Reduced Scaling Electron Correlation Methods

Emily A. Carter, University of California, Los Angeles, CA, USA
Derek Walter, University of California, Los Angeles, CA, USA

Next >

Keywords: reduced scaling; electron correlation; local correlation; pseudospectral approximation; integral direct; configuration interaction; Møller–Plesset; coupled cluster

Abstract

We give a detailed description of recent developments in reduced scaling *ab initio* methods for treating electron correlation in atoms and molecules. Since configuration interaction (CI) is the simplest electron correlation method, it is introduced first and used as a context to describe the familiar local correlation and pseudospectral approximations. The article then goes on to discuss the application of these approximations to Møller–Plesset perturbation theories. We use the context of Møller–Plesset perturbation theories to introduce the use of integral-direct techniques in electron correlation. These techniques are of particular importance as they eliminate major storage bottlenecks and pave the way for the treatment of ever larger systems. As such, we give step-by-step descriptions of the simplest integral-direct algorithms. We discuss the particular usefulness of integral-direct methods when the local correlation approximation is employed and how the combination of these two approaches can yield linear scaling electron correlation methods. Additionally, we describe some newly developed Møller–Plesset approaches based upon the direct use of highly localized atomic orbitals (AO) that exploit the same physical concepts as in the canonical approaches. The article concludes with a description of coupled cluster (CC) methods that take advantage of the full range of techniques described in the discussions of configuration interaction and Møller–Plesset theories. To enhance the accessibility of the article, the underlying theories behind the various *ab initio* electron correlation methods are described briefly prior to describing the reduced scaling versions of the methods.

Next >

Reduced Scaling Electron Correlation Methods

Emily A. Carter, University of California, Los Angeles, CA, USA
Derek Walter, University of California, Los Angeles, CA, USA

1. Introduction

The electronic structure branch of computational chemistry is concerned mainly with the formation of approximate solutions to the electronic Schrödinger equation for atoms and molecules

$$\hat{H}\Psi = E\Psi \tag{1}$$

where

$$\hat{H} = \sum_i -\frac{1}{2}\nabla^2 - \sum_{i,A} \frac{Z_{iA}}{r_{iA}} + \sum_{i>j} \frac{1}{r_{ij}} \tag{2}$$

In equation (2), the first term represents the electronic kinetic energy, the second term represents the attraction of the electrons to the nuclei, and the third term represents the electron–electron repulsion. In *ab initio* electronic structure methods, the formation of an approximate solution to equation (1) is generally a two-phase process. First, some sort of self-consistent field (SCF) calculation is carried out to obtain a zeroth-order solution to equation (1). Such SCF methods usually boil down to the minimization of the energy of a trial wave function that consists of a single antisymmetrized Hartree product, that is, a **Slater determinant**. This minimization process allows individual electrons to respond to the average charge distribution formed by the remaining electrons. However, outside of accounting for quantum mechanical exchange effects, the product form of a Slater determinant still yields independent probability distributions for individual electrons. Thus, SCF methods are generally not said to describe dynamic electron correlation, that is, the correlated motion of electrons that arises from the repulsive Coulomb potential between them (*see* Localized MO SCF Methods).

There do exist many forms of widely used multideterminantal SCF methods,[1] such as generalized valence bond (GVB),[2] RASSCF,[3] and CASSCF.[4] Such methods are generally employed with the aim of making a proper zeroth-order description of spin symmetry, bond breaking, or some other chemical situation, which is inherently multiconfigurational in nature. This type of electron correlation is referred to as static or nondynamic electron correlation. While it is true that these multiconfigurational SCF methods capture a level of dynamic electron correlation beyond what is captured by single determinant methods, the level of correlation energy recovered is rarely adequate for the purpose of obtaining quantitative accuracy from quantum chemistry calculations. Thus, some sort of treatment of dynamic electron correlation is quite often required in the formation of approximate solutions to equation (1). This comprises the second phase of the two-phase treatment of the electronic Schrödinger equation.

The most common methods for rigorously treating dynamic electron correlation involve post-SCF calculations in which the SCF wave function is used as a starting point. For a tutorial on these techniques, the reader is referred to the article of Bartlett and Stanton.[5] In this article, we will only give a brief introduction to the various methods used. The three most common methods for treating dynamic correlation are configuration

interaction (CI),[6] Møller–Plesset perturbation theory (MPx, where x denotes the order of perturbation theory),[7] and coupled cluster (CC)[8] (*see* Møller–Plesset Perturbation Theory). Although MP2 is by far the most popular *ab initio* method for treating dynamic electron correlation, all three of these methods perform admirably when used under the appropriate circumstances. Unfortunately, they all have the drawback of being very expensive computationally. The MP2 method scales as n^5 while the simplest forms of CI and CC, which involve only single and double excitations from the SCF reference, scale as n^6, where n is a variable proportional to the size of the system under consideration. More accurate versions of these methods scale even worse. Clearly, this poor scaling necessitates less expensive electron correlation methods. Recognizing this need, researchers have developed an extensive set of reduced scaling CI, MPx, and CC methods. The purpose of this article is to give a detailed description of these innovations.

The article is structured as follows. Since CI is conceptually the simplest electron correlation method, reduced scaling CI methods are discussed first and are used to exemplify many concepts that are prevalent in reduced scaling electron correlation methods. More specifically, local CI is used to demonstrate the basic ideas and algorithms behind local correlation techniques including localized single-particle bases, the use of temporary orthogonal bases, and domain selection. The context of CI is also used to describe the use of pseudospectral techniques in reduced scaling correlation methods. The article continues with a discussion of reduced scaling MPx methods. Similar to the discussion on CI, reduced scaling MPx theories based on local correlation and pseudospectral techniques will be discussed. Here, we will deviate from the pattern of using CI as a context for describing new innovations in reduced scaling electron correlation and discuss the use of integral-direct methodology within the framework of MPx. This is done because no previous work exists on reduced scaling integral-direct CI methods. The article concludes with a discussion of reduced scaling CC methods, which make use of the full arsenal of techniques described in the previous discussions of CI and MPx.

Before proceeding further, we introduce the notation we shall use, which generally conforms to accepted standards. A listing is given below:

i,j,k,l . . . . internal (occupied) orbitals
a,b,c,d . . . . virtual (unoccupied) orbitals
χ,μ,ν,λ . . . . atomic orbitals
p,q,r,s . . . . projected atomic orbitals (described in text)
n. number of electrons
n_i. number of internal orbitals
n_v. number of virtual orbitals
n_{bas}. number of basis functions
$L_{\nu x}$. contribution of atomic orbital ν to orbital x

< Previous   Next >

Reduced Scaling Electron Correlation Methods

Emily A. Carter, University of California, Los Angeles, CA, USA
Derek Walter, University of California, Los Angeles, CA, USA

2. Configuration Interaction

2.1. Conventional Configuration Interaction

The CI method is definitely the most straightforward electron correlation method to describe. Briefly, CI requires one to build a set of n-electron basis functions by replacing occupied molecular orbitals (MOs) in some SCF reference wave function with virtual (unoccupied) orbitals. The representation of the **Hamiltonian** in the resulting n-particle basis is then diagonalized. Note that we determine an orbital's designation as 'occupied' or 'unoccupied' by referring only to the number of electrons it contains in the SCF configuration, and not in any of the excited configurations. Thus, in the n-particle states, which include originally 'unoccupied' orbitals, those orbitals are now occupied (i.e., they now contain electrons in the excited configuration). By far, the most common type of CI calculation performed is based on the set of n-particle functions formed by replacing a maximum of two occupied MOs in the reference with two virtual orbitals [single and double excitation configuration interaction (SDCI)]. The SDCI wave function can then be written as

$$\Psi_{\mathrm{SDCI}} = \Phi_0 + \sum_{i,a} C_a^i \Phi_i^a + \sum_{i,j,a,b} C_{ab}^{ij} \Phi_{ij}^{ab} \tag{3}$$

Here, Φ_0 is the SCF reference function, the Φ_i^a and Φ_{ij}^{ab} are the excited n-particle states, and the C_i^a and C_{ij}^{ab} are the weights of these n-particle functions. For simplicity, we have neglected the spin variable in this discussion. Since anything beyond a double replacement will have a vanishing matrix element with Φ_0 due to the two-body nature of the Hamiltonian, SDCI generally represents the best compromise between computational cost and accuracy, which is the principle reason for its popularity compared to other CI methods.

It should be noted that the SCF reference function Φ_0 can easily be taken to be multiconfigurational in nature, in which case the method is referred to as multireference singles and doubles configuration interaction (MRSDCI). This is a particularly important point that sets apart the CI method from the MPx and CC methods in that it allows for a more straightforward treatment of nondynamic electron correlation.

To a lesser extent, CI calculations are also performed, which include the full set of n-particle wave functions that can be formed from all unique combinations of the underlying single-particle basis, subject to the constraint of a fixed particle number. Such full CI calculations give results that represent the absolute limit to the performance of the given single-particle basis. Unfortunately, the computational requirements of full CI calculations scale as $n!$ and are only practical as benchmark tools for systems having 10 or fewer electrons. For this reason, we focus here only on SDCI, and full CI methods are not discussed further. Nevertheless, it should be noted that some work on reduced scaling full CI methods has been carried out.[9]

Even in the smallest CI calculations, which are based on only single and double replacements from the reference (SDCI), the number of terms in the CI expansion can quickly exceed 10^6. The resulting Hamiltonian

matrix will then have roughly 10^{12} elements and therefore cannot be stored easily in high-speed memory or on disk. This situation forces adoption of nonconventional techniques to diagonalize the Hamiltonian. The most common of these techniques[10] relies upon the formation of a residual vector **R** that can then be used to generate an updated vector of CI expansion coefficients

$$\mathbf{R} = \mathbf{HC} - E\mathbf{C} \tag{4}$$

In equation (4), E is the current iteration's best estimate to the CI total energy and **C** is the current iteration's best estimate to the vector of CI expansion coefficients which minimizes the CI energy (referred to as the CI vector). In the CI literature, one often sees references to the **Z** or σ vectors, which are both simply the vector **HC**. In this article, we focus on the use of the residual since it is the principal computational quantity of iterative reduced scaling MPx and CC theories. The key point to recognize in equation (4) is that the full Hamiltonian matrix is never required. Rather, one simply needs to know how to form the product of the Hamiltonian and some trial vector of CI expansion coefficients. This forms the foundation for direct CI approaches[11] in which small contributions to the Hamiltonian are formed, immediately used in multiplications with the CI vector, and discarded. Of course, such methods require one to store only two vectors of length equal to the total number of terms in the CI expansion in high-speed memory, which represents an improvement of many orders of magnitude over the memory required to store the full Hamiltonian.

The most common types of direct SDCI and MRSDCI are said to be 'integral-driven': individual classes of two-electron integrals are loaded into high-speed memory and their contributions to the residual are treated in full. Generally, the integrals are grouped into classes depending on the total number of internal indices. For example, the equation for computing the contribution of the class of integrals having zero internal indices to the MRSDCI residual is

$$R_{ab}^{\lambda} \leftarrow \sum_{c,d} (ac|bd) C_{cd}^{\lambda} \tag{5}$$

A few words of explanation are in order for equation (5). The index λ denotes an $n-2$ state consisting of only internal orbitals, and the components of the CI vector are arranged into square matrices $\mathbf{C}^{\lambda}$ in which the different elements correspond to the occupation of different pairs of virtual orbitals. Thus, the quantity C_{ab}^{λ} denotes the component of the CI vector corresponding to the n-particle state with internal portion λ and having virtual orbitals a and b occupied. The spin coupling of λ and the total spin fixes the spin coupling of the virtual orbitals a and b. This external spin coupling controls the symmetry of the $\mathbf{C}^{\lambda}$ matrices: triplet coupling in the external space leads to antisymmetric matrices $(C_{ab}^{\lambda} = -C_{ba}^{\lambda})$ and singlet coupling leads to symmetric matrices ($C_{ab}^{\lambda} = C_{ba}^{\lambda}$). Similar comments can be made for the residual matrices $\mathbf{R}_{ab}^{\lambda}$.

In the case in which there is only a single reference state, a one-to-one mapping can be made between the indices λ and a pair of 'hole' orbitals that are replaced by virtual orbitals. For the multiconfigurational case, one cannot in general make such an association, necessitating the use of the λ index. In our subsequent discussions of MPx and CC theories, which are almost always based on a single-configuration reference wave function, the components of the residual shall be denoted as R_{ab}^{ij} where i and j index the hole orbitals.

Although equation (5) is quite simple in structure, it is, unfortunately, very expensive computationally, and asymptotically controls the scaling of an MRSDCI calculation. Incidentally, terms with a structure identical to equation (5) also dominate MP3, MP4, and coupled cluster with single and double excitations (CCSD) calculations. The number of residuals that must be computed scales roughly like $n_i^2 n_v^2$. For each of these residuals, a total of n_v^2 floating point operations is required, which leads to an overall scaling of $n_i^2 n_v^4 \propto N^6$.

The second-most demanding class of integrals to treat is the one having two internal indices, which contribute to the MRSDCI residual in the following manner:

$$R_{ab}^{\lambda} \leftarrow \sum_{c,\mu} B^{\lambda\mu}(ij|ac)C_{cb}^{\mu} \tag{6}$$

where the factor $B^{\lambda\mu}$ depends on the orbital occupation and spin coupling of the $n-2$ states λ and μ. Analyzing equation (6) in a manner similar to that used for equation (5), one might determine that the treatment of the *(ij|ab)* scales something like $n_i^2 n_v^3 \propto N^5$. Although this is essentially correct for large systems, the summation over the index μ can cause the treatment of the *(ij|ab)* integrals to become more costly than the treatment of the *(ab|cd)* for medium-sized molecules when a large reference space is used.

We shall not delve into the details of the equations for treating other classes of integrals in MRSDCI. For a complete discussion of this matter, the reader is referred to the review articles of Saunders and van Lenthe[12] and Duch and Karwowski.[13] The main purpose of this discussion was to briefly outline the computational procedures of MRSDCI. These procedures, in fact, bear a striking resemblance to those of MP*x* and CC theories. Thus, before going further, a brief summary of this section is warranted: The principal quantity, which must be computed to diagonalize the CI Hamiltonian, is a residual vector. This residual vector is constructed without forming the full Hamiltonian. Rather, the contributions to this residual from various classes of two-electron integrals are handled individually, using expressions similar to equations (5) and (6). Again, we emphasize that this basic procedure is followed in both MP*x* and CC theories, with the essential differences being minor variations in the expressions for the residuals.

2.2. Local Configuration Interaction: General Considerations

The local correlation approximation[14] is based on the observation that dynamic electron correlation exhibits an asymptotic r^{-6} dependence on the interelectron distances and is therefore highly *localized*. The use of highly *delocalized* canonical orbitals in conventional electron-correlation treatments, however, precludes the possibility of taking advantage of this short-ranged nature of dynamic electron correlation. The obvious solution here is to find some basis of localized MOs and work in that basis, neglecting or approximating interactions involving widely separated orbitals. This is the basic philosophy behind the local correlation approximation. Unfortunately, there are a number of technical difficulties associated with the use of localized single-particle bases. Obviously, the first difficulty is how to actually obtain localized bases. Methods of localizing the space of occupied orbitals take advantage of the fact that the total wave function arising from a manifold of closed-shell orbitals or high-spin open-shell orbitals remains invariant upon applying a unitary transformation to those orbitals. Procedures for finding a unitary transformation to localize the space of internal orbitals are all remarkably similar. Some functional of the orbitals $B\{\phi_i\}$ is defined, which provides a measure of the overall extent of orbital delocalization. This functional is then minimized or maximized through a series of pairwise rotations, each of which leads to the greatest incremental change in the value of the delocalization functional for the current iteration.[15]

A number of delocalization functionals have been proposed, the most popular of which are those used by the Boys[16] and Pipek–Mezey procedures.[17] The Boys delocalization functional is taken to be

$$B^{\text{Boys}}\{\phi_i\} = \sum_{i=1}^{n_i} \langle i(\mathbf{r}_1)i(\mathbf{r}_2)|(\mathbf{r}_1-\mathbf{r}_2)^2|i(\mathbf{r}_1)i(\mathbf{r}_2)\rangle \tag{7}$$

One can see in equation (7) that the Boys procedure aims to directly minimize the overall spatial extent of the orbitals, making the Boys procedure one of the most conceptually straightforward localization approaches. Unfortunately, the Boys procedure requires one to compute the one-electron integrals $\langle i|\mathbf{r}|i\rangle$ for all internal

orbitals. Since these integrals are first computed on an atomic orbital (AO) basis, and subsequently transformed to the MO basis, the procedure scales overall as n^3. The Pipek–Mezey procedure uses Mulliken population analyses to determine the number of atoms over which the MOs extend. The functional maximized in the Pipek–Mezey procedure is given by

$$B^{PM}\{\phi_i\} = \sum_{i=1}^{n_i} \sum_{A=1}^{n_{atoms}} (Q_A^i)^2 \tag{8}$$

where Q_A^i is the Mulliken population of orbital i on atom A. The logic behind this functional can be easily understood by noting that the quantity

$$d_i = \left\{ \sum_{A=1}^{n_{atoms}} (Q_A^i)^2 \right\}^{-1} \tag{9}$$

represents a rough estimate of the number of atoms upon which orbital i is localized. In order to enhance localization, then, one should maximize the sum of the d_i^{-1}, which is exactly equation (8). Although the cost of both the Boys and Pipek–Mezey procedures scales as n^3, the Pipek–Mezey procedure has a few advantages over the Boys procedure. First, the only external quantity required by the Pipek–Mezey method is the AO overlap matrix. Second, the Pipek–Mezey procedure generally preserves separation of σ and π orbitals better than the Boys procedure does.[17]

A few other localization functionals are worthy of note. In the Edmiston–Ruedenberg approach[15, 18] one aims to maximize the self-repulsion energy

$$B^{ER}\{\phi_i\} = \sum_i \langle i(\mathbf{r_1})i(\mathbf{r_2}) || \mathbf{r_1} - \mathbf{r_2} |^{-1} | i(\mathbf{r_1})i(\mathbf{r_2}) \rangle \tag{10}$$

Contrary to both the Pipek–Mezey and Boys procedures, the Edmiston–Reudenberg procedure requires the full set of two-electron repulsion integrals in the MO basis. The Edmiston–Ruedenberg localization method, therefore, scales as n^5 and consequently is not widely used. In the von Niessen localization scheme,[19] one minimizes the sum of the charge density overlaps:

$$B^{vN}\{\phi_i\} = \langle i(\mathbf{r_1})i(\mathbf{r_2}) | \delta(\mathbf{r_1} - \mathbf{r_2}) | i(\mathbf{r_1})i(\mathbf{r_2}) \rangle \tag{11}$$

Localization procedures based on the von Niessen functional exhibit the same poor scaling as those based on the Edmiston–Ruedenberg functional, and are also not widely used.

