

Direct calculation of electron density in density-functional theory: Implementation for benzene and a tetrapeptide

Weitao Yang

Department of Chemistry, Duke University, Durham, North Carolina 27706

(Received 10 May 1991)

A recently developed approach for the direct calculation of electron density is implemented for polyatomic molecules: benzene and a tetrapeptide with four glycine residues. The method uses the density as the basic variable, divides a system into subsystems, and determines the density for each subsystem. It is found that the method is capable of describing the electronic structure with accuracy comparable to the Kohn-Sham method. This substantiates the hope for *ab initio* calculations of large systems beyond the reach of conventional methods.

PACS number(s): 31.10.+z, 31.15.+q, 71.10.+x, 71.45.-d

For large and complex systems, first-principles electronic-structure calculations can provide reliable structural information, insight and prediction. However, despite the success for small systems, current approaches, including the Hartree-Fock-based method [1] and the Hohenberg-Kohn-Sham density functional theory, [2-8] encounter insurmountable difficulty associated with large molecules. This comes from the dependence on a set of single-electron orbitals as the basic building blocks for the wave function in the Hartree-Fock theory and for the total electron density in the Kohn-Sham (KS) theory.

In an effort to develop a theory capable of handling large systems, we have recently constructed a "divide-and-conquer" method for direct calculation of the electron density [9] based on the Hohenberg-Kohn density functional theory [2-8]. The method uses the total electron density as the basic variable, in the spirit of the Thomas-Fermi theory [2-4]. It employs the following strategy: divide a large system into small subsystems and then calculate the total energy and the density as the sums of the subsystem contributions. Thus, the building blocks are the subsystem electron densities, instead of the global Kohn-Sham single-electron orbitals; each subsystem density is determined with computational effort similar to the Kohn-Sham calculation for the subsystem alone. The final synthesis of the subsystem results is a simple three-dimensional integration.

Two questions about the method need answers: how accurate and how efficient is the method when applied to large molecules? Earlier calculations for the diatomic nitrogen molecule demonstrated its ability to describe chemical bond formation [9], a decisive advance beyond the Thomas-Fermi-type theories, which also use the electron density as the basic computational variable but which fail to describe any chemical bond [2-4].

In this report, we address the first question and examine the accuracy of the method in calculations for polyatomic molecules: benzene and a tetrapeptide with four glycine residues. For the exchange-correlation energy functional, we use the $X\alpha$ approximation and do not expect our conclusions to depend on the detail of this functional. We find that the method is capable of describing the electronic structure as accurately as the linear com-

ination of atomic orbitals (LCAO) KS method.

As described in the previous paper [9], to perform a calculation using this approach, one has to choose (i) a division of the molecule into subsystems and (ii) a set of local basis functions to represent the KS Hamiltonian for each subsystem.

To divide a molecule, we define the general normalized partition function for subsystem α as

$$p^\alpha(\mathbf{r}) = g^\alpha(\mathbf{r}) / \sum_{\alpha'} g^{\alpha'}(\mathbf{r}), \quad (1)$$

with

$$g^\alpha(\mathbf{r}) = \sum_{A \in \alpha} [\rho_0^A(|\mathbf{r} - \mathbf{R}_A|)]^2. \quad (2)$$

The summation in Eq. (1) is over all the subsystems, while the summation in Eq. (2) is over all the atoms in subsystem α . $\rho_0^A(|\mathbf{r} - \mathbf{R}_A|)$ is the spherical atomic density for atom A located at \mathbf{R}_A ; accurate KS atomic densities are used throughout this work. The partition function of Eq. (1) is thus a simple extension of what was used in our previous paper [9].

The atomic orbitals, which are used to form the local basis functions for each subsystem and which are also used in LCAO KS calculations, are constructed and optimized as follows.

