

# JCTC

Journal of Chemical Theory and Computation

## Local Hybrid Divide-and-Conquer Method for the Computation of Medium and Large Molecules

Guo-Liang Song,<sup>†</sup> Zhen Hua Li,<sup>†</sup> Zhi-Pan Liu,<sup>†</sup> Xiao-Ming Cao,<sup>†</sup> Wenning Wang,<sup>†</sup>  
Kang-Nian Fan,<sup>\*,†</sup> Yaoming Xie,<sup>‡</sup> and Henry F. Schaefer III<sup>\*,‡</sup>

*Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Department of Chemistry, and Center for Theoretical Chemical Physics, Fudan University, Shanghai 200433, China, and Center for Computational Chemistry, University of Georgia, Athens, Georgia 30602*

Received July 7, 2008

**Abstract:** A local hybrid divide-and-conquer method (LHDC) which combines the high accuracy of sophisticated wave function theory (WFT) methods and the low cost of density functional theory (DFT) has been proposed for computational studies of medium and large molecules. In the method, a large system is divided into small subsystems for which the coefficients of the exchange functional in a hybrid functional are first optimized according to the energy calculated by an accurate WFT method. The hybrid coefficients are then used to evaluate the energy of the whole system. The method not only can reproduce the total energies of the chosen WFT method in good accuracy but also provides electronic structure information for the entire system.

### 1. Introduction

With rapid developing theoretical techniques, it is becoming straightforward to perform electronic structure calculations for many complicated systems. Wave function theory (WFT) methods, such as Hartree–Fock (HF) and post-HF methods, in principle are capable of systematic convergence to the exact solution of the Schrödinger equation for a given system. For the small molecules, high level WFT methods in conjunction with a large basis set have proved to be capable of achieving chemical accuracy.<sup>1</sup> However, for medium and large systems consisting of several hundreds or thousands of atoms conventional WFT methods are still too expensive. To reduce computational costs, new strategies such as linear scaling methods,<sup>2–6</sup> have attracted much attention. By utilizing the locality of the physical properties<sup>5,6</sup> or the nearsightedness of electrons,<sup>7</sup> the computational costs of these methods promise to scale almost linearly with the size of the system instead of the conventional  $O(N^3)$  or higher. Among these linear scaling methods, the divide-and-conquer

(DC) method originally proposed for density functional methods<sup>8</sup> has become popular in the chemistry community.<sup>2</sup>

Recently, a simple and easy to implement DC method, an energy-based DC method (EDC)<sup>9–14</sup> has been proposed. The basic idea of the method is to divide a large system (entire system, ES) into small subsystems. The energy of the ES is then expanded into a sum of many-body terms where the two-, three-, and higher-body correction terms are coupling terms between the subsystems. Since a subsystem within the large system is different from the isolated one, it is important to include the effect of the surrounding environment (i.e., the coupling between the subsystems) when considering the subsystems. The higher accuracy desired, the higher-body correction terms are needed. However, including more higher-body correction terms will greatly increase computational cost, especially for highly accurate WFT methods such as the CCSD(T) method, and the linear-scaling of the method no longer holds. One drawback of the EDC method is that it does not provide information on the electronic structure of the ES since it operates only on the energy of the subsystems.

Density functional theory (DFT) methods can be applied to medium and large systems with moderate computational costs.<sup>14</sup> Many useful functionals have been developed in the past two decades.<sup>16,17</sup> Hybrid DFT methods, which are

\* Corresponding author fax: +86 21 65642978 (K.F.), (706)542-0406 (H.F.S.); e-mail: knfan@fudan.edu.cn (K.F.), sch@uga.edu (H.F.S.).

<sup>†</sup> Fudan University.

<sup>‡</sup> University of Georgia.

developed based on the adiabatic connection formula<sup>18–22</sup> and incorporate a fraction of exact exchange, are the most successful DFT methods to date. However, the challenge to DFT methods is that the exact exchange-correlation functional is unknown and has to be approximated. Unlike WFT methods, there is currently no systematic way to guarantee that DFT methods can converge to right answer. DFT methods still have some major drawbacks compared to conventional WFT methods, such as in the treatment of left-right correlation, van der Waals interactions, and  $\pi$ - $\pi$  stacking.<sup>16</sup>

Since both approaches, the WFT and DFT methods, have their merits and drawbacks, it is desirable to combine the merits of the two approaches. Ab initio DFT<sup>23–25</sup> and Görling-Levy perturbation theory<sup>26–28</sup> are the two such approaches. In the present research, we propose a novel DC approach which combines the accuracy of high-level correlation methods with the low cost of DFT methods based on the adiabatic connection formula. In this approach, an ES is first divided into several subsystems. Then the fraction ( $\eta$ ) of the local DFT exchange ( $E_x^{\text{LSD}}$ ) in a hybrid DFT functional is optimized with respect to the energy of each subsystem (local region) evaluated by a highly accurate WFT method. These  $\eta$  values (the local hybrid (LH) coefficients) are then further refined, by incorporating couplings between the subsystems, and then used to compute the energy of the ES. The new approach not only can reproduce energy profiles of the highly accurate WFT method to high accuracy but also can provide electronic structure information such as charge distributions, where such analyses with highly accurate WFT methods are prohibitive.

