Fast electronic structure methods for strongly correlated molecular systems

Martin Head-Gordon, Gregory J.O. Beran, Alex Sodt and Yousung Jung

Department of Chemistry, University of California, and, Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley CA 94720, USA.

E-mail: mhg@cchem.berkeley.edu

Abstract. A short review is given of newly developed fast electronic structure methods that are designed to treat molecular systems with strong electron correlations, such as diradicaloid molecules, for which standard electronic structure methods such as density functional theory are inadequate. These new local correlation methods are based on coupled cluster theory within a perfect pairing active space, containing either a linear or quadratic number of pair correlation amplitudes, to yield the perfect pairing (PP) and imperfect pairing (IP) models. This reduces the scaling of the coupled cluster iterations to no worse than cubic, relative to the sixth power dependence of the usual (untruncated) coupled cluster doubles model. A second order perturbation correction, PP(2), to treat the neglected (weaker) correlations is formulated for the PP model. To ensure minimal prefactors, in addition to favorable size-scaling, highly efficient implementations of PP, IP and PP(2) have been completed, using auxiliary basis expansions. This yields speedups of almost an order of magnitude over the best alternatives using 4-center 2-electron integrals. A short discussion of the scope of accessible chemical applications is given.

1. Introduction.

The motivation for the developments described in this chapter is the need within chemistry to have reliable electronic structure methods that can treat strongly correlated problems involving large molecules. These problems, exemplified by reactive molecules that have substantial diradicaloid character [1], are of interest in the assembly of solid-state organic materials with novel electrical properties, as well as in understanding transition structures, and reaction mechanism generally. For diradicaloid molecules, standard density functional theory (DFT) methods [2] fail. These molecules have small energy gaps between the nominally filled electronic levels and the nominally empty ones, which means that in their wave functions both are fractionally occupied. Standard DFT, based on a single set of fully occupied orbitals, cannot adequately treat this situation. Furthermore, while it is known in principal how to deal with this problem by taking linear combinations of all electron configurations that are significantly populated in the wave function, this cannot be done tractably at present. Such multi-configurational approaches [3, 4], exemplified by the Complete Active Space Self-Consistent Field (CASSCF) method, scale as severely as exponential with molecular size because they correspond to a valence space Schrodinger equation.

There has been a comparative paucity of methods that seek to systematically approximate the valence space Schrodinger equation. Most common is just to delete active orbitals to make the calculation tractable (up to 14 can be treated at present), but this comes at the absence of proper extensivity with the size of the molecule. Likewise, restricting allowed configurations within a linear...
model as in the Restricted Active Space (RAS) approach [5] also manifestly fails to be extensive with size. Our point of view is that we should instead seek rational approximations within the size-extensive framework of coupled cluster (CC) theory. Benchmark calculations established that even at the simplest double excitation level (CCD), good results could be obtained for bond-breaking problems in a valence active space [6]. Unfortunately even the simplest implementations of the resulting valence-optimized doubles (VOD) approach cannot scale better than 6th order with molecule size [7]. While infinitely better than exponential, applications to large systems are still out of reach without further rational approximations. The remainder of this chapter describes the formulation and efficient implementation of such approximations. The resulting pairing theories can feasibly be applied to systems containing hundreds of active electrons, and are beginning to yield new chemical insights on molecular and supramolecular systems that cannot otherwise be feasibly simulated.

2. Perfect Pairing (PP) and Imperfect Pairing (IP)

The perfect pairing active space is defined in terms of pairs of electrons. Each pair of electrons will be described by two orbitals, one (call it \( j \)) nominally occupied and the other (call it \( j^* \)) nominally correlating or empty. These orbitals are to be varied to minimize the energy associated with the chosen trial wavefunction. The exact wavefunction for a single pair of electrons in the space \((jj^*)\) can be written as a superposition of two configurations \( jj \) and \( jj^* \). Suppose the first has amplitude 1. The second configuration is a double orbital substitution from the first, whose relative amplitude could be described by a single unknown, \( t_j \). This description is sufficient to describe both the limit of a conventional electron pair bond (\( t_j \ll 1 \)), as well as partially or even fully broken bonds (\( t_j \to 1 \)), which is the fundamental reason for its appropriateness for strongly correlated systems. Let us now consider the form of coupled cluster wavefunctions for a collection of electron pairs \( \{j\} \) described within the corresponding PP active space. Such wavefunctions are of the form:

\[
\Psi = \text{exp}\left(\hat{T}\right)|0\rangle
\]

