JOURNAL
OF THE AMERICAN CHEMICAL SOCIETY

Registered in U. S. Patent Office. *© Copyright 1973 by the American Chemical Society*

VOLUME 95, NUMBER 7 APRIL 4, 1973

Simulated *ab Initio* Molecular Orbital Technique. I. Method[1]

James E. Eilers and Donald R. Whitman*

Contribution from the Chemistry Department, Case Western Reserve University, Cleveland, Ohio 44106. *Received September 7, 1972*

Abstract: A new procedure for closely approximating the results of complete *ab initio* molecular orbital calculations is presented, evaluated, and applied to a few molecules. This simulated *ab initio* molecular orbital (SAMO) method consists in transferring *ab initio* Hartree–Fock matrix elements from calculations on pattern molecules, thereby avoiding repeated lengthy calculations of two-electron repulsion integrals. The matrix elements involve conventional hybridized and unhybridized atomic orbitals and are ultimately basis independent. The SAMO applications yield MO eigenspectra and eigenfunctions which reproduce complete results with unprecedented accuracy.

In the description of electronic configurations of relatively small molecules, *ab initio* molecular orbital calculations have proved quite useful. This is because the MO method provides an intuitively convenient model, its results are generally good for expectation values of one-electron operators, and the computational procedures involved are not prohibitively lengthy or complex.

However, *ab initio* calculations for larger molecules are impractical, owing primarily to the large number of time-consuming electron repulsion integrals which must be evaluated. Despite continuing increases in computer speeds and storage capabilities, there is little prospect that successful *ab initio* calculations will soon become available for more than a scattered few of the larger molecules of chemical interest.

The *ab initio* method consists in considering each molecule as a collection of charged nuclei in fixed positions and electrons in molecular orbitals leading to minimum energy. Each molecule subjected to an *ab initio* calculation is treated as an entirely new case, even though it contains molecular fragments which also occur in many other molecules. The *ab initio* procedure ignores the fundamental fact that the physical and chemical properties of a functional group are largely independent of the particular molecule in which the functional group occurs. The spectral properties and chemical reactivities of methylene groups, for example, are quite similar in many different molecules. This fact suggests the possibility of developing a procedure in which MO calculations for large molecules could utilize information obtained from complete *ab initio* calculations on smaller molecules containing the same functional groups.

This concept was implicitly used in the work of Fitts and Orloff,[2] who showed that the diagonal matrix elements of the Hückel MO method could be obtained from *ab initio* calculations on small molecules as well as from spectroscopy. Subsequently, Lipscomb and coworkers[3] extended the idea to both σ and π electrons in their nonempirical MO approach. Moreover, certain aspects of their work can be interpreted as hinting at the transferability of off-diagonal matrix elements as well. Concurrently, the atoms in molecules,[4] localized orbitals,[5] and molecules in molecules[6] techniques applied transference of molecular orbitals themselves. Christofferson, Genson, and Maggiora[7] have optimized free-floating spherical Gaussian bases for fragments in small molecule *ab initio* calculations and then used these

---

[1] Abstracted in part from the Ph.D. thesis of J. E. E., Case Western Reserve University, 1971.

[2] M. K. Orloff and D. D. Fitts, *J. Amer. Chem. Soc.*, **85**, 3721 (1963).

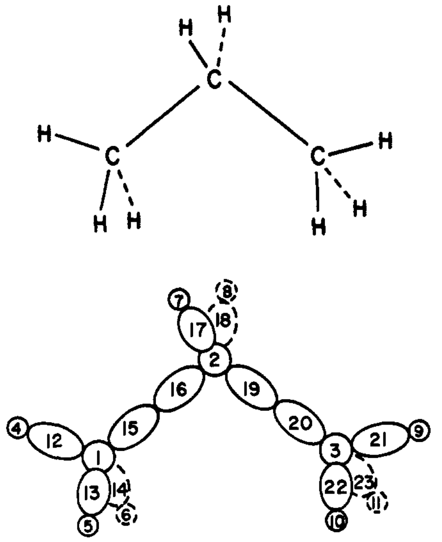
[3] M. D. Newton, F. P. Boer, and W. N. Lipscomb, *ibid.*, **88**, 2353, 2361, 2367 (1966); W. E. Palke and W. N. Lipscomb, *ibid.*, **88**, 2384 (1966); D. B. Boyd and W. N. Lipscomb, *J. Chem. Phys.*, **48**, 4955 (1968).

[4] W. Moffitt, *Proc. Roy. Soc., Ser. A*, **210**, 224, 245 (1951); **218**, 486 (1953); W. Moffitt and J. Scanlan, *ibid.*, **218**, 464 (1953); A. C. Hurley, *Proc. Phys. Soc., London, Sect. A*, **68**, 149 (1955); **69**, 49, 301, 767 (1956).

