# Structure and Stability of Molecular Carbon: Importance of Electron Correlation

Jeffrey C. Grossman,[1] Lubos Mitas,[2,3] and Krishnan Raghavachari[4]
[1]*Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801*
[2]*National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801*
[3]*Materials Research Laboratory, University of Illinois at Urbanan-Champaign, Urbana, Illinois 61801*
[4]*AT& T Bell Laboratories, Murray Hill, New Jersey 07974*

(Received 23 May 1995)

The electronic structure of medium-size molecular carbon is determined by a variety of methods, including density functional, quantum chemistry, and quantum Monte Carlo approaches. We demonstrate (i) significant differences between the mean-field methods in determining the minimum energy structure of a set of $C_{10}$ and $C_{20}$ isomers, and (ii) the crucial importance of an accurate account of electron correlation which enables us to predict, e.g., that of the available geometries for $C_{20}$ the relaxed graphite fragment (bowl) is lower in energy than either the ring or fullerene isomers.

PACS numbers: 61.46.+w, 36.40.-c, 71.45.Nt

The remarkable stability of the $C_{60}$ fullerene has motivated intense research on carbon clusters over the past decade. The mechanism of formation of fullerenes is not yet well understood, although many experimental and theoretical investigations are attempting to answer such questions. Even for $C_{20}$, which is the smallest carbon cluster that can exist as a fullerene cage, it is not known whether this structure is stable, and different studies have shown conflicting predictions [1–7]. Very recent experiments [3] have found the cyclic and bicyclic ring structures to be most prevalent, although it appears that this is predominantly an entropic effect. In addition, it has been found that another structural form, which can be considered a curved piece of graphite containing a pentagon (bowl), was found energetically competitive with the ring and cage.

Theoretical studies of these structures have also predicted different energetical orderings, depending strongly on the theory used: Hartree-Fock (HF) predicts the ring to be lowest in energy, then the bowl and then the cage, with a ring-cage difference of $\sim 4.5$ eV. The local density approximation (LDA), on the other hand, predicts the reverse ordering with a cage-ring difference of $\sim 4.0$ eV. Generalized gradient approximation (GGA) corrections to LDA generally reverse back to the HF ordering, although this depends on which functional is used [1]. Tight binding as well as recent coupled cluster (CC) calculations suggest that the LDA ordering is correct, whereas semiempirical methods tend to reproduce the HF results. The striking disparity between methods in obtaining the relative energies of these clusters demonstrates the clear need for a higher accuracy approach to this problem.

The goal of this Letter is to provide a new level of understanding via a thorough and systematic study of molecular carbon in medium-size structures with various types of bonding. To this end, we have carried out calculations of $C_{10}$ and $C_{20}$ isomers with a variety of methods, including LDA and related GGA techniques, and quantum chemistry approaches, such as CC for sizes which were not prohibitively large. For a more accurate account of correlation we used a recently developed quantum Monte Carlo (QMC) approach which is capable of describing many-body effects for a large number of valence electrons. To support and calibrate the QMC performance, we have also calculated seven hydrocarbons, ranging from methane to benzene, for which we have reproduced the experimental binding energies within 1%. The QMC results have revealed an amazingly fine interplay of Coulomb, exchange, and correlation contributions for various structures and convincingly demonstrate the necessity of an accurate treatment of electron correlation for reliable predictions. A comparison of a dozen different approaches provides an interesting view on their performance and reveals an unexpectedly low predictive power of the mean-field methods for these systems.

In our QMC approach [8], we treat the core electrons with nonlocal pseudopotentials [9], and use variational Monte Carlo (VMC) to find an optimized correlated many-body trial function $\psi_T(R)$. Then diffusion Monte Carlo (DMC) is used to remove most of the remaining variational bias. DMC is based on the property that the operator $\exp(-\tau H)$ projects out the ground state of any trial function with the same symmetry and nonzero overlap [8].

