A local correlation model that yields intrinsically smooth potential-energy surfaces

Joseph E. Subotnik[a] and Martin Head-Gordon[b]
Department of Chemistry, University of California, Berkeley, California 94720 and Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720

(Received 7 June 2005; accepted 20 June 2005; published online 18 August 2005)

We demonstrate an algorithm for computing local coupled-cluster doubles (LCCD) energies that form rigorously smooth potential-energy surfaces and which should be fast enough for application to large systems in the future. Like previous LCCD algorithms, our method solves iteratively for only a limited number of correlation amplitudes, treating the remaining amplitudes with second-order perturbation theory. However, by employing bump functions, our method smoothes the transition from iteratively solved amplitude to perturbation-treated amplitude, invoking the implicit function theorem to prove that our LCCD energy is an infinitely differentiable function of nuclear coordinates. We make no explicit amplitude domains nor do we rely on the existence of atom-centered, redundant orbitals in order to get smooth potential-energy curves. In fact, our algorithm employs only localized orthonormal occupied and virtual orbitals. Our approach should be applicable to many other electron correlation methods. © *2005 American Institute of Physics*. [DOI: 10.1063/1.2000252]

I. INTRODUCTION

The correlation of electrons above and beyond the Hartree-Fock mean-field level is crucial towards understanding bond making and bond breaking in chemistry. Hartree-Fock theory correctly predicts stable equilibrium geometries; however, the correlation energy is crucial for determining relative energies and barrier heights accurately enough to be chemically useful.

The exact computation of electron correlation energies [full configuration interaction(CI)] formally scales exponentially with the number of electrons. Approximations to full CI scale much better and usually capture the essential chemistry and correlations between electrons. Today, many correlation methods are in use, including (among many more) MP2, CCSD, and CCSD(T), which are here ordered by complexity and accuracy. Formally, these three methods scale as m^5, m^6, and m^7 respectively, where m is the number of electrons. This cost is clearly prohibitive for large molecules, e.g., $m \geqslant 1000$. Thus, in order to calculate correlation energies for systems with large numbers of degrees of freedom, one is forced to approximate standard electron correlation algorithms (which in turn already approximate full CI).

Now, most modern electron correlation algorithms require the computation of amplitudes that physically represent electrons in occupied orbitals correlating with each other and scattering into virtual orbitals. When only double excitations are considered, as in coupled-cluster doubles (CCD) which we consider here, one needs to compute only amplitudes of the form t_{ij}^{ab}. Here i and j are occupied molecular orbitals, and a and b are virtual molecular orbitals.

The basic model for local correlation theory is to localize all orbitals in three-dimensional (3D) space and then restrict attention to those sets of orbitals $\{\eta_i \eta_j \eta_a \eta_b\}$ that are close to each other. By computing only certain t_{ij}^{ab} explicitly, one hopes to dramatically speed up the computational time required. Furthermore, because the ignored sets of orbitals are separated in space, one hopes that a local correlation algorithm will capture the important correlations, and that the local correlation energy will be almost as accurate as a full correlation theory, at least for relative energies. Ideally, we should ignore only those amplitudes with numerical values so rigorously small (e.g., $<10^{-16}$) that their inclusion is guaranteed to contribute nothing more than round-off error. However, modern coupled-cluster algorithms cannot yet treat molecules large enough for such a rigorous numerical cutoff to take effect, and, as a result, we are forced to make *models* of which amplitudes to treat at a coupled-cluster level of theory.

Given this framework, there are today two specific models on how to implement local correlation approximations. The first and predominant school of thought, pioneered by Pulay and co-workers,[1–5] applies a locality criteria in order to break the localized molecular orbitals into spatial domains and then create a subset of t_{ij}^{ab} amplitudes for which one solves iteratively in a self-consistent way. The remaining amplitudes are treated with perturbation theory. This strategy was later combined with pseudospectral techniques[6,7] for computational efficiency. Most recently, Werner and co-workers[8–11] and Schütz,[12,13] have made further algorithmic improvements to yield linear scaling local algorithms for MP2, CCSD, and CCSD(T). Their algorithms have been very successful at computing local MP2 and local CCSD energies for large systems and the speed of their current algorithms is quite impressive.

In fact, the only substantial limitation of the Pulay-based

[a]Electronic mail: subotnik@post.harvard.edu
[b]Electronic mail: mhg@cchem.berkeley.edu

methods is that they behave erratically at certain points in nuclear-coordinate space, where the domain definitions and thus the calculated energies change discontinuously for infinitesimal changes in nuclear geometry. Because the basic Pulay scheme applies static domain definitions and correlates only certain domains, one produces noncontinuous potential-energy surfaces[14] when the domains are redefined. Russ and Crawford have found these discontinuities to be on the order of 1–5 milli-Hartree for small systems using the Schütz-Werner local MP2 (LMP2) and local CCSD (LCCSD) algorithms. These are not large discontinuities but, for larger systems with more degrees of freedom, ever more discontinuities will make geometric minimization and transition-state determination difficult. Werner and coworkers have shown that one can avoid such discontinuities by defining domains wisely when constructing one specific potential-energy curve.[15,10] This approach is not fundamentally satisfying, however, because it makes sense along only one path in the space of nuclear coordinates. If one switches domains along a sequence of paths, one finds hysteresis in the electron correlation energy. This approach is not a theoretical model chemistry as defined by Pople.[16]

A second school of thought for local correlation limits the allowed substitutions according to atomic criteria (without any cutoffs) and produces an energy which is an entirely differentiable function of the nuclear coordinates.[17–20] However, thus far, this approach has not scaled better than N^3, where N is the size of the basis. The primary obstacle for this technique has been handling nonorthogonal, redundant functions, which are necessary to localize all orbitals around atoms.

It is essentially self-evident that the ideal locally correlated approach should be computationally fast (linear scaling in the number of electrons) and produce an energy which is a smooth function of the nuclear coordinates. Furthermore, there should ideally be no parameterization necessary for the algorithm and, if one is forced to use parameters, their number should be as few as possible. There have been new ideas recently for local MP2 (e.g., the Laplace approach[21]), but innovation has been slower for more complicated coupled-cluster algorithms. We present here a local CCD algorithm which is rigorously smooth and which has the potential to be very fast. We have exactly six molecule-independent parameters that must be chosen to trade off speed versus accuracy for our method.

II. NOTATION

We denote by $\mathbf{n}$ the coordinates of the nuclei of our molecule. We denote by $\mathbf{t}$ the vector of t_{ij}^{ab} amplitudes. In our analysis below, we will partition the set of all t amplitudes into those *strong* amplitudes for which we solve iteratively and those *weak t* amplitudes which we treat perturbatively at the simplest second-order level originally proposed by Kapuy and co-workers.[22–27] We abbreviate the former ITA's for "iterated t amplitudes" and the latter PTA's for "perturbative t amplitudes." Finally, for the entirety of this paper, η will always be a localized orthonormal spin orbital (occupied or virtual). As usual, $ijkl$ represent occupied orbitals, and $abcd$ represent virtual orbitals.