[5] S. F. Boys, *Rev. Mod. Phys.*, **32**, 296 (1960); C. Edmiston and K. Rudenberg, *J. Chem. Phys.*, **35**, 457 (1963); W. H. Adams, *ibid.*, **42**, 4030 (1965); R. M. Rutledge and A. F. Saturno, *ibid.*, **43**, 597 (1965).

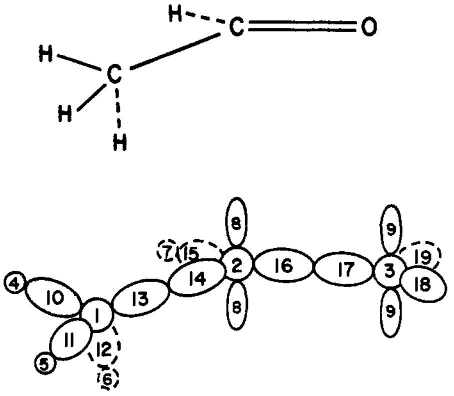
[6] W. Von Niessen, *ibid.*, **55**, 1948 (1971).

[7] R. E. Christofferson, *J. Amer. Chem. Soc.*, **93**, 4014 (1971); R. E. Christofferson, D. W. Genson, and G. M. Maggiora, *J. Chem. Phys.*, **59**, 239 (1971).



PROPANE

Figure 1. Hybridized orbital basis for propane.



ACETALDEHYDE

Figure 2. Hybridized orbital basis for acetaldehyde.

bases for conventional *ab initio* calculations in larger molecules.

This paper presents direct evidence of the transferability of *ab initio* Hartree–Fock matrix elements and estimates quantitatively the reliability of results from use of transferred elements. A simulated *ab initio* molecular orbital method is developed, with application to a few molecules. The SAMO results are shown to reproduce complete *ab initio* results with unprecedented accuracy. Finally a discussion describes the rather significant advantages of this new method.

Procedure

The Roothaan expansion technique produces molecular orbitals expressed as linear combinations of functions that have been chosen to form a basis for this expansion. Determination of the expansion coefficients for any functional basis involves the evaluation of the matrix elements of the Hartree–Fock operator with the functions of that basis. Slater-type orbitals have the advantage of efficiency in such an expansion, yet Gaussian-type functions are being used increasingly because of a substantial savings in the time required for evaluation of the electron repulsion integrals.

All of the present *ab initio* molecular orbital calculations were performed with bases of atomic optimized Gaussian-type functions,[8] which have proved useful in other molecular calculations.[9,10] The 5s 3p sets were used for both the carbon and oxygen atoms, while a 2s set optimized in an *ab initio* methane calculation was used for each hydrogen atom.

However, neither the Slater nor the Gaussian functions, with spherical harmonic angular dependence, are directly useful when considering the intermolecular transfer of Hartree–Fock matrix elements. For example, conventional p_x, p_y, and p_z functions centered on different atoms yield matrix elements critically dependent upon the precise molecular configuration and choice of coordinate system. An alternate and conceptually useful basis consists of conventional hybrid orbitals centered on each heavy atom. Such a basis facilitates accurate comparisons of corresponding matrix elements from different molecules, since these functions have the same relative orientations as simple ball-and-stick molecular models. Rothenberg[11] has, in fact, found that localized C–H bond orbitals are practically identical among some hydrocarbons and that these localized orbitals are essentially sp^3 hybrids.

All the present calculations begin with 5s 3p sets for the heavy atoms and a 2s set for each H atom. Subsequently, for each heavy atom the five s functions are contracted to a single 1s and a single 2s function; and each set of three p functions is contracted to a single p_x, p_y, or p_z atomic orbital. These 2s and 2p functions are then linearly combined to the appropriate sp^3 or sp^2 hybrid orbitals on each carbon or oxygen atom. At each H atom the two s functions are contracted to a single 1s function. The orbital exponents and contraction coefficients for these functions are given in Table I, while Figures 1 and 2 are schematic representations of the full hybridized bases for the typical molecules propane and acetaldehyde. In acetaldehyde, for example, the functions 1–7 are contracted 1s functions, 8 and 9 are unhybridized p functions, 10–13 are sp^3 hybrids, and 14–19 are sp^2 hybrids. This total basis of 19 functions produces a 19×19 Hartree–Fock matrix, each element of which is calculated for the Hartree–Fock operator with two of these standard orbitals.

Even quite similar molecules have small individual variations in geometry. However, to maintain full compatibility with the ideally hybridized bases, standard molecular geometries are used throughout. Bond angles are exactly tetrahedral or trigonal, and the bond lengths are constant at the standard values in Table II. Calculations of barriers to methyl group rotations[10] have demonstrated the quantitative as well as qualitative usefulness of this procedure.

All the *ab initio* molecular orbital calculations reported in this work were performed with a modified version of the IBMOL program of Davis and Clementi.[12]

---

(8) C. J. Hornback, Ph.D. Thesis, Case Institute of Technology, 1967; D. R. Whitman and C. J. Hornback, *J. Chem. Phys.*, **51**, 398 (1969).

