POLYATOMIC MOLECULAR DIRAC-HARTREE-FOCK CALCULATIONS WITH GAUSSIAN BASIS SETS.
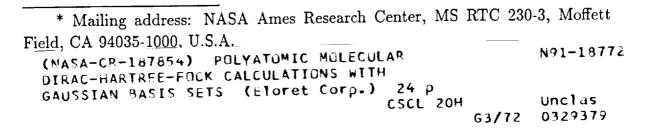
Kenneth G. Dyall†, Knut Fægri, Jr.‡ and Peter R. Taylor†.

†Eloret Institute, 3788 Fabian Way, Palo Alto,
California 94303, U.S.A. *
‡Department of Chemistry, University of Oslo,
P.O. Box 1033, Blindern, N-0315 Oslo 3, Norway.

INTRODUCTION

The majority of research in *ab initio* quantum chemistry is performed on molecules containing light atoms, and indeed a large part of chemical research generally is concerned with such systems. *Ab initio* quantum chemical calculations are limited by the performance of computer hardware and software, and advances in these areas have increased the scope of quantum chemistry to the point where it is now possible to perform accurate calculations on systems containing the lighter transition metals. However, there are many important chemical processes involving heavy elements, and here another obstacle to accurate quantum chemical calculations presents itself — the increasing importance of relativistic effects with increasing atomic number.

The chemical effects of relativity have been extensively documented (for ex-

---

   * Mailing address: NASA Ames Research Center, MS RTC 230-3, Moffett Field, CA 94035-1000, U.S.A.

haustive reviews see Pyykkö 1978, 1986, 1988, Pitzer 1979, Pyykkö and Desclaux 1979). The traditional methods of computational chemistry are clearly inadequate for the description of important phenomena such as relativistic orbital contractions and spin-orbit splitting — effects that may decisively influence structure as well as reactivity of heavy-atom molecules. Yet molecules containing heavy atoms are important in a number of areas of chemistry, such as catalysis and surface chemistry. The need for accurate quantum chemical calculations in these areas has motivated recent theoretical and methodological developments which are aimed at overcoming the obstacles to a relativistic treatment of systems containing heavy atoms.

One widely-used approach to calculations on molecules containing heavy atoms is the relativistic effective core potential (RECP) method. While this method has considerable advantages, there are situations in which it is inappropriate, and in most applications spin-orbit splitting is not included. Effective core potential methods also need calibration against *ab initio* calculations for validation. An all-electron relativistic method is clearly required, both for rigour and understanding, and in evaluating alternatives.

Numerical methods have been used successfully in atomic Dirac-Hartree-Fock (DHF) calculations for many years (Desclaux 1975, Grant et al. 1980). Some DHF calculations using numerical methods have been done on diatomic molecules (Laaksonen and Grant, 1984, Sundholm et al. 1987, Sundholm 1988), but while these serve a useful purpose for calibration, the computational effort in extending this approach to polyatomic molecules is prohibitive. An alternative more in line with traditional quantum chemistry is to use an analytical basis set expansion of the wave function. This approach fell into disrepute in the early 1980s due to problems with variational collapse and intruder states, but has recently been put on firm theoretical foundations (Grant 1986, Grant and Quiney, 1988, Quiney 1988). In particular, the problems of variational collapse are well understood, and prescriptions for avoiding the most serious failures have been developed. Consequently, it is now possible to develop reliable molecular programs using basis set methods. We describe such a program in this paper, and report results of test calculations to demonstrate the convergence and stability of the method.

THEORY

With a single determinant many-electron wave function constructed from

4-spinors $|j\,\rangle$, we may write the (unrestricted) Dirac-Fock energy as

$$E = \sum_{j=1}^{n} \langle j \,|\, \hat{h}_D \,|\, j \rangle + \frac{1}{2} \sum_{j,k=1}^{n} [\langle j\,k \,|\, \hat{g} \,|\, j\,k \rangle - \langle j\,k \,|\, \hat{g} \,|\, k\,j \rangle]. \tag{1}$$

The one-electron operator in the field of the nuclei is

$$h_D = -ic\,\alpha.\nabla + (\beta - 1)c^2 + V^{nuc}, \tag{2}$$

where $\alpha = (\alpha_x, \alpha_y, \alpha_z)$; α_x, α_y, α_z and β are 4×4 matrices,

$$\alpha_x = \begin{pmatrix} 0 & \sigma_x \\ \sigma_x & 0 \end{pmatrix}, \; \alpha_y = \begin{pmatrix} 0 & \sigma_y \\ \sigma_y & 0 \end{pmatrix}, \; \alpha_z = \begin{pmatrix} 0 & \sigma_z \\ \sigma_z & 0 \end{pmatrix}, \; \beta = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}.$$

σ_x, σ_y and σ_z are the Pauli spin matrices, and I_2 is the 2×2 unit matrix. The fully covariant electron-electron interaction can be expanded in a power series in c^{-2}. The lowest order term, which is $\mathcal{O}(c^0)$, is the Coulomb interaction,

$$\hat{g} \equiv \hat{g}(1,2) = \frac{1}{r_{12}}. \tag{3}$$

The term which contributes at the next order, $\mathcal{O}(c^{-2})$, is the Breit interaction, whose contribution to the energy comes mainly from the region near the nuclei. For present purposes, the Coulomb interaction is an adequate description of the electron-electron interaction. Writing the 4-spinors in terms of large and small component 2-spinors $\mathbf{j}^L$ and $\mathbf{j}^S$,

$$|j\rangle = \begin{pmatrix} \mathbf{j}^L \\ i\mathbf{j}^S \end{pmatrix} \tag{4}$$

where the superscripts L and S indicate large and small components respectively, we obtain for the matrix elements of the one- and two-electron operators

