

# Bond Breaking and Bond Making in Tetraoxygen: Analysis of the $O_2(X^3\Sigma_g^-) + O_2(X^3\Sigma_g^-) \rightleftharpoons O_4$ Reaction Using the Electron Pair Localization Function

Anthony Scemama,<sup>†</sup> Michel Caffarel,<sup>†</sup> and Alejandro Ramírez-Solís\*,<sup>‡</sup>

*Laboratoire de Chimie et Physique Quantiques, CNRS-IRSAMC Université de Toulouse, France, and Departamento de Física, Facultad de Ciencias, Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos 62209, México*

Received: March 5, 2009; Revised Manuscript Received: June 12, 2009

We study the nature of the electron pairing at the most important critical points of the singlet potential energy surface of the  $2O_2 \rightleftharpoons O_4$  reaction and its evolution along the reaction coordinate using the electron pair localization function (EPLF) [Scemama, A.; Chaquin, P.; Caffarel, M. *J. Chem. Phys.* **2004**, *121*, 1725]. To do that, the 3D topology of the EPLF calculated with quantum Monte Carlo (at both variational and fixed-node-diffusion Monte Carlo levels) using Hartree–Fock, multiconfigurational CASSCF, and explicitly correlated trial wave functions is analyzed. At the  $O_4$  equilibrium geometry the EPLF analysis reveals four equivalent covalent bonds and two lone pairs on each oxygen atom. Along the reaction path toward dissociation it is found that the two oxygen–oxygen bonds are not broken simultaneously but sequentially, and then the lone pairs are rearranged. In a more general perspective, the usefulness of the EPLF as a unique tool to analyze the topology of electron pairing in nontrivial chemical bonding situations as well as to visualize the major steps involved in chemical reactivity is emphasized. In contrast with most standard schemes to reveal electron localization (atoms in molecules, electron localization function, natural bond orbital, etc.), the newly introduced EPLF function gives a *direct* access to electron pairings in molecules.

## Introduction

The search for a chemically bound form of tetraoxygen has been the subject of numerous investigations during the last 30 years. Initially, the motivation was given by the analogy with the stable and well-known cyclic and chain forms of sulfur, the study of its fundamental properties and its possible use as a high energy-density material.<sup>1,2</sup> More recently, it has been proposed as a possible intermediate explaining a variety of experimental findings and appears to have been detected by neutralization reionization (NR) mass spectrometry. Given the different experimental conditions for its preparation and identification, it is important to point out that several forms of tetraoxygen are most likely involved in the different experimental setups, which could include a variety of electronic states and isomers covering a wide range of intermolecular strengths. For example, in the photoionization spectra it has been shown<sup>3,4</sup> that the  $O_4$  metastable state corresponds to an excited complex between an  $O_2$  molecule in its ground state and another in the excited state,  ${}^1\Sigma_u$ . Furthermore, convincing evidence exists that this same species could be involved in the electron transfer to  $O_4^+$  experiments of Helm and Walter.<sup>5</sup> Although these gas-phase spectroscopy experiments have suggested the existence of an  $O_4$  species, it is clear that it is not the chemically bound form studied here but rather electronically excited van der Waals complexes.<sup>6–8</sup> The same conclusion was drawn in the recent NR mass spectrometric detection of an  $O_4$  species.<sup>9</sup>

Most of the complexity found in oxygen-containing species and reactions arises from the unusual electronic structure of molecular oxygen: it has two spin-unpaired electrons occupying degenerate  $\pi_g$  orbitals which lead to a  $\Sigma_g^-$  ground state with a

total spin of  $S = 1$  and two singlet low-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<sup>10</sup> 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  $O_2$ . It is also known 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.<sup>11–14</sup>

In the solid state, oxygen is the only antiferromagnetic insulating phase among the elemental solids and 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.<sup>15–17</sup> One of the most interesting open problems is the determination and explanation of the structural and optical properties of the  $\epsilon$  phase, which is stable in a broad range of temperatures and high pressures. Some years ago Bini and co-workers<sup>15</sup> measured the infrared spectrum as a function of pressure and noticed the appearance of a new absorption band in the 300–600  $\text{cm}^{-1}$  region, in addition to the expected molecular absorption at  $\sim 1500 \text{ cm}^{-1}$ . 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<sup>16</sup> who proposed, on the basis of density functional theory, a structure

\* To whom correspondence should be addressed. E-mail: alex@uaem.mx.

<sup>†</sup> Université de Toulouse.

<sup>‡</sup> Universidad Autónoma del Estado de Morelos.

based on linear herringbone-type chains of  $O_2$  molecules consistent with the observed infrared spectra. However, very recently a crystal structure determination has been performed<sup>17</sup> which suggests that the basic unit  $O_8$  is composed of four  $O_2$  molecules, 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,<sup>2,18</sup> 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 it would be natural, given its small number of electrons, to expect that state-of-the-art computational methodologies of quantum chemistry should lead to a full knowledge of the electronic properties of  $O_4$  and its dissociation reaction path to molecular oxygen, this is not the case; a detailed understanding of the chemically bound  $O_4$  molecule still remains a significant challenge for quantum chemistry,<sup>2,3,18–20</sup> although some important progress has been done very recently.<sup>21</sup> The quantum theoretical difficulties are important and can be briefly stated as follows: the singlet  $O_4$  equilibrium structure has a relatively strong monoreference character, thus it can in principle be calculated by standard and accurate coupled cluster (CC) methods. However, describing the transition state of  $C_2$  symmetry 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/making processes, and the complicated spin recouplings necessary to describe the transition from four open-shells to a closed shell species can only be properly reproduced with large multiconfigurational wave functions. In practice, all along the reaction coordinate the smallest complete active space SCF (CASSCF) wave functions have to be built with the three 2p shells and the four valence electrons of each atom. This leads to 16 electrons within 12 active orbitals [a CASSCF(16,12) calculation], that is to say, large CASSCF expansions of 17865 and 70785 configuration state functions (CSF) at the equilibrium and transition state geometries, respectively. On top of that, it is also essential to describe in a *balanced way* the subtle electron–electron interactions, i.e., the dynamic correlation effects, arising at the various geometries corresponding to the  $O_4$  reactant, the transition state, and the  $O_2(X^3\Sigma_g^-) + O_2(X^3\Sigma_g^-)$  products, all of them lying on the lowest singlet potential energy surface.