(9) For example, H. Basch, *ibid.*, **56**, 441 (1972); J. R. Sabin and H. Kim, *ibid.*, **56**, 2195 (1972); J. B. Robert, H. Marsmann, I. Absar, and J. R. Van Wazer, *J. Amer. Chem. Soc.*, **93**, 3320 (1971).

(10) A. Liberles, B. O'Leary, J. E. Eilers, and D. R. Whitman, *ibid.*, **94**, 6894 (1972).

(11) S. Rothenberg, *J. Chem. Phys.*, **51**, 3389 (1969).

(12) D. R. Davis and E. Clementi, *ibid.*, **45**, 2593 (1966).

Table I. Basis Orbitals and Atomic Contraction Coefficients

Type	Exponents	Contraction coefficients $n = 1$	$n = 2$
	Carbon (−37.5791 hartrees)		
s	381.0000	0.023540	−0.005039
	59.1800	0.152024	−0.031347
	13.7900	0.464741	−0.128739
	3.8330	0.495882	−0.165344
	0.2946	0.018942	1.041865
p	4.2790		0.109692
	0.8700		0.460048
	0.2036		0.629634
	Oxygen (−74.5514 hartrees)		
s	736.6000	0.021637	−0.004958
	112.9000	0.144638	−0.031648
	26.0400	0.459336	−0.013746
	7.2120	0.507084	−0.171225
	0.5764	0.220000	1.043807
p	8.3560		0.119537
	1.7190		0.470553
	0.3814		0.622431
	Hydrogen (−0.48275 hartrees)		
s	1.6670	0.223484	
	0.2587	0.856922	

Table II. Standard Bond Lengths[a]

Bonded atoms	Bonding	Bond lengths, Å
Carbon–carbon	sp^3–sp^3	1.540
	sp^3–sp^2	1.520
	sp^3–sp^2 (carbonyl)	1.460
	Double bond	1.340
Carbon–oxygen	Double bond	1.230
	Single bond (carboxylic)	1.360
	Single bond (alcohol)	1.430
Carbon–hydrogen	sp^3–s	1.100
	sp^2–s	1.070
	sp^2–s (aldehyde)	1.115
Oxygen–hydrogen	sp^3–s (carboxylic)	0.960

[a] L. E. Sutton, Ed., "Tables of Interatomic Distances and Configurations in Molecules and Ions," The Chemical Society, London, 1958.

Transferability

By using standard hybridized functions as bases for MO expansions in similar molecules, corresponding Hartree–Fock matrix elements in these molecules can be compared directly. Thus, for example, the 12–15 matrix element of the methyl group of propane, Figure 1, corresponds with the 10–13 element of acetaldehyde, Figure 2. If transference of Hartree–Fock matrix information is to prove feasible, then similarities must exist between such corresponding elements.

To test for similarities of corresponding matrix elements in different molecules, complete *ab initio* calculations were performed for ethane, propane, and butane, using the standard hybridized bases and standardized geometries. A few typical converged Hartree–Fock matrix elements for these three molecules are compared in Table III. Each of the two orbitals involved in a matrix element is identified by the symbol of the atom at which it is centered, and a subscript indicating the type of orbital. s or p indicates an unhybridized orbital, 4H or 4C indicates a tetrahedral sp^3 orbital directed toward a hydrogen or a carbon atom, respectively, and 3H or 3C indicates a trigonally hybridized orbital.

Table III. Typical Equivalent Fock Matrix Elements for Alkanes (hartrees)

Propane element (Figure 1)	Orbital types	Ethane	Propane	Butane
	Methyl Intrafragment			
1–1	C_s–C_s	−11.353	−11.348	−11.347
1–4	C_s–H_s	−0.9074	−0.9071	−0.9070
1–12	C_s–C_{4H}	−0.0108	−0.0110	−0.0110
15–15	C_{4C}–C_{4C}	−0.6528	−0.6561	−0.6550
12–15	C_{4H}–C_{4C}	−0.1117	−0.1085	−0.1085
	Methylene Intrafragment			
2–2	C_s–C_s		−11.365	−11.360
2–7	C_s–H_s		−0.9081	−0.9078
2–17	C_s–C_{4H}		−0.0101	−0.0103
19–19	C_{4C}–C_{4C}		−0.6681	−0.6715
17–19	C_{4H}–C_{4C}		−0.1074	−0.1064
	Neighboring Fragments			
1–2	C_s–C'_s	−0.0121	−0.0121	−0.0121
1–7	C_s–C'_{4H}	−0.0461	−0.0461	−0.0461
1–17	C_s–C'_{4H}	0.0644	0.0645	0.0645
15–16	C_{4C}–C'_{4H}	−0.7256	−0.7256	−0.7329
15–17	C_{4C}–C'_{4H}	−0.1695	−0.1666	−0.1680
	Second-Neighbor Fragments			
1–3	C_s–C''_s		−0.0003	−0.0003
1–9	C_s–H_s''		−0.0046	−0.0046
1–21	C_s–C''_{4H}		−0.p557	−0.0557
15–20	C_{4C}–C''_{4C}		−0.3267	−0.3267
15–21	C_{4C}–C''_{4H}		−0.0458	−0.0459

