Hole Trapping at Al impurities in Silica: A Challenge for Density Functional Theories

J. Lægsgaard[1] and K. Stokbro[2]

[1]*Research Center COM, Technical University of Denmark, Building 349, DK-2800 Kgs. Lyngby, Denmark*
[2]*Mikroelektronik Centret, Technical University of Denmark, Building 345 East, DK-2800 Kgs. Lyngby, Denmark*

(Received 1 November 2000)

The atomic geometry and electronic structure around a neutral substitutional Al impurity in silica is investigated using either the unrestricted Hartree-Fock (UHF) approximation, or Beckes three-parameter hybrid functional (B3LYP). It is found that the B3LYP functional fails to describe the structural distortions around the Al impurity, while the UHF results are consistent with experimental information. We argue that the failure of the B3LYP functional is caused by the incomplete self-interaction cancellation usually present in density functional theories.

DOI: 10.1103/PhysRevLett.86.2834 PACS numbers: 61.72.Hh, 71.10.–w, 71.20.Ps, 76.30.Lh

Density functional theory (DFT) has become a standard tool in physics and chemistry for parameter-free total-energy calculations for molecules and solids [1–3]. The DFT approximation offers a considerable improvement in accuracy (compared to experimental data) over the Hartree-Fock (HF) approximation, while at the same time being much more computationally efficient than extensions of the HF method (such as configuration interaction or Møller-Plesset perturbation theory) which are cumbersome for molecules and not feasible at all in extended systems. However, DFT is based on approximations whose range of validity is difficult to estimate, and it is therefore important to clarify the limitations of the theory and gain inspiration for improved density functionals by a careful study of the rare instances where the method fails. Most known examples of DFT failure occur in systems containing transition-metal or rare-earth elements, due to the complicated electron correlation phenomena present in the d or f valence shells of these elements. Even in these cases, DFT often fails at the quantitative level only (such as the magnitude of magnetic moments in transition-metal oxides) [4,5] or for properties which are not really supposed to be described by the theory (such as excitation spectra) [6–8]. It has recently been found, however, that DFT predicts a wrong adsorption site for CO on a noble-metal surface [9], suggesting that the method is more problematic in impurity problems. In the present paper, we report a remarkable new example of DFT failure, namely, the case of a neutral substitutional Al impurity in silica. We shall demonstrate that DFT fails qualitatively in predicting a structural distortion around the Al impurity, which is correctly described in HF theory, and we shall argue why it does so.

The neutral substitutional Al impurity in α quartz (hereafter denoted as $[AlO_4]^0$) is one of the most well-studied states of Al in silica [10–23]. The state is observed in irradiated samples of natural or manufactured quartz and gives rise to a dark coloration ("smoky quartz"). As Al has one valence electron less than Si, the $[AlO_4]^0$ center will introduce a hole in the quartz bands, which for electrostatic reasons will be confined to the region around the Al impurity. Hence the center is often referred to as a "trapped hole species." Because of the presence of an unpaired electron the center can be detected with electron paramagnetic resonance spectroscopy, and this technique, together with dielectric relaxation and acoustic loss experiments, has yielded much information about the electronic structure of the impurity [10–13,15–17]. The picture emerging from these investigations is that of a hole localized on one of the four O neighbors of the Al impurity, with an accompanying distortion of the lattice structure. This view was supported by early Hartree-Fock calculations on small cluster models of the impurity, which indeed exhibited the hole localization and distortion of the AlO_4 tetrahedron [18,19]. However, recent DFT calculations in a repeated supercell geometry have provided results at variance with this picture: These calculations preserve the α-quartz symmetry, apart from an increase in the Al-O bond length compared to the Si-O bonds, and distribute the hole evenly over the four O neighbors of Al [21–23]. To clarify the reasons behind this surprising discrepancy between different theoretical treatments, we have performed calculations on a rather large cluster model using two different energy functionals: Unrestricted Hartree-Fock (UHF) and Beckes three-parameter hybrid functional (B3LYP) [24] which is a linear combination of UHF and DFT energies that in many situations has been found to improve the predictions of "pure" DFT functionals [25]. The calculations were performed using the GAUSSIAN98 code [26] on a cluster constructed from the ideal α-quartz structure by placing the Al impurity at the center and including two near-neighbor shells of both O and Si. The resulting cluster has C_2 symmetry, which we distort before the structure is allowed to relax. The dangling bonds of the outermost Si atoms were saturated by hydrogen atoms placed at the positions of the next O shell to mimic the crystal lattice. The hydrogens and the outermost Si atoms were kept fixed during the structural optimizations. Different choices of boundary conditions did not appear to change the results significantly. The geometry was optimized using the 3-21G* basis set, and

 0031-9007/01/86(13)/2834(4)$15.00

hyperfine parameters were subsequently calculated using a 6-31G* basis.