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 in various phases. The need of an accurate estimation of the heat of formation of this species has already been noted.<sup>18</sup> In a benchmark study using the CCSD(T) and CASSCF(16,12)+ACPF methods with the large aug-cc-pVQZ basis set,<sup>20</sup> 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. Therefore, in a very recent work,<sup>21</sup> we have shown that accurate thermochemical values 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. A favorable

**TABLE 1: CASSCF Wavefunction of  $O_4$  at Equilibrium Geometry,  $\Psi = \sum_i c_i | \dots \rangle^a$**

configuration	coefficient $c_i$
$a_1$ 2220, $a_2$ 220, $b_1$ 220, $b_2$ 20>	0.9081534
$a_1$ 2220, $a_2$ 202, $b_1$ 220, $b_2$ 20>	-0.0749640
$a_1$ 2220, $a_2$ 220, $b_1$ 220, $b_2$ 02>	-0.0661446
$a_1$ 2220, $a_2$ 220, $b_1$ 2-+, $b_2$ +->	-0.0600672
$a_1$ 2220, $a_2$ 220, $b_1$ 2+-, $b_2$ +->	-0.0600672
$a_1$ 2220, $a_2$ 220, $b_1$ 2-, $b_2$ ++>	0.0598511
$a_1$ 2220, $a_2$ 220, $b_1$ 2++, $b_2$ ->	0.0598511
$a_1$ 2220, $a_2$ 2-+, $b_1$ 220, $b_2$ +->	0.0580685
$a_1$ 2220, $a_2$ 2-+, $b_1$ 220, $b_2$ -+>	0.0580685
$a_1$ 2220, $a_2$ 2-, $b_1$ 220, $b_2$ ++>	-0.0577911
$a_1$ 2220, $a_2$ 2++, $b_1$ 220, $b_2$ -->	-0.0577911
$a_1$ 2220, $a_2$ 220, $b_1$ 222, $b_2$ 00>	-0.0574860
$a_1$ 2220, $a_2$ 222, $b_1$ 220, $b_2$ 00>	-0.0574154
$a_1$ 2222, $a_2$ 220, $b_1$ 220, $b_2$ 20>	-0.0523990
$a_1$ 2+2-, $a_2$ 220, $b_1$ 2-+, $b_2$ 20>	0.0510929
$a_1$ 2-2+, $a_2$ 220, $b_1$ 2+-, $b_2$ 20>	0.0510929
$a_1$ 2220, $a_2$ 2-+, $b_1$ 2-+, $b_2$ 20>	-0.0509593
$a_1$ 2220, $a_2$ 2-+, $b_1$ 2-+, $b_2$ 20>	-0.0509593
$a_1$ 2220, $a_2$ 220, $b_1$ 200, $b_2$ 22>	-0.0504427

<sup>a</sup> All coefficients  $c_i$  having a modulus greater than 0.05 are given. For the second quantized notation, see text. Molecular orbitals are grouped according to their symmetry (using the  $a_1$ ,  $a_2$ ,  $b_1$ , and  $b_2$  irreducible representations of the  $C_{2v}$  symmetry group).

comparison with the previous ab initio benchmark CASSCF+ACPF/AVQZ results<sup>20</sup> was possible. The QMC barrier to dissociation and heat of formation obtained are  $11.6 \pm 1.6$  kcal/mol and  $98.5 \pm 1.9$  kcal/mol, respectively.

Given that for the  $O_4 \rightleftharpoons O_2(X^3\Sigma_g^-) + O_2(X^3\Sigma_g^-)$  reaction the main problem found with standard ab initio methods for the accurate evaluation of the reaction barrier is the need to accurately deal with the dynamic correlation effects (on top of the strongly multiconfigurational wave function) at the transition state geometry, it would seem that the electronic distributions of the bonds around the four oxygen nuclei strongly depend on the method used to describe it. We note that, even at the equilibrium geometry of the  $O_4$  molecule, the CASSCF wave function contains 24 determinants with a coefficient larger than 0.05 (see Table 1).

Although ground-state energies are of primary interest, it is clear that a chemical understanding of a complex molecular system requires more than the knowledge of the sole ground-state energy. In general, chemists are interested in rationalizing and quantifying the structure and reactivity of the system in terms of various quantities related to electron localization. In standard computational chemistry several approaches have been developed to analyze and visualize the electronic distribution in the ordinary 3D space. Among them we can cite, e.g., the methods analyzing the deformation of densities (a build-up of charge between two atoms is interpreted as the existence of a bond),<sup>23</sup> the methods based on the topological analysis of the electron density or its Laplacian (see, for instance, Bader<sup>24</sup>), the methods studying the topography of the molecular electrostatic field<sup>25</sup> and, also, approaches using as indicator the electron localization function (ELF) describing the amount of local Pauli repulsion between electrons.<sup>26,27</sup> Of course, this list cannot be considered to be exhaustive since defining a successful and general qualitative model for the description of chemical structure is an everlasting theme in chemistry since the pioneering electron-pair model of Lewis.

In this work we propose to exploit the accurate data obtained from QMC simulations on the  $O_4 \rightleftharpoons 2O_2$  reaction to get important insights into the electron localization properties of

the tetraoxygen molecule at its equilibrium geometry, at the transition state, and also along the reaction coordinate leading to the two ground-state oxygen molecules. To do that we shall use the function, introduced by some of us recently,<sup>28</sup> describing the pairing of electrons in a molecular system. This function, called the electron pair localization function (EPLF), is built to reveal the differences in the average distances between spin-like and spin-unlike electrons. In regions where localized pairs of electrons are present (lone pairs, atomic pairs, bonds) the EPLF takes large values and displays maxima. In contrast, in regions where electrons behave essentially as an homogeneous fluid (spin-like and spin-unlike electrons being mixed together), the EPLF takes much smaller values. The form of the EPLF is simple (see next section) and has been chosen to be easily computable using QMC calculations. Originally applied to several simple atomic and molecular systems,<sup>28</sup> the EPLF appears to be a practical tool for describing electronic features in more complex molecular systems.<sup>22</sup> We shall apply it here to a quite challenging chemical problem, namely, the  $O_4 \rightleftharpoons O_2(X^3\Sigma_g^-) + O_2(X^3\Sigma_g^-)$  reaction.

**QMC.** In a QMC scheme a series of “states” or “configurations” are generated using some elementary stochastic rules. Here, a configuration is defined as the set of the  $3N$ -electronic coordinates ( $N$  number of electrons), the positions of the nuclei being fixed (Born–Oppenheimer condition)

$$\vec{R} = (\vec{r}_1, \dots, \vec{r}_N) \quad (1)$$

Stated differently, a configuration  $\vec{R}$  may be viewed as a “snapshot” of the molecule showing the instantaneous positions of each electron. Stochastic rules are chosen so that configurations are generated according to some target probability density,  $\Pi(\vec{R})$ . Many variants of QMC can be found in the literature (referred to with various acronyms: VMC, DMC, PDMC, GFMC, etc.). They essentially differ by the type of stochastic rules used and/or by the specific stationary density produced. Here, we shall employ the two most popular QMC approaches, namely the variational Monte Carlo (VMC) and fixed-node diffusion Monte Carlo (FN-DMC) methods. The basic features of these two methods useful for the following are now briefly summarized (a detailed presentation is presented in a previous work<sup>29</sup>).

