# Extended Energy Divide-and-Conquer Method Based on Charge Conservation

Guo-Liang Song, Zhen Hua Li,* and Kang-Nian Fan*

Shanghai Key Laboratory of Molecular Catalysts and Innovative Materials, Department of Chemistry, Fudan University, Shanghai 200433, P. R. China

S *Supporting Information*

ABSTRACT: The divide-and-conquer (DC) scheme, the most popular linear-scaling method, is very important in the quantum mechanics computation of large systems. However, when a chemical system is divided into subsystems, its covalent bonds are often broken and then capped by complementary atoms/groups. In this paper, we show that the charge transfer between subsystems and the complementary atoms/groups causes the nonconservation of the total charge of the whole system, and this is the main source of error for the computed total energy. On the basis of this finding, an extension of the many-body expansion method (energy-based divide-and-conquer, EDC) utilizing charge conservation (E-EDC) is proposed. In the E-EDC method, initially the total energies of the whole system at different many-body correction levels are computed according to the EDC scheme. The total charges of the whole system, that is, the sum of the charges of the subsystems without cap atoms/groups at different many-body correction levels, are also computed. Then the total energy is extrapolated to the value at which the net charge of the whole system equals to the real value. Other properties such as atomic forces can also be extrapolated in a similar way. In the test of 24 and 32 glycine oligomers, this scheme reduces the error of the total energy by about 40−70%, but the computational cost is almost the same as that of the EDC scheme.
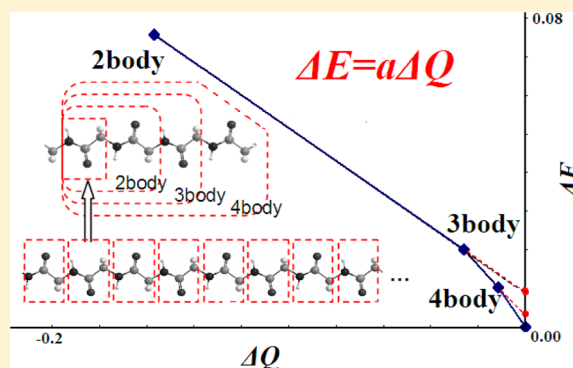
1. INTRODUCTION

At present, an empirical force field is usually adopted for the computation and simulation of large systems containing tens to thousands of atoms in material science and biologic research. Ab initio methods such as wave function theory (WFT) and density functional theory (DFT) are more accurate;[1−10] however, due to the $O(N^{2−7})$ computational scaling, they can only be applied to systems containing about several hundred atoms, even on the fastest supercomputers available today. To reduce the computational costs while still keeping high accuracy, two popular techniques, the QM/MM and the DC methods, have been developed in the last two decades.[11−31]

In 1991, Yang and Lee first imported the divide-and-conquer (DC) philosophy into the noniterative density functional theory (DFT) for enabling linear-scaling computation with respect to system size.[11] This scheme is called the density-based divide-and-conquer (DDC) method. In this method, a large system (entire system, ES) is divided into small pieces (subsystems), and each subsystem is surrounded by the so-called buffer region. Each subsystem is calculated independently. The subsystem results are then combined together to give the total energy and the charge density of the whole system. Furthermore, in order to allow the combined density to relax, a global Fermi energy has to be used for the occupation of the fragment wave functions, thus allowing charge transfer between subsystems. One of the merits of the DDC method is

that the total net charge of the ES is conserved irrespective of the dividing scheme. Later, DC-HF, DC-MP2, and DC-CC were proposed.[29−31]

Another type of DC method is the energy based divide-and-conquer (EDC) method which is based on the many-body expansions of energy.[14−24] In 1999, Kitaura et al. first proposed the FMO method[14a] and then many EDC methods have been developed such as the molecular fractionation with conjugated caps (MFCC) scheme by Zhang and Zhang,[17] the systematic fragmentation method (SFM) by Collins and Deev,[24] the isodesmic fragmentation method (IFM) by Bettens et al.[21] and the electrostatically embedded many-body expansion (EEMB) by Truhlar et al..[22] The basic idea of these EDC methods is to divide the ES into small subsystems, and the energy of the ES is then expanded into a sum of many-body correction terms where the 2-, 3-, and m-body correction terms are due to the coupling between subsystems. For covalently bonded systems, the broken bonds during the dividing are generally capped by complementary atoms/groups and their effects can be eliminated gradually by including higher-level many-body corrections. It is worth mentioning that although some of the EDC methods were applied to noncovalently bonded molecule clusters when initially proposed, in principle they can be

applied to covalently bonded systems as well. Some EDC methods have been successfully used to handle covalently bonded systems such as proteins.[23,24b]

Comparing these two kinds of methods, the effect of buffer in the DDC method is similar to that of many-body interaction in the EDC method. If a buffer of a subsystem in the DDC scheme is composed of k subsystems in the EDC scheme, then the buffer plays the same role as the many-body interactions. That is to say, the effect of buffer to the subsystem can be simplified from k-body to 2-body or 3-body interactions in the EDC method. Thus the max number of atoms in the many-body corrections is reduced and the computational cost, especially for the high order wave function methods such as MP2 or coupled cluster methods, could be dramatically reduced. However, in the EDC method, there is no charge conservation scheme, which may affect the accuracy of the computed total energy.