It can be seen from Table I that the B3LYP approximation yields geometrical results quite similar to those found in repeated-cell calculations [22,23]: The O neighbors of Al are pairwise equivalent, thus preserving C_2 symmetry, and have only minute differences in bond lengths. This is similar to the situation in the SiO_4 tetrahedrons of pure α quartz, where the (experimental) bond lengths are 1.609 and 1.614 Å, respectively. The values for the Al-O bond lengths are close to the repeated-cell calculation using the PW91 energy functional [23]. The results of the UHF treatment, on the other hand, are markedly different. In this approximation one O neighbor is displaced about 0.2 Å outward from Al, compared to the PW91 and B3LYP geometries, while the other three are pulled slightly towards the Al impurity. This difference in geometry reflects a difference in the electronic structure predicted by the various approximations. As can be seen from the Mulliken spin populations in Table II, the B3LYP approximation distributes the spin density almost evenly among the four O neighbors of Al, while in the UHF approach, the spin is essentially localized on the O atom which has been pushed away from Al. Surface plots of the hole wave functions in the two approximations are shown in Fig. 1. A convenient experimental measure of the degree of hole localization is the anisotropic ^{17}O hyperfine coupling tensor, which is roughly proportional to the p component of the hole wave function on a particular O atom. In Table III the theoretical predictions for this quantity in the various approximations are compared to the experimental values [15]. The predictions of the UHF approximation are quite accurate, while those arising from the B3LYP calculation are off by more than a factor of 4.

As the use of DFT approximations such as B3LYP, PW91, or even the local density approximation usually constitute a substantial improvement over the UHF method, it is important to determine why they fail in the present context. It was noted by Magagnini and coworkers that the barrier preventing lattice distortion in their repeated-cell calculation using the local spin-density approximation appeared to arise from the electrostatic energy terms (Hartree and ion-ion repulsions) [22]. Intuitively this appears reasonable: The concentration of charge density resulting from hole localization would be associated with a penalty in Hartree energy. However, this represents an unphysical property of the Hartree energy functional: A single particle does not interact with itself. In the UHF approach the self-interaction is exactly canceled by a corresponding self-exchange term. This is not necessarily the case in the various DFT approaches.

To formalize the argument, we set up a minimal model for the hole energy functional in the different approximations. The crucial point to realize is that the top of the valence bands in the present problem consists of nonbonding O $2p$ orbitals which are widely separated and therefore have very small overlaps. Furthermore, they are nearly degenerate in the absence of structural distortions. Focusing attention on the electrons moving in the nonbonding O $2p$ states, the total energy may be approximated by

$$E = C + \sum_i \varepsilon_i \rho_{ii} + \frac{1}{2} \sum_{ij} U_{ij} \rho_{ii} \rho_{jj} + E_{xc}[\rho], \quad (1)$$

where ρ is the density matrix, and i, j indexes the nonbonding O $2p$ states on different sites. The ε_i collects all terms linear in ρ except those arising from the xc energy, and C denotes the ρ-independent terms. The U_{ij} parameters are the Coulomb integrals between $2p$ charge distributions on sites i and j. The different total-energy approximations are distinguished by different choices of E_{xc}. Since one hole is present in the nonbonding states, the density matrix may be written as $\rho_{ij} = \delta_{ij} - \tilde{\rho}_{ij}$, with $\tilde{\rho}$ being the hole density matrix. The total energy may then be written

$$E = \tilde{C} + \sum_i \tilde{\varepsilon}_i \tilde{\rho}_{ii} + \frac{1}{2} \sum_{ij} U_{ij} \tilde{\rho}_{ii} \tilde{\rho}_{jj} + E_{xc}[\rho], \quad (2)$$

with $\tilde{C}, \tilde{\varepsilon}_i$ defined analogously to C, ε_i. In the HF case, E_{xc} will simply be the exact exchange term, which cancels the Hartree self-interaction, and only the first two terms remain. Thus, in the UHF case, any structural distortion pushing one O level above the others will lead to complete hole localization on that level. In principle, this cancellation should also occur in DFT but, as is well known, this is not the case for the approximate functionals commonly employed. A further complication is that the equivalence between a hole in a filled band and an electron in an empty band, which is inherent in the UHF equations, is lost in