$$\begin{aligned} \langle j | \hat{h}_D | j \rangle = {} & c\left[\langle \mathbf{j}^L \,|\, (\sigma.\nabla) \,|\, \mathbf{j}^S \rangle - \langle \mathbf{j}^S \,|\, (\sigma.\nabla) \,|\, \mathbf{j}^L \rangle \right] \\ & + \langle \mathbf{j}^L \,|\, V^{nuc} \,|\, \mathbf{j}^L \rangle + \langle \mathbf{j}^S \,|\, V^{nuc} - 2c^2 \,|\, \mathbf{j}^S \rangle, \end{aligned} \tag{5}$$

$$\begin{aligned} \langle j\,k \,|\, \hat{g} \,|\, j\,k \rangle = {} & (\mathbf{j}^L \mathbf{j}^L \,|\, \mathbf{k}^L \mathbf{k}^L) + (\mathbf{j}^L \mathbf{j}^L \,|\, \mathbf{k}^S \mathbf{k}^S) \\ & + (\mathbf{j}^S \mathbf{j}^S \,|\, \mathbf{k}^L \mathbf{k}^L) + (\mathbf{j}^S \mathbf{j}^S \,|\, \mathbf{k}^S \mathbf{k}^S) \end{aligned} \tag{6a}$$

$$\langle j\, k\,|\,\hat{g}\,|\,k\,j\,\rangle = (\mathbf{j}^L \mathbf{k}^L\,|\,\mathbf{k}^L \mathbf{j}^L) + (\mathbf{j}^L \mathbf{k}^L\,|\,\mathbf{k}^S \mathbf{j}^S)$$
$$+ (\mathbf{j}^S \mathbf{k}^S\,|\,\mathbf{k}^L \mathbf{j}^L) + (\mathbf{j}^S \mathbf{k}^S\,|\,\mathbf{k}^S \mathbf{j}^S). \tag{6b}$$

We expand the large and small components in a basis of 2-spinors $\{\mu^L\}$ and $\{\mu^S\}$

$$|\mathbf{j}^L\rangle = \sum_{\mu=1}^{N} c_{\mu j}^{L}\,|\mu^L\rangle; \quad |\mathbf{j}^S\rangle = \sum_{\mu=1}^{N} c_{\mu j}^{S}\,|\mu^S\rangle, \tag{7}$$

and define the nuclear potential energy, overlap, kinetic energy and density matrix elements by

$$V_{\mu\nu}^{XX} = \langle \mu^X\,|V^{nuc}|\,\nu^X\rangle, \tag{8}$$

$$S_{\mu\nu}^{XX} = \langle \mu^X\,|\,\nu^X\rangle, \tag{9}$$

$$\Pi_{\mu\nu}^{XY} = \langle \mu^X\,|\sigma.\nabla|\,\nu^Y\rangle, \tag{10}$$

$$D_{\mu\nu}^{XY} = \sum_{j=1}^{n} c_{\mu j}^{X\dagger} c_{\nu j}^{Y}, \tag{11}$$

respectively, where X and Y can be L or S, with the restriction that for the kinetic energy matrix elements, $X \neq Y$. The Dirac-Fock energy can then be written

$$E = \sum_{\mu\nu}^{N} \left[c(D_{\mu\nu}^{LS}\Pi_{\mu\nu}^{LS} - D_{\mu\nu}^{SL}\Pi_{\mu\nu}^{SL}) + D_{\mu\nu}^{LL}V_{\mu\nu}^{LL} + D_{\mu\nu}^{SS}(V_{\mu\nu}^{SS} - 2c^2 S_{\mu\nu}^{SS}) \right]$$

$$+ \frac{1}{2} \sum_{\mu\nu\kappa\lambda}^{N} \left[D_{\mu\nu}^{LL} D_{\kappa\lambda}^{LL} \{(\mu^L \nu^L\,|\,\kappa^L \lambda^L) - (\mu^L \lambda^L\,|\,\kappa^L \nu^L)\} \right. \tag{12}$$

$$+ D_{\mu\nu}^{SS} D_{\kappa\lambda}^{SS} \{(\mu^S \nu^S\,|\,\kappa^S \lambda^S) - (\mu^S \lambda^S\,|\,\kappa^S \nu^S)\}$$

$$\left. + 2D_{\mu\nu}^{LL} D_{\kappa\lambda}^{SS} (\mu^L \nu^L\,|\,\kappa^S \lambda^S) - 2D_{\mu\nu}^{LS} D_{\kappa\lambda}^{SL} (\mu^L \lambda^L\,|\,\kappa^S \nu^S) \right].$$

Differentiating with respect to the large and small component coefficients, we obtain the following matrix representation of the Dirac-Fock equations,

$$\begin{pmatrix} \mathbf{F}^{LL} - \epsilon \mathbf{S}^{LL} & \mathbf{F}^{LS} \\ \mathbf{F}^{SL} & \mathbf{F}^{SS} - \epsilon \mathbf{S}^{SS} \end{pmatrix} \begin{pmatrix} \mathbf{c}^L \\ \mathbf{c}^S \end{pmatrix} = 0. \tag{13}$$

The elements of the various blocks of the Fock matrix are defined by

$$
F_{\mu\nu}^{LL} = V_{\mu\nu}^{LL} + \sum_{\kappa\lambda}^{N} \left[D_{\kappa\lambda}^{LL} \{ (\mu^L \nu^L \mid \kappa^L \lambda^L) - (\mu^L \lambda^L \mid \kappa^L \nu^L) \} \right.
$$
$$
\left. + D_{\kappa\lambda}^{SS} (\mu^L \nu^L \mid \kappa^S \lambda^S) \right],
$$

