

# Pseudospectral localized generalized Møller–Plesset methods with a generalized valence bond reference wave function: Theory and calculation of conformational energies

Robert B. Murphy and W. Thomas Pollard  
*Schrödinger Inc., 121 SW Morrison, Suite 1212, Portland, Oregon 97204*

Richard A. Friesner  
*Department of Chemistry, Columbia University, New York, New York 10027*

(Received 2 December 1996; accepted 17 December 1996)

We describe a new multireference perturbation algorithm for *ab initio* electronic structure calculations, based on a generalized valence bond (GVB) reference system, a local version of second-order Møller–Plesset perturbation theory (LMP2), and pseudospectral (PS) numerical methods. This PS-GVB-LMP2 algorithm is shown to have a computational scaling of approximately  $N^3$  with basis set size  $N$ , and is readily applicable to medium to large size molecules using workstations with relatively modest memory and disk storage. Furthermore, the PS-GVB-LMP2 method is applicable to an arbitrary molecule in an automated fashion (although specific protocols for resonance interactions must be incorporated) and hence constitutes a well-defined model chemistry, in contrast to some alternative multireference methodologies. A calculation on the alanine dipeptide using the cc-pVTZ(-f) basis set (338 basis functions total) is presented as an example. We then apply the method to the calculation of 36 conformational energy differences assembled by Halgren and co-workers [J. Comput. Chem. **16**, 1483 (1995)], where we obtain uniformly good agreement (better than 0.4 kcal/mole) between theory and experiment for all test cases but one, for which it appears as though the experimental measurement is less accurate than the theory. In contrast, quadratic configuration interaction QCISD(T) calculations are, surprisingly, shown to fail badly on one test case, methyl vinyl ether, for which the calculated energy difference is 2.5 kcal/mole and the experimental value is 1.15 kcal/mole. We hypothesize that single reference methods sometimes have difficulties describing multireference character due to low lying excited states in carbon–carbon pi bonds. © 1997 American Institute of Physics.  
[S0021-9606(97)00312-7]

## I. INTRODUCTION

The use of multiconfigurational self-consistent field (MC-SCF) methods in quantum chemistry dates back to the earliest numerical implementations of *ab initio* methods.<sup>1</sup> Initially, such methods were derived from physical intuition, in particular valence bond formalisms which suggested important configurations for a given molecule.<sup>2</sup> Unfortunately, the use of a small number of chemically important configurations is insufficient to obtain quantitative accuracy for molecular properties. More recently, the trend has been to include a comprehensive set of configurations with a restricted active space, for example in complete active space (CASSCF) methodologies.<sup>3</sup> In conjunction with a subsequent configuration interaction or perturbation step, such methods have the potential of achieving near-chemical accuracy for a wide range of systems. The difficulty here is the exponential scaling of computational effort with the size of the active space, restricting applications to small active spaces and hence, typically, to small molecules.

This analysis suggests that it is difficult, if not impossible, to define a systematic model chemistry (in the language of Pople and co-workers) based on MC-SCF methods which is both tractable and accurate for large systems. Indeed, strong claims to this effect have been made in a recent review article,<sup>4</sup> which essentially dismisses MCSCF-based

approaches as unworthy of further consideration in quantum chemical methods development. Instead, it is suggested that the two paths worth pursuing are based on coupled cluster approaches, e.g., QCISD(T) or CCSD(T), which are capable of high accuracy, albeit at a very substantial computational cost, and density functional methods, which achieve reasonably good accuracy for certain molecular properties while retaining a modest scaling of computational effort with system size.

For the past several years, we have been pursuing an MC-SCF approach which, in contradiction to the point of view described above, is at once highly systematic and has a reasonable scaling of computational effort with system size. The approach is based on two components: (1) a generalized valence bond (GVB) formulation of the underlying MC-SCF reference system;<sup>5</sup> (2) localized MP2 (LMP2) methods<sup>6,7</sup> for carrying out perturbative corrections. Pseudospectral numerical methods are used to make both of these technologies highly efficient.<sup>8–13</sup> Results for the individual pieces of the method have been described in previous publications.<sup>14,15</sup> Here, we present for the first time computations with the combined methodology, GVB-LMP2, with specific application to molecular conformational energies.

The first part of this paper discusses the GVB-LMP2 computational methodology, including scaling of computational effort as a function of system size, which is shown to

be in the  $N^2-N^3$  range where  $N$  is the basis set size. We also investigate ways to reduce the absolute CPU time, a process which is far from completed. We find that for large systems the method is considerably more efficient than any competitive wave function-based *ab initio* approach, even when every bond is correlated at the GVB level; if only a small part of a larger system is of interest, a subset of bonds can be correlated at the GVB level, and the computational advantage will grow substantially. Application to systems of 50–100 atoms with large basis sets is thus quite feasible, a range in which it is impossible to carry out CCSD(T) calculations due to the  $N^7$  scaling of this method.

In the second section, we examine the accuracy of calculating conformational energy differences, a subject which we have previously studied using the local MP2 method.<sup>15</sup> We find that GVB-LMP2 provides uniform near-chemical accuracy; indeed, it is possible that the theoretical results are more reliable than the experimental ones. In contrast, alternative approaches do not display a high level of reliability. In particular, systems containing carbon–carbon double bonds (which have low lying pi excitations) are shown to be extraordinarily difficult to treat with single-reference methods, including DFT and, surprisingly, coupled cluster methods, which display a 1.5 kcal/mole relative conformational energy error for one molecule, methyl vinyl ether, even with a large basis set. Thus, there appear to be a class of problems, even for conformational energy differences where no bonds are made or broken, where CCSD(T) or QCISD(T) methods are qualitatively inadequate due to failure to properly handle multireference character. This observation is in substantial contradiction to assertions in many recent publications<sup>16</sup> which have suggested (although without examining anything resembling a fully representative set of molecular structures) that CCSD(T) methods provide reliable results for all but the most challenging electronic structure problems. The difficulties that we have uncovered here are likely to be exacerbated for transition states, where the alternative configurations are even closer in energy to the ground state.

## II. METHODS AND COMPUTATIONAL EFFICIENCY

The method we outline here is a synthesis of our pseudospectral implementation<sup>15</sup> of Pulay and Saebo's LMP2 theory,<sup>6,7</sup> and the generalized Møller–Plesset (GMP) theory of Pulay and Roos<sup>17–21</sup> applied to a GVB reference.

The computational and physical advantages of the local approximation, wherein excitations from the reference are made from local orbitals to a limited set of local virtual orbitals, have been discussed at length in Refs. 6, 7, and 17, and will be further supported by this work. In particular we have shown<sup>15</sup> how the local approximation is exceptionally well adapted to an efficient pseudospectral formation of the two electron integrals which can overcome the integral transformation bottleneck of all analytic-integral based formulations. This efficiency is retained in the multiconfigurational extension discussed here and is the key to retaining the  $N^3$  scaling of this theory.

The GVB wave function was chosen as a reference be-

cause it is a compact representation of the dominant ‘‘static’’ correlation effects not captured by a single determinant reference. In particular, the GVB wave function is constructed to give a good zeroth-order description of dissociative processes and other bonding situations which, on a per bond pair basis, require at the minimum a two configuration description to be described correctly. Examples of these multiconfigurational cases range from ‘‘long bonds’’ at surfaces,<sup>22</sup> correlation in transition metals,<sup>23</sup> to pi bonds in the simple organic molecules discussed here. Given that we intend to improve the reference with perturbation theory, it is essential that the reference contain all near degeneracy effects if the perturbative corrections are to be reliable. Single-reference based perturbation methods are known to diverge dramatically<sup>24</sup> or to be totally unreliable even for transition metal atoms<sup>25</sup> when a multireference character dominates the problem. Lastly, the local-orbital and contracted-CI attributes of the GVB expansion are ideally suited to an efficient local multiconfigurational treatment of dynamic correlation when coupled with local correlation methods.

The GMP theory of Pulay<sup>18,19</sup> and Roos<sup>20,21</sup> has been shown over the past several years, most notably by the work of Roos *et al.*,<sup>21</sup> to be an efficient, size-consistent, and accurate method of describing dynamic correlation at the level of the more costly multireference CI method. Applications of this theory to GVB and GVB-RCI references by Murphy and Messmer<sup>25,26</sup> have indicated that GMP theory works equally well with the less extensive GVB-RCI references as with the more complete although computationally very demanding CASSCF reference expansions. The local formulation of GMP theory using a GVB reference is shown for the first time here to be formulated as a simple extension of the single-reference LMP2 theory.