First, we use the accurate spherical atomic KS orbitals as the minimal basis set for molecular calculations; we denote this set as S , meaning a single function for each orbital. This is essentially the same as the minimal set of other KS molecular calculations [10,11]. But, in order to facilitate the construction of atomic orbitals beyond the S set, we solved self-consistently the atomic KS equations using a linear combination of the following orthogonal functions [12] as approximate radial wave functions:

$$\phi_n^l(r) = e^{-\lambda r/2} r^l L_n^{2l+2}(\lambda r), \quad n=0, 1, \dots, M-1, \quad (3)$$

where l is the angular momentum of the orbital, λ is a scale factor, the L_n^k are the associated Laguerre polynomials, and M is the number of functions. This basis set is complete if $M \rightarrow \infty$. We use $M=30$ for all atoms, $\lambda=4.0$, for H atoms and $\lambda=7.0$ for other atoms, produc-

ing total (spherical) atomic KS energies having 8 or 9 digits agreeing with corresponding converged numerical values.

Second, for each atomic angular momentum l that is occupied, one additional function is added to the S set to form a D set, meaning double functions. The additional function is constructed by projection and optimization: for each l , we project out the components of the S set in the finite M -dimensional space. In the remaining smaller space, the lowest-energy radial eigenfunction of the potential $-Z^*/r$ is chosen to add to the S set to form the D set; Z^* is nonlinearly optimized by the KS (Harris functional [13]) calculations for closed-shell diatomic molecules at equilibrium bond length (H_2 for H atom, C_2 for C atom, N_2 for N atom, and CO for O atom).

Third, the polarization basis set P is the D set plus one function of the angular momentum $l = l_{\text{occ}} + 1$, where l_{occ} is the angular momentum of the highest occupied atomic orbital. This additional function is just a hydrogenic function approximated as a linear combination of ϕ_n^l with a Z^* optimized by the diatomic calculations. Compared with the corresponding Slater basis set, the S , D , and P basis sets are more optimal in calculated KS diatomic energies.

Finally, the projection and optimization procedure can be repeated again to add more functions to each l . A set (432) constructed in this manner is used in our calculations; it contains $4s$, $3p$, and $2d$ functions for all atoms (except $3s$ and $2p$ functions for H atoms). Further details of the atomic basis set construction are to be published elsewhere.

Other technical details in our molecular calculations are the following: all the multicenter three-dimensional integrations are carried out by the method of Delley [11], with a scaled generalized Gauss-Laguerre quadrature rule for the radial coordinates [14]; the electrostatic potential is calculated by the partition method of Delley [11]; and the exchange-correlation energy functional used both in atomic and molecular calculations is the x_α approximation with $\alpha=0.7$.

Tetrapeptide of four glycine residues. We use this molecule of 31 atoms (Fig. 1) to compare the accuracy of our method with the LCAO KS method. The non-self-consistent Harris functional calculations [13] are performed for both methods; this simplifies the calculation and the comparison. The only difference between the two methods is in the approximation to the value of $2\sum_i^N \epsilon_i$, where N is the number of electrons and ϵ_i is the i th eigenvalue of the Kohn-Sham potential generated by the molecular density that is the sum of all the atomic densities. The molecular geometry is fixed in a fully extended

zig-zag flat sheet: repeating units have the same geometry; the N—H bond length is 1.015 Å, and the C—H bond length is 1.091 Å, taken from the experimental values for NH_3 and CH_4 ; four pairs of hydrogen are out of the plane, each bond between the backbone atom and the out-of-plane hydrogen forming a 120° angle with the molecular plane; the bond angle $\angle(H-N-H)=106.6^\circ$.

The molecule is divided into 13 subsystems, each containing one backbone atom and the hydrogen or the oxygen atoms attached to the backbone atom.

Before we describe the local basis sets, let us introduce the concept of *buffer atoms*: buffer atoms of subsystem α are defined as the atoms that are not part of, but are in the neighborhood of, subsystem α and that contribute their atomic orbitals to the local basis set of subsystem α to better represent the KS Hamiltonian in its density calculation. The use of a local basis set for subsystem α in place of the full atomic orbital set introduces a truncation, the effects of buffer atoms are to lessen the truncation error. This was suggested in our previous paper [9].