Because virtual orbitals have a very large number of nodes, one cannot localize them using the same methods that are used for the occupied orbitals. In fact, one generally has to forego orthogonality of the virtual orbitals in local correlation methods. The most popular type of localized virtual orbitals are projected atomic orbitals (PAO), which are formed by projecting the internal space from the individual AOs:

$$\phi_r = \sum_i^{n_i} (1 - |\phi_i\rangle\langle\phi_i|)\chi_\rho \tag{12}$$

Note here that we have generated the PAO labeled as r from the AO labeled as P in order to maintain consistency with the notation scheme we have chosen. The space of the PAOs has the important advantage of being orthogonal to the internal space. However, since one can generate a PAO from every AO and the PAOs

only span part of the space spanned by the AOs, the PAOs are linearly dependent. Both this and the fact that the PAOs are not orthogonal to one another must be taken into consideration in calculations that use PAOs.

More recently, researchers have developed MPx and CC methods in which one works directly in the AO basis.[20, 21] However, this approach has not been attempted in the CI method and we shall defer discussion of these methods to subsequent sections on MPx and CC.

An additional consideration in local CI calculations is how to appropriately choose the set of configurations that should be retained or neglected in the local CI calculation. This task can be broken down into a part involving only the internal space and a part involving both the internal and virtual orbital spaces. The former involves the elimination of configurations having excitations out of widely separated orbitals and will be referred to here as the weak-pairs (WP) approximation. The latter involves the neglect of configurations having excitations to orbitals, which are far from the hole orbital (i.e., the internal orbital that is replaced by a virtual orbital), and will be referred to here as the truncation-of-the-virtuals (TOV) approximation. In the MRSDCI method, implementation of the WP approximation is quite simple. In the local CI method of Walter et al.,[22, 23] spheres are associated with every internal orbital, and doubly excited configurations are neglected if the spheres associated with the two-hole orbitals do not overlap. The sphere radii are chosen to be the maximum separation between any two atoms that heavily contribute to the orbital, and the sphere centers are taken to be Mulliken-population weighted-average positions of those atoms. An adjustable parameter is also included in the CI calculation so that one can tune the size of the spheres and hence the accuracy of the CI calculation. In the local MPx and CC methods of Saebø and Pulay[24-26] and Werner et al.,[27, 28] configurations having excitations out of pairs of orbitals that are separated by roughly 10 times the diameter of the orbitals were neglected. It should be noted, however, that in the local MPx and CC methods developed by these researchers, pairs of internal orbitals could also be classified as weakly interacting. The contributions of configurations involving excitations out of pairs of weakly interacting orbitals, generally, were not neglected but approximated using low orders of perturbation theory or multipole approximations. As this type of approach has only been implemented for reduced scaling MPx and CC theories, a detailed discussion of these methods shall be deferred to the appropriate sections on MPx and CC.

The selection of a subset of correlating PAOs that is localized near a given internal orbital is commonly referred to as the creation of an *orbital domain*. The domain of an internal orbital i shall be denoted as [i]. For a given pair of internal orbitals ij, their domain consists of [i]∪[j] and shall be denoted as [ij]. In the local CI method of Walter et al.,[23] spheres are associated with each PAO. A PAO is then included in the domain of an internal orbital if the sphere of the PAO overlaps with the sphere of the internal orbital. As was the case for the WP method of Walter et al.,[22] the size of the spheres associated with the PAOs may be tuned in order to control the accuracy of the local CI calculation. Although this approach seems primitive, it has the advantage of being quite easy to implement. Furthermore, it has proved to be fairly successful, leading to local CI expansions that recover greater than 97% of the nonlocal CI correlation energy.[23]

Boughton and Pulay[29] developed a more involved scheme for generating orbital domains. Briefly, one attempts to expand a given internal orbital in a subset of PAOs. The average difference between the actual orbital and its approximate expansion in terms of PAOs, f_i, is then computed

$$f_i = \int (\phi_i - \phi_i')^2 d\tau \tag{13}$$

In equation (13), the orbital ϕ'_i is an approximation to ϕ_i based on the PAOs in the domain of internal orbital i

$$|\phi_i'\rangle = \sum_{r \in i} |\phi_r\rangle R_{ri}' \tag{14}$$

In the Boughton–Pulay approach to the generation of orbital domains, one takes the domain of ϕ_i to be the smallest set of PAOs that will bring the value of f_i below some prescribed threshold. To our knowledge, the Boughton–Pulay approach has not been adopted for local CI. However, the approach is not particularly difficult to implement and has been used with great success in the local correlation methods of Werner et al.[27], [28], [30], [31]

2.3. Implementation of Local Configuration Interaction

As we have noted, the residual vector is used to form an improved version of the CI vector. The most commonly used update expression in conventional CI is of the Davidson form:[32]

$$\delta C_i = \frac{R_i}{(E - H_{ii})} \tag{15}$$

where δC_i is the update to the ith component of the CI vector and H_{ii} is the ith diagonal element of the CI Hamiltonian. Unfortunately, in the local CI method the direct application of equation (15) leads to exceedingly poor convergence or, even worse, no convergence at all. The principal reason for this is that the nonorthogonality of the PAOs leads to a nonorthogonal set of n-particle states, which then, of course, mandates that one solve a generalized eigenvalue problem rather than a usual eigenvalue problem. However, equation (15) is appropriate only for the latter. The convergence problem is exacerbated by linear dependencies among the PAOs, and also by the fact that one is working with n-particle states that are no longer eigenfunctions of a zeroth-order Hamiltonian.

The convergence problems outlined above have been solved through the use of temporary orthogonal bases, which is a scheme that was implemented by Hampel and Werner in local CCSD.[27] Basically, one takes the PAOs in every orbital domain and orthogonalizes them by some means, removing the linear dependencies within each domain as a part of the orthogonalization. The resulting set of *orthogonalized* PAOs (also referred to as pseudocanonical orbitals) is then used to form a set of temporary n-particle configurations that are mutually orthogonal. The CI vector is updated in this orthogonal basis using equation (15). Note that the residual vector is still built in the basis of n-particle states formed from configurations made up of the PAOs (i.e., a nonorthogonal basis). However, once the residual has been formed, it is immediately transformed to a basis of n-particle states containing *orthogonalized* PAOs according to the following expression, where the dagger (†) indicates the adjoint of X^2:

$$\mathbf{R}^\lambda = \mathbf{X}^\lambda \mathbf{R}_{\text{PAO}}^\lambda \mathbf{X}^{\lambda\dagger} \tag{16}$$

where $\mathbf{X}^\lambda$ is a matrix that transforms the PAOs in the domain of λ to an orthogonal set

$$\phi_a^{\text{temp}} = \sum_p X_{ap}^\lambda \phi_p \tag{17}$$

Once the CI vector has been updated, it is back-transformed to the PAO basis so that a new residual may be constructed in the nonorthogonal basis. An important point to note is that the elements of the CI vector transform in a manner opposite to the individual orbitals:

$$\mathbf{C}_{\text{PAO}}^\lambda = \mathbf{X}^\lambda \mathbf{C}^\lambda \mathbf{X}^{\lambda\dagger} \tag{18}$$

This will have important implications for the structure of the CI equations in the nonorthogonal basis, as we

discuss below.

This situation whereby one must repeatedly transform the CI and residual vectors may at first seem odd, and one might propose to avoid the transformations altogether by working solely in a basis of orthogonalized PAOs. However, working in a PAO basis is advantageous because it is much more straightforward to handle the two-electron integrals than it would be if one were to work in an orthogonalized PAO basis. The latter situation would require one to compute all classes of two-electron integrals for all orthogonalized PAOs in all orbital domains – a costly and unwieldy procedure. When working in the PAO basis, on the other hand, one may take advantage of the fact that the PAO basis spans all orbital domains and simply compute all classes of two-electron integrals over all PAOs, and then use those integrals to treat all orbital domains. Furthermore, the enhanced localization properties of the PAOs may be exploited to great advantage in integral-direct local correlation treatments in which one recomputes the two-electron integrals as they are needed. [30, 33]

In addition to requiring repeated transformations of the residual and CI vectors back and forth between orthogonal and nonorthogonal bases, working in the PAO basis has the disadvantage of slightly complicating the form of the CI equations. As we noted above, the orbitals transform in a manner opposite to that in which the elements of the CI vector transform. As a result, when one attempts to transform the equations to a nonorthogonal basis, the $\mathbf{X}^\lambda$ matrices cancel out whenever an orbital index is contracted with an index of the CI vector. For example, in the case of the equation for treating the *(ab|cd)* integrals (equation 5), the structure of the equation remains unchanged in the PAO basis because both the indices *c* and *d* of the CI vector are contracted with an orbital index. The same cannot be said for the equation for treating the integrals having two external indices (equation 6), however, since the *b* index of the CI vector is not contracted with an orbital index. As a result, the equation for treating these integrals in the PAO basis requires an additional contraction with the PAO overlap matrix:

$$R_{rs}^\lambda \leftarrow \sum_{\mu,p,q} B^{\lambda\mu}(ij|rp)S_{sq}C_{pq}^\mu \tag{19}$$

where $B^{\lambda\mu}$ is a constant that depends on the spin coupling and orbital occupations of the *n* – 2 states λ and μ, and S_{pq} is the overlap integral of PAOs *p* and *q*. This additional contraction with the PAO-overlap matrix will appear in the CI equations when written in the PAO basis whenever an index of the CI vector remains uncontracted with an orbital index. The problem is most severe for the all-internal integrals, which are handled in the MO basis according to the following expression:

$$R_{ab}^\lambda \leftarrow \sum_\mu A^{\lambda\mu}(ij|kl)C_{ab}^\mu \tag{20}$$

where the indices *i*, *j*, *k*, and *l* are determined by the values of λ and μ. In the PAO basis, the *(ij|kl)* are handled according to the following expression:

$$R_{rs}^\lambda \leftarrow \sum_\mu A^{\lambda\mu}(ij|kl)S_{rp}C_{pq}^\mu S_{qs} \tag{21}$$

The additional contractions with the PAO-overlap matrix in equation (21) have the effect of increasing the formal scaling of treating the *(ij|kl)* from n_v^2 to n_v^4, on going from an orthogonal basis to the PAO basis. However, one must keep in mind that when the TOV approximation is employed, the relevant number of virtual orbitals that should be used in scaling arguments is the average number of virtual orbitals per domain, $\tilde{n}_v$, which is *independent* of the size of the system. This point is extremely important and paves the way for linear scaling electron correlation methods. It should be noted that the cost of treating the *(ij|kl)* using the TOV approximation can in many cases still exceed the cost of treating the *(ij|kl)* nonlocally, as a result of the

increased formal scaling in the PAO basis.[23] However, this is of little consequence as the size-independence of the virtual orbital domains will eventually lead to increased performance relative to the nonlocal method as the size of the system is increased.

2.4. Pseudospectral Configuration Interaction

The pseudospectral approximation[34] (see Pseudospectral Methods in Ab Initio Quantum Chemistry), as it is used in quantum chemistry,[35] is essentially a method of evaluating two-electron integrals in which one integration is handled analytically and the second integration is handled numerically. The general expression used in pseudospectral methods is

$$(ij|kl) \approx \sum_g^{n_g} Q_{ig} A_{kl}(r_g) R_{gj} \tag{22}$$

The quantities, $A_{kl}(r_g)$ and R_{gj}, appearing in equation (22) can be simply defined. The quantity $A_{kl}(r_g)$ is

$$A_{kl}(r_g) = \int \frac{\phi_k(r)\phi_l(r)}{|r - r_g|} dr \tag{23}$$

where r_g is the position of a grid point in a numerical quadrature grid and n_g is the number of grid points. Clearly, equation (23) represents an analytic integration over one of the electronic coordinates in the two-electron integral. The quantity R_{gj} is simply the value of MO j at the grid point r_g. The combination of the quantities R_{gj} and Q_{ig} represents a numerical integration over the second electronic coordinate. The quantity Q_{ig} serves to substantially enhance the accuracy of the numerical integration and its calculation is worthy of detailed discussion. Before turning to that point, however, we discuss the advantages of the pseudospectral approximation in quantum chemistry since this will motivate constraints on the **Q** matrix.

As we have noted, the treatment of the all-external integrals in conventional CI (equation 5) scales as $n_i^2 n_v^4$ and dominates the calculation as the size of the system increases. Straightforwardly inserting the pseudospectral approximation for the two-electron integrals into equation (5), one arrives at the following expression:

$$R_{ab}^\lambda \approx \sum_g Q_{bg} \left(\sum_c A_{cd}(r_g) \left(\sum_d R_{gd} C_{cd}^\lambda \right) \right) \tag{24}$$

Unlike in equation (5), the sums over c and d can be handled individually, each scaling only as $n_g n_v^2$. In addition to a small loss in precision, one must pay a computational penalty in the form of an extra sum over the number of grid points, which also scales as $n_g n_v^2$. Overall, the scaling of the pseudospectral treatment of the $(ab|cd)$ integrals is $3n_g n_v^2$. Thus, as long as $3n_g < n_v^2$ the cost of treating the $(ab|cd)$ pseudospectrally will be less than the cost of treating the $(ab|cd)$ analytically. Furthermore, the pseudospectral savings will increase as the size of the basis set increases relative to the total number of grid points. Similar results apply for the treatments of other classes of two-electron integrals but will not be discussed in detail here.

We now turn to the question of how one determines the **Q** matrix. The discussion that follows adopts many ideas from the review article of Martinez and Carter.[36] A seemingly reasonable standard for the **Q** matrix to satisfy is

$$\mathbf{S} \approx \mathbf{QR} \tag{25}$$

In the most primitive pseudospectral scheme, one might try to satisfy equation (25) by making the following

approximation for $\mathbf{Q}$:

$$Q_{jg} = R_{gj}w_g \tag{26}$$

where w_g is a quadrature weight associated with grid point r_g. The quadrature weights are, of course, chosen so that $\mathbf{Q}$ most closely satisfies equation (25). As has been pointed out by Martinez and Carter,[36] a similar approach is used in density functional theory (DFT) methods for the computation of matrix elements of the exchange correlation functionals. In such schemes, roughly 10^4 grid points per atom are required to achieve accuracy. Unfortunately, Martinez and Carter estimated that such a large number of grid points would swamp out any potential savings from the pseudospectral approach unless one carries out calculations on systems having more than 40 heavy atoms. Thus, equation (26) is not a suitable choice for $\mathbf{Q}$.

A choice for $\mathbf{Q}$, which would exactly satisfy equation (25), is

$$\mathbf{Q} = \mathbf{S}\mathbf{R}^{-1} \tag{27}$$

The disadvantage of this approach is that it requires the quantity $\mathbf{R}^{-1}$ to exist, which of course requires $\mathbf{R}$ to be square (i.e., one must use exactly as many grid points as there are basis functions). If one increases the number of basis functions in order to match the number of grid points that is necessary for numerical accuracy, the quantum chemistry calculation would become far too costly. Likewise, decreasing the number of grid points to match the number of basis functions will diminish the precision of the pseudospectral approach substantially.

A final choice for $\mathbf{Q}$ that eliminates the problems associated with the previously mentioned approaches is as follows:

$$\mathbf{Q} = \mathbf{S}(\mathbf{R}^T\mathbf{w}\mathbf{R})^{-1}\mathbf{R}^T\mathbf{w} \tag{28}$$

This $\mathbf{Q}$ matrix is rectangular, corresponding to a generalized inverse of $\mathbf{R}$, and allows one to use a number of grid points, which exceeds the number of basis functions. Note also that this choice for $\mathbf{Q}$ satisfies equation (25) exactly.