**VMC.** In a VMC calculation the probability density generated is given by

$$\Pi_{\text{VMC}}(\vec{R}) = \frac{\psi_T^2(\vec{R})}{\int d\vec{R} \psi_T^2(\vec{R})} \quad (2)$$

where  $\psi_T$  is an accurate electronic trial wave function. The most commonly used expression for  $\psi_T$  is written as a product of two terms. The first term is standard and describes the one-particle shell structure of the molecule. It is obtained from a preliminary Hartree–Fock (HF) or DFT ab initio calculation and is expressed as one (or a combination of a few) determinant(s) of single-particle spatial orbitals. The second term, the so-called Jastrow factor, is introduced to reproduce the electron–electron cusp condition of the exact wave function and, also, to incorporate some explicit coupling between electron–nucleus and electron–electron coordinates.<sup>30</sup> Here, we have chosen a standard form.<sup>31</sup>

An important step in a VMC approach is the optimization of the parameters entering the trial wave function. A number of

methods have been developed to perform efficiently such an optimization. In this work, we have used the correlated sampling method of Umrigar et al.,<sup>32</sup> an approach based on the minimization of the variance of the energy over a set of fixed configurations. Note that once the optimal parameters have been determined, the quality of the resulting trial wave function is usually good. A major part of the dynamical correlation energy (Coulomb hole) is recovered, and the gross features of the one-particle background are also correctly described via the determinantal part (i.e., the nondynamical correlation). For most atoms it is possible to recover up to 80–90% of the exact correlation energy;<sup>30</sup> for molecules the domain of variation lies usually between 30 and 90%.

The numerical method (stochastic rules) employed to generate the VMC density, eq 2, is standard and is based on the use of an improved Metropolis algorithm.<sup>33</sup>

**DMC.** In a DMC scheme the stochastic rules employed are the same as in the VMC case (Metropolis algorithm) plus a new rule corresponding to a branching (or birth-death) process. More precisely, depending on the magnitude of the local energy,  $E_L \equiv H\Psi_T/\Psi_T$ , a given configuration is destroyed (when the local energy is greater than some estimate of the exact energy) or duplicated a certain number of times (local energy lower than the exact energy). It can be shown that the stationary density resulting from these rules is now given by

$$\Pi_{\text{DMC}}(\vec{R}) = \frac{\psi_T(\vec{R})\phi_0(\vec{R})}{\int d\vec{R} \psi_T(\vec{R})\phi_0(\vec{R})} \quad (3)$$

where  $\phi_0(\vec{R})$  denotes the unknown ground-state wave function.

**Fixed-Node Error.** Actually, because the density  $\Pi_{\text{DMC}}$  is necessarily positive, as any stationary density resulting from some stochastic rules,  $\phi_0$  is not the exact ground-state wave function but some approximate one resulting from the additional constraint that  $\phi_0$  must have the same sign as the trial wave function so that the product in eq 3 is always positive. In other words, the mathematical eigenproblem solved is not the exact one but, rather, some modified one which can be written as

$$H\phi_0^{\text{FN}}(\vec{R}) = E_0^{\text{FN}}\phi_0^{\text{FN}}(\vec{R}) \quad (4)$$

where  $\phi_0^{\text{FN}}(\vec{R}) = 0$  whenever  $\psi_T(\vec{R}) = 0$ .

The fact that the nodes (points in  $3N$ -dimensional space where the wave function vanishes) of  $\psi_T$  and  $\phi_0^{\text{FN}}$  are identical leads to a so-called “fixed-node” error. However, as far as total energies are concerned, this approximation is in general very good, and the fixed-node error on total energies represents usually a small fraction of the total correlation energy. Let us emphasize that this error depends only on the quality of the nodes.<sup>34</sup>

**Multireference FN-DMC.** We have recently shown that for the problem we are interested in here, the thermochemical quantities obtained with FN-DMC (reaction barrier and dissociation energy) are very badly reproduced when the single-reference Hartree–Fock nodes are used, both for the equilibrium and for the transition-state geometries.<sup>21</sup> To get a coherent description of the nodal patterns at both geometries we proposed a scheme based on the study of the fixed-node error evolution for both structures, as a function of the size of a truncated expansion of the large CASSCF(16,12) wave function used as DMC trial wave function. Note that the basis set employed here is the large aug-cc-pVTZ basis set. The truncation is simply

done by choosing the configurations within the CASSCF expansion that have a coefficient larger than a given threshold  $\varepsilon$ , thus defining a subset of configurations at each geometry. By use of this strategy, we consider that the fixed-node DMC energy difference obtained is meaningful only when the two following conditions are fulfilled. First, the energy difference as a function of decreasing values of  $\varepsilon$  must be found almost constant within statistical fluctuations. Second, the nodal error must display some “robustness” with respect to dynamical correlation effects. By this, it is meant that the nodal patterns must not change significantly under reoptimization (energy minimization) of the coefficients of the truncated expansion  $c_i$  in the presence of the Jastrow factor.

From a more fundamental point of view, this procedure is based on the assumption that the nodal pattern of the exact wave function can be correctly described by considering *separately* the nondynamical correlation effects (CASSCF part) and the dynamical ones (Jastrow part). Clearly, this “first-order” approximation is natural but is not expected to be valid for all molecular systems.

**EPLF.** The EPLF is a local scalar function defined in the ordinary 3D-space, bounded above and below, which has been designed to reveal the localization of electron pairs. It is a good descriptive tool for chemical systems since electron pairs play a central role in our everyday interpretation of chemical structure and reactivity (Lewis model, VSEPR). The framework proposed to calculate such a localization function is that of QMC approaches. As emphasized in the introduction, QMC are techniques of a great versatility, and therefore, the definition of the EPLF described below will be of practical use for any type of wave functions (HF, post-HF, valence bond, etc.) and for any level of computation (VMC, FN-DMC, “exact”).

