IOP PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **20** (2008) 294211 (13pp) doi:10.1088/0953-8984/20/29/294211

Exploring the accuracy of relative molecular energies with local correlation theory

Joseph E Subotnik[1] **and Martin Head-Gordon**[2,3]

[1] School of Chemistry, Tel-Aviv University, 69978 Tel-Aviv, Israel
[2] Department of Chemistry, University of California, Berkeley, CA 94720, USA
[3] Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

E-mail: subotnik@post.harvard.edu and mhg@bastille.cchem.berkeley.edu

Received 2 February 2008, in final form 29 April 2008
Published 24 June 2008
Online at stacks.iop.org/JPhysCM/20/294211

Abstract

Local coupled-cluster singles–doubles theory (LCCSD) is a theorist's attempt to capture electron–electron correlation in a fast amount of time and with chemical accuracy. Many of the difficult computational hurdles have been navigated over the last twenty years, including how to construct a linear scaling algorithm and how to produce smooth potential energy surfaces. Nevertheless, there remains the question of just how accurate a local correlation model can be, and what are the chemical limits within which local models are largely applicable. Here, we investigate how accurately can LCCSD approximate full CCSD for cases of atomization energies, isomerization energies, conformational energies, barrier heights and electron affinities. Our conclusion is that LCCSD computes relative energies that are correct to within 1–2 kcal mol^{-1} of the CCSD energy using relatively aggressive cutoffs and over a broad range of different molecular environments—alkane isomers, dipeptide conformations, Diels–Alder transition states and electron attachment in charge delocalized systems. These findings should push the reach of local correlation applications into new research terrain, including molecules on metal cluster surfaces or perhaps even metal–molecule–metal clusters.

(Some figures in this article are in colour only in the electronic version)

1. Local CCSD: a small review

1.1. Historical overview

Over the past two decades, computational physicists and chemists by and large have succeeded in producing linear scaling algorithms, capable of capturing the electronic structure of large atomic systems at the mean-field level. Today, density functional (DFT) calculations are routinely done on systems with more than a thousand atoms, yielding valuable information about the band structure of large systems [1]. Nevertheless, the problem of measuring electron–electron correlation remains a daunting problem. On the one hand, if one seeks to construct the formal wavefunction for the ground state of a molecule, the computational problem scales exponentially with the number of electrons [2] and converges very slowly with basis set size, making the complete solution (full configuration interaction [CI]) unfeasible for systems with more than 10 electrons or so. On the other hand, while DFT formally solves the problem of energy levels for interacting electrons around a nuclear potential quickly and without huge basis set effects, because the exact exchange–correlation potential is unknown, DFT cannot treat highly correlated systems, it cannot be systematically improved and it does not easily yield information about magnetic field interactions. Our purpose in this paper is to review and assess the accuracy of one wavefunction-based method for overcoming the problem of exponential scaling and quickly measuring electron–electron correlation: local coupled-cluster singles–doubles theory (LCCSD).

Just as for large mean-field calculations, the basic idea of local coupled-cluster theory (CCSD) [3–11] is to invoke the locality of individual electronic orbitals in order to achieve

a huge speed-up in computational time within the context of a CCSD calculation. The coupled-cluster approach [12–15] towards calculating electronic correlation is to make the ansatz that the n-electron fully correlated ground state has the following form:

$$\Psi_{\mathrm{CCSD}} = e^{\hat{\mathbf{T}}}\Psi_0. \quad (1)$$

Here, Ψ_0 is a single Slater determinant, often the Hartree–Fock (HF) ground state, and $\hat{T}$ is an excitation operator accounting for electronic correlation. We now denote the occupied orbitals of the HF or mean-field calculation as $ijkl$ and the virtual orbitals as $abcd$, where the virtual orbitals are what remain of our original basis after we have projected out the occupied basis. Then, for the case of coupled-cluster singles–doubles (CCSD) calculation, $\hat{T}$ assumes the form $\hat{T} = \sum_{ia} t_i^a a_a^\dagger a_i + \sum_{ijab} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_i a_j$. At low order, the CCSD ansatz is

$$\Psi_{\mathrm{CCSD}} = \Psi_0 + \sum_{ia} t_{ia}\Psi_i^a + \sum_{ijab}\left(t_i^a t_j^b + t_{ij}^{ab}\right)\Psi_{ij}^{ab} + \cdots \quad (2)$$

where Ψ_i^a is the ground state Ψ_0 with occupied i replaced by virtual a, and so forth. In order to complete the calculation and solve for the electronic correlation, one plugs the CCSD ansatz (equation (2)) into the Schrödinger equation $\hat{H}\Psi = E\Psi$ and one effectively transforms the problem of electron–electron correlation into a nonlinear algebraic equation for t_i^a and t_{ij}^{ab}, the so-called t amplitudes. One needs only solve for the t amplitudes in order completely characterize the electronic correlations in the ground state (up to the doubles level). Unfortunately, when one works through the algebra [15], one finds that the CCSD algorithm scales as N^6 where N is the size of the basis set. One is forced to compute many matrix products between integrals and t amplitudes (e.g. $\sum_{ef} t_{ij}^{ef}\langle ab||ef\rangle$), which lead to the problematic sextic scaling. The strength of the CCSD ansatz, however, is that the algorithm sums up correlations to infinite order and calculates correlation energies more accurately than MP2 perturbation theory [16].

Rather than give up on expensive, wavefunction-based methods for computing electron–electron correlation, local correlation theory (originally formulated by Pulay and Saebø in the 1980s [17]) circumvents the expensive scaling by limiting the excitations to local excitations $\hat{T}_{\mathrm{local}} = \sum_{ia\ \mathrm{close}} t_i^a a_a^\dagger a_i + \sum_{ijab\ \mathrm{close}} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_i a_j$. For ia not close together, one sets $t_i^a = 0$ and the same for t_{ij}^{ab}. By limiting the number of variables (t_i^a and t_{ij}^{ab}) to a linear number (the so-called *domains*), one reduces the dimensionality of the problem so that one can more easily and quickly solve the electronic problem. In the 1990s, Friesner [18] and Carter [19] built upon the ideas of Saebø and Pulay by recognizing that one could exploit locality not just in limiting the number of t-amplitudes variables, but also in computing the necessary Hamiltonian matrix elements. For matrix products of two-electron integrals with other tensors, Friesner and Carter invoked the pseudospectral approach for calculating two-electron integrals and gained even more computational savings than before. Subsequently, over the past decade, Schütz and Werner [4, 20–31] have applied and updated these methods to a wide variety of locally correlated electronic structure methods, including LCCSD, achieving a linear scaling algorithm in certain regimes. According to their approach, one divides up the amplitudes into strong, moderate, weak and very weak amplitudes and treats the different amplitudes at different theoretical levels. By limiting the number of strong amplitudes, one can achieve enormous computational savings. In figure 1, we plot the CPU time for running alkane calculations according to our version of LCCSD [11].

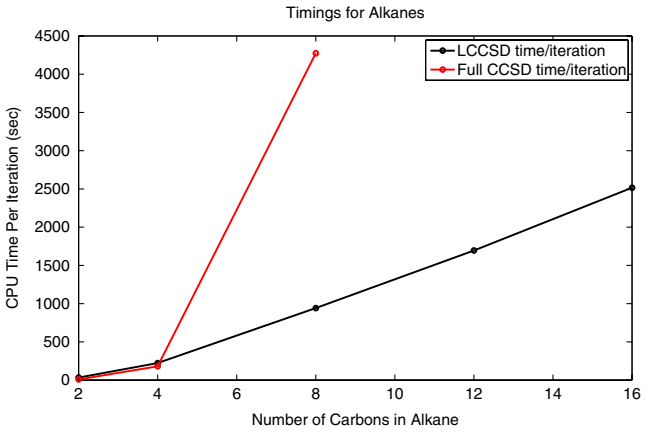

Figure 1. Computational CPU time for an LCCSD calculation (per iteration) for an alkane in a Gaussian cc-pVDZ basis, calculated according to the algorithm in [11].

Finally, once one has made all necessary local approximations for constructing a locally correlated electronic wavefunction at one fixed nuclear geometry, there remains the question of how, if possible, one can patch such wavefunctions together for different nuclear geometries. A straightforward analysis shows that, without any additional work, the local correlation approach of Pulay, Saebø, Schütz, Werner and others yields discontinuous potential energy surfaces [32]. By choosing which t_i^a and t_{ij}^{ab} to compute explicitly at which level of theory, the Pulay–Saebø–Schütz–Werner algorithm chooses coordinates to describe electronic correlation which are specifically designed for one particular choice of nuclear geometry. When the nuclei in the molecule move, the local domains of electronic interaction can change (i.e. the domains are variable), and when this occurs the local correlation algorithm often produces discontinuities in the electronic energy which cannot be smoothly patched together. As such, LCCSD as traditionally formulated is not reliable for complicated, multidimensional geometrical optimizations nor can it be used for Born–Oppenheimer dynamics, even though case-by-case solutions are sometimes possible [33] over certain regions of configuration space. This problem of smooth PES's is non-trivial to solve. In fact, in contrast to the standard Pulay variable-domain algorithms, alternative fixed-domain approaches within local correlation theory have been designed over the past ten years by Head-Gordon *et al* (e.g. DIM [34, 35], TRIM [36, 37]) specifically in order to avoid this problem of discontinuous PES's. However, when constructing smooth, well-defined, fixed-domain local correlation algorithms, redundant orbitals have been employed in the past, which slows down the approach in general.

