Theoret. Chim. Acta (Berl.) 60, 53–63 (1981)

THEORETICA CHIMICA ACTA

# Ab initio Program for Treatment of Related Molecules. I. Integral Part

Cornelia Kozmutza

Quantum Theory Group, Physics Institute, Technical University of Budapest, Budapest, Hungary

An *ab initio* integral program is described. It utilizes the local symmetries to avoid the redundant computation of integrals over spatially equivalent subsets of the basis. The integrals are grouped in a particular way to facilitate their transfer. The program is very suitable for the treatment of related systems with model geometries. The computing times of different programs are compared and the efficiency of the presented one is demonstrated.

**Key words:** *Ab initio* integral program.

## 1. Introduction

The computation of *ab initio* wavefunctions for large molecules needs huge amount of computer time, big size of central memory as well as enormous data handling [1]. Various procedures were published in the last years for avoiding the difficulties which arise in the computation of two-electron integrals. These begin with the merge, add and move routines as introduced in the program IBMOL [2] and continue by using the symmetry (program POLYATOM [3] and Ahlrichs' program [4]) or the programs which are based on the shell structure [5, 6]. Further simplifications could be obtained by avoiding the computation of sufficiently small integrals [7], by the precalculation of quantities which appear many times in the integral formulae [8] and, last but not least, by using a better and better program organization (using also machine code routines as far as possible). The importance of this latter was shown when large computer program was adopted on minicomputers [9].

In this paper a description of a new *ab initio* integral program is given. This program is devoted to make calculations, including also polarization functions,

0040-5744/81/0053/0053/$02.20

especially on extended molecules, thus the program organization is rather different from those having been published so far. This program is very efficient for computation of related systems. The reasons will be discussed in the following paragraphs in details.

## 2. Program Organization

The integrals are grouped according to the arrangement of atoms in the molecule. This idea presents an intermediate stage between the use of a complete set of symmetry operations and the total neglect of the symmetry in the calculations. It was first proposed by Clementi [7] who used such an arrangement of integrals only in order to give a simple testing procedure.

After the evaluation of one-center integrals the two-center one-electron integrals are calculated. Similarly, first the one-center, then the two-, three- and four-center two-electron integrals are evaluated. Each so-called block is formed by a certain ensemble of two-electron integrals which are needed for constructing a two-center, three-center or four-center "bonding" or a one-center block in itself. This ordering of integrals makes it possible to use these blocks as *transferable* groups. This means that the two-electron integrals of a methyl- or an ethyl-group, e.g. can be calculated separately and so keep and use it in later computations – independently of the total symmetry of the given molecule – for other, *related* molecules. As the program organization allows, furthermore, to calculate each block of integrals separately, it is possible to use even a minicomputer to construct the two-electron integrals of *any* large, extended molecule. The worst case, if one has to calculate a four-center two-electron integral block: even by the use of basis set 6-31G* [10] on all of the four atoms from the second periode it is easy to do on a small computer. The total calculation for all groups can be done step by step.

The non-redundant two-center, three-center and four-center integral blocks are determined and the redundant ones are eliminated using a program before the start of the integral calculations. The one-center two-electron integrals are computed *once* for all of the centers of the same type (i.e. the same atoms). In the case of ethane, e.g. only one carbon and one hydrogen atoms' one-center integrals are blocked and kept on tape/disc. It is worthwhile to keep the one-center integrals if standard basis set is used for a study of a series of molecules.

The total molecular symmetry is also used in the program for avoiding the calculations of vanishing integrals. Before starting the integral evaluations the molecular point group symmetry is taken into account. As very symmetric systems, especialy among large molecules, are rather rare, the program is written only for Abelian groups containing at most four symmetry elements.

When constructing the blocks of integrals, other computer time reduction procedures – besides symmetry – are still recommended: the use of relatively fast routines, an order of integrals which allows an easy further processing (i.e. the SCF supermatrix elements), a simple testing for avoiding zeros and finally an efficient output of integral blocks on tape/disc. Some possibilities will be investigated.