First, we need to introduce the two local quantities  $d_{oo}(\vec{r})$  and  $d_{o\bar{o}}(\vec{r})$  defined as follows

$$\begin{aligned} d_{oo}(\vec{r}) &\equiv \sum_{i=1}^N \langle\langle \delta(\vec{r} - \vec{r}_i) \min_{j \neq i; \sigma_j = \sigma_i} |\vec{r}_i - \vec{r}_j| \rangle\rangle \\ d_{o\bar{o}}(\vec{r}) &\equiv \sum_{i=1}^N \langle\langle \delta(\vec{r} - \vec{r}_i) \min_{j: \sigma_j \neq \sigma_i} |\vec{r}_i - \vec{r}_j| \rangle\rangle \end{aligned} \quad (5)$$

where  $\{\vec{r}_k\}_{k=1}^N$  are the positions of the  $N$  electrons for a given configuration  $R$ ,  $\sigma_i$  the spin of the  $i$ th electron ( $\sigma_i = \uparrow, \downarrow$ ), and  $\langle\langle \dots \rangle\rangle$  the stochastic average over the Monte Carlo configurations. As seen from these definitions the physical meaning of the functions  $d(\vec{r})$  is clear:  $d_{oo}(\vec{r})$  [respectively,  $d_{o\bar{o}}(\vec{r})$ ] is the average distance between an electron located at  $\vec{r}$  and the closest spin-like (respectively, spin-unlike) electron of the molecule.

The EPLF is defined as

$$\text{EPLF}(\vec{r}) = \frac{d_{oo}(\vec{r}) - d_{o\bar{o}}(\vec{r})}{d_{oo}(\vec{r}) + d_{o\bar{o}}(\vec{r})} \quad (6)$$

By definition the EPLF takes its values within the interval  $[-1, 1]$ . It gives a local indicator of electron pairing as follows. In regions of space where electrons are unpaired the average distances between spin-like and spin-unlike electrons are similar,  $d_{o\bar{o}} \approx d_{oo}$ , and the EPLF goes to zero. When spin-unlike electrons are paired we have  $d_{o\bar{o}} \ll d_{oo}$ , and EPLF goes to 1. Finally, when spin-like electrons are paired,  $d_{o\bar{o}} \gg d_{oo}$ , and thus, EPLF goes to -1. The EPLF main feature is to reveal the local differences ( $\vec{r}$ -dependence) in the average distances between

spin-like and spin-unlike electrons. In regions where localized electron pairs are present (lone pairs, atomic pairs, bonds) the EPLF takes large values and displays maxima. In contrast, in regions where electrons behave essentially as an homogeneous fluid (spin-like and spin-unlike electrons being mixed together), the EPLF takes on much smaller values. Note that for molecules with one or more open shells, as is the case with the molecular oxygen treated here, there exist regions with large spin-up (or spin-down) densities and the EPLF gives minimal values. Another point to mention is that in eq 6 defining the EPLF the important physical quantities are the  $d$  functions and not the precise definition of the EPLF in terms of such quantities. Many choices could have been possible and the simplest choice has been considered. Note that some authors have proposed alternative definitions such as the EPLFN form.<sup>22</sup> However, according to our experience the various forms do not lead to significantly different interpretations of chemical bonding. Finally, let us emphasize that the definition of EPLF is particularly well suited to QMC; eq 5 can indeed be easily computed with any QMC scheme.

## Computational Details

**Trial Wave Functions.** The oxygen basis set used is the fully decontracted Dunning aug-cc-pVTZ basis set,<sup>35</sup> and the optimized geometries chosen for the O<sub>4</sub> reactant and the transition state are those previously determined at the multireference Rayleigh–Schrödinger perturbed second-order (CASSCF-RS2) level, which are the most accurate ones to date. The determination of the critical points of the singlet potential energy surface was done through benchmark-type CASSCF-RS2 calculations with a very large monoelectronic basis of 320 molecular orbitals and taking into account up to  $5.1 \times 10^{10}$  CSF.<sup>20</sup> The trial wave functions used here consist of the truncated CASSCF(16,12) determinantal expansion discussed above multiplied by a standard Jastrow prefactor taking into account the explicit electron–electron and electron–electron–nucleus interactions.<sup>31,36</sup> Note that for a system consisting of light atoms such as O<sub>4</sub> some care has to be taken for properly reproducing the electron–nucleus cusp both for the core and valence electrons. With regard to 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.<sup>37</sup> 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  $c(1 + \alpha r^2 + \beta r^3) \exp(-\gamma r)$ , in the same spirit as a previous study.<sup>38</sup> The present FN-DMC calculations are all-electron calculations done with a very small time-step,  $\tau = 0.00015$ , to ensure a proper treatment of the nodal hypersurfaces and to reduce time-step errors. For each trial wave function, the calculations are not extensive and represent only about  $10^7$  Monte Carlo steps distributed over 20 processors. Since the EPLF gives only qualitative information, the number of steps needed to obtain relevant results is much smaller than the number of steps needed to get an acceptable error bar for the energy. Moreover, it should be stressed that, in sharp contrast with the energy, the quality of the trial wave function has little impact on the magnitude of the statistical error on the EPLF results, since the quantities which are averaged during the Monte Carlo simulation, eq 5, are not a function of the trial wave function.

In the case of the multireference trial wave functions, we have used the same wave functions as the largest reported<sup>21</sup> with a cutoff confidence interval (CI) coefficient of 0.05 in the

CASSCF expansions at both (the transition state (TS) and product) geometries, leading to 24 and 14 CSF for the equilibrium product and TS geometries, respectively. These wave functions are used to represent the determinantal part of the trial functions used in the QMC computations.

**QMC Data.** The densities relative to HF wave functions, eq 2, have been sampled by VMC simulations and are labeled HF-VMC. The densities relative to the truncated CASSCF wave functions which have been sampled by VMC simulations and are labeled MR-VMC. The densities relative to truncated CASSCF wave functions which have been sampled by multi-reference FN-DMC simulations are labeled MR-FNDMC.

The trial wave functions have also been improved by introducing an explicitly correlated Jastrow factor as described previously.<sup>31</sup> The densities corresponding to these wave functions have been sampled by VMC, labeled Jast-VMC.

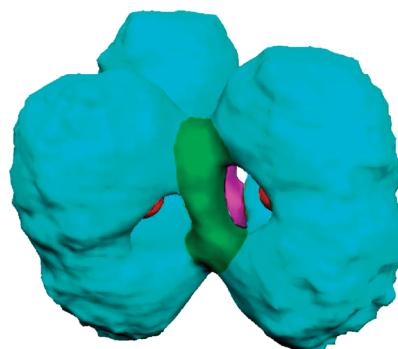
For each geometry considered, a typical simulation includes a set of 800 independent walkers and a number of Monte Carlo steps per walker of about 100000.

**EPLF Data.** The continuous 3D-space is represented using a  $80 \times 80 \times 80$  three-dimensional grid. The EPLF is calculated as follows: for each generated Monte Carlo configuration the positions of the electrons are scanned. The elementary volume of the 3D-grid occupied by each electron is determined, and the minimum distances appearing in the definition of EPLF are calculated. The noise in the localization function due to the statistical character of QMC simulations has been reduced by using a median blur filter as detailed previously.<sup>28</sup> This filter is particularly well adapted here since it is known to modify very little the regions where the gradient is large. This latter point is particularly important here since we are interested in altering as little as possible the contours of the pair localization function.