## 2. Methodology

The design of hybrid DFT functionals is based on consideration of the adiabatic connection formula,<sup>18–22</sup> which connects the Kohn–Sham noninteracting reference system to the full-interacting physical system

$$E_{xc} = \int_0^1 U_{xc}^\lambda d\lambda \quad (1)$$

where  $E_{xc}$  is the exchange correlation energy,  $\lambda$  is an interelectronic coupling-strength parameter, and  $U_{xc}^\lambda$  is the exchange-correlation potential energy at an intermediate coupling strength  $\lambda$ . In Becke's half-and-half theory,<sup>22</sup>  $E_{xc}$  is approximated as the average of the  $E_{xc}$  of the noninteracting system ( $E_{xc}^{\lambda=0} = E_x^{\text{Exact}}$ ) and that of the full interacting system ( $E_{xc}^{\lambda=1} \approx E_{xc}^{\text{LSD}}$ ), i.e.

$$E_{xc} \cong \frac{1}{2}(E_x^{\text{Exact}} + E_{xc}^{\text{LSD}}) \quad (2)$$

The B3LYP hybrid functional,<sup>29</sup> perhaps the most popular DFT functional, as implemented in Gaussian 03,<sup>30</sup> has the following form:<sup>16</sup>

$$E_{xc} \cong a_0 E_x^{\text{Exact}} + \eta E_x^{\text{LSD}} + a_x \Delta E_x^{\text{B88}} + (1 - a_c) E_c^{\text{VWN-RPA}} + a_c \Delta E_c^{\text{LYP}} \quad (3)$$

In eq 3,  $a_0$  is the coefficient of the exact exchange energy  $E_x^{\text{Exact}}$ ;  $\eta$  is the coefficient of the exchange energy under the local spin density (LSD) approximation  $E_x^{\text{LSD}}$ ;  $\eta$  is usually

taken as  $(1 - a_0)$ ;  $a_x$  is the coefficient of the gradient correction part ( $E_x^{\text{B88}}$ ) of the B88 exchange functional;  $(1 - a_c)$  and  $a_c$  are the coefficients of VWN<sup>31</sup> LSD correlation energy  $E_c^{\text{VWN-RPA}}$  and the gradient correction part of LYP correlation functional  $\Delta E_c^{\text{LYP}}$ ,<sup>32</sup> respectively. The parameters  $a_0$ ,  $a_x$ , and  $a_c$  are those used in the B3PW91 functional,<sup>33</sup> in which they are determined empirically by fitting to a set of accurate thermodynamics data. Similar to the B3LYP functional, in most hybrid functionals only a fraction of  $E_x^{\text{Exact}}$  is included (i.e., mixed with  $E_x^{\text{LSD}}$ ) in  $E_{xc}$ . Usually, the mixing coefficient  $a_0$  is fitted empirically to a set of accurate thermodynamics data and is then fixed for applications to all systems. Thus, the coefficient  $a_0$  is universal in this limited sense. However, because  $E_x^{\text{LSD}}$  and other exchange correlation functionals used in the hybrid functionals are all approximate functionals, if we know the exact solution of the Schrödinger equation and thus the exact  $E_{xc}$ , one would immediately see that the mixing coefficient and also other parameters such as  $a_x$  and  $a_c$  in the B3LYP functional must be system dependent and are probably different in the various regions of a system, i.e. dependent on  $r$ .<sup>34</sup> Thus, the mixing coefficient is not truly universal. Ideally, the mixing coefficient should be determined according to the property of each system.<sup>35</sup>

In principle, if the exact  $E_{xc}$  of a system is known, a coefficient in the hybrid functionals can be solved by fixing other coefficients. Since  $E_{xc}$  is unknown except for a few model systems, one can use  $E_{xc}$  obtained from highly accurate WFT methods, which can be improved by incorporating more correlation and enlarging basis set. Therefore, in the present work, an ES is divided into several subsystems for which a highly accurate WFT method, for example CCSD(T),<sup>36,37</sup> is applicable. The energy of each isolated subsystem (called monomer hereafter) is then computed. We fix all but one coefficient (typically  $a_0$ ), by solving, if the B3LYP functional is selected, the following equation

$$E_{xc}^{\text{WFT}} = a_0 E_x^{\text{Exact}} + \eta E_x^{\text{LSD}} + a_x \Delta E_x^{\text{B88}} + (1 - a_c) E_c^{\text{VWN-RPA}} + a_c \Delta E_c^{\text{LYP}} \quad (4)$$

In (4)  $E_{xc}^{\text{WFT}}$  is the exchange-correlation energy calculated by the accurate WFT method, and an  $a_0$  value for a monomer is then obtained. It should be noted that since electron density depends on  $a_0$  and  $E_x^{\text{Exact}}$  and other energy terms thus depend implicitly on  $a_0$ ,  $a_0$  has to be solved iteratively.

In the current study we choose B3LYP as the example functional in our local hybrid DC method (LHDC), since B3LYP is widely used and it is well-known that B3LYP (and most other DFT functionals) fails for long-range interactions such as  $\pi$ - $\pi$  stacking, which is very important in biomolecules. For the accurate WFT method, the CCSD(T) method is employed. In the B3LYP functional, we choose  $\eta$  to be varied since the exchange energy is much more important than the correlation energy.<sup>38,39</sup> Conventionally  $\eta$  is fixed to be  $1 - a_0$  to fulfill the uniform electron gas (UEG) limit, but in molecular systems it was found in practice that violation of such constraint still gives promising results, for example in the O3LYP functional.<sup>40</sup> On the other hand, usually 20%–25% ( $a_0$ ) of the exact exchange is included in hybrid functionals.<sup>33,41</sup> Therefore, we will fix  $a_0$  to be 0.20,  $a_x$  to be 0.72,  $a_c$  to be 0.81, and vary  $\eta$  using

0.80 as an initial guess. Our preliminary tests show that varying  $\eta$  instead of varying  $a_0$  with the constraint of  $\eta = 1 - a_0$  actually gives slightly better results. Further, better numerical stability is found when iteratively solving for the mixing coefficient (i.e.,  $\eta$  or  $a_0$ ) using eq 4.