In the DC method, the whole system is divided into subsystems and the chemical bonds may be broken. The broken bonds are then capped by cap atoms/groups (usually hydrogen atoms) along the direction of the original bonds. In fact, a subsystem within the large system is different from the isolated one with cap atoms/groups. The cap atoms/groups may induce different charge transfer (or population) to/from the subsystem to the rest of the ES and ultimately cause the sum of the charges of the subsystems to differ from the real total charge of the ES, that is, leading to the nonconservation of the total charge. According to DFT, energy is a functional of the electron density whose integrity is the total number of electrons. The sum of the total number of electrons and nuclear charges is the total charge. This means that if the deviation of the total charge of the ES is considerable, the effect on the total energy could not be ignored. Therefore, to obtain accurate energy of the ES, the error caused by the nonconservation of the charge should be corrected. Of course, the charge redistribution or the change of the density gradient in the subsystem induced by the cap atoms/groups may also affect the total energy, but compared with the effect of the total density change or the total charge change, usually its effect is smaller and can be ignored. This can also be inferred from the formula of DFT where energy is primarily related to density and the contribution of the generalized gradient approximation (GGA) terms to the total energy is usually much smaller than that of the local density approximation (LDA) terms in most cases.[4−10]

Theoretically, the effect of the nonconservation of the total charge on the total energy can be remedied by solving the Kohn−Sham equation of the ES. However, this solution may make the whole DC scheme useless since the ES has to be computed as a whole. Therefore it is necessary to seek a cheaper solution. In this paper we focus on the effect of the nonconservation of the total charge which is a general problem when applying the EDC method to covalently bonded systems, and therefore the simplest EDC scheme is selected: The total energy is directly expanded into m-body correction terms. All the 2-body corrections are computed while only connected m-body ($m > 2$) corrections are kept, where "connected" means the m-body subsystem is interconnected by chemical bonds. Then the total charge and the total energy are computed and their relationship is then analyzed. It should be noted that to accelerate the convergence of the EDC method, when computing the energy of a m-body subsystem some EDC schemes such as the EEMB method compute the energy of the m-body subsystem embedded in the field of the point charges at the nuclear position of the rest of the ES system except this m-body subsystem. In the present study, the embedded scheme of the EEMB method was not adopted. However, the electrostatic interaction between the subsystem and its surrounding environment is partially covered in the m-body correction terms.

The organization of this paper is as follows: In Section 2, the theoretical foundation of our method is presented. Numerical applications of the method to three glycine oligomers and two conjugated systems are presented in Section 3. Concluding remarks are given in Section 4.

2. METHODOLOGY

According to DFT, the total energy of a system can be expressed as a functional of density:

$$E = E(\rho) = T(\rho) + \int \rho(r)\, v(r)\, dr + J(\rho) + E_{cx}(\rho) \tag{1}$$

While the density at point i in space changes by $\Delta\rho$, the energy E will vary by ΔE_i. ΔE_i can be expanded by Taylor expansion:

$$\Delta E_i = E(\rho_i + \Delta\rho) - E(\rho_i) = a_i(\Delta\rho) + b_i(\Delta\rho)^2 + \ldots \tag{2}$$

where a_i and b_i are the expansion coefficients of $\Delta\rho$.

For the ground state of a system, if a small amount of density $\Delta\rho$ is transferred from point j to i, the energy change ΔE is

$$\Delta E_i = a_i(\Delta\rho) + b_i(\Delta\rho)^2 + \ldots$$

$$\Delta E_j = a_j(-\Delta\rho) + b_j(-\Delta\rho)^2 + \ldots$$

$$\Delta E = \Delta E_i + \Delta E_j = (a_i - a_j)\Delta\rho + (b_i + b_j)(\Delta\rho)^2 + \ldots \tag{3}$$

If the transfer is inversed, that is, from point i to point j, the corresponding energy change $\Delta E'$ is

$$\Delta E' = \Delta E_i' + \Delta E_j' = (a_i - a_j)(-\Delta\rho) + (b_i + b_j)(-\Delta\rho)^2 + \ldots \tag{4}$$

According to DFT, any density change will raise the ground-state energy. Therefore,

$$(a_i - a_j)(\Delta\rho) + (b_i + b_j)(\Delta\rho)^2 + \ldots \geq 0$$

$$(a_i - a_j)(-\Delta\rho) + (b_i + b_j)(\Delta\rho)^2 + \ldots \geq 0$$

Multiply these two inequations, and the following inequation is obtained:

$$-(a_i - a_j)^2(\Delta\rho)^2 + (b_i + b_j)^2(\Delta\rho)^4 - \ldots \geq 0 \tag{5}$$

Since higher order (>2) terms of $\Delta\rho$ can be ignored when $\Delta\rho$ approaches zero, to satisfy this inequality, the coefficient $(a_i - a_j)$ of the first term must be zero. That is to say, the gradient of energy to density is always identical at any point in a ground-state system:

$$a_j \equiv a_i \equiv \cdots \equiv a \tag{6}$$

If the densities at many points are changed by a little amount and the corresponding total charge and total energy changes are ΔQ and ΔE, respectively, then

$$\Delta E = \sum \Delta E_i = \sum a\Delta\rho_i + \sum b_i(\Delta\rho_i)^2 + ...$$

$$\Delta E = a\Delta Q \sum b_i(\Delta\rho_i)^2 + ... \approx a\Delta Q \qquad (7)$$

That is to say ΔE is linear with respect to ΔQ if the density change is small enough so that the second order terms can be neglected. Therefore, if ΔQ and ΔE can be obtained, the correction to the total energy of the ES can be obtained by linear extrapolation of E with respect to ΔQ.