### A. GVB reference

The reference GVB perfect-pairing (GVB-PP) wave function  $\Psi^{\text{GVB-PP}}$  is composed of  $N_{\text{GVB}}$  pairs of two singlet-coupled local orbitals  $\varphi_{i1}, \varphi_{i2}$  defining the  $i$ th GVB pair, a closed-shell core, and  $M$  high spin coupled open shell orbitals  $\phi_{i\sigma}$ . The perfect-pairing nomenclature refers to the use of only the single valence bond spin coupling among the GVB pairs,

$$\Psi^{\text{GVB-PP}} = \mathcal{A}[\{\text{core}\}\{\varphi_{11}\varphi_{12}\cdots\varphi_{N1}\varphi_{N2}\}\phi_{10}\cdots\phi_{M0}\Theta], \quad (1)$$

$$\Theta = \{(\alpha\beta - \beta\alpha)\cdots(\alpha\beta - \beta\alpha)\alpha\alpha\cdots\alpha\}$$

with  $\mathcal{A}$  the antisymmetrizing operator. The local orbital pairs  $\varphi_{i1}, \varphi_{i2}$  mutually overlap; however, a strong-orthogonality restriction is applied which forces orbitals of different pairs to be mutually orthogonal. As shown by Goddard *et al.*<sup>5</sup> this restriction and a computationally more useful form of  $\Psi^{\text{GVB-PP}}$  can be represented by expanding each local orbital pair  $\varphi_{i1}, \varphi_{i2}$  in terms of mutually orthogonal natural orbitals  $\phi_{i1}, \phi_{i2}$  via the relations

$$\begin{aligned} \varphi_{i1} &= (\sigma_{i1}^{1/2}\phi_{i1} + \sigma_{i2}^{1/2}\phi_{i2})/(\sigma_{i1} + \sigma_{i2})^{1/2}, \\ \varphi_{i2} &= (\sigma_{i1}^{1/2}\phi_{i1} - \sigma_{i2}^{1/2}\phi_{i2})/(\sigma_{i1} + \sigma_{i2})^{1/2}, \\ \sigma_{nm} &> 0; \sigma_{i1}^2 + \sigma_{i2}^2 = 1 \end{aligned} \quad (2)$$

with the pairwise normalization conditions of the last equation. The natural orbital pairs  $\phi_{i1}, \phi_{i2}$  are generally well localized between a pair of atoms with  $\phi_{i1}$  of the ‘‘bonding’’ type and  $\phi_{i2}$  having more nodal or ‘‘antibonding’’ character. This simple transformation puts  $\Psi^{\text{GVB-PP}}$  into the form

$$\Psi^{\text{GVB-PP}} = \mathcal{A}[\{\text{core}\} \{(\sigma_{11}\phi_{11}^2 - \sigma_{12}\phi_{12}^2) \cdots (\sigma_{N1}\phi_{N1}^2 - \sigma_{N2}\phi_{N2}^2)\} \Theta] \quad (3)$$

$$\Theta = \{\alpha\beta\alpha\beta \cdots \alpha\alpha\alpha\}.$$

Expansion of this expression shows that  $\Psi^{\text{GVB-PP}}$  describing  $N_{\text{GVB}}$  pairs is an expansion  $2^{N_{\text{GVB}}}$  closed-shell configuration-state functions with only  $2N_{\text{GVB}}$  variationally optimized CI coefficients  $\{\sigma\}$ .

The pseudospectral implementation of the optimization of the GVB wave function has been discussed at length in Ref. 12 where it is shown to retain a favorable  $N^3$  scaling applicable to large numbers of GVB pairs (of order 50) on workstations. Our contracted GVB-RCI wave function,<sup>14</sup> which allows for a more complete description of spin coupling effects, also retains this favorable scaling and applicability to large systems. In contrast, the state-of-the-art CASSCF codes<sup>21</sup> are limited to of order 12 electrons (6 GVB pairs).

## B. First-order wave function

The first-order wave function  $\Psi^{(1)}$  is defined by double excitations from occupied natural orbitals  $\{ij\}$  to either local AO virtual orbitals  $\{pq\}$  or other GVB natural orbitals in the case of semi-internal/internal excitations (singles do not contribute via the generalized Brillouin theorem). A particular excited term contributing to  $\Psi^{(1)}$  is thus generated as

$$\Psi_{ij}^{pq} = \hat{E}_{pi}\hat{E}_{qj}|\Psi_{\text{GVB}}\rangle \quad i \geq j, \quad (4)$$

where the excitation operators are defined by the generator state formalism implemented by Pulay *et al.*<sup>6</sup>

$$\hat{E}_{pi} = |p^\alpha\rangle\langle\phi_i^\alpha| + |p^\beta\rangle\langle\phi_i^\beta|. \quad (5)$$

Note that these excitation operators are also directly applicable as defined when the GVB reference contains high-spin open-shell occupations. To test this statement we wrote another version of the GVB-LMP2 code in which Serber<sup>27</sup> configurations were employed. The Serber and generator state function codes gave identical results. As pointed out in Refs. 6 and 7, the generator state formalism is preferred since it removes internal coupling coefficients and is more efficient than traditional CSF definitions. To illustrate the contracted nature of these expansions consider a term  $\Psi_{n1,m2}^{pq}$  generated from exciting the GVB natural orbitals  $\phi_{n1}$  and  $\phi_{m2}$  of Eq. (3), to AO virtuals  $pq$ ,

$$\Psi_{n1,m2}^{pq} = \mathcal{A}[\cdots(\sigma_{11}\phi_{11}^2 - \sigma_{12}\phi_{12}^2) \cdots \sigma_{n1}\phi_{n1} * p \cdots \sigma_{m2}\phi_{m2} * q \cdots (\ ) \cdots]. \quad (6)$$

This representative term is contracted in the sense that it is composed of order  $2^{N_{\text{GVB}}}$  determinants with weights determined by the GVB-CI coefficients  $\{\sigma\}$  of Eq. (3). This con-

traction is similar to that used in our definition of the GVB-RCI wave function.<sup>14</sup> The  $\Psi_{ij}^{(1)}$  retain the spin eigenvalue of the reference and do not have ‘‘spin contamination’’ of other spin eigenstates besides that of the reference. The configuration generated by the semi-internal excitation,  $\hat{E}_{pk}\hat{E}_{ki}$ , with  $k$  a GVB natural orbital and  $i$  any orbital outside of the  $k$ th pair can be represented in a form which retains orthogonality to the reference as

$$\hat{E}_{pk}\hat{E}_{ki}|\Psi_0\rangle = \mathcal{A}[\cdots(\sigma_{n1}\phi_{n1}^2 - \sigma_{n2}\phi_{n2}^2) \cdots (\ ) \cdots ip \cdots (\ ) \times (\sigma_{k1}\phi_{k2}^2 + \sigma_{k2}\phi_{k1}^2) \cdots]. \quad (7)$$

The excitation at pair  $k$  is identical to that used in our definition of the GVB-RCI wave function,<sup>14</sup> it is an excitation to the complementary root of the  $k$ th pair’s 2 by 2 CI space.

Introducing the square coefficient matrices  $C_{ij}$  to be determined within the GMP approximation below,  $\Psi^{(1)}$  is written as

$$\Psi^{(1)} = \sum_{i \geq j, pq} C_{ij}^{pq} |\Psi_{ij}^{pq}\rangle \quad (8)$$

with the  $pq$  sum containing both local AO and semi-internal indices. The local AO virtuals are defined exactly as in LMP2 theory, by orthogonalization of the contracted AO basis functions  $\{p_{ao}\}$  to the  $N_{\text{occ}}$  GVB, core, and open orbitals of Eq. (3)

$$|p_j\rangle = |p_{ao_j}\rangle - \sum_{i=1, N_{\text{occ}}} |\phi_i\rangle\langle\phi_i|p_{ao_j}\rangle. \quad (9)$$

We define local correlating domains for each  $ij$  pair analogously to the single determinant LMP2 method.<sup>7</sup> First the core and open-shell orbitals are independently localized via the Boys or Pipek schemes.<sup>28,29</sup> These unitary transformations leave the GVB energy unchanged. The GVB natural orbitals are by construction well localized and all local orbitals  $i$  are typically centered within a small number of atoms ( $a_1 \cdots a_n = \{a_i\}$ ,  $n=2$  for bond orbitals, and  $n=1$  for lone pairs. The correlating domain  $\{v_i\}$  for an occupied orbital  $\phi_i$  consists of the local AO virtuals  $\{p\}$  centered on the atoms  $\{a_i\}$  and the set of natural orbitals (for semi-internals) which share at least one atom with the set  $\{a_i\}$ . The occupied pairs  $ij$  defining the first order wave function are assigned a correlating space which is the union,  $\{v_i\} \cup \{v_j\}$  of the  $ij$  correlating spaces. Our code can also assign more extended correlating spaces  $\{v_i\}$  if needed for delocalized situations or situations involving resonance as discussed for one case below. However, the default mode is to base the  $\{v_i\}$  on bond pairs or lone pairs. This default assignment has been shown to be sufficient in our applications of LMP2 with the exception of transition states where an assignment of more delocalized  $\{v_i\}$  is necessary.