(14a)

$$
F_{\mu\nu}^{SS} = V_{\mu\nu}^{SS} - 2c^2 S_{\mu\nu}^{SS} + \sum_{\kappa\lambda}^{N} \left[D_{\kappa\lambda}^{SS} \{ (\mu^S \nu^S \mid \kappa^S \lambda^S) - (\mu^S \lambda^S \mid \kappa^S \nu^S) \} \right.
$$
$$
\left. + D_{\kappa\lambda}^{LL} (\mu^S \nu^S \mid \kappa^L \lambda^L) \right],
$$

(14b)

$$
F_{\mu\nu}^{LS} = c\, \Pi_{\mu\nu}^{LS} - \sum_{\kappa\lambda}^{N} D_{\kappa\lambda}^{SL} (\mu^L \lambda^L \mid \kappa^S \nu^S) = F_{\mu\nu}^{SL\dagger}.
$$

(14c)

The 2-spinors may be written as a combination of scalar functions $|a\rangle$ and $|b\rangle$ with spin functions:

$$
|\mu^X\rangle = T_{\mu a}^{X\alpha} |a\rangle \begin{pmatrix} 1 \\ 0 \end{pmatrix} + T_{\mu b}^{X\beta} |b\rangle \begin{pmatrix} 0 \\ 1 \end{pmatrix}.
$$

(15)

where $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are unit vectors in spin space corresponding to $m_s = \frac{1}{2}$ and $m_s = -\frac{1}{2}$, conventionally labelled α and β. These may be used to further reduce the Fock matrix expressions:

$$
F_{\mu\nu}^{XY} = \sum_{\sigma\tau} \sum_{ab} T_{\mu a}^{X\sigma\dagger} F_{ab}^{\sigma\tau} T_{\nu b}^{Y\tau},
$$

(16)

where σ and τ run over both spin indices. The Fock matrix elements in the scalar basis are given by the following expressions. If functions a and b belong to the same component (L or S),

$$
F_{ab}^{\sigma\sigma} = V_{ab}^{\sigma\sigma} + \sum_{cd} (ab \mid cd) \left[D_{cd}^{\sigma\sigma} + D_{cd}^{\tau\tau} \right] - \sum_{cd} (ad \mid cb) D_{cd}^{\sigma\sigma},
$$

(17a)

$$
F_{ab}^{\sigma\tau} = - \sum_{cd} (ad \mid cb) D_{cd}^{\tau\sigma}.
$$

(17b)

The sums over c and d for the direct integrals $(ab|cd)$ extend over both components, while those for exchange integrals $(ad|cb)$ extend only over the same component as a and b. For the blocks connecting the large and small components,

$$F_{ab}^{\sigma\sigma} = (2\sigma)\, c\Pi_{ab}^{0} - \sum_{cd}(ad|cb)\, D_{cd}^{\sigma\sigma} \qquad (17c)$$

$$F_{ab}^{\sigma\tau} = c\Pi_{ab}^{2\tau} - \sum_{cd}(ad|cb)\, D_{cd}^{\tau\sigma} \qquad (17d)$$

where a and d are large component functions, and b and c are small component functions. The kinetic energy matrix elements are defined by

$$\Pi_{ab}^{0} = \langle a\,|\,\frac{\partial}{\partial z}\,|\,b\rangle, \quad \Pi_{ab}^{\pm 1} = \langle a\,|\,\frac{\partial}{\partial x}\pm i\frac{\partial}{\partial y}\,|\,b\rangle. \qquad (18)$$

CONSIDERATIONS FOR IMPLEMENTATION

There are a number of alternatives to be considered in the implementation of the Dirac-Fock method in a basis set, such as choice of basis function type and nuclear model, relations between small and large component basis functions, form of spinor expansions, transformation from the scalar to the 2-spinor basis, inclusion of double group symmetry (including time-reversal symmetry), details of integral storage and transformation and SCF method. We discuss these alternatives in the following subsections.

<u>Basis function type and nuclear model</u>

The principal causes of variational collapse in the attempt to solve the Dirac equation are the failure to satisfy the boundary conditions at the nucleus (Grant and Quiney 1988), and the failure to ensure the proper relations between the large and small component basis functions (Ishikawa et al. 1983, Dyall et al. 1984, Stanton and Havriliak 1984). The second point will be addressed in the next section.

Numerical solutions of the Dirac-Fock equations have the boundary conditions built in, so that any spherically symmetric model of the nuclear charge distribution may be employed. For finite basis set approximations, these boundary conditions will determine the form of the basis functions. Thus, the choice of basis function type and nuclear model are interrelated.

The traditional model used in electronic structure calculations for the nuclear charge distribution is the point nuclear model, which gives rise to the cusp in the non-relativistic electronic wave function at the nucleus. In the solution of the electronic Dirac equation, this cusp is replaced by a singularity, which has to be modelled by the basis functions. For atoms and diatomics, Slater-type functions with *non*-integral exponents of r should be used, such as the S-spinors or L-spinors advocated by Quiney et al. (1989). While these are convenient for atomic calculations, and may also be useful for diatomic molecules, their use for polyatomic systems would be computationally intractable.

The alternative to the point charge model for the nucleus is to use some kind of charge distribution with a finite radius, for which the wave function is no longer singular at the origin. For the purposes of electronic structure calculations, the details of the model for the nuclear charge distribution are not critical, provided they approximately represent the real distribution *. The most popular models in atomic structure calculations are the uniformly charged sphere model,

$$\rho_{nuc}(r) = \rho_0, \ r \leq r_0$$
$$= 0, \ r > r_0$$

and the Fermi distribution,

$$\rho_{nuc}(r) = \rho_0[1 + \exp((r - a)/c)]^{-1}.$$

Visser et al. (1987) have investigated the use of a single Gaussian function for the nuclear charge distribution in basis set DHF calculations,

$$\rho_{nuc}(r) = \rho_0 \exp(\eta_{nuc}r^2).$$

None of these models is anything but a crude representation of the nuclear charge distribution, but they are adequate for electronic structure calculations. The justification for the use of the uniform and Fermi models is that they have been used in the fitting of nuclear scattering data to obtain gross nuclear dimensions. The parameters r_0 for the uniform distribution, a and c for the Fermi distribution, and η_{nuc} for the Gaussian distribution are determined from these fits to nuclear scattering data. The effect of the choice of finite nuclear model on the energy is not

---

* This will not necessarily be true, of course, for properties such as nuclear hyperfine structure and parity non-conservation effects, which may be sensitive to the nuclear model.

