



# Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

RECEIVED  
LAWRENCE  
BERKELEY LABORATORY

APR 5 1982

LIBRARY AND  
DOCUMENTS SECTION

## Materials & Molecular Research Division

Submitted to the Journal of Chemical Physics

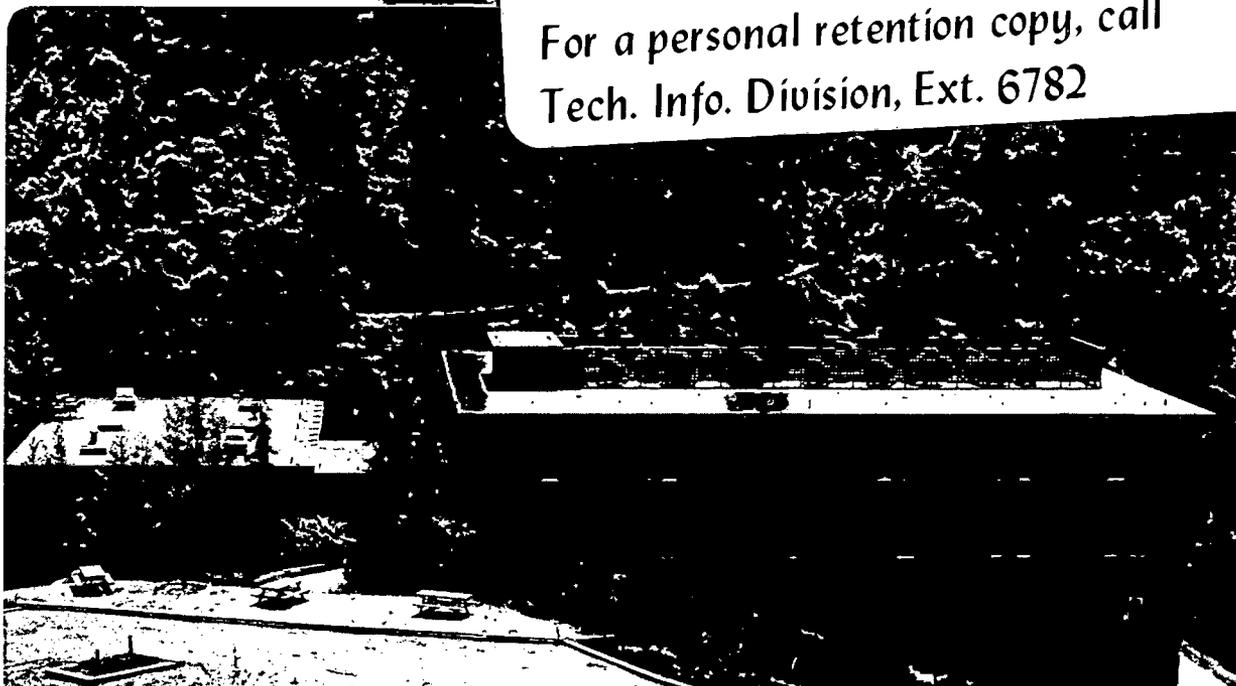
THE SHAPE-DRIVEN GRAPHICAL UNITARY GROUP APPROACH  
TO THE ELECTRON CORRELATION PROBLEM. APPLICATION  
TO THE ETHYLENE MOLECULE

Paul Saxe, Douglas J. Fox, Henry F. Schaefer III,  
and Nicholas C. Handy

March 1982

**TWO-WEEK LOAN COPY**

This is a Library Circulating Copy  
which may be borrowed for two weeks.  
For a personal retention copy, call  
Tech. Info. Division, Ext. 6782



LBL-14093  
c. 2

## DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

The Shape-Driven Graphical Unitary Group Approach to the Electron

Correlation Problem. Application to the Ethylene Molecule

Paul Saxe, Douglas J. Fox, and Henry F. Schaefer III

Department of Chemistry and Lawrence Berkeley Laboratory

University of California

Berkeley, California 94720

U.S.A.

and

Nicholas C. Handy

University Chemical Laboratory

Lensfield Road

Cambridge CB2 1EW

England

Abstract

A new method for the approximate solution of Schrödinger's equation for many electron molecular systems is outlined. The new method is based on the unitary group approach (UGA) and exploits in particular the shapes of loops appearing in Shavitt's graphical representation for the UGA. The method is cast in the form of a direct CI, makes use of Siegbahn's external space simplifications, and is suitable for very large configuration interaction (CI) wave functions. The ethylene molecule was chosen as a prototype of unsaturated organic molecules for the variational study of genuine many (i.e.,  $> 2$ ) body correlation effects. With a double zeta plus polarization

This manuscript was printed from originals provided by the authors.

basis set, the largest CI included all valence electron single and double excitations with respect to a 703 configuration natural orbital reference function. This variational calculation, involving 1,046,758 spin- and space adapted  $^1A_g$  configurations, was carried out on a minicomputer. Triple excitations are found to contribute 2.3% of the correlation energy and quadruple excitations 6.4%.

### Introduction

The problem of electron correlation in molecules has been recognized for at least 30 years as one of the central problems in theoretical chemistry.<sup>1,2</sup> During this period the state-of-the art has progressed from a six configuration wave function for the helium atom<sup>3</sup> to a 256,473 configuration variational treatment of the water molecule.<sup>4</sup> Although great strides have been made in recent years in the development of perturbation theory (most recently including complete fourth order treatments),<sup>5-7</sup> the evolution of variational methods has also accelerated<sup>8-16</sup> and it remains true that most correlated wave functions continue to adopt some sort of configuration interaction (CI) formalism. Moreover, important new methods such as the electron pair operator schemes,<sup>17-19</sup> may be shown to be related to the more conventional CI approaches. The driving force behind continuing developments is very simple and yet very powerful--namely, the desire to make quantitative theoretical predictions concerning the observable properties of ever larger molecules.<sup>20</sup>

Recently we have developed a new correlated wave function method<sup>21</sup> which simultaneously incorporates:

- (i) use of the Cooper-Nesbet algorithm<sup>22</sup> for large matrix eigenvector determination. This algorithm requires only one vector, the length of the CI expansion, to be retained in the central memory of a computer;
- (ii) all required two-electron integrals are held in the memory, enabling any matrix element to be determined without accessing peripheral storage devices;

- (iii) the many-electron expansion functions are represented as determinants, enabling simple row-by-row evaluation of the Hamiltonian matrix, with no complicated coupling parameters.

