COLOPHON

The attached paper has no connection with my current research in Computer Science and Digital Documents. It is another in my series of experiments to see how long it takes me to re-build electronic versions of my published early papers as properly re-typeset ‘PDF Normal’ rather than just as a bitmap scan.

This particular paper appeared in the proceedings of a conference held at The Science Research Council’s (now EPSRC) Atlas Computer Laboratory in April 1974. As far as I can detect, via Google, these Proceedings have not been made available electronically in any form.

The text was acquired by scanning the paper from the original Proceedings and then using Omnipage OCR on the resulting TIFF files. The paper was then re-typeset using UNIX *troff* suite to set up the correct typeface (Times) and to get the line and page breaks as accurate as possible.

The eight diagrams were extracted from the various TIFF pages and after clean-up in Adobe Photoshop, they were then vectorized using Adobe Streamline. The lettering was re-set in Adobe Illustrator, using Helvetica Bold, before exporting each diagram to version 3.2 of Illustrator’s Encapsulated PostScript (but with no TIFF preview). This Encapsulated PostScript was then incorporated into the paper using *psfig* .

The time taken to rebuild this paper was about 8 hours.

Approximate Ab Initio Calculations and the Method of Molecular Fragments

D.F.Brailsford*

A two stage approach to performing *ab initio* calculations on medium and large sized molecules is described. The first step is to perform SCF calculations on small molecules or molecular fragments using the OPIT Program. This employs a small basis set of spherical and *p*-type Gaussian functions. The Gaussian functions can be identified very closely with atomic cores, bond pairs, lone pairs, etc. The position and exponent of any of the Gaussian functions can be varied by OPIT to produce a small but fully optimised basis set.

The second stage is the molecular fragments method. As an example of this, Gaussian exponents and distances are taken from an OPIT calculation on ethylene and used unchanged in a single SCF calculation on benzene. Approximate *ab initio* calculations of this type give much useful information and are often preferable to semi-empirical approaches, since the nature of the approximations involved is much better defined.

Introduction

All-electron atomic and molecular *ab initio* calculations have been possible for about fifteen years now, and have underlined the importance of electronic computers in sustaining the momentum of quantum chemistry research. Such calculations offer a way of avoiding the many uncertainties and pitfalls that bedevil semi-empirical calculations. For this reason most *ab initio* calculations have tried to obtain high accuracy for molecular energies and properties by solving the SCF equations using a large basis set of suitably chosen functions. Little effort is usually expended in optimising such non-linear parameters as basis function exponents, because the basis is usually so large that energies better than 99% of the SCF limit can be obtained simply from the automatic determination of linear parameters (molecular orbital coefficients) that takes place in the SCF procedure. Calculations of this sort can be further refined by configuration interaction methods if required,

However, we now have a situation where there is a great difference between the accuracy and reliability of semi-empirical calculations and accurate *ab initio* work. In an attempt to bridge this gap I should like to review the work that we have been doing at Nottingham over the past five years in what might be called approximate *ab initio* calculations. By ‘approximate’, in this context, I mean that we are satisfied with energies that are about 95% of the SCF limit. The reason for this approach can best be summed up in a remark once made by the late Professor Coulson. He said that he was often asked by chemists to give them a clear description of familiar chemical concepts such as lone pairs and bond pairs, in the language of modern quantum mechanics. His reply was always to the effect that a 90% correct description is very easy to give in this way, but that describing the extra 10% is much more difficult to do and the wavefunction tends to become almost unrecognisable in chemical terms. For this reason we at Nottingham have long been interested in seeing whether a viable model for closed shell atoms and molecules can be set up using a small basis set of very simple functions.

The Frost Model

This consists of modelling a closed shell molecule using a single floating spherical gaussian function for each pair of electrons in a molecule. Thus for a $2n$ electron system there are n spherical gaussians of the form:-

$$g_i = N_i e^{-\alpha_i (r - R_i)^2}$$

where

N_i is the normalisation factor
α_i is the exponent (size factor) of the gaussian
R_i is the position of the centroid of the i^{th} gaussian relative to some centre of co-ordinates.