The fermion antisymmetry problem is treated by the commonly used fixed-node approximation [10]. The trial function used is a product of Slater determinants of single electron spin-up and spin-down orbitals times a correlation factor [11,12],

$$\psi_T = D^{\uparrow} D^{\downarrow} \exp\left[\sum_{I,i<j} u(r_{iI}, r_{jI}, r_{ij})\right], \tag{1}$$

where $I$ corresponds to the ions, $i, j$ to the electrons, and $r_{iI}, r_{jI}, r_{ij}$ to the distances. Parametrization and optimization of $u(r_{iI}, r_{jI}, r_{ij})$, which represents the electron-electron and electron-electron-ion correlations, are described elsewhere [13].

0031-9007/95/75(21)/3870(4)$06.00

For the HF calculations, we employed 6-31G* (contracted $2s,2p,1d$), 6-311G* (contracted $3s,3p,1d$) and for a few cases more accurate basis sets. We found that for energy differences the 6-311G* basis was adequate, and thus we used its pseudopotential equivalent in all our calculations. Unexpectedly, we found the 6-31G* basis to be strongly biasing for $C_{20}$ systems. In particular, the second uncontracted $s$ and $p$ functions are crucial for an accurate description of the ring and bowl triple bonds (Fig. 1), and failure to include these results in a bias which increases the ring HF energy by about 1 eV and the bowl HF energy by 0.5 eV when compared with the energy of the cage. In addition to the basis set, we also carefully checked the accuracy of our pseudopotentials [9] and found that in all cases the HF pseudopotential binding energies matched those from full core calculations to within 0.1%.

To test the accuracy of the QMC method for systems which include carbon and to assess the overall predictivity of this approach, we investigated the following seven hydrocarbon systems: $C_2H_2$, $C_2H_4$, $C_2H_6$, $C_6H_6$, $CH_4$, and $C_3H_4$ (both propyne and allene). The electronic and structural properties of these molecules are well known experimentally and theoretically [14] and a variety of carbon-carbon bonding possibilities are covered (e.g., $C_2H_2$, $C_2H_4$, $C_2H_6$, and $C_6H_6$ correspond to triple, double, single, and 1.5 bonds, respectively). Binding energies for these molecules, calculated from the HF, LDA, and QMC methods, are listed in Table I. We first note, as expected, the LDA overbinding and HF underbinding. By contrast, our QMC binding energies are all within 1% of the experimental values, indicating the high accuracy of the method. In fact, the average difference from experiment is about 0.1 eV, close to a chemical accuracy of 1–2 kcal/mol.

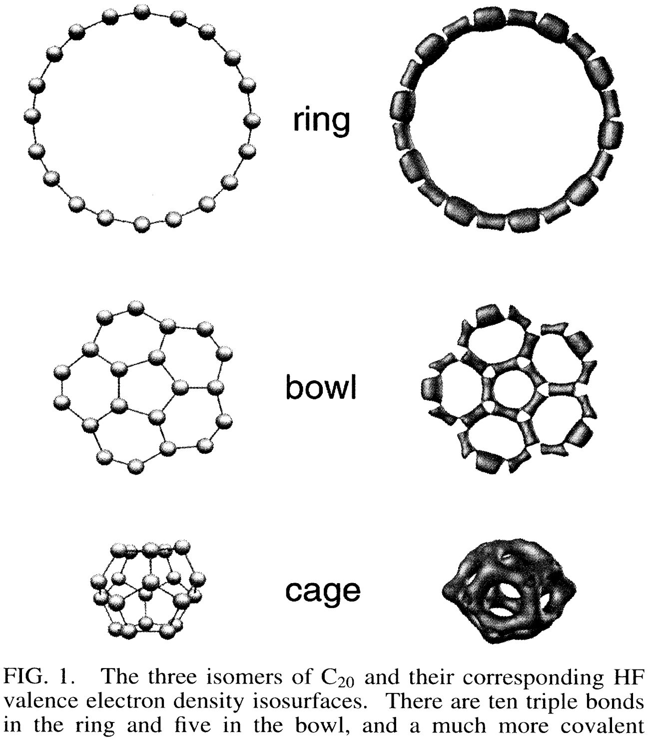

FIG. 1. The three isomers of $C_{20}$ and their corresponding HF valence electron density isosurfaces. There are ten triple bonds in the ring and five in the bowl, and a much more covalent bonding character in the cage.

