, very recently a crystal structure determination has been performed [11] which suggests that the basic unit is composed of 4 molecules, O_8, which interestingly can be reconciled with the previously proposed structures and explain the optical experiments. In all of the above examples a common theme is the nature of the intermolecular forces binding the molecular oxygen units. A radically different O_4 species has been proposed to exist on the basis of *ab initio* calculations [12–17] but its experimental detection and characterization is still lacking. The so-called chemically bound O_4 molecule has D_{2d} symmetry with four equivalent single bonds in a cyclic nonplanar structure. The bond lengths are similar to those of other oxygen single bonds. Although gas phase spectroscopy experiments [14,18] have suggested the existence of an O_4 species, it is clear that it is not the chemically bound form but rather electronically excited van der Waals complexes [14,15]. The same conclusion was drawn in the recent neutralization-reionization mass spectrometric detection of an O_4 species [19].

In view of the small size of the system it is natural to expect that state-of-the-art computational methodologies of quantum chemistry should lead to a full understanding of the electronic properties of O_4. Unfortunately, this is not the case and an accurate treatment of the chemically bound O_4 molecule still represents a significant challenge [13–17]. In a few words, the difficulties are the following. The singlet equilibrium structure has a strong monoreference character and, thus, can be calculated by standard and accurate coupled cluster methods [CCSD(T)]. However, describing the C_2 transition state leading to dissociation into two triplet oxygen molecules is much more difficult. The open-shell nature of the two diatomic molecules, the description of bond breaking or making processes, and the complicated spin recouplings necessary to describe the transition from four open shells to a closed shell species can be properly reproduced only with large multiconfigurational wave functions (nondynamical correlation energy). On top of that, it is also essential to describe in a balanced way the subtle electron-electron interactions (dynamic correlation effects) arising in the various geometries corresponding to the singlet O_4 reactant, the transition state, and the O_2(triplet) + O_2(triplet) products.

It is most important to emphasize that obtaining an accurate estimate of the barrier to dissociation is crucial since it determines the stability of the species and, therefore, its relevance for a variety of processes. Among them, we point out its potential role in atmospheric processes involving ozone and highly vibrationally excited oxygen molecules [20–23]. The need of an accurate estimation of the heat of formation of this species has already been noted recently by two of us [16,17]. In a benchmark study [17] using the CCSD(T) and CASSCF(16, 12) + ACPF methods with the large aug-cc-pVQZ basis set, it has been found that the heat of formation is significantly smaller and the barrier to dissociation larger (>9.3 kcal/mol) than previously assumed. The same study revealed that the previous theoretical estimate for the heat of formation of tetraoxygen was in error by a significant amount (18%–24%) owing to lack of accuracy of the methods then employed for evaluating the correlation energy.

The aim of this work is to show that important insights on the debate regarding the stability of the O_4 species can be obtained from a totally different but powerful electronic *ab initio* method, namely, the quantum Monte Carlo (QMC) approach. In short, the QMC approach is a potentially exact stochastic method for solving the electronic Schrödinger equation (see, e.g., [24]). When applied to realistic systems, two major sources of error arise [25]: (i) the standard statistical error present in any Monte Carlo scheme and (ii) the fixed-node error related to the use of approximate nodes for the solution of the Schrödinger equation. The first source of error, the statistical error, is not a fundamental issue since it can be reduced at will either by using better trial wave functions and/or by performing *brute force* simulations reducing the $1/\sqrt{N}$-statistical uncertainty. The second error, the fixed-node bias, is more fundamental. Regarding total energies, the fixed-node error is in general small: a few percents of the total correlation energy [24]. Accordingly, it is presently accepted that the QMC approach is the most accurate approach for calculating total ground-state energies, particularly for large systems for which standard high-level electronic approaches of similar quality are just not feasible [26]. However, a most important point is that physically interesting problems do not require precise total energies but, rather, *accurate differences* of total energies such as barriers, enthalpies, affinities, etc. As a general idea valid for any approach, accurate differences are obtained only when the systematic errors of both components nearly cancel. This is a fundamental point for all *ab initio* methods and this is of course also true for the QMC. Within the fixed-node scheme used here it means that nodal errors must almost compensate (note that, in this work, statistical errors are much smaller than the fixed-node ones and, thus, play no role). To do that, we propose a simple strategy based on the evaluation of the fixed-node error for both states when the multireference character of the nodes is varied. Using this approach we report what we believe to be the most accurate values for the thermochemical data of the $O_4 \rightarrow 2O_2$ reaction. From a wider perspective, we emphasize that the possibility of controlling the fixed-node error for a difference of energies of a few kcal/mol for a nontrivial chemistry problem such as the O_4 stability is an important result. Up to now, the use of the QMC in computational chemistry has remained confidential, mainly because of the lack of such control in small energy differences. It is our hope that the present work will open up the way to more systematic QMC studies of important problems of chemistry.