In the EDC method,<sup>9–14</sup> the total energy of an ES is expanded as

$$E_{\text{ES}} \equiv \sum_i^M E_i = \sum_i^M (E_i^{(0)} + \Delta E_i^{(2)} + \Delta E_i^{(3)} + \dots)$$

$$\Delta E_i^{(2)} = \frac{1}{2} \sum_{j \neq i}^M [E_{ij}^{(0)} - (E_i^{(0)} + E_j^{(0)})]$$

$$\Delta E_i^{(3)} = \frac{1}{3} \sum_{k \neq j \neq i}^M [E_{ijk}^{(0)} - (E_{ij}^{(0)} + E_{ik}^{(0)} + E_{jk}^{(0)} - (E_i^{(0)} + E_j^{(0)} + E_k^{(0)})]$$
 (5)

where  $E_i$  is the energy of a subsystem in the ES environment,  $E_i^{(0)}$  is the energy of a monomer,  $\Delta E_i^{(2)}$  and  $\Delta E_i^{(3)}$  are second-order and third-order corrections to  $E_i^{(0)}$  (due to the couplings between two monomers and three monomers, respectively), and  $M$  is the number of subsystems. By analogy to eq 5, the  $\eta$  value of a subsystem (local  $\eta$ ) embedded in the ES environment may be expressed as

$$\eta_i = \eta_i^{(0)} + \Delta \eta_i^{(2)} + \Delta \eta_i^{(3)} + \dots$$
 (6)

where  $\eta_i^{(0)}$  is the  $\eta$  value of a monomer determined iteratively as described above. As energy is an extensive thermodynamic variable while the hybrid coefficient  $\eta$  is an intensive one, the expressions for  $\Delta \eta_i^{(2)}$  and  $\Delta \eta_i^{(3)}$  are slightly different from those for  $\Delta E_i^{(2)}$  and  $\Delta E_i^{(3)}$

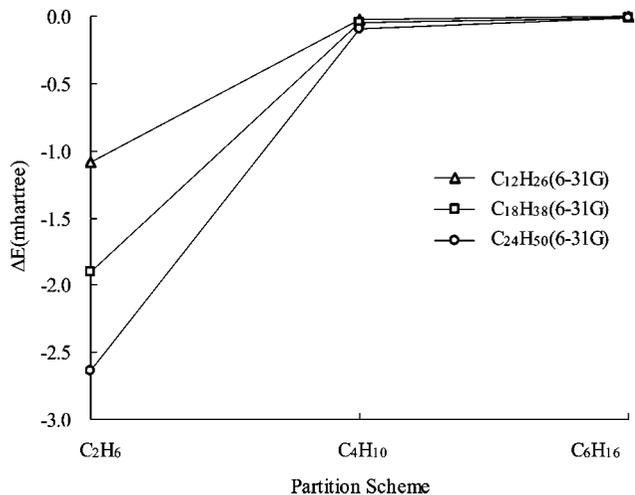
$$\Delta \eta_i^{(2)} = \sum_{j \neq i}^M (\eta_{ij}^{(0)} - \eta_i^{(0)})$$
 (7)

$$\Delta \eta_i^{(3)} = \sum_{k \neq j \neq i}^M [\eta_{ijk}^{(0)} - (\eta_{ij}^{(0)} + \eta_{ik}^{(0)}) + \eta_i^{(0)}]$$
 (8)

where  $\eta_{ij}^{(0)}$  is the  $\eta$  value of two subsystems determined in the same way as for a monomer. For example, in eq 4 all the energy terms used are calculated for two subsystems in a whole as a bigger isolated subsystem (named dimer hereafter). The use of the above expressions for  $\Delta \eta_i^{(2)}$  and  $\Delta \eta_i^{(3)}$  is to guarantee that if no truncation is applied to eq 6, the  $\eta_i$  given by eq 6 is exactly the  $\eta$  of the ES to fulfill the condition that  $\eta$  is an intensive variable.

Once all the local  $\eta$  values are determined, a DFT calculation is performed for the ES to solve the Kohn–Sham equation with these new  $\eta$  values. Since the matrix elements of the exchange–correlation potential  $\mathbf{v}_{\text{xc}}$  are evaluated in finite basis sets, when calculating element  $\langle \phi_u | \mathbf{v}_{\text{xc}} | \phi_v \rangle$  where  $\phi_u$  and  $\phi_v$  are atomic orbitals belonging to different subsystems  $i$  and  $j$ , an average value of  $\eta$ , i.e.  $(\eta_i + \eta_j)/2$  is used. Finally, the total energy and the Kohn–Sham wave function of the system are obtained.

Eq 6 is exact in reproducing the energy calculated by the accurate WFT method if all correction terms up to  $M^{\text{th}}$  order are included. In practical usage, eq 6 has to be truncated. To further reduce computational costs, contributions from



**Figure 1.** Error (relative energy to the full CCSD(T) energy, in mhartree) of the LHDC2(B3LYP:CCSD(T)) method for the longer straight-chain alkanes under three partition schemes (see text for details).

some dimers, trimers (three subsystems calculated in a whole),..., and  $m$ -mers consisting of far apart subsystems may be neglected. In the present work, the sum in eq 6 only includes contributions from those dimers, trimers,..., and  $m$ -mers in which subsystems are connected through chemical bonds (neighboring subsystems). Three different truncation schemes have been investigated, where  $\eta$  is corrected to zeroth order, second order, and third order by eq 6, respectively. It should be noted that for the conventional EDC method, there is no zeroth order method for systems that must be divided by cutting through a chemical bond, since the contributions from cap atoms cannot be canceled out in eq 5.

The model systems selected to test our LHDC methods are three straight-chain alkanes, two fully conjugated straight-chain alkenes, and the benzene dimer. The nomenclature of the new method is LHDC $n$ (DFT:WFT) in which  $n$  ( $= 0, 2, 3, \dots$ ) represents the truncation order in eq 6, DFT represents the hybrid functional used, and WFT represents the WFT method used. In the current research, DFT is B3LYP and WFT is CCSD(T). To divide ES into subsystems, one may cut through chemical bonds. In such a case, the broken bond is capped by a hydrogen atom along the direction of the original bond with a fixed C–H bond length of 1.100 Å. All the WFT calculations and the DFT calculations except the final one for ES are performed with the Gaussian 03 software package.<sup>30</sup> The iterative solution for the mixing coefficient is done utilizing the IOP option of Gaussian 03 to change the coefficients of a hybrid functional. The final DFT calculation with local  $\eta$  values is performed with our own program using integrals calculated from the Gaussian 03 program.

