Theor Chem Acc (1997) 97:10–13

Theoretical
Chemistry Accounts

Methods for rapid evaluation of electron repulsion integrals in large-scale LCGO calculations[*],[**]

Jan Erik Almlöf[1]

Institute of Theoretical Physics, University of Stockholm, Stockholm, Sweden

Abstract. A method is proposed for efficient use of molecular symmetry in the evaluation of two-electron integrals. This provides a means of avoiding the recalculation of symmetry-redundant integrals, and of symmetry-blocking matrices and supermatrices without the usual time-consuming transformation procedures. Various methods for speeding up the calculation of integrals are also discussed. Integral calculation times are given for some representative molecules.

The major problem in large nonempirical MO-LCAO calculations is the evaluation and storing of electron-repulsion integrals. The number of two-electron integrals will be roughly equal to $\frac{1}{8}N^4$, where N denotes the number of basis functions. Thus in a calculation involving 300 basis functions more than 10^9 integrals have to be evaluated and stored. Since a maximum of 5×10^6 integrals can be stored on one magnetic tape, around 200 tapes would be necessary to store these integrals. Also, their evaluation would require some 250 h on a medium-sized computer. The number of values to be stored can be substantially reduced by contracting the basis set, so reducing the number of degrees of freedom and yielding a smaller effective basis. This does not decrease the number of integrals to be evaluated, however, and calculations greater than 300–400 uncontracted or 150–200 contracted basis functions are still quite unfeasible from a computational point of view. The present paper suggests methods for making large-scale calculations more practicable, with special regard to computer time requirements and to the capacity needed for the storage of two-electron integrals.

Molecular symmetry is known to be a powerful tool for simplifying molecular orbital calculations. Probably the most well-known advantage of using symmetry is the blocking of various matrices (overlap matrix, first-order density matrix, Fock matrix, etc.) when a symmetry-adapted basis is used. This will speed up such operations as diagonalization, orthonormalization, etc., and decrease the core storage needed. However, of much greater importance in large-size calculations is the symmetry blocking of the supermatrix of two-electron integrals. In the case of a molecule with symmetry, the use of a symmetry-adapted basis will considerably reduce the number of two-electron integrals. This advantage can also be used in nonsymmetric molecules, provided that parts of the molecule are related by some local symmetry.

Caution must be exercised, however, in using a symmetry-adapted basis. A straight-forward approach would be first to construct a list of integrals over atomic orbitals, and then to transform them to a symmetry basis in a separate step. This method, involving the simultaneous handling of two large data sets, has been found to be extremely time-consuming with more than 100 basis functions. An entirely different approach is therefore suggested in this paper.

The molecular or local symmetry can also be used to relate atomic integrals which are equal by symmetry, thus avoiding the calculation of redundant integrals. The simultaneous use of symmetry in both these contexts would appear to be a nontrivial computational problem. The only computer program known to the author which combines both of these features is REFLECT [1]. In this program symmetry was used in a very limited way, taking only one two-fold symmetry element into account. However, the ideas used in REFLECT can be fur-

---

[*] This article is a translation of University of Stockholm Institute of Theoretical Physics Report 72–09, written in 1972. It is published for historical interest, but there are still many interesting and relevant things to read in the paper, both about integral evaluation and symmetry. It was prepared for publication by Peter Taylor, and we are pleased to include it in honor of the late Professor Almlöf because the article illustrates Professor Almlöf's insight at an early stage of his career. / The Editors

[**] In preparing this work for publication, I have taken the liberty of correcting a few errors and some missing text in the original manuscript. These amendments appear within [square brackets]. / Peter Taylor

[1] Deceased January 16, 1996

ther generalized [as follows]. A symmetry-adapted basis function $\phi_{\lambda a}$ belonging to the irreducible representation λ can be constructed from the atomic orbital χ_a and its symmetry-related equals according to the expression

$$\phi_{\lambda a} = \sum_i c_{i\lambda} g_i \chi_a \ . \tag{1}$$

Here $\{g_i\}$ is a set of symmetry operators transforming χ_a to all symmetry-related sites in the molecule. Only orbitals that are centred on symmetry-independent sites in the molecule should therefore be considered. The discussion will be restricted to groups with elements of second order, i.e., groups where $g_i^2 = e$ for all elements of the group, and consequently $c_\lambda = \pm 1$ apart from a common normalization factor. The sum over i [includes only] operations which are nontrivial with respect to the site of χ_a.