TABLE I. Bond lengths around the substitutional Al impurity as calculated with either the UHF or B3LYP functionals in a cluster model. The results labeled PW91 are from the repeated-cell calculation of Ref. [23].

Bond	UHF	B3LYP	PW91
Al-O(1)	1.924 Å	1.741 Å	1.720 Å
Al-O(2)	1.688 Å	1.741 Å	1.720 Å
Al-O(3)	1.703 Å	1.748 Å	1.722 Å
Al-O(4)	1.689 Å	1.748 Å	1.722 Å

TABLE II. Mulliken spin populations for Al and its nearest neighbors in the cluster. UHF and B3LYP values have been calculated in both the UHF and B3LYP equilibrium geometries.

	UHF geometry		B3LYP geometry	
Atom	UHF	B3LYP	UHF	B3LYP
Al	−0.03	−0.04	−0.14	−0.05
O(1)	1.04	0.79	0.58	0.20
O(2)	0.00	0.02	0.56	0.19
O(3)	0.00	0.04	0.02	0.19
O(4)	0.01	0.09	0.02	0.19

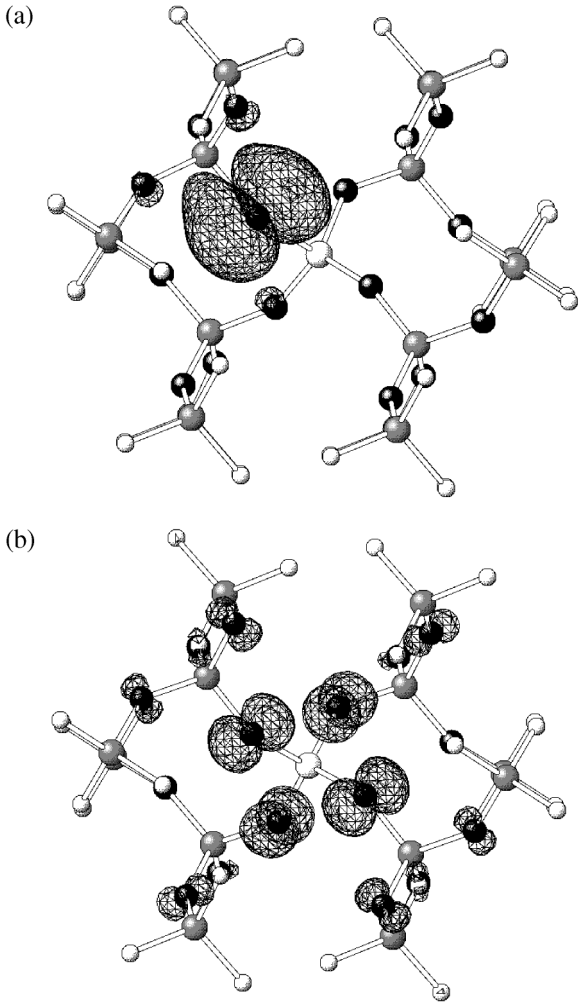

FIG. 1. Isodensity surfaces for the lowest unoccupied one-particle eigenstate calculated using either the UHF (a) or the B3LYP approximation (b), in the corresponding equilibrium geometries. Black atoms are O; grey, Si; large white, Al; and small white, H.

