Noniterative local second order Mo/ller–Plesset theory: Convergence with local correlation space

P. E. Maslen and M. Head-Gordon

Noniterative local second order Møller–Plesset theory: Convergence with local correlation space

P. E. Maslen[a)] and M. Head-Gordon[b)]
Department of Chemistry, University of California, Berkeley, Berkeley, California 94720

We extend our noniterative local correlation method [P. E. Maslen and M. Head-Gordon, Chem. Phys. Lett., **283**, 102 (1998)] by defining a hierarchy of local spaces, ranging from small to large. The accuracy of the local method is then examined as a function of the size of the local space. A medium size local space recovers 98% of the MP2 correlation energy, and reproduces fine details of the potential energy surface such as rotational barriers with an RMS error of 0.2 kcal/mol and a maximum error of 0.4 kcal/mol. A large local space recovers 99.5% of the correlation energy and yields rotational barriers with a RMS error of 0.05 kcal/mol and a maximum error of 0.1 kcal/mol, at significantly increased computational cost. © *1998 American Institute of Physics.* [S0021-9606(98)31038-7]

I. INTRODUCTION

Over the past fifteen years there has been a concerted attempt to reduce the cost of Møller–Plesset perturbation theory calculations by employing localized molecular orbitals. In terms of canonical orbitals, second order Møller–Plesset theory (MP2) has computational requirements which scale as the fifth power of molecular size. MP3 and MP4 scale as the sixth and seventh powers. This steep nonlinear cost precludes application to very large molecules, and as it has no physical origin, the development of alternative formulations based on local quantities is both feasible and desirable.

There are two distinct stages in the formulation of a local Møller–Plesset method. First, the Møller–Plesset equations must be reformulated using arbitrary orbitals in place of the usual canonical molecular orbitals. Pulay and Saebø have derived the Møller–Plesset equations in a basis of localized orbitals,[1–5] and their local correlation approach has subsequently been extended by others.[6–10] The Pulay and Saebø formulation allows the virtual orbitals to be nonorthogonal, but orthogonality is maintained among the occupied orbitals. It is sometimes advantageous to use nonorthogonal occupied orbitals, and we have recently formulated MP2, CCSD and CCSD(T) using arbitrary nonorthogonal occupied and virtual orbitals.[11,12]

In the second stage, the cost of the local Møller–Plesset calculation is reduced by working in a basis of localized molecular orbitals and discarding selected nonlocal terms, which are assumed to be small. When devising the local selection criterion it is helpful to view electron correlation as causing transient excitations from the HF reference to various excited states. These excitations tend to be spatially *local*, in that a pair of electrons in given occupied orbitals are usually promoted to virtual orbitals in the immediate vicinity. Excitations to distant virtual orbitals can be neglected,

leading to significant computational savings if localized orbitals are employed.

The Pulay and Saebø scheme for selecting local excitations provides an efficient route for large-scale Møller–Plesset calculations, and Rauhut *et al.* have recently reported MP2 calculations involving 1000 basis functions.[13] However, their method cannot readily be extended to CCSD(T), because their algorithm requires storage of the local excitation amplitudes, and there are an unmanageably large number of local triple amplitudes in CCSD(T).

We have recently proposed a local MP2 method which does not require storage of the amplitudes,[12] and which could readily be extended to CCSD(T).[14] In this article we examine the accuracy of our local MP2 results as a function of the size of the local correlation space. This paper is organized as follows: In Sec. II the MP2 energy expression is formulated in terms of arbitrary nonorthogonal orbitals. The Pulay and Saebø method for selecting a local correlation space is reviewed and compared with our local ansatz. We also elaborate on our local ansatz by defining a hierarchy of local correlation spaces, ranging from small to large. In Sec. III our local MP2 results are compared with conventional MP2, using small, medium and large local correlation spaces. The behavior of the local ansatz in the large molecule limit is probed using the linear alkanes and linear polyenes as test cases, and the ability of the local ansatz to reproduce fine details of the potential surface is assessed by examining the rotational barriers of the substituted ethylenes and substituted benzenes.

II. REVIEW OF MP2 AND LOCAL MP2

A. MP2 correlation space

In MP2 theory the correlation space is the space of all doubly-excited determinants. We denote these determinants by $|aibj\rangle$, where $i,j,k,\ldots$, are occupied spin-orbitals, $a,b,c,\ldots$, are virtual spin-orbitals, and $|aibj\rangle$ is obtained by replacing i and j with a and b in the HF wavefunction

---

[a)] Electronic mail: maslen@bastille.cchem.berkeley.edu
[b)] Electronic mail: mhg@bastille.cchem.berkeley.edu

$\Psi^{(0)}$. The MP2 correction to the HF wavefunction, $\Psi^{(1)}$, may be expanded in the basis of doubly-excited determinants,

$$\Psi^{(1)}=\sum_{aibj} T^{aibj}|aibj\rangle, \tag{1}$$

where T^{aibj} is simply the coefficient of $|aibj\rangle$, and Σ_{aibj} denotes an unrestricted sum.