The two-electron integrals are calculated and kept on tape/disc *without* label. This allows to store more integrals. In order to find the given integrals (in the SCF procedure, e.g.), on the other hand, it must be known: the two-electron integrals of each block in the program are grouped in so-called types. This special ordering in types and blocks of integrals makes finally possible the *transferability* of integral groups. This ordering allows, furthermore, an easy handling of integrals. The reason is that the centers of basis functions should not be interchanged in each step and this manifests itself in the decrease of computer time. The calculation is, therefore, faster by the manipulation of types than by the individual integrals.

(a) The one-center integrals of each block are grouped by computation of the following types (A, A1, A2, A3, A4 denotes the different basis functions belonging to the center $A$):

$$\begin{array}{ccccl}
A & A & A & A & 0 \\ \hline
A2 & A2 & A2 & A1 & 0 \\
A2 & A1 & A1 & A1 & 0 \\
\multicolumn{5}{l}{\left.\begin{array}{cccc} A2 & A1 & A2 & A1 \\ A2 & A2 & A1 & A1 \end{array}\right\}} \\ \hline
\multicolumn{5}{l}{\left.\begin{array}{cccc} A3 & A2 & A3 & A1 \\ A3 & A3 & A2 & A1 \end{array}\right\}} \\
\multicolumn{5}{l}{\left.\begin{array}{cccc} A3 & A1 & A2 & A1 \\ A3 & A2 & A1 & A1 \end{array}\right\}} \\
\multicolumn{5}{l}{\left.\begin{array}{cccc} A3 & A2 & A2 & A1 \\ A3 & A1 & A2 & A2 \end{array}\right\}} \\ \hline
\multicolumn{5}{l}{\left.\begin{array}{cccc} A4 & A3 & A2 & A1 \\ A4 & A2 & A3 & A1 \\ A4 & A1 & A3 & A2 \end{array}\right\} \ast} \\ \hline
\end{array}$$

The groups divided by line contain the same number of integrals so they can be calculated in the same DO cycle. The types marked with circle compose the supermatrix elements by themselves. The types connected with } give the contributions to the same supermatrix element therefore they are kept just one behind one. Thus, it is easy to find the corresponding elements even without label. The last group marked with asterisk is an exception where the elements and not the types are one behind one.

(b) The two-center integrals are blocked in a similar way ($A$, $A1$, $A2$, $A3$ denotes the basis functions belonging to the center $A$, $B$, $B1$, $B2$, $B3$ those belonging to $B$). The types from which a B–A two-center "bonding" can be constructed are as follows:

$$\begin{array}{ccccc} B & B & B & A & 0\\ B & A & A & A & 0\end{array}$$
$$\left.\begin{array}{cccc} B & A & B & A\\ B & B & A & A\end{array}\right\}$$

---

$$\left.\begin{array}{cccc} B1 & A2 & B1 & A1\\ B1 & B1 & A2 & A1\end{array}\right\} \qquad \left.\begin{array}{cccc} B2 & B1 & B2 & A1\\ B2 & B2 & B1 & A1\end{array}\right\}$$
$$\left.\begin{array}{cccc} B1 & A1 & A2 & A1\\ B1 & A2 & A1 & A1\end{array}\right\} \qquad \left.\begin{array}{cccc} B2 & A1 & B1 & A1\\ B2 & B1 & A1 & A1\end{array}\right\}$$
$$\left.\begin{array}{cccc} B1 & A2 & A2 & A1\\ B1 & A1 & A2 & A2\end{array}\right\} \qquad \left.\begin{array}{cccc} B2 & B1 & B1 & A1\\ B2 & A1 & B1 & B1\end{array}\right\}$$

---

$$\left.\begin{array}{cccc} B & A3 & A2 & A1\\ B & A2 & A3 & A1\\ B & A1 & A3 & A2\end{array}\right\} * \qquad \left.\begin{array}{cccc} B3 & B2 & B1 & A\\ B3 & B1 & B2 & A\\ B3 & A & B2 & B1\end{array}\right\} *$$

---

$$\left.\begin{array}{cccc} B2 & B1 & A2 & A1\\ B2 & A2 & B1 & A1\\ B2 & A1 & B1 & A2\end{array}\right\} *$$

---

The remarks are as above. It should be added that the types of the vicinal columns can be calculated using the same routine.