If we were to consider a coupled cluster doubles wavefunction, such as within the VOD method, all possible pairwise substitutions of the nominally occupied levels by the empty levels would be allowed with variable amplitudes, defining a correlation operator that using 2nd quantization is written as:

\[
\hat{T}_{\text{VOD}} = \sum_{jk \neq l^*} t_{jk}^* a_j^\dagger a_k^\dagger a_{l^*}^\dagger
\]

The number of unknown amplitudes increases with the 4th power of the number of electrons in the system. Furthermore, evaluation of the resulting amplitude equations scales with the 6th power of system size, as expected from viewing the amplitudes as an \( O^2 \) by \( O^2 \) matrix, and requiring multiplications of such matrices to yield a matrix of residuals.

The problem is to reduce this computational complexity while still keeping the essential physics that allows bonding as well as either partial or complete separation of electron pairs. The simplest way this could be attempted is to treat the electron pairs one-by-one instead of all at once. This defines the perfect pairing (PP) model of electron correlation [8], where the correlation operator now involves just one amplitude per electron pair:

\[
\hat{T}_{\text{PP}} = \sum_j t_j^* a_j^\dagger a_j^\dagger
\]

The \( O \) amplitudes that are kept are generally the most important ones, because they correspond to the correlations between electrons that are close to each other, as opposed to distant and thus more weakly interacting electron pairs. This is the first level of theory that we focus upon.

If one wishes to allow for some correlation between electron pairs (to treat situations such as multiple bonding for example, where several electron pairs are in the same spatial region), it will be necessary to expand the scope of the correlation operator beyond the PP model. The minimal step that makes no arbitrary assumptions is to permit one electron to be excited from each of two pairs simultaneously, in addition to the PP excitations defined above. This model no longer requires electrons to only correlate within pairs, and therefore we call it imperfect pairing (IP) [9]. The IP correlation operator can be written as:
With a quadratic number of amplitudes to solve for, one would be correct in assuming that the complexity of iteratively solving for these amplitudes would involve only 3rd order work. This is a dramatic reduction from the 6th order scaling of the full parent theory. Yet, as has been shown elsewhere [10], IP often recovers over 90% of the correlation energy of the full theory. Even PP is usually between 60% and 80%. Some details of our implementation are discussed in Sec. 4.

3. Second order perturbation correction to perfect pairing.

Treating the strong correlations at a sophisticated level such as the coupled cluster approach discussed above is certainly necessary to obtain even a qualitatively correct wavefunction. However, the relative energies associated with IP and especially PP will be far from quantitatively accurate, because of the neglect of many small correlations. In this section we discuss how to describe such correlations by second order perturbation theory. More detailed discussions of both the general formulation of similarity-transformed perturbation theory have already been published [11], and its specialization to the PP reference function will be published separately. Here we give just a short overview.

We divide the space of all possible electron configurations into two subspaces. The first, \( p \), is the space in which the reference coupled cluster equations have been solved. Specifically this is the space of determinants captured in the PP wave function. The second subspace, \( q \), is the remainder of the space. The perturbation theory is performed on the similarity-transformed Hamiltonian,

\[
\hat{H}_p = e^{-\hat{T}_p} \hat{H} e^{\hat{T}_p},
\]

because the PP energy is an eigenvalue of \( \hat{H} \) in the \( p \) space. \( \hat{H} \) must then be partitioned into a zero order part and a first order part so that the PP energy is an eigenvalue of the zero order part in the full space, rather than just in the \( p \) space. We believe the simplest useful choice is to take just the diagonal elements of the 1-body operator in the \( q \) space, thus defining:

\[
\hat{H}^{(0)} = \langle p | \hat{p} | p \rangle + \langle q | \hat{q} | q \rangle,
\]