To compute the total energy and charge of the ES, a simple straightforward EDC scheme is adopted to compute the corresponding properties of the subsystems. Figure 1 illustrates
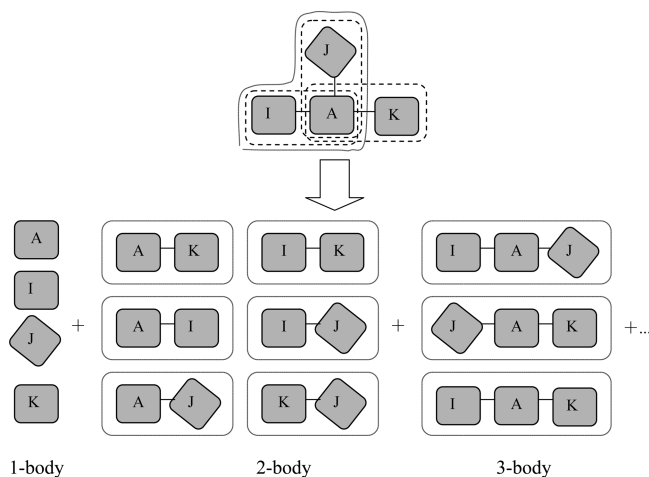
Figure 1. Dividing scheme of the EDC method adopted in the present study.

the scheme where an ES is divided into four subsystems A, I, J, and K where lines between the subsystems represent the chemical bonds between them. The 2-body subsystems include all possible combinations of the A, I, J, and K subsystems (a total of 6), while the 3-body subsystems are composed of 3 subsystems connected by chemical bonds (a total of 3). All the subsystems are capped by H atoms if there are chemical bonds broken when dividing ES into subsystems. It should be mentioned that although the energy of the capped H atoms cannot be removed if only 1-body terms are included in eqs 8−10 (see below), it is removed if higher-order correction terms are included. In practice, the EDC scheme including only 1-body terms is seldom used due to its low accuracy. In the following paragraph, the energy and charge expansion of the EDC scheme are presented in detail.

First, the energy of a subsystem A embedded in the ES (many-body corrected energy up to m-body correction level, E_A^m) is obtained by adding 2-, 3-, ..., and m-body corrections to the energy of the isolated subsystem A (E_A) according to eqs 8−10:

$$E_A^m = E_A + \sum_I^M dE_{AI}^2 + \sum_I^M \sum_J^M dE_{AIJ}^3$$

$$+ \sum_I^M \sum_J^M \sum_K^M dE_{AIJK}^4 + ... \qquad (8)$$

$$dE_{AI}^2 = (E_{AI} - E_A - E_I)/2 \qquad (9)$$

$$dE_{AIJ}^3 = (E_{AIJ} - E_{AI} - E_{AJ} - E_{IJ} + E_A + E_I + E_J)/3 \qquad (10)$$

...

where M is the total number of subsystems, E_{AI}, E_{AIJ}, and E_{AIJK} are the energies of 2-body, 3-body, and 4-body subsystems AI, AIJ, and $AIJK$, respectively, and dE_{AI}^2, dE_{AIJ}^3, and dE_{AIJK}^4 are the 2-body, 3-body, and 4-body correction terms, respectively.[22−24] The total energy of the ES up to m-body correction level (E_{ES}^m) is then obtained as

$$E_{ES}^m = \sum_A^M E_A^m \qquad (11)$$

To obtain E_{ES}^m in the EDC method, the energies of all the n-body ($1 \leq n \leq m$) subsystems selected out of a total M subsystems (a total number of $\sum_{n=1}^m C_M^n$) must be explicitly calculated by the chosen theoretical methods. If a large m is chosen for a large system, then the computational costs would be enormous. In the present study, to reduce computational costs, not all $\sum_{n=1}^m C_M^n$ n-body subsystems are computed. For $n \geq 3$ only those n-body subsystems connected by chemical bonds are computed. It should be noted that in the present EDC scheme when computing the energy of an n-body subsystem it is not embedded in the field of the point charges of the $M - n$ subsystems. Keeping all the 2-body terms is important especially for those systems in which medium- and long-range nonbonded interactions such as hydrogen bonds and long-range electrostatic interactions are too big to be ignored.

Similarly, the total charge of the ES can also be obtained in a similar way

$$Q_A^m(A) = Q_A(A) + \sum_I^M dQ_{AI}^2(A) + \sum_I^M \sum_J^M dQ_{AIJ}^3(A) + ... \qquad (12)$$

$$dQ_{AI}^2(A) = Q_{AI}(A) - Q_A(A) \qquad (13)$$

$$dQ_{AIJ}^3(A) = Q_{AIJ}(A) - Q_{AI}(A) - Q_{AJ}(A) + Q_A(A) \qquad (14)$$

...

$$Q_{ES}^m = \sum_A^M Q_A^m(A) \qquad (15)$$

where $Q_A(A)$ is the total charge of all atoms in subsystem A excluding the cap atoms, $Q_{AI}(A)$ and $Q_{AIJ}(A)$ are the total charges of all atoms belonging to subsystem A in the 2-body AI and 3-body AIJ subsystems, and $dQ_{AI}^2(A)$ and $dQ_{AIJ}^3(A)$ are the 2-body and 3-body charge corrections to the charge of subsystem A, respectively.

In many-body expansions, if all M-body correction terms are computed, E_{ES}^M and Q_{ES}^M obtained according to eqs 11 and 15 are exactly the same as the total energy and charge of the ES computed as a whole (hereafter it is referred to as the real system), respectively. If the many-body correction terms are truncated to m-body ($m < M$) level, then the effects of cap atoms and many-body interactions between subsystems cannot be fully canceled out. Charge redistribution introduced by the cap atoms cannot be fully canceled out as well. A subsystem embedded in $m - 1$ surrounding subsystems is also different from that embedded in the $M - 1$ surrounding subsystems of

the real system. Therefore deviation of E_{ES}^m and Q_{ES}^m from those of the real system, E_{ES}^M and Q_{ES}^M, respectively, is expected. Q_{ES}^m is then not conserved. As discussed above, the deviation of E_{ES}^m from E_{ES}^M is related to the deviation of Q_{ES}^m from Q_{ES}^M. If higher order terms in eq 7 can be neglected, eq 7 indicates that $E_{ES}^m - E_{ES}^M$ is linearly related to $Q_{ES}^m - Q_{ES}^M$. Therefore, E_{ES}^M can be obtained by a linear extrapolation:

$$E_{ES}^M = E_{ES}^m - \frac{E_{ES}^m - E_{ES}^{m-1}}{Q_{ES}^m - Q_{ES}^{m-1}}(Q_{ES}^m - Q_{ES}^M) \qquad (16)$$