### 3. Results and Discussion

**3.1. Straight-Chain Alkanes.** Three straight-chain alkanes with increasing sizes, C<sub>12</sub>H<sub>26</sub>, C<sub>18</sub>H<sub>38</sub>, and C<sub>24</sub>H<sub>50</sub>, have been tested as prototypes of saturated systems. For the purpose of testing the accuracy of our method on total energies, the geometries of these molecules are not optimal.

**Table 1.** Total Energies for the C<sub>16</sub>H<sub>18</sub> Molecule by the LHDC*n*(B3LYP:CCSD(T)) (*n* = 0, 2, 3) and CCSD(T) Methods

basis set	CCSD(T) <i>E</i> (a.u.)	subsystem <sup>a</sup>	$\eta^b$			$\Delta E$ (mh) <sup>d</sup>		
			LHDC0 <sup>c</sup>	LHDC2	LHDC3	LHDC0	LHDC2	LHDC3
6-31G	-617.53317	terminal C <sub>4</sub> H <sub>6</sub>	0.76689	0.76697	0.76698	6.82	-2.78	-0.54
		middle C <sub>4</sub> H <sub>6</sub>		0.76705	0.76699			
6-31G(d)	-618.35579	terminal C <sub>4</sub> H <sub>6</sub>	0.77542	0.77564	0.77569	23.00	-2.72	-0.57
		middle C <sub>4</sub> H <sub>6</sub>		0.77585	0.77574			
cc-pVDZ	-618.53343	terminal C <sub>4</sub> H <sub>6</sub>	0.77770	0.77753	0.77745	-23.20	-2.70	-0.62
		middle C <sub>4</sub> H <sub>6</sub>		0.77736	0.77738			
cc-pVTZ <sup>e</sup>	-619.02579	terminal C <sub>4</sub> H <sub>6</sub>	0.78188	0.78187	0.78196	-3.32	-2.30	-0.49
		middle C <sub>4</sub> H <sub>6</sub>		0.78187	0.78183			

<sup>a</sup> The C<sub>16</sub>H<sub>18</sub> molecule is divided into 4 C<sub>4</sub>H<sub>6</sub> subsystems. <sup>b</sup>  $\eta$  is the coefficient of the exchange energy under the local spin density (LSD) approximation  $E_X^{\text{DA}}$  and is optimized in the LHDC method. <sup>c</sup> The terminal and middle C<sub>4</sub>H<sub>6</sub> have the same  $\eta$  value for the LHDC0 method since they have the same geometry. <sup>d</sup> The error  $\Delta E$  (in mhartree) is the energy relative to the full CCSD(T) energy. <sup>e</sup> The cc-pVTZ basis set is for carbon, and the cc-pVDZ basis set is used for hydrogen.

In all the alkanes the C–C bonds are fixed at 1.523 Å; C–H bonds are fixed at 1.115 Å; CCH, CCC, and HCH bond angles are fixed at 110.0, 124.9, and 109.3 degrees, respectively; and CCCC dihedral angles are fixed at 180.0 degrees. The 6–31G(d) basis set is used for all the calculations.

In the DC methods, the accuracy of the methods is greatly affected by the choice of the partition scheme. Usually, the larger the subsystems, the higher the accuracy and the computational cost are. It is thus essential to find a balance between accuracy and cost. In the present study we have tested three partition schemes. In the first scheme, the molecules are divided into 6, 9, and 12 C<sub>2</sub>H<sub>6</sub> molecules (capped by hydrogen atoms), respectively. In the second scheme, the molecules are divided into 3 C<sub>4</sub>H<sub>10</sub>, 4 C<sub>4</sub>H<sub>10</sub> + 1 terminal C<sub>2</sub>H<sub>6</sub>, and 6 C<sub>4</sub>H<sub>10</sub> molecules, respectively. In the third scheme, the molecules are divided into 2, 3, and 4 C<sub>6</sub>H<sub>14</sub> molecules, respectively. The three partition schemes here are designated C<sub>2</sub>H<sub>6</sub>, C<sub>4</sub>H<sub>10</sub>, and C<sub>6</sub>H<sub>14</sub> partition schemes.