DFT. A simple discussion of this point can be given in the limit where the particle moves in a large number of identical orbitals, and intersite Coulomb interactions can be neglected. Putting the $\varepsilon_i = 0$ and assuming a (semi)-local form for E_{xc} the energy functional separates into a sum of local terms:

$$E = \sum_i \left(\rho_{ii}^2 \frac{U_{ii}}{2} + E_{xc}[n_i] \right), \qquad n_i = \rho_{ii} n_0 , \tag{3}$$

with n_0 being the density of a full local orbital. Comparing states in which the particle (electron or hole) is either localized on one site or evenly distributed over all N sites in the system, we find in the case of an electron in an empty band:

$$E_{\text{loc}} - E_{\text{deloc}} = \frac{U_{ii}}{2} + E_{xc}[n_0] \equiv \delta , \tag{4}$$

while in the case of a hole in a filled band

TABLE III. Principal values of the anisotropic hyperfine coupling tensor (in Gauss) for ^{27}Al and the ^{17}O atom with the highest spin population. UHF and B3LYP values have been calculated in both the UHF and B3LYP equilibrium geometries. Experimentally, only one ^{17}O signal has been resolved [15].

	UHF geometry		B3LYP geometry		
Atom	UHF	B3LYP	UHF	B3LYP	Expt.
	−0.342	−0.380	−1.212	−0.247	−0.349
Al	−0.316	−0.243	−0.018	−0.033	−0.343
	0.659	0.624	1.231	0.280	0.747
	−89.350	−72.211	−54.343	−18.339	−85.04
O(1)	44.605	36.046	26.778	9.101	41.21
	44.745	36.165	27.565	9.238	43.82

$$E_{\text{loc}} - E_{\text{deloc}} = \frac{U_{ii}}{2} + N \int d\mathbf{r} \, \frac{\delta E_{xc}}{\delta n}[n_0] \frac{n_0(\mathbf{r})}{N} - E_{xc}[n_0] . \tag{5}$$

Introducing the Hartree and xc potentials, V_H, V_{xc} as functional derivatives of the corresponding energies at full occupancy this can be written as

$$E_{\text{loc}} - E_{\text{deloc}} = \int d\mathbf{r} \, n_0(\mathbf{r}) [V_H(\mathbf{r}) + V_{xc}(\mathbf{r})] - \delta . \tag{6}$$

δ is usually on the order of 0.1–1 eV per occupied orbital [27,28], whereas the term involving the potentials in many cases can be an order of magnitude larger, as evidenced by the large shifts in Kohn-Sham eigenvalues occurring when self-interaction corrections are enforced in DFT [5,27]. Therefore, it must be expected that the spurious localization barrier in DFT is larger for a hole than an electron, even when the particle moves only in a finite set of orbitals. As the B3LYP is simply a linear combination of UHF and DFT energies the argument should apply also in this case, although the residual self-interactions will be somewhat smaller than in pure DFT.

The O atoms closest to Al are separated from each other by ~2.8 Å in the relaxed B3LYP geometry. An estimate of the hopping integrals between nonbonding O $2p$ states, which were neglected in our model discussion, is provided by the energy splitting between bonding and antibonding π states in an O_2 molecule with an internuclear distance of 2.8 Å, treated within the PW91 approximation in which the effective potential is purely local. We find this splitting to be 0.12 eV, i.e., yielding an estimate for the hopping integral of ~0.06 eV, considerably lower than the expected self-interaction error.

In order to test the validity of the above arguments the electronic structure of the relaxed UHF geometry was recalculated using the B3LYP energy functional and vice versa. The resulting Mulliken populations are also shown in Table II. Note that even the slight deviation from ideal tetrahedral geometry in the AlO_4 unit determined by

B3LYP is sufficient to cause almost complete localization of the spin density on the long-bonded O atoms within the UHF approximation. On the other hand, the B3LYP energy functional can only partly capture the hole localization even when applied to the distorted atomic configuration calculated by UHF. These results indicate that the assumptions underlying our formal arguments are well justified in the present case.

It should be stressed that, although both the PW91 and B3LYP approximations fail dramatically in the present case, it does not mean that these methods are generally useless in describing a single particle in a filled/empty band. Indeed, we have obtained quite accurate values for the hyperfine parameters of the unpaired electron in the corresponding $[PO_4]^0$ center using the PW91 energy approximation in a repeated-cell calculation [29]. As argued above, the $[AlO_4]^0$ center is a delicate problem because of the fact that the unpaired spin moves in a space of weakly coupled nearly degenerate levels. Thus, the energy scale of the effective hole Hamiltonian is small, and the residual self-interaction present in DFT can adversely affect the results.