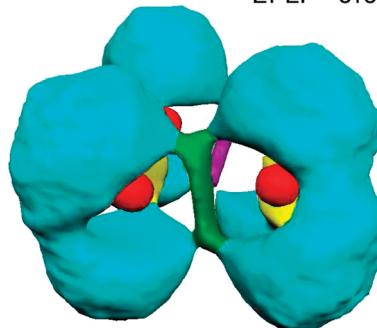
## EPLF Results

As we shall see below, the topology of the EPLF for the various systems and geometries presented here are found to be sensitive to nondynamical correlation (introduced via CASSCF expansions) but much less to dynamical correlation (via VMC simulations using explicitly correlated Jastrow factors and/or DMC simulations). Accordingly, to keep the number of figures shown reasonable, only the EPLF figures obtained from calculations done at the HF and truncated CASSCF(16,12) levels are presented.

**The  $O_4$  Equilibrium Geometry.** In Figures 1 and 2, we present the 3D-plots for the EPLF obtained with the HF and the truncated CASSCF(16,12) wave functions, respectively. In both cases, two figures are presented, one corresponding to an isosurface value of 0.08 and the other of 0.09. These two figures are given to illustrate the topological effect of using different EPLF values. The EPLF topology for the HF and CASSCF cases is very similar and leads to a clear chemical picture. As seen from these figures, the EPLF reveals the atomic shell structure: one small spherical domain around each oxygen nucleus corresponding to the  $1s$  electron pairs (red) and two separate domains on each oxygen corresponding to the lone pairs (blue). Regarding the bonding pattern, four disk-shaped domains corresponding to four equivalent  $\sigma$  bonds are observed (one in green, one in magenta, and two in yellow), these domains being exactly located at the midpoints of the four internuclear oxygen–oxygen axes. Note that when a smaller EPLF value is used to draw the isosurface, some narrow “bridges” can be seen connecting the lone pairs and the bonds between the oxygen atoms. Also clearly visible is the local tetrahedral structure around each oxygen atom between its two lone pairs and the

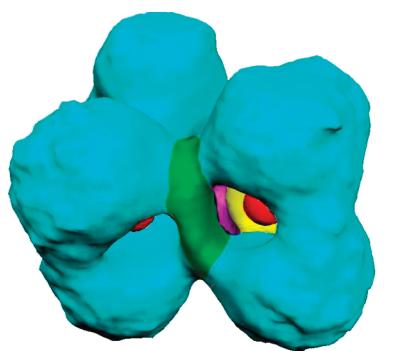


EPLF=0.08

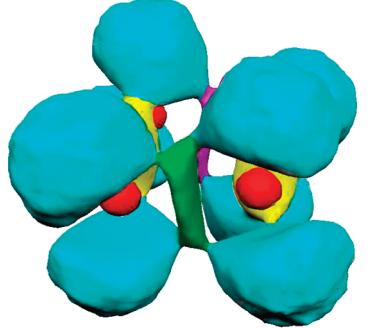


EPLF=0.09

**Figure 1.** Isosurface of the EPLF for  $O_4$  at the equilibrium geometry using a HF trial wave function. The figure in the upper part has been obtained with an EPLF isosurface value of 0.08. The figure in the lower part corresponds to a value of 0.09.



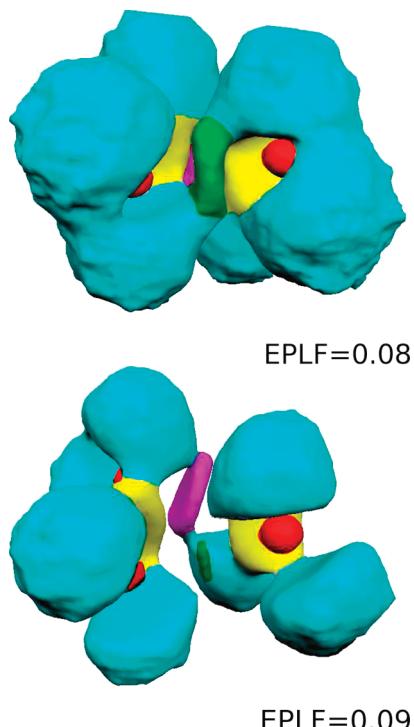
EPLF=0.08



EPLF=0.09

**Figure 2.** Isosurface of the EPLF for  $O_4$  at the equilibrium geometry using a truncated CASSCF(16,12) trial wave function. The isosurface values are identical to the HF case (see Figure 1).

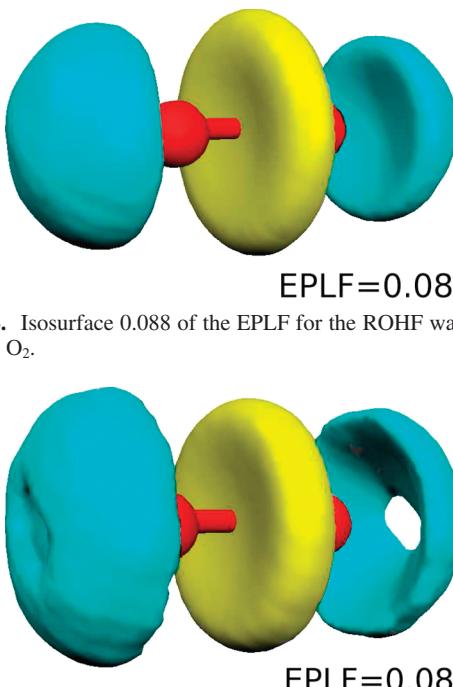
two covalent bonds formed with the two closest neighbors ( $sp^3$  hybridization). Note that the optimized interoxygen distances in this molecule are very close to other oxygen–oxygen distances corresponding to single bonds.<sup>20</sup>



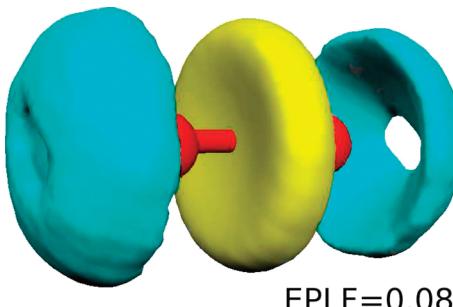
**Figure 3.** Isosurfaces of the EPLF of the transition state at the CASSCF level. Lone pairs are colored in cyan; core shells are in red. O–O bonds are colored in yellow, green, and magenta.