One issue that has not yet been addressed is dealiasing, which arises because the analytic integration over orbitals k and l in equation (23) introduces components outside of the single-particle basis. In other words, the product $A_{kl}(r_g)R_{gj}$ can be written in the following manner:

$$A_{kl}(r_g)\phi_j(r_g) = \sum_{\nu}^{\text{basis}} c_\nu \phi_\nu(r_g) + \sum_{\mu}^{\text{complement}} c_\mu \phi_\mu^{alias}(r_g) \tag{29}$$

where the first sum runs over the functions in the basis set and the second sum runs over all functions in the complement of the basis set (i.e., the set of functions orthogonal to the basis set that serve to complete the basis set). Since the $\mathbf{Q}$ matrix satisfies equation (25), one can be sure that the quadrature with the first sum will give the exact value of the integral *(ij|kl)*. However, there is no reason to assume that the quadrature with the second sum in equation (29) will vanish and not contaminate the overall value of the pseudospectral integral. This problem can be largely solved by including a small set of dealiasing functions in the $\mathbf{R}$ matrix when the $\mathbf{Q}$ matrix is computed.[37, 38] These dealiasing functions are chosen so that they contain the dominant components of the basis set complement. In doing this, one ensures that numerical quadrature with the $\mathbf{Q}$ matrix is accurate not only for the basis functions but also for the important parts of the complement of the basis set. As a result, the aliasing terms in equation (29) should largely vanish.

2.5. Overview of Performance of Reduced Scaling CI Methods

The first modern reduced scaling CI methods based upon the local correlation approximation were implemented by Saebø and Pulay.[39, 40] These initial studies focused mainly on the use of the TOV approximation, although a few sample calculations were done that made use of a very severe WP approximation. The results of their calculations were encouraging: they found that roughly 98% of the nonlocal SDCI correlation energy could be recovered using virtual orbital spaces only half as large as the full virtual orbital spaces. Furthermore, they found that the CPU time for local CI calculations could be roughly 40-fold less than the CPU times for nonlocal CI calculations if one was satisfied with recovering only 94% of the correlation energy. However, these initial calculations were carried out only on small systems having four or fewer heavy atoms and only a single closed-shell reference was used. The work of Saebø and Pulay in the area of reduced scaling CI was discontinued shortly thereafter, in favor of a focus on reduced scaling perturbation theory.

Subsequent work in reduced scaling CI focused on pseudospectral CI methods. Martinez and Carter first implemented an MRSDCI method in which the integrals having two, three, and four external indices were treated pseudospectrally.[41] For single reference SDCI calculations, significant savings were observed. A pseudospectral SDCI calculation on glycine using a 6-31G** basis set was 3.7-fold faster than the fully analytic version. For small MRSDCI calculations, the speedups were somewhat less (e.g., 2.2-fold speedup in difluoroethylene using a 6-31G** basis) but still significant. The pseudospectral method was also very precise, giving energies that differed from the analytic energies by only a few tenths of a millihartree.

Later, the work of Martinez and Carter was augmented by Reynolds et al.[42] to include the local correlation approximation. This was first done by simply including the WP approximation on top of the pseudospectral method. Unfortunately, it was determined that the WP approximation had a much more severe effect on the precision of the method compared to the pseudospectral method. Generally, only meager computational savings could be obtained from a WP pseudospectral CI method if one required 1 kcal mol^{-1} precision in total energies. Fortunately, the only quantities of physical interest are energy differences. For this purpose, it was determined that local pseudospectral calculations could reproduce to within 2% analytic nonlocal C–N conformational energy differences in the glycine molecule, for example, sixfold faster than the analytic calculations. No MRSDCI calculations were carried out using this method, and no applications more sophisticated than bond rotation were explored.

Reynolds and Carter later incorporated the TOV approximation into this WP pseudospectral SDCI method.[43] The results from this method showed a significant improvement over the use of the WP approximation alone. For example, it was shown that for the glycine molecule, the local pseudospectral CI method could recover over 98% of the nonlocal analytic CI correlation energy with a fivefold reduction in computational expense. An important point should be made regarding the use of the pseudospectral approximation in this local CI method. The pseudospectral approximation was used only to treat the *(pr|qs)* integrals, and this calculation was not done using truncated virtual orbital spaces, as is usually the case in the TOV approximation. Rather, the contributions of the *(pr|qs)* integrals to the residual were computed in the full AO basis pseudospectrally and then transformed to the PAO basis for incorporation into a final total residual. This procedure can be handled with minimal changes to the computer program since the structure of the equation for treating the *(pr|qs)* (equation 5) is invariant to a change of basis as a result of the cancellation of transforms (see Section 2.3). Another way to see this is to consider the expression for the contributions of the *(pr|qs)* in which the PAO expansion coefficients are extracted from the integrals

$$R_{rs}^{\lambda} \leftarrow \sum_{p,q,\mu,\nu,\chi,\sigma} (\mu\nu|\chi\sigma) C_{pq}^{\lambda} L_{\mu p} L_{\nu r} L_{\chi q} L_{\sigma s} \tag{30}$$

One can then move the transforms for p and q onto the CI expansion coefficients to give

$$R_{rs}^{\lambda} \leftarrow \sum_{\mu,\nu,\chi,\sigma} (\mu\nu|\chi\sigma) C_{\mu\chi}^{\lambda} L_{\nu r} L_{\sigma s} \tag{31}$$

Finally, the quantity $R_{\nu\sigma}^{\lambda}$ can be computed as

$$R_{\nu\sigma}^{\lambda} \leftarrow \sum_{\mu,\chi} (\mu\nu|\chi\sigma) C_{\mu\chi}^{\lambda} L_{\nu r} L_{\sigma s} \tag{32}$$

and then be back-transformed to the PAO basis:

$$R_{rs}^{\lambda} \leftarrow \sum_{\nu,\sigma} R_{\nu\sigma}^{\lambda} L_{\nu r} L_{\sigma s} \tag{33}$$

The sole reason for handling the *(pr|qs)* in this manner was to eliminate the need to generate the *(pr|qs)* from the AO integrals (an n^5 scaling procedure) and subsequently store the *(pr|qs)* on disk. This was done by simply storing the individual pseudospectral units on disk, the number of which scales approximately as n^3. In the case of CI calculations on the glycine molecule using a 6-31G** basis set, this reduced the disk storage requirements over 40-fold.

One is inclined to ask why the pseudospectral method is not used in combination with truncated virtual orbital spaces to achieve savings beyond what can be obtained by the TOV approximation alone. Unfortunately, to obtain savings from the pseudospectral approximation, the square of the number of virtual orbitals must exceed the number of grid points. If truncated virtual orbital spaces are used, one must use extremely small quadrature grids to satisfy this requirement, which will severely diminish the accuracy of the pseudospectral method. Thus, it is quite difficult to make efficient use of the pseudospectral and TOV approximations simultaneously.

The most recent work on reduced scaling CI methods was done by Walter et al.[44] and focused on the development of multireference CI methods for the accurate calculation of **bond dissociation energies** (BDEs) (see Bond Dissociation Energy (BDE)). For the determination of BDEs in small organic molecules having no more than two heavy atoms (systems where the local correlation approximation is not useful), the use of PS methods to treat the *(ij|ab)* and *(ab|cd)* integrals led to binding energies that differed from analytically determined binding energies by an average of only 0.2 kcal mol^{-1}. The introduction of the WP approximation to the PS CI for the purpose of computing binding energies in molecules having four or more heavy atoms led to a small decrease in the precision of the method: BDEs computed using the WP-PS CI method generally differed from nonlocal analytic BDEs by slightly more than 1 kcal mol^{-1}. Although this error is significant, it is acceptable for most applications. Furthermore, the savings from the combined WP-PS method were quite substantial. For the case of the determination of the central C–C bond in *n*-hexene, for example, the combined WP-PS method was over sevenfold faster than the nonlocal analytic method.

Despite the above-mentioned success of pseudospectral CI methods, they are fundamentally limited in that they can at best scale as $n_g n_v^2 \approx n^3$. The situation is different for reduced scaling methods based on the simultaneous use of the TOV and WP approximations. This can be seen most easily through an analysis of the expression for treating the *(pr|qs)* integrals in a CI calculation based on a single closed-shell reference

$$R_{rs}^{ij} \leftarrow \sum_{p,q} (pr|qs) C_{pq}^{ij} \tag{34}$$

When the WP approximation is used, one must compute for a fixed value of i the contributions to the R_{rs}^{ij} arising only from a subset of all possible orbitals j that are localized near orbital i. The number of such j is independent of the size of the system. Thus, the total number of pairs ij scales linearly with the size of the system. Similarly, when the TOV approximation is employed, the orbitals a, b, c, and d may only be PAOs that are localized near the orbitals i and j. Thus, for each possible ij pair an $\tilde{n}_v^4$ scaling calculation must be done. Since $\tilde{n}_v$ is independent of the size of the system, the overall cost of treating the (pr/qs) integrals using the local correlation approximation should scale linearly with the system size. Note that we have neglected the cost of generating the (pr/qs) integrals in this discussion. However, as we will see later, this task can also be completed in a linear scaling fashion. In light of this, the prospects for reduced scaling CI methods based strictly on local correlation methods seem more optimistic than those based on the use of pseudospectral methods.

Walter et al. have recently completed a study[23] of a local MRSDCI method that focused on the use of local correlation for computing BDEs and mapping out potential energy surfaces (PESs). The performance of the method was more or less in line with the performance of previously mentioned reduced scaling CI methods, although the error from the TOV approximation was more significant than the error arising from the WP approximation alone. In general, the use of the TOV approximation led to BDEs that were roughly 2 kcal mol^{-1} less than those determined by the fully nonlocal method. In an extreme case (the $H_2N – C(CH_3)_3$ BDE), the fully local BDE was 3.9 kcal mol^{-1} lower than the nonlocal BDE. This error was attributed to the neglect of long-ranged electron correlation. In order to map out PESs for bond cleavage, the virtual orbital domains associated with the relevant active orbitals were allowed to include PAOs associated with both fragments. For cleavage of a single bond in n-butane and a double bond in $trans$-2,3-hexene, this procedure led to fairly smooth surfaces. However, the local D_es were slightly lower than the analytic D_es due again to the neglect of long-ranged electron correlation (here, D_e is the difference between the total energies of the dissociated fragments and the molecule's absolute minimum of the potential energy surface describing the bond). The computational savings from this method were also quite significant. The treatment of the (ab/cd) integrals, which scales as n^6 in nonlocal CI, was found to scale slightly worse than n^2 in local CI. For most classes of integrals, similar improvements were observed. The treatment of the (ij/kl) integrals, however, was observed to scale slightly worse in the local treatment ($n^{4.7}$) than in the local treatment (n^4). This increase is attributable to the extra contractions with the PAO-overlap matrix that are required in local CI. Fortunately, the overall time for treating the (ij/kl) is quite small and does not dominate the calculation. Overall, the fully local MRSDCI method of Walter et al. was observed to scale somewhere between n^3 and n^4, which falls well short of linear scaling. This shortcoming is very likely a result of the fact that Walter et al. were not able to examine systems having more than eight heavy atoms because of an inability to handle the generation of the two-electron integrals in the PAO basis for larger systems. Further calculations on larger systems would probably yield reduced overall scalings. This has in fact been observed in sophisticated local CC and MPx methods that generate the required two-electron integrals on the fly and are therefore able to treat extraordinarily large systems.[28, 30] These methods will be discussed in detail in the following sections on CC and MPx.

< Previous   Next >

Reduced Scaling Electron Correlation Methods

Emily A. Carter, University of California, Los Angeles, CA, USA
Derek Walter, University of California, Los Angeles, CA, USA

3. Møller–Plesset Perturbation Theory

3.1. Conventional Møller–Plesset Perturbation Theory

Conventional MPx is covered in detail in a large number of quantum chemistry texts[45, 46] (*see* Møller–Plesset Perturbation Theory). Indeed, it is perhaps the most common and popular electron correlation method. For this reason, we will not give a highly detailed description of MPx. Instead, we will briefly summarize the theory and then give a detailed description of certain points that are most relevant to our discussion of reduced scaling MPx and are not covered in the basic textbooks.

The zeroth-order Hamiltonian in MPx is taken to be the sum of the single particle Fock operators:

$$\hat{H}_0 = \sum_i \hat{f}(i) \tag{35}$$

and the perturbation is taken to be the difference between the full Hamiltonian and the zeroth-order Hamiltonian

$$\hat{H}' = \hat{H} - \hat{H}_0 \tag{36}$$

The energies and wave functions at the desired orders of perturbation theory are then usually determined by solving the Rayleigh–Schrödinger perturbation theory equations. The first-order wave function, for example, is determined by

$$(\hat{H}_0 - E_0)\Phi_1 = -\hat{H}'\Phi_0 \tag{37}$$

and the second-order energy is determined by

$$E_2 = \langle \Phi_1 | \hat{H}' | \Phi_0 \rangle \tag{38}$$

The first-order wave function is generally expanded in a basis of n-particle wave functions orthogonal to Φ_0 and inner products of equation (37) with this basis are then taken in order to determine the expansion coefficients. As in the case of CI, equation (37) is commonly solved iteratively by using the residual (i.e., the difference between the left- and right-hand sides of equation (37) when evaluated using an approximate Φ_1) in order to obtain an updated set of expansion coefficients.

An extremely popular n-particle basis is spanned by the so-called first-order interacting space, which consists of the set of functions having nonzero matrix elements with the reference configuration. This basis can be generated by applying the orbital excitation operators to the reference wave function

$$\Phi_{ij}^{ab} = \hat{E}_{ai}\hat{E}_{bj}\Phi_0 = (\hat{a}_a^{\alpha\dagger}\hat{a}_i^{\alpha} + \hat{a}_a^{\beta\dagger}\hat{a}_i^{\beta})(\hat{a}_b^{\alpha\dagger}\hat{a}_j^{\alpha} + \hat{a}_b^{\beta\dagger}\hat{a}_j^{\beta})\Phi_0 \tag{39}$$

In equation (39) the $\hat{a}_a^{\alpha\dagger}$ creates an electron in orbital a with spin α and the $\hat{a}_i^{\alpha}$ annihilates an electron in orbital i with α spin. These states have the property that $\Phi_{ij}^{ab} = \Phi_{ji}^{ba}$. Consequently, in the following expansion of Φ_1 in terms of the first-order interacting space, we must make the restriction that $C_{ab}^{ij} = C_{ba}^{ji}$

$$\Phi_1 = \frac{1}{2}\sum_{i,j,a,b} C_{ab}^{ij}\Phi_{ij}^{ab} \tag{40}$$

As was done for CI, the coefficients are arranged into matrices that are labelled by the internal orbitals i and j. The restriction on the coefficients is equivalent to requiring

$$\mathbf{C}^{ij} = \mathbf{C}^{ji\dagger} \tag{41}$$

The expression for the MP2 energy when evaluated using the first-order interacting space is quite compact:

$$E^2 = \sum_{i,j} Tr\left[\mathbf{K}^{ij}(2\mathbf{C}^{ij} - \mathbf{C}^{ji})\right] \tag{42}$$

where $K_{ab}^{ij} = (ia|jb)$.

Although using the first-order interacting space has a number of computational advantages that were pointed out by Pulay, Saebø, and Meyer,[47] the individual states in the first-order interacting space have the drawback of not being orthogonal to one another. The first-order interacting space is therefore commonly used in conjunction with a set of reciprocal n-particle states defined by

$$\tilde{\Phi}_{ij}^{ab} = \frac{1}{6}(2\Phi_{ij}^{ab} + \Phi_{ji}^{ab}) \tag{43}$$

Rather than projecting against the first-order interacting space, equation (37) is projected against the set of reciprocal states when equation (40) is adopted for Φ_1. The reciprocal states are ideal for this purpose as they satisfy the properties

$$\langle\tilde{\Phi}_{ij}^{ab}|\Phi_{kl}^{cd}\rangle = \delta_{ac}\delta_{bd}\delta_{ik}\delta_{jl} + \delta_{ad}\delta_{bc}\delta_{il}\delta_{jk} \tag{44}$$

$$\langle\tilde{\Phi}_{ij}^{ab}|\Phi^1\rangle = R_{ab}^{ij} \tag{45}$$

that led to relatively simple expressions for the residual matrices. In the case in which orthogonal virtual orbitals are used, the MP2 residuals take the form

$$\mathbf{R}^{ij} = \mathbf{K}^{ij} + \mathbf{f}_{\mathbf{virt}}\mathbf{C}^{ij} + \mathbf{C}^{ij}\mathbf{f}_{\mathbf{virt}} - \sum_k \left(f_{ik}\mathbf{C}^{kj} + \mathbf{C}^{ik}f_{kj}\right) \tag{46}$$

where $\mathbf{f}_{\mathbf{virt}}$ is the virtual–virtual block of the Fock matrix. If canonical Hartree–Fock orbitals are used, the Fock matrix is diagonal and equation (46) can be solved straightforwardly in a noniterative manner

$$C_{ab}^{ij} = K_{ab}^{ij}/(\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j) \tag{47}$$

This is important to note as a number of successful MP2 methods are based upon this approach.[48, 49] If some kind of nonorthogonal virtual orbital basis is used, such as the PAOs, the expression for the MP2 residuals is then given by[24]

$$
\begin{aligned}
\mathbf{R}^{ij} = {}& \mathbf{K}^{ij} + \mathbf{f}_{\mathbf{virt}} \mathbf{C}^{ij} \mathbf{S}_{\mathbf{virt}} + \mathbf{S}_{\mathbf{virt}} \mathbf{C}^{ij} \mathbf{f}_{\mathbf{virt}} \\
& - \sum_k \mathbf{S}_{\mathbf{virt}} (f_{ik} \mathbf{C}^{kj} + f_{kj} \mathbf{C}^{ik}) \mathbf{S}_{\mathbf{virt}},
\end{aligned}
\tag{48}
$$

where $\mathbf{S_{virt}}$ is the virtual–virtual block of the overlap matrix. This latter formulation is, of course, most relevant for local MP2.