Table 1. The effect of the nuclear model on the $1s$ eigenvalue of Hg^{79+}, given in E_h.

Nuclear model	Eigenvalue E	E − E(point)
Point	−3532.191 849	
Uniform	−3530.174 275	2.017 574
Fermi	−3530.182 156	2.009 693
Gaussian	−3530.193 999	1.997 850

large, as shown by some numerical calculations for Hg^{79+} using a modification of the GRASP program (Dyall et al. 1989), given in table 1. The nuclear size effect is of the order of 2 E_h, but the effect of the shape of the nuclear boundary is only of the order of 20 mE_h.

The relation between the choice of basis functions and the nuclear charge model may be clarified by the following analysis. The nuclear potential for any finite nuclear charge distribution may be expanded in a power series about the origin,

$$V^{nuc}(r) = v_0 + v_2 r^2 + v_3 r^3 + \ldots.$$

Note that there is no term linear in r, so that for small r, regardless of the details of the nuclear model, the potential is approximately harmonic. The solutions of the Schrödinger equation for a harmonic potential are Hermite Gaussian functions; the solutions of the Dirac equation are not, but may be represented by Gaussian functions. At very short range, then, Gaussian functions will be appropriate basis functions. If all terms of odd order in the series expansion of the nuclear potential have zero coefficients, then the solutions of the electronic Dirac equation for a many-electron atom are either pure even or pure odd functions of r. Both the uniform and the Gaussian nuclear charge distribution have such an expansion inside the nuclear radius. The series for the uniform model is only valid inside the nuclear radius, but for the Gaussian nuclear model the series is valid for all r. Solutions of the Dirac equation with a uniform nuclear model will have discontinuities in the higher derivatives at the nuclear boundary, but for a Gaussian nuclear model, the solutions will be continuous and differentiable to all orders for all r. The Fermi distribution, on the other hand, has terms of odd order in the series expansion, which will introduce elements in the solutions which have cusps in the higher derivatives at the origin. Use of the Gaussian nuclear model thus leads to solutions which

are mathematically more well-behaved than the other two models, and as a result, expansion of the solutions in a finite Gaussian basis set is likely to have superior convergence properties.

One further, practical consideration in the choice of nuclear model needs mention. The calculation of many-centre nuclear attraction integrals would have to be done numerically for the uniform and Fermi models, whereas for the Gaussian model, the integrals may be evaluated using existing technology, since they require only straightforward changes to the expressions for point nuclear integrals.

Kinetic balance

Many of the problems with variational collapse and intruder states disappeared, once it was established that the small component basis functions must at least be related to the large component functions (Ishikawa et al. 1983, Dyall et al. 1984, Stanton and Havriliak 1984) by the condition now known as kinetic balance,

$$\{\mu^S\} \supseteq \{\sigma.\nabla\,\mu^L\}. \tag{19}$$

Most of the remaining problems relate to the nuclear boundary conditions discussed above. Kinetic balance is, however, only a zeroth-order or non-relativistic approximation, which is useful when the potential V^{nuc} is much smaller than $2c^2$. A relation which more accurately represents the small component where the potential is large (Ishikawa et al. 1983, Dyall et al. 1984), derived from the one-electron Dirac equation, is

$$\{\mu^S\} \supseteq \{\,[2c^2 + \epsilon - V^{nuc}]^{-1}\,\sigma.\nabla\,\mu^L\,\}. \tag{20}$$

With a Gaussian nucleus, kinetically balanced Gaussian basis functions do not satisfy this relation, but they still provide a good representation of the small component.

Applying $\sigma.\nabla$ to jj-coupled 2-spinors composed of spherical harmonic Gaussian functions of the form

$$\chi_{n,\ell,m}(r,\theta,\phi) = \mathcal{N}r^{n-1}\exp(-\zeta r^2)Y_{\ell m}(\theta,\phi),$$

with $n = \ell + 1$, yields the following results. For the $j = \ell + \frac{1}{2}$ spin-orbit components of the large component 2-spinors, the small component 2-spinors are composed of functions $\chi_{n+1,\ell+1,m'}$, with appropriate values of m'. The small component basis functions in this case have one *more* unit of angular momentum than the large component basis functions. For the $j = \ell - \frac{1}{2}$ spin-orbit components of the large component 2-spinors, the small component 2-spinors are composed of functions $\chi_{n+1,\ell-1,m'}$ and $\chi_{n-1,\ell-1,m'}$. In this case, the small component basis functions have one *less* unit of angular momentum than the large component basis functions, but they are also composed of *two* radial functions rather than just one. Thus, the $1s_{1/2}$, $2p_{3/2}$, $3d_{5/2}$, ...large component 2-spinors generate $2p_{1/2}$, $3d_{3/2}$, $4f_{5/2}$, ...small component 2-spinors; but the $2p_{1/2}$, $3d_{3/2}$, $4f_{5/2}$, ...large component 2-spinors generate $1s_{1/2}$, $2p_{3/2}$, $3d_{5/2}$, ... and $3s_{1/2}$, $4p_{3/2}$, $5d_{5/2}$, ...2-spinors. Kinetic balance also requires equal exponents for corresponding large and small component Gaussian functions.