(c) The three-center integral groups contain less types than the one- or two-center integral blocks but the number of integrals over the contracted basis functions for the formers are much larger (see Appendix). The notations are as above, the types are the following:

$$\left.\begin{array}{cccc} C & B & C & A\\ C & C & B & A\end{array}\right\}$$

---

$$\left.\begin{array}{cccc} C & B & B & A\\ C & A & B & B\end{array}\right\}$$

---

$$\left.\begin{array}{cccc} C & A & B & A\\ C & B & A & A\end{array}\right\}$$

---

$$\left.\begin{array}{llll} C & B & A2 & A1 \\ C & A2 & B & A1 \\ C & A1 & B & A2 \end{array}\right\} *$$

---

$$\left.\begin{array}{llll} C & B2 & B1 & A \\ C & B1 & B2 & A \\ C & A & B2 & B1 \end{array}\right\} *$$

---

$$\left.\begin{array}{llll} C2 & C1 & B & A \\ C2 & B & C1 & A \\ C2 & A & C1 & B \end{array}\right\} *$$

---

The comments are as above.

(d) The four-center integrals are calculated by using only one type and its permutation, namely:

$$\left.\begin{array}{llll} D & C & B & A \\ D & B & C & A \\ D & A & C & B \end{array}\right\} *$$

---

The number of integrals which are to be calculated by the use of a given type can be determined in a simple way, the formulae are given in the Appendix. It is easy to prove that the well-known formula

$$M1 = \frac{N(N+1)}{2} \quad \text{and} \quad M2 = \frac{M1(M1+1)}{2}$$

can be expressed using the types given in the Appendix, $N$ is the number of contracted basis functions on a center and $M2$ the number of two-electron integrals. The following equality holds:

$$M2 = N + 4\binom{N}{2} + 6\binom{N}{3} + 3\binom{N}{4}.$$

Calculating the exact figures for some commonly used basis set (6-31$G^*$, e.g.) using the formulae presented in the Appendix the results give information what one could gain in the actual calculations ignoring the redundant integral groups. Some numerical results will be given later.

It should be emphasized that the formulae by which the integrals actually are calculated (using $s, p, d$ type basis functions) are not new, they originate from the integral program IBMOL-IV [2]. Only a few simplifications were made as will be pointed out in the next paragraph. The name of this program is SYCETY as it uses the symmetry (SY), the non-redundant groups of centers (CE) and different types (TY) for the computation and the handling of two-electron integrals. Some of these features are used also in the one-electron integrals' computation although the difficulties, as well known, are not as serious there.

## 3. Special Features

Some ways for reducing the time of integral calculations and/or avoiding the redundant manipulations will now be discussed. Attention will also be paid to making the input easy and "safe" what is rather important in cases of large molecules. There are some features already having been used in other programs, but even those are mentioned for the sake of completeness.

(a) A simple generating program precedes the integral evaluation. That program generates the non-redundant two-center, three-center and four-center groups which are to be calculated. This procedure is very easy and fast as only the atoms are subjected to the symmetry operations. The construction of non-redundant blocks helps to avoid at least a part of manual errors in the input data.

(b) The basis set – contracted as well as non-contracted – is read in by cards only once per each different atom. The least possible memory is, therefore, needed for the exponents, coefficients, etc. This is efficient also because the same input data can thus be used for any other related system containing atoms of the same type.

(c) A similar simplification to the above is applied in the following: the exponential parts of the overlaps for the non-contracted functions of non-redundant blocks (called $r_{ij}$ in [4] and SABOO and SCDOO in [2]) are in core. This makes it possible to avoid superfluous computations and, at the same time, to keep the size of memory on a rather low value.