We will not make any attempt to derive expressions for MP3. Rather, we will simply list the equations and then point out the computationally dominant components. The MP3 energy is computed as

$$
E^3 = \sum_{i,j} Tr \left[\mathbf{K}^{ij} (2\mathbf{T}^{ij} - \mathbf{T}^{ji}) \right]
\tag{49}
$$

where the components of the T^{ij} matrices consist of the projections of the vector $\left(\hat{H} - E_0 - E_1 \right) |\Phi_0 + \Phi_1\rangle$ onto the components of the first-order interacting space. These matrices are explicitly given by[25]

$$
\begin{aligned}
\mathbf{T}^{ij} = {}& \mathbf{K}^{ij} + \mathbf{K}[\mathbf{C}^{ij}] + \mathbf{f}_{\mathbf{virt}} \mathbf{C}^{ij} \mathbf{S}_{\mathbf{virt}} \\
& + \mathbf{S}_{\mathbf{virt}} \mathbf{C}^{ij} \mathbf{f}_{\mathbf{virt}} + \mathbf{B}^{ij} \mathbf{S}_{\mathbf{virt}} + \mathbf{S}_{\mathbf{virt}} \mathbf{B}^{ji^{\dagger}} \\
& + \mathbf{S}_{\mathbf{virt}} (\mathbf{G}^{ij} + \mathbf{G}^{ji^{\dagger}}) \mathbf{S}_{\mathbf{virt}}
\end{aligned}
\tag{50}
$$

where

$$
\mathbf{B}^{ij} = \sum_k \left[\left(\mathbf{K}^{ik} - \frac{1}{2} \mathbf{J}^{ik} \right) (2\mathbf{C}^{kj} - \mathbf{C}^{jk}) - \frac{1}{2} \mathbf{J}^{ik} \mathbf{C}^{jk} - \mathbf{J}^{jk} \mathbf{C}^{ik} \right]
\tag{51}
$$

$$
\mathbf{G}^{ij} = \sum_{k \geq l} (1 + \delta_{kl})^{-1} \left[(ij|kl) - \delta_{lj} f_{ik} - \delta_{ki} f_{jl} \right] \mathbf{C}^{kl}
\tag{52}
$$

and

$$
\mathbf{K} \left[\mathbf{C}^{ij} \right]_{ab} = \sum_{c,d} (ac|bd) C_{cd}^{ij}
\tag{53}
$$

Here, the $\mathbf{J}^{ij}$ matrices are analogous to the $\mathbf{K}^{ij}$ matrices except where the Coulomb-style *(ij|ab)* integrals are used. The most expensive part of equation (50) is the computation of the $\mathbf{K}[\mathbf{C^{ij}}]$ matrices. Note that this calculation is identical to the calculation that is the bottleneck in SDCI calculations.

The MP4 energy is computed as

$$
E^4 = \langle \Phi_1 | \hat{H}' - E_1 | \Phi_2 \rangle - E^2 \| \Phi_1 \|^2
\tag{54}
$$

where

$$\Phi_2 = -(\hat{H}_0 - E_0)^{-1}(\hat{H}' - E^1)\Phi_1 \tag{55}$$

Similar to MP2, the second-order wave function is expanded in a basis of *n*-particle states and the expansion coefficients are then obtained through an iterative solution of equation (55). We will not go into any detail on how this is done, except to note that the most costly part of the calculation is again the computation of the $\mathbf{K}[\mathbf{C^{ij}}]$ matrices. For a more thorough listing of the MP4 equations, the reader is referred to the article of Saebø and Pulay.[25]

3.2. Iterative Local and Pseudospectral MP*x*
3.2.1. Iterative Local and Pseudospectral MP2
An enormous amount of work has been done in the area of local MP2. Similar to local CI, this work was pioneered by Saebø and Pulay in the mid-1980s[24-26] and work in the field is ongoing.[50] As we have indicated would be the case, the computational strategy followed in Saebø and Pulay's local MP2[50] is quite similar to that followed in local CI. The residuals are only computed for pairs of internal orbitals *i* and *j* that are located near one another. Furthermore, the sets of allowed virtual orbitals are restricted to be PAOs that are in [*ij*]. These restrictions lead to a total number of C_{pq}^{ij} coefficients that scales linearly with the size of the system since for each internal orbital *i*, the *j*, *p*, and *q* must correspond to orbitals localized near *i*. Similar to local CI, the residual is initially formed in the PAO basis and then transformed to a temporary orthogonal basis where it is used to generate an updated set of expansion coefficients. The updated coefficients are then transformed back to the PAO basis, and the cycle is iterated to convergence.

Once the $\mathbf{K^{ij}}$, Fock, and overlap matrices are in place, the calculation of the MP2 residuals is not particularly costly. Since the number of coefficients scales linearly with the size of the system, so do the number of residuals that must be computed. The overall scaling will be asymptotically linear; therefore, the cost of computing each residual is independent of the system size. This can be seen as follows. Each of the matrix multiplications in equation (48) involve matrices of dimension equal to the size of the virtual orbital domain for a pair of internal orbitals, which is a quantity that is independent of the size of the system. Furthermore, the number of *k* that must be included in the sum is also independent of the size of the system, and can be reduced further by screening with values of the Fock matrix. Similar remarks can be made for the computation of the MP2 energy (equation 42).

One additional refinement that has been included in the local MP2 method by various researchers is an enhanced stratification of pairs of internal orbitals. Of course, there are categories for pairs of internal orbitals that are localized close by one another and there is a category for pairs that are localized extremely far from one another. The former are treated without approximation and the latter are neglected entirely. An additional category is characterized as 'distant' and, in the scheme of Saebø and Pulay,[50] consists of pairs of orbitals that are separated by four to five times their diameters. Such pairs individually make a small contribution to the correlation energy, but their overall sum is quite significant. These pairs have been handled through the use of a multipole expansion.[33, 51]

The pseudospectral approximation has also been applied to the MP2 method.[52] However, it has been applied in a manner slightly different from how it is applied in CI. Rather than using the pseudospectral method to accelerate the contraction of MO integrals with expansion coefficients, it is used to accelerate the integral transforms required to form the $\mathbf{K^{ij}}$ matrices. This is done by transforming the individual pseudospectral units separately and then assembling the exchange integrals from the transformed pseudospectral units. The transform for the $\mathbf{Q}$ operator is given by

$$Q_{ig} = \sum_{\nu} L_{\nu i} Q_{\nu g} \qquad (56)$$

and scales as $n_{bas} n_i n_g$. A similar expression holds for the matrix **R**, but the scaling is $n_{bas} n_v n_g$ since a virtual index must be transformed. Each index of the A operator can be transformed as follows:

$$A_{j\sigma}(g) = \sum_{\nu} L_{\nu j} A_{\nu\sigma}(g) \qquad (57)$$

This operation scales as $n_{bas}^2 n_i n_g$, and a similar expression holds for the transformation of the second index. The dominant step in the formation of the $\mathbf{K^{ij}}$ matrices, however, is the final sum over grid points

$$\mathbf{K^{ij}}_{ab} = \sum_{g} Q_{ig} R_{ga} A_{bj}(g) \qquad (58)$$

that scales as $n_i^2 n_v^2 n_g$. Comparing this to the cost of the dominant step in the analytic transform, $n_i n_{bas}^4$, one can see that the pseudospectral treatment will be cheaper as long as $n_i n_v^2 n_g < n_{bas}^4$. Generally, for large molecules, the number of grid points is about an order of magnitude larger than the number of basis functions, and this pseudospectral approach clearly does not lead to a scaling advantage. However, if one uses localized virtual orbitals, one may truncate the sum over the grid points since the values of R_{ga} will quickly approach zero. This approach can lead to significant computational savings, as we will show below. It should be noted that this approach does not require any truncation of the virtual space. The only requirement is that one use some set of localized virtual orbitals.

The prospects for pseudospectral MP2 improve significantly if one also takes advantage of the TOV approximation. In this case, the quantity n_v becomes independent of the size of the system and the formal scaling reduces to $n_i^2 n_g \approx n^3$. This approach has been implemented by Murphy et al.[53] Of course, one could easily take advantage of the WP approximation in this approach as well, which would yield a method that scaled roughly quadratically. To our knowledge, such an approach has not been implemented.

Drawing upon the multiconfigurational Møller–Plesset theories of Wolinski et al.[54, 55] and Andersson et al.,[56] Murphy et al. implemented a version of local pseudospectral MP2 suitable for use with GVB perfect-pairing style reference wave functions.[57] This development is significant as it represents one of the few efforts to develop a reduced scaling electron correlation method for use with multiconfigurational reference functions. Procedurally, the pseudospectral local GVB-MP2 of Murphy et al. is very similar to the single reference version of pseudospectral local MP2. However, some significant differences arise because one cannot simply take the zeroth-order Hamiltonian to be the sum of single-particle Fock operators as is done in equation (35). For the local GVB-MP2 method, the zeroth-order Hamiltonian is taken to be

$$\hat{H}_0 = \hat{h} + \sum_{kl} \rho_{kl} \left(\hat{J}_{kl} - \frac{1}{2} \hat{K}_{kl} \right) \qquad (59)$$

where $\hat{h}$ is the sum of electron kinetic energy and nuclear-attraction operators, ρ_{kl} is an element of the occupied–occupied block of the reference wave function density matrix, and $\hat{J}$ and $\hat{K}$ are the usual Coulomb and exchange operators, respectively. The virtues of this choice of zeroth-order Hamiltonian have been discussed by Andersson et al.[56] The analog of equation (46) when equation (59) is used for the zeroth-order Hamiltonian is given by

$$\mathbf{R}^{ij} = \tilde{\mathbf{K}}^{ij} + \sum_{i',j'} A_{ij,i'j'}(\mathbf{f_{virt}}\mathbf{C}^{i'j'}\mathbf{S_{virt}} + \mathbf{S_{virt}}\mathbf{C}^{i'j'}\mathbf{f_{virt}})$$

$$+ \mathbf{S_{virt}}\left(\sum_{i',j'} B_{ij,i'j'}\mathbf{C}^{i'j'}\right)\mathbf{S_{virt}}$$

(60)

$$\tilde{\mathbf{K}}^{ij} = \sum_{i',j'} D_{ij,i'j'}\mathbf{K}^{i'j'}$$

(61)

Here, the $A_{ij,\ i'j'}$ and $D_{ij,\ i'j'}$ are coupling constants similar to those that appear in MRSDCI (see equation 19, for example), and the $B_{ij,\ i'j'}$ are linear combinations of Fock matrix elements. With the exception of the extra couplings, this equation mirrors the structure of equation (48). Furthermore, it has been argued that the cost of treating these additional couplings is nominal.[57] This method was further generalized by Dunietz et al. for use with GVB restricted configuration interaction (GVB-RCI)[58] wave functions.[59] The equations for such a pseudospectral multiconfigurational local MP2 method are similar to equation (60), but the couplings between different components of the first-order wave function that must be taken into account are more numerous and complicated.

3.2.2. Integral-direct Methods in Local MP2

An extremely important aspect of local MP2 is that integral-direct techniques[60, 62] have been incorporated into the method. These techniques allow one to generate the two-electron integrals in the required MO basis on the fly for immediate use in the MPx calculation. This development allows one to examine extraordinarily large systems since it bypasses the traditional n^5 transformation of the AO integrals to the MO basis and their subsequent storage on disk, an n^4 scaling requirement. Let us consider how this might work in detail for MP2, which requires only the $\mathbf{K}^{ij}$ matrices in addition to the Fock and overlap matrices (equation 48). We will take as an example the scheme of Saebø and Pulay[50] for computing the $\mathbf{K}^{ij}$ matrices, which consists of a pair of two-index transforms that are each handled in a single step. A similar algorithm was also implemented by Rauhut et al.[63] For clarity, it should be noted that this scheme was not implemented in a fully direct manner. Rather, half-transformed integrals were stored on disk and then read into memory as needed for the subsequent half-transform. We begin with a discussion of this scheme since it is relatively straightforward and will pave the way for a discussion of a more complicated integral transform. The scheme of Saebø and Pulay is summarized in Figure 1. The outer loops of the scheme run over pairs of basis function shells (groups of AOs consisting of basis functions on the same center and having the same angular momentum). Immediately inside these loops is an AO integral-prescreening test that is done to eliminate entire evaluations of integral blocks associated with a pair of shells. The prescreening criteria take into account both the magnitudes of the MO transformation coefficients and also the estimated magnitudes of the AO integrals. The latter can be computed beforehand through the use of a Schwarz inequality:

$$|(\mu\rho|\nu\sigma)| \leq |(\mu\rho|\mu\rho)|^{\frac{1}{2}}|(\nu\sigma|\nu\sigma)|^{\frac{1}{2}}$$

(62)

The prescreening in Saebø and Pulay's scheme mandates that the product

$$D_{\mu\rho}^{\mathbf{Max}}(\mu\nu|\rho\sigma)_{est}$$

(63)

exceeds some threshold. Here, $D_{\mu\rho}^{\mathbf{Max}}$ denotes the maximum of $(|L\mu_i L\rho_j|)$ over all orbitals i,j and $(\mu\nu|\rho\sigma)_{est}$ is the maximum value of $(\mu\nu|\lambda\sigma)$ based on equation (62). It should be noted that these quantities are computed not for the individual basis functions, but are further maximized at the shell level. This is done to

accommodate the outer loops over shells, which is the most convenient way to drive the integral evaluation. Should the prescreening criteria be satisfied, a block of $(\mu\nu|\rho\sigma)$ integrals corresponding to all possible values of ν and σ and all values of μ and ρ in the shells M and R will be evaluated.

Figure 1. [Full View]

text

The next layer of loops runs over the integrals in this block. Prescreening then takes place using criteria very similar to equation (63), except now the quantities used to form the prescreening criteria need not be maximized over shells. Unlike the initial prescreening, this fine-grained level of prescreening only serves to eliminate additional contractions of AO integrals with MO expansion coefficients. An important point is that the MO coefficients are sorted in ascending order prior to the transformation. This allows one to use the inner prescreening tests to exit 'do' loops once and for all as soon as the prescreening quantity falls below the prescribed threshold, since one can be certain that subsequent coefficients will be too small for the prescreening quantity to exceed the threshold. Such an arrangement obviates the need to do threshold testing for every single orbital.

A more sophisticated scheme for generating the $\mathbf{K}^{ij}$ matrices was introduced by Schütz et al.[28, 30] and is summarized in Figure 2. This scheme carries out the entire integral transformation on the fly and is fully direct (i.e., no disk storage is utilized). Similar to the scheme of Saebø and Pulay, the outer loops run over AO shells. However, unlike the scheme of Saebø and Pulay, the transform of Schütz et al. begins with a set of loops over a quartet of shells. The evaluation of AO integrals is then prescreened using the same criteria as in the Saebø and Pulay transform. At this stage, the scheme of Schütz et al. diverges significantly from the scheme of Saebø and Pulay. A single index of the AO integrals is transformed first, taking advantage of the permutational symmetry of the indices ρ and σ. This quarter-transform makes heavy use of a sophisticated prescreening mechanism. Obviously, one should consider the sparsity of the vector of MO expansion coefficients for orbital j in attempting to reduce the cost of the first quarter-transform. However, the scheme of Schütz et al. goes one step further and also considers the sparsity of the MO expansion coefficients that are required in subsequent transformation steps. For the second quarter-transformation, for example, this is done by considering the quantity

$$L_{\rho i}^{\max} = \underset{j \in P(ij)}{\mathrm{Max}} L_{\rho j} \tag{64}$$

which is the maximum value of the $L_{\rho j}$ considering all internal orbitals j, which form a strong pair with orbital i. Similarly, for the third and fourth quarter-transforms, which are for the virtual orbitals, one may consider the quantity

$$P_{\mu i}^{\max} = \underset{r \in UP1(i)}{\mathrm{Max}} L_{\mu r} \tag{65}$$

Here $UP1(i)$ denotes the union of PAOs contained within all the internal orbital pair domains involving orbital i. The actual prescreening procedure is quite complicated and involves eliminating rows and columns of

Figure 1.