Analogously, if higher order terms in eq 7 can be neglected, other properties of the system such as atomic forces can also be approximately obtained with a similar extrapolation scheme. Taking the atomic force $F_{xi}(ES)$ on the ith atom as an example, it is the first order derivative of energy with respect to the coordinate x_i:

$$F_{xi}(ES) = -\frac{\partial E_{ES}^M}{\partial x_i} = -\frac{\partial E_{ES}^m}{\partial x_i} + \frac{\partial a}{\partial x_i}(Q_{ES}^m - Q_{ES}^M)$$
$$+ a\frac{\partial Q_{ES}^m}{\partial x_i} \qquad (17)$$

where a $[= (E_{ES}^m - Q_{ES}^{m-1})/(Q_{ES}^m - Q_{ES}^{m-1})]$ is the linear coefficient in eq 16.

In the above equation, the third term $\partial Q_{ES}^m/\partial x_i$ can be neglected based on the following two arguments: The first is that Q_{ES}^m is very close to Q_{ES}^M. As Q_{ES}^M is a constant value, Q_{ES}^m also hardly change especially at high m-body correction level. The second is that the change in Q_{ES}^m is mainly caused by the charge transfer between cap atoms and subsystem. In addition, the charge transfer is mostly determined by the atoms in the subsystem directly bonding to the cap atoms. Since the broken bonds are often far away from the ith atom especially at high m-body correction level (it should be noted that even if the ith atom in subsystem A is connected to a cap atom, it is not in a m-body ($m > 1$) subsystem), changing x_i of the ith atom in subsystem A will hardly influence the charge of the cap atom. In the test (Section 3), it is also shown that the change of the total charge with respect to geometry change is very small. Therefore, neglecting $\partial Q_{ES}^m/\partial x_i$ will only introduce a very small error in the calculated force. Thus,

$$F_{xi}(ES) \approx -\frac{\partial E_{ES}^M}{\partial x_i} + \frac{\partial a}{\partial x_i}(Q_{ES}^m - Q_{ES}^M)$$
$$= F_{xi}^m(ES) + \frac{\partial a}{\partial x_i}(Q_{ES}^m - Q_{ES}^M) \qquad (18)$$

where $F_{xi}^m(ES) = -(\partial E_{ES}^m/\partial x_i)$ is the x component of the force acting on the ith atom given by the EDC method and it can be deduced from the differential of E_{ES}^m as expressed by eq 11. Therefore, $F_{xi}(ES)$ can also be obtained by a linear extrapolation as the total energy E_{ES}^M:

$$F_{xi}(ES) = F_{xi}^m(ES) - \frac{F_{xi}^m(ES) - F_{xi}^{m-1}(ES)}{Q_{ES}^m - Q_{ES}^{m-1}}(Q_{ES}^m - Q_{ES}^M) \qquad (19)$$

As geometry optimization and molecular dynamics simulation are sensitive to the consistency of the force and the grad of the energy, Q_{ES}^m and Q_{ES}^{m-1} are suggested to be constant values which can be obtained at the first point of a geometry optimization or a molecular dynamics simulation. Then in eq

17 there is no variable introduced except energy and then atomic force (eq 19) is strictly equal to the negative grad of the energy (eq 16). This scheme is called as E-EDC(FixQ) in this paper. However, this approximation can introduce some errors in both the extrapolated energy and force.

Of course, neglecting the higher-level terms in eq 7 may not always work. In the practice since the computational costs for $m > 3$ are quite expensive, a two-point linear extrapolation with $m = 3$ using eq 17 is preferred. The reliability of eq 16 can be verified by performing an additional 4-body EDC calculation but with a smaller basis set. If the linearity in E_{ES}^m with respect to $Q_{ES}^m - Q_{ES}^M$ is not good enough, larger subsystem size is suggested so that the higher order terms in eq 7 can be neglected.

3. TEST

The present method is assessed by five tests. The first is a linear 24 glycine oligomer (Gly24) for which the geometry is not optimized and is available from the Supporting Information. Gly24 is divided into 24 subsystems, as Figure 2 shows. The
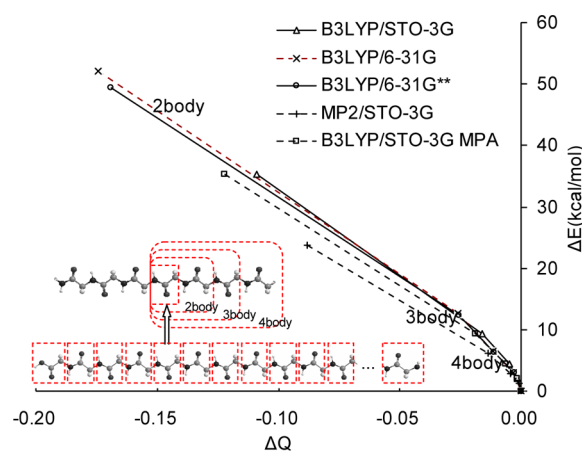
Figure 2. Relationship between the change of the total energy (ΔE) and the total net charge (ΔQ) for the Gly24 molecule computed by the EDC method. The population analysis method is the NPA unless otherwise noted.