III. THEORY

A. The full CCD equations

The CCD equations describe the correlation between pairs of electrons which couple together and scatter. We work below in a basis of orthonormal localized spin orbitals $\{\eta\}$. f denotes the Fock matrix. The CCD equations have the following form:

$$0 = I_{ijab} + A_{ijab}^{(d)} t_{ij}^{ab} + R_{ijab}(\mathbf{t}), \tag{1}$$

where

$$\begin{aligned} R_{ijab}(\mathbf{t}) = & \sum_{klcd} A_{ijab,klcd}^{(nd)} t_{kl}^{cd} + \frac{1}{2}\sum_{cd} I_{abcd} t_{ij}^{cd} + \frac{1}{2}\sum_{kl} I_{klij} t_{kl}^{ab} \\ & - \sum_{kc} \left(I_{bkjc} t_{ki}^{ac} + I_{akic} t_{kj}^{bc} - I_{bkic} t_{kj}^{ac} - I_{akjc} t_{ki}^{bc} \right) \\ & + \sum_{klcd} I_{klcd} \begin{pmatrix} \frac{1}{4} t_{ij}^{cd} t_{kl}^{ab} - \frac{1}{2} t_{ij}^{ac} t_{kl}^{bd} + \frac{1}{2} t_{ij}^{bc} t_{kl}^{ad} \\ - \frac{1}{2} t_{ik}^{ab} t_{jl}^{cd} + \frac{1}{2} t_{jk}^{ab} t_{il}^{cd} \\ + \frac{1}{2} t_{ki}^{ac} t_{lj}^{bd} + \frac{1}{2} t_{kj}^{bc} t_{li}^{ad} - \frac{1}{2} t_{kj}^{ac} t_{li}^{bd} - \frac{1}{2} t_{ki}^{bc} t_{lj}^{ad} \end{pmatrix}, \end{aligned} \tag{2}$$

$$I_{pqrs} = \langle \eta_p \eta_q \| \eta_r \eta_s \rangle, \tag{3}$$

$$A_{ijab}^{(d)} = A_{ijab,ijab}^{(d)} = f_{aa} + f_{bb} - f_{ii} - f_{jj}, \tag{4}$$

$$\begin{aligned} A_{ijab,klcd}^{(nd)} = & f_{ac}\delta_{bd}\delta_{ik}\delta_{jl} + f_{bd}\delta_{ac}\delta_{ik}\delta_{jl} - f_{ik}\delta_{ac}\delta_{bd}\delta_{jl} \\ & - f_{jl}\delta_{ac}\delta_{bd}\delta_{ik}. \end{aligned} \tag{5}$$

If we define $\mathbf{A} = \mathbf{A}^{(d)} + \mathbf{A}^{(nd)}$, then $\mathbf{A}$ is the Hylleraas[28] bilinear form for standard MP2. Here d stands for the diagonal and nd for the nondiagonal part of $\mathbf{A}$. In all that follows, we may calculate MP2 (rather than CCD) energies simply by ignoring all terms in $\mathbf{R}$ except the $\mathbf{A}$ term.

For CCD, Eq. (1) is solved iteratively. One pulls out the $A^{(d)}$ term and computes

$$-A_{ijab}^{(d)} t_{ij}^{ab} = I_{ijab} + R_{ijab}(\mathbf{t}) \tag{6}$$

repeatedly until convergence. It follows that traditional coupled-cluster doubles theory scales as m^6, where m is the number of electrons. There are proportionally $O(m^4)$ amplitudes to solve, and each equation require $O(m^2)$ flops from matrix multiplication. If no algebraic simplification or transformation of these equations can be made, then any attempt to improve the scaling of this algorithm must reduce the number of relevant amplitudes so that both the number of iterative equations and the cost of each equation are reduced. Again, the local ansatz is that one chooses localized orbitals $\{\eta_i, \eta_j, \eta_a, \eta_b\}$ and then solves iteratively only for those t_{ij}^{ab} for which $\{\eta_i, \eta_j, \eta_a, \eta_b\}$ are all close.

B. The smoothed local CCD equations

The problem is how to construct a local CCD algorithm for which the energy is a smooth function of nuclear coordi-

nates. We want to partition the set of t_{ij}^{ab} amplitudes into those for which we solve iteratively (ITA's) and those which we treat perturbatively (PTA's). The essential difficulty is our requirement that when the nuclear geometry changes and the number of ITA's changes noncontinuously, the energy must change differentiably. We propose to solve this problem by modifying the CCD amplitude equations [see Eq. (6) above] as follows:

$$-A_{ijab}^{(d)}t_{ij}^{ab} = I_{ijab} + b_{ijab}R_{ijab}(\tilde{\mathbf{t}}), \tag{7}$$

$$\tilde{t}_{ijab} = b_{ijab}t_{ijab}. \tag{8}$$

Following standard terminology in differential geometry, we call b_{ijab} a bump function and we show how it is constructed below. For the moment, assume only that

- b is a smooth, infinitely differentiable function, $0<b<1$;
- $b=1$ when $\{\eta_i, \eta_j, \eta_a, \eta_b\}$ are close together, forming a highly correlated set of occupied and virtual orbitals; and
- $b=0$ when $\{\eta_i, \eta_j, \eta_a, \eta_b\}$ are far apart.

Before we analyze the local approximations made in Eqs. (7) and (8), let us first show that these equations produce rigorously smooth t amplitudes. We initially return to the standard CCD equations (1) and (2), which we know produce smooth potential-energy curves from empirical experience. From an abstract perspective, these equations may be rewritten as follows:

$$0 = \mathbf{g}(\mathbf{t},\mathbf{n}), \tag{9}$$

$$\mathbf{g}(\mathbf{t},\mathbf{n}) = \mathbf{I}(\mathbf{n}) + \mathbf{A}^{(d)}(\mathbf{n})\cdot\mathbf{t} + \mathbf{R}(\mathbf{t},\mathbf{n}), \tag{10}$$

where $\mathbf{n}$ represents nuclear coordinates and $\mathbf{t}$ is given implicitly as the solution to these nonlinear equations. Here, dependence on $\mathbf{n}$ is through the integrals and the Fock matrix, both of which depend smoothly on the choice of local orbitals (which in turn depend smoothly on the nuclear geometry). In other words,

$$\mathbf{f}(\mathbf{n}) = \mathbf{f}(\boldsymbol{\eta}(\mathbf{n})), \tag{11}$$

$$\mathbf{I}(\mathbf{n}) = \mathbf{I}(\boldsymbol{\eta}(\mathbf{n})), \tag{12}$$

$$\mathbf{A}^{(d)}(\mathbf{n}) = \mathbf{A}^{(d)}(\mathbf{f}(\boldsymbol{\eta}(\mathbf{n}))), \tag{13}$$

$$\mathbf{R}(\mathbf{t},\mathbf{n}) = \mathbf{R}(\mathbf{t},\mathbf{I}(\boldsymbol{\eta}(\mathbf{n}))),(\mathbf{f}(\boldsymbol{\eta}(\mathbf{n}))). \tag{14}$$