Significant conclusions can be drawn from the information condensed in Table III. The matrix elements represented in the table are strikingly similar for these three molecules. This is particularly remarkable since the number of nuclear attraction integrals and two-electron repulsion integrals vary widely among the molecules. Ethane, for example, has 9 occupied molecular orbitals while butane has 17, yet most of the corresponding Fock matrix elements are nearly identical. There is a small, but distinct, difference between the C_s–C_s elements of the methyl groups and those of the methylene groups. This difference, of about 0.01 hartree or 0.3 eV, is of the magnitude expected for the difference of carbon 1s electron energies. Generally, however, corresponding elements from methyl and methylene groups are quite similar in magnitude, again confirming an intuitive expectation. Also, most matrix elements are observed to decrease in magnitude with increasing separation of the functions involved. However, this is not always the case, as with element 15–17 which is larger in magnitude than the elements 12–15 or 17–19. Finally, it should be noticed that at larger separations some elements are positive, indicating net repulsions.

The basic premises of the transference concept are further substantiated by the results of complete *ab initio* calculations for ethene and propene, and acetaldehyde and acetic acid. Representative functional group matrix elements for these molecules are presented in Table IV. Here again corresponding elements have quite similar values. However, distinct differences do occur (between the two carbonyls, in particular). Again this is not unexpected, since the chemical and physical properties of the carbonyl and carboxyl groups are quite distinguishable.

Another conclusion of matrix element comparisons is not revealed in the data of Tables III and IV. Upon

Table IV. Typical Equivalent Fock Matrix Elements for Alkenes and Carbonyls (hartrees)

Methylene element	Orbital types	Ethene	Propene
	C_s–C_s	−11.383	−11.364
	C_s–H_s	−0.9640	−0.9626
	C_s–C_{3H}	−0.0100	−0.0105
	C_{3C}–C_{3C}	−0.8901	−0.8840
	C_{3H}–C_{3C}	−0.1088	−0.1168
	C_p–C'_p	−0.2380	−0.2370
Acetaldehyde element (Figure 2)	**Orbital types**	**Acetaldehyde**	**Acetic acid**
3–3	O_s–O_s	−20.615	−20.569
17–17	O_{3C}–O_{3C}	−1.0877	−1.0767
2–2	C_s–C_s	−11.498	−11.567
14–14	C_{3C}–C_{3C}	−0.7816	−0.8275
16–16	C_{3O}–C_{3O}	−1.0502	−1.0893
8–9	C_p–O'_p	−0.3871	−0.3783

partitioning matrix elements into one- and two-electron components, it is found that neither of these components is invariant between corresponding elements in similar molecules. Thus, transference of the *ab initio* information requires transference of the full Fock matrix elements rather than of individual components. Compensatory variations of the one- and two-electron terms occur between similar molecules.

The differences between corresponding Hartree–Fock matrix elements in comparable molecules are observed rarely to be as large as 0.01 hartree, with most differences smaller than 0.005 hartree. It would be quite useful to obtain a quantitative estimate of the effect of matrix element deviations of this order of magnitude upon calculated eigenvalues. Such an estimate can be obtained by a statistical treatment of perturbation results.

The Roothaan[13] formulation of the matrix Hartree–Fock equations is

$$\mathbf{F}\mathbf{C}_i = \epsilon_i \mathbf{S}\mathbf{C}_i$$

where F is the matrix of the Hartree–Fock operator in the chosen basis, S is the overlap matrix of the functions of that basis, and ϵ_i and $\mathbf{C}_i$ are the eigenvalue and expansion coefficients of the *i*th molecular orbital.

Perturbation theory shows that the effect of a perturbing matrix of small elements **f** upon $\mathbf{C}_i$ is

$$\Delta\epsilon_i = \mathbf{C}_i^{+}\mathbf{f}\mathbf{C}_i$$

In the case of transference of matrix elements from one molecule to another, the **f** elements are the deviations of the transferred elements from the true ones which would have resulted from a precise *ab initio* calculation. While the values of these deviations are of course unknown in any real application, there will be a large number of them, approximately randomly distributed and of order 0.005 hartree. Applying the general variance law[14] to the perturbation expression for $\Delta\epsilon_i$

$$\sigma_{\Delta\epsilon}{}^2 = \sigma_f{}^2\sum_{j,k}(\mathbf{C}_{ij}\mathbf{C}_{ik})^2 + \sigma_c{}^2\sum_j(\sum_k \mathbf{f}_{jk}\mathbf{C}_{ik})^2 \leq \sigma_f{}^2\sum_j \mathbf{C}_{ij}{}^2\sum_k \mathbf{C}_{ik}{}^2 + \sigma_c{}^2\sum_k \mathbf{C}_{ik}{}^2\sum_{j,k}\mathbf{f}_{jk}{}^2$$