Over the last three years, the authors have suggested and demonstrated an alternative approach to LCCSD theory from the standard Pulay–Saebø–Schütz–Werner algorithm, and this method achieves linear scaling while not forfeiting the smoothness of the PES's [9–11]. The details of this approach will be summarized below in section 1.2.2, where we also provide plots of the correlation energy along a rotational degree of freedom. The ability to produce a fast and smooth LCCSD algorithm has been a significant step forward for local correlation theory, and in the future we expect there will be standard LCCSD geometric optimizations and, we hope, eventually Born–Oppenheimer dynamics on a locally correlated CCSD electronic surface.

1.2. Theoretical nuts and bolts

The design of an accurate and efficient local correlation algorithm requires first and foremost a criterion by which one decides which quartets of integrals $ijab$ to correlate together among all the t_{ij}^{ab} in equation (2). Second, there is the question of how to obtain smooth potential energy surfaces. Third, efficiency requires that one approximate the integrals $\langle ij||ab\rangle$ in some fashion, just as one approximates the amplitudes t_{ij}^{ab}. We now treat these ideas in turn.

1.2.1. Selection criteria. Because we want to capture only local correlation, we need a selection criteria for measuring locality. In fact, it is natural to construct a function g_{ijab} where g_{ijab} is equal to 1 when $ijab$ are deemed close together, and equal to 0 when $ijab$ are far apart. For reasons to be discussed later, we call g_{ijab} a four-electron bump function, though for the moment, it appears to be just a standard characteristic function. Several options exist for constructing g. For computational efficiency, one certainly wants to pick a selection criteria which works pairwise in terms of a two-particle bump function:

$$g_{ijab} = g_{ij}^{(1)} g_{ab}^{(2)} g_{ia}^{(3)} g_{jb}^{(4)} g_{ib}^{(5)} g_{ja}^{(6)}. \tag{3}$$

Even so, several options still exist. For instance, most obviously, one could choose to select orbital pairs (ia) by an orbital exchange integral $(ia|ia)$ [9] or the distance between centroids $(|r_i - r_A|)$ [10] or some other similar criteria. The standard Boughton–Pulay criterion [38] chooses pairings based on the contributions of localized molecular orbitals into atomic orbitals.

We have decided to select pairs of orbitals by a hybrid criteria of distance and spatial extent, which we call the *bubble* method. We select the orbital pair ia according to a criterion ϕ_{ia}, defined by

$$\sigma_i = \left\langle i\left|(\mathbf{r} - \langle i|\mathbf{r}|i\rangle)^2\right|i\right\rangle^{1/2} \tag{4}$$

$$\sigma_a = \left\langle a\left|(\mathbf{r} - \langle a|\mathbf{r}|a\rangle)^2\right|a\right\rangle^{1/2} \tag{5}$$

$$\phi_{ia} = |\mathbf{r}_i - \mathbf{r}_a| - \lambda_{ia}\left(\sigma_i + \sigma_a\right). \tag{6}$$

Here λ_{ia} is a parameter which tells us how to model the spatial extent of localized orbitals. For the present paper, we choose $\lambda_{ia} = 1$.

Accordingly, the two-electron bump function g_{ia} can be defined by some cutoff (c_*):

$$g_{ia}(\phi_{ia}) = 1 \qquad \phi_{ia} < c_* \tag{7}$$

$$g_{io}(\phi_{ia}) = 0 \qquad \phi_{ia} > c_*. \tag{8}$$

Finally, as first suggested by Pulay and Saebø, local correlation theory amounts to locally approximating the CCSD ansatz in equation (1):

$$\Psi_{\text{Full CCSD}} = e^{\hat{\mathbf{T}}}\Psi_0 \longrightarrow \Psi_{\text{LCCSD}} = e^{\mathbf{g}\cdot\hat{\mathbf{T}}}\Psi_0 \tag{9}$$

and then solving the Schrödinger equation straightforwardly to find the amplitudes $\mathbf{T}$:

$$\left\langle \Psi_{\text{excited}}\left|e^{-\mathbf{g}\cdot\mathbf{T}}\mathbf{H}e^{\mathbf{g}\cdot\mathbf{T}}\right|\Psi_o\right\rangle = 0. \tag{10}$$

Here $|\Psi_{\text{excited}}\rangle = |\Psi_i^a\rangle$ or $|\Psi_{ij}^{ab}\rangle$, which closes our equations for the t amplitudes, $\mathbf{T}$. Now, because g_{ia} is either 1 or 0 depending on the overlap of i and a, we now have a basic local correlation theory.

1.2.2. Smooth potential energy surfaces. According to the standard Pulay–Saebø–Schütz–Werner approach, the local domains of electronic interaction can change as the nuclear geometry of a molecule changes. When this occurs, the local correlation algorithm often produces discontinuities in the electronic energy. Russ and Crawford found these discontinuities to be of the order of 1–5 mHartree for small covalent systems [32] undergoing dissociation.

Over the past three years, the authors have shown that the solution to the problem of discontinuous PES's lies in 'bumping' the amplitudes in the amplitude equations rather than enforcing a strict cutoff as in equation (7) [9–11]. In so doing, one can achieve mathematically smooth potential energy surfaces while retaining the computational savings of a linear scaling local correlation algorithm. More precisely, the 'bumping' procedure is a three-step procedure. First, one must be sure that, as the molecular geometry changes, one can localize all occupied and virtual orbitals smoothly. For the occupied orbitals, smooth and unique localized orbitals can be found by employing a slight variation of the Boys localization method [10, 39, 40]. Smooth localization of the virtual orbitals is more complicated because the virtual space is rugged, but smoothly localized orbitals can be produced using tricks designed around a minimal basis separation [41].

Second, we modify the bump function, so that it no longer has discontinuities, as in equation (7), by redefining g_{ia}:

$$g_{ia}(\phi_{ia}) = 1 \qquad \phi_{ia} < c_1 \tag{11}$$

$$g_{ia}(\phi_{ia}) = \frac{1}{1 + \mathrm{e}^{-\frac{2|c_1-c_0|}{c_1-\phi_{ia}} + \frac{|c_1-c_0|}{\phi_{ia}-c_0}}} \qquad \phi_{ia} \in (c_1, c_0) \tag{12}$$

$$g_{io}(\phi_{ia}) = 0 \qquad \phi_{ia} > c_0. \tag{13}$$

Here, we choose parameters c_1, c_0 so that $|c_0 - c_1|$ is a bumping window needed for smooth PES's. We choose the following parameters for our windows:

$$c_1^{\text{strong}} = 0.58\ \text{Å} \qquad c_0^{\text{strong}} = 1.48\ \text{Å} \tag{14}$$

$$c_1^{\text{medium}} = 2.12\ \text{Å} \qquad c_0^{\text{medium}} = 2.65\ \text{Å}. \tag{15}$$

There are two parameters because, in practice, we use two bumping functions. See [11] and the appendix for more details.

Third, instead of using equation (10) as a starting point, we begin with the *full* coupled-cluster ansatz from equation (1) and plug it into the Schrödinger equation, arriving at the standard nonlinear equations for all of the t amplitudes without yet any local approximations:

$$\left\langle \Psi_{\text{excited}} | e^{-\mathbf{T(t(n))}} \mathbf{H(I(n), F(n))} e^{\mathbf{T(t(n))}} | \Psi_o \right\rangle = 0. \tag{16}$$

We now rearrange as follows:

$$\mathbf{I(n)} + \mathbf{A}^{(\mathrm{d})}\mathbf{(n)} \cdot \mathbf{t} + \mathbf{R(t, I(n), F(n))} = 0. \tag{17}$$

Between equations (16) and (17), we have pulled out the largest, zeroth-order terms in the equations. Here **n** represents nuclear coordinates, **t** represents the t amplitudes for which we must solve, **I(n)** represents integrals dependent on the nuclear coordinates, $\mathbf{A}^{(\mathrm{d})}\mathbf{(n)}$ is a diagonal matrix dependent on the nuclear coordinates and **R(t, n)** is a complicated expression, involving matrix products of integrals **I**, the Fock matrix **F** and t amplitudes **t**. We now 'bump' the amplitude equations (equation (17)) by multiplying the bump function g by the t amplitudes in two places: (i) the **t** term in the **R** term and (ii) the entire **R** term on the outside:

$$\mathbf{I(n)} + \mathbf{A}^{(\mathrm{d})}\mathbf{(n)} \cdot \mathbf{t} + \mathbf{G} \cdot \mathbf{R(G \cdot t, I(n), F(n))} = 0. \tag{18}$$

Here **G** is a diagonal matrix made up of bump functions. Effectively, we have changed the Hamiltonian to force it to admit a locally correlated wavefunction, rather than guessing a locally correlated wavefunction in the beginning. Locality is enforced because g_{ijab} has to equal 1 when $ijab$ are deemed close together, and equal 0 when $ijab$ are far apart, and g_{iajb} almost always multiplies t_{ij}^{ab}. Note, though, that we never set t_{ij}^{ab} to zero even for t_{ij}^{ab} very weak. Instead, it goes naturally to the perturbative limit:

$$\mathbf{t}_{\text{weak}} = \mathbf{A}^{(\mathrm{d})}\mathbf{(n)}^{-1} \cdot \mathbf{I(n)} \tag{19}$$

where the energetic sum of all of these contributions can be summed up exactly in one swoop. This is the reason why this approach yields smooth potential energy surfaces.