Pseudocode for the Saebø and Pulay scheme for two-index integral transformations in local MP2. Upper case letters *M* and *R* correspond to shells of AOs. Lower case letters *m*, *n*, *r*, and *s* correspond to AOs and lower case letters *i* and *j* correspond to internal MOs. Integral storage arrays are denoted by *Qx* where *x* is the number of transformed indices.

```
do M = 1, # of shells
    do R = 1, M
          o Prescreen for evaluation of (Mn|Rs) (equation 63 in text)
          o Evaluate (Mn|Rs) integral block if necessary

          do loop over all m in shell M
              do loop over all r in shell R
                  do n = 1, # of basis functions
                      do s = 1, # of basis functions

                            o if ( |(mn|rs)|*Ln,max*Ls,max < threshold)
                              then exit loop over s
                            do i = 1, # of internal orbitals
                                 o if ( |(mn|rs)|*L(n,i)*Ls,max < threshold)
                                   then exit loop over i

                                 do loop over required j


                                     t = (mn|rs)*L(n,i)*L(s,j)
                                     o if ( t < threshold) then exit j loop
                                     Q2(mi|rj) + Q2(mi|rj) + t

                                 end do

                            end do

                      end do
                  end do
              end do
          end do
    end do
end do
```

integral matrices $\mathbf{I}^{\mu\rho}{}_{\nu\sigma} = (\mu\rho|\nu\sigma)$. First, the elimination is done using a coarse-grained filter based upon the product of the quantity $P_{\mu j}^{\max}$ maximized over shells and orbitals j and the quantity $D_{\rho\sigma}^{\text{Max}}$. Briefly, rows and columns of the integral matrices are thrown out if all their elements multiplied by the prescreening quantities fall below a threshold. We refer to this prescreening as coarse-grained since it is based upon the maximum values of prescreening quantities. As a by-product of this coarse-grained procedure, one is able to obtain the maximum values of the integral matrices $\mathbf{I}^{\mu\rho}$. These quantities can then be subsequently used in finer-grained prescreening based directly upon the quantities $L_{\rho i}^{\max}$ and $P_{\mu i}^{\max}$.

<div align="center">

Figure 2. [Full View]

text

</div>

Although the number of half-transformed integrals ($\mu i|\nu j$) asymptotically scales linearly, it has been observed that the onset of linear scaling is quite late and that the number of these integrals is often very large, leading to a significant memory requirement. In the transform scheme of Schütz et al., this memory requirement is avoided by carrying out the second and third-quarter transforms simultaneously. The penalty for this is that the third quarter-transform must be done repeatedly since it takes place within the loop for the second quarter-transform. Fortunately, when the local correlation approximation is used, the total number of shells N, which contributes for a given M, is asymptotically independent of the molecular size and this handling of the third quarter-transformation therefore does not affect the overall scaling of the algorithm. Similar to the first quarter-transformation, prescreening with the $P_{\mu j}^{\max}$ and $L_{\rho j}^{\max}$ (maximized over shells) is employed to reduce the cost of the combined second- and third-quarter transform. The fourth and final quarter-transform is handled straightforwardly by the outer loop over shell M. The algorithm described here does not take advantage of the permutational symmetry ($\mu\nu|\rho\sigma) = (\rho\sigma|\mu\nu)$. However, this permutational symmetry, in fact, can be exploited when only the $\mathbf{K}^{ij}$ matrices are needed and this has been implemented for MP2.[28, 62]

A number of points should be made regarding the integral transforms described above. The integral-prescreening techniques described above can reduce the scaling of a canonical integral transformation from n^5 scaling to roughly n^3 scaling. However, as we have alluded to above, a more dramatic reduction is intimately connected to the use of the local correlation approximation, by which we mean the combination of localized orbitals and the use of both the TOV and WP approximations. This point has been made in great detail by Schütz et al.[28] The local correlation approximation first has the effect of localizing the integral evaluation prescreening density $D_{\mu\lambda}^{\text{Max}}$ to the point that the number of required integral evaluations scales linearly with the size of the system. Similarly, for a given j, the number of signficant elements in the products $L_{\rho j}^{\max}L_{\sigma j}$ and $P_{\mu j}^{\max}P_{\nu j}^{\max}$ becomes independent of the size of the system when the local correlation approximation is used. This leads to linear scaling in both the CPU and memory requirements for the first quarter-transform in Figure 2. Similar remarks can be made for the subsequent transformation steps.

3.2.3. Overview of Performance of Local and Pseudospectral MP2
Through the use of the methods described in Sections 3.2.1 and 3.2.2, MP2 calculations on extraordinarily large systems have become possible. For example, the local integral-direct MP2 method of Schütz et al.[28] has enabled calculations on systems such as a chain of 22 glycine molecules. Such a calculation involved 492 correlated electrons and 1586 basis functions but required only about 8 h of CPU time on a modest Pentium III computer. Furthermore, for calculations on glycine chains having more than about 15 subunits, the required

Reduced Scaling Electron Correlation Methods

Figure 2.

Pseudocode for the Schütz and Werner scheme for integral transformations in local MP2. Upper case letters *M*, *N*, *R*, and *S* correspond to AO shells. Lower case letters *i* and *j* correspond to internal MOs. Lower case letters *r* and *s* correspond to PAOs (note that this notation differs from Figure 1). Integral storage arrays are denoted by *Qx* where *x* is the number of transformed indices. The loops over individual basis functions are neglected below but are handled as matrix multiplies in the actual algorithm.

```
do M = 1, # of shells
    do N = 1, # of shells

        o Reinitialize Q1 memory
        do R = 1, # of shells
            do S = 1, R
                o Prescreen for (MN|RS) integral evaluation (equation 63 of text)
                o Evaluate (MN|RS) integral block
                o Prescreen integral block for first quarter-transform
                  (equations 64 and 65 of text)
                do i = 1, # of internal orbitals
                    Q1(MN|Ri) = Q1(MN|Ri) + (MN|RS)*L(S,i)
                    Q1(MN|Si) = Q1(MN|Si) + (MN|RS)*L(R,i)
                end do
            end do
        end do

        do loop over computed (MN|Ri) blocks
            o Prescreen for second and third quarter-transforms
            do loop over required j
                do loop over r in [ij]
                    Q3(jM|ri) = Q3(jM|ri) + {Q1(MN|Ri)*L(N,j)}*L(R,r)
                end do
            end do
        end do
    end do
    do loop over strong pairs ij
        o Prescreen for fourth quarter-transform
        do loop over s in [ij]
            Q4(js|ri) = Q4(js|ri) + Q3(jM|ri)*L(M,s)
        end do
    end do
end do
```

Encyclopedia of Computational Chemistry
Published by John Wiley & Sons, Ltd

CPU time exhibits linear scaling. Similar calculations were performed on polyglycines by Saebø and Pulay[50] to benchmark (*see* Benchmark Studies on Small Molecules) their local MP2 method and the results were comparable to those of Schütz et al. Calculations on polyglycine molecules are ideal for showcasing the local correlation approximation since they are well extended in a linear chain. Calculations on less extended systems, such as $(H_2O)_{60}$, also become amenable through the use of the local correlation approximation. However, the linear scaling regime becomes more difficult to achieve and Schütz et al. only obtained slightly less than quadratic scaling. This is still a considerable achievement.

With regards to precision, the local MP2 method also performs quite well. For molecules having 16 to 30 heavy atoms, Saebø and Pulay[50] were able to recover all but a few hundredths of millihartree of the correlation energy when about half of the total orbital pairs were correlated. Very similar errors were observed by Schütz et al.[28] in benchmark calculations on a polyglycine molecule having eight subunits. In this case, the local MP2 energy differed from the nonlocal energy by only 413 μE_h. For the purposes of examining chemical processes, these errors are entirely negligible.

For large molecules, the pseudospectral variants of MP2 do not offer the same level of computational savings that the local versions of MP2 offer. However, for smaller systems in which the local correlation approximation is not particularly effective, pseudospectral MP2 is roughly fivefold less costly than analytic MP2.[36, 52] Furthermore, the computational savings will grow as one increases the ratio of the number of basis functions to the number of grid points. An additional advantage of the pseudospectral approach is that no approximations are made to the wave function, making pseudospectral MP2 energies more comparable to analytic MP2 energies. However, the pseudospectral method used by itself does not lead to a formal scaling advantage for MP2 calculations.

The situation is of course different when the TOV approximation is incorporated into pseudospectral MP2, as has been done by Murphy et al.[53] This method exhibited scaling slightly better than n^3. While this scaling is not as good as the linear scaling observed in fully local MP2 methods, the prospects for a pseudospectral MP2 method that takes advantage of both the WP and TOV approximations remain positive.

Similar to the CI method, the pseudospectral MP2 is observed to give total energies, which differ from the analytic energies by no more than a few tenths of a millihartree. The error is reduced by roughly an order of magnitude more when special analytic corrections[64] are applied to the pseudospectral integrals. When the local correlation approximation is introduced, the errors are of course increased. However, an exhaustive study[53] was done by Murphy et al. to measure the accuracy of the local pseudospectral MP2 method for the determination of conformational energy differences, and the results were very encouraging. Briefly, the energies of conformational changes in 36 organic molecules were considered and the values determined by local MP2 differed from experimentally determined values by an average of only 0.25 kcal mol^{-1}. This level of accuracy is sufficient for the overwhelming majority of applications, and the local pseudospectral method has been included in the commercially available Jaguar quantum chemistry package.[65] Furthermore, when the local pseudospectral MP2 method is used in conjunction with GVB perfect-pairing references this accuracy improves by 0.07 kcal mol^{-1}.[57]

3.2.4. Iterative Local and Pseudospectral MPx x > 2
Compared to MP2, far less work has been done on higher order perturbation theory: Saebø and Pulay considered local MP3 and MP4 in detail[24-26] and Martinez and Carter implemented a pseudospectral version of MP3.[36, 52] We shall briefly describe each of these studies, starting with the local MP3 and MP4 of Saebø and Pulay.

Overall, the local MP3 and MP4 methods of Saebø and Pulay are very similar to their local MP2 method. The principal difference is their handling of weak pairs. The most important point is that all pairs of orbitals, regardless of their separation, are *at least* included at the MP2 level. Without additional treatment of the WP, this approach still neglects coupling between WP and coupling between strong and WP. Thus, higher orders of perturbation theory still retain contributions from the WP, although more significant approximations are made. At the MP3 level of theory, for example, the diagonal terms

$\mathbf{K}[\mathbf{C^{ij}}] - \mathbf{J^{jj}C^{ij}S_{virt}} - \mathbf{S_{virt}C^{ij}J^{ii}} + (ii|jj)\mathbf{S_{virt}C^{ij}S_{virt}}$ are neglected in equation (50) if the pair ij is weak.[25]

Additionally, one neglects the terms $\mathbf{C^{ik}}$ from the expression for $\mathbf{B}^{ij}$ (equation 51) if the pairs of orbitals ik, jk, and ij are all characterized as weak.[25] The treatment of each of these terms exhibits n^6 scaling so the computational savings can be substantial. At the MP4 level, the contributions of the WP are neglected entirely, and the quadruple excitations were handled using a more sophisticated set of approximations that are similar to those used at the MP3 level of the theory.[25]

Compared to the savings in the local SDCI method of Saebø and Pulay, the savings from their local MP3 and MP4 methods were more modest.[26] For calculations on difluoroethylene, for example, the local MP4 method reduced the required CPU time by at most fourfold when a triple-ζ basis set was used. At the time of their work, Saebø and Pulay did not have the computational resources to carry out nonlocal MP4 on significantly larger systems. Thus, they were unable to measure the computational savings of the local MP4 method for larger systems in which the local correlation approximation is expected to be most effective. Nevertheless, one could be certain that even for larger systems the overall savings would still be less than what was observed in the SDCI case. This is attributable to the method they adopted for computing the $\mathbf{K}[\mathbf{C^{ij}}]$ matrices. In the most straightforward approach, one would precompute the required block of integrals in the PAO basis for every pair of orbitals ij and store them on disk in a direct access file. These matrices would then be read into memory as needed and contracted with the first-order wave function expansion coefficients. Unfortunately, this approach relies heavily on the use of disk storage, which was a limiting factor for Saebø and Pulay. An alternative approach is to compute the $\mathbf{K}[\mathbf{C^{ij}}]$ matrices in the full AO basis. Although it is possible to take advantage of the local correlation approximation in such an approach, this was not done by Saebø and Pulay, and the calculation was dominated by the n^6 scaling calculation of the $\mathbf{K}[\mathbf{C^{ij}}]$ matrices. The only savings obtained were derived from the WP approximation. This is offset slightly by the fact that unlike SDCI, MP3 and MP4 do not require repeated formation of the $\mathbf{K}[\mathbf{C^{ij}}]$ matrices.

Clearly, Saebø and Pulay were very conservative in their application of the local correlation approximation to MP3 and MP4. As a result, their method was quite precise: for a calculation on difluoroethylene with a 6-311G** basis, the correlation energy was only a few hundredths of a percent higher than the nonlocal correlation energy.[26]

In the pseudospecral MP3 method of Martinez and Carter[52] the $\mathbf{K}[\mathbf{C^{ij}}]$ matrices were formed in a manner identical to equation (24), and a very similar approach was adopted for the $\mathbf{B}^{ij}$ terms. As we have indicated in Section 2.4, this leads to a scaling advantage of approximately n_v^2/n_g in the treatment of these terms. Furthermore, since the computation of these terms consumes roughly 90% of the CPU's time, this leads to a similar scaling advantage for the overall method. An extension of the method to include a local treatment of the WP would be trivial and would further enhance the computational savings. However, this has not been implemented. Similar to the pseudospectral MP2 method, the errors incurred by the use of pseudospectral methods in MP3 were quite small. In a series of 10 test calculations carried out by Martinez and Carter on systems ranging in size from water to glycine, the pseudospectral MP3 approach gave energies that differed from the analytic energies by no more than 0.06 millihartrees.

3.3. Noniterative Local MP2

As we have noted in Section 3.1, one can easily solve the MP2 equations analytically when canonical HF orbitals are used. The resulting MP2 energy expression takes the fairly simple form

$$E^2 = \sum_{i,j,a,b} \frac{(ia|jb)[2(ia|jb) - (ib|ja)]}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \tag{66}$$

Direct evaluation of this expression scales as n^5 owing to the integral transformation required to generate the $(ai|bj)$. Despite this steep scaling, a significant amount of research has been devoted to the direct evaluation of equation (66). Such methods generally proceed by first rewriting equation (66) in some sort of localized basis in which one can exploit the decay properties of the basis. A particularly ingenious approach to doing this was devised by Almlöf,[48] who proposed using a Laplace transform to rewrite equation (66) as

$$E^2 = \int_0^\infty \sum_{i,a,j,b} (i(t)a(t)|j(t)b(t))[2(i(t)a(t)|j(t)b(t)) \\ - (i(t)b(t)|j(t)a(t))] \tag{67}$$

where

$$i(t) = ie^{((\epsilon_i - \epsilon_F)t/2)} \tag{68}$$

$$a(t) = ae^{(-(\epsilon_a - \epsilon_F)t/2)} \tag{69}$$

and ϵ_F is a parameter that can be chosen to enhance numerical stability. When written in this form, the denominator of equation (66) no longer blocks the use of linear transformations to re-express the equation in some localized basis. Of course, the penalty to be paid is the additional integration associated with the Laplace transform. Very efficient quadrature schemes that require 10 or fewer points have been devised for this purpose,[66, 67] however, so the penalty is manageable. Introducing the numerical quadrature, equation (67) becomes

$$E^2 = \sum_\alpha w_\alpha e_\alpha \tag{70}$$

where w_α is a quadrature weight and

$$e_\alpha = \sum_{i,a,j,b} (i_\alpha a_\alpha|j_\alpha b_\alpha)[2(i_\alpha a_\alpha|j_\alpha b_\alpha) - (i_\alpha b_\alpha|j_\alpha a_\alpha)] \tag{71}$$

Here, i_α is simply defined as $i(t_\alpha)$. From this point forward, we will only consider the evaluation of a single e_α. The most obvious localized basis in which to rewrite equation (71) is simply the underlying AO basis. Introducing the MO expansion coefficients and absorbing the exponentials into them, equation (71) can be written as

$$e_\alpha = \sum_{\mu,\nu,\chi,\lambda} (\bar{\nu}\underline{\mu}|\bar{\chi}\lambda)[2(\nu\mu|\chi\lambda) - (\nu\lambda|\chi\mu)] \tag{72}$$

where

$$(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda}) = \sum_{\mu',\nu',\lambda',\chi'} (\nu'\mu'|\chi'\lambda') \bar{D}^\alpha_{\nu'\nu} \underline{D}^\alpha_{\mu'\mu} \bar{D}^\alpha_{\chi'\chi} \underline{D}^\alpha_{\lambda'\lambda} \tag{73}$$

and

$$\underline{D}^\alpha_{\mu\nu} = \sum_{i}^{occupied} L_{\mu i} L_{\nu i} e^{(\epsilon_i - \epsilon_F)t_\alpha} \tag{74}$$

$$\bar{D}^\alpha_{\mu\nu} = \sum_{a}^{virtual} L_{\mu a} L_{\nu a} e^{(\epsilon_a - \epsilon_F)t_\alpha} \tag{75}$$