where the Schwartz inequality has been invoked. The magnitude of the right-hand side of this expression can be approximated by reasonable but conservative estimates of the various quantities. The deviation, $\mathbf{f}_{jk}$, of the transferred Hartree–Fock matrix elements from the true *ab initio* values are assumed randomly distributed, with mean zero and standard deviation σ_f = 0.005 hartree. The coefficients of the *i*th normalized eigenvector can be assumed to have a mean of 0.1, and standard deviation $\sigma_c = 0.1$. Then, with the assumption that $\Sigma_k \mathbf{C}_{ik}{}^2 \approx 1$, and recognizing that $\Sigma_{j,k}\mathbf{f}_{jk}{}^2 = n^2\sigma_f{}^2$ since there are n^2 elements

$$\sigma_{\Delta\epsilon}{}^2 \leq \sigma_f{}^2(1 + n^2\sigma_c{}^2)$$

For the propane calculation as an example, $n = 23$, and consequently the expected variance of MO energies as a result of Hartree–Fock matrix element deviations is $\sigma_{\Delta\epsilon}{}^2 \leq 157 \times 10^{-6}$. Thus, $\sigma_{\Delta\epsilon} \approx 0.013$ hartree, and two-thirds of calculated orbital energies can be expected to differ from precise *ab initio* values by no more than this amount. For larger molecules n^2 of course increases, but $\sigma_c{}^2$ decreases simultaneously, and it is reasonable to conclude that $\sigma_{\Delta\epsilon}{}^2$ changes little.

While this statistical treatment involves assumptions, it permits estimating the order of magnitude of effects to be expected from the matrix element approximations invoked in this (or any other) MO technique for large molecules. Subsequent experience with actual calculations has shown that the estimate of 0.013 hartree for $\sigma_{\Delta\epsilon}$ is in fact conservative, with orbital energy deviations rarely this large.

Simulated *ab Initio* Molecular Orbitals

As a consequence of the demonstrated transferability of corresponding Hartree–Fock matrix elements between molecules containing the same molecular fragments, the intriguing possibility emerges of constructing Hartree–Fock matrices for larger molecules entirely by transference of the appropriate information from smaller molecule *ab initio* calculations.

The simulated *ab initio* molecular orbital, SAMO, method involves using converged Hartree–Fock matrix elements from complete *ab initio* calculations on pattern molecules to construct a simulated *ab initio* Hartree–Fock matrix for a larger molecule. This simulated matrix is combined with the true overlap matrix for the larger molecule to obtain the Roothaan equations. These equations are solved by standard matrix techniques for the SAMO eigenvalues ϵ_i and eigenvectors $\mathbf{C}_i$. Since the matrix elements from the pattern molecule calculations are already converged Hartree–Fock matrix elements, they simulate the converged values in the larger molecule, and the Roothaan equations are solved only once, without the necessity of iteration to convergence.

The SAMO method has been tested by application to the alkane series of ethane, propane, and butane. A complete *ab initio* calculation with the standard hybridized bases was performed for each of these three molecules; these calculations serve not only as the source of *ab initio* matrix elements for use in SAMO calculations, but also as the reference results for testing the validity of the SAMO method. As an initial application, the *ab initio* matrix elements of the propane calculation were used to construct a simulated *ab initio* Hartree–Fock matrix for the ethane molecule.

---

(13) C. C. J. Roothaan, *Rev. Mod. Phys.*, 23, 69 (1951).

(14) T. B. Crumpler and J. H. Yoe, "Chemical Computations and Errors," Wiley, New York, N. Y., 1940, Chapter 10.

This is of course working in the "wrong" direction—from a larger to a smaller molecule—but this case is important as a test of the basic method, since every matrix element required in the ethane calculation is directly available from propane, while the converse is not true. Once the simulated Hartree–Fock matrix for ethane was constructed, the true ethane overlap matrix was calculated, and the Roothaan equations were solved.

The results of this application to ethane are compared in Table V with the ethane results from complete *ab initio* calculations with the same basis.

Table V. Comparison of Complete *ab Initio* and SAMO Results for Ethane Molecular Orbital Energies

	Orbital energies, hartrees	
Orbital	Complete *ab initio*	SAMO
a_{1g}	−11.3539	−11.3509
a_{2u}	−11.3539	−11.3509
a_{1g}	−1.0619	−1.0632
a_{2u}	−0.8713	−0.8721
e_u	−0.6240	−0.6248
a_{1g}	−0.5387	−0.5393
e_g	−0.5081	−0.5070
Total energy	−78.8196	−78.8049

The eigenspectrum of orbital energies is reproduced well by the SAMO calculation, with entirely correct ordering. The largest deviation of an orbital energy is 0.003 hartree, or 0.08 eV, which is small compared with the fundamental errors inherent in any molecular orbital method. Thus, the SAMO ethane calculation yields results only insignificantly different from the complete *ab initio* results, yet with a substantial reduction in computing time required.