It is usual in the Frost [1,2] method to optimise the exponents α_i and positions R_i of all the gaussian functions involved. In most cases one of the gaussians will come to rest on each heavy nucleus in the molecule and can be kept fixed there. Figure 1 shows

* *Department of Mathematics, University of Nottingham, University Park, Nottingham, NG7 2RD*

Frost model basis sets for methane and ethylene. In the case of methane the CH bond functions are equivalent by symmetry and so the number of independent non-linear parameters is only three i.e. the core and bond gaussian exponents, and the distance of the bond gaussian along the CH bond (denoted αC, αCH and d_{CH} respectively). In ethylene we have five independent parameters, αC, αCH, d_{CH},α_{π} and d_{π}. The last two parameters denote the exponents, and distance from the internuclear axis, of the two spherical gaussian functions that model the π electron system.

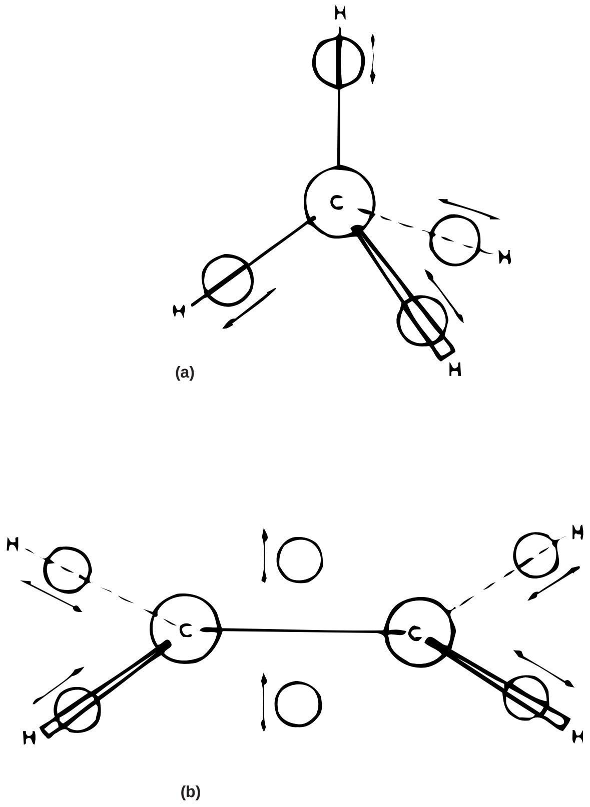

Figure 1: Frost models for (a) methane and (b) ethylene: arrows show the direction of movement of the floating gaussians

For each setting of the non-linear parameters indicated by the optimisation routine an energy is calculated. This is extremely easy in the Frost model, for the number of basis functions is equal to the number of electron pairs, and there is no need for an iterative scheme to find the energy. In fact, the density matrix in the Frost model turns out to be just the inverse of the overlap matrix [1]. Also, the choice of spherical gaussians makes all integrals easy to evaluate [3]. When non-linear parameters are optimised the Frost model gives, typically, 85% of the SCF energy limit. This low accuracy, however, is not so much of a drawback as the inherent instability of the model under certain circumstances. The main contribution to the molecular energy arises from the 'heavy' carbon atoms in the molecule. It follows, therefore, that it the optimisation routine can improve the description of the heavy atoms then a lower energy will probably be obtained. This is achieved by moving the bond pairs in methane (say) closer in to the carbon nucleus. If they coalesce, with equal exponents, then the overlap matrix becomes singular with off-diagonal elements of unity. This extreme form of behaviour does not usually occur in saturated systems, but is almost certain to occur in unsaturated molecules. For example, the gaussian functions simulating the π orbital system in ethylene will coalesce if their positions are allowed to optimise fully.