**The TS.** The EPLF obtained for the CASSCF(16,12) wave function at the TS geometry (with a truncation threshold of 0.05 for the CI coefficients) are displayed in Figure 3; as previously done for the equilibrium geometry, we have made two isosurfaces with different EPLF values of 0.08 and 0.09, and we have used the same perspective to show the nonequivalence of the two pairs of oxygen atoms; recall that there are two short equivalent O–O distances, one intermediate and another which is longer. A few words are in order here to explain the main difference with the equilibrium geometry. Note that, while in Figure 2 the EPLF is completely equivalent as seen from the midpoint of any O–O bond, in Figure 3 there is a clear distinction of a couple of atoms which lie further away from each other than the other three couples; this is reflected in the fact that one of the “bridges” (shown in green) connecting the lone pairs and the bond between these two atoms is significantly smaller and nearly gone. The core 1s (red) and three of the interatomic  $\sigma$  bond domains are practically the same as those seen in Figure 2 in good agreement with the analysis presented in ref 18 where one of the O–O bonds appears to be already broken at the TS geometry.

**The Ground-State Triplet Oxygen Molecule.** The ROHF and the CASSCF valence wave functions of the triplet oxygen molecule are very similar and can be well described (in second-quantized notation) as a single  $|222 + +\rangle$  determinant, where the last two are degenerate  $\pi$  molecular orbitals. The EPLF obtained respectively for the ROHF and for the CASSCF wave functions are displayed in Figures 4 and 5. For both wave functions, the EPLF exhibits a maximum value describing the 1s orbital of each oxygen atom (red). A central “pancake-like” domain (yellow) located at the midpoint along the internuclear axis corresponds to an axial  $\sigma$  bond superimposed on the two orthogonal  $\pi$  bonds. The EPLF values corresponding to the  $\pi$  bonds are lowered by the presence of the unpaired electrons (located in the  $\pi$  antibonding orbital). The superposition of these two bonds leads to a maximum value located at the center of



**Figure 4.** Isosurface 0.088 of the EPLF for the ROHF wave function of triplet  $O_2$ .



**Figure 5.** Isosurface 0.088 of the EPLF for the CASSCF wave function of triplet  $O_2$ .

the molecule since the  $\sigma$  contribution to the EPLF is higher than the  $\pi$  one. In the ROHF case, the oxygen lone pairs (blue) are characterized by two maxima of the EPLF located on the internuclear axis, outside of the O–O bond. These maxima are produced by the contributions of the doubly occupied orbitals with large coefficients on the  $2p_z$  atomic orbitals. In the CASSCF case, the second most important configuration is a  $|22 + + 22\rangle$  determinant. In this determinant, one electron of each lone pair has been promoted to an antibonding  $\pi^*$  orbital. The contribution of this configuration changes the EPLF by lowering its value along the molecular axis outside of the O–O bond. This results in a “hole” made in the lone pair domain which is visible in Figure 5. Note that the presence of this hole, which is absent in the ROHF case, Figure 4, illustrates very clearly the impact of nondynamical correlation effects on the pairing distribution.

**Bond Forming/Bond Breaking.** The analysis of the EPLF along the reaction path connecting the three cases just discussed ( $O_4$  molecule, TS, and the two triplet  $O_2$  molecules) turns out to be more challenging. One could naïvely think that the identification of the bond-forming and bond-breaking processes of this reaction is a straightforward matter that could be simply done by analyzing the evolution of the wave function along the reaction coordinate. At the HF level it could theoretically be done in terms of the HF canonical orbitals, while for the truncated CASSCF case this could be achieved in terms of the natural orbitals and the CSF. However, the presence of a large number of configurations, both at the  $O_4$  equilibrium geometry and at the TS, casts serious doubts on the validity of the HF description of this complex reaction. In practice, the identification of the bond-forming and bond-breaking processes is quite a tough and tedious task. In particular, an important necessary step is to localize the molecular orbitals, a task for which there does not exist a general criterium-free scheme and for which some fragment reference is needed. Given the equivalence of the four oxygen atoms along the reaction coordinate close to the TS, there is a practical problem to define coherently the reference fragments on both the negative (toward  $O_4$ ) and the positive (toward two  $O_2$  molecules) reaction coordinate. We

**TABLE 2: Singlet CASSCF Wavefunctions at the TS Geometry,  $\Psi = \Sigma_i c_i |...i\rangle^a$** 

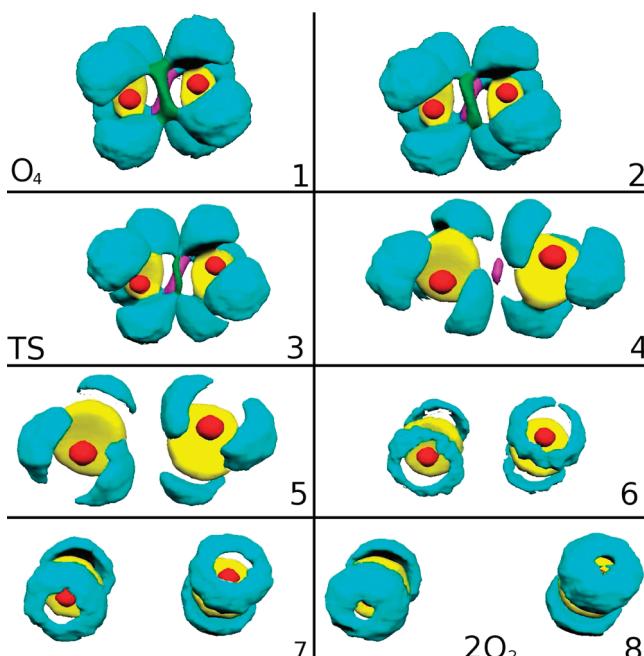
configuration	coefficient $c_i$
$ 222220\ 222000\rangle$	0.8670236
$ 222200\ 222200\rangle$	-0.3348786
$ 222020\ 222020\rangle$	-0.1620731
$ 222220\ 220002\rangle$	-0.0811422
$ 22+22-\ 22-00+\rangle$	-0.0779391
$ 22-22+\ 22+00-\rangle$	-0.0779391
$ 220222\ 222000\rangle$	-0.0731238
$ 222000\ 222220\rangle$	0.0723549
$ 222222\ 220000\rangle$	-0.0710389
$ 22+22+\ 22-00-\rangle$	0.0620807
$ 22-22-\ 22+00+\rangle$	0.0620807
$ 220220\ 222002\rangle$	-0.0566175

<sup>a</sup> All coefficients  $c_i$  having a modulus greater than 0.05 are given. For the second quantized notation in two irreps, see text.  $C_2$  symmetry is used.