Formulated as such, the method allowed the most complete variational study to date of the  $\text{CH}_2$  singlet-triplet energy separation,<sup>21b</sup> as well as the 256,473 configuration full CI treatment<sup>4</sup> of water mentioned earlier.

However, it is clear that there are also serious disadvantages associated with the above method.<sup>21</sup> These include

(a) the fact that it is not straightforward to extend to electronic states other than the lowest state of a given symmetry;

(b) the number of two-electron integrals, of the order of  $n^4/8$ , where  $n$  is the number of orbitals directly involved in the CI, becomes very large when symmetry is reduced or the number of basis functions becomes large;

(c) the use of determinants means that the length of the CI expansion may be twice or more longer than necessary;

(d) the method was suitable only to certain types of correlated wave functions, specifically full CI; single and double excitations relative to a single self-consistent-field (SCF) reference configuration; or single and double replacement CI (CISD) with respect to a multi-reference complete active space.<sup>23</sup> Thus the method could not treat the important open-shell case of CISD within the Hartree-Fock interacting space.<sup>24</sup> Furthermore CISD with respect to a restricted set of reference configurations was not possible.

The great strength of the method was its great simplicity, leading to a very short program with highly efficient code.

Here will be introduced a new formalism which retains much of the efficiency of the original method,<sup>21</sup> and yet overcomes the disadvantages

outlined above. The new method makes extensive use of our previous experience with the loop-driven graphical unitary group approach (LDGUGA)<sup>8,9,25,26</sup> and would appear to combine several of the advantages of both schemes. After setting out the new method in adequate detail, a number of applications to the ethylene molecule are reported. These yield considerable information concerning the role (in the correlation energy) of configurations differing by more than two electrons from the Hartree-Fock reference configuration. Although the water molecule has served as a model for such studies in the past,<sup>4</sup> it may be that in the future ethylene will prove a more suitable prototype for studies of the effects of electron correlation in molecules.

### Outline of the Method

There are five essential elements of the new method, although it is to be emphasized that none of these is unique in itself:

- (i) the method returns to the Davidson technique<sup>27</sup> for large matrix eigenvector determination. This allows one at least in principle to treat arbitrary excited electronic states.
- (ii) the method requires an  $ij$ -block of electron repulsion integrals to be held in the central memory. The term  $ij$ -block designates the collection of all integrals  $(ij|k\ell)$ , with  $j \leq i, \ell \leq k$ , and  $(k\ell) \leq (ij)$ , for fixed  $i$  and  $j$ . There will be at most on the order of  $n^2/2$  such integrals in an  $ij$ -block.
- (iii) the method uses the graphical unitary group approach<sup>8</sup> for the representation of the many-electron expansion functions and the determination of Hamiltonian matrix elements. Much more detail will be given in this regard in the following section of the paper.
- (iv) the method partitions external and internal parts of the Shavitt graph<sup>8</sup> as proposed by Siegbahn.<sup>10</sup> Partial loops of a similar shape share coupling coefficients and contribute to sequential elements of the  $\sigma$  ( $\equiv HC$ ) vector so that they may be identified with simple program loops. The number of related elements can reach 300 for even a modest sized calculation. This fact paired with the simple loop structure makes vectorization possible.

- (v) the sequential nature of these elements points to an underlying block structure to  $H$  which has been utilized to select only those portions of  $\sigma$  and  $C$  that are needed in core. This effectively removes the length of the configuration list as a limitation to be considered.

Taken together, these features yield an efficient theoretical framework capable of treating single and double replacement calculations (CISD) out of any set of reference functions. One of the wave functions reported here for ethylene includes variationally more than one million spin- and space-adapted configurations.

The Shape Driven Graphical Unitary Group Approach (SDGUGA)

The Davidson diagonalization technique<sup>27</sup> requires the construction of the  $\underline{\sigma}$  vector, which for the  $k^{\text{th}}$  iteration is given by

$$\sigma_{m'}^{(k)} = \sum_m H_{m',m} C_m^{(k-1)} \quad (1)$$

Rewriting this in the formalism of GUGA,<sup>8</sup> one obtains

$$\sigma_{m'}^{(k)} = \sum_{ijm} \langle m' | E_{ij} | m \rangle \langle i | h | j \rangle C_m^{(k-1)} + \frac{1}{2} \sum_{ijklm} \langle m' | e_{ijkl} | m \rangle [ij;kl] C_m^{(k-1)} \quad (2)$$

The matrix elements of the one- and two-body unitary group operators  $E_{ij}$  and  $e_{ijkl}$  are determined as the product of segment shapes  $W$  of loops formed by walks  $m$  and  $m'$  as follows

$$\langle m' | E_{ij} | m \rangle = \prod_{k=i}^j W(T_k, b_k)$$

$$\langle m' | e_{ijkl} | m \rangle = \{ \prod W(T_k, b_k) \} \sum_{x=0,1} \{ \prod W(T_k, b_k) \} \quad (3)$$

where the factorization of  $e_{ijkl}$  is that given recently by Shavitt<sup>28</sup> to avoid summation over intermediate states. The first method developed by Brooks and Schaefer<sup>9</sup> used the loop-driven algorithm to simultaneously locate and evaluate via equation (3) all loops required for a specific wave function. Subsequently the loop contributions  $\langle m' | E_{ij} | m \rangle \langle i | h | j \rangle$  or

$\langle m' | e_{ijkl} | m \rangle [ij;kl]$  were written on a peripheral (disc) file as a diagonalization tape.<sup>25</sup>

The above approach, though extraordinarily general in many respects<sup>26</sup> (the choice of configurations is virtually unrestricted, for example), is not suitable when implemented as a direct CI. In a direct CI framework the loop-driven methodology becomes inefficient, requiring typically twice as long to generate all the loops as to perform the multiplication by the CI vector. Although this amount of effort is insignificant if the diagonalization tape is created once and stored, it becomes impractical if repeated on every diagonalization iteration, of which there are typically  $\sim 7$ . Siegbahn has proposed<sup>10,30</sup> a way of finding and evaluating most loops with much less effort, for certain types of CI expansions, by implicitly finding all loops or parts of loops in the external space. In essence, the products in equation (3) are broken into two terms--one a product in the external space and the other in the internal space, the total matrix element being the product of the two terms. Because the external space is so repetitively simple for expansions that allow one or two electrons in external orbitals, the code to explicitly handle external parts of loops can be generated without too much difficulty. The more complex internal space is then handled by a general method such as the loop-driven technique.