Finally, let us write explicitly the bumped and smooth local equations for one case, the MP2 example. In terms of the Fock matrix (i.e. the mean-field Hamiltonian **F**), the exact, unapproximated equation for the amplitudes is

$$\langle ij||ab\rangle + \sum_e t_{ij}^{ae} f_{be} - \sum_e t_{ij}^{be} f_{ae} - \sum_m t_{im}^{ab} f_{mj} + \sum_m t_{jm}^{ab} f_{mi} = 0. \tag{20}$$

We then separate out the largest diagonal terms and put these amplitude equations in the form of equation (17):

$$\langle ij||ab\rangle + (f_{aa} + f_{bb} - f_{ii} - f_{jj}) t_{ij}^{ab} + \left\{ \begin{array}{l} \sum_e t_{ij}^{ae} f_{be}\,(1-\delta_{be}) - \\ \sum_e t_{ij}^{be} f_{ae}\,(1-\delta_{ae}) - \\ \sum_m t_{im}^{ab} f_{mj}\,(1-\delta_{mj}) + \\ \sum_m t_{jm}^{ab} f_{mi}\,(1-\delta_{mi}) \end{array} \right\} = 0. \tag{21}$$

The local, bumped equation corresponding to equation (18) is

$$\langle ij||ab\rangle + (\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j) t_{ij}^{ab} + g_{ijab} \left\{ \begin{array}{l} \sum_e g_{ijae} t_{ij}^{ae} f_{be}\,(1-\delta_{be}) - \\ \sum_e g_{ijbe} t_{ij}^{be} f_{ae}\,(1-\delta_{ae}) - \\ \sum_m g_{imab} t_{im}^{ab} f_{mj}\,(1-\delta_{mj}) + \\ \sum_m g_{jmab} t_{jm}^{ab} g_{mi} f_{mi}\,(1-\delta_{mi}) \end{array} \right\} = 0 \tag{22}$$

where $\epsilon_i = f_{ii}$.

Using this idea, the mathematical smoothness of our LCCSD potential energy surfaces is guaranteed by the implicit function theorem [42], which says that we can invert equation (22) to find **t** smoothly as a function of **n**. Moreover, in a recent article [11], we have shown that, with our parameters, certain bond-making and bond-breaking trajectories will not admit artificial points x in configuration space and tangent directions $\vec{y}$ such that $\nabla|_x \cdot \vec{y} = 0$ or $\nabla\nabla|_x \cdot \vec{y} = \vec{0}$. In other words, in one dimension, our smoothed, local PES will have no artificial maxima, minima or inflection points: $f'(x) \neq 0$, $f'(x) \neq 0$. This defines a notion of *chemically smooth* potential energy surfaces, which should be meaningful for future energy optimizations and Born–Oppenheimer dynamics. To convince the reader, in figure 2, we draw graphs of the potential energy surface and the gradient of the surface for C–C bond rotation in propane. Here, one curve has been computed following the traditional approach of ignoring smoothness, while the other has been computed following our prescription for producing smooth PES's. The smoothed graphs are indeed smooth. The unsmoothed graphs are nearly smooth to the naked eye, though the first derivative shows clear discontinuities of the order of 0.05 kcal/mol/degree. These discontinuities will be much bigger for local approximations tighter than those we make, and there are much larger discontinuities (1–5 kcal mol^{-1}) in the energy (and, consequently, a chaotic derivative) for bond-breaking and bond-breaking events [9, 32]. In all cases of bond-breaking checked so far, however, our procedure retains the chemical smoothness of the PES [11].

1.2.3. Approximating the integrals. The final consideration in a local correlation algorithm is if and how to approximate the integrals that enter the calculation. Much tedious work suggests that, if one approximates the amplitudes, one can and, in fact, should approximate the two-electron integrals and Fock matrix elements. As a practical matter, this means that in the amplitude equations, one should replace

$$\langle ij||ab\rangle \longrightarrow g_{ijab} \langle ij||ab\rangle \tag{23}$$

$$f_{pq} \longrightarrow g_{pq} f_{pq}. \tag{24}$$

We have found that this approximation leads to both a faster and, paradoxically, more accurate algorithm. See [11] and the data in section 2.1.

1.3. Remaining doubts: the question of accuracy

Having achieved a fast and smooth algorithm for calculating energies according to a local coupled-cluster ansatz, there

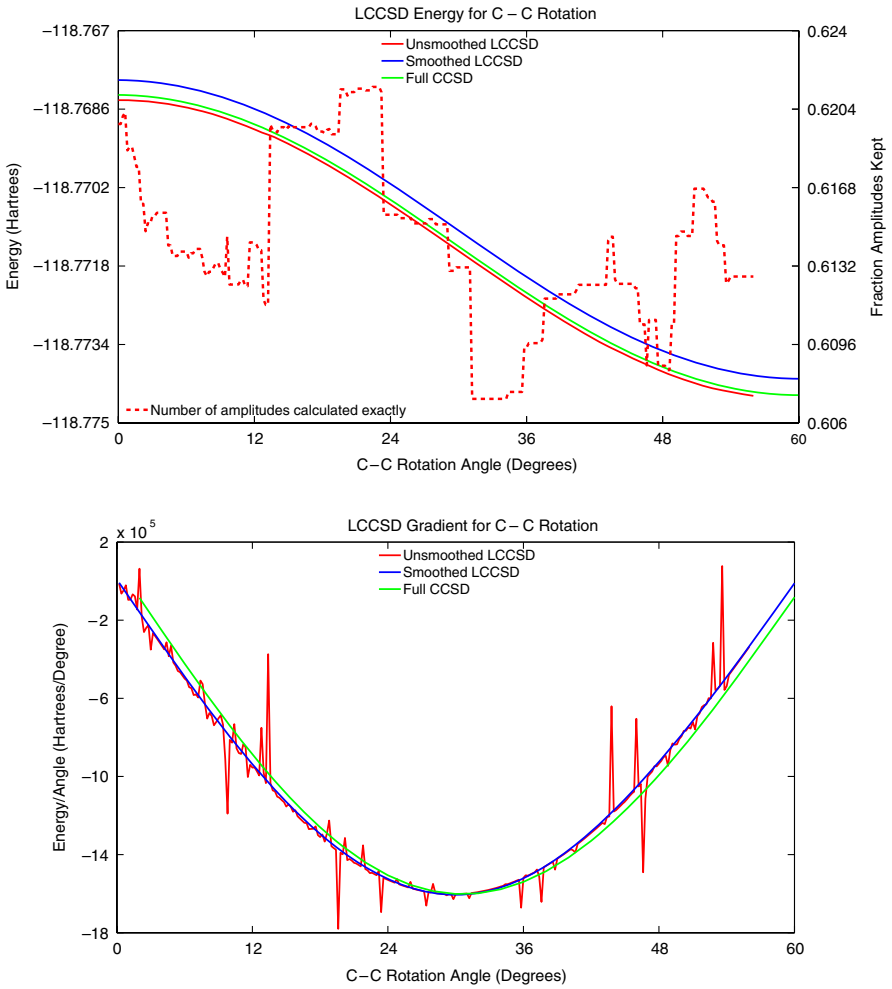

Figure 2. Smoothed and unsmoothed LCCSD curves for the rotation of propane around a C–C bond in a cc-pVDZ basis. The smoothed curves set $c_1^{\text{strong}} = 0.58$ Å and $c_0^{\text{strong}} = 1.48$ Å, and the unsmoothed curves set $c_1^{\text{strong}} = c_0^{\text{strong}} = 1.48$ Å.

remains the task of verifying the accuracy of the LCCSD method and identifying the limits of its applicability. Auer and Nooijen [43] have done extensive calculations showing that one must use domains much larger than ours if we seek complete chemical accuracy, where atomization errors of medium-sized molecules are always less than 1 kcal mol^{-1} and decay monotonically to zero as the constraints are relaxed.

Nevertheless, for our purposes, we are willing to make more severe approximations, allowing the atomization energies of molecules to have errors of 1–5 kcal mol^{-1}, provided that relative molecular energies (of many different types) are smaller than 1 kcal mol^{-1}. Our purpose in the present paper is to test our LCCSD algorithm against a background of different chemical problems, as we attempt to learn (i) how to estimate the accuracy of a local model and (ii) to discover what are the chemical limits wherein a local model is applicable.