An important aspect of this theory is that not all electrons must be correlated at the GVB-LMP2 level. In the present implementation the valence closed shell space (defined by electrons not GVB paired in the reference) is treated at the LMP2 level of correlation as in conventional LMP2. A closed shell orbital will however use any semi-internal orbital which falls into its correlation domain. In addition we

can easily implement a version which leaves some region of the molecule at the Hartree–Fock or GVB level (i.e., a region without MP2 correlation). This ability to locally define a level of correlation is important for treating large systems efficiently.

### C. GVB-LMP2 second-order solution

The basic equation to be solved at second-order for the first-order wave function  $\Psi^{(1)}$  is<sup>6</sup>

$$\mathbf{T}^{(2)} = (\mathbf{H}_0 - E_0)\Psi^{(1)} + \mathbf{H}'\Psi_0 = 0, \quad (10)$$

where  $\mathbf{T}^{(2)}$  is the second-order residuum function,  $\mathbf{H}_0$  is the zeroth-order GMP Fock operator,  $\mathbf{H} = \mathbf{H}_0 + \mathbf{H}'$  is the full Hamiltonian, and  $\Psi_0 = \Psi_{\text{GVB}}$  here. The operator  $\mathbf{H}_0$  is defined as the generalized Fock operator of Pulay and Roos<sup>18,20</sup>

$$\mathbf{H}_0 = \mathbf{h} + \sum_{kl} \rho_{kl} \left( \mathbf{J}_{kl} - \frac{1}{2} \mathbf{K}_{kl} \right). \quad (11)$$

Here  $\mathbf{h}$  is the bare one electron operator,  $\rho_{kl}$  is the occupied–occupied orbital density matrix of the reference, and  $\mathbf{J}, \mathbf{K}$  are the Coulomb and exchange operators within this density. For the GVB wave function  $\rho_{kl}$  is diagonal with elements of  $2\sigma_i^2$  for the GVB natural orbitals and 2 or 1 for closed/open orbitals, respectively. The  $\mathbf{J}_{ii}$ ,  $\mathbf{K}_{ii}$  matrices in AO space for  $i$  a GVB natural orbital or open shell orbital and the closed shell Fock operator are formed with  $N^3$  scaling during the GVB optimization and hence do not have to be regenerated. The generalized Fock operator in AO space is subsequently transformed via a two index transform to the occupied/virtual orbital space with  $N^3$  scaling.

At this point it is possible to solve for  $\Psi^{(1)}$  with a procedure similar to the single-reference LMP2 formalism. The set of coupled equations for the  $\mathbf{C}_{ij}$  matrices of Eq. (8) can be derived by the projection of the contravariant function  $\tilde{\Psi}_{ij}$  on the residuum function of Eq. (10). As defined within the generator state formalism the contravariant function  $\tilde{\Psi}_{ij}$  is simply

$$\tilde{\Psi}_{ij}^{pq} = \mathcal{N} (2\Psi_{ij}^{pq} - \Psi_{ij}^{qp}), \quad (12)$$

where  $\mathcal{N}$  is a normalization factor. This formula applies to the high-spin open-shell limit as well. Note that this contravariant definition is used only in the formation of the residual.

A particular residual matrix  $\mathbf{T}_{ij}$  obtained by this projection satisfies the equation for the  $\mathbf{C}_{ij}$  matrices,

$$\mathbf{T}_{ij}^{(2)} = \langle \tilde{\Psi}_{ij} | \mathbf{H}_0 - E_0 | \Psi^{(1)} \rangle + \langle \tilde{\Psi}_{ij} | \mathbf{H}' | \Psi_0 \rangle = 0. \quad (13)$$

The difference with the single-reference limit arises from a slightly more complicated form of the matrix elements in the residual expression. The general expression for the contribution to the residual from the external excitations to AO virtuals (the AO–AO block) can be written in terms of overlap  $\mathbf{S}$ , Fock  $\mathbf{F} = \mathbf{H}_0$ , and exchange integral matrices  $\mathbf{K}_{ij}$  as

$$\begin{aligned} \mathbf{T}_{ij}^{\text{AO-AO}} = & \tilde{\mathbf{K}}_{ij} + \sum_{i'j'} A_{ij,i'j'} (\mathbf{F} \mathbf{C}_{i'j'} \mathbf{S} + \mathbf{S} \mathbf{C}_{i'j'} \mathbf{F}) \\ & + \mathbf{S} \left[ \sum_{i'j'} B_{ij,i'j'} \mathbf{C}_{i'j'} \right] \mathbf{S}, \\ \tilde{\mathbf{K}}_{ij} = & \sum_{i'j'} D_{iji'j'} \mathbf{K}_{i'j'}. \end{aligned} \quad (14)$$

In the notation above bold faced matrices in each operation have dimensions of the AO–AO part of  $\mathbf{C}_{i'j'}$  in that operation, while  $\mathbf{T}_{ij}$ ,  $\tilde{\mathbf{K}}_{ij}$ , and  $\mathbf{K}_{i'j'}$  are matrices with the dimension of the  $ij$  virtual space. An element  $K_{ij}^{pq}$  of the exchange matrix is the exchange integral  $(ip|jq)$ . The formulae for the coupling coefficients ( $A, B, D$ ) are supplied in the Appendix. In comparison to a single-reference LMP2 calculation, the GVB expression for the residual only generates more couplings than the single-reference case when  $ij$  or  $i'j'$  are natural orbitals of the same GVB pair. For example, in the GVB expression we have nonzero  $A_{iaia,ibib}$  parameters for  $ia, ib$  natural orbitals of pair  $i$ , while in the single-reference case  $A_{ij,i'j'} = \delta_{ii'} \delta_{jj'}$ . Thus, aside from the increase in occupied orbitals in going from the single reference to GVB, the increase in the couplings among the pairs in the residual expression is nominal. Furthermore, the expressions for the ( $A, B, D$ ) coupling coefficients are quite simple (independent of the AO indices) and can be quickly computed as needed.

The expression for the part of the residual involving one or two GVB natural orbital semi-internal indices is similar to that in Eq. (15) with additional projection factors. For example, consider the following part of the residual element  $T_{ij}^{kp}$  coming from the coupling to the term represented by  $C_{i'j'}^{q'q}$ . Here  $k$  is a GVB natural orbital from the  $k$ th GVB pair, and  $qq'$  are AO virtuals. Writing out the bra and ket we can see that if the  $k$ th pair is not in the  $iji'j'$  sites, then the occupation of  $k$  by the semi-internal excitation forces the occupation of the complementary natural orbital  $\bar{k}$  at the  $k$ th pair in both the bra and ket,

$$\begin{aligned} T_{ij}^{kp} = & + \dots \langle \dots ik \dots jp \dots \sigma_{\bar{k}k} \bar{k} | \mathbf{H}_0 \\ & - E_0 | \dots i' q' \dots j' q \dots \sigma_{\bar{k}k} \bar{k} \rangle C_{i'j'}^{q'q} + \dots \end{aligned} \quad (15)$$

The net effect of this complementary occupation is an overlap factor  $\sigma_{\bar{k}k}^2$  which weights this contribution to the residual in addition to the other coupling factors in the ( $A, B, D$ ) coefficients above. These additional semi-internal projection factors are dependent on the relationship between the  $k$ th pair site and the  $iji'j'$  sites. If in the above example  $j' = k_a$  is the first natural orbital of the  $k$ th pair and  $k = k_b$  is the second natural orbital, no additional projection factor is needed beyond that which is built into the ( $A, B, D$ ) coefficients. This site dependence of the projection factor complicates the evaluation of these GVB semi-internal terms. With an outer loop on the  $ij$  residual label, for each  $iji'j'$  term we scale the semi-internal block of  $\mathbf{C}_{i'j'}$  with the appropriate projection factors, and then use this scaled coefficient matrix in the matrix operations. In this manner the speed of the matrix multiply in the semi-internal space can be retained at

TABLE I. CPU times (minutes) for pseudospectral GVB [ $T$ -GVB $g^{**}$  for 6-31G $^{**}$  basis,  $T$ -GVB $cc$  for cc-pVTZ(-f) basis using the 6-31G $^{**}$  initial guess], exchange integral generation ( $T$ - $Kij$ ), and iterative solve ( $T$ -Solv) on a single IBM-SP2 390 thin mode.

Molecule	$N_{\text{bas}}$	$N_{\text{pair}}$	$T$ -GVB $g^{**}$	$T$ -GVB $cc$	$T$ - $Kij$	$T$ -Solv
Alanine Dipeptide	338	29	222	793	213	393
Methylcyclohexane	287	21	76	550	107	103
Cyclohexane	246	18	43	340	67	81
Methylvinylether	146	12	12	83	16	30

the cost of the pre-scaling. For the timings presented here we have not fully optimized this procedure and significant improvements are certainly possible.