We want to know when can we invert Eq. (9) so that the amplitude vector $\mathbf{t}$ is a differentiable function of $\mathbf{n}$, i.e., $\mathbf{t}=\mathbf{t}(\mathbf{n})$. The answer is given by the implicit function theorem from the field of differential geometry; we must require not only that $\mathbf{g}$ be smooth, but also that the derivative matrix of $\mathbf{g}$ be of full rank (i.e., invertible):[29]

$$\det\left(\frac{\partial \mathbf{g}}{\partial \mathbf{t}}\right) \neq 0. \tag{15}$$

Because CCD potential-energy curves are smooth, Eq. (15) must be satisfied for full CCD (though we cannot prove it exactly). This will be further explained below.

Now let us apply the implicit function theorem to our LCCD method. According to Eqs. (7) and (8), we have modified the amplitude equations so that they now read as

$$0 = \mathbf{g}_b(\mathbf{t},\mathbf{n}), \tag{16}$$

$$\mathbf{g}_b(\mathbf{t},\mathbf{n}) = \mathbf{I}(\mathbf{n}) + \mathbf{A}^{(d)}(\mathbf{n})\cdot\mathbf{t} + \boldsymbol{\Lambda}_b(\mathbf{n})\cdot\mathbf{R}(\boldsymbol{\Lambda}_b(\mathbf{n})\cdot\mathbf{t},\mathbf{n}). \tag{17}$$

Here $\boldsymbol{\Lambda}_b$ is a diagonal matrix of bump functions: $\Lambda_{ijab,klcd}=b_{ijab}\delta_{ik}\delta_{jl}\delta_{ac}\delta_{bd}$. We assume that the off-diagonal linear pieces and the quadratic piece of the LCCD amplitude equations are small compared with the diagonal linear piece. In other words,

$$\frac{\partial \mathbf{g}_b}{\partial \mathbf{t}} \approx \mathbf{A}^{(d)}. \tag{18}$$

We conclude that so long as we keep the diagonal piece of the $\mathbf{A}$ matrix,

$$A_{ijab,ijab}^{(d)} = f_{aa} + f_{bb} - f_{ii} - f_{jj}, \tag{19}$$

then we may smoothly set (or bump) $\mathbf{R}$ to 0 as $\eta_i, \eta_j, \eta_a, \eta_b$ move farther away, just as we do in Eqs. (7) and (8). In the end, by such a procedure, the t amplitudes will still be smooth functions of the nuclear coordinates. This, of course, also justifies why Eq. (15) is true for full CCD.

Before finishing this section, we note that one needs not look at the implicit function theorem as a black box for deciding whether or not the t amplitudes are smooth functions of the nuclear coordinates. In fact, our result is obvious when applied to local MP2 calculations. For MP2, the Hylleraas formulation is

$$\mathbf{g}(\mathbf{t},\mathbf{n}) = \mathbf{A}(\mathbf{n})\cdot\mathbf{t} + \mathbf{I}(\mathbf{n}) = 0. \tag{20}$$

Now, suppose we want to bump the $\mathbf{A}$ matrix and make it more sparse in order to speed up calculations. Let $\tilde{\mathbf{A}}$ be a modified $\mathbf{A}$ form. Clearly, in order to prove that $\mathbf{t}$ is a smooth function of the nuclear coordinates, we need only to know that $\tilde{\mathbf{A}}(\mathbf{n})$ is a smooth function of nuclear coordinates and that $\tilde{\mathbf{A}}^{-1}$ exists [i.e., $\det(\tilde{\mathbf{A}})\neq 0$]. In that case, $\mathbf{t}(\mathbf{n})=-\tilde{\mathbf{A}}^{-1}(\mathbf{n})\cdot\mathbf{I}(\mathbf{n})$. [Furthermore, if $\tilde{\mathbf{A}}$ is diagonally dominant, we really only require that $(\tilde{\mathbf{A}}^{(d)})^{-1}$ exists.] The implicit function theorem is then only a generalization of this result to the case where $\mathbf{g}(\mathbf{t},\mathbf{n})$ is not linear in $\mathbf{t}$.

C. The locality of the smoothed local equations

We now return to Eq. (7). If $b_{ijab}>0$, then t_{ij}^{ab} is a member of the ITA's and we must solve for all ITA's iteratively together. This is a very expensive procedure and we need to construct a meaningful bump function b, for which there are only a linear number of orbital sets $\{\eta_i, \eta_j, \eta_a, \eta_b\}$ such that $b_{ijab}>0$. These ITA's are then analogous to the linear num-

ber of strong pairs in the Pulay-Werner procedure. Below, we describe our choice of bump function, and in the Results section, we show that our bump function works for the case of N_2.

Now when $b_{ijab}=0$, the R_{ijab} term in Eq. (7) is effectively zero and we need not iterate to find t_{ij}^{ab}. We have immediately

$$t_{ij}^{ab} = \frac{\langle \eta_i \eta_j \| \eta_a \eta_b \rangle}{f_{ii} + f_{jj} - f_{aa} - f_{bb}}. \tag{21}$$

This is the most basic approximation for the value of the t_{ij}^{ab} amplitude. It was suggested long ago by Kapuy and co-workers.[22–27] who proposed using local orthonormal occupied and virtual orbitals, and constructing amplitudes of the form shown in Eq. (21). One may then compute a MP2-type energy (which we call KMP2),

$$E_{\mathrm{KMP2}} = \sum_{ijab} t_{ij}^{ab} I_{ijab}, \tag{22}$$

and hope that this energy is close to the MP2 result.

Note that the perturbative equations (21) and (22) are highly dependent on the choice of localized orbitals. Clearly, our local method (like Kapuy and co-workers' KMP2 method) is not invariant to unitary transformations of the orbitals. Furthermore, as Subotnik and Head-Gordon[30] showed, the KMP2 energy is close to the MP2 energy only for orbitals which have been optimized to have small off-diagonal Fock matrix elements. In general, when one employs orbitals localized by standard techniques [e.g., Boys, Pipek-Mezey, and Edmiston-Ruedenberg (ER)], the KMP2 energy is a poor approximation of the MP2 energy.

For this reason, because our algorithm uses standard localized orthonormal orbitals, we compute amplitudes via Eq. (21) only for amplitudes which we expect to be very small. We call these amplitudes PTA's; Eq. (21) is really just a perturbative correction for small amplitudes. Though we expect each PTA to be small, we anticipate the sum of all PTA's to add accuracy to our algorithm.