theoretical methods used for the single-point-energy calculations are B3LYP and MP2, and the basis sets are STO-3G, 6-31G, and 6-31G**. For the MP2 method, due to the high computational cost of the real system, only the STO-3G basis set is used since the calculations of the real system using the other two basis sets are all out of memory or disk space. Population analysis is performed using the natural population analysis (NPA) and the Mulliken population analysis (MPA) methods.[32,33] All the energies obtained by linear extrapolation employing eq 16 are compared to those from the conventional B3LYP or MP2 calculations for the real system. The plots of $\Delta E(= E_{ES}^m - E_{ES}^M)$ vs $\Delta Q(= Q_{ES}^m - Q_{ES}^M)$ are presented in Figure 2 where $Q_{ES}^M = 0$ since the Gly24 molecule is neutral. All the computations are performed with the Gaussian 09 software package.[34]

As shown in Figure 2, ΔE is almost linear to ΔQ whatever the basis set is. This confirmed our idea proposed in Section 2. Using eq 16, the extrapolated total energy of the ES is obtained and the results are listed in Table 1. The total energy computed by the EDC method at the 2-body, 3-body, 4-body, and 5-body correction levels are also listed in Table 1. It can be inferred

Table 1. Absolute Errors (kcal/mol) of the E-EDC Method and the EDC Method in Calculating Total Energy[a]

m-body/ method	B3LYP STO-3G	B3LYP 6-31G	B3LYP 6-31G**	MP2 STO-3G	B3LYP STO-3G/MPA
		E-EDC (eq 16)			
3	4.81	5.70	5.68	2.89	4.59
4	2.17	1.61	1.81	1.46	1.87
		EDC (eqs 8−11)			
2	35.30	52.13	49.36	23.76	35.30
3	9.32	12.61	12.35	6.11	9.32
4	4.37	6.45	6.31	2.93	4.37
5	2.01	3.03	3.00	1.33	2.01

[a]The NPA method is used unless otherwise noted.

from the results in Table 1 that the errors in the total energies given by the E-EDC method are significantly lower than those computed by the EDC method. For example, for the E-EDC method the errors of the total energies at the 3-body correction level are reduced to less than 5.7 kcal/mol, which decrease by 50%−70% compared to that of the EDC method with the same many-body correction level. For the EDC method, to reach a similar accuracy, at least a 4-body correction level is needed. If a 4-body correction level is used for the E-EDC method, the error is usually similar or less than that of the EDC method at the 5-body correction level. Therefore, the accuracy of the E-EDC method is usually comparable or even slightly higher than that of the EDC method at $(m + 1)$-body correction level. It should be noted that the E-EDC and EDC methods have almost the same computational cost at the same correction level since all the many-body correction terms up to m-body are needed to be computed to apply eqs 8−11.

We have also tested the choice of population analysis method on the E-EDC method. The results presented in Table 1 (the second and sixth column for the NPA and MPA population analysis method at the B3LYP/STO-3G level, respectively) and Figure 2 indicate that the two population analysis methods give similar results. Therefore in the following tests only the NPA method is adopted for the population analysis.

The second test is a branched oligomer of 32 glycine residues (Gly32) of which two conformers are considered. One is uncurled and the other is curled as shown in Figure 3. The Gly32 molecule is divided into 32 subsystems. The method used is B3LYP/6-31G. The results are shown in the left panel of Figure 3.

From Figure 3 it can be seen that the linearity of ΔE to ΔQ is kept for the uncurled conformer of Gly32. This result is similar to that of the linear Gly24 molecule. However, the linearity fails for the curled conformer of Gly32, since the 2-body corrected total energy (E_{ES}^2) is almost equal to the 3-body corrected total energy (E_{ES}^3) which leads to almost zero linear coefficient in eq 16, and the correction to E_{ES}^m does not work at all. Therefore, the error of the extrapolated energy is still as large as 9 kcal/mol. A close examination of the results indicates that the failure is mainly due to the hydrogen bonds between the capped H atoms to the oxygen atoms of neighboring subsystems, especially in the curled conformation. The hydrogen bonds have larger effect on 2-body correction than on 3-body correction since the number of the capped H atoms is reduced in the latter case which leads to the failure of the linearity. This problem can be solved simply by increasing the size of the subsystem to decrease the number of the unwanted hydrogen bonds and to enlarge the distance between the capped H atoms and the oxygen atoms. The Gly32 molecule is then divided into 16 subsystems, and each subsystem contains 2 glycine residues. The results are shown in the right panel of Figure 3. It can be seen that the linearity of ΔE to ΔQ is improved and the error of the extrapolated energy at the 3-body correction level decreases to 1.5 kcal/mol. It should be noted that increasing the subsystem size can not only improve the linearity of ΔE to ΔQ but also decrease the error of the total energy given by the EDC method.

The above tested molecules are all neutral molecules. Then the E-EDC method is further tested on charged systems. For the third test three charged Gly24 are tested: one with a protonated terminal amino group (GLY24n), one with a deprotonated carboxylic group (GLY24p), and the third being a zwitterion with one protonated terminal amino group on one end of the chain and one deprotonated carboxylic group on the other end of the chain (GLY24np). All the tested molecules are divided into 24 subsystems. The computation method is B3LYP/6-31G. The population analysis method is NPA. The results are shown in the left panel of Figure 4. From these results, it can be seen that ΔE has a similar linear relationship with ΔQ as that of the neutral Gly24. The error of the extrapolated energy is also greatly reduced.