The FN-DMC method.—The QMC method used is standard and is known as the fixed-node diffusion Monte Carlo (FN-DMC) approach. In a few words, a population of "configurations" or "walkers" ($3n$-dimensional vectors, n number of electrons) is introduced and, at each Monte Carlo step, a set of simple probabilistic rules are applied on each member of the population. A first step consists in moving the walkers according to a time-discretized stochastic equation

$$\mathbf{R}(t+\tau) = \mathbf{R}(t) + \mathbf{b}[\mathbf{R}(t)]\tau + \Delta\mathbf{W}, \tag{1}$$

where $\Delta\mathbf{W}$ is the free diffusion Brownian process and $\mathbf{b}[\mathbf{R}(t)]\tau$ a deterministic drifted move. In this formula the drift vector, $\mathbf{b}$, is given by

$$\mathbf{b} \equiv \nabla\psi_T/\psi_T, \tag{2}$$

where ψ_T is some approximate trial wave function. Then, a so-called birth-death step is performed. It consists in duplicating (or deleting) each walker with a probability proportional to the weight $e^{-(E_L - \langle E_L \rangle)\tau}$ (branching process), where $E_L \equiv H\psi_T/\psi_T$ is the local energy and $\langle E_L \rangle$ its statistical average. Finally, after some transient regime, the population equilibrates and quantum averages of interest can be calculated as statistical Monte Carlo averages.

Control of the fixed-node error.—For bosonic systems the ground-state does not vanish at finite distances and the DMC algorithm just described is exact within statistical uncertainties. In contrast, for fermions the algorithm is biased because of the approximate nodes of ψ_T which play the role of infinitely repulsive barriers for the diffusing walkers [the drift vector diverges, Eq. (2)]. For molecules, the fixed-node error for the total energy is small but it can be sufficiently large to lead, in some cases where the cancellation of errors is not controlled, to large errors in small energy differences. Here, to keep coherent the nodal error both for the equilibrium and transition state geometries of O_4, we propose to build trial wave functions from truncated expansions issued from the large Complete Active Space SCF (CASSCF) (16,12) wave functions. In a previous work, it has been shown that such a description provides zeroth-order reference wave functions that include the nondynamic correlation effects in a *balanced way* for all geometries [17]. Our hope is that the nodal structure of the trial wave functions will also display such a balanced behavior. Regarding the dynamic correlation effects directly related to the interelectronic interaction, the DMC approach is expected to be particularly well adapted.

For both geometries we have performed all-electron FN-DMC calculations with multireference trial wave functions comprising all the determinants of the CASSCF expansion having a configuration interaction (CI) coefficient whose absolute value is greater than a varying given threshold, ϵ. In the case of the smallest threshold considered here, $\epsilon = 0.05$, the different numbers of determinants are 24, 14, and 7 for the equilibrium geometry of O_4, the transition state, and the triplet state of O_2 at its equilibrium geometry, respectively. The curves representing the FN-DMC total energies as a function of the cutoff threshold for both geometries are shown in Fig. 1. As expected from the difference in nondynamical character of the equilibrium and TS states, the fixed-node energy difference calculated with SCF nodes (only one configuration for each geometry) is found to be quite large as compared with other differences. However, as soon as the most important configurations are taken into account, the fixed-node energy difference as a function of the number of determinants is found to converge quite rapidly. It thus appears not to be necessary to consider very large truncated expansions to get a coherent nodal structure for both geometries. To test whether this important result is still valid when the dynamical correlation effects are taken into account, we have studied the stability of our results when an explicitly correlated term is introduced into the trial wave function (the so-called Jastrow factor). The CI coefficients, which in the presence of the Jastrow factor are no longer optimal, have been reoptimized using a technique similar to what has been proposed recently by Riley and Anderson [27]. After reoptimization of the CI coefficients at both geometries, the variational energies have been found to improve. However, this is not true for the fixed-node DMC energies, which are not changed within statistical uncertainties.