OPIT

The first version of this program was written in 1969 for a KDF9 computer. As its name implies it is both optimising and iterative. It has recently been completely rewritten and implemented on ICL 1900 series computers and the CDC 7600 machine [4,5]. The program is a general SCF-MO floating gaussian program capable of giving single determinant closed shell wavefunctions to any required degree of accuracy provided a large enough basis set is specified. However, its main use so far has been to extend the scope of the Frost model by using two independently optimised spherical gaussians at each ’heavy’ nucleus. Under these conditions the cusp condition at the nucleus is much better approximated, and the energy obtained is now 95% of the SCF limit. The extra core gaussian greatly reduces the tendency of lone pair or bond pair gaussians to coalesce on the nucleus during the

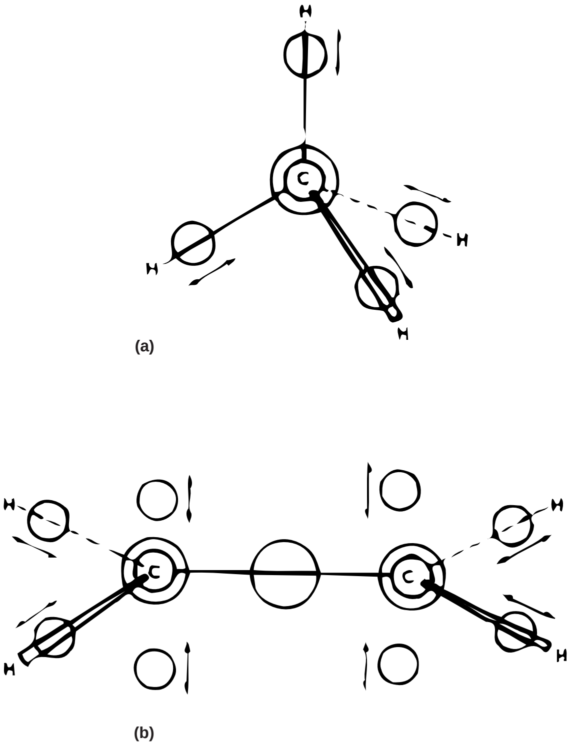

Figure 2: OPIT models for (a) methane and (b) ethylene

optimisation. But, since the basis set is now larger than the Frost minimal set, a SCF calculation is needed to calculate the energy for every setting of the non-linear parameters. Possible OPIT models for methane and ethylene are shown in figure 2 and calculations of this type are reported in the paper by Ford, Hall and Packer [6]. A coherent description of saturated organic molecules can be obtained from models of this sort. Bond lengths and molecular angles are predicted semi-quantitatively. The energy ordering of conformers e.g. the staggered and eclipsed forms of ethane, is given correctly but the predicted barrier heights are rather high. However, when one studies OPIT values for barrier heights, geometrical features and other first order properties [7] a pleasing pattern emerges that the predicted values are a constant fraction of the true values.

Despite the many advantages of OPIT over a simple Frost treatment, the problem of modelling π electron systems still remains. For, even with an extra core gaussian, any attempt to model a molecule such as ethylene, in the manner depicted in figure 2b, would still fail due to the tendency of the spherical gaussians representing the π system to coalesce, with identical exponents, on to the carbon nuclei. In fact, figure 3 shows the effect of fixing all ethylene parameters at the values shown, except for the distance, *d*, of the π spherical gaussians from the carbon atoms. This distance is then allowed to optimise and the best result is clearly obtained as $d \rightarrow 0$! Study of the molecular orbital eigenfunctions and the behaviour of the overlap matrix for small *d* [8], shows a clear tendency for the two pairs of gaussians to combine antisymmetrically to give functions of *p* type symmetry.