Our prescription for LMP2 or LCCD is then quite different from that of Pulay, Schütz, and Werner. In a broad sense, both schemes have a hierarchy of amplitudes which are treated with progressively more accuracy. For instance, Werner applies LCCSD to the "strong" amplitudes and LMP2 to "weak" and "distant" amplitudes, and ignores very distant amplitudes. However, the following are the differences between the Pulay-Werner approach and ours.

(1) We work entirely with localized orthonormal orbitals. Pulay and Werner employ localized orthonormal occupied orbitals, but nonorthogonal redundant orbitals for the virtual space.
(2) We have no predetermined domains when calculating potential-energy surfaces. Though our smoothed local amplitude equations allow us to partition the t amplitudes into ITA's and PTA's, this is a dynamic partitioning that is done for every individual geometry. For the Pulay-Werner approach, one must define domains once and for all before calculating potential-energy curves, in order that the surfaces will be smooth, or merge domains together.
(3) Our local CCD equations for the t amplitudes are smooth functions of the nuclear coordinates, i.e., $\mathbf{R}(\mathbf{t},\mathbf{n})$ is a differentiable function. By contrast, the Pulay-Werner approach never writes down general equations for all amplitudes at once. In fact, by picking which amplitudes to correlate with given levels of theory, the analogy of $\mathbf{R}(\mathbf{t},\mathbf{n})$ in their equations would be a discontinuous step function.
(4) For weak amplitudes, we make the most basic KMP2 approximation. According to the Pulay-Werner approach, however, one performs LMP2 for these orbitals. Though we could presumably improve our accuracy by performing LMP2 on weak pairs, by contrast, it is unclear whether the KMP2 approximation would be applicable to the Pulay-Werner model, where one is working with a redundant set of nonorthogonal localized virtual orbitals.
(5) Finally, we repeat that our amplitudes and energies are rigorously smooth functions of nuclear geometry, which will allow analytical gradients to be applicable even when there are large changes in nuclear geometry.

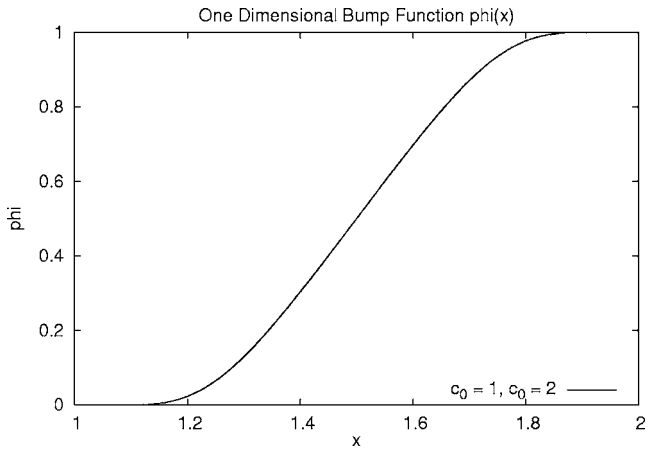

FIG. 1. A plot of the one-dimensional bump function given by Eq. (23) in the text. Here, $c_0=1$ and $c_1=2$.

D. The bump function

We now define our choice of the bump function b_{ijab}. Of course, the form of b is not unique. b_{ijab} can be any smooth function that goes from 1 (when $\{\eta_i, \eta_j, \eta_a, \eta_b\}$ are close) to 0 (when $\{\eta_i, \eta_j, \eta_a, \eta_b\}$ are far apart).

Construction of b requires one-dimesional functions that smoothly go from 0 to 1 over the interval $[c_0, c_1]$. Such functions are common in differential geometry. In our work, we define the one-dimensional bump function ϕ as

$$\phi(x) = 0, \quad x < c_0,$$

$$\phi(x) = \frac{1}{1 + e^{-(c_1-c_0/c_1-x)+(c_1-c_0/x-c_0)}}, \quad x \in (c_0, c_1), \tag{23}$$

$$\phi(x) = 1 \quad x > c_1,$$

where ϕ is infinitely differentiable everywhere. ϕ is plotted in Fig. 1 for $c_0 = 1$ and $c_1 = 2$.

Now, given ϕ, it is easy to construct b_{ijab}. We determine how close or far apart are the orbitals $\{\eta_i, \eta_j, \eta_a, \eta_b\}$ according to their two-electron integrals $[(\eta_i\eta_i|\eta_j\eta_j), (\eta_a\eta_a|\eta_b\eta_b), (\eta_i\eta_a|\eta_i\eta_a)$, and $(\eta_j\eta_b|\eta_j\eta_b)]$. Here, we use the standard notation:

$$(\eta_p\eta_q|\eta_u\eta_v) = \int \eta_p(r_1)\,\eta_q(r_1)\frac{1}{r_{12}}\eta_u(r_2)\,\eta_v(r_2)dr_1dr_2, \tag{24}$$

$$\langle\eta_p\eta_q|\eta_u\eta_v\rangle = \int \eta_p(r_1)\,\eta_q(r_2)\frac{1}{r_{12}}\eta_u(r_1)\,\eta_v(r_2)dr_1dr_2. \tag{25}$$

We could, of course, use other criteria here, including orbital-center separation or even symmetry. Our reasoning, however, is that the size of t_{ij}^{ab} is roughly determined by

$$I_{ijab} = \langle\eta_i\eta_j||\eta_a\eta_b\rangle = \langle\eta_i\eta_j|\eta_a\eta_b\rangle - \langle\eta_i\eta_j|\eta_b\eta_a\rangle. \tag{26}$$

See Eq. (1). Furthermore, by Cauchy-Schwarz,

$$\langle\eta_i\eta_j|\eta_a\eta_b\rangle^2 \leqslant (\eta_i\eta_i|\eta_j\eta_j)(\eta_a\eta_a|\eta_b\eta_b), \tag{27}$$

$$\langle\eta_i\eta_j|\eta_a\eta_b\rangle^2 \leqslant (\eta_i\eta_a|\eta_i\eta_a)(\eta_j\eta_b|\eta_j\eta_b). \tag{28}$$

With this intuition, we have used the following explicit formula for b_{ijab}. For *ia* spin alpha and *jb* spin beta, we define

$$b_{ijab} = \phi_{vv}((\eta_a\eta_a|\eta_b\eta_b))\,\phi_{oo}((\eta_i\eta_i|\eta_j\eta_j))\,\phi_{ov}((\eta_i\eta_a|\eta_i\eta_a) \times(\eta_j\eta_b|\eta_j\eta_b)). \tag{29}$$

For *ijab* of the same spin, we define

$$b_{ijab} = \phi_{vv}((\eta_a\eta_a|\eta_b\eta_b))\,\phi_{oo}((\eta_i\eta_i|\eta_j\eta_j)) \times\phi_{ov}\left(\frac{(\eta_i\eta_a|\eta_i\eta_a)(\eta_j\eta_b|\eta_j\eta_b) + (\eta_i\eta_b|\eta_i\eta_b)(\eta_j\eta_a|\eta_j\eta_a)}{2}\right). \tag{30}$$

Finally, it remains to choose explicit values for the six parameters c_0^{oo}, c_0^{vv}, c_0^{ov}, c_1^{oo}, c_1^{vv}, and c_1^{ov}. The c_0 parameters define how many t amplitudes will be treated iteratively (i.e., the number of ITA's) and hence the speed of our algorithm. The c_1 parameters determine the domain or width of our bump function. In our algorithm, one must decide once and for all on the optimal choice of c_0 and c_1 such that our algorithm will be fast (large c_0), accurate (small c_1), and smooth (large $c_1 - c_0$). These conflicting forces will be discussed further in the Discussion section.