The exchange terms  $K_{ij;i'j'}^{pq}$  in the residual are composed of linear combinations of exchange integrals of the form  $K_{ij}^{pq} = (ip|jq)$ , where the  $pq$  range over the virtual space of the  $ij$  pair. We use the pseudospectral formalism discussed in detail in Ref. 15 to form these exchange matrices with better than  $N^3$  scaling. Relative to LMP2, GVB-LMP2 simply requires extending the  $ij$  indices to include GVB natural orbitals or open-shell orbitals and the  $pq$  indices to include the semi-internal dimensions, and finally making linear combinations to form the contracted  $\tilde{\mathbf{K}}$  matrices. In the limit that all pairs are GVB correlated, the formation of the  $\mathbf{K}_{ij}$  matrices is only four times as expensive as the corresponding LMP2 calculation. This exceptional scaling of the exchange algorithm is the key to keeping the over all GVB-LMP2 scaling in the  $N^3$  regime. As shown below the formation of the exchange matrices is significantly faster than our current GVB reference optimization timing though both exhibit  $N^3$  scaling. Elements of the contracted exchange matrices involving semi-internals require extra consideration such as the insertion of projection factors discussed above.

Finally, note that linearly dependent terms and Pauli excluded terms can arise in coefficient matrices. For example,  $C_{k_a k_a}^{pq}$  for  $k_a$  the first natural orbital of GVB pair  $k$ , represents the same state as  $C_{k_b k_b}^{pq}$  with  $k_b$  the second natural orbital. In this case we zero out the  $C_{k_b k_b}^{pq}$  matrix. We also zero out elements corresponding to pure single excitations ( $j \rightarrow x$ ) such as  $C_{kj}^{kq}$ . A Pauli excluded term is of the form  $C_{ij}^{k_a k_b}$  where  $ij$  are not in pair  $k$ . Such elements among others are set to zero.

The iterative solution of Eqs. (15) for the  $\mathbf{C}_{ij}$  matrices is achieved by a modification of our iterative LMP2 solver by including the additional GVB-LMP2 coupling coefficients ( $A, B, D$ ) in the construction of the residual as well as separate routines for handling the semi-internal blocks of the residual. Once the residual matrices are formed we use the updating scheme outlined by Pulay<sup>7</sup> which transforms to and from a temporary orthogonal basis within each pair space, with the exception that we use a DIIS accelerator instead of a conjugate gradient accelerator. We have found that the use of the generator state formalism greatly reduces convergence problems. However, the GVB-LMP2 equations do exhibit slower convergence behavior than the single-reference LMP2 case as expected from the larger number of couplings

in GVB-LMP2. Ten to fifteen iterations are typically required to achieve convergence which is similar to the situation seen in the QCISD(T) calculations.

Once the  $\mathbf{C}_{ij}$  matrices have been found the second order energy  $E^{(2)}$  is given by a sum of traces

$$E^{(2)} = \sum_{i \geq j} \langle \mathcal{H}_{ij} \mathbf{C}_{ij} \rangle, \quad (16)$$

where  $\mathcal{H}_{ij}$  resembles  $\tilde{\mathbf{K}}_{ij}$  with an additional exchange contribution (see the Appendix),

$$\mathcal{H}_{ij}^{pq} = \sum_{i'j'} \mathcal{D}_{ij,i'j'} K_{i'j'}^{pq} + \mathcal{D}_{ij,i'j'}^\dagger K_{i'j'}^{qp}. \quad (17)$$

We are presently extending this formalism to GVB-RCI<sup>14</sup> reference wave functions in a similar although technically more complicated manner.

### III. TIMINGS

Preliminary timings for GVB-LMP2 in a cc-pVTZ (-f) basis are shown in Table I for up to 338 basis functions and 29 GVB pairs. In every case all GVB pairs are correlated. In these tests we follow the protocol of running the GVB to convergence with a 6-31G $^{**}$  basis and use these results as an initial guess for the cc-pVTZ(-f) GVB initial guess. This protocol is roughly twice as fast as using a cc-pVTZ(-f) initial guess. This test set displays an overall scaling of approximately  $N_{\text{bas}}^{2.9}$  in comparison to the HF-LMP2 scaling of  $N_{\text{bas}}^{2.6}$ . The GVB-LMP2 times reported here in which all GVB pairs are correlated are roughly eight times more expensive than the corresponding pseudospectral LMP2 times and with further optimization we expect them to be approximately five times as expensive in the all pair correlated limit. The GVB-LMP2 times are dominated by the self-consistent GVB calculation in the cc-pVTZ(-f) basis ( $T$ -GVB $cc$ ). The cc-pVTZ(-f) GVB times can be reduced by a factor of 2 with improvements to our SCF algorithms in progress. The generation of the local exchange integrals ( $T$ - $Kij$ ) is comparatively very efficient. This is to be contrasted to analytic methods for which this step is rate limiting. The time to solve for the pair CI coefficients ( $T$ -Solv) in roughly 10–15 iterations is about 20% of the calculation cost and roughly the same cost as generating the local exchange integrals. In comparison, the iterative solver time for HF-LMP2 is only 20% of  $T$ - $Kij$ . As discussed above the logic associated with the semi-internals has additional costs

which we have not made a serious attempt to optimize yet. We expect that with a proper optimization *T-Solv* should approach 50% of *T-Kij*.

Thus, with a relatively small effort we expect the timings presented here can be improved by approximately a factor of 2. More extensive efforts involving the development of multigrad techniques will result in even greater speedups and will certainly lower the overall scaling below the current value of 2.9. This type of performance is to be compared with methods of comparable accuracy such as quadratic configuration interaction [QCISD, QCISD(T)] which formally scale as  $N^6$  and  $N^7$ , respectively. Using the data from a recent publication<sup>30</sup> we obtain a scaling of  $N^{5.85}$  for QCISD in a cc-pVDZ basis and  $N^{5.58}$  for the corresponding local version of this theory using analytic integration. Furthermore, the disk usage of GVB-LMP2 is modest in comparison to that of QCISD methods which require *all* two-electron integrals on disk. The disk usage of GVB-LMP2 is given by the maximum used for the GVB or that used independently by the LMP2 section. The GVB storage requires storing roughly  $5+4N_{\text{pair}}$  square matrices of dimension  $N_{\text{bas}}$  on disk. The LMP2 portion requires storing five square matrices of dimension  $N_{\text{bas}}$  and  $5N_{\text{corr}}^2$  square matrices of dimension  $N_{\text{loc}}$  with  $N_{\text{corr}}$  the number of correlated orbitals and  $N_{\text{loc}}$  the size of the local correlating space, on average 60 for a cc-pVTZ (-f) basis. The dipeptide calculation above only requires 2 gigabytes of disk space while the corresponding QCISD(T) calculation would have inordinate disk and CPU requirements.

## IV. RESULTS

### A. GVB-LMP2 protocol

We shall consider a protocol here in which all bonds and lone pairs are correlated at the GVB-PP level. We have carried out some experimentation with reducing the number of GVB pairs, and it is likely that some approach along these lines can be worked out, particularly for larger molecules; however, initial efforts for the small molecule database we investigate here did not lead to a completely robust scheme. This direction will be pursued in later publications.

A key issue is the treatment of the correlation space for the local MP2 part of the calculation. The analogy of the standard LMP2 protocol that we have used above is to employ the usual union of orthogonalized atomic basis sets associated with the bonds being correlated, and add to this semi-internal excitations into GVB natural orbitals which meet the same localization criteria. The arguments for this

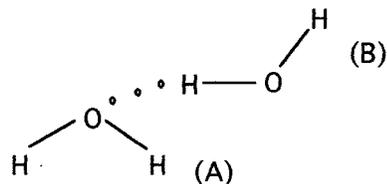


FIG. 1. Hydrogen bonding of two water molecules. GVB-LMP2 requires a delocalization of the semi-internal excitations from the OH bond of monomer *B* to include the semi-internal orbitals of monomer *A*.

approach are: (a) reduction in computation time; (b) elimination of basis set superposition error.

An alternative approach is to retain the localized atomic space but include all semi-internal GVB orbitals. Tests of this method indicated that for the vast majority of molecules in our test suite, this approach is either equivalent or inferior to the use of localized semi-internals. In most cases, the differences in conformational energy separations amounted to 0.1–0.2 kcal/mole, with the localized semi-internal approach most often closest to experiment. Therefore, we chose the localized protocol as the default methodology for assembling the bond and lone pair correlation spaces.

However, there are specific cases in which the localized protocol is found to be qualitatively inadequate. We emphasize that in each of these cases there is a specific chemical reason for a targeted delocalization of the virtual space. Examples along this line can also be found in local MP2, where, for example, delocalization to three centers is in many cases necessary to accurately compute transition state energies. The particular examples we focus on here are problems that arise specifically in GVB-LMP2 calculations. Such delocalizations are acceptable as part of a well defined model chemistry provided that the protocol is uniformly applied every time the relevant chemical structure is identified. The procedures defined below straightforwardly meet this criterion.