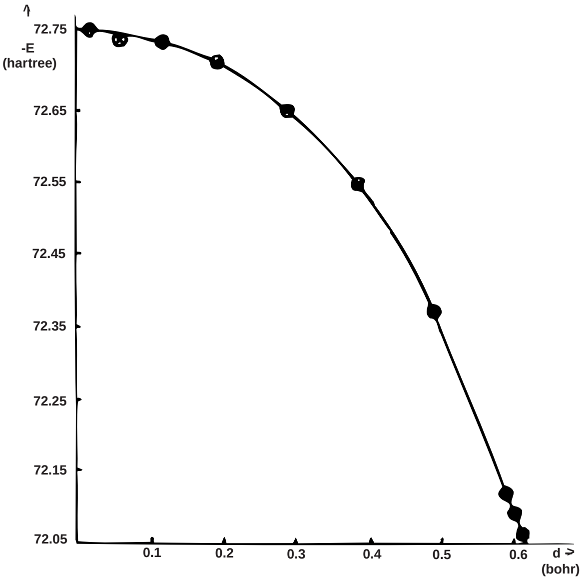

Figure 3: Plot of energy *versus* distance (d) of the π spherical gaussian from the C atom, in the OPIT ethylene model. The remaining parameter values are fixed at: α_C(inner) = 40.0, α_C(outer) = 10.0, α_{CH} = 0.45 α_{CC} = 0.5, α_π = 1.0, d_{CH} = 1.336

The symmetric combination of the functions, however, also contributes more strongly to the σ system around the C–C bond, as *d* becomes small. This helps to lower the total energy of the system and, thus, configurations with small *d* values are favoured by the optimisation routine.

For this reason an attempt was made to add into the Fock matrix only the antisymmetric contributions from these functions. This, unfortunately, only had the effect of slowing down the coalescence of the functions, and gives a worse energy for the system because of the extra constraint. Now it has always been part of the OPIT philosophy that there should be no constraint in optimising any non-linear parameter, if it is desired to do so. Faced with this situation we had to capitulate and include 'genuine' p-type gaussians as defined by Boys [9] into the program. The integrals over these functions were performed using the relationships given by Huzinaga *et al.* [10] but in the form of the specific formulae given by Clementi and Davis [11]. Note that, in the interests of keeping the basis set small, we include only those p_z type orbitals that are essential for describing the π orbital system.

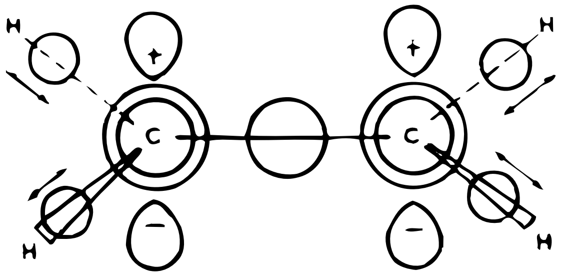

Figure 4: OPIT model for ethylene using *p*-type gaussians

Table 1: Ethylene molecular orbital energies [*]

Species	OPIT	Moskowitz *et al.* [12]
$1a_g$	-11.064	-11.240
$1b_{1u}$	-11.052	-11.238
$2a_g$	-1.019	-1.040
$2b_{1u}$	-0.759	-0.780
$1b_{2u}$	-0.495	-0.655
$3a_g$	-0.405	-0.581
$1b_{3g}$	-0.392	-0.514
$1b_{3u}$ (π)	-0.207	-0.374
$1b_{2g}$ (π)	+0.354	
Total Energy	-74.6083	-78.0062

---

[*] All energies in hartree

The model for our ethylene calculation is now shown in figure 4. In accordance with the philosophy of full optimisation, the *p* functions were even left free to drift off the carbon nuclei. However, the amount of drift was found to be negligible, and so in this, and all subsequent calculations using *p* functions, they are anchored to the heavy nuclei. The results for this ethylene calculation are shown in table 1, the ordering of the orbital energies agrees with that obtained in a much more accurate calculation performed by Moskowitz *et al.* [12].

Similar calculations, with full optimisation, were performed on *cis* and *trans* butadiene with basis sets as shown in figure 5, and also for benzene. The energies obtained are shown in tables 2 and 3. Note how the energies obtained are about 95% of those obtained in the accurate *ab initio* calculations that are included for comparison. In agreement with our general experience of OPIT, we predict the energy difference between the *cis* and *trans* conformers of butadiene to be too high by a factor of two.