with this partitioning, a perturbative expansion for the exact ground state energy and wave function is then written, taking the reference coupled cluster wave function to be zeroth order. The first non-zero energy correction arises at second order, and can be written as:

\[
E^{(2)} = \langle \langle 0 | \hat{H}^{(1)} | R^{(1)} \rangle + \langle \langle \Lambda^{(0)} | \hat{H}^{(1)} | R^{(1)} \rangle
\]

and hence we call this the PP(2) method. Space precludes giving full algebraic details here, but the important computational outcome is the following. Though this second-order correction scales with the seventh power of system size when applied to a full coupled cluster doubles wave function, PP(2) scales only with the fifth power, which is formally the same as simple second order perturbation theory starting from just one determinant (MP2 theory).

The PP(2) correction includes six classes of terms. These include MP2-like singles and doubles that are coupled to the reference wavefunction, \( 0 \), and singles, doubles, triples and quadruples that are coupled to the \( \Lambda \) vector (i.e. doubles). The MP2 doubles, which dominate the PP(2) correction both energetically and computationally, are like the canonical MP2 expression, except that the integrals and energy denominators are replaced by the corresponding dressed elements of \( \hat{H} \). These terms provide an MP2-like treatment of all pair correlations not described by PP, and there are \( OV \) such terms, where \( O \) and \( V \) are the numbers of occupied and empty orbitals. The Lambda doubles occur for a smaller set of pair correlations, involving at least one orbital in a PP correlation pair. The Lambda doubles tend to occur with the opposite sign of the MP2 doubles and are largest when the PP amplitudes are large. They help to compensate for overestimation of the correlation by the MP2 doubles in cases with very large PP amplitudes. Finally, the Lambda triple and quadruple excitations provide dynamical correlation for the excited configurations generated in the PP wave function.

Thus, at first glance, PP(2) behaves like a hybrid doubles level treatment of the electron correlation, where a linear number of the most important terms are treated at the coupled cluster level using PP, and the rest of the doubles are treated at the MP2 level. However, the MP2-like treatment couples to
the correlation of the potentially large PP amplitudes through the similarity transformation and the counter-balancing Lambda singles and doubles terms. It is of course the coupling of amplitudes that differentiates CCSD from MP2 and makes the former much more reliable. These singles and doubles terms on top of PP could then be said to provide a simple and inexpensive approximation to CCSD [12]. PP(2) goes further, though. The triples and quadruples terms then provide dynamical correlation for those doubly excited configurations contained in the PP wave function. This makes it more suitable to the treatment of systems with strong electron-electron correlations than MP2.

4. Efficient implementation based on the resolution of the identity (RI) approximation.

Using conventional implementations based on 4-center 2-electron matrix elements over Gaussian basis functions leads to computational effort for the electron correlation energy that even at the simplest MP2 level dominates the preceding calculation of the mean-field energy. A similar situation would apply to both the PP and IP methods, as well as the PP(2) correction. Fortunately, at the MP2 level it is known that this can be effectively overcome by using auxiliary basis expansions [13], which is also synonymously termed the resolution of the identity (RI) or density fitting (DF) approximation. Following Ahlrichs and co-workers, who developed standardized auxiliary basis sets for MP2 calculations [14], which we shall utilize in calculations using our new methods, we use the RI name. Using these standardized auxiliary basis sets, the RI approximation introduces almost negligible errors into computed correlation energies. For example, for the molecules of the G2 set using the VDZ basis, the RI approximation (using the algorithm described in Sec. 4.1) introduces an RMS error of 29 $\mu$H to the PP energy, or 8 $\mu$H per atom. The accuracy of the method is fairly uniform over the set of molecules, with the greatest error being 31 $\mu$H per atom.