The first important case that we have encountered is for hydrogen bonded structures. Tests on the water dimer indicate that it is necessary to treat the hydrogen bond as a “bond” when constructing the delocalization protocol for excitations into the GVB semi-internal space. Consider the water dimer geometry shown in Fig. 1. The virtual space for excitation from the O–H bond of monomer *A* must contain the GVB natural orbitals of the hydrogen bonded O–H bond of monomer *B*. Results of the calculation with and without inclusion of these terms are shown in Table II, along with

TABLE II. Water dimer binding energies (kcal/mole). All results use the counterpoise correction of the corresponding reference wave function. GVB-LMP2 uses the local protocol and GVB-LMP2-d delocalizes the OH semi-internal excitations.

HF	LMP2	GVB	GVB-LMP2	GVB-LMP2-d	LMP4 <sup>a</sup>	LCEPA2 <sup>a</sup>	Expt.
4.17	5.03	3.15	3.62	4.67	4.61	4.68	5.4 (±.7)

<sup>a</sup>Local MP4 and local CEPA2 results from Ref. 17.

results from other calculations.<sup>17</sup> It can be seen that the exclusion of the semi-internal excitations results in a hydrogen bonding strength that is clearly too low. The justification of the extension of the semi-internal space is straightforward. The AO virtual space is orthogonalized to the GVB natural orbitals and hence this region of phase space is excluded from the AO virtual space (in contrast to LMP2 where exclusion of virtual space cannot occur since it is totally contained in the AO virtual space with no semi-internal space). Thus, if the excitation is physically important, it must be included explicitly. Excitation of an electron across the hydrogen bond is clearly a potentially important excitation and it is therefore unsurprising that elimination of this state from the virtual basis results in an underestimation of the hydrogen bond energy. Addition of a single orbital to the pair virtual space is highly unlikely to lead to large basis set superposition errors, which typically arise from enhanced dynamical correlation of the monomer which requires a large number of functions to describe. The semi-internal delocalized result is in very good agreement with the local MP4(SDQ) and local CEPA2 values of Ref. 17.

We have therefore incorporated a protocol in our GVB-LMP2 code in which delocalization into GVB natural orbitals in hydrogen bonding situations is implemented. As stated above, this protocol is entirely automatic and uniformly applied to any chemical structure. The only parametrization that is required is a definition of the cutoff distance for the hydrogen bond. A distance is used at which the effect is very small, thus avoiding discontinuities in the potential energy surface.

The second case is where there is a strong resonance within a particular functional group. This poses particular problems for GVB methods in which the orbitals localize into a single resonance structure. This localization can be corrected quite effectively by the perturbation theory component of the methodology; however, in order to do so it is necessary to build a correlation space delocalized over the resonant functional group.

As in the hydrogen bonded case, it is straightforward to define an automatic protocol once a relevant chemical group is identified. In this paper, we have encountered only one situation in which resonance presents significant accuracy problems; this is the carboxylic acid group in the glyoxylic acid molecule, which also makes an internal hydrogen bond. A detailed discussion of this case is presented below. In a subsequent publication, we will describe a survey of chemical functionalities in which resonance plays a role along with a protocol for each case where resonance presents difficulties. Only in this manner can a robust methodology of uniform accuracy be constructed from a local orbital approach.

All calculations were carried out with the Dunning correlation consistent cc-pVTZ(-f) basis set.<sup>31</sup> Our tests indicate that this basis set provides a good balance between accuracy and computational tractability for the level of precision we are seeking here ( $\approx 0.5$  kcal/mole maximum error). Smaller basis sets are incapable of providing this sort of reliability (at least in our hands), while our initial tests with larger basis sets, e.g., those including *f* functions or

diffuse functions, do not appear to yield significant improvement in the theory/experiment comparisons (this may, of course, be due to inaccuracies in the experimental data). Thus, our view is that extensive studies with larger basis sets will require better quality control with regard to the experimental data.

## B. Overview of results for conformational energies

As in our previous work, we study here the database of small molecule conformational energies assembled by Halgren, Kollman, and co-workers.<sup>32</sup> Table III presents results for conformational energy differences of all of these molecules at the HF, LMP2, and GVB-LMP2 levels. In addition, we have carried out a number of QCISD(T)/cc-pVTZ(-f) calculations for selected cases, the reasons for which will be described below (we did not perform such computation for all test molecules due to the formidable computational expense of such an undertaking). These latter calculations were performed with the GAUSSIAN 92<sup>33</sup> suite of programs. All GVB-LMP2 numbers reported in Table III utilized the local semi-internal protocol described above with the exception of glyoxylic acid, where the delocalized protocol was employed. (The localized value is provided in Table V but is not used in computation of average or rms errors.)

The most significant result of Table III is the remarkably good agreement between theory and experiment for the GVB-LMP2 calculations. Numerous improvements as compared to the LMP2 results can be seen, most strikingly methyl vinyl ether where the error is reduced by a factor of  $\approx 5$ . This case and others are discussed in detail below.

In Ref. 15, we attempted to ascertain which of the larger theory/experiment disagreements might be due to problems with the experimental value. We shall proceed along the same lines here, armed with substantial additional high level theoretical data from both GVB-LMP2 and QCISD(T) calculations. The first case we consider is formic acid. Here, all theoretical results beyond the HF level (including gradient corrected DFT calculations, not shown here) yield an energy difference of 4.5 kcal/mole, whereas the experimental result is 3.9 kcal/mole. In view of the exceptionally good agreement between the QCISD(T) and GVB-LMP2 results, and the lack of large perturbation from the MP2 results, it seems overwhelming likely that the experimental value is in error here (by  $\approx 15\%$ , a very reasonable error bar when the conformational energy difference is large). We therefore eliminate this case from calculation of the average and rms error. The second case is that of cyclohexanol, where computation of the energy difference of the two conformers as we interpreted them in Ref. 32 yields a result of  $\approx -0.1$  kcal/mole, in contrast to the experimental energy difference of 0.52 kcal/mole. However, examination of the original experimental literature indicates that Ref. 32 contained an error with regard to labeling of the conformational difference relevant to the experiment. The correct difference is that between the lowest energy axial conformer and the lowest energy equatorial conformer. We have carried out the appropriate calculations to determine this, and our GVB-LMP2 result, 0.6 kcal/mole,

TABLE III. Relative conformational energies (kcal/mole) from HF (HF), local MP2 (LMP2), GVB, GVB-LMP2, and QCISDT(T) calculations with a cc-pVTZ(-f) basis set.

Molecule	HF	LMP2	GVB	GVB-LMP2	Expt.
ethanol	-0.18	0.07	0.20	0.19	0.12
piperidine	0.78	0.56	0.89	0.56	0.40
isopropylamine	0.31	0.26	0.33	0.41	0.45
isopropanol	0.31	0.31	0.29	0.28	0.28
2,3-dimethylbutane	-0.07	-0.11	0.09	0.07	0.17
propylamine	0.54	0.18	0.46	0.38	0.42
cyclohexanol	0.85	0.54	0.74	0.61	0.52
methylcyclohexane	0.89	1.80	2.15	1.73	1.75
methoxycyclohexane	0.91	0.30	0.80	0.61	0.45
butanone	1.18	0.90	1.13	1.04	1.15
isoprene	2.80	2.68	1.34	2.49	2.65
1,3-butadiene	3.64	3.18	2.07	2.31	2.49
methyl vinyl ether	1.54	2.62	0.48	1.45	1.15
methyl vinyl ether					(QCISDT=2.53)
<i>N</i> -methylacetamide	2.49	1.89	2.53	2.14	2.30
formic acid	5.01	4.30	4.29	4.52	3.90
formic acid					(QCISDT=4.49)
<i>N</i> -methylformamide	1.00	0.90	1.13	1.35	1.45
<i>N</i> -methylformamide					(QCISDT=0.94)
ethyl formate	0.69	0.29	0.96	0.56	0.19
2-butene	1.74	1.21	1.63	1.15	1.00
acrolein	2.39	2.20	1.82	2.25	2.00
butane	1.09	0.73	0.97	0.86	0.75
methyl ethyl ether	1.69	1.22	1.70	1.53	1.50
methyl formate	5.47	5.35	4.66	5.09	4.75
ethyl ether	1.75	1.12	1.73	1.39	1.10
<i>N</i> -methyl piperidine	3.97	3.42	3.95	3.38	3.15
cyclohexane	7.00	6.04	6.25	5.85	5.50
glyoxylic acid	0.40	1.05	-0.50	0.93	1.20
cyclohexamine	1.29	0.80	1.19	0.78	1.10
propionaldehyde	0.85	0.67	0.57	0.80	0.95
dimethyl dioxane	1.30	0.85	1.00	1.02	0.90
methyl acetate	8.82	7.72	8.14	7.91	7.5-8.5
1-butene	0.74	0.76	0.80	0.22	0.53
fluoropropane	-0.02	0.39	0.03	0.14	0.35
chloropropane	0.43	-0.16	0.36	-0.21	-0.05
1,2-difluoroethane	0.07	0.56	0.57	0.58	0.80
1,2-dichloroethane	2.02	1.31	1.90	1.35	1.20
methoxytetrahydropan	0.49	1.25	0.68	1.22	1.05

now agrees with experiment to within 0.1 kcal/mole. This is the value that has been entered into Table III.