Table 2: Total energies* for *cis* and *trans* butadiene

	OPIT	Buenker and Whitten [13]
cis	-148.2463	-154.7023
trans	-148.2609	-154.7103
ΔE	0.0146 (9.1 kcal mole^{-1})	0.0080 (5 kcal mole^{-1})

* Energies are in hartree except where stated. Geometry is as described in [13]

Table 3: Total energies for benzene

OPIT	Buenker *et al.* [14]
-221.027 hartree	-230.375 hartree

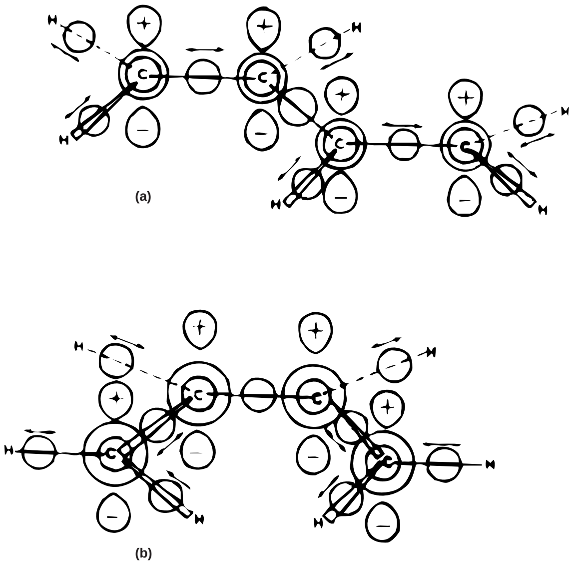

Figure 5: OPIT model for (a) *trans* and (b) *cis* butadiene

The most interesting aspect of these calculations on simple π orbital systems, lies in the values of the optimised non-linear parameters. These results are collected in table 4. To a good approximation the core and bond exponents do not vary from molecule to molecule, with the exception of the *p* orbital exponents. We see a clear effect, for the *p* functions, in the case of butadiene, to become less diffuse on the central carbon atoms than they would have been in ethylene, while those on the terminal carbon atoms become more diffuse. The difference between the central atom and terminal atom *p* orbital exponents is greater in *trans* than in *cis* butadiene reflecting, perhaps, the different steric factors in the two molecules. As might be expected, the greater delocalisation of the π electron system in benzene leads to the *p* orbital being rather more diffuse than in ethylene and its exponent value differs from ethylene by 10%.

Table 4: OPIT optimal parameter values for ethylene, butadiene and benzene

	Ethylene	Butadiene		Benzene
		cis	*trans*	
α_C (inner)	45.90	45.945 45.915	45.958 45.918	45.93
α_C (outer)	6.63	6.639 6.635	6.642 6.635	6.64
α_π	0.33	0.33 0.458	0.290 0.493	0.36
α_{CC}	0.40	0.433 0.386	0.436 0.386	0.41
α_{CH}	0.377	0.378 0.379 0.378	0.377 0.378 0.377	
d_{CH}	1.255 1.279	1.254 1.255 1.280	1.239 1.244	1.276
d_{CC}	-	1.251	1.249	

α Values are optimal gaussian exponents

d Values are optimal distances (in bohrs) measured from the heavy atom along the bond direction

The multiple parameter values for butadiene reflect the number of symmetrically distinct parameters of each type in the molecule; parameters for the terminal atoms and bonds are given first in each set followed by the values for central atoms or bonds

These calculations confirm my earlier remark that 95% accurate calculations can fit in with chemical notions in a very simple and straightforward way. Table 4 shows us the importance of π orbitals, the amount of their delocalisation, and lends weight to the hypothesis of $\sigma-\pi$ separability, since the σ systems remain essentially constant while the π systems alter appreciably. It is particularly pleasing that these results arise from an all electron *ab initio* calculation.

The high degree of parameter transferability that is evident so far, prompts the use of OPIT as a tool for optimising non-linear parameters in small molecules, with a view to their later use in the method of molecular fragments described by Christoffersen *et al.* [15,16]. The idea here is that, as molecules become bigger, it is neither necessary nor desirable to optimise fully all the non-linear parameters. Instead, one fixes the parameters at values obtained from smaller species and then a ‘one-off’ SCF calculation is performed without any attempt at reoptimising the non-linear parameters. As the larger molecules have little or no symmetry, a special program (SCOFF–**S**elf Consistent **C**alculation Using **O**ptimised **F**ixed **F**ragments) has been developed. In essence it resembles the very first function evaluation of OPIT but with special techniques to cope with low symmetry, and the large number of near-zero integral values that occur simply because of large inter-atomic distances. The essential features of OPIT modelling (i.e. single spherical gaussians for lone pairs and bond pairs; two spherical gaussians on each heavy nucleus with *p* type gaussians for π systems) are retained in SCOFF.