Now that the MP2 energy has been re-expressed in a suitable localized basis, one can attempt to evaluate the expression at a reduced cost through agressive use of integral prescreening. First, one can eliminate evaluations of the integral $(\nu'\mu'|\chi'\lambda')$ unless a precomputed upper bound to its overall contribution to the MP2 energy exceeds some threshold. Such an upper bound can be computed as

$$\mathbf{Q}_{\nu'\mu'}\mathbf{Q}_{\chi'\lambda'}[2\mathbf{Z}_{\nu'\mu'}\mathbf{Z}_{\chi'\lambda'} + \mathbf{Z}_{\nu'\lambda'}\mathbf{Z}_{\chi'\mu'}] \tag{76}$$

where

$$\mathbf{Q}_{\nu\mu} = (\nu\mu|\nu\mu)^{\frac{1}{2}} \tag{77}$$

$$\mathbf{X}_{\nu\mu} = (\bar{\nu}\mu|\bar{\nu}\mu)^{\frac{1}{2}} \tag{78}$$

$$\mathbf{Y}_{\nu\mu} = (\underline{\nu}\mu|\underline{\nu}\mu)^{\frac{1}{2}} \tag{79}$$

$$\mathbf{Z}_{\mu\nu} = min\left(\sum_\chi \mathbf{X}_{\nu\chi}|\underline{D}_{\chi\mu}|, \sum_\chi \mathbf{Y}_{\mu\chi}|\bar{D}_{\chi\nu}| \right) \tag{80}$$

One can easily show that equation (76) represents an upper bound to the contribution of the integral $(\nu'\mu'|\chi'\lambda')$ to the energy by noting that $\mathbf{Q}_{\nu\mu}\mathbf{Q}_{\chi\lambda} \geq |(\nu\mu|\chi\lambda)|$ and $\mathbf{Z}_{\nu\mu}\mathbf{Z}_{\chi\lambda} \geq |(\bar{\nu}\underline{\mu}|\bar{\chi}\underline{\lambda})|$. The former relationship can be derived from the Schwarz inequality, while the latter inequality can be proven through the use of both the Schwarz and triangle inequalities. If this thresholding test mandates the computation of $(\nu'\mu'|\chi'\lambda')$, this integral can then be used in the place of $\mathbf{Q}_{\nu'\mu'}\mathbf{Q}_{\chi'\lambda'}$ in equation (76) to form a quantity that can be used to prescreen the transform of the $(\nu'\mu'|\chi'\lambda')$ to $(\bar{\nu}\mu'|\chi'\lambda')$. One can then prescreen the second quarter-transformation using the quantity

$$|(\bar{\nu}\mu'|\chi'\lambda')|[2\mathbf{Y}_{\mu'\nu}\mathbf{Z}_{\chi'\lambda'} + \mathbf{Y}_{\lambda'\nu}\mathbf{Z}_{\chi'\mu'}] \tag{81}$$

Analogous prescreening quantities can be used at every stage in the construction of equation (71). An approach that is more or less identical to this has also been explored for MP4.[66]

Using the approach described above, Häser developed an MP2 method, which was observed to scale only as n^4.[67] Later, Ayala and Scuseria showed that this Laplace–MP2 method, in fact, asymptotically scaled quadratically in systems with large HOMO-LUMO gaps.[68] This fact was proven by considering two facts. First, they noted that the integrals $(\lambda\mu|\chi\nu)$ decay exponentially with $\lambda-\mu$ separation and $\chi-\nu$ separation. This reduces the number of significant $(\lambda\mu|\chi\nu)$ to something on the order of n^2. Second, they noted that in systems with large HOMO-LUMO gaps, the number of significant elements in the $D^\alpha_{\mu\nu}$ and $\bar{D}^\alpha_{\mu\nu}$ matrices is somethng on the order of n. The combination of these two facts reduces the significant $(\bar{\mu}\nu|\bar{\lambda}\chi)$ integrals to those in the class where ν is located near μ and χ is located near λ. The number of such integrals scales asymptotically as n^2, which insures that the Laplace–MP2 method asymptotically scales quadratically. Of course, the most significant surviving $(\bar{\mu}\nu|\bar{\lambda}\chi)$ integrals will be the $(\bar{\mu}\underline{\mu}|\bar{\lambda}\underline{\lambda})$, which was shown by Ayala and Scuseria to decay in magnitude as $1/R^3_{\lambda\mu}$.[68]

The $1/R^3_{\lambda\mu}$ decay of the $(\bar{\mu}\nu|\bar{\lambda}\chi)$ was further exploited by Ayala and Scuseria to implement a linear scaling Laplace–MP2 method as follows. In the vein of the local correlation approximation, a spherical interaction domain was defined for each basis function μ. The radius of the interaction domain for μ was taken to be the distance from μ of the furthest $\gamma\delta$ charge distribution satisfying

$$(\underline{D}^\alpha_{\mu\gamma}S_{\gamma\delta}\bar{D}^\alpha_{\delta\mu})^2 \geq \epsilon \tag{82}$$

(the absolute positions of the $\gamma\delta$ charge distributions were defined as atomic number weighted averages of the positions of the atoms upon which γ and δ were centered). These interaction domains were chosen so that the orthogonality between the occupied and virtual spaces, expressed as

$$\underline{D}^\alpha \mathbf{S}\bar{D}^\alpha = 0 \tag{83}$$

was satisfied to within some tolerance for each basis function μ. Essentially, equation (82) picks out all the significant contributions to equation (83), which insures that

$$\left| \sum_{\gamma\delta\in\mu} \underline{D}^\alpha_{\mu\gamma}S_{\gamma\delta}\bar{D}^\alpha_{\delta\mu} \right| \leq \epsilon_2 \tag{84}$$

Armed with well-defined interaction domains, linear scaling MP2 can be achieved by neglecting integrals such as $(\bar{\mu}\underline{\mu}|\bar{\lambda}\underline{\lambda})$ if μ and λ belong to interaction domains whose edges are separated by further than some predefined distance criteria.

Using the scheme outlined above, Ayala and Scuseria observed linear scaling of the computational effort for calculations on polyglycines having greater than 12 subunits.[68] The same was observed for water clusters including more than 96 individual molecules. It should be stressed that this method should not be regarded as an approximation to canonical MP2. Rather, it should be regarded as a formulation of the MP2 method in the AO basis that reproduces the canonical MP2 energies to within a very tight tolerance. Scuseria and Ayala, in fact, made detailed arguments regarding how the error in the MP2 method could be systematically controlled through the adjustment of ϵ and the distance beyond which interactions between domains were neglected. For the previously described calculations on polyglycines and water clusters, the error of Scuseria and Ayala's linear scaling MP2 relative to canonical MP2 was no more than a few hundredths of a microhartree. A drawback of this method is that one must compute e_α for each quadrature point, which causes the method to scale linearly with a very steep prefactor. Furthermore, the method does not become linear scaling until the size of the molecule reaches a regime where the number of significant $(\mu\nu|\lambda\chi)$ is of the order n^2.

Another approach to the direct evaluation of the MP2 energy was developed by Head-Gordon et al.[20, 49, 69] Similar to the approach of Ayala and Scuseria, both nonorthogonal occupied and virtual orbital spaces were used. As long as the occupied and virtual spaces remain orthogonal to one another, the MP2 energy can be expressed as

$$E^2 = -\frac{1}{4}\vec{V}^T \cdot \mathbf{\Delta}^{-1} \cdot \vec{V} \tag{85}$$

where

$$V_{aibj} = (ai|bj) - (bi|aj) \tag{86}$$

$$\begin{aligned}
\Delta_{aibj;a'i'b'j'} = {} & f_{aa'}S_{ii'}S_{bb'}S_{jj'} \\
& - S_{aa'}f_{ii'}S_{bb'}S_{jj'} \\
& + S_{aa'}S_{ii'}f_{bb'}S_{jj'} \\
& - S_{aa'}S_{ii'}S_{bb'}f_{jj'}
\end{aligned} \tag{87}$$

The expression for Δ can also be expressed as a direct product

$$\begin{aligned}
\Delta = {} & \mathbf{f_{virt}} \otimes \mathbf{S_{occ}} \otimes \mathbf{S_{virt}} \otimes \mathbf{S_{occ}} \\
& + \mathbf{S_{virt}} \otimes \mathbf{f_{occ}} \otimes \mathbf{S_{virt}} \otimes \mathbf{S_{occ}} \\
& + \mathbf{S_{virt}} \otimes \mathbf{S_{occ}} \otimes \mathbf{f_{virt}} \otimes \mathbf{S_{occ}} \\
& + \mathbf{S_{virt}} \otimes \mathbf{S_{occ}} \otimes \mathbf{S_{virt}} \otimes \mathbf{f_{occ}}
\end{aligned} \tag{88}$$

This can be further factorized as

$$\Delta = F^2 \otimes S^2 + S^2 \otimes F^2 \tag{89}$$

where

$$F^2_{ai;a'i'} = f_{aa'}S_{ii'} - S_{aa'}f_{ii'} \tag{90}$$

$$S^2_{ai;a'i'} = S_{aa'}S_{ii'} \tag{91}$$

If one simultaneously diagonalizes both F^2 and S^2, then the matrix Δ will also be diagonal and evaluation of equation (85) becomes trivial. Unfortunately, if the full space of doubly excited configurations is used, the diagonalization of F^2 and S^2 is still a formidable problem. Thus, the set of doubly excited configurations must be reduced somehow for this method to become viable. Furthermore, they must be reduced in such a way that Δ can still be written in the direct product form of equation (89). This can be achieved if the set of doubly excited configurations, denoted as $\{aibj\}$, also has a direct product structure[70, 71]

$$\{aibj\} = \{ai\} \otimes \{bj\} \tag{92}$$

In other words, for every internal orbital i, a set of allowed virtual orbitals is defined. The full space of allowed

doubly excited configurations is then formed by taking all possible pairs of these pairs of orbitals. In the work of Head-Gordon et al.,[20, 49, 69] localized occupied and virtual orbitals were first associated with individual atoms. For a given localized internal orbital, the allowed set of virtual orbitals was taken to be the set of virtuals associated with the same atoms as the internal orbital. Thus, the configurations are defined on an atom-by-atom basis, and Head-Gordon et al. therefore dubbed their approach the 'atoms-in-molecules' method. Compared to the previously discussed local correlation approaches,[14] the local correlation space in the atoms-in-molecules approach does not include charge transfer terms in which both of the virtual orbitals in a doubly excited configuration are localized near only one of the internal orbitals. Such terms are important for capturing short-range correlations, and one must systematically relax the criteria for determining whether or not a virtual orbital is localized near a particular internal orbital to correct for this deficiency of the atoms-in-molecules approach.

A number of approaches can be taken to associate localized occupied and virtual orbitals with individual atoms. For the virtual orbitals, the simplest approach is to simply use the PAOs, which are naturally associated with a particular atom by virtue of their generation from atom-centered basis functions. Similarly, one can generate occupied orbitals that are associated with a given atom by projecting the virtual space from individual AOs. This approach, as well as more sophisticated approaches, has been explored by Head-Gordon et al.[69, 72]

The size of the configuration space in the atoms-in-molecules method can be tuned by loosening the criteria used to determine whether or not a given occupied orbital is associated with a particular atom. For example, one can consider internal orbitals associated with nearest-neighbor and next-nearest-neighbor atoms to also be associated with a particular atom. Using a nearest-neighbor criterion, the atoms-in-molecules MP2 method consistently recovers greater than 98% of the correlation energy. Including up to the next-nearest neighbors, increased this fraction to 99.8%. When no neighbor atoms are used, the fraction of the correlation energy drops significantly to about 92%, so such calculations are not particularly useful. No CPU times were reported in the work of Head-Gordon et al., so it is difficult to state how well the atoms-in-molecules MP2 method performs relative to various formulations of canonical MP2. However, in the case that nearest-neighbor atoms are used to associate internal orbitals with a given atom, the number of configurations was reduced by slightly more than fivefold relative to nonlocal MP2. Furthermore, Head-Gordon et al. point out that their method does not require the storage of any wave function amplitudes. This may enable the atoms-in-molecules approach to be used in correlation methods that include triple excitations where the storage of all the amplitudes can become prohibitive even when the local correlation approximation is used. Such a CC method will in fact be discussed below.

Reduced Scaling Electron Correlation Methods

Emily A. Carter, University of California, Los Angeles, CA, USA
Derek Walter, University of California, Los Angeles, CA, USA

4. Coupled Cluster Theory

4.1. Conventional Coupled Cluster Theory

Although MP2 is still the most popular electron correlation method, more expensive CC methods have become increasingly popular as computers have become faster. In fact, CC methods may soon become the most popular electron correlation techniques available. Here, we briefly introduce the relevant aspects of CC theory needed for our subsequent discussion of reduced scaling CC. For a comprehensive review of conventional CC theory, the reader is referred to the excellent review article of Crawford and Schaefer.[73]

All CC methods are based upon an exponential ansatz for the wave function

$$\Phi_{CC} = e^{\hat{T}}\Phi_0 \tag{93}$$

where

$$\hat{T} = 1 + \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots \tag{94}$$

Here, the $\hat{T}_i$ are operators that generate various levels of excited configurations. Similar to MPx, the most efficient way to formulate the CC method is in terms of the first-order interacting space. In this case, the $\hat{T}_2$ operator can be written as

$$\hat{T}_2 = \frac{1}{2}\sum_{i,j,a,b} T_{ab}^{ij}\hat{E}_{ai}\hat{E}_{bj} \tag{95}$$

Unlike MP2, we denote CC expansion coefficients as T_{ab}^{ij} rather than C_{ab}^{ij}. This is done not only to distinguish the CC amplitudes from the MP2 amplitudes but also because this is a universal convention for denoting CC amplitudes. The exponential ansatz is employed in order to insure proper scaling of the correlation energy with the size of the system, a property not held by the CI wave function.

As in the case of CI, the most common CC method is CCSD, which is defined by truncating equation (94) after $\hat{T}_2$. The CCSD amplitudes are then determined by projecting the Schrödinger equation against the reciprocal n-particle states defined in equation (43). This leads to the following set of nonlinear equations, which, as in the case of CI and MPx, can be solved in an iterative manner through repeated computation of residuals

$$\langle \tilde{\Phi}_i^a |(\hat{H} - E^{\mathrm{CCSD}})(1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^2 + \hat{T}_1\hat{T}_2 + \frac{1}{3!}\hat{T}_1^3)|\Phi_0\rangle = 0 \tag{96}$$

$$\langle \tilde{\Phi}_{ij}^{ab} | (\hat{H} - E^{\text{CCSD}})(1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^2 + \hat{T}_1\hat{T}_2 + \frac{1}{3!}\hat{T}_1^3 + \frac{1}{2}\hat{T}_2^2$$
$$+ \frac{1}{4}\hat{T}_1^4)|\Phi_0\rangle = 0 \tag{97}$$

We will not list the expressions for the residuals. In subsequent sections, we will simply highlight individual terms in order to discuss important aspects of reduced scaling CC methods. The interested reader can refer to the article of Crawford and Schaefer[73] for a listing of the CC equations in a canonical-orbital basis and the article of Hampel and Werner[27] for a listing of the equations in a basis having a nonorthogonal virtual space. Once the amplitudes have been determined, the CCSD energy can be computed as

$$E^{\text{CCSD}} = \langle \Phi_0 | \hat{H}(1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2}\hat{T}_1^{\,2})|\Phi_0\rangle \tag{98}$$

The connection between perturbation theory and CC is well understood. For example, if one retains all the terms in the CCSD equations, which are of second-order in the Møller–Plesset sense, one arrives at the MP2 equations. The same comment can be made for the third-order terms in the CCSD equations and MP3. However, one is not able to reproduce the MP4 equations from the fourth-order terms of the CCSD equations. For this task, the 'connected' triple excitations that arise from the $\hat{T}_3$ operator are also needed. These facts strongly suggest that one could improve CCSD by computing the contribution of the triples to the MP4 energy[74] and then adding this to E^{CCSD}. This is the general principle behind a large number of corrections to the CCSD method, the most popular of which is the CCSD(T) method.[75] In CCSD(T), the coefficients of the triple excitations are solved in terms of the first-order doubles coefficients, which are in turn approximated as converged CCSD amplitudes. The CCSD(T) approximation also includes an additional term that first appears in the fifth order of perturbation theory. This term was proposed by Raghavachari et al.,[75] and there is a great deal of empirical evidence supporting its use. In fact, the CCSD(T) method is considered by many to be the most accurate *ab initio* quantum chemistry method that can be practically applied to closed-shell main-group molecules. For molecules outside this realm, the accuracy degrades substantially. We shall not yet show the CCSD(T) equations in full. For now, we simply note that the most common formulation of the CCSD(T) method involves the computation of a large number of terms similar to

$$W_{ijk}^{abc} \leftarrow \sum_d T_{ad}^{ij}(bd|ck) \tag{99}$$

Calculation of this term exhibits n^7 scaling, making the CCSD(T) correction extremely expensive. A slightly different formulation has been adopted for the purpose of implementing local CCSD(T),[31, 76] and we shall discuss this in more detail below.

4.2. Local CCSD
Local CCSD was first fully implemented by Hampel and Werner.[27] Their method shared many of the features of previously discussed local correlation techniques. The virtual space consisted of PAOs, and the residuals were updated in a temporary orthogonal basis. Furthermore, there was no neglect of weak pairs. At the very least, such pairs were treated at the MP2 level of theory. Overall, the accuracy of the method was quite high. In calculations on organic molecules ranging in size from water to hexatriene using a 6-31G** basis set, greater than 98% of the correlation energy was always recovered, and in most cases greater than 99% of the correlation energy was recovered. The computational savings were also significant. For a calculation on octane, for example, the total CPU time decreased from 89 066 to 12 088 s on going from conventional CCSD to local CCSD with perturbative treatment of the weak pairs. In the local CCSD calculation, roughly 90% of the CPU time was consumed by the construction of the $\mathbf{K}[\mathbf{T}^{ij}]$ matrices (defined as in equation (53) except

using the CC amplitudes), which were handled in the AO basis in a manner identical to that described in Section 2.5. Thus, in the treatment of the all-external integrals, the only savings were derived from the WP approximation. This ultimately limited the scaling reductions that could be obtained from this local CCSD method.