Next we carried out calculations on four structures of $C_{10}$: a linear chain, cumulenic and acetylenic rings, and a compact structure which consists of distorted facing pentagons and has all threefold coordinated atoms. Geometries were taken from previous calculations [15], except for the compact structure which we optimized within HF. The moderate size of these clusters (40 valence electrons) allowed us to compare a host of methods (cf. Table II), which provides some insight into the physical reasons behind the ability of a given method to describe relative energies. For example, in HF the high spin (triplet) linear molecule is only $\approx$ 1.0 eV above the rings, while in LDA this energy difference is $\approx$ 3.7 eV. The correct answer, as is the case for many systems, lies somewhere in between HF and LDA. Of the quantum chemistry methods studied, CCSD(T) (CC with single, double, and perturbationally triple excitations) performs the best, although its accuracy is strongly dependent on, and severely limited by, the basis set. The small energy difference between the two ring structures is difficult to evaluate, and subtle changes in geometry could cause a shift in energy by $\approx$ 1.0 eV [15,16]. We note that CC recovers about 75% of the DMC correlation energy, while, using our trial functions, VMC recovers about 85%.

For $C_{20}$, we used geometries from previous calculations [1] which were optimized within the HF method. Very high quality trial functions were obtained through several reoptimizations of the correlated trial wave functions, and several independent DMC runs were carried out in order to avoid statistical data correlation and to estimate the error bars more accurately. From our results (see Fig. 2), it is evident that correlation energy has a profound effect on the final energies, and that the ordering of the structures is a result of the subtle interplay between electrostatic and correlation terms. As one might expect, the cage has the largest correlation since it has the highest number of bonds, all of which are essentially $sp^2$ hybrids. The bowl has the second largest correlation and, because of its intermediate electrostatic contribution, is the lowest in total energy of the geometries studied. Although the ring has very favor-

TABLE I. Binding energies (eV) of small hydrocarbons calculated using the HF, LDA, and DMC methods. Experimental values are listed for comparison.

| | HF | LDA | DMC | Exp. |
|---|---|---|---|---|
| Methane ($CH_4$) | 14.20 | 20.59 | 18.28(5) | 18.19 |
| Acytylene ($C_2H_2$) | 12.70 | 20.49 | 17.53(5) | 17.59 |
| Ethylene ($C_2H_4$) | 18.54 | 28.19 | 24.44(5) | 24.41 |
| Ethane ($C_2H_6$) | 23.87 | 35.37 | 31.10(5) | 30.85 |
| Allene ($C_3H_4$) | 22.63 | 35.87 | 30.36(5) | 30.36 |
| Propyne ($C_3H_4$) | 22.70 | 35.70 | 30.55(5) | 30.45 |
| Benzene ($C_6H_6$) | 44.44 | 70.01 | 59.2(1) | 59.24 |

TABLE II. Energies (eV) of $C_{10}$ isomers relative to the cumulenic ring structure ($D_{10h}$) for a variety of methods.

| | HF | LDA | MP2 | MP4 | CCSD(T) | DMC |
|---|---|---|---|---|---|---|
| Ring ($D_{10h}$) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Ring ($D_{5h}$) | −0.05 | 0.25 | 0.48[a] | ... | −0.02[a] | −0.3(1) |
| Chain ($D_{4h}$) | 0.99 | 3.629 | 4.44 | 2.26 | 2.22 | 2.6(1) |
| Compact ($C_1$) | 7.45 | 6.21 | 7.76 | 6.31 | 6.59 | 7.2(1) |

[a] Reference [15].

able Coulomb and exchange contributions, its correlation energy is substantially [3.4(2) eV] less than that of the cage, making it the second most stable structure overall.