With this set of decisions in place, the average errors and rms errors of all computations are presented in Table IV. We have also included the DFT results from Ref. 32 for comparison. It can be seen that the GVB-LMP2 performance is superior to all other methods by a significant margin. Most

TABLE IV. Mean absolute deviation (MAD) and absolute RMS deviations (RMS) of the conformational energy differences (kcal/mole) of the previous table from experiment.

Error	HF	LMP2	NLSDA	GVB	GVB-LMP2
Full MAD	0.50	0.25	0.33	0.41	0.19
Full RMS	0.62	0.37	0.46	0.54	0.23
Filtered <sup>a</sup> MAD	0.49	0.25	0.32	0.41	0.18
Filtered <sup>a</sup> RMS	0.60	0.37	0.45	0.54	0.21

<sup>a</sup>The filtered average removes formic acid (see text) for which the experimental value is questioned.

importantly, the largest errors are of order 0.35 kcal, such as in the formates and in the glyoxylic acid case discussed below. These results suggest that, for the first time, a robust method for the determination of conformational energy differences (at least of organic molecules) has been developed. Note that this is not the case for QCISD(T) which apparently makes a 1.35 kcal/mole error for methyl vinyl ether. This is despite the fact that, if one simply looked at the MP2 and QCISD(T) results, one would believe that the answer was converged unproblematically. We should point out that, in addition to the agreement with experiment of the GVB-LMP2 results, there are two independent experiments<sup>34,35</sup> (performed, in fact, with different types of experimental apparatus) on methyl vinyl ether, both of which yield nearly identical answers for the conformational energy difference. Because this result is so surprising, further theoretical and experimental investigations are nevertheless warranted.

A closer examination of Table III reveals that the largest errors in the LMP2 results occur for systems with one or

more carbon–carbon double bond, e.g., methyl vinyl ether and 1,3 butadiene. Interestingly, these are also the systems for which gradient corrected DFT displays the poorest results,<sup>32</sup> and as discussed above, QCISD(T) cannot properly treat methyl vinyl ether. A simple interpretation of this observation is that a C=C moiety has a low lying pi excited state and hence possesses some important multireference character, in particular multireference effects appear to occur when the carbon–carbon double bond interacts with a second functional group, which in the case of methyl vinyl ether is the lone pairs on the ether oxygen. The conclusion of our study is that such character cannot be accurately represented by intrinsically single-reference theories, whether that theory is DFT, QCISD(T), or MP2. This problem of course occurs in a more extreme form in molecules such as ozone or, more generally, in studying transition states. Further examples of the break down of single-reference methods can be found in Refs. 22, 25, and 26. GVB-LMP2, as the only multireference method with a scaling in the  $N^2$ – $N^3$  range (the lower end requires treatment of only part of the molecule at the GVB level, which is plausible when treating larger systems), represents a systematic, automated approach to building in the requisite multireference character at a modest computational cost.

It should be noted that other LMP2 cases with somewhat smaller errors are also improved significantly by the GVB-LMP2 calculations. These include cyclohexane, *N*-methyl formamide [where the QCISD(T) result is not particularly accurate either], *N*-methyl acetamide, methyl ethyl ether, and 2,3-dimethyl butane. In no case does the GVB-LMP2 treatment lead to a qualitative increase in the theory/experiment deviations. However, the LMP2 results are qualitatively reasonable for all of these cases as opposed to methyl vinyl ether or (to a lesser extent) 1,3 butadiene.

### C. Resonance effects: Glyoxylic acid

Figures 2(a) and (b) present the two conformers of glyoxylic acid studied in this paper. The ground state conformation *B* has an internal hydrogen bond between the carboxylic acid hydrogen and the adjacent carbonyl; indeed, this is the only internal hydrogen bond in our test suite. This hydrogen bond, which is not present in conformer *A*, stabilizes a resonance structure of the acidic group shown in Fig. 2(c), in which the oxygen on the carbonyl becomes negatively charged and the the OH group positively charged. The positively charged OH group also strengthens the internal hydrogen bond. Experimental evidence for resonance in the acidic group can be inferred<sup>36</sup> from the CO and OH bond lengths being longer/shorter than in a typical ketone/alcohol, respectively.

From Table V it can be seen that the GVB reference predicts a 0.5 kcal/mole energy difference in the *wrong* direction as compared to experiment. The corresponding GVB-RCI<sup>14</sup> calculation which includes intra pair open-shell configurations and additional spin couplings does not improve the situation. We hypothesize that this occurs because the GVB description of the carbonyl of the acid OCOH

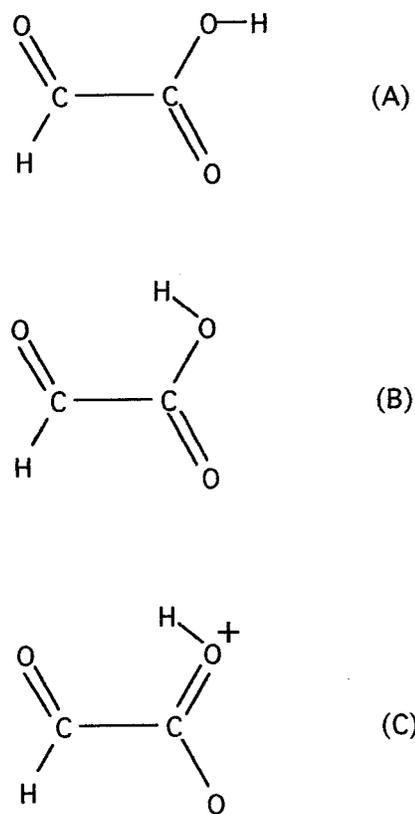


FIG. 2. Two conformers of glyoxylic acid studied; (b) is the hydrogen bonded ground state. (c) depicts a resonance structure of (b) which stabilizes the internal hydrogen bond.

group, which mandates a sigma-pi localized orbital description, fails to properly describe the resonance structure in Fig. 2(c). Ideas like this have been presented extensively in the previous literature, for example by Goddard and co-workers.<sup>37</sup>

We next examine what happens when the GVB-PP wave functions are correlated at the GVB-LMP2 level. First of all, within the local space protocol discussed above the GVB-LMP2 result of  $-0.27$  kcal, although an improvement over GVB, is not satisfactory. From a study of the water dimer at the GVB-LMP2 level, we discovered that it is essential to allow the semi-internal spaces of the H bonded units to be shared among these units rather than localized, as was discussed above. The delocalization of the CO and OH hydrogen bonded semi-internals among each other (i.e., for the CO pairs we include the semi-internals of the OH pairs and vice-versa) brings the energy difference to the correct sign and within 0.7 kcal of experiment. Finally, to account for the resonance in the OCOH group as depicted in Fig. 2(c), we delocalize the virtual space of pairs within the OCOH unit to include all virtuals (both AO and semi-internal) in the OCOH unit. This additional delocalization along with the hydrogen bond semi-internal delocalization brings the energy difference to 0.93 kcal/mole, which is quite acceptable. We made a similar LMP2 calculation in which the AO virtuals were delocalized among the OCOH unit and obtained

TABLE V. Energy differences (kcal/mole) between the two conformers of glyoxylic acid with HF, GVB, GVB-RCI, and GVB-LMP2 within the local protocol (GVB-LMP2-1), with H bonding accounted for (GVB-LMP2-*h*), and with both H bonding and resonance effects accounted for (GVB-LMP2-*hr*).

HF	LMP2	GVB	GVB-RCI	GVB-LMP2-1	GVB-LMP2- <i>h</i>	GVB-LMP2- <i>hr</i>	Expt.
0.4	1.05	-0.5	-0.15	-0.23	0.53	0.93	1.2 ( $\pm 5$ )

an energy difference of 1.05 kcal which is identical to the local protocol result for LMP2. This critical test indicates that the effect we see from delocalization in the GVB-LMP2 calculation is not a result of basis set superposition error.

## D. Conclusion

We have presented the first multireference methodology which is capable of achieving chemical accuracy for conformational energies, can be applied in an automated fashion, and has a scaling with system size suitable for large molecule applications. It is clear from our results that multireference character is essential in describing important classes of chemical problems, even for closed-shell ground state species of apparently innocuous character. Furthermore, the CPU times, disk space, and memory required for the calculations are quite reasonable and will allow application in production situations where high accuracy is desired.