It is worthwhile to highlight one additional point regarding the calculation of the local CCSD residuals. Classes of terms with the structure

$$\mathbf{G^{ij}} \leftarrow \mathbf{S_{virt}} \sum_k \mathbf{T^{ki}J^{kj}} \tag{100}$$

are required quite often. In conventional CCSD, it is obviously most efficient to first carry out the sum over k and then do the matrix multiplication with $\mathbf{S_{virt}}$. When the local correlation approximation is used, the situation can be slightly different. If the sum over k is carried out first, the side length of the resulting matrix $\sum_k \mathbf{T^{ki}J^{kj}}$ is equal to the dimension of the union of all domains that involve orbital i. This matrix must then of course be multiplied by an enlarged block of the PAO-overlap matrix $\mathbf{S_{virt}}$. Depending on how large this matrix becomes, it may in fact be more efficient to simply move the matrix multiplication with $\mathbf{S_{virt}}$ inside the sum over k, which will then only require matrix multiplications involving small matrices of dimensions equal to typical domain sizes. Most local correlation programs include subroutines for computing these types of terms in either manner, and the programs generally determine the most efficient route automatically prior to the start of the calculation.

The work of Hampel and Werner was recently extended by Schütz and Werner,[30] who developed a local CCSD method that exhibits linear scaling in CPU, disk, and memory requirements. In addition to employing the local correlation approximation, their method makes use of integral-direct techniques (Section 3.2.2) to handle almost all integrals. Furthermore, a multipole expansion can be used to evaluate the exchange integrals in the MP2 treatment of distant pairs (an additional class of less widely separated pairs is handled using analytical local MP2, and the strong pairs are of course treated using full CCSD).

Most of the terms on the right-hand side of the expressions for the CCSD $\mathbf{R}^{ij}$ explicitly involve index i or j. For these terms, it is generally straightforward to see how linear scaling can be achieved using arguments similar to those used in previous sections. However, the CCSD residual expressions also involve terms such as

$$\mathbf{R^{ij}} \leftarrow \mathbf{S_{virt}T^{ij}} \sum_{k,l} \mathbf{K^{kl}T^{lk}S_{virt}} \tag{101}$$

Since indices l and k are completely different from indices i and j, one might conclude that it is only possible to achieve quadratic scaling in the evaluation of these terms. However, Schütz and Werner[30] showed that it is in fact possible to achieve linear scaling in the evaluation of these terms. Their argument was made by considering an individual element of the $\mathbf{R}^{ij}$ matrix:

$$R_{rs}^{ij} \leftarrow \sum_{t,u\in[ij]} \sum_{k,l} \sum_{v,w\in[kl]} S_{rt} T_{tu}^{ij} (uk|lv) T_{vw}^{lk} S_{ws} \tag{102}$$

Consider the case that kl is far from ij in equation (102). Under these circumstances, r, s, t, and u are far from v, w, k, and l. As a result, both S_{ws} and $(uk|lv)$ will become very small since they decay exponentially with w–s and u–k separation, respectively. Thus, the sum over k, l in equation (102) need not include terms in which the pair kl is far from pair ij, allowing one to recover linear scaling in the evaluation of equation (101).

An additional noteworthy aspect of the linear scaling CCSD method of Schütz and Werner is their handling of the $\mathbf{K}[\mathbf{T}^{ij}]$ matrices. Similar to the local CCSD method of Hampel and Werner and the local MP3 and MP4 methods of Saebø and Pulay, the terms are computed in the AO basis and subsequently transformed to the PAO basis. However, Schütz and Werner incorporated a great deal of prescreening into their evaluation of the $\mathbf{K}[\mathbf{T}^{ij}]$ matrices. The prescreening procedure begins with the computation of a test density D_{RS} as

$$D_{RS} = \underset{ij}{\mathrm{Max}} \; \underset{\rho \in R, \sigma \in S}{\mathrm{Max}} \; T_{\rho\sigma}^{ij} \tag{103}$$

where R and S denote shells of AOs. This quantity can be used in combination with a Schwarz inequality estimate of the largest possible integral involving functions in the shell quartet $MRSN$, I_{MRSN}, to eliminate evaluations of the block of AO integrals corresponding to the shell quartet $MRSN$. Since the orbitals i and j are localized, the value of D_{RS} will be small unless R and S correspond to shells located near orbitals i and j. Furthermore, since the two-electron integrals decay exponentially with M–R and S–N separation, this prescreening step is sufficient to give linear scaling in the number of integrals that must be computed. In a subsequent prescreening step, the contractions of the AO integrals with the amplitudes are skipped for a given pair of internal orbitals ij if the quantity $\mathrm{Max}_{\rho \in R, \sigma \in S} T_{\rho\sigma}^{ij}$ multiplied by I_{MRSN} is below some threshold. This prescreening step finally yields linear scaling in the $\mathbf{K}[\mathbf{T}^{ij}]$ evaluation. We note in passing that Schütz has recently developed an additional semi-direct algorithm for computation of the $\mathbf{K}[\mathbf{T}^{ij}]$ that avoids storage of redundant integrals.[77]

A linear scaling CCSD method was also developed by Ayala and Scuseria.[21] This method was quite similar to their linear scaling MP2 method in that the general strategy was to rewrite the relevant equations in an AO basis and then make heavy use of screening techniques. However, their CCSD method differed from their MP2 method in many respects. The most straightforward way to understand their CCSD method is to consider the transformation to the AO basis of the expression for the contribution of the T_{ab}^{ij} to the CCSD energy

$$E \leftarrow \sum_{i,j,a,b} T_{ab}^{ij}(2(ia|jb) - (ib|ja)) \tag{104}$$

$$= \sum_{i,j,a,b} T_{ab}^{ij} \omega_{ab}^{ij} \tag{105}$$

It should be noted that the correlation energy expression of any method can be cast in the form of equation (105) with a proper choice of the T_{ab}^{ij} amplitudes. Thus, the discussion that follows is relevant for MP2 as well. As we will see, this point is of central importance because a number of interesting properties held by the AO-based MP2 expansion coefficients can be uncovered through a comparison with the Laplace MP2 expressions.[21]

By extracting the MO expansion coefficients from the integrals in equation (105) and moving them onto the T_{ab}^{ij} amplitudes, equation (105) can be rewritten as

$$E \leftarrow \sum_{\mu,\nu,\lambda,\sigma} T_{\mu\nu}^{\lambda\sigma} \omega_{\lambda\sigma}^{\mu\nu} \tag{106}$$

where

$$T_{\mu\nu}^{\lambda\sigma} = \sum_{i,j,a,b} L_{\mu i} L_{\nu j} L_{\lambda a} L_{\sigma b} T_{ab}^{ij} \tag{107}$$

At this point, it is instructive to compare equation (106) with equations (70) and (72). Equating the coefficients of the $\omega_{\lambda\sigma}^{\mu\nu}$ in both sets of equations, one arrives at the following expression for the MP2 $T_{\mu\nu}^{\lambda\sigma}$

$$T_{\mu\nu}^{\lambda\sigma} = \sum_\alpha w_\alpha (\bar{\lambda}\underline{\mu}|\bar{\sigma}\underline{\nu}) \tag{108}$$

This expression states that the MP2 wave function amplitudes can be written as linear combinations of the $(\bar{\lambda}\underline{\mu}|\bar{\sigma}\underline{\nu})$, and therefore have the same properties as the $(\bar{\lambda}\underline{\mu}|\bar{\sigma}\underline{\nu})$. These properties include $1/R^3$ decay with $\lambda\mu$–$\sigma\nu$ separation and compliance with a Schwarz inequality

$$|T_{\mu\nu}^{\lambda\sigma}| \leq |T_{\mu\mu}^{\lambda\lambda}||T_{\nu\nu}^{\sigma\sigma}| \tag{109}$$

Strictly speaking, these qualities apply only to the MP2 amplitudes. However, it turns out that the extent to which the CCSD amplitudes do not also share these properties is negligible. One, therefore, may use these properties as a foundation for a reduced scaling CCSD method in which insignificant amplitudes are neglected.

As has been shown by Ayala and Scuseria, the use of the screening criteria described above leads to linear scaling in the number of significant $T_{\mu\nu}^{\lambda\sigma}$ amplitudes.[68, 78] However, to achieve linear scaling for the overall CCSD calculation, each amplitude must be computed in an amount of time that is independent of the size of the system. For this matter, it is advantageous to work in an AO representation different from the one described above. The $T_{\mu\nu}^{\lambda\sigma}$ are further transformed as

$$\theta_{\mu\nu}^{\lambda\sigma} = \sum_{\alpha,\beta,\gamma,\delta} S_{\mu\alpha} S_{\nu\beta} S_{\lambda\gamma} S_{\sigma\delta} T_{\mu\nu}^{\lambda\sigma} \tag{110}$$

and the $(\bar{\lambda}\underline{\mu}|\bar{\sigma}\underline{\nu})$ integrals for $t_\alpha = 0$ (t_α is defined in Section 3.3) are used in place of the $(\lambda\mu|\sigma\nu)$. Since the overlap matrices decay exponentially, this transformation of the $T_{\mu\nu}^{\lambda\sigma}$ does not alter their decay properties. Furthermore, as has been shown by Ayala and Scuseria, the $(\bar{\nu}\underline{\mu}|\bar{\chi}\lambda)$ integrals decay as $R_{\lambda\mu}^{-3}$,[68] which is a significant improvement over the $R_{\lambda\mu}^{-1}$ decay of the $(\nu\mu|\chi\lambda)$. Thus, when working in this basis, the long-range contributions to the residuals and energies decay as R^{-6}. This allows one to screen very efficiently the contributions to the residuals and evaluate each of them in an amount of time that is independent of the size of the system.

Unlike in their MP2 work,[68] Ayala and Scuseria did not aim to reproduce CCSD energies to within a well-defined tolerance. Instead, they aimed to consistently reproduce a fixed fraction of the CCSD correlation energy. Through a proper choice of neglect threshold, Ayala and Scuseria were able to capture greater than 99.9% of the CCSD correlation energy and still maintain linear scaling in test calculations on a series of molecules ranging in size from water to dodecane. Furthermore, the recovery of such a high fraction of the correlation energy allowed them to generate an extremely smooth PES for the binding of two water molecules.[21]

4.3. Reduced Scaling CCSD(T)
A number of corrections to the CCSD method have been proposed. Almost all of these corrections aim to capture the most important contributions of the triple excitations for reasons outlined above. Here, we shall

focus on the CCSD(T)[75] correction to CCSD since it is the most common CCSD correction. For completeness, we note in passing that Schütz has also implemented local versions[79] of the iterative CCSDT-1a and CCSDT-1b[80] methods for treating the contributions of the connected triples.

In a canonical-orbital basis, the CCSD(T) correction may be computed in a noniterative manner. The expression for the CCSD(T) energy in the basis of canonical orbitals may be neatly summarized as

$$E^{(T)} = \frac{1}{3} \sum_{i,j,k,a,b,c} (W_{ijk}^{abc} - V_{ijk}^{abc}) \frac{X_{ijk}^{abc}}{D_{ijk}^{abc}} \tag{111}$$

where

$$W_{ijk}^{abc} = \mathcal{P}_{ijk}^{abc} \left[\sum_f (ia|jb) T_{cf}^{kj} - \sum_m (ia|jm) T_{bc}^{mk} \right] \tag{112}$$

$$V_{ijk}^{abc} = t_i^a (bj|ck) + t_j^b (ai|ck) + t_k^c (ai|bj) \tag{113}$$

$$D_{ijk}^{abc} = f_{aa} + f_{bb} + f_{cc} - f_{ii} - f_{jj} - f_{kk} \tag{114}$$

$$X_{ijk}^{abc} = 4W_{ijk}^{abc} + W_{kij}^{abc} + W_{jki}^{abc} - 2W_{kji}^{abc} - 2W_{ikj}^{abc} - 2W_{jik}^{abc} \tag{115}$$

$$\mathcal{P}_{ijk}^{abc} x_{ijk}^{abc} = x_{ijk}^{abc} + x_{jki}^{bca} + x_{kij}^{cab} + x_{ikj}^{acb} + x_{kji}^{cba} + x_{bac}^{jik} \tag{116}$$

In equation (116), x is some arbitrary six-index quantity and $\mathcal{P}$ is an operator that cyclically permutes the indices of x. Because the D_{ijk}^{abc} reside in the denominator of equation (111), the W_{ijk}^{abc} must be computed as full six-index quantities, leading to the characteristic n^7 scaling of the CCSD(T) method. However, just as in the MP2 method, the D_{ijk}^{abc} are vulnerable to a Laplace transform. This strategy has been implemented by Constans et al.,[81] and it allows one to carry out CCSD(T) calculations mainly in terms of four-index quantities. This leads to an n^6 scaling CCSD(T) method. Additional scaling reductions are possible by transforming the CCSD(T) expressions to a localized basis and then screening negligible contributions. To this end, Constans et al. used a basis of generalized CCSD natural orbitals. This led to an n^5 scaling CCSD(T) method and exhibited sub-millihartree precision. They also outlined the necessary equations that would be used in an AO-Laplace CCSD(T) method. However, no results were presented for this method.

Maslen et al. implemented an atoms-in-molecules CCSD(T) method[82] that was quite similar to the atoms-in-molecules MP2 method. The canonical CCSD(T) energy expression can be written in the matrix form

$$E^{(T)} = \frac{1}{3}(\mathbf{W} + \mathbf{V}^T)\mathbf{D}^{-1}\mathbf{X}. \tag{117}$$

Here, $\mathbf{W}$, $\mathbf{V}$, and $\mathbf{X}$ are vectors of length $n_i^3 n_v^3$ and $\mathbf{D}$ is a matrix of sidelength $n_i^3 n_v^3$ that is diagonal in the

canonical MO basis. In a localized basis, equation (117) can be approximated as

$$E^{(T)} \approx \frac{1}{3}(\mathbf{W} + \mathbf{V}^T)\mathbf{D}^{-1}\mathbf{P}^T\mathbf{X} \tag{118}$$

where $\mathbf{P}$ is a projector onto the space spanned by the localized basis. Formation of $\mathbf{P}$ entails the construction of the generalized inverse of the matrix $\mathbf{U}$ that transforms the triply excited n-particle states from the canonical MO basis to some localized basis:

$$|i'j'k'a'b'c'\rangle = \sum_{i,j,k,a,b,c} U_{ijkabc;i'j'k'a'b'c'} |ijkabc\rangle \tag{119}$$

where the localized functions are denoted with primes. To achieve reduced scaling, the space of triply excited configurations must be truncated. Similar to the inversion of the Δ matrix in MP2, facile inversion of $\mathbf{U}$ depends upon being able to write the truncated space of triply excited configurations as a direct product. When the space of triply excited states is chosen according to the usual atoms-in-molecules approach as described in Section 3.3, equation (118) yields only about 70% of the conventional nonlocal CCSD(T) energy. Compared with the recovery of greater than 90% of the correlation energy when the same local model is applied to MP2, this indicates that the contribution of the triples to the correlation energy is significantly more nonlocal in character. To take this into account, Maslen et al. computed CCSD(T) energies in which enlarged direct product spaces were used to define the space of triply excited states. Namely, ionic configurations in which one of the excitations was left entirely unrestricted were added to the space. Similar to equation (92), this can be written as

$$\{a'i'b'j'c'k'\} = \{a'\} \otimes \{i'\} \otimes \{b'j'\} \otimes \{c'k'\} \tag{120}$$

When such spaces were used, the fraction of the CCSD(T) energy recovered increased to 99%.

No CPU times or any other demonstrations of reduced scaling were given in the CCSD(T) work of Maslen et al. However, the method does truncate the configuration space and does not require the storage of any triples amplitudes, so the prospects for reduced scaling are hopeful.

A linear scaling integral direct local CCSD(T) method was also implemented by Schütz.[31, 76] This method makes use of all the techniques used in the local MP2 and CCSD methods of Schütz et al.[28, 30] However, there are important differences between the local CCSD(T) method of Schütz and other local correlation techniques that deserve additional commentary. Of course, the most obvious difference is that one now must deal with triplets of local orbitals. Certainly, one could determine the strong triples by referring to the list of strong pairs. For example, if ij, ik, and jk are strong pairs then it is natural to form the strong triple ijk. This approach is advantageous because it allows one to reuse integrals that were formed in the local CCSD calculation. Unfortunately, when this approach is taken, less than 70% of the CCSD(T) energy is recovered. This is again a result of the extended nonlocal character of the CCSD(T) energy that was also observed by Maslen et al. in their atoms-in-molecules approach.[82] In order to handle this, Schütz introduced an additional class of 'close pairs' that is slightly further separated than the strong pairs but less separated than most of the weak pairs. The close pairs were then used in combination with the strong pairs in the construction of the list of strong triples. It should be noted that the close pairs are never treated explicitly in the local CCSD calculation, so these CCSD amplitudes are not available for use in a subsequent CCSD(T) calculation. We discuss below how this is dealt with.