The occupied and virtual orbitals may be localized or delocalized, orthogonal or nonorthogonal and linearly independent or overcomplete. The only essential restriction is that $i,j,\ldots$, must lie entirely within the occupied space while $a,b,\ldots$, must lie entirely within the virtual space.[11]

B. MP2 energy

The MP2 energy is invariant to the choice of orbitals, provided that the occupied orbitals span the occupied space and the virtual orbitals span the virtual space.[11] The MP2 energy may be obtained by minimizing the Hylleraas energy functional[15] with respect to $\Psi^{(1)}$,

$$E[\Psi^{(1)}]=\langle\Psi^{(1)}|\hat{F}-E^{(0)}|\Psi^{(1)}\rangle+\langle\Psi^{(1)}|\hat{V}|\Psi^{(0)}\rangle$$
$$+\langle\Psi^{(0)}|\hat{V}|\Psi^{(1)}\rangle, \tag{2}$$

where $\hat{F}$ is the Fock operator, $E^{(0)}=\langle\Psi^{(0)}|\hat{F}|\Psi^{(0)}\rangle$ and $\hat{V}$ is the electron repulsion operator, $1/r_{12}$. Evaluating the determinantal matrix elements one obtains

$$E[\vec{T}]=\vec{T}\cdot\boldsymbol{\Delta}\cdot\vec{T}+\vec{T}\cdot\vec{V}. \tag{3}$$

In this expression $\vec{T}$ is a vector containing the amplitudes T^{aibj} and $\vec{V}$ is a vector containing the electron repulsion integrals V_{aibj},

$$V_{aibj}=(ai|bj)-(bi|aj), \tag{4}$$

where

$$(ai|bj)=\int\int d\mathbf{r}_1 d\mathbf{r}_2 a(1)i(1)1/r_{12}b(2)j(2). \tag{5}$$

If a canonical basis (i.e., one which diagonalizes the Fock matrix) is used, then $\boldsymbol{\Delta}$ is a huge diagonal matrix with diagonal elements $(F_{aa}-F_{ii}+F_{bb}-F_{jj})$. In an arbitrary nonorthogonal basis the expression for $\boldsymbol{\Delta}$ becomes[12]

$$\Delta_{aibj;a'i'b'j'}=F_{aa'}S_{ii'}S_{bb'}S_{jj'}-S_{aa'}F_{ii'}S_{bb'}S_{jj'}$$
$$+S_{aa'}S_{ii'}F_{bb'}S_{jj'}-S_{aa'}S_{ii'}S_{bb'}F_{jj'}, \tag{6a}$$

where $\mathbf{S}$ and $\mathbf{F}$ are the overlap and Fock matrices in the molecular orbital basis. The plethora of indices can be avoided by expressing $\boldsymbol{\Delta}$ in terms of matrices rather than individual components,

$$\boldsymbol{\Delta}=F_v\otimes S_o\otimes S_v\otimes S_o-S_v\otimes F_o\otimes S_v\otimes S_o$$
$$+S_v\otimes S_o\otimes F_v\otimes S_o-S_v\otimes S_o\otimes S_v\otimes F_o, \tag{6b}$$

where F_v denotes the matrix representation of the Fock operator in the virtual basis and F_o refers to the occupied basis.

For future reference we rewrite the expression for $\boldsymbol{\Delta}$ in the factorized form,

$$\boldsymbol{\Delta}=F^2\otimes S^2+S^2\otimes F^2, \tag{7}$$

where

$$F^2_{ai,a'i'}=F_{aa'}S_{ii'}-S_{aa'}F_{ii'} \tag{8}$$

and

$$S^2_{ai,a'i'}=S_{aa'}S_{ii'}. \tag{9}$$

$E[\vec{T}]$ is a quadratic function of $\vec{T}$ and hence has a single minimum corresponding to

$$\vec{T}=-\frac{1}{2}\boldsymbol{\Delta}^{-1}\cdot\vec{V}, \tag{10}$$

and

$$E=-\frac{1}{4}\vec{V}\cdot\boldsymbol{\Delta}^{-1}\cdot\vec{V}. \tag{11}$$

If a canonical basis is used then $\boldsymbol{\Delta}$ is diagonal and so it is easy to form $\boldsymbol{\Delta}^{-1}$. If another basis (such as localized orbitals) is used then $\boldsymbol{\Delta}$ is not diagonal and $\boldsymbol{\Delta}^{-1}\cdot\vec{V}$ is usually evaluated via an iterative solution of

$$\boldsymbol{\Delta}\cdot\vec{T}=-\frac{1}{2}\vec{V}. \tag{12}$$

C. The Pulay and Saebø local correlation space

The cost of an MP2 calculation may be considerably reduced if the correlation space is restricted to a small subspace of the doubly excited determinants. One may reasonably expect that the dominant contribution to the correlation energy comes from determinants $|aibj\rangle$ where a and b lie in close proximity to the occupied orbitals i and j, and this provides an efficient selection criterion if the orbitals are localized.