If one is willing to accept errors of  $\approx 0.5$  kcal/mole for small molecules, a reasonable approach would be to utilize GVB-LMP2 when the molecule contains C=C double bonds, and LMP2 otherwise. The LMP2 calculations are considerably less expensive than the GVB-LMP2 calculations, and not all applications require the kind of precision that GVB-LMP2 can supply. On the other hand, as the cost/performance of computing continues to be drastically reduced, one will have the option of using an accurate method such as GVB-LMP2 for an increasing number of problems. The scaling with system size is sufficiently modest so that large molecules can be treated on workstations even at the present time.

An interesting question, not addressed in this paper, is whether it is possible to develop hierarchical methods, such as those presented by Petterson and co-workers,<sup>38</sup> in which a large basis set is used at the Hartree-Fock level and smaller basis sets are employed for the correlated calculations. Petterson and co-workers were examining bond energetics and, in cases where a small basis set was used for the correlation part of the calculation, were interested in an accuracy on the order of 2 kcal/mole for quite small molecules (the *G2* data base). We have carried out some preliminary tests which indicate that the present situation, involving higher precision and larger molecules, may be qualitatively different. Nevertheless, this is a research direction which certainly needs to be pursued, as the computational savings could in principle be substantial.

## ACKNOWLEDGMENTS

This work was supported in part by grants to R.A.F. from the NSF and NIH (GM-40526) and by an SBIR grant (GM-51740) to Schrödinger, Inc.

## APPENDIX

We tabulate here the residual coupling coefficients ( $A, B, D$ ) of Eq. (15) and the energy coefficients  $\mathcal{D}$ ,  $\mathcal{D}^\dagger$  of Eq. (17). In these equations GVB orbitals are denoted by  $(i_x, j_y)$  or for the natural orbital pairs  $(i_a, i_b)$ , open-shell orbitals are denoted by  $(i_o, j_o)$ , and closed-shell orbitals by  $(i, j)$ ,  $\sigma_{i_x}$  and  $\sigma_{j_y}$  denote GVB pair coefficients from pairs  $i, j$ , respectively [Eq. (4)],  $F_{ij}$  denotes a matrix element of the generalized Fock operator, and  $K_{ij}^{pq}$  is the two electron integral  $(ip|jq)$ . The sections are organized with respect to the pair indices occurring in  $T_{ij}$ , for example  $ij$  both closed,  $i$  closed  $j_y$  a GVB orbital, etc. Unless otherwise noted primed indices are assumed to be distinct from their corresponding unprimed indices. For the sake of brevity we have only tabulated the unique cases here and have not presented the details of the semi-internal projection factors which occur as outlined above or a few of the more complicated semi-internal exchange elements.

### 1. Closed-shell-closed-shell couplings

$$A_{ij, i'j'} = \delta_{ii'} \delta_{jj'} = D_{ij, i'j'}, \quad (\text{A1})$$

$$B_{ij, i'j'} = -\delta_{ii'} F_{jj'} - \delta_{jj'} F_{ii'}, \quad (\text{A2})$$

$$B_{ij, ij'_y} = -\sigma_{j_y}^2 F_{j, j'_y}, \quad (\text{A3})$$

$$B_{ij, i'_xj} = -\sigma_{i_x}^2 F_{i, i'_x}, \quad (\text{A4})$$

$$B_{ij, ij'_0} = -F_{j, j'_0}. \quad (\text{A5})$$

### 2. GVB-GVB couplings

In this section the intrapair element  $F_{ip}$  is defined by

$$F_{ip} = 2(\sigma_{i_a}^2 F_{i_a i_a} + \sigma_{i_b}^2 F_{i_b i_b}); \quad (\text{A6})$$

$$A_{i_x, i_x, i_y, i_y} = \sigma_{i_x} \sigma_{i_y} = D_{i_x, i_x, i_y, i_y}, \quad (\text{A7})$$

$$A_{i_x, j_y, i'_x j'_y} = \delta_{ii'} \sigma_{i_x}^2 \sigma_{j_y}^2 = D_{i_x, j_y, i'_x j'_y}, \quad (\text{A8})$$

$$B_{i_x, i_x, i_y, i_y} = -\sigma_{i_x} \sigma_{i_y} F_{ip}, \quad (\text{A9})$$

$$B_{i_x i'_x, i'_x i_y} = -\sigma_{i_x} \sigma_{i'_x}^2 \sigma_{i_y} F_{i_y i'_x}, \quad (\text{A10})$$

$$B_{i_x i_x, i_y j_y'} = -\sigma_{i_x}^2 \sigma_{j_y}^2 \sigma_{i_y} F_{i_y j_y'}, \quad (\text{A11})$$

$$B_{i_x i_x, i_y j_0'} = -\sigma_{i_x} \sigma_{i_y} F_{i_y j_0'}, \quad (\text{A12})$$

$$B_{i_x i_x, i' i_y} = -\sigma_{i_x} \sigma_{i_y} F_{i_y i'}, \quad (\text{A13})$$

$$B_{i_x j_y, i_x j_y} = \sigma_{i_x}^2 \sigma_{j_y}^2 (F_{i_x i_x} + F_{j_y j_y} - F_{i_p} - F_{j_p}), \quad (\text{A14})$$

$$B_{i_x j_a, i_x j_b} = \sigma_{i_x}^2 \sigma_{j_a} \sigma_{j_b} F_{j_a j_b}, \quad (\text{A15})$$

$$B_{i_x j_y, i_a j_a} = -\sigma_{i_x}^2 \sigma_{j_y} \sigma_{j_a} F_{i_x j_y}, \quad (\text{A16})$$

$$B_{i_x j_y, i' j_y} = -\sigma_{i_x}^2 \sigma_{j_y}^2 F_{i' i_x}, \quad (\text{A17})$$

$$B_{i_x j_y, i' j_y} = -\sigma_{i_x}^2 \sigma_{j_y}^2 \sigma_{i_x}^2 F_{i_x i_x}, \quad (\text{A18})$$

$$B_{i_x j_y, i_x j_0'} = -\sigma_{i_x}^2 \sigma_{j_y}^2 F_{j_y j_0'}. \quad (\text{A19})$$

### 3. GVB-closed-shell couplings

$$A_{ij_y, i' j_y'} = \delta_{ii'} \delta_{j_y j_y'} \sigma_{j_y}^2 = D_{ij_y, i' j_y'}, \quad (\text{A20})$$

$$B_{ij_y, i j_y} = \sigma_{j_y}^2 (F_{j_y j_y} - F_{j_p} - F_{ii}), \quad (\text{A21})$$

$$B_{ij_a, i j_b} = \sigma_{j_a} \sigma_{j_b} F_{j_a j_b}, \quad (\text{A22})$$

$$B_{ij_y, j_x i_x} = -\sigma_{j_y} \sigma_{j_x} F_{ij_y}, \quad (\text{A23})$$

$$B_{ij_y, ii} = -\sigma_{j_y}^2 F_{ij_y}, \quad (\text{A24})$$

$$B_{ij_y, i' j_y} = -\sigma_{j_y}^2 F_{i' i}, \quad (\text{A25})$$

$$B_{ij_y, i' j_y} = -\sigma_{j_y}^2 \sigma_{i_x}^2 F_{ii'}, \quad (\text{A26})$$

$$B_{ij_y, i j_y'} = -\sigma_{j_y}^2 F_{j_y j_y'}, \quad (\text{A27})$$

$$B_{ij_y, i j_y'} = -\sigma_{j_y}^2 \sigma_{j_y}^2 F_{j_y j_y'}, \quad (\text{A28})$$

$$B_{ij_y, i j_0'} = -\sigma_{j_y}^2 F_{j_y j_0'}, \quad (\text{A29})$$

$$B_{ij_y, j_y j_y'} = -\sigma_{j_y}^2 \sigma_{j_y}^2 F_{ij_y'}, \quad (\text{A30})$$

$$B_{ij_y, j_y j_0'} = -\sigma_{j_y}^2 F_{ij_0'}. \quad (\text{A31})$$

### 4. Closed-shell-open couplings

$$A_{ij_0 i' j_0'} = \delta_{ii'} \delta_{j_0 j_0'} = D_{ij_0 i' j_0'}, \quad (\text{A32})$$

$$B_{ij_0, i j_0} = -F_{ii} - F_{j_0 j_0}, \quad (\text{A33})$$

$$B_{ij_0, ii} = -F_{ij_0}, \quad (\text{A34})$$

$$B_{ij_0, i' j_0} = -F_{i' i}, \quad (\text{A35})$$

$$B_{ij_0, i' j_0} = -\sigma_{i_x}^2 F_{i' i}, \quad (\text{A36})$$

$$B_{ij_0, i' j_0} = -F_{i_0' j_0}, \quad (\text{A37})$$

$$B_{ij_0, i' i} = -F_{i' j_0}, \quad (\text{A38})$$

$$B_{ij_0, i j_y'} = -\sigma_{j_y}^2 F_{j_y j_0'}, \quad (\text{A39})$$

$$B_{ij_0, i j_0'} = -F_{j_0' j_0}, \quad (\text{A40})$$

$$B_{ij_0, j_0 j_0'} = -F_{i, j_0'}. \quad (\text{A41})$$

### 5. GVB-open couplings

$$A_{i_x j_0 i' j_0'} = \sigma_{i_x}^2 \delta_{i_x i_x'} \delta_{j_0 j_0'} = D_{i_x j_0 i' j_0'}, \quad (\text{A42})$$