We propose here a method driven by the shapes of the external parts of loops. In this regard the two loops seen in Figure 1 help to illustrate how the loop shape "drives" the method. Note that these two loops share the same internal partial loop but differ in the external shape. If there are  $N_{\text{ext}}$  external orbitals and no point group symmetry,

there will be  $N_{\text{ext}}(N_{\text{ext}}-1)/2$  loops of type A that differ only in the levels a and b, and  $N_{\text{ext}}$  loops of type B. In a modest size calculation, this one internal part of a loop will give rise to about 1000 loops which will have one of two values for the coupling coefficient  $\langle m' | e_{ijkl} | m \rangle$ , depending on whether this loop has the shape of A or the shape of B. Therefore one can write two short sections of code that, when passed sufficient information about the internal part of the loop will construct all these loops very efficiently. The two different values of the coupling coefficient are known immediately upon completion of the internal part of the loop, so the first section of the algorithm need only loop over the levels a and b, which will define not only the walks  $m'$  and  $m$  but also the integrals  $[ia,ib]$  that will be used in equation (2). The second section can then process all the loops of type B. By specializing the method to this degree, it is possible to order the walks and integrals so that a set of pointers (which are simply incremented) specify the locations in the two vectors and the storage location of the needed integrals.

To accomplish the above, the numbering scheme of the walks must be slightly modified from the normal lexical order.<sup>8,9</sup> The only change required is the renumbering of those walks with two electrons singlet coupled in the external space, so that all the external walks with a doubly-occupied orbital from a given internal walk are given sequential indices. The integrals can be held in the order in which they will be used by sorting them according to the number of repeated indices. The eight lists of integrals formed in this way are then sorted into the necessary order, storing doublets or triplets of integrals with the

same indices together, as proposed by Brooks. We have not yet implemented this ordering, as it is not critical on a scalar machine. For a vector machine this ordering will enable the code to be run in vector mode.

With the above in mind, each of the one hundred or so different possible external parts of loops can be used to generate explicit code, of which the specific example given earlier will be one section. The external part of the coupling coefficient is determined analytically by equation (3) and is explicitly coded into each section of the program. Note that the loops constructed by one such section of code will contribute to one rectangular submatrix of  $H$  that has maximum dimensions of about  $N_{\text{ext}}^2/2$ . In the example given, the block had dimensions of  $N_{\text{ext}}(N_{\text{ext}}+1)/2$  and unity, so that only those sections of the CI vector and  $\underline{\sigma}$  vector that correspond to this submatrix need be in central memory. Because the  $H$  matrix is symmetric and one creates effectively only the lower half matrix, the core requirements are space for an ij-block of integrals and about  $2 N_{\text{ext}}^2$  elements of the two vectors. The current program actually requires  $m^3$  groups of integrals, but it would not be difficult to modify the program so that only  $n^2$  groups of integrals, with two common indices, are needed at one time, and thus modified the program could handle much larger basis sets. However, our present minicomputer (the Harris Corporation Series 800) is not sufficiently powerful to justify the extensions to cases significantly larger than the 1,046,758 configuration variational treatment of ethylene presented in this paper.

One final point about organizing the program in this fashion needs to be emphasized. Although the method is quite effective on a scalar machine such as the Harris 800 or CDC 7600, it also seems possible to adapt the SDGUGA formalism to vector processors such as the Cray machines or the CDC Cyber 205. This combined with the minimal amount of input/output needed per iteration (essentially one read of the integral file) suggests that the method might be fast compared to previous direct CI programs if run on a supercomputer. Even on a machine as primitive as the Harris 800, the simplicity of the code enhances performance due to the modest pipelining feature of the central processor.

The procedure used to generate the distinct row table, which is the numerical form of the Shavitt graph, has been reformulated to accommodate the SDGUGA method. The new program generates distinct row tables for an arbitrary excitation level from reference configurations given as input. If only a single reference or certain specific multi-reference treatments are desired, the configuration list can be limited to the Hartree-Fock interacting space.<sup>24</sup> Multireference cases such as CISD relative to a CASSCF reference set<sup>23</sup> or CISD relative to a more limited set of reference configurations (e.g., single and double excitations within a small set of valence orbitals) can be handled if the full spin space is allowed. The full exploitation of symmetry is limited to Abelian groups or subgroups, as in our previous work<sup>9,25,26</sup> based on the UGA.<sup>31</sup>

### Preliminary Results

A preliminary implementation of the present method indicated that it was at least as efficient as the earlier LDGUGA method<sup>25</sup> for small cases. This first attempt did not treat symmetry at all, but as the innermost sections of the program remained unchanged as symmetry was incorporated, little change in performance was anticipated. Indeed, identical calculations on systems with no elements of point group symmetry showed the later version of the program to be quite similar in timing characteristics. The inherent structure of the method makes the inclusion of symmetry quite simple, most cases amounting to restricting the ranges of DO statements.

Timings also indicated that typically just over 50% of the central processor time was used in the multiplication of  $H$  by  $C$ . The bulk of the remaining time must be spent multiplying integrals by coupling coefficients to form matrix elements, with very little effort expended for index calculations. Also, since the innermost loops run over the number of external orbitals, the method should become more efficient for larger calculations.<sup>10,16,30</sup>

One of the first test cases was a CISD treatment of the transition state for formaldehyde dissociation<sup>32</sup> to  $H_2+CO$ . Using a standard double zeta plus polarization (DZ+P) basis set and point group  $C_s$  (the molecular plane is the only element of symmetry), this case involves 10,221 configurations when the core (carbon and oxygen 1s-like) orbitals are deleted from the CI procedure. The identical calculation was run with SDGUGA in Cambridge on the IBM 370/165 and in Berkeley on the Harris 800. Both calculations converged to within  $1 \times 10^{-8}$  hartree on the

seventh diagonalization interaction, and the total energies agree to within  $1 \times 10^{-7}$  hartree. Timings were 7 minutes for the IBM machine and 21 minutes for the Harris 800.