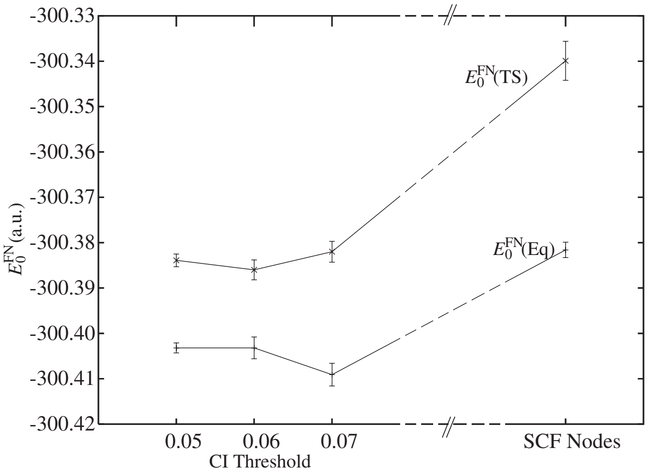

FIG. 1. Convergence of FN-DMC total energies for the equilibrium geometry of O_4 and the transition state as a function of the threshold for absolute values of the CI coefficients of the trial wave function used, see text. Different thresholds correspond to different nodal structures.

Computational details and discussion of results.—The basis set used here is the fully decontracted Dunning aug-cc-pVTZ basis set and the optimized geometries chosen for the reactants, the transition state, and the products are those previously determined [17] at the multireference Rayleigh-Schrödinger perturbed second-order (CASSCF-RS2) level, which are the most accurate ones to date. The trial wave functions used here consist of the truncated CASSCF determinantal expansion discussed above multiplied by a standard Jastrow prefactor taking into account the explicit electron-electron and electron-electron-nucleus interactions (see, e.g. [24]). For a system consisting of light atoms such as O_4 some care has to be taken for properly reproducing the electron-nucleus cusp both for the core and valence electrons. Regarding the core region, we have replaced the four $1s$ atomic orbitals expanded over the Gaussian basis set by the $1s$ Slater-type orbital given in the Clementi and Roetti's tables [28]. On the other hand, the valence molecular orbitals are also modified at short nuclear distances to impose the nuclear cusp; we do that using a short-r representation of the radial part of orbitals under the form $c_1 \exp(-\gamma_1 r) + c_2 r^2 \exp(-\gamma_2 r)$, in the same spirit as Ref. [29]. The FN-DMC calculations are all-electron calculations done with a very small time step, $\tau = 0.00015$, to insure a proper treatment of the nodal hypersurfaces and to avoid the finite time-step error. For each geometry and for each CI-coefficient threshold, the calculations are quite extensive and represent about 10^{10} Monte Carlo steps distributed over a large number of processors. Tables I and II show the MR-FN-DMC values of the barrier to dissociation and the heat of formation

TABLE I. Barrier to dissociation in kcal/mol, O_4-TS.

Method	AVDZ basis	AVTZ basis	AVQZ basis
CASPT2:	4.7	6.5	6.6
CCSD(T)[a]	7.9	NA	NA
CASSCF + ACPF	7.7	9.1	9.3*
FN-DMC (HF nodes)		26.2 ± 2.9	
MR-FN-DMC (CAS nodes)		11.4 ± 1.6	

[a] From Ref. [13].

TABLE II. Heat of formation in kcal/mol, O_4-$2O_2$.

Method	AVDZ basis	AVTZ basis	AVQZ basis
CASPT2	105	100.5	101.1
CCSD(T)[a]	98.1	93.2	92.7
CASSCF + ACPF	99.8	95.2	95.0
FN-DMC (HF nodes)		140.2 ± 3.1	
MR-FN-DMC (CAS nodes)		98.5 ± 1.9	

[a] From Ref. [13].

compared with the previous *ab initio* CASSCF-ACPF benchmark values. The MR-FN-DMC values listed there correspond to the smallest CI threshold of 0.05 (the largest multireference trial wave functions). Note that a dramatic improvement in both thermochemical values is obtained when the truncated CASSCF expansions are used as multireference trial wave functions as compared with the HF trial ones; the FN-DMC values obtained with the HF trial wave functions are far away from the benchmark *ab initio* values, with errors of more than 100% for the barrier and more than 40% for the heat of formation. This shows the crucial role played by the nodes of the trial wave functions on the description of the correlation effects for the various nuclear geometries relevant in this reaction. Note that both state-of-the-art MR-FN-DMC thermochemical values are actually quite close to the benchmark *ab initio* ones but, thanks to the accuracy achieved, these data give a strong indication that the chemically bound O_4 species can actually be detected experimentally through the neutralization-reionization technique. Our results provide significant improvements on the previous estimates for the barrier and the heat of formation of this species, which is relevant when invoking its possible role in atmospheric processes.