In the local correlation method devised by Pulay and Saebø,[1–5,13,16–21] one retains all $|aibj\rangle$ such that a lies near i and b near j, or both a and b lie near i or both a and b lie near j. This method has been extensively tested and has the virtue that it recovers a large fraction of the correlation energy while significantly reducing the basis set superposition error (BSSE).[22] Pulay and Saebø argue convincingly that BSSE arises from determinants $|aibj\rangle$ where a or b is well separated from both i and j, and these determinants are excluded from their local method.

D. Iterative calculation of the local correlation energy

In a local treatment of electron correlation only the amplitudes T^{aibj} and integrals V_{aibj} corresponding to determinants $|aibj\rangle$ in the local correlation space need be calculated, so the length of the amplitude and integral vectors ($\vec{T}$ and $\vec{V}$) are considerably shortened and the side-length of the matrix $\boldsymbol{\Delta}$ is reduced by the same amount.

The local correlation ansatz introduces two complications into the calculation of the MP2 energy. First, nonca-

nonical localized orbitals are employed because canonical orbitals are too delocalized. As a consequence $\mathbf{\Delta}$ is not diagonal and the formation of $\mathbf{\Delta}^{-1}$ becomes a nontrivial problem.

Second, in a full (nonlocal) calculation the correlation space is of the form $\{a\}\otimes\{i\}\otimes\{b\}\otimes\{j\}$, and as a consequence the set of doubly-excited determinants $|aibj\rangle$ form a fourth rank tensor. This tensorial structure is exploited in the conventional transformation of $\vec{V}$ from the atomic-orbital basis to the molecular-orbital basis, enabling the four indices to be transformed one at a time. In contrast, the subspace of determinants lying within the Pulay and Saebø local space do not form a fourth rank tensor, and devising an efficient four-index transformation for local correlation calculations is a considerable challenge.

The local ansatz also destroys the tensorial structure of $\mathbf{\Delta}$, which can no longer be expressed in the direct-product form of Eq. (7). Because $\mathbf{\Delta}$ is not diagonal and has lost its tensorial structure, Pulay and Saebø calculate $\mathbf{\Delta}^{-1}\cdot\vec{V}$ iteratively via Eq. (12). Although this does not significantly add to the cost of the calculation, which is dominated by the four-index transformation, it does mean that the local amplitudes must be stored. This precludes the straightforward extension of the local method to CCSD(T), because there are generally an unfeasibly large number of local triples amplitudes to be stored.

E. A noniterative local MP2 method

In order to devise a noniterative local MP2 (LMP2) method, we have used an alternative local ansatz which retains the direct-product structure of Eq. (7). By exploiting this direct-product structure it is possible to invert $\mathbf{\Delta}$ explicitly without iteration.[12]

Our local basis is defined as follows. First we choose all pairs $\{ai\}$ where a lies sufficiently close to i. The local basis $|aibj\rangle$ then consists of all possible pairs of pairs, and has the direct-product structure,

$$\{aibj\}=\{ai\}\otimes\{bj\}. \tag{13}$$

As a consequence of this choice the set of local doubly excited determinants form a second-rank tensor. In this local basis a lies near i, and b lies near j. Compared with the Pulay and Saebø local basis our basis is lacking ''ionic'' terms where both a and b lie near i, or both a and b lie near j. These terms are important for a short-range correlation, but not for a long-range correlation. To mitigate the deficiency in our ansatz we enlarge the local pair-basis $\{ai\}$ somewhat by increasing the cutoff distance which determines whether a lies near i.

This enlarged basis includes the short-range ionic terms, but also contains a significant number of nonlocal long-range terms which add to the cost of the calculation without improving the accuracy. Ideally one would use a large local-correlation space for a short-range correlation and a smaller local-correlation space for a long-range correlation, but it is not possible to do this without destroying the tensorial structure of the basis [Eq. (13)]. Instead, we use the enlarged basis

for both short- and long-range correlations and aim to offset the additional cost with the savings which accrue from the tensorial structure of our local basis.

In order to invert $\mathbf{\Delta}$ we first construct S^2 and F^2 in Eq. (7) in the local pair-basis $\{ai\}$. These matrices are small enough to be diagonalized explicitly, provided the local pair-basis is considerably smaller than the full nonlocal pair-basis. By transforming to a basis where S^2 and F^2 are simultaneously diagonal $\mathbf{\Delta}$ is also brought into diagonal form, and formation of $\mathbf{\Delta}^{-1}$ is then a trivial operation.

The resulting local MP2 method is variational, i.e., it recovers no more than 100% of the full MP2 correlation energy.

F. Size of the local correlation space

In a previous article we reported LMP2 energies for the substituted ethenes and benzenes.[12] While those calculations produced highly accurate results, recovering over 99.8% of the correlation energy, the local correlation space was fairly large and hence the calculations were relatively expensive. In this work we systematically vary the size of the local space and examine the accuracy of the LMP2 results as a function of the size of the local basis.