E. Orbital choice

In our LCCD algorithm, we have used only orthonormal localized orbitals (both occupied and virtual). This is in contrast to the traditional local correlation approach due to Pulay. There one invokes localized orthonormal occupied orbitals and the fully redundant set of atomic orbitals (AO's) projected onto the virtual space. We chose orthonormal localized orbitals so the equations would be as simple as possible, without needing to consider a redundant basis.

For our local correlation energy and amplitudes to be smooth functions of nuclear position, it is crucial that the orbitals vary smoothly with nuclear position. See Eqs. (11)–(14). To that end, we used the prescription of Subotnik, Dutoi, and Head-Gordon[31] with one modification. In brief, this algorithm works as follows: One first creates a minimal basis space for the molecule by looking for the subspace of the AO basis that most resembles a STO-3G basis and that exactly contains the occupied space. By definition, this minimal basis space is the direct sum of the occupied space and the valence virtual space. One then projects out this minimal subspace from the AO basis atom by atom, throwing away the smallest orbitals (i.e., those that lie closest to the minimal basis space). The remaining linearly independent orbitals are denoted as atom-centered hard or extra-valence virtuals. These hard-virtual orbitals are orthonormalized using a symmetric orthogonalization, thus keeping them as local as possible. Lastly, one performs separate Boys localizations on the occupied and valence virtual spaces.

In the present context, we implemented one modification of the algorithm by Subotnik, Dutoi, and Head-Gordon.[31] Because the occupied and valence virtual Boys orbitals for N_2 (our example molecule) are not unique and are defined up to a rotational invariance, we could not use a standard Boys localization. We required well-defined, rotationally fixed, smoothly varying orbitals. Hence, we localized the occupied and valence virtual orbitals with a Boys-like localization function that maximized not just the sum of the variances of the orbitals, but also effectively the sum of the squares of their second moments. We called our localization function ζ_{BoysQuad}.

$$\zeta_{\text{Boys}}(\eta_1,\ldots,\eta_m) = \sum_{i=1}^{m}\langle\eta_i|x|\eta_i\rangle^2 + \sum_{i=1}^{m}\langle\eta_i|y|\eta_i\rangle^2 + \sum_{i=1}^{m}\langle\eta_i|z|\eta_i\rangle^2, \tag{31}$$

$$\zeta_{\mathrm{BoysQuad}}(\eta_1,\ldots,\eta_m)=\zeta_{\mathrm{Boys}}+\sum_{i=1}^{m}\langle\eta_i|x^2|\eta_i\rangle^2+\sum_{i=1}^{m}\langle\eta_i|y^2|\eta_i\rangle^2+\sum_{i=1}^{m}\langle\eta_i|z^2|\eta_i\rangle^2. \tag{32}$$

This choice of $\zeta_{\mathrm{BoysQuad}}$ broke rotational symmetry and produced unique, well-isolated, smooth localized orthonormal occupied and valence virtual orbitals. When combined with the smooth hard-virtual space produced by the algorithm in Ref. 31, we had smoothly varying localized orthonormal orbitals that allowed our algorithm to proceed.

One caveat should now be made regarding the procedure (from Ref. 31) used to obtain *globally* smooth, orthonormal, atom-localized hard-virtual orbitals. As discussed in the original article, this method must necessary fail when the AO basis itself is linearly dependent (i.e., there are large numbers of basis functions on every atom). In such a case, one can at best compute well-defined, localized, hard-virtual orbitals that are *locally* smooth functions of nuclear coordinates (with occasional discontinuities). The effects of these discontinuities in the hard-virtual orbitals would necessarily lead to discontinuities in the local correlation energy when the nuclear geometry is changed. The sizes of such discontinuities would presumably be small, however, given that our basis is reaching linear dependence and the most redundant orbitals should be contributing only numerical noise to the correlation energy. Furthermore, we note that provided the AO basis is not linearly dependent, the prescription for localized virtual orbitals in Ref. 31 thus far has been stable and reliable.

IV. THE SKETCH OF AN ALGORITHM

For concreteness, we now sketch how the most simplistic implementation of our algorithm would work.

(1) Localize orbitals into orthonormal occupied (η_i) and virtual (η_a) orbitals using Eq. (32).
(2) Form all integrals $\langle\eta_p\eta_q||\eta_r\eta_s\rangle$ and the Fock matrices f_{ij} and f_{ab}.
(3) Loop through all quartets of significant *ijab* and calculate b_{ijab}.
(4) When $b_{ijab}=0$, set

$$t_{ij}^{ab}=\frac{\langle\eta_i\eta_j||\eta_a\eta_b\rangle}{f_{ii}+f_{jj}-f_{aa}-f_{bb}}.$$

Call these amplitudes $t^{(\mathrm{PTA})}$.
(5) We define the ITA's to be the set of t_{ij}^{ab} for which $b_{ijab}>0$. Solve explicitly and self-consistently for all of the $t^{(\mathrm{ITA})}$ according to Eq. (7).
(6) Calculate

$$E_{\mathrm{LCCD}}=\Sigma_{ijab}t_{ij}^{ab}\langle\eta_i\eta_j||\eta_a\eta_b\rangle=\mathbf{t}^{(\mathrm{ITA})}\cdot\mathbf{I}^{(\mathrm{ITA})}+\mathbf{t}^{(\mathrm{PTA})}\cdot\mathbf{I}^{(\mathrm{PTA})}.$$

Our algorithm should be fast because we restrict the number of ITA's to be very small. Hence, we expect that our LCCD algorithm (like good LMP2 algorithms) should be capable of optimally fast scaling.