We wish to thank the UAEM supercomputing center (No. SEP-FOMES2000), IDRIS (CNRS Orsay), CALMIP (Université de Toulouse), and CNSI (UC Santa Barbara) for unlimited superscalar CPU time. R. H. L. thanks CONACYT (México) through Project No. 44117E. A. R. S. thanks CONACYT (México) through Project No. 45986 and UCMEXUS/CONACYT for sabbatical support at UC Santa Barbara.

---

*Corresponding author.
alex@buzon.uaem.mx.
On sabbatical leave from Departamento de Física, Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos 62209, México.

[1] G. N. Lewis, J. Am. Chem. Soc. **46**, 2027 (1924).
[2] C. A. Long and G. E. Ewing, J. Chem. Phys. **58**, 4824 (1973).
[3] J. Goodman and L. E. Brus, J. Chem. Phys. **67**, 4398 (1977).
[4] P. E. S. Wormer and A. van der Avoird, J. Chem. Phys. **81**, 1929 (1984).
[5] B. Bussery and P. E. S. Wormer, J. Chem. Phys. **99**, 1230 (1993).
[6] V. Aquilanti, D. Ascenzi, M. Bartolomei, D. Cappelletti, S. Cavalli, M. de Castro Vitores, and F. Pirani, Phys. Rev. Lett. **82**, 69i (1999).
[7] V. Aquilanti, D. Ascenzi, M. Bartolomei, D. Cappelletti, S. Cavalli, M. de Castro Vitores, and F. Pirani, J. Am. Chem. Soc. **121**, 10 794 (1999).
[8] R. Hernández-Lamoneda, M. Bartolomei, M. I. Hernández, J. Campos-Martínez, and F. Dayou, J. Phys. Chem. A **109**, 11 587 (2005).
[9] F. A. Gorelli, L. Ulivi, M. Santoro, and R. Bini, Phys. Rev. Lett. **83**, 4093 (1999).
[10] J. B. Neaton and N. W. Ashcroft, Phys. Rev. Lett. **88**, 205503 (2002).
[11] H. Fujihisa, Y. Akahama, H. Kawamura, Y. Ohishi, O. Shimomura, H. Yamawaki, M. Sakashita, Y. Gotoh, S. Takeya, and K. Honda, Phys. Rev. Lett. **97**, 085503 (2006).
[12] V. Adamantides, D. Neisius, and G. Verhaegen, Chem. Phys. **48**, 215 (1980).
[13] E. T. Seidl and H. F. Schaefer, III, J. Chem. Phys. **96**, 1176 (1992).
[14] D. S. Peterka, M. Ahmed, A. G. Suits, K. J. Wilson, A. Korkin, M. Nooijen, and R. J. Bartlett, J. Chem. Phys. **110**, 6095 (1999).
[15] D. Schröder, Angew. Chem., Int. Ed. Engl. **41**, 573 (2002).
[16] R. Hernández-Lamoneda and A. Ramírez-Solís, J. Chem. Phys. **113**, 4139 (2000).
[17] R. Hernández-Lamoneda and A. Ramírez-Solís, J. Chem. Phys. **120**, 10 084 (2004).
[18] H. Helm and C. W. Walter, J. Chem. Phys. **98**, 5444 (1993).
[19] F. Cacace, G. de Petris, and A. Troiani, Angew. Chem., Int. Ed. **40**, 4062 (2001).
[20] H. Park and T. G. Slanger, J. Chem. Phys. **100**, 287 (1994).
[21] R. Hernández, R. Toumi, and D. C. Clary, J. Chem. Phys. **102**, 9544 (1995).
[22] R. L. Miller, A. G. Suits, P. L. Houston, R. Toumi, J. A. Mack, and A. M. Wodtke, Science **265**, 1831 (1994).
[23] D. M. Lauvergnat and D. C. Clary, J. Chem. Phys. **108**, 3566 (1998).
[24] B. L. Hammond, W. A. Lester, Jr., and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, Singapore, 1994).
[25] There also exist some other sources of errors (finite time step, finite population, etc.), but all these errors are easily extrapolated to zero and, thus, are not discussed here.
[26] A. Aspuru-Guzik, O. El Akramine, J. C. Grossman, and W. A. Lester, Jr., J. Chem. Phys. **120**, 3049 (2004).
[27] K. E. Riley and J. B. Anderson, Mol. Phys. **101**, 3131 (2003).
[28] E. Clementi and C. Roetti, At. Data Nucl. Data Tables **14**, 177 (1974).
[29] S. Manten and A. Lüchow, J. Chem. Phys. **115**, 5362 (2001).