$$B_{i_x j_0, i_x j_0} = \sigma_{i_x}^2 (F_{i_x i_x} - F_{i_p} - F_{j_0 j_0}), \quad (\text{A43})$$

$$B_{i_x j_0, i_b j_0} = \sigma_{i_x} \sigma_{i_b} F_{i_a i_b}, \quad (\text{A44})$$

$$B_{i_x j_0, i_a i_a} = -\sigma_{i_x} \sigma_{i_a} F_{i_x j_0}, \quad (\text{A45})$$

$$B_{i_x j_0, i' j_0} = -\sigma_{i_x}^2 F_{i_x i'}, \quad (\text{A46})$$

$$B_{i_x j_0, i' j_0} = -\sigma_{i_x}^2 \sigma_{i_x}^2 F_{i_x i'}, \quad (\text{A47})$$

$$B_{i_x j_0, i' j_0} = -\sigma_{i_x}^2 F_{i_x i'}, \quad (\text{A48})$$

$$B_{i_x j_0, i_x j_0'} = -\sigma_{i_x}^2 F_{j' j_0}, \quad (\text{A49})$$

$$B_{i_x j_0, i_x j_y'} = -\sigma_{i_x}^2 \sigma_{j_y}^2 F_{j_y j_0'}, \quad (\text{A50})$$

$$B_{i_x j_0, i_x j_0'} = -\sigma_{i_x}^2 F_{j_0 j_0'}. \quad (\text{A51})$$

### 6. Open-open couplings

$$A_{i_0 j_0 i' j_0'} = \delta_{i_0 i_0'} \delta_{j_0 j_0'} = D_{i_0 j_0 i' j_0'}, \quad (\text{A52})$$

$$B_{i_0 j_0, i_0 j_0} = -F_{i_0 i_0} - F_{j_0 j_0}, \quad (\text{A53})$$

$$B_{i_0 j_0, i' j_0} = -F_{i' i_0}, \quad (\text{A54})$$

$$B_{i_0 j_0, i' j_0} = -\sigma_{i_x}^2 F_{i_x i_0'}, \quad (\text{A55})$$

$$B_{i_0 j_0, i' j_0} = -F_{i_0' i_0}. \quad (\text{A56})$$

The general relations for the  $\mathcal{D}$ ,  $\mathcal{D}^\dagger$  coefficients of Eq. (17) are

$$\mathcal{D}_{ij, i' j'} = \gamma D_{ij, i' j'} \mathcal{D}_{ij, i' j'}^\dagger = \delta \mathcal{D}_{ij, i' j'}, \quad (\text{A57})$$

$$ij' i' j' \text{ core or GVB orbitals: } \gamma=4, \delta=-2, \quad (\text{A58})$$

$$ii' \text{ core/GVB } jj' \text{ open orbitals: } \gamma=2, \delta=-1, \quad (\text{A59})$$

$$iji' j' \text{ open orbitals: } \gamma=1, \delta=-1. \quad (\text{A60})$$

<sup>1</sup>G. Das and A. C. Wahl, J. Chem. Phys. **56**, 3532 (1972).

<sup>2</sup>R. C. Ladner and W. A. Goddard III, J. Chem. Phys. **51**, 1073 (1969).

<sup>3</sup>B. Roos, Adv. Chem. Phys. **69**, 399 (1987).

<sup>4</sup>M. H. Gordon, J. Phys. Chem. **100**, 13213 (1996).

<sup>5</sup>F. W. Bobrowicz and W. A. Goddard, in *Methods of Electronic Structure Theory*, edited by H. F. Schaefer III (Plenum, New York, 1977), p. 79.

<sup>6</sup>P. Pulay, S. Saebo, and W. Meyer, J. Chem. Phys. **81**, 1901 (1984).

<sup>7</sup>S. Saebo and P. Pulay, J. Chem. Phys. **86**, 914 (1984).

<sup>8</sup>R. A. Friesner, Chem. Phys. Lett. **116**, 39 (1985).

<sup>9</sup>R. A. Friesner, J. Chem. Phys. **85**, 1462 (1986).

<sup>10</sup>R. A. Friesner, J. Chem. Phys. **86**, 3522 (1987).

<sup>11</sup>R. A. Friesner, J. Phys. Chem. **92**, 3091 (1988).

- <sup>12</sup>J. M. Langlois, R. P. Muller, T. R. Coley, W. A. Goddard III, M. N. Ringnalda, Y. W. Won, and R. A. Friesner, *J. Chem. Phys.* **92**, 7488 (1990).
- <sup>13</sup>B. H. Greeley, T. V. Russo, D. T. Mainz, R. A. Friesner, J. M. Langlois, W. A. Goddard III, R. E. Donnelly, Jr., and M. N. Ringnalda, *J. Chem. Phys.* **101**, 4028 (1994).
- <sup>14</sup>R. B. Murphy, R. A. Friesner, M. N. Ringnalda, and W. A. Goddard III, *J. Chem. Phys.* **101**, 2986 (1994).
- <sup>15</sup>R. B. Murphy, M. D. Beachy, R. A. Friesner, and M. N. Ringnalda, *J. Chem. Phys.* **103**, 1481 (1995).
- <sup>16</sup>K. Raghavachari and J. B. Anderson, *J. Phys. Chem.* **100**, 12960 (1996).
- <sup>17</sup>S. Saebo, W. Tong, and P. Pulay, *J. Chem. Phys.* **98**, 2170 (1993).
- <sup>18</sup>K. Wolinski and P. Pulay, *J. Chem. Phys.* **90**, 3647 (1989).
- <sup>19</sup>K. Wolinski, H. L. Sellers, and P. Pulay, *Chem. Phys. Lett.* **140**, 225 (1987).
- <sup>20</sup>K. Andersson, P. A. Malmqvist, and B. O. Roos, *J. Chem. Phys.* **96**, 1218 (1992).
- <sup>21</sup>B. O. Roos, K. A. Anderson, M. P. Fulsher, P. A. Malmqvist, L. S. Andres, K. Pierloot, and M. Merchan, *Adv. Chem. Phys.* Vol. XCII, 219 (1996).
- <sup>22</sup>C. H. Patterson and R. P. Messmer, *Chem. Phys. Lett.* **192**, 277 (1992).
- <sup>23</sup>B. H. Botch, T. H. Dunning, Jr., and J. F. Harrison, *J. Chem. Phys.* **75**, 3466 (1981).
- <sup>24</sup>N. C. Handy, P. J. Knowles, and K. Somasundram, *Theo. Chim. Acta* **68**, 68 (1985).
- <sup>25</sup>R. B. Murphy and R. P. Messmer, *J. Chem. Phys.* **97**, 4170; 4974 (1992).
- <sup>26</sup>R. B. Murphy and R. P. Messmer, *Chem. Phys. Lett.* **183**, 443 (1991).
- <sup>27</sup>R. Pauncz, *Spin Eigenfunctions-Construction and Use* (Plenum, New York, 1979).
- <sup>28</sup>S. F. Boys, in *Quantum Theory of Atoms, Molecules and the Solid State*, edited by P. O. Löwdin (Academic, New York, 1968), p. 253.
- <sup>29</sup>J. Pipek and P. G. Mezey, *J. Chem. Phys.* **90**, 4916 (1989).
- <sup>30</sup>C. Hampel and H.-J. Werner, *J. Chem. Phys.* **104**, 6286 (1996).
- <sup>31</sup>T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989).
- <sup>32</sup>A. St-Amant, W. Cornell, P. Kollman, and T. Halgren, *J. Comput. Chem.* **16**, 1483 (1995).
- <sup>33</sup>M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, *GAUSSIAN 92*, Gaussian, Inc., Pittsburgh, PA, 1992.
- <sup>34</sup>N. L. Owen and N. Sheppard, *Trans. Faraday Soc.* **6**, 634 (1964).
- <sup>35</sup>N. L. Owen and H. M. Seip, *Chem. Phys. Lett.* **5**, 162 (1970).
- <sup>36</sup>D. S. Kemp and F. Vellaccio, *Organic Chemistry* (Worth, New York, 1980), p. 325.
- <sup>37</sup>A. F. Voter and W. A. Goddard III, *J. Chem. Phys.* **57**, 253 (1981).
- <sup>38</sup>J. W. Ochterski, G. A. Petersson, and J. A. Montgomery, Jr., *J. Chem. Phys.* **104**, 2598 (1996).