### Many-Body Correlation Effects in Ethylene

During the past five years theoretical interest has intensified<sup>4-7,33,34</sup> with respect to the importance of true many-body correlation effects, i.e., configurations differing by more than two electrons from the Hartree-Fock reference configuration. It has of course been known for more than a decade that as diatomic molecules separate to their atomic dissociation limits, these higher order correlation effects become very important.<sup>35</sup> Moreover, the importance of certain types of higher excitations (unlinked clusters) is well established and several relatively simple but reliable schemes exist for estimating these effects.<sup>36-39</sup> However, in recent years the spectre has been raised<sup>6,7</sup> that even for stable closed-shell molecules near their equilibrium geometries, it may be necessary to evaluate the importance of triple excitations, as well as both unlinked and linked quadruple excitations. For example, Frisch, Krishnan, and Pople<sup>34</sup> state that "the contribution of triple substitutions to the electron correlation energy is fairly large, particularly for multiply bonded molecules." Thus it is of critical importance to have detailed variational studies of these higher excitation correlation effects for at least a few small molecules. Whereas much attention has recently focused (quite properly) on the water molecule,<sup>4,6,7,33</sup> we turn in this study to a

slightly larger system, the ethylene molecule.

Since it is possible that the present work will establish  $C_2H_4$  as something of a benchmark for the study of correlation effects in unsaturated organic molecules, it is important to state the precise details of the geometry and basis set used here. The assumed ethylene geometry<sup>40</sup> involved  $r(C=C) = 1.330 \text{ \AA}$ ,  $r(C-H) = 1.076 \text{ \AA}$ ,  $\theta(HCH) = 116.6^\circ$ . In atomic units this places the carbon atoms at cartesian positions  $(0, 0, \pm 1.25666814)$  and the hydrogen atoms at  $(0, \pm 1.72999314, \mp 2.32513368)$  resulting in a nuclear repulsion energy of 33.51358956 hartrees. It is hard to imagine that one could reproduce this nuclear repulsion energy to high precision without having the correct ethylene geometry.

The basis set used was a standard double zeta plus polarization (DZ+P) set of contracted gaussian functions.<sup>41</sup> For carbon the (9s5p) primitive gaussian basis of Huzinaga<sup>43</sup> was contracted to (4s2p) following Dunning.<sup>44</sup> In like manner the (4s) hydrogen basis was contracted to (2s) and scaled by a factor of 1.2, i.e., all the original hydrogen orbital exponents were multiplied by 1.44. To each carbon atom was then appended a set of six cartesian d functions with orbital exponent 0.75. Similarly  $p_x$ ,  $p_y$ , and  $p_z$  functions with orbital exponent  $\alpha = 1.0$  were added to each hydrogen atom. The technical designation for the DZ+P basis thus described is C(9s5p1d/4s2 1d), H(4slp/2slp), and it includes 88 primitive and 52 contracted gaussian functions. The authors will provide a computer printout of this basis set and geometry to persons interested in pursuing these technical details further.

The electronic ground state of  $C_2H_4$  is of  ${}^1A_g$  symmetry and arises from the electron configuration

$${}^1A_g \quad 1a_g^2 \quad 1b_{1u}^2 \quad 2a_g^2 \quad 2b_{1u}^2 \quad 1b_{2u}^2 \quad 1b_{3g}^2 \quad 3a_{ag}^2 \quad 1b_{3u}^2 \quad (4)$$

The additional (unoccupied) valence orbitals are the  $1b_{2g}$ ,  $3b_{1u}$ ,  $2b_{2u}$ ,  $4a_g$ ,  $2b_{3g}$ , and  $4b_{1u}$  orbitals, and these are expected on intuitive grounds to be the most important virtual orbitals. However, in the context of conventional single configuration self-consistent-field (SCF) theory, these six virtual orbitals do not necessarily take on their intuitively desirable characteristics.<sup>41</sup> Therefore a transformation to natural orbitals<sup>42</sup> was carried out, based on a CI including single and double excitations relative to (4). With the  $1a_g$  and  $1b_{1u}$  core-like orbitals constrained to be always doubly occupied, there are 5057  ${}^1A_g$  configurations included in the CISD treatment. The occupation numbers from the resulting natural orbital transformation are given in Table I and confirm the breakdown of orbitals suggested above. There is a very large gap between the occupation number of the highest occupied molecular orbital ( $1b_{3u} = 1\pi$ , occupation 1.943) and that of the lowest unoccupied orbital ( $1b_{2g} = 1\pi^*$ , occupation 0.046). A much smaller but still very evident separation is seen between the occupation numbers of the last valence orbital ( $4b_{1u}$ , 0.014) and the first orbital falling outside the intuitive valence space ( $2b_{3u}$ , 0.008). Thus the classification of the orbitals in Table I into three general types is seen to be justifiable.

To variationally consider the nature of true many-body correlation effects, it would be desirable to include all triple and quadruple excitations (as well as the 5057 single and double excitations) in a

CI. However, such a CI, even with the core  $1a_g$  and  $1b_{1u}$  orbitals constrained to be doubly occupied, includes 10,593,385 configurations and is clearly beyond the scope of the present research. Accordingly, we have set a more realistic goal, namely the inclusion of all single and double excitations plus a limited (although rather large) number of triple and quadruple excitations. Only the valence shell correlation energy was investigated, i.e., the  $1a_g$  and  $1b_{1u}$   $1s$ -like orbitals were doubly-occupied in every configuration. More importantly, only those triple and quadruple excitations relative to (4) have been variationally included in which no more than two electrons occupy orbitals beyond the set of (two core and) twelve valence orbitals. There are a total of 1,046,758 configurations obtained in this manner.

An alternate way of looking at the present treatment of triple and quadruple excitations readily presents itself. Within the space of the twelve valence orbitals, there are 702 single and double excitations relative to the Hartree-Fock configuration (4). If we consider these 703 (HF+S+D) configurations to be reference functions, then the 1,046,758 configurations are all distinct single and double replacements arising from any of these 703 reference configurations. This procedure is well-defined for any molecule and basis set and guaranteed to include the most important triple and quadruple excitations. Moreover, the procedure automatically excludes higher than quadruple excitations, unlike methods including all single and double excitations relative to the members of a CASSCF.<sup>23</sup> In few-electron cases such (as the ethylene ground state) where the Hartree-Fock approximation is

qualitatively reasonable, it is probably wise to systematically exclude the relatively unimportant<sup>4</sup> quintuple, sextuple, and higher excitations.