If we consider the scalar Gaussian basis functions of which the 2-spinors are composed, there are $2\ell + 1$ large component basis functions for each ℓ value, if the two spin-orbit components share the same basis functions. By the above rules, these generate $6\ell + 1$ small component basis functions, except for s, which generates 3 basis functions. The total number of basis functions for a given ℓ shell is then $8\ell + 2$, which represents an approximately *fourfold* increase in the basis set size over the corresponding nonrelativistic basis set size of $2\ell + 1$ spherical harmonic Gaussians. If the two spin-orbit components do *not* share the same basis functions, we must add $2\ell + 1$ functions to the large component basis, making a total of $10\ell + 3$ basis functions, which is an almost *fivefold* increase.

For example, consider a non-relativistic basis set consisting of 15 s, 11 p, 6 d and 3 f functions, which has a total of 99 basis functions. With common basis functions for the spin-orbit components in the large component set, the corresponding relativistic basis set would have these 99 functions for the large component set, and the small component set would consist of 11 s, 21 p, 14 d, 6 f and 3 g pure spherical functions, and 11 $3s$, 6 $4p$ and 3 $5d$ "contaminants" — a total of 257 basis functions for the small component. Including both components, the relativistic basis set would consist of 356 basis functions. Without the use of common basis functions for the spin-orbit components, the relativistic basis set would consist of 440 basis functions in total.

In a contracted basis set, it would generally be necessary to have different contracted functions for the two spin-orbit components of a given shell. Both the large and small component functions would have different contraction coefficients, including the contaminants (which belong to one spin-orbit component) and the pure spherical functions (which belong to the other). If we contracted the above non-relativistic basis to $6s5p3d2f$, for example, the relativistic contracted basis would have 80 large component basis functions and 130 small component basis functions, a total of 210 basis functions, compared with 43 basis functions for the nonrelativistic contracted basis. Thus, contraction as a space-saving mechanism is not quite as advantageous in a relativistic calculation as compared with a nonrelativistic calculation, owing to the need to duplicate the large component set for all basis function types except s.

Choice of spinor expansion

The expansion of the components of the 4-spinors $|j\rangle$ in a basis set can be done in three ways: independent expansions for each of the four components,

expansion in terms of 2-spinors, as outlined in this paper, and expansion in 4-spinors, where each component is fixed in relation to the others. Since the last of these has not been much used, we will focus on the first two. Atomic finite basis calculations (see Pyykkö 1986 for references) have usually exploited atomic symmetry, and expanded only the radial parts of the large and small components in a basis. These use equal length expansions for the large and small components. Some molecular calculations have also used a 2-spinor basis (Hegarty and Aerts 1987).

Early molecular calculations (Mark et al. 1980, Lee and McLean 1982) used separate expansions for each of the four spinor components. This corresponds to carrying out the entire calculation in the scalar basis. It inevitably means that the basis set expansions for the large and small components will have different lengths. For example, consider the case of a single s function for the large component. This requires a p-type function in the small component, which consists of three functions: p_1, p_0 and p_{-1}. If these are coupled to the spin to form jj-coupled 2-spinors, we get two functions for the $p_{1/2}$ and four for the $p_{3/2}$ spin-orbit components of the p shell. Only the $p_{1/2}$ set are used to represent the small component for the large component s function. The $p_{3/2}$ functions form a basis for a "negative-energy" (or positron) state of symmetry $j = 3/2$, which will also appear in the spectrum along with the desired states. If the basis is ill-chosen, these extra states may not even be confined to the negative continuum, but appear instead as intruder states in the bound state region or in the positive continuum. A classic example of intruder or spurious states is the calculation of the hydrogenic states for $j = \ell - \frac{1}{2}$ by Drake and Goldman (1982), who observe a state degenerate with the state of the same j, but different ℓ value — for instance, a $p_{1/2}$ state degenerate with the $1s_{1/2}$ state. The features of this approach, then, are that it gives an unbalanced representation of the two branches of the spectrum, with a better description of the negative continuum, which is of no interest for molecular properties, than of the bound states and positive continuum. More than that, the extra functions will give rise to extra eigenvalues, which may occur as intruder states in the parts of the spectrum which are of interest, as Grant and Quiney (1988) have warned.

Some more recent 2-spinor based calculations, both on atoms (Ishikawa and Sekino 1990) and molecules (Visscher et al. 1990), have also employed different expansion lengths for the large and small components. Their expansions have a better representation of the negative energy states than the positive energy states. Experience has generally shown that, provided the basis functions are appropriate for the boundary conditions, the intruder states are to be found in the negative

continuum. However, the physical states of interest may be distorted by the higher density of states in the negative continuum (see Stanton and Havriliak 1984), and mask problems of variational collapse or bounds failure. A more serious problem arises in the treatment of electron correlation if the negative continuum is to be included, for example in perturbation sums. Calculations of this type go beyond the no-pair approximation, i.e. they include contributions from creation of electron-positron pairs (see Grant and Quiney 1988 for a discussion). While the inclusion of the negative continuum is not likely to be of importance for molecular properties, the distortion of the positive spectrum, which results from the higher density of states in the negative continuum, may even affect calculations that do not include it.

Though the seriousness of the problems that may arise from the use of different length expansions will depend on the nature of the calculations, we believe it is wise to avoid any possible problems by ensuring that the large and small component spinor basis sets are of the same length, which means using jj-coupled atomic 2-spinors as a basis for both large and small components in matched sets.