An electron repulsion integral over symmetry-adapted basis functions can now be constructed:

$$\begin{aligned}(\phi_{\lambda a}\phi_{\mu b}|\phi_{\nu c}\phi_{\xi d}) &= \sum_i \sum_j \sum_k \sum_l c_{i\lambda} c_{j\mu} c_{k\nu} c_{l\xi} \\ &\quad \times (g_i\chi_a \ g_j\chi_b | g_k\chi_c \ g_l\chi_d) \ . \end{aligned} \tag{2}$$

The atomic orbital integrals [on the right-hand side] of this expression are invariant under operations on the whole molecule by any of the members of the molecular point group. If such a transformation is carried out with the operator g_i^{-1}, the resulting expression is

$$\sum_i \sum_j \sum_k \sum_l c_{i\lambda} c_{j\mu} c_{k\nu} c_{l\xi} (\chi_a \ g_i^{-1} g_j \chi_b | g_i^{-1} g_k \chi_c \ g_i^{-1} g_l \chi_d) \tag{3}$$

Introducing

$$g_J = g_i^{-1} g_j, \quad g_K = g_i^{-1} g_k, \quad g_L = g_i^{-1} g_l \ , \tag{4}$$

the coefficients c will transform as

$$c_{j\mu} = c_{J\mu} c_{i\mu}, \quad c_{k\nu} = c_{K\nu} c_{i\nu}, \quad c_{l\xi} = c_{L\xi} c_{i\xi} \ , \tag{5}$$

and the total expression for the integral now becomes

$$\sum_i c_{i\lambda} c_{i\mu} c_{i\nu} c_{i\xi} \sum_J \sum_K \sum_L c_{J\mu} c_{K\nu} c_{L\xi} (\chi_a \ g_J \chi_b | g_K \chi_c \ g_L \chi_d) \ . \tag{6}$$

If the product of the four irreducible representations λ, μ, ν, and ξ is the totally symmetric representation, the product $c_{i\lambda} c_{i\mu} c_{i\nu} c_{i\xi}$ will equal unity regardless of the index i. The case in which the product representation is not totally symmetric is trivial, as this will render the whole integral zero by symmetry. In all cases of interest, therefore, the whole summation over i is redundant and may be replaced by a constant multiplicative factor.

This procedure of simplification can be further extended if the orbital χ_a is centred on a certain symmetry element, whereas any of the orbitals χ_b, χ_c, or χ_d [are not]. The resulting formulae provide a way of avoiding the recalculation of integrals related by the effective molecular symmetry.

Given the four atomic orbitals χ_a, χ_b, χ_c, and χ_d, the list of symmetry-adapted integrals $(\phi_{\lambda a}\phi_{\mu b}|\phi_{\nu c}\phi_{\xi d})$ for all allowed combinations of λ, μ, ν, and ξ must be kept in core until the loops over J, K, and L are completed. This requires a maximum of M^3 integrals to be handled and stored simultaneously, M being the maximum number of symmetry-equivalent orbitals that can be constructed from one atomic orbital by using the set of operators $\{g_i\}$.

If the integrals are to be used in a Hartree-Fock calculation only, the size of the list can be reduced further. The only integrals needed in this case are those of type $(\lambda\lambda|\mu\mu)$ and $(\lambda\mu|\lambda\mu)$, where λ and μ denote irreducible representations. This makes it possible to delete a considerable number of integrals from the list in calculations involving more than two irreducible representations.

The P, Q-supermatrix formalism [2] can be used to combine these integrals in a convenient way. For a closed-shell case, the expression for the Fock matrix element $F_{\lambda ab}$ becomes

$$F_{\lambda ab} = T_{\lambda ab} + V_{\lambda ab} + \sum_{\mu cd} D_{\mu cd} P_{\lambda ab, \mu cd} \ , \tag{7}$$

where

$$P_{\lambda ab, \mu cd} = (ab|cd) - \tfrac{1}{4}(ac|db) - \tfrac{1}{4}(ad|cb) \ . \tag{8}$$

In this way the information needed from these three integrals may be stored in one single number. The integral list is thus further shortened in all cases with more than one irreducible representation.

The use of the P, Q-supermatrix formalism also simplifies the construction of the Fock matrix considerably. In practice, the P and Q supermatrices can be obtained without any transformation or reordering procedure involving I/O by [computing] simultaneously the integrals $(ab|cd)$, $(ac|db)$, and $(ad|cb)$, and by arranging the loops so that recalculation of identical elements is avoided. This will complicate the calculation somewhat as the maximum number of elements to be stored simultaneously in core increases by a factor of 3. These procedures can reduce the list to a reasonable size even for quite large calculations. However, the computer time needed to evaluate the many billions of integrals used in a large calculation is still a troublesome factor. Several procedures for speeding up this part of the calculation have therefore been investigated.

Assuming a Gaussian-type basis, the expression for an electron repulsion integral over primitive Gaussians can be written formally as

$$(ij|kl) = S_{ij} S_{kl} \sum_N R_N(i,j,k,l) T_N(i,j,k,l) \ . \tag{9}$$

In this expression the S, R, and T factors are functions of the orbital exponents and nuclear coordinates. S_{ij} and S_{kl} are exponentials, and their product determines the order of magnitude of the integral, as R and T are quite slowly varying functions. A simple test can therefore be made on the arguments of the exponentials, whether it is necessary to calculate the integrals or not. It is found that, especially in large calculations, a majority of the

Table 1 Examples of integral calculation times for some representative molecules. The times are given in minutes and refer to a UNIVAC 1108 computer