Although the present approach does include the most important triple and quadruple excitations, one should appreciate that some of the less important configurations included are less important than some of those excluded. For example, consider a quadruple excitation of the type

$$i^2 j^2 \rightarrow 2b_{3g}^2 13b_{1u}^2 \quad (5)$$

where  $i$  and  $j$  are occupied in the Hartree-Fock picture, while the virtual orbital  $13b_{1u}$  has the lowest occupation number (0.0000007) of all the natural orbitals. This configuration is included in the present CI, but likely to be rather unimportant, not only because of the unimportance of the  $13b_{1u}$  orbital, but also the fact that the  $2b_{3g}$  is the least important (see Table I) of the ethylene valence orbitals. The configuration

$$i^2 j^2 \rightarrow 2b_{3u}^2 5a_g^2 \quad (6)$$

is likely to be more important than (5), even though it is not included in our CI, since both the  $2b_{3u}$  and  $5a_g$  orbitals (although not valence orbitals) do have large (0.008 and 0.007) natural orbital occupation numbers. Although more elaborate schemes could be concocted to exclude the likes of (5) or include configurations such as (6), the simplicity of the present a priori selection technique commends itself.

To look specifically at the importance of triple excitations,<sup>4,6,7,21b</sup> a separate CI treatment was carried out. This begins with a 37 configuration reference function including all single excitations relative to (4) within the valence space. Then all single and double excitations relative to this 37 configuration reference function were included, a total of 109,473 configurations. As with the above-discussed quadruples procedure,<sup>8</sup> this method guarantees the variational inclusion of all the most important triple excitations.

Table II summarizes the results of the present study of the ethylene  ${}^1A_g$  ground state. It may be noted initially that the valence natural orbitals, as expected, are very similar to the canonical occupied Hartree-Fock orbitals. The energy of the first natural configuration (4) is only about one millihartree above the SCF energy. Moreover, the CISD energy based on natural orbitals is only about 0.1 millihartrees above that based on the canonical SCF orbitals. Computation times on the Harris 800 minicomputer were (per diagonalization iteration, of which 7 are typically required), 1 minute for the 5057 CI, 46 minutes for the  $10^5$  CI, and 13 hours for the  $10^6$  CI cited in Table II.

The bulk of the correlation energy

$$E_{\text{CORR}} = E_{\text{CI}} - E_{\text{SCF}} \quad (7)$$

within the present DZ+P basis set is of course the 0.27747 hartrees obtained at the CISD level with natural orbitals. A very similar correlation energy was reported earlier by Brooks<sup>45</sup> in the context of his study of the  $\pi \rightarrow \pi^*$  excited singlet state of ethylene. The best previous variational calculation on ethylene is that of Ahlrichs, Lischka, Zurawski, and Kutzelnigg,<sup>46</sup> who reported SCF

and CI energies of -78.05357 and -78.33196 hartrees using a triple zeta plus polarization (TZ+P) basis set C(9s5p1d/5s3p1d), H(5s1p/3s1p) only slightly larger than our own. Their correlation energy of 0.27840 is sufficiently close to our CISD result that an evaluation of their CEPA treatment<sup>36</sup> of unlinked clusters is possible.

It should be explicitly noted that the purpose of this research is not to contend for the lowest variational energy to date for ethylene, although this was accidentally achieved. Our purpose is rather to probe the importance of triple and quadruple substitutions in a full variational treatment. To this end, we now take as our reference energy the natural orbital CISD result of -78.32800 hartrees. The addition of 104,416 triple excitations to the CISD list of 5057 lowers the total energy by 7.00 millihartrees. Triple excitations are also difficult to deal with via perturbation and the only results to date for systems with more than one heavy atom are those of Pople and co-workers.<sup>6,34,47</sup> For ethylene with their 6-316\* basis set Frisch, Krishnan, and Pople<sup>34</sup> find the contribution of triples to be 8.15 millihartrees in fourth order. The 6-31G\* basis, however is significantly smaller than our DZ+P basis, which includes a set of p functions on each hydrogen atom. Therefore the difference between the 7.00 millihartree variational result and 8.15 perturbation result will certainly increase when strictly comparable basis sets are used. Nevertheless, the agreement is encouraging. We suspect that the remaining discrepancy of  $\sim 30\%$  (attempting to correct for the difference in basis sets from Pople's more extensive work<sup>47</sup> on acetylene) is due to a combination of our neglect of the less important triples and to the overestimation of their importance in fourth order, the first order of perturbation theory in which triples are treated.

Triples and quadruples taken together (final entry in Table II

minus the NO CISD result) contribute 26.51 millihartrees to the correlation energy. Subtracting the 7.00 millihartrees previously attributed to triple excitations, we can ascribe 19.51 millihartrees to the effects of quadruples. Thus in the variational picture, quadruple excitations are seen to be nearly three times as important as triples, a result consistent with traditional thinking.<sup>41</sup> However, comparison with perturbation theory is not straightforward, since perturbation theory views a significant part of the variational quadruples energy as a correction<sup>5</sup> to the fact that CI with double excitations is not "size consistent". We prefer the variational breakdown, but will not press for a perhaps artificial comparison here.

The total variational contribution due to triple and quadruple excitations, 26.51 millihartrees, is remarkably close to that (26.07 millihartrees) predicted by the Davidson correction.<sup>37</sup> This must be considered fortuitous since the Davidson correction makes no attempt to incorporate the results of triple excitations. This simple correction rather represents (in many body perturbation theory) the unlinked fourth order contribution of double excitations that must be cancelled by unlinked contributions from quadruple excitations.<sup>37</sup> In any case it appears that the fortuitous agreement between  $E(\text{SDTQ})$  and  $E(\text{Davidson corrected CISD})$  is primarily due to an overestimate of these unlinked clusters by the Davidson formula. However, it must be recognized that although more than 900,000 quadruple excitations (including all of the most important ones) have been included variationally, many less important quadruples have

been neglected and it is difficult to evaluate the importance of configurations such as (6). Ahlrichs CEPA calculation<sup>46</sup> with a TZ+P basis actually predicts an even larger role (31.35 millihartrees) for unlinked clusters than does the Davidson correction, but the slightly larger basis of Ahlrichs causes some equivocation in this respect.

It is possible to make a crude estimate of the relative importance of the neglected quadruple excitations using the natural orbital occupation numbers. If  $i$  and  $j$  are orbitals occupied in (4) then the total contribution of all quadruples of the type

$$i^2 j^2 + x^2 y^2 \quad (8)$$

(all  $i$ , all  $j$ , one  $x$ , one  $y$ ) is probably roughly proportional in some sense to the product of the occupation numbers  $b_x$  and  $b_y$  of natural orbitals  $x$  and  $y$ . In this respect, the sum of products of occupation numbers

$$\sum_x \sum_{y>x} b_x b_y \quad (9)$$

is a rough measure of the total importance of such "diagonal" quadruple excitations. For our entire natural orbital basis set (beyond the orbitals occupied in the FNC) the sum (9) has the value 0.0157. However, in our treatment of quadruples either  $x$  or  $y$  can lie beyond the valence space (i.e., occupation number  $b < 0.014$ ), but not both. The appropriate sum (9) going over the included diagonal quadruples is 0.0138, or 88% of the complete summation.