In order to define the local correlation space we require a set of occupied orbitals, a set of virtual orbitals and a prescription for choosing local pairs $\{ai\}$. The set of doubly-excited determinants $|aibj\rangle$ is then defined via Eq. (13). Both the occupied and the virtual bases are defined atom by atom, as described below. We then retain all local pairs $\{ai\}$ where both a and i belong to the same atom.

The virtual basis $\{a\}$ for a given atom is obtained by projecting all atomic orbitals $\{\alpha\}$ belonging to that atom onto the virtual space,

$$|a\rangle=(\hat{1}-\hat{P})|\alpha\rangle, \tag{14}$$

where $\hat{P}$ is a projection operator defining the occupied space. We note in passing that virtual orbitals on different atoms are not orthogonal, and the virtual basis contains several exact linear dependencies. Indeed, in the large-basis limit where the atomic-orbital basis on each atom is complete, the virtual basis on a given atom spans the virtual space for the entire molecule, and our LMP2 algorithm yields the same energy as a full nonlocal MP2 calculation.

In order to vary the size of the local space, the set of occupied orbitals associated with a given atom is systematically increased from small to large. The method is outlined here and the details are described in the Appendix.

To obtain a small set of occupied orbitals for a given atom, we first localize all the occupied orbitals about the current atom, and then select those orbitals with a sufficiently large Mulliken charge on this atom. This usually yields a ''minimal basis,'' i.e., one occupied orbital per hydrogen atom, five per carbon atom, etc. The localization method is noniterative, requiring a single matrix diagonalization.

To obtain a medium-size basis we additionally consider those atoms bonded to the current atom, which we refer to as nearest neighbors. The Mulliken charge is then defined as the
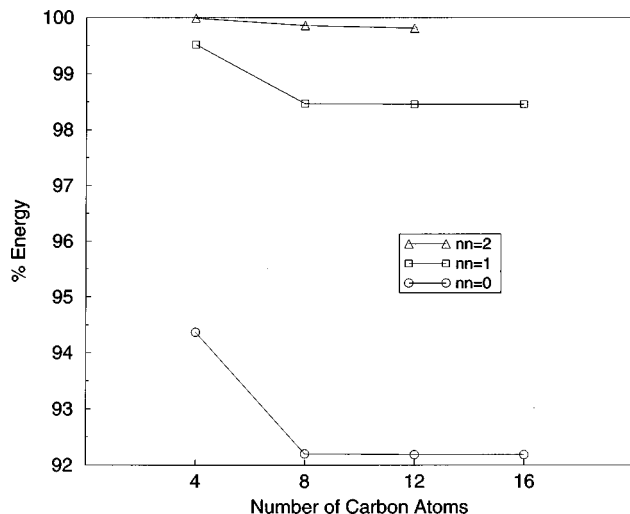
FIG. 1. % correlation energy per additional C_4H_m unit for the linear alkanes $C_{4n}H_{8n+2}$. The local MP2 calculations were performed using small ($nn=0$), medium ($nn=1$) and large ($nn=2$) local correlation spaces. The geometries are not fully optimized, but are close to the equilibrium structures. The calculations use a 6-31G* basis, and all electrons are correlated, including the $1s$ orbitals.

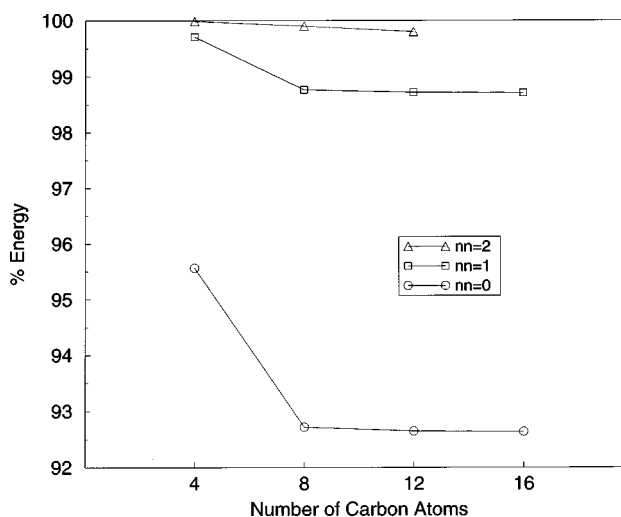
FIG. 2. % correlation energy per additional C_4H_m unit for the linear polyenes $C_{4n}H_{4n+2}$. The local MP2 calculations were performed using small ($nn=0$), medium ($nn=1$) and large ($nn=2$) local correlation spaces. The geometries are not fully optimized, but are close to the equilibrium structures. The calculations use a 6-31G* basis, and all electrons are correlated, including the $1s$ orbitals.

total charge on the current atom and its nearest neighbors. The occupied-orbitals are localized so as to maximize this enlarged Mulliken charge, and those with a sufficiently large charge are retained. To obtain a large occupied basis we include the current atom, the nearest neighbors and the nearest neighbors of the nearest neighbors (the second-nearest neighbors).