The total molecular energy for both the complete *ab initio* and SAMO methods is the sum of the orbital energies and the one-electron energies.[13] Since ideally hybridized bases and standard geometries are used throughout, one-electron integrals and overlap integrals are identical, whether recalculated or transferred between molecules. For such simple integrals recalculation for each molecule is actually easier, since it avoids the manipulations involved in transference. The total molecular energy of the SAMO calculation for ethane is in excellent agreement with the complete *ab initio* result.

A test of the SAMO method was made on propane, using ethane *ab initio* matrix elements. In this case certain elements occur in propane which are not found in ethane, since ethane contains no methylene group. The necessary methylene elements can be reasonably approximated by appropriate ethane methyl group elements, but clearly propane elements between nonneighboring carbon atoms are unavailable from ethane and must be assumed negligible.

Comparison of the propane results from complete *ab initio* calculation and from the SAMO method using ethane matrix elements is presented in Table VI.

The results of the SAMO propane calculation are distinctly poorer than those for ethane, with orbital energy deviations as large as 0.09 hartree, and with two orbital energies in misplaced order. This result, however, is in a sense satisfying since it confirms the sensitivity of the SAMO method to the precision of the matrix elements transferred. In this case some elements transferred were of questionable applicability, and all elements between non-nearest-neighbor groups were necessarily assumed to be zero. It is legitimate to conclude that if matrix elements of mediocre quality are used in the SAMO method, then mediocre results are obtained.

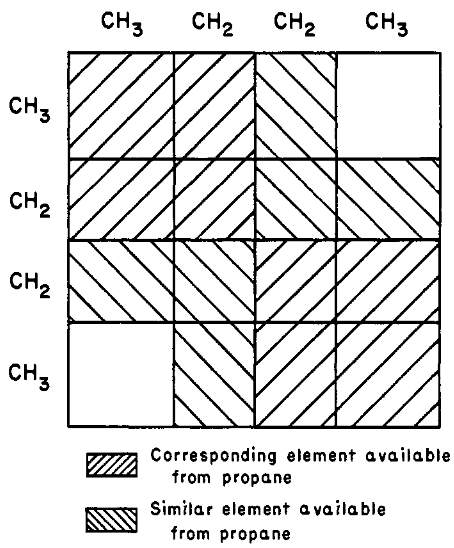


Figure 3. Schematic Hartree–Fock matrix for butane.

Table VI. Comparison of Complete *ab Initio* and SAMO Calculations of Propane Molecular Orbital Energies

	Orbital energies, hartrees	
Orbital	Complete *ab initio*	SAMO
a_1	−11.3659	−11.3515
b_1	−11.3492	−11.3518
a_1	−11.3491	−11.3509
a_1	−1.1000	−1.0909
b_1	−0.9606	−0.9608
a_1	−0.8306	−0.8263
b_2	−0.6530	−0.6263
a_1	−0.6248	−0.6142
b_1	−0.5771	−0.6602[a]
a_2	−0.5542	−0.6056
a_1	−0.5047	−0.4554[a]
b_1	−0.5011	−0.4929
b_2	−0.4917	−0.4648
Total energy	−117.6776	−117.6275

[a] Note incorrect energy ordering.

The critical test of the SAMO method is application to the butane molecule. Here all matrix elements through next-nearest-neighbor groups are accessible from propane, although some methylene terms must be approximated by methyl terms. Only the longest distance, methyl–methyl terms are assumed negligible. Figure 3 depicts schematically the source of Hartree–Fock matrix elements for the butane calculation. Each block represents a number of matrix elements, of course, since there are eight functions of the basis associated with each methyl group, and seven with each methylene. The complete *ab initio* and SAMO results for butane are compared in Table VII.

The butane results show excellent agreement between the complete *ab initio* and SAMO energies. The largest

Table VII. Comparisons of Complete *ab Initio* and SAMO Results for *n*-Butane Molecular Orbital Energies

Orbital	Orbital energies: Complete *ab initio*	Orbital energies: SAMO
a_g	−11.3613	−11.3658
b_u	−11.3613	−11.3660
b_u	−11.3478	−11.3489
a_g	−11.3479	−11.3489
a_g	−1.1184	−1.1190
b_u	−1.0202	−1.0205
a_g	−0.8916	−0.8924
b_u	−0.8194	−0.8178
b_g	−0.6675	−0.6653
a_g	−0.6135	−0.6151
b_u	−0.6026	−0.6053
a_u	−0.5930	−0.5981
b_u	−0.5391	−0.5375
b_g	−0.5170	−0.5185
a_g	−0.5015	−0.5055
a_u	−0.4864	−0.4849
a_g	−0.4839	−0.4768
Total energy	−156.5353	−156.5479

orbital energy deviation is only 0.0071 hartree, well below the estimated standard deviation of 0.013 hartree found above. Furthermore, there are no inversions of energy levels as occurred with propane. The striking validity of these SAMO results for butane is attributable to the use of *ab initio* results from a pattern molecule (propane) which contains each of the two different functional groups (methyl and methylene) found in butane. In addition, each functional group in propane has the same heavy atom nearest neighbors as in butane, and it is not necessary to neglect next-nearest-neighbor terms. Additional evidence of the importance of pattern molecules with three adjoining correct heavy atoms occurs below.

