Theoret. chim. Acta (Berl.) 28, 307—310 (1973)

Relatio

No N^4-Dependence in the Calculation of Large Molecules

Volker Dyczmons
Lehrstuhl für Theoretische Chemie der Universität Göttingen
Göttingen, Germany

Received September 18, 1972

It is shown that the time needed for computing large molecules is proportional to $N^2 \cdot (\ln N)^2$, when Gaussian functions are used and the molecule is larger than a well-definable limit. Some proposals to overcome this limit are made.

The time T needed for computing a molecule is proportional to the number of integrals one has to calculate. Now let us consider a four-center integral I_{abcd}:

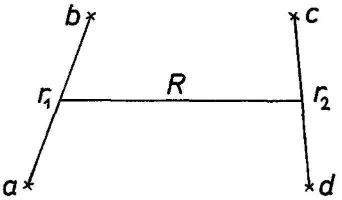

Fig. 1

Then, for large R (cf. Fig. 1)

$$I_{abcd} = \int \varphi_a(1)\, \varphi_b(1) \frac{1}{r_{12}} \varphi_c(2)\, \varphi_d(2)\, d\tau_1\, d\tau_2 \rightarrow S_{ab} S_{cd} \frac{1}{R}\,. \tag{1}$$

If we now use as basis-functions pure Gaussian functions (GF's):

$$\varphi_a(1) = \exp[-\eta_a(\boldsymbol{r} - \boldsymbol{r}_a)^2] \tag{2}$$

we get [1] (cf. Fig. (1)):

$$I_{abcd} \rightarrow \exp(-c_1 r_1^2) \cdot \exp(-c_2 r_2^2) \frac{1}{R} \tag{3}$$

with

$$c_1 = \frac{2\eta_a \eta_b}{\eta_a + \eta_b}, \quad c_2 = \frac{2\eta_c \eta_d}{\eta_c + \eta_d}\,. \tag{3a}$$

We now consider a linear molecule, which lies on the x-axis of a 3-dimensional coordinate system (Fig. 2).

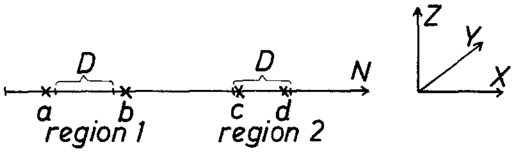

Fig. 2

Let N be the number of basis-functions which are used to describe this molecule. We choose the region 1 with length D in such a way that a and b lie both either within or outside the region. When a and b lie outside, the region should be between a and b. Let the same be true for region 2 with the same length D with regard to c and d.

Now one does not want to calculate all very small integrals. However, we require that the error due to neglect of these integrals ought to be smaller than ε.

Looking at Fig. 2 we note: If a, b lie both outside of region 1, then from (3) and with $R > 1$, one gets:

$$I_{abcd} < \exp(-c_1 D^2)\,. \tag{4}$$

We include in the calculation only such integrals where a, b, and c, d lie within their respective regions. The sum of all neglected integrals ΔI ought to be smaller than ε. This means:

$$\Sigma \Delta I < N^4 \cdot \exp(-c_M D^2) = \varepsilon \tag{5}$$

where c_M is the minimum of all c_1, c_2. It then follows:

$$D = \left[\frac{1}{c_M}\left(4 \ln N + \ln \frac{1}{\varepsilon}\right)\right]^{1/2} \propto \sqrt{\ln N} \tag{6}$$

summin over all positions of the regions 1 and 2 gives an additional factor N^2. The time T necessary for computing a linear molecule is then:

$$T \propto N^2 D^4 \propto N^2 \cdot (\ln N)^2\,. \tag{7}$$

For a nonlinear or nonplanar molecule we can repeat the same considerations for the y- and/or z-directions (see Fig. 2). Regions 1 and 2 then become spheres with the diameter D. With $N = N_x \cdot N_y \cdot N_z$ the total number of basisfunctions Eq. (5) is still valid. By summing over all positions one gets a factor $N_x^2 \cdot N_y^2 \cdot N_z^2 = N^2$ and the result is again Eq. (7).