2. Testing the accuracy of the LCCSD algorithm

In order to assess the accuracy of the LCCSD approach, the algorithm described above (and in [11]) has been implemented in the Q-Chem quantum chemistry package [44]. There is an inherent difficulty in measuring the accuracy of local methods because the severity of a local approximation can only be tested on relatively large molecules, but such molecules are often too big for a standard CCSD calculation (remembering the sixth-order scaling). Thus, in this paper, we have focused on examples with 8–10 heavy atoms for which both full and local CCSD calculations can be completed and compared. In order to assess the accuracy of our method, we have measured a variety of conformational and isomerization energies, barrier height energies and electron-attachment energies. In all of the examples tested (with the parameters from section 1.2.2), we find that the atomization energies of the LCCSD algorithm are within 6 kcal mol^{-1} of the CCSD energy (and usually under 4); moreover, relative energies of different isomers, conformers, transition states and charge states are usually correct to within 1 kcal.

2.1. Isomerization energies, conformational energies and periodic trends

In table 1, we report the energetic gap between two different conformations of alanine dipeptide, one ‘globular’ and one ‘extended’, as shown in figure 3. We have

Table 1. The ground-state energies of globular and extended conformations of alanine dipeptide treated at different levels of theory for electron–electron correlation. LCCSD version 0 does not bump two-electron integrals and is described in [10], while LCCSD version 1 does bump integrals and is our recommended algorithm from [11]. '% Ampls.' signifies the percentage of the amplitudes t_{ij}^{ab} for which $g_{ijab}^{\text{strong}} > 0$ in equation (3), so that these amplitudes must be calculated explicitly by the algorithm according to equation (9). Similarly, '% Ints.' signifies the percentage of the two-electron integrals $\langle pq||rs\rangle$ for which $g_{pqrs}^{\text{strong}} > 0$, and so, according to equation (23), these integrals must be calculated explicitly and stored. The basis is cc-pVDZ and the molecules have been optimized at the RIMP2 level of theory [45].

Dipeptide	Full MP2 (Hartree)	Full CCSD (Hartree)	Local CCSD Vers. 0 (Hartree)	Error ($E_{\text{LCCSD}} - E_{\text{CCSD}}$) (kcal mol^{-1})	Local CCSD Vers. 1 (Hartree)	Error ($E_{\text{LCCSD}} - E_{\text{CCSD}}$) (kcal mol^{-1})
Extended	−494.355 30	−494.425 20	−494.417 83	4.6	−494.422 39	1.8
Globular	−494.437 73	−494.513 20	−494.505 71	4.7	−494.509 65	2.2
Diff.	−51.7	−55.2	−55.1		−54.8	
(kcal mol^{-1}) $E_{\text{glob}} - E_{\text{ext}}$						
%Ampls. Ext.	100	100	7.2		7.3	
%Ampls. Glob.	100	100	8.3		8.5	
%Ints. Ext.	100	100	60		9.4	
%Ints. Glob.	100	100	69		11	

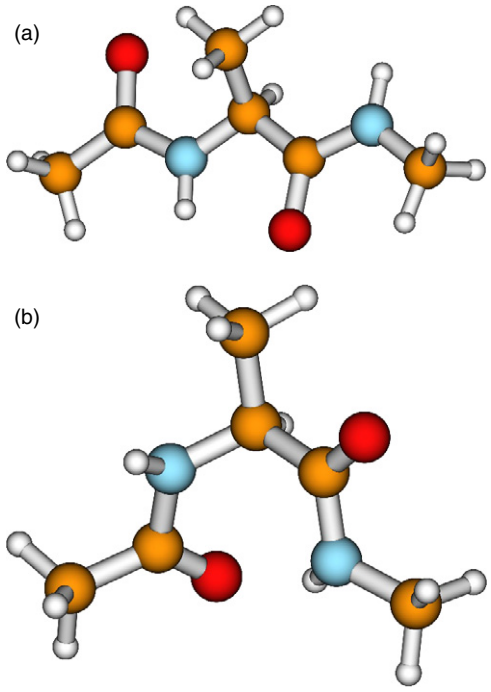

Figure 3. (a) The extended alanine dipeptide conformation corresponding to table 1. (b) The globular alanine dipeptide conformation corresponding to table 1.

previously calculated the CCSD and LCCSD energetics of these molecules [10], using an earlier version of LCCSD. The essential difference between our older algorithm [10] and our newer algorithm [11] is that, as discussed in section 1.2.3, the newer version bumps the amplitudes and the integrals in order to gain computational efficiency, whereas before we included all necessary integrals and approximated only the amplitudes.

All relevant data is included in table 1, including energetics and the sizes of the approximations. From the data, one sees the remarkable fact that, even though the newer version makes more approximations than the older version, the energetics from the newer version are closer to the CCSD energetics than the older version. While the older version has a serendipitous cancellation of errors to get the relative conformational energy correct to within 0.1 kcal, the newer version appears far more accurate and stable, cutting in half the error in atomization energies. This empirical fact—that bumping the integrals and amplitudes together greatly increases the accuracy of the atomization energies relative to bumping the amplitudes alone—has been widely reproduced in our LCCSD calculations. It appears to be a rare example where one can gain computational speed and accuracy at the same time.

In table 2a, we report the energetic gap between a branched and a linear alkane with eight carbon atoms. This system has been studied in depth by Grimme [46], who pointed out the difficulty in getting the correct isomerization energy, where (according to experiment) the branched isomer is about 2 kcal mol^{-1} more stable than the linear molecule. Note that, for all calculations, intramolecular basis set superposition error (BSSE) ought to lower the energy of the branched alkane relative to the linear alkane, as a Gaussian basis will saturate function space much faster for a compact molecule rather than for an extended molecule. Nevertheless, Hartree–Fock incorrectly predicts that the linear isomer is more stable by a very large margin, whereas MP2 over-corrects and assumes the branched structure is much too stable relatively. Even coupled-cluster gets the relative stability wrong in a double-zeta basis. One finds the correct solution only by going to a triple-zeta basis and correcting between an MP2 and CCSD calculation. As table 2a demonstrates, our local methods parallel the full methods every step of the way, producing a corrected answer which is less than 1 kcal away from the full, corrected solution.

Finally, in order to assess how sensitive our algorithm is to our choice of molecules with lighter, first row atoms, we repeat the same calculation of branched versus linear 8-mer for a silane rather than an alkane in table 2b. This is a crucial test of our approach, for our current algorithm chooses which localized orbitals to correlate together based on the distance of one from the other and the size of the orbital (measured as $\langle(\mathbf{r} - \langle\mathbf{r}\rangle)^2\rangle$). The hope is, that by accounting for the size of the orbitals, the theory should extend naturally to molecules with heavier atoms, which have more diffuse electron clouds around

Table 2a. Calculations for the ground-state energy of branched and linear eight-alkane isomers. Both molecules have been optimized at the RIMP2 level of theory in a cc-pVDZ basis [45].

Molecule	Basis	Branched octane (2,2,3,3-tetramethyl butane) (Hartree)	Linear octane (*n*-octane) (Hartree)	$E_{\text{branched}} - E_{\text{linear}}$ (kcal mol^{-1})
HF	cc-pVDZ	−313.441 28	−313.458 53	10.82
Full MP2	cc-pVDZ	−314.649 55	−314.643 40	−3.9
Local MP2	cc-pVDZ	−314.646 20	−314.640 60	−3.5
Full CCSD	cc-pVDZ	−314.755 95	−314.756 13	0.1
Local CCSD	cc-pVDZ	−314.752 91	−314.753 84	0.6
HF	cc-pVTZ	−313.525 25	−313.543 12	11.2
Full MP2	cc-pVTZ	−315.096 24	−315.087 40	−5.6
Local MP2	cc-pVTZ	−315.092 12	−315.083 89	−5.2
Corrected CCSD[a]	cc-pVTZ	−315.202 64	−315.200 14	−1.6
Corrected local CCSD[a]	cc-pVTZ	−315.198 83	−315.197 13	−1.1
Experimental [46]				-2 ± 1

[a] The correct relative stability is found only by combining the CCSD double-zeta energy with the MP2 triple-zeta energy: $E_{\text{corrected}} = E_{\text{CCSD}}(\text{DZ}) + (E_{\text{MP2}}(\text{TZ}) - E_{\text{MP2}}(\text{DZ}))$.

Table 2b. Branched versus linear molecules. The basis is cc-pVDZ and all molecules have been optimized at the level of RIMP2/cc-pVDZ.

Molecule	Basis	Branched octasilane (2,2,3,3-tetrahydryl tetrasilane) (Hartree)	Linear octasilane (*n*-octasilane) (Hartree)	$E_{\text{branched}} - E_{\text{linear}}$ (kcal mol^{-1})
HF	cc-pVDZ	−2321.929 88	−2321.925 27	−2.9
Full CCSD	cc-pVDZ	−2322.978 22	−2322.964 90	−8.4
Local CCSD	cc-pVDZ	−2322.968 96	−2322.954 10	−9.3

them. Nevertheless, in practice, the size of an orbital is a rather nebulous construction, and by calculating size according to dipole and quadrupole matrix elements, one ignores entirely the tails of the localized orbitals (which, in reality, decay with linear exponent, unlike the Gaussian atomic orbital basis set). Thus, testing the applicability of a local correlation algorithm to heavier elements is crucial. From the data, one may surmise that LCCSD performs quite well at measuring the energetic gap between the isomers, though the errors of the absolute numbers (i.e. atomization energies) are more than they usually are for lighter elements. Future testing of the accuracy of LCCSD methods with ECP's (i.e. pseudopotentials) will be another crucial test of our method.