Usually, the EDC method is less accurate for conjugated system. The forth test is on conjugated polyacetylenes with a carboxyl terminal group, one is neutral ($CH_2{=}CH{-}(CH{=}$
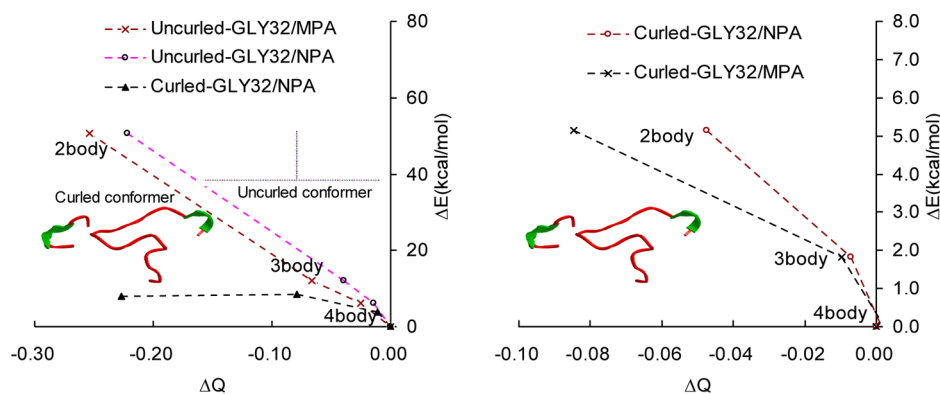
Figure 3. Relationship between the change of the total energy (ΔE) and the total net charge (ΔQ) of Gly32. In the left panel, Gly32 is divided into 32 subsystems and each subsystem contains one glycine residue, while in the right panel, it is divided into 16 subsystems and each subsystem contains two glycine residues. The method used is B3LYP/6-31G.
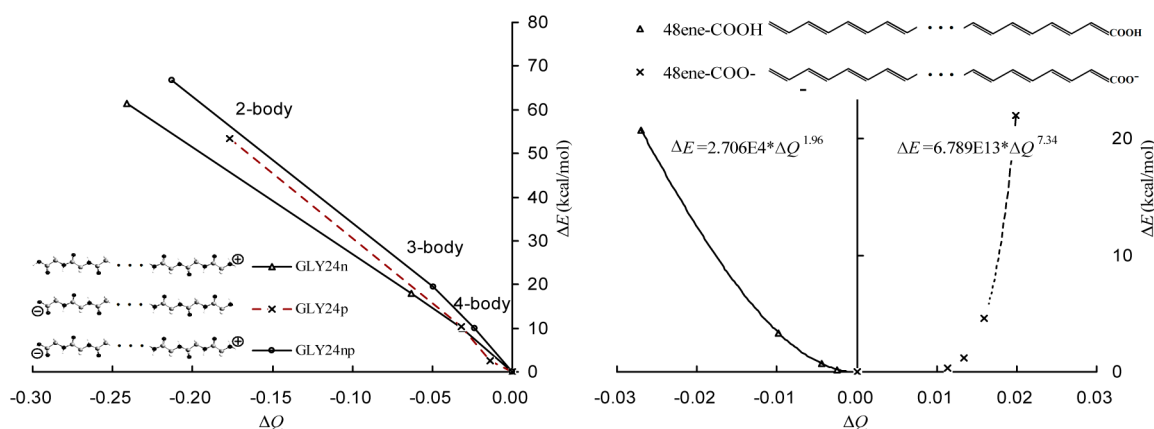
Figure 4. Relationship between the change of the total energy (ΔE) and the total net charge (ΔQ) of three charged GLY24 molecules (left panel) and two polyacetylene molecules (right panel). The method used is B3LYP/6-31G.
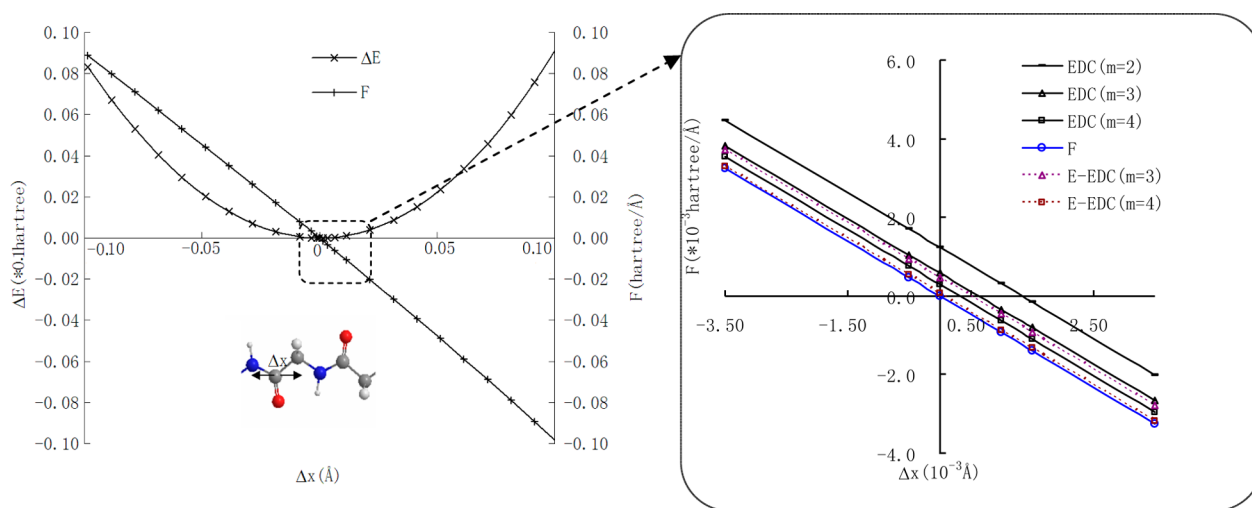

Figure 5. Atomic force of the C atom in the 12th subsystem of the Gly24 molecule calculated by different methods.