Although this measure is admittedly a very naïve one, it does suggest with plausibility that the bulk of the important quadruple excitations have been included variationally in this research.

To conclude this section of the paper, it may be stated that our most complete CI wave function for ethylene, including more than one million configurations treated with full variational freedom, gives a correlation energy of 303.98 millihartrees. Of this 277.47 millihartrees or 91.3% of the correlation energy is due to single and double excitations. 7.00 millihartrees or 2.3% is due to triple excitations, while 19.51 millihartrees or 6.4% of the correlation energy may be attributed to quadruple excitations. It must be emphasized that although all these results arise from variational treatments, the actual importance of triple and quadruple replacements will be somewhat greater, since only the most important have been included here. It is our hope that this work will establish DZ+P ethylene as something of a benchmark for the study of electron correlation in small molecules, and in particular that various perturbation methods will be applied to this model unsaturated organic system.

### Discussion

The primary reason the shape driven method (SDGUGA) is more efficient than our previous techniques<sup>9,21,25,26</sup> is the same reason the method is able to deal simultaneously with an entire rectangular block of  $H$ , the Hamiltonian matrix. Namely, the present method ignores the number of upper walks from a loop and only uses the number of lower walks when needed. Previous implementations have broken down the upper and lower walks from a loop and used each loop everywhere it occurs. The very simplicity of the external space that Siegbahn's approach<sup>10,30</sup> exploits guarantees that most loops have one lower walk. Furthermore, the majority of loops open near the top of the graph, so they cannot have many upper walks. In this sense previous implementations spent considerable time setting up literally millions of DO statements ranging from one to one.

Since our previous LDGUGA used reverse lexical ordering<sup>9</sup> instead of lexical,<sup>8</sup> the lower walks determined the location in the  $\sigma$  vector of series of sequential upper walks. On those occasions where the number of lower walks was greater than one the resulting contributions were randomly spaced over the  $\sigma$  vector and random access of the entire vector was required. The nature of either lexical or reverse lexical order guarantees that upper or lower walks, respectively, will change the index by large amounts making simultaneous breakdown of both upper and lower walks impractical unless the entire vector is held in central memory. The evaluation of loops is sufficiently rapid in

an externally based method that is much faster to ignore the number of upper walks from a loop and re-evaluate the loop for each upper walk.

The use of lexical, rather than reverse-lexical, order is crucial to the success of the current implementation of SDGUGA. The use of lexical order simplifies greatly the extension of the method to cases including higher than double-excitations; is mandatory for large CI expansions, where the  $C$  and  $\sigma$  vectors are too long to be held simultaneously in central memory; and is essential if there is any hope of vectorizing the code. The search for valid loops always begins at the graph head. Since all loops originating at one level in the graph are processed before continuing at the next level, it is necessary to find all possible paths from the graph head to the level of interest. There are two advantages to this approach. First, it eliminates any need to break down the number of upper walks from a loop, and thus eliminates the need for a reordering vector spanning the set of internal walks. Without the reordering vector, it is possible to variationally treat higher excitations within a direct CI framework. For example, the current program can handle cases involving all single, double, triple and quadruple excitations relative to a set of one or more reference configurations. In order to accomplish this, the "external space" is redefined for each loop as the space where both walks have two or less electrons. The external space cannot easily be redefined in this fashion if a reordering vector is employed. For cases involving higher excitations, the SDGUGA retains most of its efficiency since the loops which do not enter a redefined external space before terminating will in general have large numbers

of lower walks. The second advantage to beginning at the graph head is that the walks from graph head to loop head are found in lexical order. Because of this, the Hamiltonian matrix is evaluated in an orderly fashion starting from the upper left corner and proceeding toward the lower right corner. Since only those portions of the  $\underline{C}$  and  $\underline{g}$  vectors corresponding to the portion of the Hamiltonian actually being constructed need be in central memory it is straightforward to transfer the appropriate sections of the vectors from peripheral storage to the central memory in cases where both vectors cannot simultaneously be held in the central memory. This process is extremely efficient due to the methodical fashion in which the Hamiltonian matrix is constructed. Except in very large cases, there is no need to transfer a particular portion of the vectors into central memory more than a few times per iteration of the Davidson method. The SDGUGA is therefore able to handle any problem that can be treated via the LDGUGA (which is significantly more flexible than most other UGA forms), but with all the advantages of a direct CI method.<sup>29</sup> Furthermore, in most cases the SDGUGA already requires less time than the LDGUGA, and even on a scalar machine such as our Harris minicomputer, it will be worth reordering the integrals so that the code is vectorizable, as even less time will be expended computing indices.

From the tenor of this report it will be apparent that great progress has been made during the past two years by our group and others<sup>5-8,10-16,18,19,28,30,34</sup> in the theoretical treatment of electron correlation in molecules. How long this exciting trend will continue

is difficult to predict. Hopefully, ten years from now the million configuration variational treatment of ethylene reported here will be considered routine. Should this be the case, the benefits for chemistry will be very great indeed.

#### Acknowledgments

We thank Professor I. Shavitt for a challenge that provided much of the inspiration for the present research. This work was supported by the U.S. National Science Foundation under Grant CHE-8009320, and by the Director, Office of Energy Research, Office of Basic Energy Sciences, Chemical Sciences Division of the U.S. Department of Energy under Contract Number DE-AC03-76SF00098.