(d) The parity check is used in order to neglect the calculation of zero integrals as is done in other programs (see, e.g. [4]). It is worthwhile, therefore, to compute and store the integrals of types $A4A3A2A1$, $BA3A2A1$, $B2B1A2A1$, $B3A2B1A$, $CBA2A1$, $CB2B1A$, $C2C1BA$ and $DCBA$, respectively, together with their permutation as if the first of three is zero by symmetry then all of them are zero. This order, on the other hand, is not so efficient for other types (because of the interchange of centers), as it was pointed out in the preceding paragraph.

(e) The symmetry is used also in avoiding superfluous calculation for zero integrals of one-center integral blocks. These groups are invariant under three symmetry operations even if the total symmetry of molecular point group has less than three operations. Not only the one-center, but also some two-center (especially between heavy atoms if the symmetry allows) integral blocks can be calculated faster using three (+identity) symmetry operations. The results given later will affirm this argument.

(f) The idea for avoiding the calculation of small enough integrals is also used. The primitive functions for each contracted group are listed in an order of increasing exponents. In this case it is sufficient to test for zero by magnitude only for the first exponential part of overlaps: if it is zero then the program can skip to the next contracted function. Due to the special integral order as described above, this is valid also for the whole integral group. In the case of type DCBA, e.g. if the atoms are far enough apart from each other, the calculation of the total block can be skipped.

(g) Some recalculations can also be avoided. The $T$ values (as used in [2] or $t_{ij}$ in [4]), are independent from the exponents in certain cases and, if it is so, the error function is not recalculated but kept in core.

(h) The handling with zeros – as the integrals are not labelled – is different from that of other programs. There are two versions in the program. One can keep all integrals on tape/disc (including zeros) or the zeros are neglected and the first and last indices of the given integrals between of which the zeros occur are kept in a record, together with the integral record. This version is more efficient than the "label-formalism" as the zeros are often concentrating at certain places on the integral file, especially if the integrals are grouped according to the distribution of atoms in molecule, like in this program.

(i) The memory is used in a dynamic way (may be as used by [11]). This storage allows to make calculations even on a CDC 3300 for determination of wavefunction of molecules containing at most three atoms from the second periode, including also polarization functions.

(j) Some FORTRAN routines were made faster by transferring them to machine code. Certain summations were used instead of products to make the symmetry check faster, etc.

(k) The program allows to calculate not only the non-redundant integrals. The generating program determines also the symmetry factors which are to be used for projecting the redundant ones. Some results are given later. This procedure may be useful if, e.g. a localization procedure or correlation energy calculations will follow the integral +SCF procedure.

(l) It is also possible to prepare in advance some time-consuming tables and values: the error function in some cases, the overlaps for $s$-type orbitals, the one-center integrals for a given basis set, etc. Some integral groups can also be determined in advance and used in calculations for any related system.

It is very advantageous to calculate certain groups in advance and keep them on tape/disc. Then one can simply copy the integrals for a calculation of any related molecule with the same model geometry. Keeping methyl and ethyl groups, e.g. the related methylamine, ethylamine, methylfluorid, ethylfluorid etc., can be "constructed" by saving significant amount of computer time.

## 4. Results

The integral program SYCETY was tested on several computers. Most of calculations were performed on an IBM 3031 and on a CDC 3300 (Hungarian Academy of Sciences, Budapest). The computations on the IBM were made in double precision. The molecules water and ethane were taken as samples and calculated using basis set 6-31G*. The effects of the following procedures were investigated:

(a) calculations of different number of blocks (only the non-redundant ones or all)
(b) use of three symmetry operations for the calculations of the one-center two-electron integrals
(c) three symmetry operations taken into account for the two-electron integral computations of heavy atom – heavy atom "bond"
(d) checking the parity of orbitals
(e) varying the thresholds
(f) calculations of all two-electron integrals by projecting the redundant blocks from the non-redundant ones.