These three correlation spaces may be compactly denoted as $nn=0$ (small), $nn=1$ (medium) and $nn=2$ (large). The exact result corresponds to the largest neighbor value possible for the molecule, nn_{max}. Convergence of the LMP2 results can thus be assessed by increasing the nn parameter.

III. RESULTS

A. MP2 correlation energy in the linear alkanes and polyenes

Local correlation methods work artificially well for small molecules, where nn_{max} is typically equal to 1 or 2, and the effects of the local approximation only become apparent as the molecule size is increased. In order to determine the accuracy of our local ansatz in the large-molecule limit, we have studied the linear alkanes, C_nH_{2n+2}, and the linear polyenes, $C_{2n}H_{2n+2}$.

Figures 1 and 2 show the percentage of the correlation energy recovered by LMP2 for every fourth alkane and polyene, from C_4 to C_{16}, using a 6-31G* basis. In order to exhibit the behavior of LMP2 with respect to increasing molecule size, Figs. 1 and 2 are based on the correlation energy increment per additional C_4H_m unit, rather than the total correlation energy. The quantity of interest is the limiting value of correlation energy recovery as the chain length is increased. This limit is reached for a chain length of eight carbon atoms for $nn=0$ and $nn=1$, while $nn=2$ is slightly slower to converge.

The fraction of the correlation energy recovered by the LMP2 schemes varies with the size of the local correlation space. For $nn=0$, 1 and 2 the fraction of correlation energy recovered is $>92\%$, $>98\%$ and $>99.8\%$, respectively, exhibiting exponential convergence with respect to nn. Interestingly, the LMP2 calculations perform equally well on the alkanes and the polyenes, even though one might expect the polyenes to be more delocalized than the alkanes. This exponential convergence is essentially a result of the exponential decay of the density matrix with distance in insulators, and should thus be a general phenomenon in molecules.

The cost of the LMP2 calculations depends on the number of local pair functions, N_{pair}; inversion of Δ [Eqs. (7) and (11)] requires the diagonalization of two $N_{pair} \times N_{pair}$ matrices, S^2 and F^2, and the cost of a local four-index transformation also increases with N_{pair}. Figures 3 and 4, together with Table I, display the number of local pair functions used in each LMP2 calculation, and the number of pair functions employed in standard MP2.

On a current workstation diagonalization of a 1000×1000 matrix takes less than a minute, while diagonalization of a $10,000 \times 10,000$ matrix takes several hours, so we regard a calculation involving 1000 pair functions as small and a calculation involving 10,000 pair functions as large. By extrapolating Figs. 3 and 4 we can determine the largest molecule which could be treated with LMP2 using 10,000 pair functions. For $nn=0$, 1 and 2 correlation spaces the largest molecules are $C_{125}H_m$, $C_{50}H_m$ and $C_{25}H_m$, respectively, using a 6-31G* basis. These estimates are based on all-electron calculations. If the $1s$ occupied orbitals are excluded from the correlation calculation then the maximum molecule size increases to $C_{150}H_m$, $C_{60}H_m$ and $C_{30}H_m$, respectively. The $nn=0$ and 1 correlation spaces are clearly of
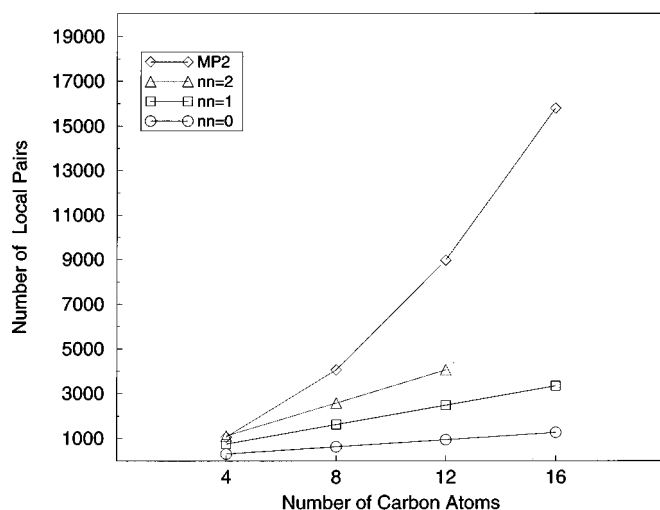
FIG. 3. The size of the local pair basis $\{ai\}$ for the linear alkanes $C_{4n}H_{8n+2}$. The local MP2 calculations were performed using small ($nn=0$), medium ($nn=1$) and large ($nn=2$) local correlation spaces. The calculations use a 6-31G* basis, and all electrons are correlated, including the $1s$ orbitals.

most interest, while calculations using $nn=2$ will generally be very expensive.