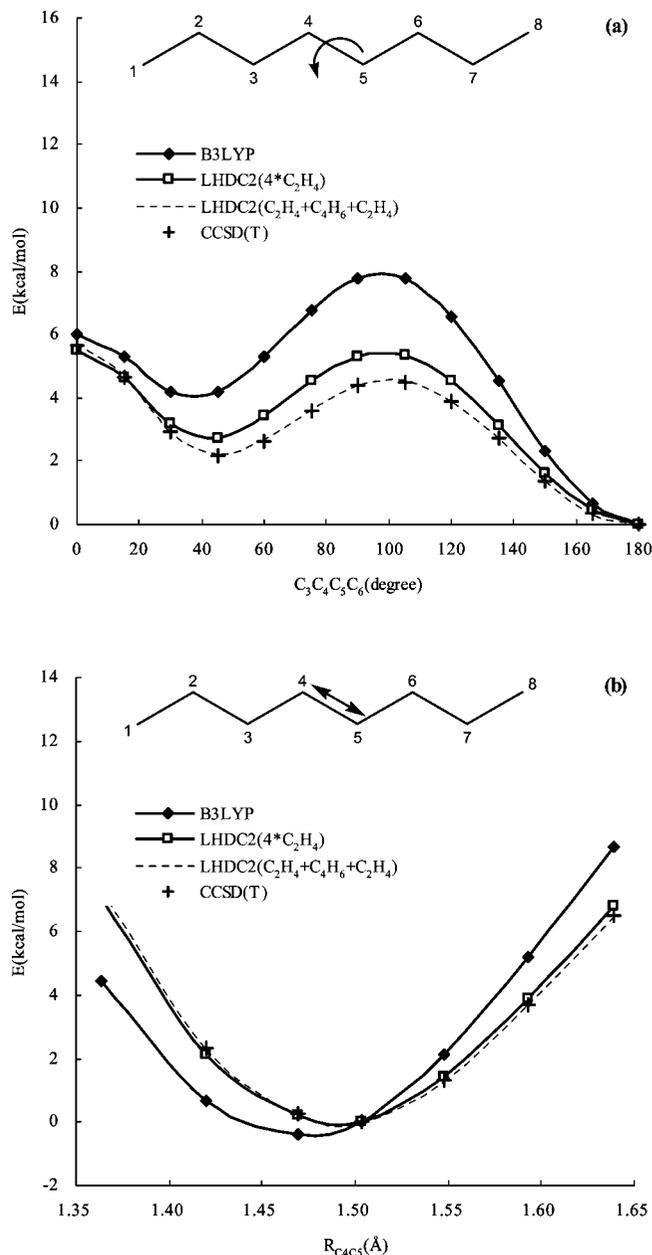
In Figure 1 the error (the energy difference between the LHDC and the full CCSD(T) methods) of the second-order LHDC method is plotted as a function of the size of the subsystems. It can be seen that the LHDC2(B3LYP:CCSD(T)) method with the C<sub>2</sub>H<sub>6</sub> partition scheme can already reproduce the full CCSD(T) energy within 3 mhartree (1.9 kcal mol<sup>-1</sup>). As the size of the subsystem enlarged from C<sub>2</sub>H<sub>6</sub> to C<sub>4</sub>H<sub>10</sub>, the error is greatly decreased, down to below 0.1 mhartree, which is just 0.06 kcal mol<sup>-1</sup>. For the C<sub>4</sub>H<sub>10</sub> partition scheme, the most timing consuming step is the calculation of dimers (i.e., C<sub>8</sub>H<sub>14</sub>) at the CCSD(T) level, which is still affordable on a common personal computer (PC) with a moderate basis set. For the C<sub>6</sub>H<sub>14</sub> partition scheme, although the error is very small (0.013 mhartree), it requires the calculation of a large C<sub>12</sub>H<sub>26</sub> molecule, which is perhaps a little expensive for PC computation and is not recommended. It is worth noting that the error of the LHDC method increases with the size of the alkane and is always negative, indicating the energy calculated by the LHDC method is systematically lower than that of the WFT method.

**3.2. Straight-Chain Alkenes.** To test our method on conjugated systems, two fully conjugated straight-chain alkenes, C<sub>16</sub>H<sub>18</sub> and C<sub>8</sub>H<sub>10</sub> ((3E,5E)-octa-1,3,5,7-tetraene), have been tested. In both alkenes, C–C single and double bonds are arbitrarily fixed at 1.500 Å; C–H bonds are fixed at 1.100 Å; and CCC and CCH bond angles are both fixed at 120 degrees. For C<sub>16</sub>H<sub>18</sub>, we have also tested the effect of

basis set. Four basis sets have been tested: 6–31G, 6–31G(d), cc-pVDZ, and cc-pVTZ\*. In the latter the cc-pVTZ basis set is used for carbon, while the smaller cc-pVDZ basis set is used for hydrogen to reduce computational times. The results for C<sub>16</sub>H<sub>18</sub> are presented in Table 1, where C<sub>16</sub>H<sub>18</sub> is divided into 4 C<sub>4</sub>H<sub>6</sub>.

From Table 1, it may be seen that energies given by the LHDC0 method can already reproduce the full CCSD(T) energy to an accuracy of about 14 kcal mol<sup>-1</sup> (23 mhartree). For the LHDC2 and LHDC3 methods, the energies calculated are systematically lower than the full CCSD(T) energies, by –2.78 to –2.30 mhartree for the LHDC2 method and by –0.62 to –0.49 mhartree for the LHDC3 method. However, for the LHDC0 method including no higher-order corrections, the error is not systematic. For the LHDC2 and LHDC3 methods, the basis-set dependence of the error is rather weak, and a moderate decrease is observed when the basis set varies from simple to complex. Compared with saturated hydrocarbons, the error in the total energy for the fully conjugated hydrocarbons is at least an order of magnitude larger when alkene and alkane are both divided into subsystems containing four carbon atoms. Therefore, to achieve a similar accuracy in total energy as that for saturated hydrocarbons, either higher-order corrections should be included or larger subsystems should be used for conjugated systems. This is not unexpected since electrons in conjugated systems are more delocalized.