In the RI approximation, two-center function pairs are replaced with a sum of one-center auxiliary basis functions. The one-center expansions are chosen such that they minimize the Coulomb self-interaction of the difference between the function pair and its expansion,

$$(\mu \nu - \mu \nu | \mu \nu - \mu \nu)$$

This leads to the following (optimal) expression for the 4-center 2-electron integrals in terms of a set of 3 center quantities, the B tensors, which are given both in terms of atomic orbitals (Greek letters) and auxiliary basis functions ($L,M,...$):

$$(\mu \nu | \lambda \sigma) \approx \sum_{LMN} B_{\mu \nu}^{L} B_{\lambda \sigma}^{N} = \sum_{LMN} (\mu \nu | L)(L | M)^{-1/2} (M | N)^{-1/2} (N | \lambda \sigma).$$

By expanding two-center functions in terms of one-center auxiliary functions, explicit four-center integrals are never needed. The number of two-center function pairs, which we call NFP, formally scales linearly with system size, due to the fact that Gaussian AOs have limited spatial extent, and therefore the product of two well-separated AOs will be negligible. Thus the number of required two-electron integrals will be reduced (but will still scale quadratically with system size).

4.1. RI algorithm for perfect and imperfect pairing

The PP and IP methods involve outer iterations that vary the nominally occupied orbitals, and their nominally empty correlating counterparts, and inner loops that solve for the correlating amplitudes given the present orbitals. The inner iterations are, from the standpoint of computational cost, essentially free. The rate-determining step in the outer loops is the construction of a set of Coulomb-like (J) and exchange-like matrices (K) for each electron pair [15], for example,

$$J_{\mu \nu}^{ii} = (ii | \mu \nu),$$

$$K_{\mu \nu}^{ii} = (i\mu | i\nu),$$

and likewise $J_{\mu \nu}^{ii}$, $J_{\mu \nu}^{ii}$, $K_{\mu \nu}^{ii}$, and $K_{\mu \nu}^{ii}$. The construction of these integrals requires MO transformations over four-center atomic orbitals (AOs), the bottleneck step of the method. It is this step which we have dramatically speeded up by developing an RI algorithm.

The following is an outline of the RI algorithm, with the scaling of the step indicated in parentheses.
1a. Form \((L \mid M)^{-1/2}\) (X)

2a. Form \((\mu \nu \mid M)\) (NFP X)

3a. Contract: \(B^L_{\mu \nu} = \sum_m (\mu \nu \mid M)(M \mid L)^{-1/2}\) (NFP X)

4a. Contract: \(B^L_{\mu \nu \mid i, j'} = \sum_{\mu'} B^L_{\mu \nu \mid i} C_{\mu' \mid i, j'}\) (NFP X O)

5a. Contract: \(B^L_{\mu \nu \mid i, i'} = \sum_{\mu'} B^L_{\mu \nu \mid i} C_{\mu' \mid i, i'}\) (X N O)

6a. Contract: \(K^{(a, a', \lambda \sigma)}_{\mu \nu \mid i, i', \mu' \nu'} = \sum_L B^L_{\mu \nu \mid i, i'} C_{\lambda \sigma \mid i, i'}\) (X N O)

7a. Contract: \(J^{(a, a', \lambda \sigma)}_{\mu \nu \mid i, i', \mu' \nu'} = \sum_L B^L_{\mu \nu \mid i, i'} B^L_{\mu' \nu' \mid i, i'}\) (NFP X O)

Steps marked with an asterisk need only be computed once per calculation; otherwise the step must be done each time MOs are updated. Comma-separated indices in square brackets are independent and their contributions to the total cost of the step are therefore mutually additive. \(O\) is the number of occupied levels, \(N\) is the number of atomic orbitals, and \(X\) is the auxiliary basis dimension.

Contrast this with the alternative algorithm [15] for creating the \(J\) and \(K\) matrices using four-center AO integrals without the RI approximation:

1b. Form \((\mu \nu \mid \lambda \sigma)\) (NFP3)

2b. Contract: \(K^{(a, a', \lambda \sigma)}_{\mu \nu \mid i, i', \mu' \nu'} = \sum_{\lambda \sigma} (\mu \lambda \nu \sigma) C_{\lambda \mu \nu \sigma} C_{\lambda \sigma \mu \nu} C_{\lambda \sigma \mu' \nu'}\) (NFP3 O)