B. Rotational barriers in the substituted ethylenes and substituted benzenes

While LMP2 calculations using an $nn=2$ local space can recover 99.8% of the MP2 correlation energy, it is not obvious that this will be adequate for properties which depend on fine details of the potential surface, such as rotational barriers. For comparison, a typical rotational barrier of 1 kcal/mol is equivalent to only 0.1% of the correlation energy.
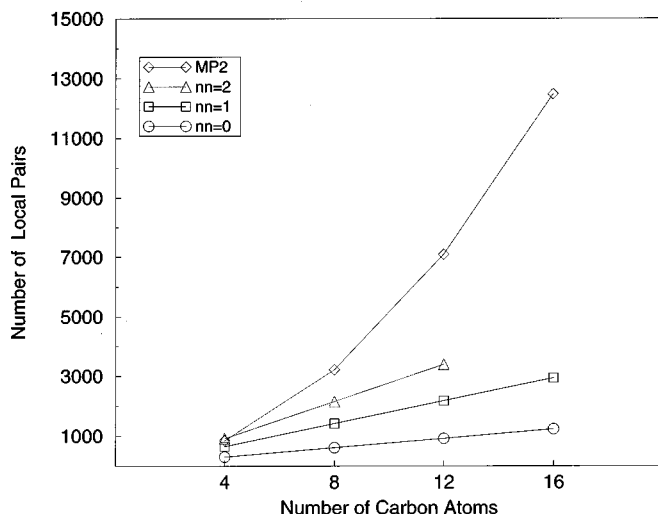

FIG. 4. The size of the local pair basis $\{ai\}$ for the linear polyenes $C_{4n}H_{4n+2}$. The local MP2 calculations were performed using small ($nn=0$), medium ($nn=1$) and large ($nn=2$) local correlation spaces. The calculations use a 6-31G* basis, and all electrons are correlated, including the $1s$ orbitals.

TABLE I. Size of the local pair basis $\{ai\}$ for the linear alkanes and linear polyenes.

Molecule	Ne^b	N^c	$N_{pair}{}^d$		
			$nn=0$	$nn=1$	$nn=2$
C_4H_{10}	34	80	320	760	1112
C_8H_{18}	66	156	636	1620	2580
$C_{12}H_{26}$	98	232	952	2480	4048
$C_{16}H_{34}$	130	308	1268	3340	
C_4H_6	30	72	312	660	922
C_8H_{10}	58	140	620	1420	2158
$C_{12}H_{14}$	86	208	928	2180	3394
$C_{16}H_{18}$	114	276	1236	2940	

[a]A 6-31G* basis was used. All electrons were correlated, including the $1s$ orbitals.
[b]Number of electrons.
[c]Number of atomic orbitals.
[d]Number of local pair-functions $\{ai\}$. The LMP2 calculations were performed using $nn=0$, $nn=1$ and $nn=2$ local-correlation spaces.

To test whether our local ansatz recovers a constant fraction of the correlation energy across the potential surface, we have calculated the rotational barriers for the substituted ethylenes and substituted benzenes, using a 6-311G** basis. Table II contains the absolute energies of the equilibrium conformations while Table III contains the rotational barriers. The LMP2 barriers are more accurate for the substituted ethylenes than for the substituted benzenes; this is to be expected since the substituted ethylenes are smaller and so the local approximation has less effect ($nn_{max}\approx3$).

The LMP2 barriers calculated with an $nn=2$ local space are very accurate for all molecules in the table, with a RMS error of 0.05 kcal/mol and a maximum error of 0.1 kcal/mol. We note in passing that the LMP2 calculations which we reported previously[12] correspond approximately to the $nn=2$ calculations in the current work.

The $nn=0$ barriers are wildly in error, with a RMS error of 1.6 kcal/mol and a maximum error of 4 kcal/mol. The small local space lacks the "ionic" configurations, and the current results indicate that these ionic terms are very important for relative energies as well as for absolute energies.

The $nn=1$ rotational barriers are in good agreement with MP2 for the substituted ethylenes, with a maximum error of 0.2 kcal/mol. They do not perform quite so well for the substituted benzenes. The rotational barriers are in semi-quantitative agreement with MP2, with a RMS error for the substituted benzenes of 0.25 kcal/mol, but the maximum error is 0.4 kcal/mol.

We have also performed several exploratory LMP2 calculations employing Pipek–Mezey occupied orbitals.[24,25] The $nn=1$ rotational barriers calculated using Pipek–Mezey orbitals are less accurate than those described above, with errors of up to 0.56 kcal/mol for the substituted ethylenes and larger errors for the substituted benzenes, so we do not discuss them further.