3. Barrier heights: a Diels–Alder reaction

Beyond the ground-state energy problem, another test of the limit of local correlation theory is the applicability of the theory to measuring barrier heights, where one must consider nuclear geometries at which electronic orbitals are often delocalized as bonds are being made and broken. Following the recent work of Jones *et al* [47], we focus on the Diels–Alder reaction of cyclopentadiene with cyano-substituted ethylene derivatives.

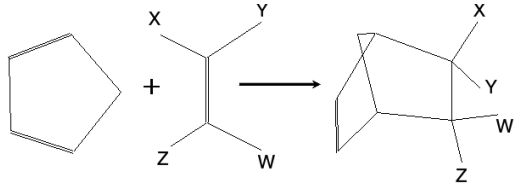

Table 3a. Experimental data from [48] showing the clear trend in reaction rates, as one increases the number of cyano substituents. Data taken in dioxane solution at 20 °C.

Dienophile	Rate of reaction 10^6 k_2 mol^{-1} s^{-1}(20 °C)
CN	10.4
trans-1, 2-$(CN)_2$	81
cis-1, 2-$(CN)_2$	91
1, 1-$(CN)_2$	455 000
1, 1, 2-$(CN)_3$	4 800 000
1, 1, 2, 2-$(CN)_4$	430 000 000

Benchmark experiments on this system were made in a classic experiment by Sauer *et al* [48] in 1964, who found that the reaction rate increases by a factor 10^7 as one traverses the different cyanoethylenes, beginning with the mono-substituted molecule and ending with the tetra-substituted one. Their classic data is replicated in table 3a.

In a recent report [47], Jones, Guner and Houk suggested a mechanism for this Diels–Alder reaction whereby each substituted ethylene passes through a single transition state (shown in both figures 4 and 5) before reacting to form the product. To this model, the authors then applied: (i) gas-phase data yielding the activation energy of mono-substituted ethylene reacting with cyclopentadiene [49] at 550 K and (ii) the Sauer reaction-rate data in dioxane solution at 293 K. By assuming that the reaction rate behaves in an Arrhenius fashion in both cases, and by further assuming that the Arrhenius constants are 'close

Table 3b. Relevant energies and enthalpies in the Diels–Alder reaction of cyclopentadiene with cyano-substituted ethylene. (a), (b), (d) from [47]. Basis sets are: (a) 6-31G(d), (b) 6-31G+(d, p) and (c) 6-31G(d, p). Geometries of (c) are from [47], i.e. optimized according to MPW1K/6-31G + (d, p). Note that these numbers are not corrected for BSSE (see the text). Although we provide energies instead of enthalpies in (c) columns, one can show that the contributions of vibrations and rotations is rather constant and less than 1 kcal mol^{-1} in all cases (at least within the harmonic approximation at room temperature). See the text for an explanation of how the experimental numbers in column (d) are calculated. Because these experimental numbers require the comparison of data from gas-phase and liquid-phase experiments, the absolute barrier heights may have a systematic error, although the relative energies should be quite accurate.

Molecule	HF[(a)] ΔH	B3LYP[(a)] ΔH	MPW1K[(b)] ΔH	MP2[(c)] ΔE	CCSD[(c)] ΔE	LCCSD[(c)] ΔE	Extrap. Expt. ΔH[(d)]
CN	37.8	18.9	16.3	5.8	19.4	20.1	15.5
trans-1, 2-$(CN)_2$	34.4	16.5	12.9	0.9	16.0	17.8	12.8
cis-1, 2-$(CN)_2$	34.5	17.0	13.4	1.5	16.3	18.2	12.8
1, 1-$(CN)_2$	30.0	11.2	8.9	0.5	13.6	15.0	9.0
1, 1, 2-$(CN)_3$	28.0	11.9	7.7	−3.3	11.5	13.9	7.6
1, 1, 2, 2-$(CN)_4$	26.1	12.1	5.6	−8.4	8.5	10.5	5.0

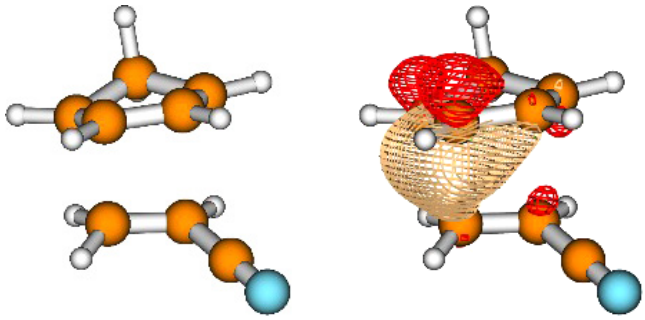

Figure 4. The transition structure for mono-cyanoethylene reacting with cyclopentadiene, and the largest localized orbital.

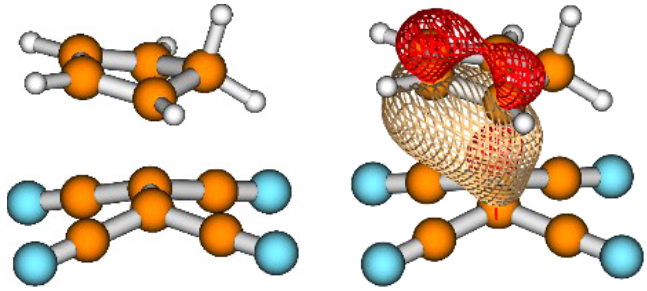

Figure 5. The transition structure for tetracyanoethylene reacting with cyclopentadiene, and the largest localized orbital.

enough' between the different temperatures and gas/liquid media, one may use the Sauer data to estimate the gas-phase activation barriers for the Diels–Alder reaction by $\Delta\Delta H^{\dagger} = RT \ln(k_2/k_1)$. These experimental estimates of the activation barriers are reported in column (d) of table 3b. Alternatively, Jones *et al* also use the COSMO polarization continuum model [50] to model the effect of bulk solvent, and do not find any large effect. Surely, ignoring the effects of the real quantum-mechanical liquid on a transition-state energy barrier is certainly a drastic approximation—but, nevertheless, it does give us a zeroth approximation to the activation energies of the differently substituted reactants in the gas phase. Furthermore, the relative barrier heights of mono-through tetra-cyano-substituted ethylene can certainly be computed rigorously in the gas phase, giving us an interesting example for theoretical benchmark calculations.

Besides the experimental data in column (d) of table 3b, we also include computational data from Jones *et al* showing that the B3LYP DFT functional entirely misses the correct trend in electronic structure responsible for the reaction rates in table 3a. Of the functionals tested by Jones *et al*, the only one found to work was the MPW1K [51] functional from the Don Truhlar group. Notice that the data presented here are formally enthalpies, not energies. Nevertheless, Jones and coworkers argue that nuclear motion plays almost no role in determining the relative rates of this reaction, which are determined by the increase in electron affinity as one adds more cyano groups to ethylene. This suggestion has been consistent with all of our frequency calculations.

Our interest in this molecular reaction was to test the accuracy of our local CCSD algorithm relative to the full CCSD algorithm. The reaction barrier heights of these Diels–Alder reactions are a good test of a local correlation model for two reasons. First, for a pericyclic reaction, multiple bonds are being made and broken at the same time and the degree of electronic correlation should be large, and the delocalization of electronic orbitals should change greatly between reactants and transition state. Second, two competing factors could introduce systematic error into our LCCSD calculations. On the one hand, as the number of cyano groups grows, there should be more electron donation from the diene to the dienophile at the transition state, giving us larger orbitals (see figures 4 and 5), and thus more amplitudes to include exactly in the calculation. On the other hand, as the number of cyano groups grows, the local approximation should exclude more amplitudes from the calculation. If our LCCSD is to be stable and accurate, it must be able to balance these effects naturally and correctly judge the energy of each configuration. The data shown in table 3b shows that our LCCSD algorithm is, in fact, able to reproduce the CCSD numbers to an accuracy of 2–3 kcal mol^{-1}.

For all but one of the calculations in table 3b, the basis set of choice is double-zeta (either 6-31G(d) or 6-31G(d, p)) with no augmented, diffuse basis functions. We expect that these calculations may be compared, one to the other, without worrying too much about the effect of basis set—polarization functions on the hydrogens are unlikely to be

Table 3c. In the first two columns, the counterpoise correction for BSSE has been applied to the data in table 3b, raising the barrier heights. The BSSE correction makes the LCCSD barrier height accurate to within 1 kcal mol^{-1} of the CCSD answer in a double-zeta basis. In the third and fourth columns, we further correct our barrier heights to account for a triple-zeta basis, using $E_{\text{corrected}} = E_{\text{CCSD}}(\text{CP, DZ}) + (E_{\text{MP2}}(\text{CP, TZ}) - E_{\text{MP2}}(\text{CP, DZ}))$, which lowers the barrier heights.