The molecule $H_2O$ belongs to the $C_{2v}$ molecular point group having three symmetry operations: two mirror-planes and one rotation. The molecule lies in the $(xz)$ plane. $C_2H_6$ belongs to the $D_{3d}$ point group with a mirror-plane, one inversion and one rotation. The carbon-carbon bond coïncides with the $(z)$ axis.

The CPU time for the calculation of the integrals of water was 10 minutes on CDC 3300 while that for ethane 45 minutes on IBM 3031. These are the computer times which are needed for the calculations if *all* possible time reducting factors – as mentioned above – are used in the two-electron integral evaluations. All times will be given in CPU minutes. The results obtained have shown that a different amount of CPU time can be gained (or lost) due to the use of the various procedures.

(a) The calculation of only the non-redundant blocks results in an effective decrease of CPU time. The numbers of non-redundant integral blocks for water, ethane and propane, e.g. are given as follows (for a comparison in parantheses the total ones are also presented):

| | $H_2O$ | $C_2H_6$ | $C_3H_8$ |
|---|---|---|---|
| One-center | 2 (3) | 2 (8) | 2 (11) |
| Two-center | 2 (3) | 11 (28) | 20 (55) |
| Three-center | 1 (1) | 17 (56) | 52 (165) |
| Four-center | – | 24 (70) | 99 (330) |

Three symmetry operations help to reduce the number of integral groups to be calculated for all molecules shown. The reduction in CPU time, on the other hand, depends not only on the number of integral blocks but on the number of integrals which belong to the given block as well. More details are given in the Appendix. When all integral blocks were calculated for $H_2O$, the total evaluation of integrals took 15.4 min: more than 50% increase in CPU time because of the addition of one two-center (O—H) block and one one-center (H) block (this latter has much smaller effect). The CPU time for the calculation of all blocks for $C_2H_6$ using basis 6-31G* was more than one a half hour on an IBM 3031 (the evaluation of the non-redundant blocks took only 45 min on the same machine), in spite of the inclusion of symmetry and other time-reducting factors.

(b)–(c) These calculations refer to the same reduction possibility. The following CPU times (in minutes) demonstrate their effect:

| | $H_2O$ | $C_2H_6$ |
|---|---|---|
| using all three operations (see text) | 10.0 | 45.0 |
| not using three operations for the oxygen/carbon one-center integrals | 17.5 | 61.9 |
| not using three operations for the carbon–carbon bond | — | 51.8 |
| not using three operations for the carbon one-center and for the C—C bond | — | 69.1 |

Thus, it becomes clear that using the mirror-planes and one rotation for the one-center and (if the symmetry allows) for some two-center integral groups is very efficient. This effect is especially large in the case of one-center integrals: the reduction in CPU time for water is about 75% while that for ethane 38%.

(d) Checking the parity of orbitals results also in the decrease of CPU time. The calculations have shown that the gain is 3.7 min (37%) for water but only 2.8 min (6%) for ethane if the parity check is done. If neither this check is performed nor the symmetry is used the increase in CPU time is more than 104% for $H_2O$ and 63% for $C_2H_6$. If only one symmetry operation is used instead of three, the loss in CPU time is still larger, for ethane, e.g. 68%.

(e) The threshold was varied only for ethane, but even for this molecule the effect is not too large. If a threshold 1.0D-13 corresponds to 100% of the total time, a threshold 1.0D-08 decreases the time by 4.5%, while a threshold 1.0D-06 is more efficient in the time-reduction (14.3%). This last threshold, however, causes some errors in the integral evaluation (see, e.g. [4]).

(f) The calculation of the non-redundant two-electron integral blocks implies a significant decrease in CPU time. When the redundant blocks were calculated (i.e., the O—$H_2$ bond in the case of $H_2O$, e.g.) the CPU time is longer by about 26%. For the water molecule, on the other hand, the total evaluation took only 12.8 min by using basis 6-31G*, when the O—$H_2$ bond was simply projected (i.e. transferred by symmetry) from the O—$H_1$ bond. This procedure can be proposed if all integrals are needed in a further study. In this case *all* integrals are kept on disc or tape. The total evaluation of integrals by projection is much faster (by about 25%) than the direct calculation of all blocks. If the number of integrals is very large, the projected integrals can be kept separately.