References

1. D. P. Craig, Proc. Roy. Soc. 200A, 474 (1950); 202A, 498 (1950).
2. S. F. Boys, Proc. Roy. Soc. 200A, 542 (1950).
3. G. F. Taylor and R. G. Parr, Proc. Natl. Acad. Sci. U.S. 38, 154 (1952).
4. P. Saxe, H. F. Schaefer, and N. C. Handy, Chem. Phys. Lett. 79, 202 (1981).
5. R. J. Bartlett and G. D. Purvis, Physica Scripta 21, 255 (1980).
6. R. Krishnan, M. J. Frisch, and J. A. Pople, J. Chem. Phys. 72, 4244 (1980).
7. S. Wilson and M. F. Guest, Chem. Phys. Lett. 72, 49 (1980); 73, 607 (1980).
8. I. Shavitt, Intern. J. Quantum. Chem. Symp. 11, 131 (1977); 12, 5 (1978).
9. B. R. Brooks and H. F. Schaefer, J. Chem. Phys. 70, 5092 (1979).
10. P. E. M. Siegbahn, J. Chem. Phys. 70, 5391 (1979).
11. R. J. Buenker, section 1.5 of Quantum Chemistry into the 1980's, editor P. G. Burton, Proceedings of the Molecular Physics and Quantum Chemistry Workshop, Wollongong, Australia, February, 1980.
12. B. Liu and M. Yoshimine, J. Chem. Phys. 74, 612 (1981).
13. E. R. Davidson, L. E. McMurchie, and S. J. Day, J. Chem. Phys. 74, 5491 (1981).
14. A. F. Voter and W. A. Goddard, J. Chem. Phys. 75, 3638 (1981).
15. C. Zirz and R. Ahlrichs, J. Chem. Phys. 75, 4980 (1981).
16. H. Lischka, R. Shepard, F. B. Brown, and I. Shavitt, Int. J. Quantum Chem. Symp. 15, 91 (1981).

17. W. Meyer, J. Chem. Phys. 64, 2091 (1976); C. E. Dykstra, H. F. Schaefer and W. Meyer, J. Chem. Phys. 65, 2740 (1976).
18. S. M. Bachrach, R. A. Chiles, and C. E. Dykstra, J. Chem. Phys. 75, 2270 (1981).
19. H.-J. Werner and E.-A. Reinsch, J. Chem. Phys. 76, 0000 (1982).
20. W. G. Richards, Nature 278, 507 (1979).
21. (a) N. C. Handy, Chem. Phys. Lett. 74, 280 (1980); (b) P. Saxe, H. F. Schaefer, and N. C. Handy, J. Phys. Chem. 85, 745 (1981).
22. R. K. Nesbet, J. Chem. Phys. 43, 311 (1965).
23. B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, Chem. Phys. 48, 157 (1980).
24. A. Bunge, J. Chem. Phys. 53, 20 (1970); C. F. Bender and H. F. Schaefer, J. Chem. Phys. 55, 4798 (1971).
25. B. R. Brooks, W. D. Laidig, P. Saxe, N. C. Handy, and H. F. Schaefer, Physica Scripta 21, 312 (1980).
26. B. R. Brooks, W. D. Laidig, P. Saxe, J. D. Goddard and H. F. Schaefer, pages 158-176 of The Unitary Group for the Evaluation of Electronic Energy Matrix Elements, editor J. Hinze, Vol. 22 of "Lecture Notes in Chemistry" (Springer-Verlag, Berlin, 1981).
27. E. R. Davidson, J. Comput. Phys. 17, 87 (1975).
28. I. Shavitt, pages 51-99 of the book cited in reference 26.
29. B. Roos, Chem. Phys. Lett. 15, 153 (1972).
30. P. E. M. Siegbahn, J. Chem. Phys. 72, 1647 (1980).
31. For a definitive discussion of point groups in the context of GUGA, see I. Shavitt, Chem. Phys. Lett. 63, 421 (1979).
32. For background on this chemical problem, see J. D. Goddard, Y. Yamaguchi, and H. F. Schaefer, J. Chem. Phys. 75, 3459 (1981).

33. W. D. Laidig, P. Saxe, and H. F. Schaefer, *J. Chem. Phys.* 73, 1765 (1980).
34. M. J. Frisch, R. Krishnan, and J. A. Pople, *Chem. Phys. Lett.* 75, 66 (1980).
35. P. S. Bagus, B. Liu, A. D. McLean, and M. Yoshimine, pages 99-118 of Wave Mechanics: The First Fifty Years, editors W. C. Price, S. S. Chissick, and T. Ravendale (Butterworths, London, 1973).
36. For a discussion of the CEPA approach to unlinked clusters, see W. Meyer, *J. Chem. Phys.* 58, 1017 (1973).
37. E. R. Davidson, pages 17-30 of The World of Quantum Chemistry, editors R. Daudel and B. Pullman (Reidel, Dordrecht, Holland, 1974); E. R. Davidson and D. W. Silver, *Chem. Phys. Lett.* 52, 403 (1977).
38. J. A. Pople, R. Seeger, and R. Krishnan, *Intern. J. Quantum Chem. Symp.* 11, 149 (1977).
39. P. E. M. Siegbahn, *Chem. Phys. Lett.* 55, 386 (1978).
40. J. S. Binkley, J. A. Pople, and W. J. Hehre, *J. Amer. Chem. Soc.* 102, 939 (1980).
41. H. F. Schaefer, The Electronic Structure of Atoms and Molecules: A Survey of Rigorous Quantum Mechanical Results (Addison-Wesley, Reading, Massachusetts, 1972).
42. P. O. Löwdin, *Adv. Chem. Phys.* 2, 207 (1959).
43. S. Huzinaga, *J. Chem. Phys.* 42, 1293 (1965).
44. T. H. Dunning, *J. Chem. Phys.* 53, 2823 (1970).
45. B. R. Brooks and H. F. Schaefer, *J. Chem. Phys.* 68, 4839 (1978).
46. R. Ahlrichs, H. Lischka, B. Zurawski, and W. Kutzelnigg, *J. Chem. Phys.* 63, 4685 (1975).

47. R. Krishnan, M. J. Frisch, J. A. Pople, and P. R. Schleyer,  
Chem. Phys. Lett. 79, 408 (1981).
48. See, for example, R. P. Hosteny, R. R. Gilman, T. H. Dunning,  
A. Pipano, and I. Shavitt, Chem. Phys. Lett. 7, 325 (1970).

Table I. Natural orbital occupation numbers (greater than 0.001) for the ground state of ethylene. These natural orbitals were obtained from a CI including single and double excitations from the Hartree-Fock reference configuration and were used in all subsequent calculations.