Of significant importance is the impact of the geometry (and how it was optimized) on the energy differences. To assess the impact of using a particular set of geometries, we calculated both HF and LDA binding energies for four different sets of geometries (cf. Table III). Two sets were optimized within LDA and two were optimized within HF theory. It may initially seem as though the cage would have the most widely varying energies, since it is the least symmetric structure and originates from perfect $I_h$ geometry, which has a large Jahn-Teller instability. We find, however, that in fact the cage energies vary by less than 0.5 eV, whereas the variation of ring energies is more than twice as much. This can be understood by considering the important role of the triple bond, as LDA favors more homogeneous and less dimerized structures. A comparison of different geometries within correlated calculations requires further investigation. However, within the scope of this Letter we have concentrated on energy differences between various isomers for geometries previously published [1].

One of the most important results obtained is the large difference in predictions for carbon systems between LDA, LDA-related approaches, and other methods. We believe that the problem has to do with the enormous variability of carbon structural forms, where systems with different types of bonding and, in this case, even dimensionality, can form competing structures. The character of the LDA errors is very interesting; LDA highly favors the compact structure with larger density against the more extended ring with lower density. This trend has been shown for another set of first row element systems, namely, the nitrogen atom, molecule, and solids [17]. Very interestingly, for systems with a smoother electronic density such as various medium range structures of silicon clusters, LDA provided excellent predictions of relative energies [18]. This shows, in accord with recent observations [17,19], that there is a density dependent systematic error in the currently used LDA functional (the exchange correlations of Hedin and Lundquist and Vosko, Wilk, and Nusair led to almost identical LDA pictures).

The GGA's [20], which were designed to fix this deficiency, are evidently not very reliable. We have evaluated the relative energies of the three $C_{20}$ isomers with a variety of such gradient-corrected functionals. Using the previously optimized HF geometries in all these calculations, we find, surprisingly, that the different functionals behave very differently and yield completely different energy orderings (cf. Table IV). In particular, the B-LYP functional (Becke-88 gradient corrected exchange functional, Lee-Yang-Parr correlation functional) was previously shown to essentially bring back the HF picture, which is the complete reverse of the LDA energy ordering. The more sophisticated B3-LYP functional, which has been extremely successful in reproducing the bond energies of small molecules, also behaves in a similar manner. However, the inclusion of the correlation functionals of Perdew and Wang (PW91) yields a different picture. The bowl is now the most stable isomer, in qualitative agreement with the QMC results,

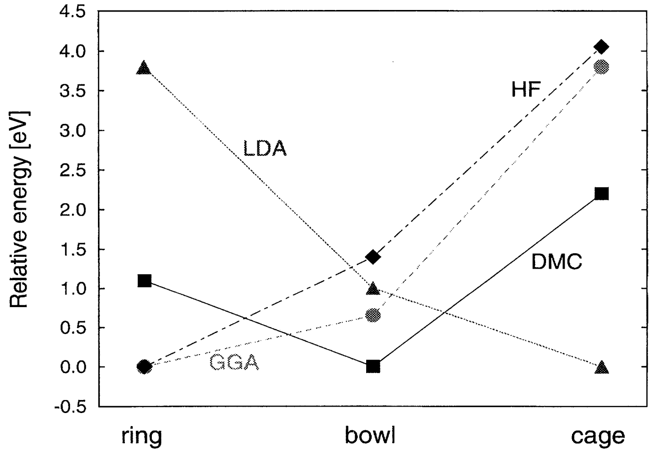

FIG. 2. The relative energy differences for the ring, bowl, and cage $C_{20}$ isomers (DMC error bars are 0.2 eV). For each theory the lowest energy structure is taken as a reference. BLYP refers to calculations done with the Becke-Lee-Yang-Parr functional.

TABLE III. HF and LDA binding energies (eV) of $C_{20}$ isomers for four different sets of geometries optimized within HF and LDA.

| | HF[a] | HF[b] | LDA[b] | LDA[c] |
|---|---|---|---|---|
| Ring (HF) | 84.70 | 84.67 | 83.37 | 83.08 |
| Bowl (HF) | 83.06 | 83.05 | 82.58 | 82.55 |
| Cage (HF) | 80.22 | 79.98 | 80.01 | 79.79 |
| Ring (LDA) | 154.20 | 154.38 | 155.31 | 155.33 |
| Bowl (LDA) | 157.38 | 157.43 | 157.40 | 157.68 |
| Cage (LDA) | 158.49 | 158.89 | 158.90 | 158.99 |

[a] Reference [1].