Transformation to 2-spinors

The considerations of the previous section affect only the final representation of the eigenstates of the system under study. In the process of the construction of these eigenstates, it is not necessary to keep all quantities in the 2-spinor basis. Indeed, there may be some advantages in using the scalar basis at various stages in the calculations. Computationally, the transformation to 2-spinors is not expensive, as the transformation matrix is very sparse. The transformation will scale only as the number of indices to be transformed. The point at which the transformation is done will then be determined by the practical considerations of storage and efficiency.

In a direct SCF procedure, the transformation can be done while the Fock matrix is being constructed; in conventional SCF, there are two points at which the transformation to 2-spinors may be done. The first is at the stage of two-electron integral generation; the second is after construction of the Fock matrix.

The decision of whether or not to transform the integrals will be influenced by a number of factors, but the overriding consideration will be that of disk space, given the much larger basis sets needed in a relativistic calculation. Most molecular integral codes generate integrals over real basis functions, which may be Cartesian or spherical functions, atom-centred or symmetry-adapted functions. Since the jj-coupled 2-spinors are complex, the permutational symmetry of the integrals will be reduced after the transformation, and hence the number of values to be stored will increase. Time-reversal symmetry regains one degree of permutational freedom. If the integrals are generated over symmetry functions, then the fact that the 2-spinor functions are generally distributed between fewer irreducible representations in the double group than the scalar functions are in the single group will, in general, mean an increase in the number of values stored. These increases are offset by the reduction in size of the small component basis upon transformation. It is not immediately clear that transformation will inevitably increase the total storage requirements, but it may do so. The transformation would be implemented either at the same stage as the spherical harmonic transformation, since it is of a similar nature, or after the symmetry transformation (if the integral code includes this).

If the integrals are kept in the scalar basis, then the Fock matrix must be transformed instead. Since the construction of the Fock matrix is the most time-consuming step in a conventional SCF calculation, it is important to make it efficient. Use of unordered integrals in a non-relativistic calculation does not permit vectorization; to achieve this, ordered integrals or supermatrices are used. Due to

the large basis sets needed for relativistic calculations, the extra disk space required for integral ordering may not be available. When constructing the Fock matrix in the scalar basis, however, each $(ab|cd)$ integral with distinct indices contributes in 36 unique places. It is then possible to vectorize the construction of the Fock matrix using sparse vector operations. While the gain in speed from vectorization is smaller than that obtained from the use of ordered integrals or supermatrices, it is nevertheless substantial.

IMPLEMENTATION

Our program uses for the scalar basis spherical harmonic Gaussian functions, which are symmetry-adapted for D_{2h} and its subgroups. The jj-coupled 2-spinors constructed from these are symmetry functions for the corresponding double group. Linear molecules are treated as a special case. The nuclear charge distribution is a single Gaussian with an exponent chosen to match the rms radii of the Gaussian and the nucleus, given by

$$\eta_{nuc} = 3/(2r_{rms}^2).$$
(21)

The nuclear rms radius is fitted to a function of the nuclear mass:

$$r_{rms} = \langle r^2 \rangle_{nuc}^{1/2} = 0.836A^{1/3} + 0.57$$
(22)

where A is in amu and r_{rms} is in fm. This formula is appropriate for the Fermi 2-parameter distribution. The one-and two-electron integrals, which are generated by an adaption of MOLECULE (Almlöf and Taylor, unpublished), are kept in the scalar basis. The Fock matrix is constructed in the scalar basis, then transformed to the 2-spinor basis. Each (double group) symmetry block of the Fock matrix is symmetrically orthonormalized, and diagonalized using standard EISPACK routines. Density damping has been implemented to accelerate convergence.

RESULTS

Testing of any new program requires comparison of a number of calculations of various types with known results. For atoms, the adapted GRASP program (Dyall et al. 1989) was used to obtain numerically accurate Dirac-Fock-Coulomb energies with a Gaussian nuclear charge distribution for comparison with results from the present program. For molecules, fewer results exist for comparison. The numerical diatomic calculations of Laaksonen et al. (1984) and Sundholm (1988) were used for comparison. All calculations were done with uncontracted basis sets.

One-electron calculations

The first sets of calculations were performed on one-electron systems. Tables 2 and 3 give results for the 1s orbital of H and Hg^{79+}. The basis sets for H were taken from Partridge (1989), while those for Hg^{79+} were obtained by scaling the H sets. The results converge from above on the exact solution in both cases. For H, the error in the total energy, due to basis set truncation, converges more slowly than

Table 2. Convergence of energies for hydrogen. Values given in μE_h.

Basis	ΔE	$\Delta(\Delta E_{rel})$	% error
$7s$	16.724	0.021	0.32
$8s$	5.448	0.010	0.15
$9s$	1.869	0.005	0.07
$10s$	0.670	0.002	0.03
$11s$	0.250	0.000	0.01
$12s$	0.097	0.000	0.00
$13s$	0.039	0.000	0.00
$14s$	0.016	0.000	0.00
$15s$	0.007	0.000	0.00

Numerical results: $E = -0.500\,006\,656\ E_h$, $\Delta E_{rel} = -6.656\,59\ \mu E_h$.

Table 3. Convergence of energies for Hg^{79+}. Values given in mE_h.

Basis	ΔE	$\Delta(\Delta E_{rel})$	% error
$7s$	2938.266	2839.618	0.86
$8s$	1229.427	1199.215	0.36
$9s$	458.401	448.863	0.14
$10s$	138.589	135.432	0.041
$11s$	32.612	31.475	0.010
$12s$	12.081	11.635	0.004
$13s$	9.370	9.197	0.003
$14s$	4.058	3.996	0.001
$15s$	0.540	0.504	0.0002

Numerical results: $E = -3530.193\,999\ E_h$, $\Delta E_{rel} = -330.482\,066\ E_h$.

Table 4. Convergence of energies for Hg^{79+} with optimized scale factors. Values given in mE_h.