IV. CONCLUSION

We have previously developed a local MP2 method which is noniterative and does not require the storage of the

TABLE II. Absolute energies for the substituted ethylenes and substituted benzenes: HF and correlation contributions.[a]

Molecule conformation	HF	MP2 (full)	MP2 (local)[b]		
			$nn=0$	$nn=1$	$nn=2$
Vi–C_2H_3 trans	−154.960121	−0.6358153	−0.604408 (95.06)	−0.632642 (99.50)	−0.635614 (99.96)
Vi–CFO trans	−289.718733	−0.8935754	−0.860216 (96.26)	−0.888521 (99.43)	−0.893323 (99.97)
Vi–CHO trans	−190.812152	−0.6794832	−0.649911 (95.64)	−0.676504 (99.56)	−0.679304 (99.97)
Vi–NH_2 pyramidal	−133.103964	−0.5236071	−0.497971 (95.10)	−0.521618 (99.62)	−0.523549 (99.98)
Vi–NO_2 planar	−281.578165	−0.9474502	−0.905774 (95.60)	−0.942183 (99.44)	−0.947002 (99.95)
Vi–OH syn	−152.937831	−0.5399657	−0.517413 (95.82)	−0.538452 (99.71)	−0.539915 (99.99)
Ph–C_2H_3 planar	−307.653473	−1.2488944	−1.160028 (92.88)	−1.232288 (98.67)	−1.247140 (99.86)
Ph–CFO planar	−442.418148	−1.5077998	−1.408201 (93.39)	−1.486873 (98.61)	−1.504066 (99.75)
Ph–CHO planar	−343.511145	−1.2923023	−1.199674 (92.83)	−1.275399 (98.69)	−1.290246 (99.84)
Ph–NH_2 pyramidal	−285.800480	−1.1350916	−1.052182 (92.69)	−1.120332 (98.69)	−1.134030 (99.90)
Ph–NO_2 planar	−434.276812	−1.5637473	−1.453595 (92.95)	−1.540225 (98.49)	−1.557855 (99.62)
Ph–OH planar	−305.634634	−1.1500996	−1.070010 (93.03)	−1.136271 (98.79)	−1.149326 (99.93)

[a]In atomic units; 1 a.u.=627.5 kcal/mol. The results are for the lowest energy conformation of each molecule. The geometries, from Ref. 23, are optimized at the HF/6-31G* level. A larger 6-311G** basis was used for the HF, MP2, and LMP2 energy calculations. The MP2 and LMP2 calculations correlated all electrons, including the $1s$ orbitals.

[b]Percentage relative to the full calculation in parentheses. The LMP2 calculations were performed using $nn=0$, $nn=1$ and $nn=2$ local-correlation spaces.

MP2 amplitudes.[12] The current work examines the accuracy of the local method as a function of the size of the local correlation space for the linear alkanes, polyenes, substituted ethylenes and substituted benzenes.

The local correlation space is defined by increasing the range over which occupied functions are chosen to combine with the virtual functions on a given atom, via a neighbor number parameter, nn. $nn=0$ corresponds to atomic excitations from an occupied minimal basis to the virtuals derived from all atomic orbitals on that atom. $nn=0$ recovers 92% of the correlation energy, and appears inadequate for chemical applications. The remaining 8% of the correlation energy arises from ionic configurations corresponding to electron transfer, and is very important for the description of rotational barriers and other fine details of the potential surface.

A medium size correlation space ($nn=1$) recovers 98% of the correlation energy, which may be adequate for many purposes but is marginally accurate for the prediction of rotational barriers, where it introduces a RMS error of 0.2 kcal/mol and a maximum error of 0.4 kcal/mol ($\sim$0.8 m-Hartree). A large correlation space ($nn=2$) recovers 99.5% of the correlation energy and reduces the RMS error in the rotational barriers to 0.05 kcal/mol, but at a significantly increased computational cost.

TABLE III. Rotational barriers for the substituted ethylenes and substituted benzenes: HF barrier and MP2 correlation correction.[a]

Molecule	Description[b]	HF	MP2 (full)	MP2 (local)[c]		
				$nn=0$	$nn=1$	$nn=2$
Vi–C_2H_3	rotation	6.058	−0.288	−0.316	−0.188	−0.263
Vi–C_2H_3	gauche	3.182	−0.529	−0.264	−0.334	−0.446
Vi–CFO	rotation	7.604	−0.930	−2.291	−1.192	−0.926
Vi–CHO	rotation	8.006	−0.085	−1.361	−0.290	−0.101
Vi–NH_2	rotation	6.477	0.806	1.075	0.753	0.825
Vi–NO_2	rotation	6.955	−1.834	−2.712	−1.826	−1.812
Vi–OH	rotation	4.573	0.998	1.140	0.944	0.985
Vi–OH	anti	1.991	0.059	0.133	0.060	0.053
Ph–C2H3	rotation	2.738	−0.506	0.188	−0.224	−0.569
Ph–CFO	rotation	8.744	−1.423	−5.225	−1.840	−1.420
Ph–CHO	rotation	8.809	−0.494	−4.017	−0.771	−0.557
Ph–NH2	rotation	3.746	0.571	1.723	0.783	0.526
Ph–NH2	inversion	1.458	0.741	0.026	0.628	0.770
Ph–NO2	rotation	7.154	−2.314	−4.192	−2.525	−2.437
Ph–OH	rotation	2.570	0.809	1.177	0.897	0.786
RMS error				1.576	0.201	0.049
Max. error				3.802	0.417	0.123

[a]In kcal/mol. The geometries, from Ref. 25, are optimized at the HF/6-31G* level. A larger 6-311G** basis was used for the HF, MP2, and LMP2 rotational barrier calculations. The MP2 and LMP2 calculations correlated all electrons, including the $1s$ orbitals.