[b] Reference [5].

[c] G. Galli (private communication).

although B-PW91 and B3-PW91 yield different orderings of the cage and the ring. The overall performance of the gradient corrected functionals is difficult to judge considering the differences between the functionals used.

On the purely practical level of computational demands and feasibility, our calculations have revealed another interesting and quite remarkable feature of QMC. After optimization of the trial function, the VMC runs (error bar $\sim 0.01$ eV/atom) did not take much longer than the HF runs with a 6-311G* basis (using the GAMESS package). In addition, CCSD(T) runs, which often recover less correlation energy than VMC, were extremely demanding even for only 10 carbon atoms. For $C_{20}$, computational barriers simply prevented us from carrying out CCSD(T) calculations with a 6-311G* basis set, a modest basis for correlation purposes. By limiting the basis set to cc-pVDZ, which is qualitatively similar to 6-31G*, Taylor *et al.* [5] were able to perform coupled cluster calculations on the $C_{20}$ structures. Their results predict the bowl to be most stable for our geometries, although their extrapolation, to account for limited basis set effects, places the cage with the lowest energy. Usually, however, a much larger basis set is necessary for the CC method to produce unambiguous predictions [15,16] (see also calculations of $C_4$ isomers by Stanton *et al.* [21]). A comparison of the wave functions shows clear advantages of the built-in many-body character of QMC trial functions against the slow convergence in the basis set expansion of the CC method (20 VMC variational parameters as opposed to $19.5 \times 10^6$ self-consistently and $29 \times 10^9$ perturbationally CC variational parameters [5]).

Finally, we believe that our calculations, which provide a much cleaner picture of the correlation in these systems, also show that even for *sp* systems LDA should be used with caution, especially when molecular structures with very different bonding character are studied. At the same time we do not claim absolute accuracy even in our QMC approach, as it includes two approximations: the fixed-node error and projection of the nonlocal part of pseudopotentials in the DMC method [8]. We estimate from calculations of smaller carbon systems, including the hydrocarbons, that the amount of valence correlation energy which is recovered by the DMC method is about 95%. However, all indications which we have so far show an increase in accuracy by a factor of 5 to 10 when compared with mean-field methods (which are the only practical alternatives for larger systems). We consider the results to be an important step forward in understanding molecular carbon and its bonding capabilities.

TABLE IV. Relative energies (eV) of $C_{20}$ isomers with different functionals.

| | LDA | B-LYP | B3-LYP | B-PW91 | B3-PW91 |
|---|---|---|---|---|---|
| Ring | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Bowl | −3.18 | 1.07 | 0.40 | −0.47 | −0.82 |
| Cage | −4.28 | 3.43 | 2.33 | 0.31 | −0.10 |

Further, we believe that our calculations convincingly demonstrate the importance of high accuracy treatment of the correlation. Because molecular carbon is so abundant and so important in many disciplines such as organic chemistry and biology, additional research in this direction can bring far-reaching consequences. It is evident that these carbon systems provide an excellent "laboratory of electron correlation," especially for testing new methods. To facilitate this for other researchers, the geometries, basis sets, energies, and other relevant data are available on the world wide web at http://www.ncsa.uiuc.edu/Apps/CMP/c20.html.

We thank D. M. Ceperley and R. M. Martin for support and encouragement, and J. Weare, G. Scuseria, and G. Galli for providing us with data and for useful discussions. We are grateful to A. J. Freeman, E. Wimmer, M. F. Jarrold, and O. Sankey for useful discussions, and to V. Balaji for help with the CC calculations. This research was supported by NCSA and by DOE Grant DEFG02-91-ER45439.