Concluding this paragraph two comparisons are given concerning the calculations by other ab initio programs. The original IBMOL–IV program on a CDC 3300 requires 37.4 min for $H_2O$ (using basis set 6-31G*, i.e. 36 orbitals contracted to 19). Using the Ahlrichs' program (as taken from [12]) for $H_2O$ (Gauss–lobe functions, 47 contracted to 15) the integral procedure takes about 4 min on a SIEMENS 4004, which is about two times faster than a CDC 3300. As the program SYCETY runs 10 minutes for $H_2O$ (by the use of basis 6-31G*), it may be faster even than the latter integral program.

*Acknowledgements.* Thanks are due to Professor E. Kapuy for many valuable discussions. I am indebted to Zs. Ozoróczy and J. Pipek for some help during the program writing.

## Appendix

How to calculate the number of integrals for the different groups

(a) Types of integrals, containing 1, 2, 3 or 4 different basis functions:

| | |
|---|---|
| one | 1111 |
| two | 2221, 2211, 2121, 2111 |
| three | 3321, 3231, 3211, 3121, 3221, 3122 |
| four | 4321, 4231, 4132 |

(b) Let *NI*, *NJ*, *NK*, *NL* the number of basis functions on centers *I*, *J*, *K* and *L*, respectively. Thus from the types given above it follows:

– number of one-center two-electron integrals on center *I*

$$N1CT = NI + 4\binom{NI}{2} + 6\binom{NI}{3} + 3\binom{NI}{4}$$

– number of two-center two-electron integrals on centers *I* and *J*

$$N2CT = 4\,NI \times NJ + 6\binom{NI}{2}NJ + 6\,NI\binom{NJ}{2} + 3\binom{NI}{2}\binom{NJ}{2} + 3\,NI\binom{NJ}{3} + 3\binom{NI}{3}NJ$$

– number of three-center two-electron integrals on centers *I* and *J* and *K*

$$N3CT = 6\,NI \times NJ \times NK + 3\binom{NI}{2}NJ \times NK + 3\,NI\binom{NJ}{2}NK + 3\,NI \times NJ\binom{NK}{2}$$

– number of four-center two-electron integrals on centers *I* and *J* and *K* and *L*

$$N4CT = 3\,NI \times NJ \times NK \times NL$$

## References

1. Kapuy, E., Török, F.: Quantum theory of atoms and molecules (in Hungarian). Budapest: Akadémiai Kiadó 1975
2. Veillard, A.: IBMOL–IV, special IBM Technical Report. San José, California 1968
3. Neumann, D. B., Basch, H., Kornegay, R. L., Snyder, L. S., Moskovitz, J. W., Hornback, C., Liebmann, S. P.: POLYATOM-2, QCPE 199, Indiana University, Bloomington
4. Ahlrichs, R.: Theoret. Chim. Acta (Berl.) **33**, 157 (1974)
5. Newton, M.D., Lathan, W. A., Hehre, W. J., Pople, J. A.: J. Chem. Phys. **51**, 3927 (1969)
6. Dupuis, M., Rys, J., King, H. F.: J. Chem. Phys. **65**, 111 (1976)
7. Clementi, E.: Proc. Nat. Acad. Sci. USA **69**, 2942 (1972)
8. Pople, J. A., Hehre, W. J.: J. Comput. Phys. **27**, 161 (1978)
9. Sparks, R. A.: Int. J. Quantum Chem. Symp. **12**, 191 (1978)
10. Hariharan, P. C., Pople, J. A.: Theoret. Chim. Acta (Berl.) **28**, 213 (1973)
11. Pulay, P.: Theoret. Chim. Acta (Berl.) **50**, 299 (1979)
12. Laurinc, V.: private communication, Bratislava, 1980

Received January 20, 1981/April 27, 1981