In conclusion, we have shown that the B3LYP approximation to DFT is qualitatively inconsistent with experiments when applied to cluster models of the $[AlO_4]^0$ impurity state in α quartz, whereas the UHF approximation captures the essential physics of the system. Furthermore, we have argued that the failure of B3LYP is connected with the presence of unphysical self-interactions in these energy functionals, and that such problems are likely to exist whenever a single particle (especially a hole) is moving in a space of levels which are weakly coupled and near degenerate compared to the energy scale of the residual self-interactions in DFT. It has been noted in other problems involving localized defect states in ionic solids that HF theory leads to much larger structural relaxations than DFT, but to our knowledge the role of residual self-interactions has not been discussed in this connection [30–32]. It seems appropriate to reconsider these cases in light of the above analysis.

After submission of this manuscript it has come to our attention that a similar work has recently been performed by Pacchioni *et al.* [33].

The authors would like to acknowledge J. K. Nørskov and A. Svane for useful discussions and a critical reading of the manuscript.

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. Sham, Phys. Rev. **140**, A1133 (1965).
[3] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).
[4] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).
[5] A. Svane and O. Gunnarsson, Phys. Rev. Lett. **65**, 1148 (1990).
[6] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983); L. J. Sham and M. Schlüter, *ibid.* **51**, 1888 (1983).
[7] K. Terakura, T. Oguchi, A. R. Williams, and J. Kübler, Phys. Rev. B **30**, 4734 (1984).
[8] D. M. Bylander and L. Kleinman, Phys. Rev. B **49**, 1608 (1993).
[9] P. J. Feibelman *et al.* (to be published).
[10] J. H. E. Griffiths, J. Owen, and I. M. Ward, in *Proceeding of the Conference on Defects in Crystalline Solids, Bristol* (Physical Society, London, 1954).
[11] M. C. M. O'Brien, Proc. R. Soc. London A **231**, 404 (1955).
[12] A. L. Taylor and G. W. Farnell, Can. J. Phys. **42**, 595 (1964).
[13] R. Schnadt and J. Schneider, Phys. Kondens. Mater. **11**, 19 (1970).
[14] O. F. Schirmer, Solid State Commun. **18**, 1349 (1976).
[15] R. H. D. Nuttall and J. A. Weil, Solid State Commun. **35**, 789 (1980).
[16] L. E. Halliburton, N. Koumvakalis, M. E. Markes, and J. J. Martin, J. Appl. Phys. **52**, 3565 (1981).
[17] R. H. D. Nuttall and J. A. Weil, Can. J. Phys. **59**, 1696 (1981).
[18] M. J. Mombourquette, J. A. Weil, and P. G. Mezey, Can. J. Phys. **62**, 21 (1984).
[19] F. Sim, C. R. A. Catlow, M. Dupuis, and J. D. Watts, J. Chem. Phys. **95**, 4215 (1991).
[20] M. Martini, A. Paleari, G. Spinolo, and A. Vedda, Phys. Rev. B **52**, 138 (1995).
[21] A. Continenza and A. D. Pomponio, Phys. Rev. B **54**, 13 687 (1996).
[22] M. Magagnini, P. Giannozzi, and A. D. Corso, Phys. Rev. B **61**, 2621 (2000).
[23] J. Lægsgaard and K. Stokbro, Phys. Rev. B **61**, 12 590 (2000).
[24] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[25] J. Song, R. M. van Ginhoven, L. R. Corrales, and H. Jonsson, Faraday Discuss. **117**, 303 (2001).
[26] M. J. Frisch *et al.,* GAUSSIAN 98 (Revision A.7), Gaussian, Inc., Pittsburgh, PA, 1998.
[27] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[28] A. Svane, Phys. Rev. B **53**, 4275 (1996).
[29] J. Lægsgaard and Kurt Stokbro (to be published).
[30] E. A. Kotomin, R. I. Eglitis, A. V. Postnikov, G. Borstel, and N. E. Christensen, Phys. Rev. B **60**, 1 (1999).
[31] M. M. Kuklja, E. V. Stefanovich, E. A. Kotomin, A. I. Popov, R. Gonzalez, and Y. Chen, Phys. Rev. B **59**, 1885 (1999).
[32] Q. S. Wang and N. A. W. Holzwarth, Phys. Rev. B **41**, 3211 (1990).
[33] G. Pacchioni, F. Frigoli, D. Ricci, and J. A. Weil, Phys. Rev. B **63**, 054102 (2001).