Basis	ΔE	$\Delta(\Delta E_{rel})$	% error
$7s$	2337.469	2238.815	0.71
$8s$	937.087	906.869	0.27
$9s$	329.167	319.623	0.097
$10s$	94.710	91.547	0.028
$11s$	23.785	22.642	0.007
$12s$	11.674	11.205	0.003
$13s$	8.477	8.299	0.002

Table 5. Results for 2p states of Hg^{79+}.

2$p_{1/2}$ orbital energies and differences in E_h.

Basis	ϵ	ΔE	$\Delta E(3s)^a$
9p	−904.269 615	0.550 647	0.122 584
10p	−904.653 964	0.166 299	0.034 308
11p	−904.784 643	0.035 620	
11p^b	−904.802 234	0.018 029	0.014 425
numerical	−904.820 263		

2$p_{3/2}$ orbital energies and differences in E_h.

Basis	ϵ	ΔE
9p	−817.805 038	0.002 446
10p	−817.806 484	0.001 000
11p	−817.806 640	0.000 844
11p^b	−817.806 318	0.001 166
numerical	−817.807 484	

[a] Extra energy difference when 3s functions omitted from small component spinors.
[b] Previous basis with most diffuse function replaced by tight function.

the error in the relativistic correction, as the basis set size is increased. Clearly, non-relativistic basis sets for H do not need modification for use in the Dirac equation. For Hg^{79+}, however, the error in the total energy is almost entirely due to the error in the relativistic correction. The basis set clearly does not describe the relativistic contraction of the 1s orbital well. We optimized the scale factor for each basis, with results as given in Table 4. Some 25% of the missing energy is recovered by this procedure, but the error in the relativistic correction still dominates the error in the total energy. This indicates that nonrelativistic basis sets need to be re-optimized, at least for the core orbitals, if they are to be used in DHF calculations.

Results of calculations on the 2$p_{1/2}$ and 2$p_{3/2}$ orbitals of Hg^{79+} with a few different basis sets are given in table 5. The basis sets were taken from atomic calculations on Pb (Fægri 1987). Here, the effect of spin-orbit splitting on the basis sets may be seen. The error for the 2$p_{1/2}$ is much larger than that for the 2$p_{3/2}$,

even when the most diffuse function is replaced by a tighter function. Matsuoka and Okada (1989) found it necessary to add two tight p functions to the non-relativistic basis sets of Fægri (1987) in order to sufficiently reduce the eigenvalue errors for the neutral $6p$ block elements. Clearly, the basis set requirements for the two spin-orbit components are different, which implies that, in a contracted basis, it will be necessary to use different contractions for the $2p_{1/2}$ and $2p_{3/2}$ functions.

Strict kinetic balance requires the $2p_{1/2}$ small component basis functions to be composed of a $1s$ and a $3s$ function in a fixed ratio. It may be argued that the $3s$ could be represented as a linear combination of $1s$ functions, and therefore may be omitted. We have investigated the effect of omitting the $3s$ part of the small component spinor for each basis set size. The results are also given in Table 5. The importance of the $3s$ decreases with increasing basis set size, as is expected. With large enough basis sets, we conclude that the $3s$ may be omitted, but for smaller basis sets, which may be necessary for molecular calculations, it may still be beneficial to retain the $3s$. It is important also to note that omission of the $3s$ does *not* cause bounds failure: to the contrary, the energy increases when it is left out. We should emphasize that the coefficients of the $1s$ and $3s$ functions in our calculations are *not* independently varied, as in the recent work by Ishikawa and Sekino (1990). Our approach is equivalent to their method 2.

Finally, some calculations were done on H_2^+ and HeH^+. The s basis sets were taken from Partridge (1989), and the p basis sets were chosen in the range normally used for polarization functions on H. The results are given in Tables 6 and 7, along with the numerical results of Laaksonen and Grant (1984). Though these calculations do not approach the basis set limit, they at least demonstrate that there are no problems with variational collapse. Even when the difference between the $7s$ and the $15s$ basis results is added to the $7s4p$ result, as an estimate of the truncation error in the s set, the energy is still above the numerical value.

Many-electron systems

We have done several series of calculations on many-electron atoms, to investigate the trends in basis set errors with basis set size and with atomic number. Results of a representative set of calculations, for 10-electron systems with $Z = 10$, 20, 40 and 80, are given in Table 8. The Ne basis sets were taken from Partridge (1989). The basis sets for the ions were energy optimized in non-relativistic finite nucleus atomic calculations. Our results for Ne are consistent with those of Hegarty and Aerts (1987). The results are presented in terms of the basis set truncation

Table 6. Eigenvalues in E_h for H_2^+ at $R = 2$ bohr.

Basis	$\epsilon_{1/2g}$	$\epsilon_{1/2u}$
$7s$	$-1.090\ 734$	$-0.667\ 302$
$15s$	$-1.090\ 954$	$-0.667\ 337$
$7s1p$	$-1.101\ 289$	$-0.667\ 438$
$7s3p$	$-1.102\ 384$	$-0.667\ 516$
$7s4p$	$-1.102\ 405$	$-0.667\ 517$
numerical	$-1.102\ 642$	$-0.667\ 553$

Table 7. Lowest eigenvalue in E_h for HeH^{2+} at $R = 2$ bohr.

Basis	$\epsilon_{1/2}$
$7s4p$	$-2.511\ 977$
numerical	$-2.512\ 296$

error and the error in the predicted relativistic correction for each basis set. The trends with atomic number demonstrate again that for the heavy atoms, the error in the total energy comes mainly from the non-optimal nature of the basis for relativistic calculations. However, the fractional error in the relativistic correction increases slowly with atomic number: for example, the error for Hg is only twice that for Ne.