3b. Contract: \(J^{(a, a', \lambda \sigma)}_{\mu \nu \mid i, i', \mu' \nu'} = \sum_{\lambda \sigma} (\mu \lambda \nu \sigma) C_{\lambda \mu \nu \sigma} C_{\lambda \sigma \mu \nu} C_{\lambda \sigma \mu' \nu'}\) (NFP3 O)

Our RI algorithm formally scales with the fourth power of system size (Step 6a), while without the RI approximation the algorithm scales only with the third power of system size (steps 2b and 3b). However, for systems of a size for which either algorithm might be feasible, few function pairs can be neglected. NFP is thus comparable to \(N^2\), yielding effectively fifth-order scaling without the RI approximation.

Each step of the RI algorithm may be broken up such that it can be computed using limited memory, that scales quadratically with system size. Mass storage requirements scale cubically with system size. We can contrast these modest requirements with the coupled cluster methods that do not employ local truncations, which have large (quartic) disk requirements and large (sixth order) computational requirements.

4.2. RI algorithm for the PP(2) correction.

The computation of PP(2) is dominated by the fifth order steps involved in the construction and transformation of the molecular orbital basis integrals, just like MP2. Unlike MP2, for which only \(\langle ij \mid ab \rangle\) integrals (\(i, j\) occupied, \(a, b\) virtual) are needed [14], PP(2) requires almost the full set of two-electron integrals. This, combined with the fact that the integrals must be dressed with a linear number of terms (to form \(\tilde{T}\)) and that there are many more types of terms than in standard MP2 theory, PP(2) will scale the same as MP2 but with a larger prefactor.

The computation of an electronic structure method that is dominated by the construction of the two-electron integrals is the ideal situation for employing the RI approximation. In addition to the computational savings with minimal loss of accuracy, the RI approximation facilitates the development of efficient algorithms. As above, we form the integrals as:

\[
\langle ij \mid ab \rangle = \langle ia \mid jb \rangle = \sum_L B_{ia}^L B_{jb}^L
\]

We are able to form and store the \(B\) matrices on disk, using only cubic storage (without even exploiting sparsity), instead of the quartic storage needed for the full set of two-electron integrals. The integrals can be formed as needed by reading in blocks of \(B\) in batches and performing matrix multiplies. The integrals are then used and discarded.
Consider the MP2 doubles terms, which are the rate-limiting step. Our algorithm is nearly identical to RI-MP2, except for the integral dressing steps. It proceeds as written below. Our PP(2) algorithm is therefore fifth order, dominated by the matrix multiplication to form the integrals from \( \mathbf{B} \). It also requires quartic I/O to read \( \mathbf{B} \) from disk. The loops are designed to utilize only quadratic memory, but they can be optionally batched to take advantage of all of the available memory.

Dress and store \( \mathbf{F}_{pq} \) in memory
loop over \( i \)
  Read \( \mathbf{B}(L,a,i) \) for all \( L \) and \( a \) from disk
loop over \( j \)
  Read \( \mathbf{B}(L,b,j) \) for all \( L \) and \( b \) from disk
  Form \( \langle a|ib \rangle = \sum_i B(L,a,i)B(L,b,j) \), antisymmetrize
  Dress \( \langle a|ib \rangle \) to form \( \mathcal{W}_{ij}^{ab} \) and \( \mathcal{W}_{ab}^i \)
loop over \( a, b \)
  \( E_{\text{MP2 doubles}}^{(2)} + = \mathcal{W}_{ij}^{ab}\mathcal{W}_{ab}^i / 4 \left( \mathcal{F}_{ii} + \mathcal{F}_{jj} - \mathcal{F}_{aa} - \mathcal{F}_{bb} \right) \)
end loop
end loop

In fact, the Lambda doubles terms require many of the same integrals, particularly the dressed \( \mathcal{W}_{ij}^{ab} \), where orbital \( I \) is active. Rather than rebuilding these integrals later or storing them on disk, we simply evaluate the Lambda doubles energy contribution simultaneously with the MP2 doubles. Though this complicates the coding slightly, it does not alter the algorithm fundamentally. The Lambda triples terms are also computationally intensive, since, if all valence electrons are active, their number is comparable to the \( O^{2V^2} \) terms for the MP2 doubles. This a dramatic reduction from the \( O^{3V^3} \) triples that arise in the (2) correction to a doubles vector that includes all amplitudes.