Molecule	CCSD ΔE (6-31g*) w/CP	LCCSD ΔE (6-31g*) w/CP	CCSD ΔE w/CP and TZ Corr.	LCCSD ΔE w/CP & TZ Corr	Extrap. Expt. $\Delta H^{(a)}$
CN	27.1	26.2	19.4	18.6	15.5
trans-1, 2-$(CN)_2$	24.0	24.1	16.4	16.5	12.8
cis-1, 2-$(CN)_2$	24.4	24.6	16.7	16.9	12.8
1, 1-$(CN)_2$	21.2	21.0	14.1	13.8	9.0
1, 1, 2-$(CN)_3$	19.8	20.3	12.5	12.9	7.6
1, 1, 2, 2-$(CN)_4$	17.9	17.7	10.0	9.8	5.0

[a] See text and table 3b for an explanation of the experimental values.

crucial. However, the MPW1K data (of Jones *et al*) was calculated with augmented functions (6 − 31 + G(d, p)), which could give the algorithm a relatively unfair advantage when calculating barrier heights, for diffuse functions may strongly help stabilize a transition-state geometry with partially formed bonds.

Note that the figures in table 3b have not been corrected for basis set superposition error (BSSE). In the first two columns of table 3c, we make the counterpoise correction for BSSE and we find that, while the ordering of the relative barrier heights is unmoved, in fact, the LCCSD energies become much closer to the CCSD energies. The changes in the overall sizes of the barrier heights due to basis set superposition error are large in the 6-31g(d, p) basis: the DFT and HF barrier heights go up rather uniformly by around 4–5 kcal mol^{-1}, the MP2 barriers are raised about 9 kcal mol^{-1} and the CCSD/LCCSD numbers are raised about 7 kcal mol^{-1}. These counterpoise corrections which add to the barrier height are, of course, compensated by the effect of basis set size: the double-zeta basis is incomplete here. In fact, MP2 calculations in a triple-zeta basis give barrier heights that are 7–8 kcal mol^{-1} lower than those in a double-zeta basis (data not shown). If we further correct the counterpoise-corrected numbers by this basis set correction, we end up with the values in the third and fourth columns of table 3c. In summary, certainly more accurate calculations should be done in the future to better estimate the barrier heights involved for bigger basis sets. Nevertheless, our data suggests that, when we account for BSSE, our LCCSD algorithm is able to reproduce CCSD barrier heights to well within 1 kcal mol^{-1}. Finally, we mention that our best CCSD/LCCSD numbers are not so close to the best DFT numbers, although they are still far lower than the overbound HF reaction. Our calculations imply that the experimental numbers in table 3b may suffer from a systematic error because they compare liquid-and gas-phase data together. Alternatively, either our CCSD energies are not converged with basis set or there are significant differential contributions from neglected three-body or higher correlations not present in CCSD (i.e. triples, quadruples, etc). These are the most logical ways to explain the discrepancy between our CCSD numbers and the experimental values.

4. Relevance to charge delocalization

Our final example of the applicability of local CCSD theory is for the case of measuring the electron affinity of tetracyanoethylene (TCNE, experimentally, 66.2 kcal mol^{-1})[4]. Our data is found in table 4. For this example, the mean-field solution is far from correct. The HF electron affinity badly overbinds the extra electron, arising as the result of spin contamination. Rather than the expected value of $\langle S^2 \rangle = 0.75$, we find for the unrestricted HF calculation, $\langle S^2 \rangle = 0.92$ in the 6-31g* basis and 0.84 in the cc-pVTZ basis.

Within the context of our LCCSD calculations, we now point out that one can construct two very different approaches for locally measuring the electron–electron correlation of a radical anion. Consider that the anion has 33 alpha electrons and 32 beta electrons. One can 'localize' the electronic orbitals in two different ways. First, one can localize all 33 alpha electrons together and hope that good local orbitals emerge. Second, one can localize the 32 tightly bound alpha electrons, which should correspond to the 32 tightly bound beta electrons in the closed-shell molecule, and then leave the 32nd attached electron totally delocalized, and to be treated separately. The first algorithm produces larger localized alpha orbitals than the second algorithm (ignoring the one loosely held electronic orbital which is purposefully kept delocalized). Examples of the most diffuse, localized occupied orbitals are shown in figures 6(a) and (b).

Once one has fixed the localized orbitals, one can next run a standard LCCSD calculation as described above. This provides two very different approaches for computing LCCSD energies, which are listed in table 4 as versions 1 and 2. Remarkably, one finds that both methods are successful at treating the correlation problem, though version 1 is slightly more accurate than version 2 in this case. Of course, by allowing the loosely attached electron to remain delocalized, one must pay a higher computational cost, so version 1 is the recommended algorithm of choice.

Nevertheless, the success of version 2 at capturing electronic correlation is noteworthy and deserves mention. Let us label i as the loosely attached electronic orbital in figure 6(b) which is not localized, but suppose we localize all of the other occupied orbitals. Because i is delocalized, i overlaps

[4] Lide D (ed) 2008 CRC Handbook of Chemistry and Physics.

Table 4. The ground-state energies of neutral and anionic tetracyanoethylene (TCNE) treated at different levels of theory for electron–electron correlation. The molecule has been optimized at the level of 6-31G*/RIMP2. The experimental answer is 66.2 kcal mol^{-1}.

Method	Basis	Neutral TCNE (Hartree)	Anion TCNE (Hartree)	Electron affinity (kcal mol^{-1})	Error $E_{\text{calc}} - E_{\text{expt}}$ (kcal mol^{-1})	% Strong t-amps.	% Integr. included
HF	6-31G*	−444.887 57	−445.012 03	78.1	11.9	—	—
MP2	6-31G*	−446.303 20	−446.362 80	37.4	−28.8	100	100
CCSD	6-31G*	−446.300 69	−446.391 86	57.2	−9.0	100	100
LCCSD version 1	6-31G*	−446.299 75	−446.391 239	57.4	−8.8	9.7	16.5
LCCSD version 2	6-31G*	−446.299 75	−446.390 138	56.2	−10.0	10.2	16.7
CCSD(T)	6-31G*	−446.372 47	−446.455 21	51.9	−14.3	100	100
HF	cc-pVTZ	−446.025 07	−446.158 69	83.8	17.6	—	—
MP2	cc-pVTZ	−446.841 75	−446.927 60	53.9	−12.3	100	100
CCSD[a,c]	cc-pVTZ	−446.816 69	−446.935 44	74.5 [69.2]	8.3 [3.0]	100	100
LCCSD[a,b] version 1	cc-pVTZ	−446.797 52	−446.913 15	72.6 [67.1]	6.4 [0.9]	2.4	5.6
LCCSD[a,b] version 2	cc-pVTZ	−446.797 52	−446.913 37	72.7 [68.5]	6.5 [2.3]	2.5	5.5

[a] The numbers in brackets represent our final, triple-zeta, triples-corrected estimate of the electron affinity: $EA = E_{\text{CCSD}}(\text{TZ}) + E_{\text{CCSD}(T)}(\text{DZ}) - E_{\text{CCSD}}(\text{DZ})$.

[b] In order to minimize computational time, we used more severe cutoffs for these LCCSD calculations than usual, setting $c_1^{\text{strong}} = c_0^{\text{strong}} = 0.58$ Å and $c_1^{\text{medium}} = c_0^{\text{medium}} = 2.12$ Å. These are much stronger cutoffs than usual: see section 1.2.2. This explains our rather large errors in atomization energy, even though the relative energies are still accurate.

[c] Note that a full CCSD calculation is possible here only because of the high point-group symmetry of TCNE.

with all localized occupied orbitals jk and localized virtual orbitals $abef$. If we use a good selection criteria for electronic amplitudes which accounts for spatial extent, we must select all ij and ia for strong amplitude coupling, where j and a are localized orbitals. Thus, if we consider the formal CCSD equations which must be solved, we find that we must compute a large number of virtual two-electron integrals $\langle ab||ef\rangle$ with a and e far away—since a and e are coupled together by $R_{ij}^{ab} = t_{ij}^{ef}\langle ab||ef\rangle$ (in equation (17)) and i couples to e and a (since i is delocalized). We conclude that, if we want to formally allow delocalization and also achieve a fast linear scaling algorithm, we should bump (i.e. approximate) the two-electron integrals, as we suggested earlier in section 1.2.3. Indeed, the two-electron integrals were indeed bumped in the calculations labeled version 2 in table 4. Note that, for tetracyanoethylene, the correlation energy of the neutral molecule is 1.413/1.776 Hartree and the correlation energy of the anion is 1.380/1.792 Hartree (at the DZ/TZ level), and this large difference in correlation energy suggests that the extra electron is strongly correlated to other electrons in the system. Thus, the fact that versions 1 and 2 give very similar correlation energies, with totally different ansatzes for the wavefunctions—and very different approximations of the integrals—proves that bumping of the two-electron integrals is not necessarily a bad approximation.