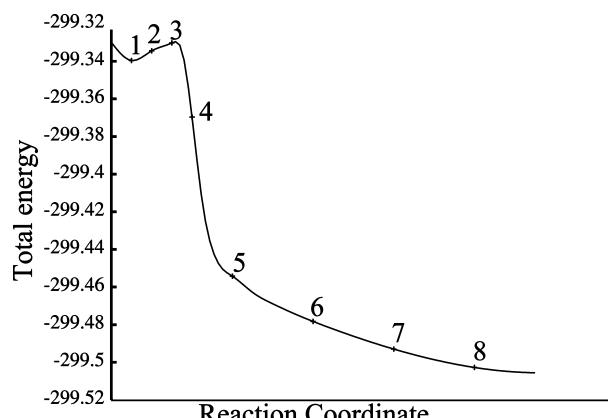
emphasize that for the truncated CASSCF the situation is even more difficult than in the HF case, because of the strongly multiconfigurational character of the wave function; this point is clearly illustrated in Table 2 where the coefficients of the CSF whose magnitude is greater than 0.05 are presented.

As mentioned above for the TS, one can remark that the shortest O–O bonds are equivalent. Then, the two other O–O bonds are not equivalent: the values of the EPLF of these O–O domains are different. Therefore, the bond breaking process occurs in three steps: (i) the bond represented by the green domain in Figure 3 is broken, (ii) then the bond represented by the domain in magenta is broken, and (iii) then the last stage involves a topological rearrangement of the lone pair domains on each oxygen atom separately leading to two open-shells in both  $O_2$  molecules. The first step occurs in the “uphill” region of the PES from the equilibrium geometry to the TS (with a cost of around 11 kcal/mol according to our best multireference QMC estimate<sup>21</sup>); the second and third stages take place during the “downhill” descent on the PES from the TS to the separate  $O_2$  molecules.

To understand bond-forming and bond-breaking processes along the reaction path going from the  $O_4$  molecule at its equilibrium geometry to the isolated oxygen molecules, we have performed CASSCF calculations at several (equally spaced) geometries following the positive reaction coordinate as given in Figure 4a of ref 18 up to 6 units.<sup>18</sup> In Figure 6 the EPLF isosurfaces are plotted for every geometry used along the reaction path. To better visualize what chemical situation is associated each snapshot we present in Figure 7 the total energy profile as a function of the reaction coordinate and we indicate on the curve the labels (from 1 to 8) corresponding to the various snapshots of Figure 6 for a value of the EPLF kept equal to 0.085 (intermediate between 0.08 and 0.09). The series of snapshots reveals beautifully the nontrivial synchronous intermolecular bond-breaking and intramolecular bond-making processes that lead to bimolecular dissociation. Although difficult to perceive with a single planar perspective, we note that the topology of the EPLF at the equilibrium geometry (snapshot 1) is perfectly symmetrical as seen from any of the four equivalent oxygen–oxygen midpoints. To facilitate the analysis, in the various pictures of Figure 6 we have highlighted in red the core domains, in cyan the lone pair domains, in yellow the two O–O bond domains which lead to the O–O bonds of the separated ground-state oxygen molecules. The O–O bonds which are broken along the reaction path are highlighted in green and magenta. Note that starting from the four equivalent  $\sigma$  bonds



**Figure 6.** Isosurface 0.085 of the EPLF along the reaction coordinate at the CASSCF level. Lone pairs are colored in cyan; core shells are red. O–O bonds are colored yellow, green, and magenta.



**Figure 7.** Total energy profile corresponding to the reaction path shown in Figure 6. The labels on the curve (from 1 to 8) correspond to the various snapshots presented in Figure 6.

of  $O_4$  at the equilibrium geometry (first snapshot of Figure 6), two of them (the yellow-colored regions) remain almost unchanged, while the other two  $\sigma$  bonds are sequentially broken (the green and magenta domains).

We summarize three sequential important facts: (a) the narrow bridges connecting the (different fragments) lone pair domains are the first to be broken (snapshots 1–3). The decrease of the values of the EPLF between the bond domains and the lone pair domains illustrate the decrease of delocalization of the electron pairs between the lone pair domains and the O–O bond domains. From the purely energetic point of view this EPLF interpretation explains the cost of having the same number of electrons confined in a spatial region almost like the one of the equilibrium geometry, but with one O–O bond missing, a situation which needs slightly over 10 kcal/mol. (b) Then the domains that connect the two dissociating  $O_2$  moieties are broken one after the other: The green bond is broken in snapshot 4, and the magenta bond is broken in snapshot 5. (c) Finally, the couple of separate lone pair domains around each oxygen

atom get thinner and fuse to form a single lone pair domain (two at both ends of each O<sub>2</sub> molecule) of the dissociated fragments.

It should be emphasized that in this series of snapshots EPLF has allowed us to visualize the precise sequence of topological transformations of the electron pairings which is at the root of the overall decrease of more than 90 kcal/mol from the equilibrium singlet configuration to the separate couple of O<sub>2</sub> triplet fragments.<sup>18</sup> Clearly, such a scenario could not have been extracted just by looking at the CSF composition of the truncated CASSCF wave function of the transition state as given in table Table 2 and its evolution along the reaction coordinate. Finally, we stress that the introduction of the explicit interelectronic Jastrow prefactor has been found here of little importance to address the evolution of electron pairing along the reaction coordinate. This is actually not surprising since the very strong multiconfigurational nature of the wave function is clearly the dominant feature. To introduce or not the Jastrow term can a priori be expected to be less important than the use of a zeroth-order MCSCF description. At this point it is interesting to note that the VMC total energy obtained for a SCF wave function with a Jastrow factor is slightly lower than the truncated CASSCF(16,12) one. The short-range Jastrow factor allows the electrons to better avoid each other locally, thus significantly diminishing the single-reference energy; however, the truncated CASSCF(16,12) description, which is diabatically correlated with the bimolecular dissociation fragments (ground-state molecular oxygen) is more physical, as shown by recent extensive MR-FNDDMC calculations.<sup>21</sup>

## Summary

In this paper we have used the EPLF to study the electronic pairing distribution along the reaction path of the O<sub>4</sub>  $\rightleftharpoons$  2O<sub>2</sub> reaction. At the equilibrium geometry of O<sub>4</sub> the EPLF reveals four equivalent covalent bonds and a couple of lone pairs surrounding each oxygen atom. This is true for all levels of theory. The analysis of the EPLF along the reaction path allows to obtain a detailed understanding of the evolution of the electronic structure during the dissociation process: the two dissociating O–O bonds are not broken simultaneously but sequentially, and then the lone pairs are rearranged.