[1] K. Raghavachari *et al.*, Chem. Phys. Lett. **214**, 357 (1993).
[2] C. J. Brabec *et al.*, Phys. Rev. B **46**, 7326 (1992).
[3] H. Handschuh *et al.*, Phys. Rev. Lett. **74**, 1095 (1995).
[4] K. B. Shelimov, J. M. Hunter, and M. F. Jarrold, Int. J. Mass Spectrosc. Ion Proc. **138**, 17 (1994).
[5] P. R. Taylor *et al.*, Chem. Phys. Lett. **235**, 558 (1995); J. Weare (private communication).
[6] G. von Helden *et al.*, Chem. Phys. Lett. **204**, 15 (1993).
[7] M. Sawtarie, M. Menon, and K. R. Subbaswamy, Phys. Rev. B **49**, 7739 (1994).
[8] L. Mitas, E. L. Shirley, and D. M. Ceperley, J. Chem. Phys. **95**, 3467 (1991).
[9] W. J. Stevens, H. Basch, and M. Krauss, J. Chem. Phys. **81**, 6026 (1984).
[10] J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975); P. J. Reynolds *et al.*, *ibid.* **77**, 5593 (1982); J. W. Moskowitz *et al.*, *ibid.* **77**, 349 (1982).
[11] K. E. Schmidt and J. W. Moskowitz, J. Chem. Phys. **93**, 4172 (1990).
[12] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).
[13] L. Mitas, in *Computer Simulation Studies in Condensed-Matter Physics V*, edited by D. P. Landau, K. K. Mon, and H. B. Schuttler (Springer, Berlin, 1993).
[14] L. A. Curtiss *et al.*, J. Chem. Phys. **94**, 7221 (1991).
[15] V. Parasuk and J. Almlöf, Theor. Chim. Acta **83**, 227 (1992).
[16] J. D. Watts and R. J. Bartlett, Chem. Phys. Lett. **190**, 19 (1992).
[17] L. Mitas and R. M. Martin, Phys. Rev. Lett. **72**, 2438 (1994).
[18] J. C. Grossman and L. Mitas, Phys. Rev. Lett. **74**, 1323 (1995).
[19] D. M. Bylander and L. Kleinman, Phys. Rev. Lett. **74**, 3660 (1995).
[20] M. J. Frisch *et al.*, *Gaussian 94* (Gaussian Inc., Pittsburgh, PA, 1995).
[21] J. F. Stanton *et al.*, Int. J. Quantum Chem. Symp. **26**, 879 (1992).

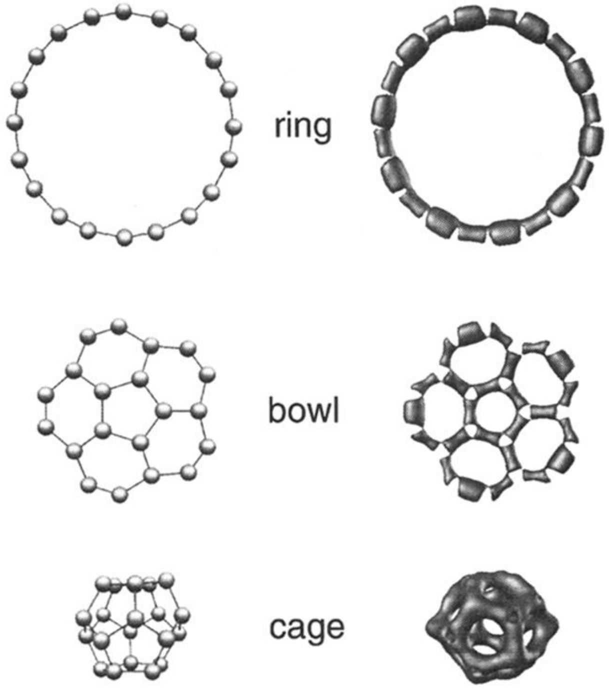

FIG. 1. The three isomers of $C_{20}$ and their corresponding HF valence electron density isosurfaces. There are ten triple bonds in the ring and five in the bowl, and a much more covalent bonding character in the cage.