In our calculations for each type (S , D , or P) of atomic orbital, we use six schemes for the local basis sets: scheme 0 for each subsystem is just the union of atomic orbitals from all the atoms in the subsystem considered, with no buffer atoms; scheme 1, in addition to scheme 0, includes as buffer atoms the first-nearest-neighbor atoms, which are connected to the subsystem by one chemical bond; scheme 2, in addition to scheme 0, includes as buffer atoms up to second-nearest-neighbor atoms, which are connected to the subsystem by up to two chemical bonds; etc. The most extensive is scheme 5, including as buffer atoms up to fifth-nearest-neighbor atoms.

The calculated Harris-energy-functional results are summarized in Table I. The second column contains the LCAO KS Harris functional energies; the values in the third column, $E_{\text{KS}} - E(432)$, are lower-bound estimates of the error of the LCAO KS energy [$E(432) = 898.91349$ a.u. is the result of the basis set (432) described in the preceding text].

From Table I we note that (i) scheme 0, not having any buffer atom, results in substantial deviation from the LCAO values; (ii) scheme 1, having first-nearest-neighbor buffer atoms, already reduces the errors of the present method to the values that are in magnitude smaller than or similar to the errors inherent in the corresponding LCAO results (second column); (iii) further use of buffer atoms in schemes 2–5 continues to decrease the deviation from the corresponding LCAO results. Such a decrease is particularly fast for the calculations with minimal atomic orbitals (S)—the deviation is practically nil when

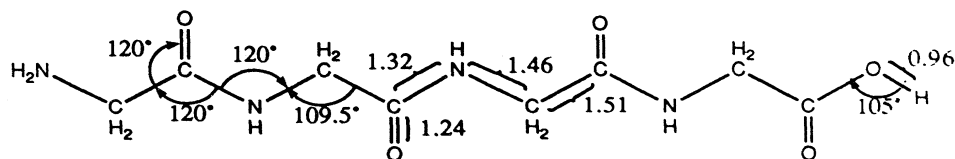


FIG. 1. The tetrapeptide, with bond angles and bond lengths (Å) for the backbone indicated.

up to fifth-nearest-neighbor buffer atoms are included. This has to do with the fact that the atomic orbitals in the *S* set are shorter ranged than in the *D* and *P* sets. It is satisfying to see two very different calculations converge. (iv) There are two numbers in each combination of atomic orbitals and construction schemes of the local basis sets for the subsystems. The first number is from the calculation at $\beta=100$ a.u. (β being the inverse temperature for the Fermi function); the second is from the calculation at $\beta=200$ a.u. They do not differ in a significant way. This establishes the reliability of calculations at finite value of β , instead of the infinite value corresponding to the absolute-zero temperature.

We have also investigated the use of two other forms of partition functions, different from Eq. (2); they are

$$g^\alpha(\mathbf{r}) = \sum_{A \in \alpha} \rho_0^A(|\mathbf{r} - \mathbf{R}_A|),$$

$$g^\alpha(\mathbf{r}) = \sum_{A \in \alpha} \rho_0^A(|\mathbf{r} - \mathbf{R}_A|)/|\mathbf{r} - \mathbf{R}_A|.$$

Our observations about Table I equally apply to the results with these two partition functions. The differences between results of different partition functions decrease with the increasing use of buffer atoms. Our current preferred choice is Eq. (2), which give an overall slightly smaller derivation from the LCAO result.

These calculations confirm that the new method can be as accurate in calculating total energy as the LCAO KS method, as long as a certain degree of buffer atoms are used.

To appreciate the distinction and the efficiency of our method, consider the following. The LCAO KS procedure requires construction and diagonalization of one matrix of dimension: 99, 181, and 308, respectively, for *S*, *D*, and *P* atomic-orbital sets. The method we use requires construction and diagonalization of 13 matrices of dimensions: 11–20, 20–36, and 33–56, respectively, for *S*, *D*, and *P* atomic orbital sets, all with scheme 1 for the local basis set for subsystems (which gives comparable results to the LCAO KS calculations).