Molecule	Basis set	Contracted basic set	Number of GTOs	Number of CGTOs	Number of irr reps	Integral time
Biphenyl $C_{12}H_{10}$	(C/7,3) (H/3)	(C/2,1) (H/1)	222	70	4	37.2[a] 47.6[a]
Bicyclopenta-dienyl $C_{10}H_{10}$	(C/7,3) (H/3)	(C/2,1) (H/1)	190	60	2 4	48.9 26.7
Cyclopenta-dienyl $C_5H_5^-$	(C/8,4) (H/4)	(C/2,1) (H/1)	120	30	2 4	12.4 14.6
Pyrrole C_4H_5N	(C,N/7,3) (H/3)	(C,N/4,2) (H/2)	95	60	4	15.7
Pyrrole C_4H_5N	(C,N/7,3) (H/3)	(C,N/2,1) (H/1)	95	30	4	8.2
Benzene C_6H_6	(C/8,4)	(C/2,1)	144	36	1 2 4 8	60.9 32.1 19.2 20.4

[a] The different values refer to a planar and nonplanar conformation, respectively

integrals need not be calculated at all as they are too small to contribute significantly to the final result.

This way of simplifying calculations leads to a considerable reduction of the integral evaluation time for large calculations. It is easily programmed, and it is employed in several MO-LCAO programs which are in current use [1, 3].

From a computational point of view, the factors $R_N(i,j,k,l)$ are quite complicated, involving square roots, gamma functions, etc. On the other hand, they do not depend on the directional properties of the basis functions involved (i.e., x, y, or z for p-orbitals). $T_N(i,j,k,l)$, in contrast, are quite easily calculated but differ for different directions. It is therefore advantageous first to compute a set of R_N, then to evaluate all integrals related in this way [through the factors T_N] without recalculating the R factors. This will yield a maximum of $(2L_{\max}+1)^4$ related integrals at the same time, $L_{\max}$ being the maximum azimuthal quantum number. The number of symmetry integrals that must be simultaneously kept in core is accordingly increased by this factor.

When contracted Gaussians are used as a basis, the expression for a two-electron integral becomes

$$(ab|cd) = \sum_{i\in a}\sum_{j\in b} S_{ij}C_{ai}C_{bj}\sum_{k\in c}\sum_{l\in d} S_{kl}C_{ck}C_{dl} \times \sum_N R_N(i,j,k,l)T_N(i,j,k,l) \quad , \tag{10}$$

i, j, k, and l being primitive Gaussians constituting the contracted orbitals a, b, c, and d, respectively [with contraction coefficients C_{ai} etc]. As the R and T factors are slowly varying compared to S, the integral may be expressed approximately as

$$(ab|cd) = \sigma_{ab}\sigma_{cd}\sum_N \bar{R}_N\bar{T}_N \tag{11}$$

where $\sigma_{ab} = \sum_{i\in a}\sum_{j\in b} S_{ij}C_{ai}C_{bj}, \sigma_{cd} = \sum_{k\in c}\sum_{l\in d} S_{kl}C_{ck}C_{dl}$, [and] $\bar{R}_N$ and $\bar{T}_N$ are evaluated using average values for the orbital exponents. σ_{ab} and σ_{cd} depend on two indices only, and a list of σs can easily be precomputed and saved in core. The product of σ_{ab} and σ_{cd} hence provides a rapid check on whether the integral should be put equal to zero, approximated according to Eq.(11), or evaluated with full accuracy as a sum of integrals over primitive Gaussians.

The full use of symmetry, the P,Q-supermatrix formalism and the use of the simplifications due to expression (9) result in a very complicated order of the final supermatrix elements. It would therefore be most inconvenient to try to arrange the elements on the output unit in a canonical order. A separate reordering step after the integral calculation is also undesirable, as this would require the simultaneous handling of two large data sets. It has therefore been found useful to assign two indices to each supermatrix element, so making it possible to write them out in an arbitrary order. When indexed two-electron integrals are used, four or five indices are needed for each integral, and the construction of the Fock matrix becomes quite a tedious procedure. In contrast, only two indices are needed when P and Q elements are used, and the construction of the Fock matrix is in fact speeded up somewhat by this indexing.

The above ideas have been implemented in the Hartree-Fock part of the program system MOLECULE, which is being developed [at the Institute of Theoretical Physics, University of Stockholm]. The CI part of this program system has been described earlier [4]. The program has been used in a number of Hartree-Fock calculations, and the gain in simplicity and computer time is considerable compared to earlier programs. This is especially so for medium- and large-size calculations. The time spent in the integral section for some representative molecules is given in Table1.

Acknowledgements. The program was partly debugged and tested at the CNRS computer centre in Strasbourg, and I am greatly indebted to Prof. Veillard for his hospitality and for free computer time. I am also grateful to my colleagues Dr. B. Roos and Dr. P. Siegbahn for many fruitful discussions and suggestions. The work was supported by the Swedish Board for Technical Development and the Swedish Natural Science Research Council.

References

1. Siegbahn P (1970) Chem Phys Lett 8:245
2. Roothaan CCJ, and Bagus PS (1963) Methods in computer physics, vol 2. Academic Press, New York
3. Veillard A (1968) IBMOL, Version IV. IBM Special Technical Report, San Jose
4. Roos B (1972) Chem Phys Lett 15:153