From a more general perspective, we would like to emphasize the usefulness of EPLF which is both simple and easily computable using QMC. A major advantage of QMC is the possibility of evaluating the EPLF at various levels of accuracy (HF, MCSCF, VB, DFT, VMC with explicitly correlated trial wave functions, FN-DMC, etc.). Such a possibility is particularly interesting to get new insights into the nature of the pairing and localization of electrons at these various levels of description and, particularly, to understand more deeply the role of the dynamical (Coulomb hole) and the nondynamical (near-degeneracy) correlation effects.

**Acknowledgment.** M.C. and A.S. would like to thank IDRIS (CNRS, Orsay), CCRT (CEA/DAM, Ile-de-France), and CALMIP (Université de Toulouse) for computational support. A.R.S. wishes to thank the FOMES2000 “Cómputo Científico” Project

for CPU time on the IBM-p690 supercomputer and support from CONACYT Project No. 45986-E. A.R.S. also is thankful for unlimited CPU time on the CNSI facility at UC Santa Barbara.

## References and Notes

- (1) Adamantides, V.; Neisius, D.; Verhaegen, G. *Chem. Phys.* **1980**, 48, 215.
- (2) Seidl, E. T.; Schaefer, H. F. *J. Chem. Phys.* **1992**, 96, 1176.
- (3) Peterka, D. S.; Ahmed, M.; Suits, A. G.; Wilson, K. J.; Korkin, A.; Nooijen, M.; Bartlett, R. J. *J. Chem. Phys.* **1999**, 110, 6095.
- (4) Peterka, D. S.; Ahmed, M.; Suits, A. G.; Wilson, K. J.; Korkin, A.; Nooijen, M.; Bartlett, R. J. *J. Chem. Phys.* **1999**, 111, 5279.
- (5) Helm, H.; Walter, C. W. *J. Chem. Phys.* **1993**, 98, 544.
- (6) Long, C. A.; Ewing, G. E. *J. Chem. Phys.* **1973**, 58, 4824.
- (7) Goodman, J.; Brus, L. E. *J. Chem. Phys.* **1977**, 67, 4398.
- (8) Wormer, P. E. S.; van der Avoird, A. *J. Chem. Phys.* **1984**, 81, 1929.
- (9) Cacace, F.; de Petris, G.; Troiani, A. *Angew. Chem., Int. Ed.* **2001**, 40, 4062.
- (10) Lewis, G. N. *J. Am. Chem. Soc.* **1924**, 46, 2027.
- (11) Bussery, B.; Wormer, P. E. S. *J. Chem. Phys.* **1993**, 99, 1230.
- (12) Aquilanti, V.; Ascenzi, D.; Bartolomei, M.; Cappelletti, D.; Cavalli, S.; de Castro Vitores, M.; Pirani, F. *Phys. Rev. Lett.* **1999**, 82, 69.
- (13) Aquilanti, V.; Ascenzi, D.; Bartolomei, M.; Cappelletti, D.; Cavalli, S.; de Castro Vitores, M.; Pirani, F. *J. Am. Chem. Soc.* **1999**, 121, 10794.
- (14) Hernández-Lamoneda, R.; Bartolomei, M.; Hernández, M. I.; Campos-Martínez, J.; Dayou, F. *J. Phys. Chem. A* **2005**, 109, 11587.
- (15) Gorelli, F. A.; Ulivi, L.; Santoro, M.; Bini, R. *Phys. Rev. Lett.* **1999**, 83, 4093.
- (16) Neaton, J. B.; Ashcroft, N. W. *Phys. Rev. Lett.* **2002**, 88, 205503.
- (17) Fujihisa, H.; Akahama, Y.; Kawamura, H.; Ohishi, Y.; Shimomura, O.; Yamawaki, H.; Sakashita, M.; Gotoh, Y.; Takeya, S.; Honda, K. *Phys. Rev. Lett.* **2005**, 97, 085503.
- (18) Hernández-Lamoneda, R.; Ramírez-Solís, A. *J. Chem. Phys.* **2000**, 113, 4139.
- (19) Schröder, D. *Angew. Chem., Int. Ed. Engl.* **2002**, 41, 573.
- (20) Hernández-Lamoneda, R.; Ramírez-Solís, A. *J. Chem. Phys.* **2004**, 120, 10084.
- (21) Caffarel, M.; Scemama, A.; Hernández-Lamoneda, R.; Ramírez-Solís, A. *Phys. Rev. Lett.* **2007**, 99, 153001.
- (22) Amador-Bedolla, C.; Salomón-Ferrer, R.; Lester, W. A., Jr.; Vásquez-Martínez, J. A.; Aspuru-Guzik, A. *J. Chem. Phys.* **2007**, 126, 204308.
- (23) Coppens, P.; Hall, M. B. *Electron Distributions and Chemical Bonds*; Plenum: New York, 1982.
- (24) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Clarendon: Oxford, 1990.
- (25) Gadre, S. R.; Kulkarni, S. A.; Shrivastava, I. H. *J. Chem. Phys.* **1992**, 96, 5253.
- (26) Becke, A. D.; Edgecombe, K. E. *J. Chem. Phys.* **1990**, 92, 5397.
- (27) Silvi, B.; Savin, A. *Nature* **1994**, 371, 683.
- (28) Scemama, A.; Chaquin, P.; Caffarel, M. *J. Chem. Phys.* **2004**, 121, 1725.
- (29) Hammond, B. L.; Lester, W. A., Jr.; Reynolds, P. J. *Monte Carlo Methods in Ab Initio Quantum Chemistry*; World Scientific Lecture and Course Notes in Chemistry; Vol. 1, 1994.
- (30) Schmidt, K. E.; Moskowitz, J. W. *J. Chem. Phys.* **1990**, 93, 4172.
- (31) Assaraf, R.; Caffarel, M. *J. Chem. Phys.* **2000**, 113, 4028.
- (32) Umrigar, C. J.; Wilson, K. G.; Wilkins, J. W. *Phys. Rev. Lett.* **1988**, 60, 1709.
- (33) Scemama, A.; Lelièvre, T.; Stoltz, G.; Cancès, E.; Caffarel, M. *J. Chem. Phys.* **2006**, 125, 114105.
- (34) Flad, H. J.; Caffarel, M.; Savin, A. *Recent Advances in Quantum Monte Carlo Methods*; World scientific publishing edition, 1997.
- (35) Dunning, T. H. *J. Chem. Phys.* **1989**, 90, 1007.
- (36) Filippi, C.; Umrigar, C. J. *J. Chem. Phys.* **1996**, 105, 213.
- (37) Clementi, E.; Roetti, C. *Atomic Data and Nuclear Data Tables* **1974**, 14, 177.
- (38) Manten, S.; Lüchow, A. *J. Chem. Phys.* **2001**, 115, 5362.

JP902028G