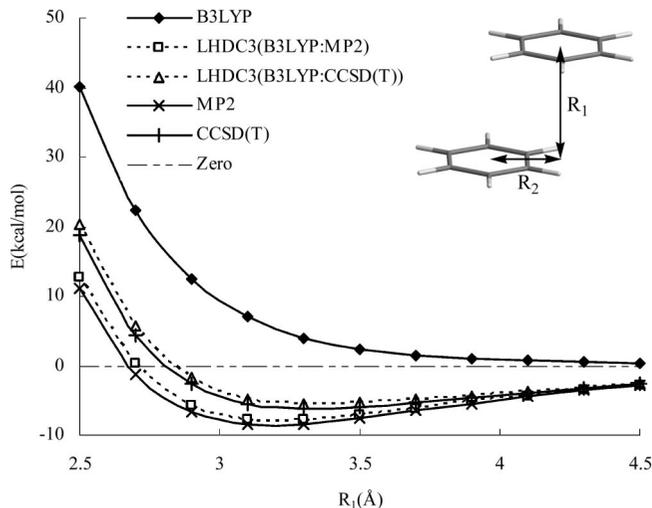
For C<sub>8</sub>H<sub>10</sub>, we have investigated the potential energy profiles with respect to rotation around and stretching along the central C4C5 bonds, while keeping all other geometrical parameters fixed. The resulting energy profiles are presented in Figure 2. Since one of the major deficiencies of the B3LYP method is the calculation of long-range interactions, which play an important role in the C<sub>8</sub>H<sub>10</sub> rotation barrier, including second-order corrections is necessary to give good results. Therefore, the results of the LHDC0 method are not presented here. Figure 2 indicates that LHDC methods greatly improve the results of the original B3LYP method. The results indicate that although the error in the total energy computed by the LHDC2 method is large (see also Table 1), the error in the calculated relative energies is much smaller (less than 1 kcal mol<sup>-1</sup>) even when C<sub>8</sub>H<sub>10</sub> is divided into subsystems as small as C<sub>2</sub>H<sub>4</sub>. On the other hand, the error for the original B3LYP method is as large as 3.4 kcal mol<sup>-1</sup>. When enlarged subsystems are used, the energy



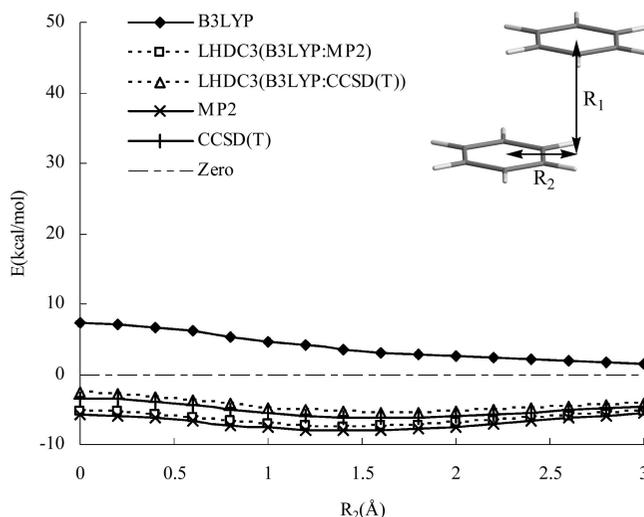
**Figure 2.** Potential energy profiles of (3E,5E)-octa-1,3,5,7-tetraene: a) rotating around the central C4C5 bond, where the zero of the energy is that of the structure with a C3C4C5C6 dihedral angle of 180 degrees and b) stretching along the central C4C5 bond, where the zero of the energy is that of the structure with a C4C5 bond length of 1.5 Å. Two partition schemes have been investigated: dividing into 4 C<sub>2</sub>H<sub>4</sub> and 2 terminal-C<sub>2</sub>H<sub>4</sub> + 1 C<sub>4</sub>H<sub>6</sub>.

profiles calculated by the LHDC2 method nearly overlap with those calculated by the full CCSD(T) method. Therefore, the error in the total energies calculated by the LHDC methods except LHDC0 is most likely systematic. Since relative energies are more meaningful than total energies, it seems that even for fully conjugated systems, the LHDC2 method with a rather fine partition scheme is acceptable.

**3.3. Benzene Dimer.** The benzene dimer is a notorious system for most DFT functionals.<sup>42</sup> For example, it is unbound at the B3LYP level.<sup>43</sup> In the present LHDC calculations, benzene is divided into three C<sub>2</sub>H<sub>2</sub> subsystems, and the third-order LHDC method is used since benzene

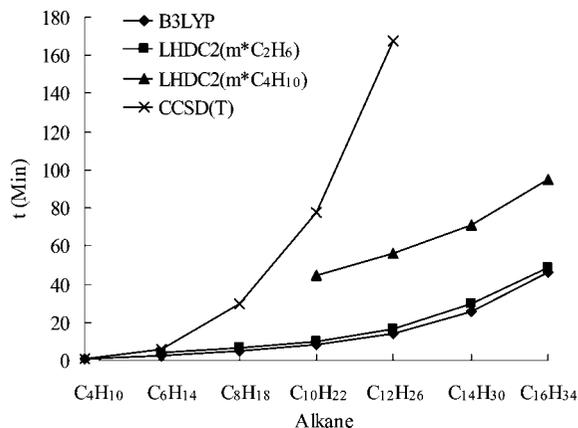


**Figure 3.** Potential energy profiles for the parallel displaced configuration of the benzene dimer, with horizontal separation  $R_2$  fixed at 1.6 Å. The zero of the energy is that for two isolated benzenes, each with the geometry in the dimer.

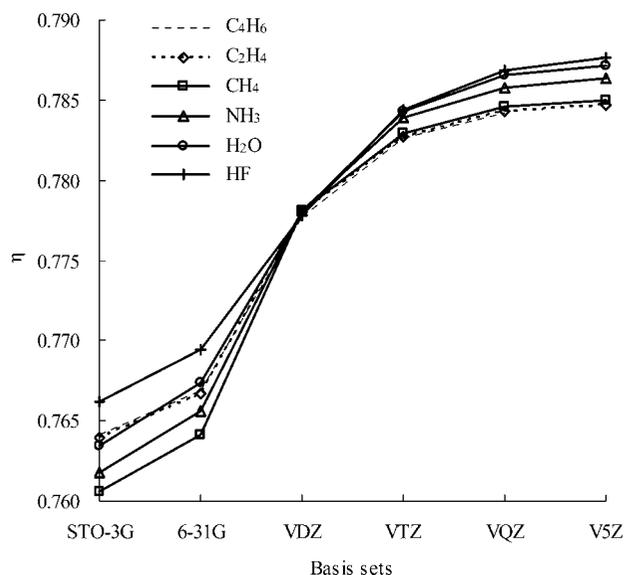


**Figure 4.** Potential energy profiles for the parallel displaced configuration of the benzene dimer, with vertical separation  $R_1$  fixed at 3.4 Å. The zero of energy is that for two isolated benzenes, each with the geometry in the dimer.