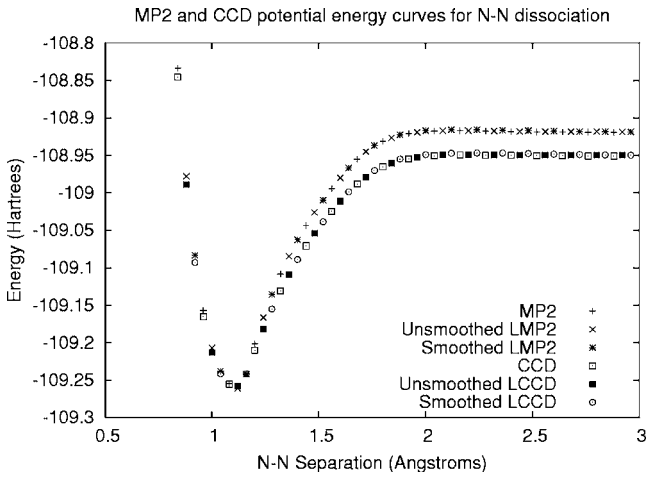

FIG. 2. A plot of the MP2 and CCD energies, full and local, for N_2 separation. Here we plot both bumped (i.e., smoothed) and unbumped (i.e., unsmoothed) energies. The basis is 6-31G*. The calculation is unrestricted. See text for list of c_0 and c_1 parameters.

V. RESULTS

We implemented the algorithm above using the Q-CHEM quantum chemistry package.[32] Thus far in our calculations we have not made an exhaustive attempt to optimize the parameters c_0 and c_1. We have found the following parameterizations to be suitable compromises between speed and accuracy when bumping (i.e., smoothing) amplitude equations:

$$c_0^{oo}=c_0^{vv}=0.2, \tag{33}$$

$$c_1^{oo}=c_1^{vv}=0.25, \tag{34}$$

$$c_0^{ov}=0.001, \tag{35}$$

$$c_1^{ov}=0.004. \tag{36}$$

For nonbumped calculations below (i.e., calculations with a step-function or nonsmooth cutoff), we always used

$$c_0^{oo}=c_0^{vv}=c_1^{oo}=c_1^{vv}=0.2, \tag{37}$$

$$c_0^{ov}=c_1^{ov}=0.001. \tag{38}$$

Because c_0 was constant for bumped and unbumped calculations, both types of calculations ran at the same speed.

In order to assess the accuracy of our LCCD method, we measured the energy of N_2 as the two nitrogen atoms collided and dissociated over the range of 0.8–5 Å. Our LCCD energy was computed after an unrestricted Hartree-Fock solution, and we worked in a 6-31G* basis. Our data is shown in Figs. 2–5.

Figure 2 is a plot of the potential-energy curve for N_2 dissociation calculated by MP2 and CCD (full and local) from 0.5 to 3 Å. Notice that one cannot distinguish between CCD and LCCD (or MP2 and LMP2) on this scale. This demonstrates the accuracy of our LCCD and LMP2 methods. In fact, with our choice of parameters, we calculate that $\max_R|E_{\mathrm{LCCD}}(R)-E_{\mathrm{CCD}}(R)|\approx 0.003$ mhartree. This error is less than 1% of the equilibrium correlation energy. This error

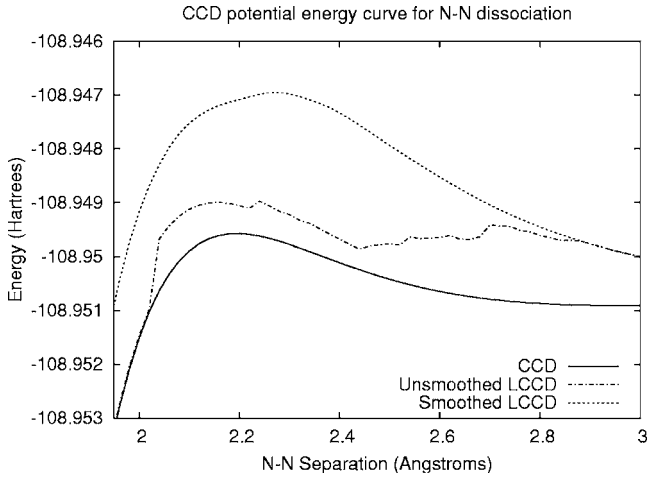

FIG. 3. A plot of the CCD energies, full and local, for N_2 separation. This figure blows up the data from Fig. 2. Here, we focus on the region 1.95–3 Å to demonstrate the discontinuities in the potential-energy surface when the amplitudes are not bumped (i.e, smoothed) and the smoothness of the bumped solution. The basis is 6-31G*. The calculation is unrestricted. See text for list of c_0 and c_1 parameters.

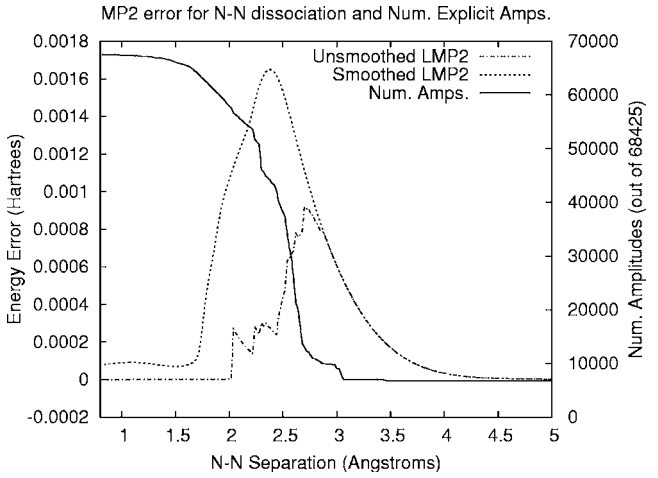

FIG. 5. The error $E_{\mathrm{LMP2}}-E_{\mathrm{MP2}}$ for N_2 separation is plotted on the vertical grid on the left. The number of iterative t amplitudes (ITA's) is plotted on the vertical grid on the right. (It is the same for the smoothed and unsmoothed curves.) See text for list of c_0 and c_1 parameters. Note that the smoothed (i.e., bumped) energy has twice the energy deviation as the unsmoothed (or unbumped) energy—this is the price we pay for a smooth potential-energy surface. The basis is 6-31G*. The calculation is unrestricted.

(while small) is predictably maximized when N_2 is in the process of dissociating. See also Figs. 3 and 4.

Figure 3 zooms in on the CCD curves from Fig. 2 over the domain 1.95–3 Å. This scale allows better resolution of the area where N_2 dissociates and one can see the difference between smoothed and unsmoothed potential-energy curves. When no smooth bump function is applied to accommodate changes in the number of ITA's, there are discontinuities on the order of millihartrees when the N–N distance changes. However, these discontinuities are smoothed away when we use our smooth bump function. See also Fig. 4.

Figure 4 is our most elucidating graphical result. On the left-hand side of the vertical axis, we plot the error $\epsilon_{\mathrm{CCD}}=E_{\mathrm{LCCD}}-E_{\mathrm{CCD}}$ vs N_2 separation. We do this both for a smoothed (or bumped) solution and for a nonsmoothed (or unbumped) solution. On the right-hand side of the vertical axis, we plot the number of ITA's treated as a function of N–N separation. In Fig. 5, we show the exact same graph only for LMP2/MP2 instead of CCD [i.e., we set $\mathbf{R}=0$ in Eq. (7)]. The similarities between these two figures (Figs. 4 and 5) show that our bumping scheme should behave similarly for a variety of different local correlation approximations.