Work Performed with SCOFF

As a check on the viability of the SCOFF method we have recalculated many systems that had previously been done, with OPIT. As an example of this, table 5 shows the results of recalculating *cis* and *trans* butadiene and benzene using ethylene parameters and ignoring the effects such as differing C–C lengths in the species involved. We See that the energies obtained by the two methods are in remarkably good agreement although, of course, the better values are given by OPIT. The stabilities of the butadiene conformers are still predicted in the correct order and the energy difference is close to that obtained by OPIT.

Table 5: Comparison of energies* for butadiene and benzene obtained from a full parameter optimisation (OPIT) and a molecular fragments approach using ethylene parameters (SCOFF)

	OPIT	SCOFF
butadiene		
E *cis*	-148.2463	-148.2317
E *trans*	-148.2609	-148.2440
ΔE	0.0146	0.0123
	(9.1 kcal mole^{-1})	(7.7 kcal mole^{-1})
benzene		
E	-221.0268	-221.0145

* Energies are in hartree except where stated. Effects due to bond length differences between butadiene, benzene and the ethylene fragment have been ignored

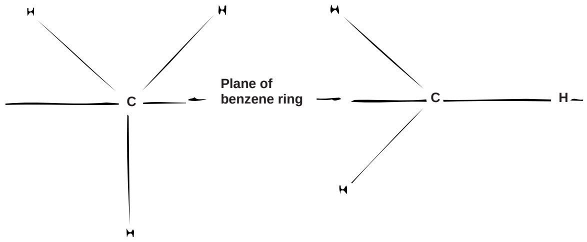

Figure 6.. Energies for two toluene conformers obtained from SCOFF

Finally, as examples of the possibilities opened up by a molecular fragments program like SCOFF, I shall quote two recently obtained results. The first is summarised in figure 6 and shows the energies of two toluene conformers. The form in which the CH_3 group is fully staggered with respect to the benzene ring is predicted to be fractionally more stable than the one in which a hydrogen atom is eclipsed. Secondly, we have looked at the drug amphetamine in its protonated form. The important torsional angles τ_1, τ_2 and τ_3 are shown in figure 7. We have calculated energies for this molecule for $\tau_1 = 90^\circ$ and various τ_2. The relative stabilities of the various rotated forms are shown in figure 8 together with interpolated results for the same species calculated by Pullman *et al.* [17] using semi-empirical methods. The overall shape of the plot is the same, but SCOFF appears to differentiate between the three equivalent minima predicted by the PCILO method and indicates that the rotamer with the NH_3^+ group folded back over the benzene ring, but the CH group extended away from the ring, will be fractionally more stable.

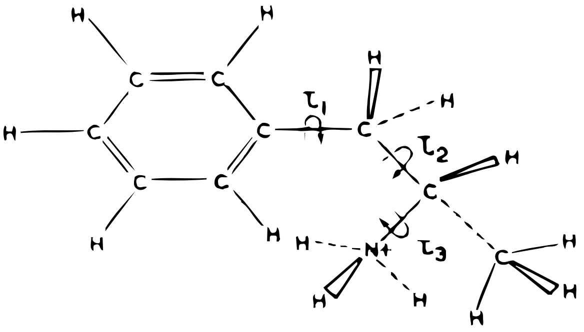

Figure 7: The amphetamine ion

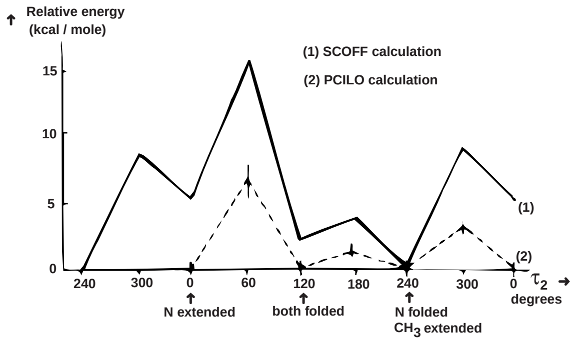

Figure 8: Relative energy of rotameric forms of amphetamine as a function of the angle τ_2. τ_1 is fixed at 90° (see figure 7): the PCILO calculation is reported in [17]