contains three C<sub>2</sub>H<sub>2</sub> units and its  $\pi$  electrons are highly delocalized. It should be noted that all trimers have been included for this special case of benzene dimer. The energy profiles of the parallel configuration of the benzene dimer are studied, where the CC and CH bonds are fixed at 1.391 Å and 1.080 Å,<sup>44</sup> respectively, and the  $R_1$  and  $R_2$  geometrical parameters as indicated in Figures 3 (varying  $R_1$  with  $R_2$  fixed at 1.6 Å) and 4 (varying  $R_2$  with  $R_1$  fixed at 3.4 Å) are varied. The purpose of this test is to see if our LHDC method can successfully overcome the deficiency of the B3LYP method and how accurate it is in reproducing the energy profiles predicted by the sophisticated WFT method. Therefore, the energy profiles presented in Figures 3 and 4 are calculated with a moderate aug-cc-pVDZ basis set and are not corrected by basis set superposition error (BSSE). The zero of energy in Figures 3 and 4 is the two isolated benzene monomers with the geometry found for the dimer.



**Figure 5.** Timings ( $t$ ) in minutes for the calculation of straight-chain alkanes on a single 3.0-GHz Xeon CPU. Two partition schemes for the LHDC2(B3LYP:CCSD(T)) methods are investigated: the molecule is divided into  $C_2H_6$  and  $C_4H_{10}$  subsystems.



**Figure 6.** Basis-set dependence of the hybrid coefficient  $\eta$  for six small molecules. Basis sets considered are STO-3G, 6-31G(d), VDZ (cc-pVDZ), VTZ (cc-pVTZ), VQZ (cc-pVQZ), and V5Z (cc-pV5Z).

Two LHDC methods have been tested, LHDC3(B3LYP:MP2) and LHDC3(B3LYP:CCSD(T)) methods, to reproduce the energy profiles calculated by the full MP2 and CCSD(T) methods, respectively. The results indicate that both methods can accurately reproduce the energy profiles calculated by the WFT methods. Except for the first point in Figure 3, the differences between the energies calculated by the LHDC3 methods and the corresponding full WFT methods are small, all below 1 kcal mol<sup>-1</sup>. It can also be seen that benzene dimer is slightly over bound under the LHDC treatments compared to the full WFT methods.

**3.4. Cost and Efficiency.** The computational requirements of the LHDC method are greater than a conventional DFT calculation since in addition to the calculation of ES at the DFT level it also requires additional computations to obtain local  $\eta$  values. The additional cost consists of two parts: the first part is the energy calculation of  $m$ -mers ( $m = 0, 2, 3, \dots, n$ ) by the WFT method, and the second part is the iterative

solution for local  $\eta$  values by the DFT method. The iterative procedure usually finishes in less than 4 iterations. It is known that the DFT method is much faster than correlation WFT methods, and thus the costs for the DFT calculations for the  $m$ -mers are negligible compared to the expensive WFT methods. In addition, for the demanding WFT methods, the calculation of an  $m$ -mer ( $m > 1$ ) is much more expensive than that of an  $(m-1)$ -mer. Therefore, the cost for the LHDC method mainly comes from two calculations, the calculation of the largest  $m$ -mers by the WFT method and the final DFT calculation for the ES. In the LHDC2 methods, the calculation of  $M(M-1)/2$  dimers is required if all second-order corrections are included. However, this number can be greatly reduced by neglecting contributions from dimers in which two monomers are far apart. For example, in the present computations for straight-chain alkanes and alkenes, this number is reduced to  $M-1$  by excluding those dimers in which monomers are not connected through chemical bonds.

In Figure 5 we have compared the computation time for the straight-chain  $C_nH_{2n+2}$  ( $n = 4-16$ ) alkanes which are divided into  $C_2H_6$  or  $C_4H_{10}$  subsystems. The 6-31G basis set is used, and the calculations are performed on a single 3.0-GHz XEON CPU. The comparison indicates that the costs of the LHDC2(B3LYP:CCSD(T)) method with an  $C_2H_6$  partition scheme are just slightly higher than those of the full B3LYP method. The results for the alkenes indicate that the LHDC2(B3LYP:CCSD(T)) method with a fine partition scheme in which alkanes and alkenes are divided into subsystems containing two carbon atoms is probably accurate enough to study the energy profiles of common molecules. For highly conjugated systems containing aromatic rings, one may need to use the more expensive LHDC3 method for satisfactory accuracy. However, the LHDC3(B3LYP:CCSD(T)) method is still much less expensive than the full CCSD(T) method.

### 3.5. Basis-Set Dependence of the Hybrid Coefficient $\eta$ .

Six sample molecules including methane, ethene, 1,3-dibutene, ammonia, water, and hydrogen fluoride are selected to test basis set dependence of  $\eta$ . The geometries of these molecules are all optimized using the B3LYP/6-31G(d,p) method. The results are presented in Figure 6. Two conclusions may be drawn from our results. First,  $\eta$  for each molecule appears systematically convergent to the basis set limit when the basis set is enlarged up to cc-pV5Z, and the changes from the smallest basis set to the largest basis set are only about 3%. Though the basis set dependence of  $\eta$  is not significant, we have found the change of the total energy for a given molecule is significant. However, since the dependence of  $\eta$  on the basis set is systematic, this will not introduce nonsystematic errors for relative energies, which are more meaningful than total energies. Second, the  $\eta$  values for different molecules are close to each other. For a given basis set, the difference between the  $\eta$  values for the molecule with the largest  $\eta$  and the molecule with the smallest  $\eta$  is only 0.006.