An additional difference between the local CCSD(T) method of Schütz and conventional CCSD(T) methods is the manner in which the second-order triples amplitudes are obtained. This issue intertwines heavily with the

issue of determining the strong triples, as we will see shortly. In conventional CCSD(T) methods that use a canonical MO basis, the equation for the required second-order triples coefficients can be trivially solved as

$$T_{abc}^{ijk} = \frac{-W_{abc}^{ijk}}{D_{abc}^{ijk}} \tag{121}$$

This straightforward relation relies upon the absence of off-diagonal terms in the Fock matrix. The representation of the Fock matrix in a typical localized basis, however, has a great many off-diagonal elements. The second-order triples amplitudes therefore must be obtained through the iterative solution of a complicated equation analogous to equation (121) that is appropriate for the localized nonorthogonal basis. This equation has been given in full by Schütz[76] and we shall not repeat it here, except to say that it requires the first-order doubles amplitudes. In CCSD(T), these amplitudes are generally approximated as converged CCSD amplitudes. However, as we have noted, local CCSD calculations never include the close pairs that are used to determine the strong triples. The missing amplitudes can be dealt with using a few different approaches. First, they can be neglected entirely in the determination of the triples amplitudes. This leads to recovery of 80 to 85% of the conventional CCSD(T) correlation energy depending on the cutoff radii used to determine strong and close pairs. Second, they can be approximated as local MP2 amplitudes. This increases the percentage of the CCSD(T) correlation energy recovered by about an additional 8%. Thus, the percentage of the correlation energy recovered when the local correlation approximation is used to treat the triples never quite reaches the level observed in methods that only require double excitations.

Despite this shortcoming, the local CCSD(T) method has significant merit, as the computational savings from the local CCSD(T) method are enormous. For example, a local CCSD(T) calculation that treated over 300 correlated electrons using 1018 basis functions required just less than 5 h on a modern computer. The corresponding conventional CCSD(T) calculation was estimated to require an astounding 56 years.

< Previous   Next >

Copyright© 1999-2005 *John Wiley & Sons, Inc.* All rights reserved.

Reduced Scaling Electron Correlation Methods

Emily A. Carter, University of California, Los Angeles, CA, USA
Derek Walter, University of California, Los Angeles, CA, USA

5. Conclusions

We have presented here a summary of reduced scaling methods for treating electron correlation. A few points are clear. First, the problem of accurately treating electron correlation in large organic molecules, which are described well by a single reference determinant, is largely solved. These problems can be handled through the use of well-tested local CCSD and local CCSD(T) methods that work quite well in this realm. At this point, it is likely that the methods we have described here will slowly start to become available for this purpose in commonly used quantum chemistry packages. As we have noted earlier, this has already happened for pseudospectral MP2 methods.[53, 65] In fact, applications of reduced scaling quantum chemistry methods to closed-shell molecules probably represent the next major avenue of research in the field of *ab initio* quantum chemistry applications.

Second, although it is possible to compute smooth PESs[21, 23] and even energy gradients[83, 84] using reduced scaling techniques, these methods still have not been perfected for the treatment of systems that are inherently multiconfigurational and or open shell in nature. In particular, to the authors' knowledge, reduced scaling techniques have not been applied to systems that include transition metal centers. A few PESs describing bond cleavage in organic molecules have been mapped out using small reference spaces, but basically nothing more complicated has been performed.

The problem of treating systems that are multiconfigurational in nature using reduced scaling techniques presents an interesting dilemma to quantum chemists. On the one hand, it is most straightforward to treat such systems using reduced scaling multireference configuration interaction (MRCI) techniques. Unfortunately, truncated CI does not offer size extensivity, which becomes increasingly important as the size of the system increases. Coupled cluster methods and perturbation theory, on the other hand, do offer rigorous size extensivity but are not easily generalized to multiconfigurational references, although there has been some work on local multireference MP2.[57, 59] At first glance, a way out of this dilemma is to simply use one of the many size-extensive modifications of the MRSDCI method that are available, such as average coupled pair functional (ACPF). Work in this area is in fact ongoing.[85] However, yet another problem arises. In local CC methods, long-range correlation between WP can be handled easily and efficiently through the use of local MP2, which is a rigorous subset of local CC. By contrast, it is not clear how long-range correlation should be handled in CI. This is especially true for CI methods based upon a multiconfigurational reference where the use of perturbation theory is not straightforward. It may be possible in the future to include a multireference perturbation theory (MR-PT) approximation to account for long-range correlation. At present, CI methods are slowly being subjugated by the increasing popularity of CC methods, especially for applications to closed-shell organic molecules. However, the CC methods run into problems treating radicals, bond breaking (due to the incipient open-shell character of the fragments), and molecules with near degeneracies (e.g., transition metals). Thus, the continued advancement of reduced scaling electron correlation methods for general application may hinge upon the development, practical implementation, and widespread acceptance of some sort of reduced scaling multireference CC technique. Alternatively, reduced scaling MRCI, with size

extensivity corrections and MR-PT treatment of long-range correlation, may be a satisfactory approach. Both avenues of reduced scaling correlation, local MRCI and local MRCC, are likely to continue to evolve in the coming years, affording vastly cheaper means of including electron correlation *ab initio*, and thereby facilitating the examination of larger and ever more complex molecules.

< Previous   Next >

About Wiley InterScience | *About Wiley* | *Privacy* | *Terms & Conditions*
Copyright© 1999-2005 *John Wiley & Sons, Inc.* All rights reserved.

Reduced Scaling Electron Correlation Methods

Emily A. Carter, University of California, Los Angeles, CA, USA
Derek Walter, University of California, Los Angeles, CA, USA

6. Acknowledgments

Reduced Scaling Electron Correlation Methods

Emily A. Carter, University of California, Los Angeles, CA, USA
Derek Walter, University of California, Los Angeles, CA, USA

References

1. R. Shepard, "The multiconfiguration self consistent field method", in *Ab Initio Methods in Quantum Chemistry – II*, ed. K. P. Lawley, John Wiley & Sons, 1987, p. 63.

2. F. Bobrowicz, and W. A. Goddard III, "The self-consistent field equations for generalized valence bond and open-shell hartree-fock wave functions", in *Methods of Electronic Structure Theory*, Vol. **3**, ed. H. F. Schaefer III, Plenum Press, New York, 1977, p. 79.

3. P. Å. A. Malmqvist, A. Rendell, and B. O. Roos, J. Phys. Chem., 1990, **94**, 5477.

4. B. O. Roos, Adv. Chem. Phys., 1987, **69**, 399. Links

5. R. J. Bartlett, and J. F. Stanton, "Applications of post-hartree-fock methods: a tutorial", in *Reviews in Computational Chemistry*, Vol. **V**, eds K. B. Lipkowitz, and D. B. Boyd, VCH Publishers, New York, 1994, p. 65.

6. I. Shavitt, Mol. Phys., 1998, **94**, 3. Links

7. C. Møller, and M. S. Plesset, Phys. Rev., 1934, **46**, 608.

8. R. J. Bartlett, Annu. Rev. Phys. Chem., 1981, **32**, 359. Links

9. T. J. Martinez, A. Mehta, and E. A. Carter, J. Chem. Phys., 1992, **97**, 1876. Links

10. E. R. Davidson, Comput. Phys. Commun., 1989, **53**, 49.

11. B. O. Roos, Chem. Phys. Lett., 1972, **15**, 153.

12. V. R. Saunders, and J. H. van Lenthe, Mol. Phys., 1983, **48**, 923. Links

13. W. Duch, and J. Karwowski, Comput. Phys. Rep., 1985, **2**, 93.

14. S. Saebø, and P. Pulay, Annu. Rev. Phys. Chem., 1993, **44**, 213.

15. C. Edmiston, and K. Ruedenberg, Rev. Mod. Phys., 1963, **35**, 457. Links

16. S. F. Boys, *Quantum Theory of Atoms, Molecules, and the Solid State*, Academic Press, New York, 1966, p. 263.

17. J. Pipek, and G. Mezey, J. Chem. Phys., 1989, **90**, 4916. Links

18. C. Edmiston, and K. Ruedenberg, J. Chem. Phys., 1965, **43**, S97. Links

19. W. von Niessen, J. Chem. Phys., 1972, **56**, 4290. Links

20. P. E. Maslen, and M. Head-Gordon, Chem. Phys. Lett., 1998, **283**, 102. Links

21. G. E. Scuseria, and P. Y. Ayala, J. Chem. Phys., 1999, **111**, 8330. Links

22. D. Walter, and E. A. Carter, Chem. Phys. Lett., 2001, **346**, 177. Links

23. D. Walter, A. Venkatnathan, and E. A. Carter, J. Chem. Phys., 2003, **118**, 8127. Links

24. P. Pulay, and S. Saebø, Theor. Chim. Acta, 1986, **69**, 357.

25. S. Saebø, and P. Pulay, J. Chem. Phys., 1987, **86**, 914.

26. S. Saebø, and P. Pulay, J. Chem. Phys., 1988, **88**, 1884.

27. C. Hampel, and H. -J. Werner, J. Chem. Phys., 1996, **104**, 6286. Links

28. M. Schütz, G. Hetzer, and H. -J. Werner, J. Chem. Phys., 1999, **111**, 5691. Links

29. J. W. Boughton, and P. Pulay, J. Comput. Chem., 1993, **14**, 736. Links

30. M. Schütz, and H. -J. Werner, J. Chem. Phys., 2001, **114**, 661. Links

31. M. Schütz, and H. -J. Werner, Chem. Phys. Lett., 2000, **318**, 370. Links

32. E. R. Davidson, J. Comput. Phys., 1975, **17**, 87.

33. G. Hetzer, P. Pulay, and H. -J. Werner, Chem. Phys. Lett., 1998, **290**, 143. Links

34. S. A. Orszag, Stud. Appl. Math., 1972, **51**, 253.

35. R. A. Friesner, Annu. Rev. Phys. Chem., 1991, **42**, 341. Links

36. T. J. Martinez, and E. A. Carter, "Pseudospectral methods applied to the electron correlation problem", in *Modern Electronic Structure Theory – II*, ed. D. R. Yarkony, World Scientific, Singapore, 1995, p. 1132.

37. R. A. Friesner, J. Chem. Phys., 1986, **85**, 1462. Links

38. C. C. Tazartes, C. R. Anderson, and E. A. Carter, J. Comput. Chem., 1998, **19**, 1300. Links

39. P. Pulay, Chem. Phys. Lett., 1983, **100**, 151. Links

40. S. Saebø, and P. Pulay, Chem. Phys. Lett., 1985, **113**, 13.

41. T. J. Martinez, and E. A. Carter, J. Chem. Phys., 1995, **102**, 7564. Links

42. G. Reynolds, T. J. Martinez, and E. A. Carter, J. Chem. Phys., 1996, **105**, 6455. Links

43. G. Reynolds, and E. A. Carter, Chem. Phys. Lett., 1997, **265**, 660. Links

44. D. Walter, A. B. Szilva, K. Niedfeldt, and E. A. Carter, J. Chem. Phys., 2002, **117**, 1982. Links

45. A. Szabo, and N. S. Ostlund, *Modern Quantum Chemistry*, New York, McGraw-Hill, 1989.

46. R. McWeeny, *Methods of Molecular Quantum Mechanics*, London, Academic Press, 1992.

47. P. Pulay, S. Saebø, and W. Meyer, J. Chem. Phys., 1984, **81**, 1901. Links

48. J. Almlöf, J. Chem. Phys., 1991, **96**, 489.

49. M. S. Lee, P. E. Maslen, and M. Head-Gordon, J. Chem. Phys., 2000, **112**, 3592. Links

50. S. Saebø, and P. Pulay, J. Chem. Phys., 2001, **115**, 3975. Links

51. G. Hetzer, M. Schütz, H. Stoll, and H. J. Werner, J. Chem. Phys., 2000, **113**, 9443. Links

52. T. J. Martinez, and E. A. Carter, J. Chem. Phys., 1994, **100**, 3631. Links

53. R. B. Murphy, M. D. Beachy, R. A. Friesner, and M. N. Ringnalda, J. Chem. Phys., 1995, **103**, 1481. Links

54. K. Wolinski, and P. Pulay, J. Chem. Phys., 1989, **7**, 3647.

55. K. Wolinski, H. L. Sellers, and P. Pulay, Chem. Phys. Lett., 1987, **140**, 225. Links

56. K. Andersson, P. A. Malmqvist, and B. O. Roos, J. Chem. Phys., 1992, **96**, 1218. Links

57. R. B. Murphy, W. T. Pollard, and R. A. Friesner, J. Chem. Phys., 1997, **106**, 5073. Links

58. R. B. Murphy, R. A. Friesner, and W. A. Goddard III, J. Chem. Phys., 1994, **101**, 2986. Links

59. B. D. Dunietz, and R. A. Friesner, J. Chem. Phys., 2001, **115**, 11052. Links

60. M. Schütz, R. Lindh, and H. -J. Werner, Mol. Phys., 1999, **96**, 719. Links

61. P. R. Taylor, Int. J. Quantum Chem., 1987, **31**, 521. Links

62. S. Saebø, and J. Almlöf, Chem. Phys. Lett., 1989, **154**, 83.

63. G. Rauhut, P. Pulay, and H. -J. Werner, J. Comput. Chem., 1998, **1998**, 1241.

64. B. H. Greeley, V. Russo, D. T. Mainz, J. -M. Langlois, W. A. Goddard III, R. E. Donnelly Jr., and M. N. Ringnalda, J. Chem. Phys., 1994, **101**, 4028. Links

65. *Jaguar 4.1*, Schrödinger, Inc., Portland, Oregon, 2000.

66. M. Häser, and J. Almlöf, J. Chem. Phys., 1992, **96**, 489.

67. M. Häser, Theor. Chim. Acta, 1993, **87**, 147.

68. P. Y. Ayala, and G. E. Scuseria, J. Chem. Phys., 1999, **110**, 3660. Links

69. P. E. Maslen, and M. Head-Gordon, J. Chem. Phys., 1998, **109**, 7093. Links

70. M. Head-Gordon, M. S. Lee, T. Van Voorhis, and S. R. Gwaltney, "Tensors in electronic structure theory: basic concepts and applications to electron correlation models", in *Modern Methods and Algorithms of Quantum Chemistry*, ed. J. Grotendorst, NIC Directors, Juelich, 2000, p. 593.

71. M. Head-Gordon, P. E. Maslen, and C. A. White, J. Chem. Phys., 1998, **108**, 616. Links

72. M. S. Lee, and M. Head-Gordon, Int. J. Quantum Chem., 2000, **76**, 169. Links

73. T. D. Crawford, and H. F. Schaefer III, "An introduction to coupled cluster theory for computational chemists", in *Reviews in Computational Chemistry*, eds K. B. Lipkowitz, and D. B. Boyd, John Wiley & Sons, 2000, p. 33.

74. R. Krishnan, M. J. Frisch, and J. A. Pople, J. Chem. Phys., 1980, **72**, 4244. Links

75. K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett., 1989, **157**, 479. Links

76. M. Schütz, J. Chem. Phys., 2000, **113**, 9986. Links

77. M. Schütz, Phys. Chem. Chem. Phys., 2002, **4**, 3941.

78. P. Y. Ayala, and G. E. Scuseria, J. Comput. Chem., 2000, **21**, 1524.

79. M. Schütz, J. Chem. Phys., 2002, **116**, 8772. Links

80. M. Urban, J. Noga, S. J. Cole, and R. J. Bartlett, J. Chem. Phys., 1985, **83**, 4041. Links

81. P. Constans, P. Y. Ayala, and G. E. Scuseria, J. Chem. Phys., 2000, **113**, 10451. Links

82. P. E. Maslen, M. S. Lee, and M. Head-Gordon, Chem. Phys. Lett., 2000, **319**, 205. Links

83. R. Guntram, A. E. Azhary, F. Eckert, U. Schumann, and H. -J. Werner, Spectrochim. Acta A, 1999, **55**, 647. Links

84. A. E. Azhary, G. Rauhut, P. Pulay, and H. -J. Werner, J. Chem. Phys., 1998, **108**, 5185. Links

85. A. Venkatnathan, A. Szilva, D. Walter, R. J. Gdanitz, and E. A. Carter, J. Chem Phys, 2003, in press. Links

‹ Previous   Next ›

About Wiley InterScience | *About Wiley* | *Privacy* | *Terms & Conditions*

Reduced Scaling Electron Correlation Methods

Emily A. Carter, University of California, Los Angeles, CA, USA
Derek Walter, University of California, Los Angeles, CA, USA

‹ **Previous**

Related Articles

Complete Active Space Self-consistent Field (CASSCF) Second-order Perturbation Theory (CASPT2)

Natural Bond Orbital Methods

Møller–Plesset Perturbation Theory

Spectroscopy: Computational Methods

Benchmark Studies on Small Molecules

NMR Chemical Shift Computation: Ab Initio

Relativistic Effective Core Potential Techniques for Molecules Containing Very Heavy Atoms

Wavefunctions and Chemical Bonding: Interpretation

R12-Dependent Wavefunctions

Gradient Theory

Localized MO SCF Methods

‹ **Previous**