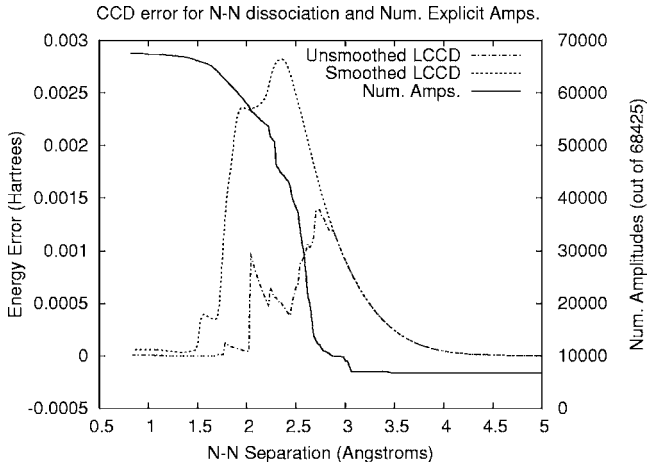

FIG. 4. The error $E_{\mathrm{LCCD}}-E_{\mathrm{CCD}}$ for N_2 separation is plotted on the vertical grid on the left. The number of iterative t amplitudes (ITA's) is plotted on the vertical grid on the right. (It is the same for the smoothed and unsmoothed curves.) See text for list of c_0 and c_1 parameters. Note that the smoothed (i.e., bumped) energy has twice the energy deviation as the unsmoothed (or unbumped) energy—this is the price we pay for a smooth potential-energy surface. The basis is 6-31G*. The calculation is unrestricted.

The potential-energy difference curves in Fig. 4 clearly show that by imposing a bump function, one does enforce smoothness of the t amplitudes while there are discontinuities without such a bump function. Unfortunately, these difference curves also show how the absolute magnitude of the correlation energy error grows when one smoothes out the amplitude equations. In both our CCD and MP2 curves, by bumping the amplitudes, one increases the error (ϵ) by a factor of 2. This result had been anticipated; smoothing out the LCCD equations effectively changes the physics underlying the CCD equations and must lead to larger errors. However, we had always hoped that such effects should be small because only small amplitudes are bumped by our procedure. Indeed, our data clearly supports this hypothesis. Even for the aggressive cutoffs we chose above, absolute errors from the smoothed solution are still in the regime of millihartrees and only twofold larger than the absolute errors of the unbumped discontinuous solution. Of course, for cutoffs slightly smaller than those given above (e.g., $c_0^{oo}=0.1$ instead of $c_0=0.2$), the error ϵ decreases dramatically (usually by orders of magnitude) for both smoothed and unsmoothed LCCD curves. In such a case, the effect of smoothing on the absolute error of the energy becomes miniscule and smoothing is totally unnecessary; when the error is on the order of microhartree's, the discontinuities in the amplitudes are no more than numerical noise and not important. Thus, the trade off between smoothness and absolute energy error is relevant only when aggressive cutoffs are applied

(i.e., large c_0), and, for that case, future work must choose an optimal window size (i.e., $|c_1-c_0|$) to find the right balance.

The plot of the number of ITA's in Fig. 4 demonstrates how very few amplitudes are needed to correctly describe the CCD energy of N_2 at large separation distances when the interatomic interaction becomes weak. For the molecule N_2 in 6-31G*, there are 68 425 t_{ij}^{ab} amplitudes, provided one ignores zero amplitudes of the form t_{ii}^{ab}, t_{ij}^{aa}. At long distances (e.g., N_2 nuclei separated by more than 3.3 Å), our method selected 6806 ITA's, which correspond essentially to every occupied orbital on each N atom correlated to all virtuals on that atom. (There were actually a few intra-atomic t amplitudes that were not considered ITA's because their exchange integrals were too small.) The number of ITA's decreases tenfold as the N–N distance increases from 1.5 to 3 Å. Along this dissociation pathway, our LCCD energy curve remains smooth, showing that one may aggressively cut down the number of ITA's while retaining smoothness of the amplitudes and energies.

VI. DISCUSSION

The N_2 separation curve above demonstrates one chemical example where we generate accurate and smooth potential-energy curves from a limited number of ITA's. It remains to check atomization energies and potential-energy curves on a broader set of molecules.

Notwithstanding our limited data, we may make several reasonable predictions at this time based on previous local correlation theory results and our own experience thus far. Regarding the accuracy of our local CCD energy, we expect that our local correlation scheme, by selecting t amplitudes from integral cutoffs, should be no more or less accurate than the standard Pulay-Schütz-Werner local correlation energies. Both methods apply the standard CCD equations to a limited number of amplitudes. Moreover, just as Schütz and Werner may pick the size of their domains (and larger domains yield more accurate energies), we choose values for c_0 and c_1 (which determine how many amplitudes are ignored and hence the accuracy of our method). Of course, in the same way that the choice of domain size for Schütz and Werner determines the speed of their algorithm, our choice of c_0 will determine the speed of our algorithm. Now, it is likely that our localized orthonormal virtual orbitals will be less local than the nonorthogonal redundant projected AO's that are typically used by Pulay-based methods; and hence we will need to include more amplitudes for the same energy accuracy. We note, however, that because our algorithm does not use redundant sets of virtual orbitals, we automatically have fewer amplitudes than Pulay-based methods and we do not believe orthonormality of our virtual orbitals will be computationally problematic. Thus, in the end, we need only to decide how much energy deviation is acceptable and how fast we require our algorithm to be when choosing our cutoffs. The experience of Schütz and Werner suggests that one can choose rather severe cutoffs and still end up with acceptable relative energies.

Regarding the shape and smoothness of potential-energy curves, one significant concern must now be voiced. Our reasoning demonstrates rigorously that the LCCD curves from this algorithm will be smooth. However, smoothness for a mathematician does not mean smoothness for the practicing chemist. On the one hand, for a mathematician, smoothness is a purely local property which makes sense only in the context of open sets: for example, a bump function may look like a discontinuous step function from afar, but then become rounded at the edges over the interval $(-\epsilon, \epsilon)$. Such a function is mathematically smooth. On the other hand, a practicing chemist, who does geometric minimizations, seeks potential-energy curves that "look" smooth on a fixed length scale, say 0.01Å, over which minimization steps take place. Furthermore, regarding the broader shape of potential energy curves, a chemist wants to avoid artificial maxima or minima arising from a local ansatz. A plethora of artificial stationary points (which can certainly be smooth) would necessarily make geometry optimization unviable. Thus, the goal of our (or any) LCCD method is to be fast, accurate, and smooth over steps of 0.01 Å without the introduction of any spurious maxima or minima. It remains to find the correct parameters to make this happen for a broad array of molecules. From the N_2 data presented here, we are optimistic that such c_0 and c_1 parameters can be found. In particular, because such large cutoffs may be used with N_2 while the error of the method does not grow much, we do expect that one can indeed produce very smooth curves while attaining large speed ups in general.