[b]See Ref. 25 for a detailed description of the conformations.

[c]The LMP2 calculations were performed using $nn=0$, $nn=1$ and $nn=2$ local-correlation spaces.

ACKNOWLEDGMENTS

APPENDIX: LOCALIZED OCCUPIED ORBITALS

The occupied basis for a given atom A is obtained by selecting localized occupied orbitals with a Mulliken population exceeding a threshold of 0.1 a.u. on atom A. If a larger set of occupied orbitals is required then we consider the sum of the Mulliken populations of several atoms centered on A.

To obtain the localized orbitals we begin with the usual expression for the molecular charge in the atomic orbital basis,

$$\text{charge} = \text{Tr}(PS), \tag{A1}$$

where P is the projection operator defining the occupied space and S is the overlap matrix. By making use of idempotency, $P = PSP$, and the identity $\text{Tr}(AB) = \text{Tr}(BA)$, Eq. (A1) can be rewritten as

$$\text{charge} = \text{Tr}(S^{-1/2}SPWPSS^{-1/2}), \tag{A2}$$

where W is equal to the overlap matrix S.

If W, initially equal to S, is set to zero except for a single atomic block on atom A, W_{AA}, then Eq. (A2) yields the Mulliken charge on atom A. Similarly, if W is set to zero except for the off-diagonal atomic blocks W_{AB} and W_{BA} then Eq. (A2) yields the Mulliken bonding charge connecting atoms A and B.

The matrix $S^{-1/2}SPWPSS^{-1/2}$ may be diagonalized, yielding a series of eigenvalues, which are Mulliken charges, and eigenvectors in the orthogonalized atomic orbital basis. (The orthogonalized atomic orbitals ϕ' are related to the atomic orbitals ϕ by $\phi' = S^{-1/2}\phi$.)

To obtain a small set of occupied orbitals for atom A all blocks of W are set to zero except for W_{AA}. One then obtains a few large eigenvalues, corresponding to orbitals which are localized on atom A, and many small or zero eigenvalues, corresponding to orbitals localized on other atoms. To obtain a medium set of occupied orbitals for atom A we retain W_{AA}, W_{AB}, W_{BA} and W_{BB}, where B is any atom directly bonded to atom A (i.e., B is a nearest neighbor of A). To obtain a large set of orbitals we allow B to be a nearest neighbor of A or a nearest neighbor of a nearest neighbor.

[1] P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
[2] S. Saebø and P. Pulay, J. Chem. Phys. **86**, 914 (1987).
[3] S. Saebø, W. Tong, and P. Pulay, J. Chem. Phys. **98**, 2170 (1993).
[4] P. Pulay, J. Chem. Phys. **85**, 1703 (1986).
[5] P. Pulay and S. Saebø, Theor. Chim. Acta. **69**, 357 (1986).
[6] C. Hampel and H. J. Werner, J. Chem. Phys. **104**, 6286 (1996).
[7] R. Murphy, B. Beachy, R. A. Friesner, and M. N. Ringnalda, J. Chem. Phys. **103**, 1481 (1995).
[8] B. Kirtman and C. E. Dykstra, J. Chem. Phys. **85**, 2791 (1986).
[9] B. Kirtman, Int. J. Quantum Chem. **55**, 103 (1995).
[10] D. Bernholdt and R. J. Harrison, J. Chem. Phys. **102**, 9582 (1995).
[11] M. Head-Gordon, P. E. Maslen, and C. A. White, J. Chem. Phys. **108**, 616 (1998).
[12] P. E. Maslen and M. Head-Gordon, Chem. Phys. Lett. **283**, 102 (1998).
[13] G. Rauhut, P. Pulay, and H. J. Werner, J. Comput. Chem. **19**, 1241 (1998).
[14] P. E. Maslen and M. Head-Gordon, work in progress.
[15] W. Kutzelnigg, Theor. Chim. Acta. **68**, 445 (1985).
[16] S. Saebø and P. Pulay, Chem. Phys. Lett. **113**, 13 (1985).
[17] S. Saebø and P. Pulay, Chem. Phys. Lett. **131**, 384 (1986).
[18] S. Saebø and P. Pulay, J. Chem. Phys. **88**, 1884 (1988).
[19] S. Saebø and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).
[20] A. El Azhary, G. Rauhut, P. Pulay, and H. J. Werner, J. Chem. Phys. **108**, 5185 (1998).
[21] G. Rauhut, in preparation.
[22] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).
[23] M. Head-Gordon and J. A. Pople, J. Phys. Chem. **97**, 1147 (1993).
[24] J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989).
[25] J. Boughton and P. Pulay, J. Comput. Chem. **14**, 736 (1993).