The most difficult aspect in implementing PP(2) efficiently is the proper treatment of the dressing terms to minimize I/O bottlenecks. In a loop structure like the one given above, forming \( \mathcal{W}_{ij}^{ab} \) for all \( a, b \) with a given \( i, j \) will include dressing terms for one particular \( a, b \) element where \( a \) and \( b \) correspond to some particular active orbitals from PP pairs. A naïve implementation of these terms requires two disk seeks to load two vectors to form a single integral, which is inefficient. However, there are usually no more than a quadratic number of such terms, so an efficient algorithm must build these in advance using fewer disk seeks and store them in memory or on disk until they are needed. Our implementation is partially optimized with regard to this issue, though further work could be done to bring down the computational time. Nevertheless, our current implementation demonstrates the affordability of PP(2) as compared to MP2 or other coupled cluster methods.

5. Tests of computational performance.
To assess the performance of our implementation of the PP, IP, and PP(2) methods, we compare against the full coupled cluster theory that these methods are approximating, as well as more affordable alternatives, in Figure 1 below. The left-hand curve is for the singles and doubles coupled cluster method (CCSD), which scales with the 6th power of molecular size. These 6th order steps must be executed on every iteration of a CCSD calculation. This limits it to small systems, even with the use of high-performance computers because doubling the system size requires 64 times more computing time. By contrast, PP(2) has only 4th order work per iteration, plus a single non-iterative 5th order step. It is the second curve from the left on Figure 1, and is clearly applicable to far larger systems. The relative sizes of the iterative 4th order steps and the non-iterative 5th order for PP(2) can be seen by comparing the PP(2) curve with the PP curve, which is the CPU time for just the iterative 4th order steps. For systems in the size regime of roughly 15-20 non-hydrogen atoms, these two steps are comparable in size. As system size increases further, the 5th order step will steadily become
dominant. We conclude that PP(2) calculations are viable using modern desktop computers for molecules at least 3 to 5 times larger than those that can be treated using CCSD. Given the 6th order cost scaling of the latter, this gap will not be readily bridged. It is also instructive to compare PP(2) and PP timings against RI-MP2 and Hartree-Fock (HF) timings, because the PP(2) is an analog of MP2 theory for a multiconfigurational reference (the PP reference). From Figure 1, the cost difference between RI-MP2 and RI-PP(2) is between a factor of 3 and 5 for the size regime shown. This means that if an MP2 calculation is feasible, then it will usually be the case that PP(2) is also feasible.

Figure 1. CPU times for linear alkanes, with the cc-pVDZ basis, as a function of size. All calculations were unrestricted, with no use of point group symmetry. They were done on a 2 GHz IBM 970fx cpu. Timings are reported for the full CCSD method, as well as for PP(2), and just the PP reference itself. The latter timings can be compared against RI-MP2, as well as just the Hartree-Fock reference.

It is also pertinent to assess the role that the resolution of the identity algorithms are playing in making the overhead of PP(2) so low relative to MP2. This can be seen from Table 1 below, where we compare the CPU time associated with RI-PP (and RI-IP) calculations against the CPU time required for a non-RI implementation. The algorithms are as described in Section 4.1. It is evident from Table 1 that the RI algorithms reduce CPU time by almost an order of magnitude.