Armed with lots of theory and a few examples suggesting that one can indeed produce fast, smooth LCCD algorithms, we currently need to implement our algorithm so that it runs as quickly as possible. Given that our technique should require a linear number of ITA's, our algorithm should potentially scale linearly (just as the Schütz-Werner LCCD algorithm does) in the regime where the ITA calculation dominates. Furthermore, for large systems, in the future we will need to introduce a third class of t amplitudes. This will be the class of numerically negligible t amplitudes, where t_{ij}^{ab} is negligible if $|\langle \eta_i \eta_j || \eta_a \eta_b \rangle| < 10^{-\alpha}$ (for some suitable parameter α). For these amplitudes, the Kapuy energy correction should be rigorously small enough so that their inclusion would add only numerical noise. We will then have a linear number of ITA's, a quadratic number of PTA's (quadratic, not linear, because of the slow decay of the Coulomb operator), and a quartic number of negligible amplitudes for large systems.

Finally, further in the future, after efficiently implementing and parameterizing our basic LCCD algorithm, we will want to include the singles correction (for LCCSD) and, subsequently, construct analytical gradients and perform geometry optimizations. After all, the strength of our approach is that our fast local algorithm is also smooth. Hence optimization via analytical gradients should be possible over a global potential-energy surface. This is in contrast to the potential-energy surfaces (PES's) of Werner and co-workers, which are smooth only locally. Now, because we compute the relevant ITA's on the fly, at every single nuclear configuration, a geometry optimization algorithm for our LCCD theory will require more creativity in organization, memory architecture, and input/output in order to achieve the same efficiency that

Werner has achieved for his fixed domain model.[33,34] We are optimistic, however, that such computational obstacles can be overcome and that optimizations over globally smooth potential-energy surfaces from local correlation algorithms will be possible in the future.

VII. CONCLUSIONS

This paper demonstrates that one can write down and implement local CCD and MP2 theories which generate LCCD and LMP2 energies that are smooth functions of nuclear coordinates, with potential for very fast implementation. Such methods should find use in modern correlation calculations where systems are large and require local techniques, but where one would like to calculate optimized geometries. Our hope is that the method proposed in this paper, i.e., bumping the amplitude equations and invoking the implicit function theorem, will inspire a new generation of local correlation theories where one can achieve both speed and smoothness.

ACKNOWLEDGMENTS

We thank Greg Beran, Alex Sodt, and Tony Dutoi for much help and interesting discussions. One of the authors (J.E.S.) thanks Alan Weinstein for mathematical instruction, and S. Breaker Garrett and C. Cakes Chen for their graphical support. This article is dedicated to Jeremy K. Burdett. (J.E.S.) was supported by the Fannie and John Hertz Foundation. This work was supported by a grant from the Department of Energy through the Computational Nanoscience Program. Another author (M.H.G.) is a part owner of Q-Chem, Inc.

[1] P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
[2] S. Saebo and P. Pulay, Chem. Phys. Lett. **113**, 13 (1985).
[3] P. Pulay and S. Saebo, Theor. Chim. Acta **69**, 357 (1986).
[4] S. Saebo and P. Pulay, J. Chem. Phys. **86**, 914 (1987).
[5] S. Saebo and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).
[6] R. B. Murphy, M. D. Beachy, and R. A. Friesner, J. Chem. Phys. **103**, 1481 (1995).
[7] G. Reynolds, T. Martinez, and E. Carter, J. Chem. Phys. **105**, 6455 (1996).
[8] M. Schütz, G. Hetzer, and H. J. Werner, J. Chem. Phys. **111**, 5691 (1999).
[9] H. J. Werner, F. R. Manby, and P. J. Knowles, J. Chem. Phys. **118**, 8149 (2003).
[10] M. Schütz and H. J. Werner, J. Chem. Phys. **114**, 661 (2001).
[11] M. Schütz and H. J. Werner, Chem. Phys. Lett. **318**, 370 (2000).
[12] M. Schütz, J. Chem. Phys. **113**, 9986 (2000).
[13] M. Schütz, J. Chem. Phys. **116**, 8772 (2002).
[14] N. Russ and T. D. Crawford, J. Chem. Phys. **121**, 691 (2004).
[15] M. Schütz, G. Rauhut, and H. J. Werner, J. Phys. Chem. A **102**, 5997 (1998).
[16] J. A. Pople, in *Energy, Structure and Reactivity*, edited by D. W. Smith and W. B. McRae (Wiley, New York, 1973), p. 51.
[17] M. Head-Gordon, P. E. Maslen, and C. A. White, J. Chem. Phys. **108**, 616 (1998).
[18] P. E. Maslen and M. Head-Gordon, J. Chem. Phys. **109**, 7093 (1998).
[19] M. S. Lee, P. E. Maslen, and M. Head-Gordon, J. Chem. Phys. **112**, 3592 (2000).
[20] P. E. Maslen, A. D. Dutoi, M. S. Lee, Y. H. Shao, and M. Head-Gordon, Mol. Phys. **103**, 425 (2005).
[21] P. Y. Ayala and G. E. Scuseria, J. Chem. Phys. **110**, 3660 (1999).
[22] E. Kapuy, Z. Csepes, and C. Kozmutza, Int. J. Quantum Chem. **23**, 981 (1983).
[23] E. Kapuy, F. Bartha, F. Bogar, Z. Csepes, and C. Kozmutza, Int. J. Quantum Chem. **38**, 139 (1990).
[24] E. Kapuy, F. Bogar, and E. Tfirst, Int. J. Quantum Chem. **52**, 127 (1994).
[25] C. Kozmutza, E. Kapuy, E. Evleth, and E. Kassab, J. Mol. Struct. **332**, 141 (1995).
[26] C. Kozmutza, E. Kapuy, E. Evleth, J. Pipek, and L. Trezl, Int. J. Quantum Chem. **57**, 775 (1996).
[27] J. Pipek and F. Bogar, Top. Curr. Chem. **203**, 43 (1999).
[28] E. A. Hylleraas, Z. Phys. **65**, 209 (1930).
[29] J. M. Lee, *Introduction to Smooth Manifolds* (Springer, New York 2002).
[30] J. E. Subotnik and M. Head-Gordon, J. Chem. Phys. **122**, 034109 (2005).
[31] J. E. Subotnik, A. D. Dutoi, and M. Head-Gordon, J. Chem. Phys. (accepted).
[32] J. Kong, M. S. Lee, A. M. Lee *et al.*, J. Comput. Chem. **21**, 1532 (2000).
[33] A. El Azhary, G. Rauhut, P. Pulay, and H. J. Werner, J. Chem. Phys. **108**, 5185 (1998).
[34] M. Schütz, H. J. Werner, R. Lindh, and F. Manby, J. Chem. Phys. **121**, 737 (2004).