CH)$_{47}$—COOH, abbreviated as 48ene-COOH), and the other one is negatively charged (CH$_2$=CH—(CH=CH)$_{47}$—COO$^-$, abbreviated as 48ene-COO$^-$). The molecular structure is shown in the right panel of Figure 4. The polyacetylene molecules are divided into 24 subsystems, and each subsystem consists of two —CH=CH— subunits. The computation method is B3LYP/6-31G. The results show that the linearity of the $\Delta Q - \Delta E$ relationship cannot be kept. From Figure 4, it can be inferred that the higher order of the charge density cannot be neglected. When a conjugated system is computed as a real system, a small disturbance in the charge density at one point can be redistributed to the whole system and the gradient of density is thus close to zero. On the other hand, for computing subsystems using the EDC method, the charge is localized to the subsystem containing the charged group. Therefore the gradient of the density in the lower level EDC treatment is much larger than that in the higher level EDC treatment since in the latter the larger conjugated subunit is more efficient in redistributing charge. As a result, the hypothesis that higher order terms in eq 7 can be neglected is no longer valid. From Figure 4 it can be seen that for a conjugated system the dependence of ΔE on ΔQ for the 48ene-CHOOH molecule is close to second order while for the charged 48ene-COO$^-$ it is close to seventh order. In order to keep the simplicity of the E-EDC scheme, nonlinear

extrapolation schemes are not recommended, although in principle there is no technical difficulties in applying such schemes.

To demonstrate the power of the E-EDC method, the atomic forces of the atoms in Gly24 are also computed by employing eq 19 where NPA charge is used. The C atom in the 12th subsystem is selected as an example since its force has the maximum error in all the atoms of Gly24 if no extrapolation is applied. The x-Cartesian coordinate of the C atom is changed by −0.1 Å to 0.1 Å from its original position, and the corresponding changes in the forces computed at different many-body correction levels using the EDC method and the E-EDC method (eq 19) are shown in Figure 5. As the force error is very small and all the curves are overlapped, the figure is zoomed in near zero of the abscissa (Δx). The RMS errors of the force for all the atoms computed by different methods are shown in Figure 6. It can be seen from Figures 5 and 6 that the extrapolation also decreases the error in the computed forces by 40−70%. The extrapolation of atomic force is slightly less prominent than that of energy. This may be due to the neglecting of $\partial Q_{ES}^m / \partial x_i$ in eq 17. Another result should be noted is that the results of E-EDC when Q_{ES}^m is fixed are very similar to the results of the extrapolation scheme neglecting $\partial Q_{ES}^m / \partial x_i$ (eq 19). The difference in the atomic forces obtained by the two different methods to extrapolate force is less than 10^{-6} Hartree/
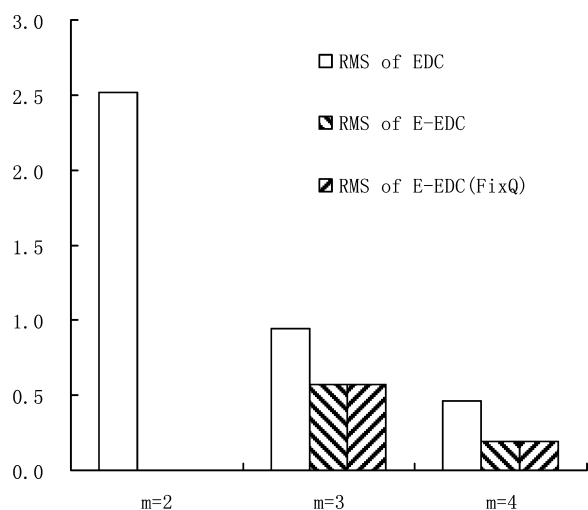
Figure 6. RMS error (mHartree/Å) of the atomic force of the Gly24 molecule calculated by different methods.

Å (see also Figure 6). The results indicate that fixing Q_{ES}^m does not introduce much error in the computed atomic forces but keeps the strict consistency between the atomic force and the energy grad.

To verify if Q_{ES}^m is almost constant when geometry changes (Δx), Q_{ES}^m at different x-Cartesian coordinates computed by the EDC method at different m-body correction levels are listed in Table 2. From the results in Table 2 it can be concluded that

Table 2. Change of Total Charge Q_{ES}^m Computed by the EDC Method When Changing the x-Cartesian Coordinate of the C Atom in the 12th Subsystem of the Gly24 Molecule[a]

m	Q_{ES}^m (au) $\Delta x = -0.1$ Å	Q_{ES}^m (au) $\Delta x = 0.0$ Å	$\partial Q_{ES}^m / \Delta \Delta x$ (au/Å)
2	−0.17495	−0.17448	0.0047
3	−0.02597	−0.02599	−0.0002
4	−0.01136	−0.01143	−0.0007

[a]The computation level is B3LYP/6-31G and the NPA method are used for charge analysis.

Q_{ES}^m is almost constant when geometry changes, especially when $m > 2$: $\Delta Q_{ES}^m / \Delta \Delta x$ is −0.0002 au/Å at 3-body correction level and is −0.0007 au/Å at 4-body correction level. Thus, the approximation in eq 18 is reliable.

4. CONCLUSION

In summary, in the many-body expansion based (in this paper called energy-based) divide-and-conquer method, the chemical bonds are often broken when the whole system is divided into subsystems and complementary atoms/groups are added to cap the dangling bond. In this process, the nonconservation of charge due to the charge transfer between the cap atoms/groups and subsystem is the main source of error in the computed total energy. In the present study a linearly extrapolated version of the EDC method, that is, the E-EDC method, works very well for nonconjugated systems, while for conjugated systems a higher order extrapolation scheme is needed.