Table 1. CPU times comparing resolution of the identity (RI) algorithms for the PP and IP methods, against a non-RI PP algorithm on linear alkanes. Calculations were performed using a 2.3 GHz IBM 970fx processor in an Apple XServe. The basis set used is cc-pVDZ with its RI-MP2 fitting basis.

| Chain length | RI-PP CPU (s) | RI-IP CPU (s) | PP CPU (s) |
|--------------|--------------|--------------|------------|
| 2            | 8.0          | 8.8          | 25.0       |
| 4            | 54.1         | 63.2         | 281.1      |
| 6            | 170.4        | 198.7        | 1120.1     |
| 8            | 321.1        | 402.6        | 3061.8     |

6. Emerging application areas.
Here we shall briefly summarize some of the chemical systems that can be treated using the methods we have described, based on our applications experience to date. The most obvious applications area is to apply these methods to large molecules which are believed to have strong correlations and thus significant diradicaloid character. From the experimental perspective, diradical character can only be assessed indirectly through spectroscopy (the lack of an ESR signal indicating a singlet state; a low
lying electronic transition in the visible or even infrared suggesting a small gap between highest filled molecular orbital (HOMO) and lowest empty orbital (LUMO), etc. Experimental chemists continue to seek to design synthetically accessible molecules that have high degrees of diradicaloid character, both because of fundamental interest in having available a partly or even fully broken bond to characterize by the methods used to probe stable molecules, as well as the exciting implications that such species can have for materials design, with respect to both optical, electrical and spin properties.

Calculation of a wavefunction that includes occupation numbers for the HOMO and the LUMO permits direct assessment of the extent of diradical character through the LUMO occupation number. After all, in the limit of a broken bond, the HOMO and the LUMO become equal in energy, and thus equally occupied (1 electron each), and this LUMO occupation number corresponds to 100% diradicaloid character. Thus the PP and IP methods can be immediately applied to characterize the nature of chemical bonding in novel materials, where chemical intuition fails, and the experimental evidence is limited. One example [16] is illustrated below in Figure 2. This inorganic species was first isolated experimentally in 2002, and was believed by its synthetic creators to be an example of an indefinitely stable singlet diradical. We characterized its LUMO occupation number by large-scale PP calculations, and found it to instead be 17% diradicaloid. We were able to understand the origin of this surprisingly low diradicaloid character based on through-space interactions with neighboring groups, which cause the originally small energy gap between HOMO and LUMO to increase by stabilization of the HOMO. This has implications for the rational design of materials with controlled gaps [17].

![Figure 2](image.png)

**Figure 2.** The structure of the 4-membered BPBP ring system. The left-hand panel shows the highest occupied orbital (HOMO), which has an occupation number of 1.83 electrons based on PP calculations correlating all 94 valence electrons. It is nominally bonding (through space) between the 2 boron atoms, but also exhibits significant (through bond) neighboring group interactions. The right-hand panel shows the corresponding lowest unoccupied orbital (LUMO), which is calculated to have occupation number 0.17, corresponding to 17% diradicaloid character.

A variety of other interesting chemical applications are possible to systems where chemical bonding may be diradicaloid in nature, and therefore beyond the reach of standard density functional theory methods. Other examples that we have recently studied include the dimerization of stable radicals, such as phenalenyl [18] and tetracyanoethylene anion [19], to yield stable crystal structures characterized by C-C distances that are too short for conventional Van der Waals complexes but too far too long for conventional covalent bonds. Our calculations were able to reproduce the experimental observables, such as bond length and bond strength, and then also provide an explanation for their origin. DFT found the complexes to be unbound. This application requires second order perturbation theory in addition to the PP or IP treatment of strong correlations in order to reproduce the bond strength, which emerges from the interplay of attractive covalent bonding and antibonding effects (in the strong correlations), attractive Van der Waal interactions (from the weak correlations), balanced against strong repulsions between filled orbitals as the monomers approach, with additional
important electrostatic effects depending on total charge. We believe that there are many other
interesting large systems with hitherto incompletely understood electronic structure where these new
methods may now permit calculations that can yield new insight.