Now what is the optimal way in calculating large molecules?:

a) First of all we note that GF's are of course better than Slater functions (for which $T \propto N^2 \cdot (\ln N)^4$).

b) There is a special kind of GF's which is optimal as basis-functions. They are grouped pure GF's with a *fixed* number of different η-values (the last point saves about 60% computing time).

c) One has to optimise molecular basis-sets for the different bonds (see e.g. [2]) and s-p-hybrids, as e.g. Huzinaga has done for atomic wave-functions [3].

d) One has to avoid small η-values as far as possible; cf. Eq. (6) and the definition of c_M in Eq. (3a).

After all this has been done one has to write computer programs in such a way that the estimate in Eq. (7) really holds. But this is easily done. One has only to test for which *a*, *b* (respectively *c*, *d*)

$$\int |\varphi_a(1)|\,|\varphi_b(1)|\,d\tau_1 > \frac{\varepsilon}{N^4} \tag{8}$$

holds. A better approximation than (5) shows that in (8) one gets const$\cdot\varepsilon/N$ instead of ε/N^4 [6].

In those cases where the correlation energy plays an important part, a SCF-calculation is not sufficient. For large molecules the following scheme is probably optimal as far as total computing time is concerned:

1) SCF-calculation.

2) *Complete* localization of the SCF-MO's; this is only possible by using non-orthogonal MO's.

3) Calculation of the pair-energies (see e.g. [4]). Interpair energies are calculated for neighbouring pairs only. The correlation energies of all these pairs are calculated only in the HF-field of neighbouring SCF-MO's. This is opposed to some methods which calculate the correlation energy for the molecule as a whole.

4) Calculation of the errors which arise by the additivity assumption for the pair-energies [5].

The great advantage of this scheme is that the time needed for the calculation of the correlation energy is *only* proportional to N, as compared to an N^2-dependence for the SCF-calculation.

This approach is probably good enough for most molecules. It may become possible to use tabulated values of intrapair and interpair energies and of pair additivity errors of small molecules for the computation of the correlation energy of larger systems.

Now we want to determine the molecular size for which Eq. (7) holds. For this purpose we consider a linear molecule: if it is larger than $2D$ (see Fig. 2), then Eq. (7) holds. For example, let us determine the critical size $2D$ for a chain of C-atoms with a C-C-distance of 2.5 a. u. With the choice of $\varepsilon = 10^{-6}$, 30 basis functions per C-atom and $c_M = 0.3$ we obtain with Eq. (6) a minimal number of 9 C-atoms. But Eq. (6) is not a very good approximation for this case since Eq. (6) was derived for the limit of molecule much larger than $2D$. 9 carbon atoms is a crude upper bound on the size of the molecule for which Eq. (7) applies. A better estimate of D in this case can be determined by the usual requirement that the smallest integral to be calculated should be equal to 10^{-8}. This means that $10^{-8} = S_{ab} = \exp(-c_M D^2)$. We also take $c_M = 0.3$ and get then $D = 7.8$ a. u. Therefore Eq. (7) holds if we have more than 6 C-atoms.

The essential point is this: Where is the limit from which the computing time increases quadratically? It is easily shown [6] that this limit is smaller than the one obtained above (i. e. about 5 C-atoms).

Another very important question is: How must the numerical accuracy of the individual integrals depend on N? In [6] it will be shown that this accuracy goes with $1/N$.

In conclusion, the techniques discussed in this paper should make *ab initio* calculations on larger molecules possible within a few years time.

I am grateful to Prof. Dr. W. A. Bingel for his encouragement of this work and to the „Deutsche Forschungsgemeinschaft" for its support.

References

1. Boys, S. F.: Proc. Roy. Soc. (London) A **200**, 542 (1950).
2. Dyczmons, V., Staemmler, V., Kutzelnigg, W.: Chem. Physics Letters **5**, No. 6, 361 (1970).
3. Huzinaga, S.: J. chem. Physics **42**, 1293 (1967).
4. Jungen, M., Ahlrichs, R.: Theoret. chim. Acta (Berl.) **17**, 339 (1970).
5. Dyczmons, V., Kebabcian, R.: to be published.
6. Dyczmons, V.: to be published.

V. Dyczmons
Lehrstuhl für Theoretische Chemie
der Universität Göttingen
D-3400 Göttingen, Bürgerstr. 50a
Federal Republic of Germany