#### 4. Concluding Remarks

We have presented a novel LHDC method, which combines highly accurate WFT methods with low-cost DFT methods, to compute the energy and electronic structure of medium and large molecules. The method requires one to divide an entire system into small subsystems and to optimize the local hybrid coefficient  $\eta$  of a hybrid DFT functional for each subsystem according to the energy of the subsystem calculated by a high level WFT method. Further refinement to  $\eta$  (local hybrid coefficient) is made by incorporating couplings between the subsystems. The local hybrid coefficients are then used to calculate the energy and electronic structure of the entire system with the hybrid DFT functional. The new method can accurately reproduce the energy profiles of the highly accurate WFT methods.

Since for small subsystems, highly accurate WFT methods may be applied, our LHDC method provides a route to systematically converge DFT results to “exact” results for medium and large molecules which are too large to apply the accurate WFT methods to. The cost of the LHDC method equals the cost of a usual DFT calculation for the entire system plus additional efforts to obtain the local hybrid coefficients. Our results indicate that a second-order LHDC method which is just slightly more expensive than a DFT method can already give satisfactory results on potential energy profiles for highly delocalized systems such as fully conjugated alkenes.

From a WFT viewpoint, our method is a DC method using DFT as “glue” to bind subsystems but with the capability of reproducing the WFT energy and providing electronic structure information for the entire system that is not addressed by conventional EDC methods. From a DFT viewpoint, the hybrid coefficients of subsystems (local regions) are optimized according to the energy of the accurate WFT method, and thus our method can be viewed as a new local hybrid DFT method.<sup>34</sup> Thus the “local hybrid (LH)” in the name LHDC has a double meaning, the *hybrid* of DFT and WFT through the optimization of local *hybrid* coefficient  $\eta$  of the DFT exchange.

**Acknowledgment.** This work is supported by the National Natural Science Foundation of China (20433020, 20673024, 20828003) and by the U.S. National Science Foundation (CHE-0749878).

#### References

- (1) Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.
- (2) Galli, G. *Curr. Opin. Solid State Mater. Sci.* **1996**, *1*, 864.
- (3) Morokuma, K. *Phil. Trans. R. Soc. London A* **2002**, *360*, 1149.
- (4) Bowler, D. R.; Miyazaki, T.; Gillan, M. J. *J. Phys.: Condens. Matter* **2002**, *14*, 2781.
- (5) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085.
- (6) Goedecker, S.; Scuseria, G. E. *Comput. Sci. Eng.* **2003**, *5*, 14.
- (7) Kohn, W. *Phys. Rev. Lett.* **1996**, *76*, 3168.
- (8) Yang, W. *Phys. Rev. Lett.* **1991**, *66*, 1438.
- (9) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832.
- (10) Li, S. H.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7215.
- (11) Stoll, H.; Paulus, B.; Fulde, P. *J. Chem. Phys.* **2005**, *123*, 144108.
- (12) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.
- (13) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.
- (14) Fedorov, D. G.; Ishida, T. *J. Comput. Chem.* **2007**, *28*, 1476.
- (15) Kohn, W. *Mod. Phys.* **1999**, *71*, 1253.
- (16) Scuseria, G. E.; Staroverov, V. N. Progress in the Development of Exchange- Correlation Functionals In *Theory and Application of Computational Chemistry: The First 40 Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 669–724.
- (17) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
- (18) Harris, J.; Jones, R. O. *J. Phys. F* **1974**, *4*, 1170.
- (19) Gunnarsson, O.; Lundqvist, B. I. *Phys. Rev. B* **1976**, *13*, 4274.
- (20) Langreth, D. C.; Perdew, J. P. *Phys. Rev. B* **1977**, *15*, 2884.
- (21) Harris, J.; Jones, R. O. *Phys. Rev. A* **1984**, *29*, 1684. Yang W. *J. Chem. Phys.* **1998**, *109*, 10107.
- (22) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1327.
- (23) Bartlett, R. J.; Lotrich, V. F.; Schweigert, I. V. *J. Chem. Phys.* **2005**, *123*, 062205.
- (24) Bartlett, R. J.; Schweigert, I. V.; Lotrich, V. F. *J. Mol. Struct.* **2006**, *771*, 1.
- (25) Grabowski, I.; Lotrich, V.; Bartlett, R. J. *J. Chem. Phys.* **2007**, *127*, 154111.
- (26) Görling, A.; Levy, M. *Phys. Rev. A* **1993**, *47*, 13105.
- (27) Görling, A.; Levy, M. *Phys. Rev. A* **1994**, *50*, 196.
- (28) Görling, A.; Levy, M. *Int. J. Quantum Chem. Symp.* **1995**, *29*, 93.
- (29) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
- (30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 03, Revision C. 02*; Gaussian, Inc.: Pittsburgh, PA, 2003.
- (31) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys* **1980**, *58*, 1200.
- (32) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (34) Jaramillo, J.; Scuseria, G. E.; Ernzerhof, M. *J. Chem. Phys.* **2003**, *118*, 1068.

- (35) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.
- (36) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. *Int. J. Quantum Chem.* **1978**, *14*, 545.
- (37) Bartlett, R. J.; Purvis, G. D. *Int. J. Quantum Chem.* **1978**, *14*, 516.
- (38) Ivanov, S.; Bartlett, R. J. *J. Chem. Phys.* **2001**, *114*, 1952.
- (39) Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673.
- (40) Cohen, A. J.; Handy, N. C. *Mol. Phys.* **2001**, *99*, 607.
- (41) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1997**, *274*, 242.
- (42) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656.
- (43) Ye, X. Y.; Li, Z. H.; Wang, W. N.; Fan, K. N.; Xu, W.; Hua, Z. Y. *Chem. Phys. Lett.* **2004**, *397*, 56.
- (44) Gauss, J.; Stanton, J. F. *J. Phys. Chem. A* **2000**, *104*, 2865.

CT800265P