Finally, we point out how close our final, triple-zeta, triples-corrected electron affinities are to experiment. There is certainly some luck associated with our triple-zeta calculations. As discussed in table 4, for our LCCSD triple-zeta calculations, we have used much more severe cutoffs than usual in order to save time. This results in a rather large error in atomization energy, though the relative energy of the anion to the neutral remains remarkably constant. Most likely, we have found a fortuitous cancellation of errors, though it is conceivable that this error cancellation is the result of neglecting a constant long-distance correlation.

5. Discussion

This paper has surveyed a small number of molecular examples where computational chemists would like to calculate the electronic structure of molecules to within chemical accuracy (i.e. less than 1 kcal mol^{-1} from the exact value). For the most part, we have found that local correlation theory, in the form of local coupled-cluster theory, *can* accurately and reliably calculate the effect of electron–electron correlation as well as full CCSD for a variety of molecular geometries: equilibrium ground-state geometries, anion and transition-state geometries, and even molecules with heavier elements (or at least those in the third periodic row, including silicon). Indeed, one can measure electronic correlation energies using LCCSD which are usually within 1 kcal mol^{-1} of the exact CCSD energy even for very subtle differences, including isomerization energies. Of course, the trick is how strictly does one choose the cutoffs. We emphasize that, for this paper, for all double-zeta calculations, we have used the cutoffs of our previous paper [11], wherein the cutoffs and bumping regions were chosen so as to have smooth potential energy surfaces. Although we have been forced (for computational reasons) to *increase* the severity of our cutoffs for the TZ calculations in table 4, we have never relaxed our cutoffs in such a way as to increase the accuracy of the calculations reported in this paper. Moreover, the reader who is generally skeptical of local correlation methods should be comforted by the simple fact that, when performing local correlation

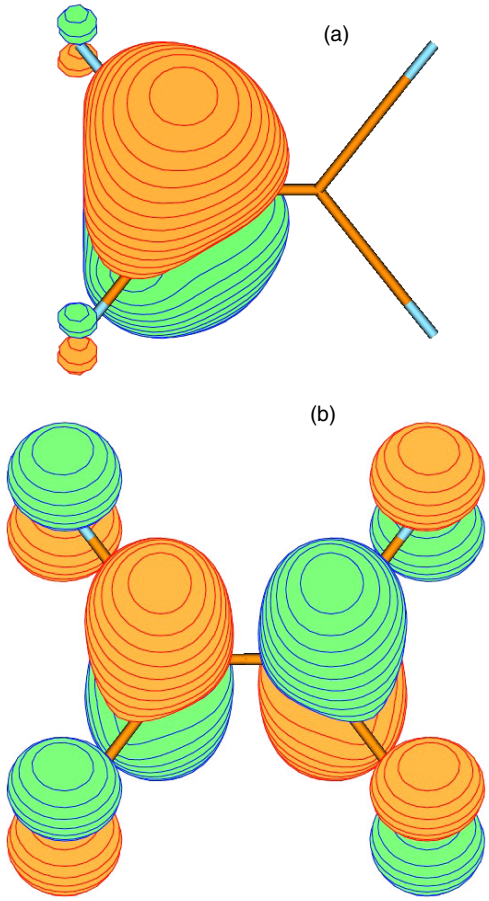

Figure 6. (a) The most diffuse localized orbital from anionic TCNE assuming all alpha, occupied electronic orbitals are localized. This orbital is used in version 1 of table 4. (b) The loosely attached, but occupied, electronic orbital of anionic TCNE which is not localized in version 2 of table 4.

calculations, one cannot avoid a penalty by relaxing the cutoffs of the calculation for increased accuracy. Indeed, we have shown [11] that there is a stiff penalty that must be paid for insisting on smooth potential energy surfaces in LCCSD, requiring the large bumping regions mentioned in section 1.2.2. If one ignores the smoothness of the PES's, one can gain at least a factor of 2–3, if not more, in computational speed.

We must admit, however, that our LCCSD method does not have the complete chemical accuracy which is the aim of the LCCSD model of Auer and Nooijen [43]. While Auer and Nooijen seek to locally approximate the CCSD by converging the local correlation energy to a stable, monotonically decaying energetic limit, we make no claim to have a completely converged solution. Although Auer and Nooijen will certainly achieve a more stable algorithm than ours, they must sacrifice so much in computational time that the benefits of locality do not become paramount until one reaches quite large systems. Our philosophy is that we want to approximate the CCSD correlation energy by a LCCSD energy that (i) is fast and (ii) has an error relative to CCSD which is less than the innate CCSD error relative to full CI:

$$|E_{\mathrm{LCCSD}} - E_{\mathrm{CCSD}}| < |E_{\mathrm{CCSD}} - E_{\mathrm{Full\ CI}}|. \qquad (25)$$

We believe that our LCCSD algorithm satisfies this constraint of accuracy using our choice of parameters given in this paper. Perhaps the largest source of error in our algorithm is that the moderate amplitude correlations depend on the strong amplitude correlations, but not vice versa. This essentially corresponds to the weak-coupling approximation as discussed by Auer and Nooijen [43]. Correcting such a unidirectional flow of information should be explored in the future. Nevertheless, it is very interesting that this nonsymmetric algorithm for computing LCCSD energies does not lead to very large errors. It is empirically true that, if we compute the LCCSD energies of strong amplitudes first, and then we compute the LMP2 energies of moderate amplitudes as a function of the strong amplitudes, our final energy is close to the exact CCSD energy. This is not *at all* true if one defines strong and moderate amplitudes in the context of an LMP2 calculation. In other words, if one defines strong LMP2 amplitudes as those amplitudes that should give the strongest LMP2 contributions, and then one computes moderate LMP2 amplitudes as a function of the strong LMP2 amplitudes, one finds that the final answer is usually off by 10–20 kcal. A good LMP2 algorithm cannot have any strong LMP2 amplitudes, only moderate LMP2 amplitudes, and all of the moderate LMP2 amplitudes must be computed self-consistently. Thus, either the strong LCCSD amplitudes have a much smaller correlation length than the 'strong' LMP2 amplitudes, or perhaps there is a natural cancellation of errors between the strong LCCSD amplitudes and moderate LMP2 amplitudes. In any event, the unidirectional flow of information from strong to moderate amplitudes does not appear to be a large source of error in our algorithm. Furthermore, in our limited experience with the strong-coupling limit of Auer and Nooijen [43], where the strong amplitudes are allowed to depend on the moderate amplitudes, we found there was often numerical instability when using our current cutoff parameters.

Moving beyond the calculation of ground-state or transition-state energies for insulators, our success at approximating the electron affinity of tetracyanoethylene by two different algorithms suggests many new and exciting directions for the LCCSD algorithm. If one can use local coupled-cluster theory to treating molecules with true charge delocalization, then perhaps one might even be capable of describing the correlations of an insulating molecule on a small metallic cluster or perhaps the electronic correlation on a metal–molecule–metal junction. That would be a remarkable application of a local theory. There must, however, be a physical limit to the applicability of a local CCSD theory, however, as the size of a metal cluster grows, because for a large enough metal cluster, a mean-field HF or DFT ground state cannot be a good-enough zeroth-order approximation for a correlated non-periodic wavefunction. In that macroscopic limit, the bandgap shrinks to zero and the CCSD equations cannot be inverted (or at least not in the standard formalism).

We must now mention that the most disturbing and worrisome piece of data presented in this paper comes from the branched and linear silane 8-mer (table 2b). One notes that the error in atomization energy for the silanes is double the error in atomization for the alkanes, even though the correlation

energy for the silanes is less than the correlation energy for the alkanes (tables 2a, 2b). This suggests that, in the future, more work must be done to either (i) verify that the bubble method correctly describes the sizes and correlation lengths of localized orbitals with heavy atoms, or (ii) invent a new, but still differentiable, function of orbitals which will determine locality. Such a function should necessarily work with ECP's (i.e. pseudopotentials) as well.

Finally, we mention that the next phase of development for our LCCSD algorithm should be the construction of the gradient and the application of the triples (T) correction. If these advances can be implemented, we hope that our LCCSD algorithm will become more widely used in the computational chemistry community. More distantly in the future, we note that the size of the systems which can be treated at present by LCCSD are still nowhere near as large as those that can be treated at the mean-field DFT level for several reasons. First, local CCSD has not yet been efficiently parallelized to the knowledge of the authors. Second, there has not been any development of a CCSD or LCCSD algorithm that can be applied to periodic systems in more than one dimension [52]. Many theoretical and computational challenges remain in the field of localized electron–electron correlation theory, and we hope to see much more progress in the years to come.

6. Conclusions

This paper has investigated the accuracy of local coupled-cluster theory as a computational, quantum-mechanical tool designed to capture electronic correlation with chemical accuracy ($<$1 kcal mol^{-1}) and with broad applicability. Over a broad range of difficult chemical problems, including isomerization energies, conformational energies, electron affinities and barrier heights, the accuracy of the LCCSD holds up. The only worrisome piece of data is that the error in atomization energy for the branched and linear silanes is double the error for the corresponding alkanes. This suggests that more calculations must be done on molecules with heavy elements beyond the third row of the periodic table in order to ascertain how accurately LCCSD treats very diffuse electronic orbitals around very heavy nuclei, or equivalently, electronic orbitals in the presence of a pseudopotential (i.e. ECP's). Finally, we have presented here data suggesting that LCCSD theory can successfully capture the electronic correlation of anions and transition states, molecular geometries which often have a great deal of charge delocalization. This data raises the prospect that local correlation theory may be extended far beyond standard quantum chemical processes. We will investigate this claim much more thoroughly in the future.