Conclusions

Although we have used OPIT in many contexts to obtain a simple and readily understandable wavefunction, the emphasis in this paper has been on its use for obtaining molecular fragment parameters for later use in the SCOFF program. Approximate *ab initio* calculations of this type give about 95% of the Hartree-Fock SCF limiting energy. Most of the remaining energy deficit can be ascribed to the poor description of atomic nuclei in this model. In general, valence electron and bond properties are well predicted, but even when such properties are given incorrectly, the results obtained are usually a constant fraction of the true values. The wavefunction obtained, being of *ab initio* initio type is much more amenable to error analysis than semi-empirical results.

The place of semi-empirical and accurate *ab initio* calculations is already well established in quantum chemistry. We believe that approximate *ab initio* molecular fragment techniques are well suited to bridge the gap between these extreme approaches in areas such as the study of medium sized molecules of biological interest.

Acknowledgements

In a review paper of this sort 1 have had to draw heavily on the experience and research work of other colleagues. I should like to thank Professor G.G. Hall for his continued interest in this project, and also Brian Ford, Jane Hylton, Colin Miller, Gary Schnuelle (who wrote the SCOFF program) and Andrew Tait. These colleagues have all contributed in some measure to the results reported here. We thank the Cripps Computer Centre, University of Nottingham, and the Science Research Council Atlas Computer Laboratory, Chilton, for allocations of computer time.

References

[1] FROST, A.A. (1967). *J. Chem. Phys.*, **47**, 3707.

[2] ___________ (1967). *J. Chem. Phys.*, **47**, 3714.

[3] SHAVITT, I. (1963). *Methods Comp. Phys.*, **2**, 1.

[4] PACKER, J.C. and BRAILSFORD, D.F. (1973). *Comp. Phys. Comm.*, **5**, 123.

[5] BRAILSFORD, D.F. and PRENTICE, J.A. (1973). *Comp. Phys. Comm.*, **5**,136.

[6] FORD, B., HALL, G.G. and PACKER, J.C. (1970). *Int. J. Quantum Chem.*, **4**, 533.

[7] DIXON, M. and TAIT, A.D. *Mol. Phys.*, (submitted for publication).

[8] FORD, B. and HALL, G,G. (1974). *Comp. Phys. Comm.*, **8**, 337.

SCHNUELLE, G.W., (unpublished work).

[9] BOYS, S.F. (1950). *Proc. Roy. Soc. (London),* **A200**, 542.

[10] TAKETA, H., HUZINAGA, S. and O-OHATA, K. (1966). *J. Phys. Soc, Japan*, **21**, 2313.

[11] CLEMENTI, E. and DAVIS, D.R. (1966). *J. Comp. Phys.*, **1**, 223.

[12] SCHULMAN, J.M., MOSKOWITZ, J. and HOLLISTER, C. (1967). *J. Chem. Phys.*, **46**, 2759.

[13] BUENKER, R.J. and WHITTEN, J.L. (1968). *J. Chem. Phys.*, **49**, 5381.

[14] BUENKER, R.J., WHITTEN, J.L. and PETKE, J.D. (1968). *J. Chem. Phys.*, **49**, 2261.

[15] CHRISTOFFERSEN, R.E., GENSON, D.W. and MAGGIORA, G. (1971). *J. Chem. Phys.*, **54**, 239.

[16] CHRISTOFFERSEN. R.E. (1971). *J. Am. Chem. Soc.*, **93**, 4104.

[17] PULLMAN, B., COUBEILS, J-L., COURRIERE, P. and GERVOIS, J-P. (1971). *J. Medicinal Chem.*, **15**, 17.