The E-EDC method requires performing two EDC calculations to obtain the total energies of the whole system at two different many-body correction levels employing eqs 8−11. The total charge of the whole system is then computed

at each many-body correction level using similar equations (eqs 12−15) as those of the total energy (eqs 8−11). It has been shown that an approximate linear relationship exists between the total energy and the total charge for nonconjugated systems. Therefore a linear extrapolation scheme employing eq 16 can be used to obtain a total energy with improved accuracy.

The test calculations on glycine oligomers show that the proposed E-EDC method can substantially improve the accuracy of the computed total energies and atomic forces for nonconjugated systems, reducing the errors in total energy by 40−70% no matter whether they are neutral or charged. For nonlocalized systems the linear extrapolation is not suitable and a higher order extrapolation scheme is needed. Further more, more investigations are needed to explore the applicability of our simple extrapolation scheme in other systems such as noncovalently bonded systems.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

Cartesian coordinates of the testing molecules and detailed results of the computed total energies, total charges, and atomic forces. This information is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail lizhenhua@fudan.edu.cn; fax +86-21-65641740 (Z.H.L.). E-mail knfan@fudan.edu.cn; Fax +86-21-65641740 (K.-N.F.).

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Pople, J. A. *Rev. Mod. Phys.* **1999**, *71*, 1267.
(2) Galli, G. *Curr. Opin. Solid State Mater. Sci.* **1996**, *1*, 864.
(3) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, 1133.
(4) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(5) Yang, W. T. *Phys. Rev. Lett.* **1991**, *66*, 1438.
(6) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
(7) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
(8) Kohn, W. *Phys. Rev. Lett.* **1996**, *76*, 3168.
(9) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(10) (a) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415. (b) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
(11) Yang, W. T.; Lee, T. S. *J. Chem. Phys.* **1995**, *103*, 5674.
(12) Warschkow, O.; Dyke, J. M.; Ellis, D. E. *J. Comput. Phys.* **1998**, *143*, 70.
(13) Gogonea, V.; Westerhoff, L. M.; Merz, K. M. *J. Chem. Phys.* **2000**, *113*, 5604.
(14) (a) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. *Chem. Phys. Lett.* **1999**, *313*, 701. (b) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2004**, *120*, 6832. (c) Pruitt, S. R.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. *J. Chem. Theory Comput.* **2010**, *6*, 1. (d) Nagata, T.; Fedorov, D. G.; Kitaura, K. *Chem. Phys. Lett.* **2012**, *544*, 87. (e) Nakata, H.; Fedorov, D. G.; Nagata, T.; Yokojima, S.; Ogata, K.; Kitaura, K.; Nakamura, S. *J. Chem. Phys.* **2012**, *137*, 044110. (f) Fedorov, D. G.; Nagata, T.; Kitaura, K. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562.

(15) Tanaka, S.; Watanabe, C.; Okiyama, Y. *Chem. Phys. Lett.* **2013**, *556*, 272.

(16) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. *PLoS one* **2012**, *7*, e41117.

(17) (a) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599. (b) Mei, Y.; Zhang, D. W.; Zhang, J. Z. H. *J. Phys. Chem. A* **2005**, *109*, 2.

(18) (a) Li, S. H.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7215. (b) Li, W.; Li, S. H. *J. Chem. Phys.* **2005**, *122*, 194109. (c) Li, W.; Fang, T.; Li, S. H. *J. Chem. Phys.* **2006**, *124*, 154102. (d) Li, W.; Li, S. H.; Jiang, Y. S. *J. Phys. Chem. A* **2007**, *111*, 2193. (e) Yang, Z.; Hua, S.; Li, S. H. *J. Phys. Chem. A* **2010**, *114*, 9253. (f) Hua, S.; Li, S. H. *ChemPhysChem* **2013**, *14*, 108.

(19) Sodt, A.; Subotnik, J. E.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 194109.

(20) Stoll, H.; Paulus, B.; Fulde, P. *J. Chem. Phys.* **2005**, *123*, 144108.

(21) (a) Bettens, R. P. A; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777. (b) Le, H. A.; Tan, H. J.; Ouyang, J. F.; Bettens, R. P. A. *J. Chem. Theory Comput.* **2012**, *8*, 469.

(22) (a) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46. (b) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342. (c) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1. (d) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33. (e) Kurbanov, E. K.; Leverentz, H. R.; Truhlar, D. G.; Amin, E. A. *J. Chem. Theory Comput.* **2012**, *8*, 1.

(23) Fedorov, D. G.; Ishimura, K.; Ishida, T.; Kitaura, K.; Pulay, P.; Nagase, S. *J. Comput. Chem.* **2007**, *28*, 1476.

(24) (a) Deev, V. A.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102. (b) Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104. (c) Collins, M. A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7744.

(25) Kohn, W. *Rev. Mod. Phys.* **1999**, *71*, 1253.

(26) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.

(27) Reuter, N.; Dejaegere, A.; Maigret, B.; Karplus, M. *J. Phys. Chem. A* **2000**, *104*, 1720.

(28) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760.

(29) Akama, T.; Kobayashi, M.; Nakai, H. *J. Comput. Chem.* **2007**, *28*, 2003.

(30) Kobayashi, M.; Imamura, Y.; Nakai, H. *J. Chem. Phys.* **2007**, *127*, 074103.

(31) (a) Kobayashi, M.; Nakai, H. *J. Chem. Phys.* **2008**, *129*, 044103. (b) Kobayashi, M.; Nakai, H. *J. Chem. Phys.* **2009**, *131*, 114108.

(32) (a) Foster, J. P.; Weinhold, F. *J. Am. Chem. Soc.* **1980**, *102*, 7211. (b) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736. (c) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735. (d) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.

(33) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.