Table VIII. Comparisons of Complete *ab Initio* and SAMO Eigenfunction Coefficients for *n*-Butane

Basis orbital type	$1a_g$ Complete	$1a_g$ SAMO	$6b_u$ Complete	$6b_u$ SAMO	$12a_u$ Complete	$12a_u$ SAMO	$17a_g$ Complete	$17a_g$ SAMO
				Methyl				
C_s	0.0017	0.0072	−0.0370	−0.0372	0	0	−0.0022	−0.0021
C_{4C}	0.0043	0.0044	0.2241	0.2245	0	0	0.2238	0.2590
C_{4H}	0.0007	0.0008	0.1920	0.1893	0	0	−0.2852	−0.2737
C_{4H},C_{4H}	0.0008	0.0008	0.1850	0.1838	±0.2696	±0.2667	0.0660	0.0561
H_1	0.0002	0.0002	0.0925	0.0937	0	0	−0.2740	−0.2800
H_2,H_3	−0.0001	−0.0001	0.0905	0.0921	±0.2069	±0.2067	0.0640	0.0474
				Methylene				
C_s	−0.7078	−0.7078	−0.0211	−0.0215	0	0	0.0014	0.0021
C_{4C}	−0.0043	−0.0044	0.1944	0.1807	0	0	0.1934	0.2361
C_{4C}	−0.0008	−0.0008	0.0439	0.0441	0	0	−0.4146	−0.3984
C_{4H},C_{4H}	−0.0042	−0.0043	0.1002	0.1001	±0.1562	±0.1571	0.0636	0.0415
H_4,H_5	0.0042	0.0043	0.0483	0.0507	±0.1129	±0.1188	0.0714	0.0438

Comparisons of complete and SAMO eigenfunction coefficients for four representative *n*-butane molecular orbitals are presented in Table VIII. The unique basis orbitals are identified by the symbolism introduced for Table III, while coefficients of the other basis orbitals are apparent from symmetry. The lowest and highest energy occupied molecular orbitals are included, along with two intermediate orbitals of different symmetries. The correspondence of the complete and SAMO coefficients is excellent, with the most significant deviations occurring with the highest energy orbital, as would be expected. The large discrepancy in the methyl 1s coefficient for the lowest energy orbital arises from the small difference of large numbers in symmetry factoring, and is almost entirely compensated in subsequent orbitals.

The potential of the SAMO method for nonalkane molecules was tested by calculations on acetic acid. In addition to the complete *ab initio* calculation which served as reference, two different SAMO calculations for acetic acid were performed using complete *ab initio* matrix elements from the pattern molecules formic acid and acetaldehyde. The first acetic acid SAMO application consisted in using pattern molecule elements for all intrafragment and nearest-neighbor terms, with all non-nearest neighbor terms assumed zero. This assumption was due to the fact that the non-nearest-neighbor terms between the methyl and hydroxyl groups do not both occur in either pattern molecule. The second acetic acid SAMO calculation entailed improved values for these non-nearest-neighbor terms. The methyl–carboxyl oxygen terms came from acetaldehyde, and the carboxyl oxygen–hydroxyl terms from formic acid, while the methyl–hydroxyl terms were borrowed from the complete *ab initio* calculation on acetic acid. This latter step is justified only by the fact that no other estimate of these terms was available, and the presumption that had another source existed, the elements transferred would have been very near to the values actually used.

These two levels of SAMO calculation are compared in Table IX with the complete *ab initio* results. Although both SAMO calculations yield the correct ordering of energy levels, and reasonably good total energy estimates, the use of next-nearest-neighbor matrix elements in the level 2 calculation produces distinctly superior orbital energies particularly for the higher states. The SAMO level 1 result for the total energy of acetic acid is below the complete result, just as for butane in Table VII, in an apparent violation of the variation theorem. However, the variation theorem does not apply strictly to the SAMO procedure, since matrix elements are not calculated exactly with the correct Hamiltonian, but instead are closely approximated by transference. In the acetic acid level 1 calculation the

Table IX. Comparisons of Acetic Acid Orbital Energies from Complete *ab Initio* Calculations with SAMO Calculations Neglecting (Level 1) and Including (Level 2) Next-Nearest-Neighbor Matrix Elements