Acknowledgments

We thank Stefan Grimme, Vered Ben-Moshe and Robert Distasio Jr for provocative conversations and helpful suggestions. JES was supported by the NSF International Research Fellowship. MHG is a part owner of Q-Chem.

Appendix

For completeness, we now write down how and why we have two bump functions in equation (14). The $\mathbf{R}$ matrix in equation (18) can be naturally split up into the perturbative MP2 piece ($\mathbf{R}_{\mathrm{MP2}}$) and the higher-order, coupled-cluster piece ($\mathbf{R}_{\mathrm{CC}}$): $\mathbf{R} = \mathbf{R}_{\mathrm{MP2}} + \mathbf{R}_{\mathrm{CC}}$. The MP2 piece is almost always found to decay more slowly. When one makes this division, it follows that, for more accuracy, one can choose two different bump functions: g^{strong} (also referred to as $\mathbf{G}_{\mathrm{CC}}$) corresponding to $\mathbf{R}_{\mathrm{CC}}$ and g^{weak} (also referred to as $\mathbf{G}_{\mathrm{MP2}}$) corresponding to $\mathbf{R}_{\mathrm{MP2}}$. One then has two different classes of t amplitudes, those amplitudes for which $g^{\mathrm{strong}} > 0$ and those for which $g^{\mathrm{moderate}} > 0$.

If we insist that the weak amplitudes depend only on the strong amplitudes, but not vice versa, we get the equations:

$$\mathbf{I}(\mathbf{n}) + \mathbf{A}^{(\mathrm{d})}(\mathbf{n}) \cdot \mathbf{t} + \mathbf{G}_{\mathrm{MP2}} \cdot \mathbf{R}_{\mathrm{MP2}}(\mathbf{G}_{\mathrm{CC}} \cdot \mathbf{t}, \mathbf{I}(\mathbf{n}), \mathbf{F}(\mathbf{n})) + \mathbf{G}_{\mathrm{CC}} \cdot \mathbf{R}_{\mathrm{CC}}(\mathbf{G}_{\mathrm{CC}} \cdot \mathbf{t}, \mathbf{I}(\mathbf{n}), \mathbf{F}(\mathbf{n})) = 0.$$

These are the precise definitions of the bump-function constants in equation (14). If one furthermore bumps the integrals and Fock matrix elements, one arrives at our final, smooth LCCSD equations:

$$\mathbf{I}(\mathbf{n}) + \mathbf{A}^{(\mathrm{d})}(\mathbf{n}) \cdot \mathbf{t} + \mathbf{G}_{\mathrm{MP2}} \cdot \mathbf{R}_{\mathrm{MP2}}(\mathbf{G}_{\mathrm{CC}} \cdot \mathbf{t}, \mathbf{G}_{\mathrm{CC}} \cdot \mathbf{I}(\mathbf{n}), \times \mathbf{G}_{\mathrm{MP2}} \cdot \mathbf{F}(\mathbf{n})) + \mathbf{G}_{\mathrm{CC}} \cdot \mathbf{R}_{\mathrm{CC}}(\mathbf{G}_{\mathrm{CC}} \cdot \mathbf{t}, \times \mathbf{G}_{\mathrm{CC}} \cdot \mathbf{I}(\mathbf{n}), \mathbf{G}_{\mathrm{MP2}} \cdot \mathbf{F}(\mathbf{n})) = 0.$$

References

[1] Martin R M 2004 *Electronic Structure* (Cambridge: Cambridge University Press)
[2] Szabo A and Ostlund N 1996 *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (New York: Dover)
[3] Scuseria G E and Ayala P Y 1999 *J. Chem. Phys.* **111** 8330
[4] Schütz M and Werner H J 2001 *J. Chem. Phys.* **114** 661
[5] Li S, Ma J and Jiang Y 2002 *J. Comput. Chem.* **23** 237
[6] Flocke N and Bartlett R J 2004 *J. Chem. Phys.* **121** 10935
[7] Schütz M 2002 *Chem. Phys. Phys. Chem.* **4** 3941
[8] Schütz M 2003 *Chem. Phys. Phys. Chem.* **5** 3349
[9] Subotnik J E and Head-Gordon M 2005 *J. Chem. Phys.* **123** 064108
[10] Subotnik J E, Sodt A and Head-Gordon M 2006 *J. Chem. Phys.* **125** 074116
[11] Subotnik J E, Sodt A and Head-Gordon M 2008 *J. Chem. Phys.* **128** 034103
[12] Cízek J 1966 *J. Chem. Phys.* **45** 4256
[13] Cízek J 1969 *Adv. Chem. Phys.* **14** 35
[14] Cízek J and Paldus J 1971 *Int. J. Quantum Chem.* **5** 359
[15] Crawford T D and Schaefer H F 2000 *Rev. Comput. Chem.* **14** 33
[16] Jensen F 1999 *Introduction to Computational Chemistry* (Chichester: Wiley)
[17] Saebo S and Pulay P 1993 *Annu. Rev. Phys. Chem.* **44** 213
[18] Murphy R B, Beachy M D, Ringnalda M N and Friesner R A 1995 *J. Chem. Phys.* **103** 1481
[19] Reynolds G, Martinez T and Carter E 1996 *J. Chem. Phys.* **105** 6455
[20] Hampel C and Werner H J 1996 *J. Chem. Phys.* **104** 6286
[21] Azhary A E, Rauhut G, Pulay P and Werner H J 1998 *J. Chem. Phys.* **108** 5185

[22] Rauhut G, Azhary A E, Eckert F, Schumann U and Werner H J 1999 *Spectrochim. Acta* A **55** 647
[23] Schütz M, Hetzer G and Werner H J 1999 *J. Chem. Phys.* **111** 5691
[24] Schütz M and Werner H J 2000 *Chem. Phys. Lett.* **318** 370
[25] Schütz M 2000 *J. Chem. Phys.* **113** 9986
[26] Rauhut G and Werner H J 2001 *Phys. Chem. Chem. Phys.* **3** 4853
[27] Schütz M 2002 *J. Chem. Phys.* **116** 8772
[28] Rauhut G and Werner H J 2003 *Phys. Chem. Chem. Phys.* **5** 2001
[29] Werner H J, Manby F R and Knowles P J 2003 *J. Chem. Phys.* **118** 8149
[30] Schütz M, Werner H J, Lindh R and Manby F 2004 *J. Chem. Phys.* **121** 737
[31] Hrenar T, Rauhut G and Werner H J 2006 *J. Phys. Chem.* A **110** 2060
[32] Russ N and Crawford T D 2004 *J. Chem. Phys.* **121** 691
[33] Mata R A and Werner H J 2006 *J. Chem. Phys.* **125** 184110
[34] Lee M S, Maslen P A and Head-Gordon M 2000 *J. Chem. Phys.* **112** 3592
[35] Distasio R A, Jung Y and Head-Gordon M 2005 *J. Chem. Theor. Comput.* **1** 862
[36] Maslen P E and Head-Gordon M 1998 *J. Chem. Phys.* **109** 7093
[37] Maslen P E and Head-Gordon M 1998 *Chem. Phys. Lett.* **283** 102
[38] Boughton J W and Pulay P 1993 *J. Comp. Chem.* **14** 736
[39] Edmiston C and Ruedenberg K 1963 *Rev. Mod. Phys.* **35** 457
[40] Boys S F 1966 *Quantum Theory of Atoms, Molecules and the Solid State* ed P Lowdin (New York: Academic) p 253
[41] Subotnik J E, Dutoi A D and Head-Gordon M 2005 *J. Chem. Phys.* **123** 114108
[42] Lee J M 2002 *Introduction to Smooth Manifolds* (Berlin: Springer)
[43] Auer A A and Nooijen M 2006 *J. Chem. Phys.* **125** 024104
[44] Shao Y *et al* 2006 *Phys. Chem. Chem. Phys.* **8** 3172
[45] Distasio R A, Steele R, Rhee Y M, Shao Y and Head-Gordon M 2007 *J. Comput. Chem.* **28** 839
[46] Grimme S 2006 *Angew. Chem. Int. Edn* **45** 4460
[47] Jones G O, Guner V A and Houk K N 2006 *J. Phys. Chem.* A **110** 1216
[48] Sauer J, Wiest H and Mielert A 1964 *Chem. Ber.* **97** 3183
[49] Walsh R and Wells J M 1976 *J. Chem. Soc. Perkin Trans.* 2, 52
[50] Klamt A and Schüürmann G 1993 *J. Chem. Soc. Perkin Trans.* 2, 799
[51] Lynch B J, Patton L F, Harris M and Truhlar D G 2000 *J. Phys. Chem.* A **104** 4811
[52] Hirata S, Podeszwa R, Tobita M and Bartlett R J 2004 *J. Chem. Phys.* **120** 2581