Benzene. The foregoing study encourages the use of only the first-nearest-neighbor buffer atoms in the construction of the local basis sets for subsystems. We now investigate this idea in a stringent test: benzene, an aromatic compound. Benzene is well known for its completely delocalized π bonds that cause its aromaticity. The application of the procedure used here means that completed delocalized π bonds no longer contribute to the calculation of the density and the energy. Can the method stand such a test?

We divide the benzene molecule into three subsystems, each having two connected carbon and two hydrogen atoms attached to the carbon atoms. Two first-nearest-neighbor carbon atoms are included in the local basis sets as buffer atoms for each subsystem. Total self-consistent energies as functions of the symmetric stretching of the C—C and C—H bonds are plotted in Figs. 2 and 3. All calculations are done with the inverse temperature $\beta=100$ a.u. We observe that the deviations of our method from the KS method are all within the error of the LCAO KS solution, and that these deviations decrease with the use of better atomic orbitals. The method

TABLE I. Comparison of errors in total energies (a.u.) via Harris functional by the present work (PW) for the tetrapeptide, with three atomic basis sets and six schemes of using buffer atoms. The numbers 0, 1, ..., 5 on top of the columns indicate the schemes. E_{KS} is the LCAO energy with the corresponding atomic basis set. Two errors for each combination are given: the first one for $\beta=100$ a.u. and the second for $\beta=200$ a.u. (Numbers in square brackets are the powers of ten.)

Basis set	E_{KS}	$E_{KS} - E(432)$	$\Delta E = E_{PW} - E_{KS}$					
			0	1	2	3	4	5
<i>S</i>	-896.658 40	2.248	-3.502	4.53[-1]	3.06[-2]	1.73[-3]	2.17[-4]	1.96[-6]
			-3.503	4.52[-1]	3.05[-2]	1.67[-3]	1.01[-4]	-1.62[-5]
<i>D</i>	-898.508 42	4.05[-1]	-9.65[-1]	-2.22[-1]	-1.25[-1]	-1.34[-2]	-5.21[-3]	-9.41[-4]
			-9.69[-1]	-2.26[-1]	-1.25[-1]	-1.36[-2]	-5.25[-3]	-9.60[-4]
<i>P</i>	-898.875 73	3.78[-2]	-2.88[-1]	-3.61[-2]	-8.45[-2]	-1.18[-2]	-5.26[-3]	-1.10[-3]
			-2.97[-1]	-4.01[-2]	-8.53[-2]	-1.21[-2]	-5.33[-3]	-1.13[-3]

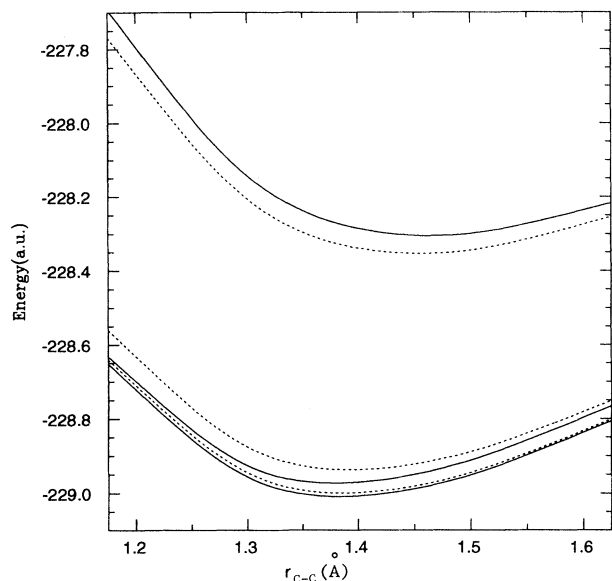


FIG. 2. Comparison of energies as a function of the C—C bond length for benzene between self-consistent calculation by the LCAO Kohn-Sham method (dashed lines) and by the present method (solid lines). The C—H bond length is fixed at experimental value of 1.084 Å. The curves for both methods, from top to bottom, are, respectively, the results of *S*, *D*, and *P* atomic orbital sets.

can accurately describe an aromatic system, even without the use of completely delocalized π bonds.