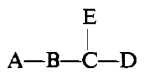
	Orbital energies, hartrees		
	Complete	SAMO	
Orbital	*ab initio*	Level 1	Level 2
1	−20.6062	−20.6119	−20.6110
2	−20.5699	−20.5927	−20.5910
3	−11.5676	−11.5704	−11.5705
4	−11.3912	−11.3809	−11.3809
5	−1.4766	−1.4774	−1.4828
6	−1.3616	−1.4493	−1.3679
7	−1.0551	−1.0765	−1.0439
8	−0.8209	−0.8304	−0.8191
9	−0.6971	−0.7286	−0.6918
10	−0.6709	−0.6662	−0.6758
11	−0.6656	−0.6546	−0.6629
12	−0.5908	−0.6421	−0.5863
13	−0.5775	−0.6308	−0.5748
14	−0.5404	−0.6005	−0.5401
15	−0.4632	−0.5986	−0.4667
16	−0.4566	−0.4699	−0.4609
Total energy	−226.8398	−227.3163	−226.7862

inherent deviation from the complete result is accentuated by the neglect of the non-nearest-neighbor matrix elements, many of which are positive and consequently largely repulsive.

Discussion

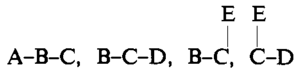
The SAMO procedure has considerable advantages for molecular orbital calculations on large molecules. Once an adequate library of transferable matrix elements is available, the actual computation time for a SAMO calculation is only a small fraction of the time required for a corresponding complete *ab initio* calculation on the same molecule. The SAMO eigenvalue spectra and orbital expansion coefficients are in excellent agreement with the complete *ab initio* values. The procedure is applicable to any molecule for which appropriate pattern molecule calculations are available, or can be made. Finally, the Fock matrix elements which are transferred from pattern molecule calculations are ultimately entirely basis independent. If workers use Slater-type orbitals or other larger Gaussian bases for expansion of molecular orbitals, then upon transformation to the hybridized orbital bases they will obtain only slightly different (and presumably better) approximations to the true Hartree–Fock matrix elements than those of Tables III and IV. Thus, it is entirely feasible to build cooperatively an extensive catalog of Hartree–Fock matrix elements from pattern molecule calculations done at different laboratories with different bases. Implementation of a SAMO calculation for a molecule would then entail only looking up the required diagonal and off-diagonal matrix elements in this catalog, calculating the required overlap matrix elements, and solving the secular equations.

Comparisons of matrix elements, as in Tables III and IV, and examination of SAMO results indicate that completely satisfactory transference of matrix elements occurs only if each of the functional groups of the target molecule is contained in one of the pattern molecules with the same heavy atom nearest neighbors. Thus, for the hypothetical molecule



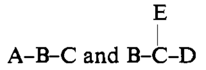
$$A\text{—}B\text{—}\overset{\overset{E}{|}}{C}\text{—}D$$

where each of the five letters represents a different group, totally satisfactory SAMO results would require complete *ab initio* information from the patterns



$$A\text{–}B\text{–}C,\ B\text{–}C\text{–}D,\ B\text{–}\overset{\overset{E}{|}}{C},\ \overset{\overset{E}{|}}{C}\text{–}D$$

Alternatively, still more relable information would be found from



$$A\text{–}B\text{–}C \text{ and } B\text{–}\overset{\overset{E}{|}}{C}\text{–}D$$

Once these patterns are completed, however, a number of additional simlar molecules become accessible with SAMO, with little additional computation.

Each of the calculations reported here has been done for a standardized molecular geometry, with ideally hybridized orbitals. Actual molecules rarely conform precisely to the ideal, but a few tests have shown that small deviations from idealized geometry lead only to small changes in SAMO results. Ultimately it is probable that a perturbation technique can be used to estimate changes in two-electron contributions to Hartree–Fock matrix elements upon geometric distortions. The changes in one-electron and nuclear repulsion terms are of course easily and quickly calculable precisely.

Application of the SAMO method to quite large molecules is immediately possible, and results with condensed-ring systems are the subject of a subsequent paper. However, a significant problem arises in application to larger molecules, and that is the one of efficiently processing the matrix elements being transferred. For an anthracene calculation, as an example, the total basis consists of 80 standard hybridized functions, yielding a total of 6400 Hartree–Fock matrix elements. The basic symmetry reduces the number of distinct elements, and each of these elements can be located from an appropriate butadiene pattern calculation.

Nonetheless, the amount of labor involved in accurately finding and inserting thousands of elements is considerable, and the chances for error are many. Consequently, a reliable system for computer storage of pattern molecule data and subsequent retrieval and deployment is essential for work with very large molecules. Some preliminary steps have been taken in this direction, but the problem is by no means solved.

Conclusions

The SAMO procedure reproduces complete *ab initio* results to well within the limits of validity of the entire molecular orbital model. It permits large savings in computation time and data storage space and is ultimately basis independent. In addition, procedures and results of the method are expressed in language of conventional hybridized orbitals and thus assure maximum physical interpretability. It appears that the SAMO method should be extensively developed, in order to permit reliable MO calculations for large molecules.