7. Conclusions and outlook.
Our goal at the outset was to develop theory, algorithms, and finally efficient computer programs to
implement electronic structure methods that are capable of describing large molecules that contain
strong electronic correlations such as singlet diradicaloids. We have taken a 2-step approach to this
problem that first approximates the Schrodinger equation in a valence space of bonding and anti-
bonding orbitals that contains the strong correlations. The resulting PP and IP methods, when
implemented with resolution of the identity (RI) algorithms, are efficient enough to be applied to
systems with hundreds of active electrons. The RI algorithms scale only with the \(4^{th}\) power of
molecular size. By contrast, complete active space methods, which exactly solve the valence space
Schrodinger equation, can scarcely handle more than a dozen correlated electrons. The second step
then re-introduces all remaining (weaker) correlations via second order perturbation theory. We have
formulated and implemented such an approach from the PP reference, as PP(2) theory. PP(2) energies
involve computational work that scales with the \(5^{th}\) power of molecule size, which, with the aid of an
efficient RI-based implementation, is still favorable enough to be applied to molecules with over 50
non-hydrogen atoms, and over 200 correlated electrons. This is a substantial step forward in
computational capability relative to what is routinely possible with other accurate correlated methods.

A note of caution must be sounded in conclusion. While the approximations we have made to
develop and implement these methods are physically well-founded and reasonable, it is necessary to
perform significant additional testing to fully determine the extent of applicability of PP, IP and PP(2).
As discussed in Section 6, it appears that many interesting molecules with potential diradicaloid
character can be satisfactorily treated by these methods. However, there are also likely to be classes of
systems for which these new theories exhibit limitations and inadequacies. We have a hint that one
such class of systems will be those involving conjugated systems with multiple resonance structures.
Because the PP and IP models are fundamentally based on the electron pair as the basic quantity, they
will tend to break symmetry in systems with multiple resonance structures, as has been known from
earlier theories built on similar ideas. These limitations will need to be explored and documented, and,
hopefully, they will serve as the inspiration for further theoretical improvements in the future.

Acknowledgements
This research was supported by the Department of Energy, Office of Basic Energy Sciences, SciDAC
Computational Chemistry program (Grant No. DE-FG0201ER403301), with additional support from
Basic Energy Sciences. Computer resources from NERSC are gratefully acknowledged.

References
[1] Salem L and Rowland C 1972 Ang. Chem.-Intl. Ed. 11 92-&
[2] Kohn W, Becke A D and Parr R G 1996 J. Phys. Chem. 100 12974-12980
[3] Shepard, R. (1987) in Ab Initio Methods in Quantum Chemistry II, ed. Lawley, K. P.
(John Wiley and Sons, Ltd., Chichester), pp. 64-200.
[4] Schmidt M and Gordon M 1998 Ann. Rev. Phys. Chem. 49 233-266
[5] Malmqvist P, Rendell A and Roos B 1990 J. Phys. Chem. 94 5477-5482
[6] Van Voorhis T and Head-Gordon M 2000 J. Chem. Phys. 113 8873-8879
[7] Krylov A I, Sherrill C D, Byrd E F C and Head-Gordon M 1998 J. Chem. Phys. 109
10669-10678
[8] Cullen J 1996 Chem. Phys. 202 217-229
[9] Van Voorhis T and Head-Gordon M 2000 Chem. Phys. Lett. 317 575-580
[10] Head-Gordon M, Van Voorhis T, Beran G J O and Dunietz B 2003 Computational
Science - ICCS 2003, Pt IV. Proceedings 2660 96-102
[11] Gwaltney S R, Sherrill C D, Head-Gordon M and Krylov A I 2000 J. Chem. Phys. 113 3548-3560
[12] Beran G J O, Gwaltney S R and Head-Gordon M 2002 J. Chem. Phys. 117 3040-3048
[13] Dunlap B 2000 Phys. Chem. Chem. Phys. 2 2113-2116
[14] Weigend F, Haser M, Patzelt H and Ahlrichs R 1998 Chem. Phys. Lett. 294 143-152
[15] Van Voorhis T and Head-Gordon M 2002 J. Chem. Phys. 117 9190-9201
[16] Jung Y S and Head-Gordon M 2003 ChemPhysChem 4 522-525
[17] Jung Y S and Head-Gordon M 2003 J. Phys. Chem. A 107 7475-7481
[18] Small D, Zaitsev V, Jung Y, Rosokha S, Head-Gordon M and Kochi J 2004 J. Am. Chem. Soc. 126 13850-13858
[19] Jung Y and Head-Gordon M 2004 Phys. Chem. Chem. Phys. 6 2008-2011