<u>Occupied in Hartree-Fock Picture</u>		<u>Valence Orbitals</u> <u>Unoccupied in Hartree-Fock Picture</u>		<u>Most Important Additional</u> <u>Natural Orbitals</u>	
1a <sub>g</sub>	2.000	1b <sub>2g</sub>	0.046	2b <sub>3u</sub>	0.008
1b <sub>1u</sub>	2.000	3b <sub>1u</sub>	0.017	5a <sub>g</sub>	0.007
2a <sub>g</sub>	1.982	2b <sub>2u</sub>	0.017	3b <sub>2u</sub>	0.007
2b <sub>1u</sub>	1.976	4a <sub>g</sub>	0.015	6a <sub>g</sub>	0.006
1b <sub>2u</sub>	1.973	2b <sub>3g</sub>	0.014	5b <sub>1u</sub>	0.005
3a <sub>g</sub>	1.969			1b <sub>1g</sub>	0.005
1b <sub>3u</sub>	1.943			3b <sub>3u</sub>	0.004
				3b <sub>3g</sub>	0.003
				1a <sub>u</sub>	0.003
				4b <sub>2u</sub>	0.002
				2b <sub>2g</sub>	0.002
				6b <sub>1u</sub>	0.002
				7a <sub>g</sub>	0.002
				4b <sub>3g</sub>	0.001
				5b <sub>2u</sub>	0.001

Table II. Summary of correlated wave functions and energies (in hartrees) for the ground state of ethylene near the equilibrium geometry. A standard double zeta plus polarization basis set (see text) was used throughout. As discussed in the text, each correlated wave function involves single and double excitations with respect to one or more reference configurations. In the final column,  $C_0$  is the coefficient of the Hartree-Fock configuration in the CI wave function.

	Number of Reference Configurations	Total Number Configurations	Total Energy	$C_0$
Self-Consistent-Field	-	1	-78.050 53	1.0
Canonical SCF Orbitals	1	5057	-78.328 11	0.9516
Above, Davidson Corrected	-	-	-78.354 34 <sup>a</sup>	-
First Natural Configuration	-	1	-78.049 42	1.0
Natural Orbitals	1	5057	-78.328 00	0.9521
Above, Davidson Corrected	-	-	-78.354 07 <sup>a</sup>	-
Triple Excitations	37	109,473	-78.335 01	0.9481
Triples and Quadruples	703	1,046,758	-78.354 51	0.9354

<sup>a</sup>Not a variational result. See reference 37.

Figure Caption

Figure 1. Sketch intended to illustrate the sense in which the SDGUGA is "shape driven". Note that the internal space includes orbitals occupied in one or more of the CI reference configurations, while the external space includes orbitals not occupied in any of the reference configurations, with respect to which various excitation levels of CI may be carried out.

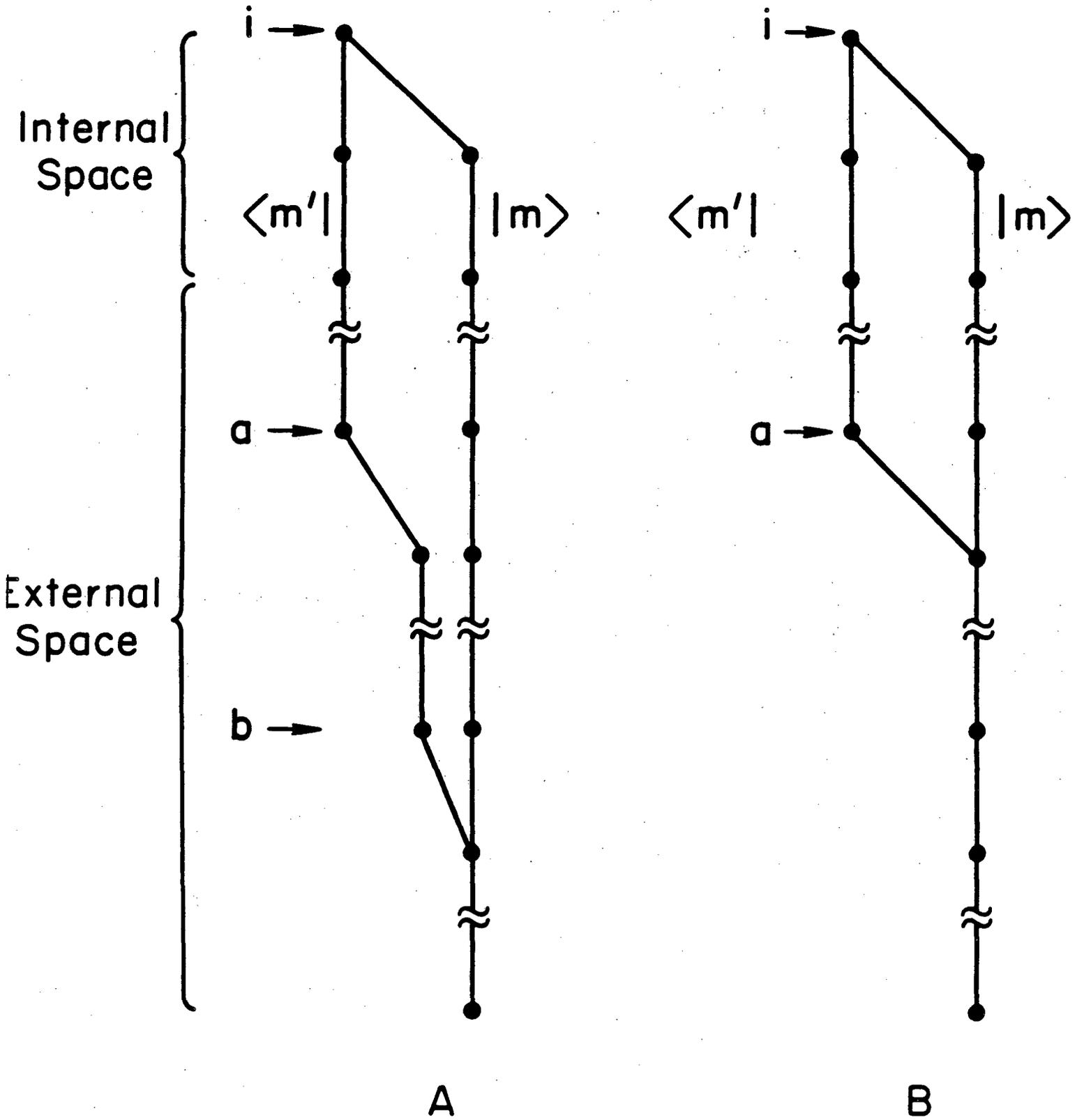


Figure 1

This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

TECHNICAL INFORMATION DEPARTMENT  
LAWRENCE BERKELEY LABORATORY  
UNIVERSITY OF CALIFORNIA  
BERKELEY, CALIFORNIA 94720