The confirmation of accuracy being the focus, we do not address the efficiency. Our computer program, coded from scratch, is still in its developmental phase. A comparison of timing would be more meaningful at a later stage.

We anticipate the implementation of the method on coarse-grained massively parallel computers with each processor handling a subsystem of several atoms. Harnessing the additive power of many such processors

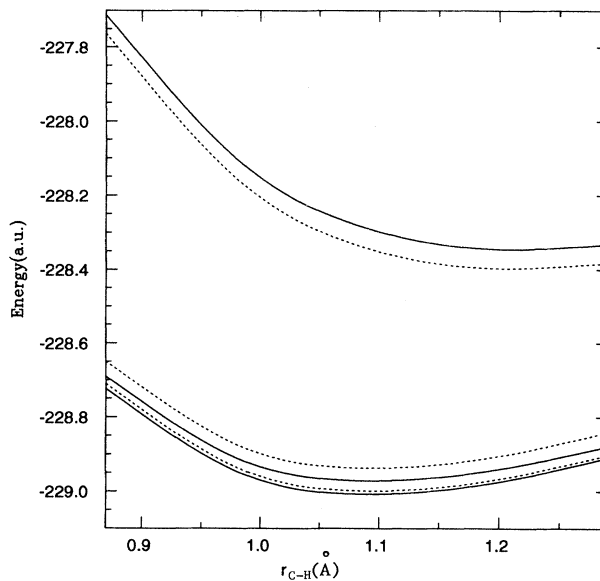


FIG. 3. Comparison of energies as a function of the C—H bond length. The C—C bond length is fixed at experimental value of 1.397 Å. Other specifications are the same as in Fig. 2.

would tremendously enhance the capacity of the *ab initio* quantum theory in its application to large systems.

The author wishes to thank Professor R. G. Parr for his constant encouragement. Discussions with Dr. L. Bartolotti, Dr. Z. Zhou, and Mr. A. Maynard have been helpful. Acknowledgment is made to the Petroleum Research Fund, administered by the American Chemical Society, for partial support of this research. The author is grateful to the North Carolina Board of Science and Technology for financial support, and to the North Carolina Supercomputing Center for a CPU grant. This research was also partially funded by the National Science Foundation.

- [1] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (McGraw-Hill, New York, 1989).
- [2] *Theory of the Inhomogeneous Electron Gas*, edited by S. Lundquist and N. H. March (Plenum, New York, 1983).
- [3] *The Single-Particle Density in Physics and Chemistry*, edited by N. H. March and B. M. Deb (Academic, New York, 1987).
- [4] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
- [5] M. Levy and J. Perdew, in *Density Functional Methods in Physics*, edited by R. H. Dreizler and J. da Providencia (Plenum, New York, 1985).
- [6] R. O. Jones and O. Gunnarsson, *Rev. Mod. Phys.* **61**, 689

- (1989).
- [7] E. S. Kryachko and E. V. Ludena, *Energy Density Functional Theory of Many-Electron Systems* (Kluwer Academic, Dordrecht, 1990).
- [8] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [9] W. Yang, *Phys. Rev. Lett.* **66**, 1438 (1991).
- [10] F. W. Averill and D. E. Ellis, *J. Chem. Phys.* **59**, 6412 (1973).
- [11] B. Delley, *J. Chem. Phys.* **92**, 508 (1990).
- [12] G. Arfken, *Mathematical Methods for Physicists* (Wiley-Interscience, New York, 1986).
- [13] J. Harris, *Phys. Rev. B* **31**, 1770 (1985).
- [14] W. Yang, *J. Chem. Phys.* **94**, 1208 (1991).