Most of the testing has been on atomic systems. We present also a few calculations on molecular systems, H_2 and H_2O, in Tables 9 and 10. For H_2, the numerical results were taken from Sundholm (1988). The basis set for H_2, which is the same as that used for H_2^+ is modest in size — the basis set truncation error is 5.5 mE_h, but the relativistic correction is accurate to 0.4 μE_h. Using the previous 7s basis results for H atom, we obtain a result for the relativistic contribution to the binding energy of 1.1 μE_h, in agreement with the numerical result to better than 0.1 microhartree. Thus, while the absolute value of the relativistic correction may not be accurate due to an inadequate description of the region near the nucleus, the relativistic contribution to the binding energy is well predicted. Of course, relativistic effects are very small for H_2, but nevertheless, this result may carry over to other molecules. It may be that the re-optimization of the core basis sets, found to

Table 8. Comparisons of results for Ne, Ca^{10+}, Zr^{30+} and Hg^{70+}.

Percentage error in relativistic correction.

Basis	Ne	Ca^{10+}	Zr^{30+}	Hg^{70+}
$7s3p$	2.50	2.54	2.80	5.12
$9s5p$	0.53	0.41	0.46	1.12
$10s6p$	0.29	0.20	0.22	0.58
$12s7p$	0.11	0.07	0.08	0.24
$13s8p$	0.06	0.03	0.04	0.13

Error in relativistic correction as a percentage of error in total energy.

Basis	Ne	Ca^{10+}	Zr^{30+}	Hg^{70+}
$7s3p$	1.4	8.7	32.2	81.5
$9s5p$	3.9	32.3	61.8	95.5
$10s6p$	6.7	29.9	72.3	97.5
$12s7p$	10.3	43.4	82.7	98.9
$13s8p$	13.6	55.6	88.2	99.4

be necessary in the calculations on Hg^{79+} in order to obtain accurate core energies, may not be as critical for relativistic contributions to chemical binding. Indeed, Schwarz et al. (1989) comment that the principal differential screening effects that cause expansion or contraction of atomic valence orbitals due to relativity come from the valence and sub-valence shell, and not from orthogonality tails. This implies that a relatively poorer description of the core may not significantly affect calculated chemical properties, and that non-relativistic basis sets may be used in relativistic calculations without re-optimization. The results of Matsuoka and Okada (1989) on heavy atoms show that the valence eigenvalue errors are of the order of a few mE_h when a large nonrelativistic basis is used without modification, and the improvement on correction of the major deficiency in the core basis gives at most $1mE_h$ improvement. Moreover, the eigenvalue errors are all in the same direction, so that some cancellation of basis set truncation errors occurs in the relative energies. Further investigation of the use of non-relativistic basis sets is necessary, particularly in molecular calculations.

While there are no calculations for H_2O to which we can compare directly,

Table 9. Results for H_2 at R = 1.4 a_0. Energies in E_h, differences in μE_h.

Basis	E_{rel}	E_{NR}	ΔE_{rel}
$7s3p$	$-1.133\ 095\ 73$	$-1.133\ 081\ 64$	-14.0
numerical[a]	$-1.133\ 643\ 97$	$-1.133\ 629\ 57$	-14.4

[a]Sundholm (1988)

Table 10. Results for H_2O at 0.96Å, 104.5°. Energies in E_h, differences in mE_h.

Basis	E_{rel}	E_{NR}	ΔE_{rel}
$4s2p, 2s$	$-75.168\ 415$	$-75.120\ 506$	-47.909

the computed relativistic correction is reasonable when compared with the sum of the relativistic corrections for the atoms, which is dominated by the contribution of 56 mE_h from oxygen. The correction is consistent with the results obtained for atoms, that the relativistic correction to the energy is underestimated in basis set calculations, and converges from above on the true value as the basis set is enlarged.

CONCLUSIONS

In this article we have discussed some of the principles underlying finite basis Dirac-Fock calculations, and described how these have been implemented in a computer program. Although dealing mostly with atoms, our test calculations, as well as results obtained by others using similar approaches, clearly demonstrate the feasibility of carrying out high-quality Dirac-Fock calculations for molecules. The main obstacle to such calculations today appears to be the lengthy basis set expansions required. At the Dirac-Fock level the storage problem this creates can be largely overcome by resorting to direct SCF methods. We feel confident that, with further development, routine quantum chemical calculations of high accuracy for molecules containing heavy atoms will soon be a reality.

REFERENCES

Desclaux, J.-P., 1975, Comput. Phys. Commun. 9: 31.

Drake, G. W. F., and Goldman, S. P., 1982, Phys. Rev. A 25: 1230.

Dyall, K. G., Grant, I. P., Johnson, C. T., Parpia, F. A., and Plummer, E. P., 1989 Comput. Phys. Commun. 55: 425.

Dyall, K. G., Grant, I. P. and Wilson, S. 1984 J. Phys. B 17: L45, 493, 1201.

Fægri, K., 1987, Theor. Chim. Acta 72: 297.

Grant, I. P., McKenzie, B. J., Norrington, P. H., Mayers, D. F., and Pyper, N. C., 1980, Comput. Phys. Commun., 21: 207.

Grant 1986, J. Phys. B, 19: 3187.

Grant, I. P., and Quiney, H. M., 1988, Adv. At. Mol. Phys., 23: 37.

Hegarty, D., and Aerts, P. J. C., 1987, Physica Scripta, 36: 432.

Ishikawa, Y., Binning, R. C., and Sando, K. M., 1983, Chem. Phys. Lett., 101: 111.

Ishikawa, Y., and Sekino, H., 1990, Chem. Phys. Lett., 165: 243.

Laaksonen, L., and Grant, I. P., 1984, Chem. Phys. Lett., 109: 485, 112: 157.

Laaksonen, L., Grant, I. P., and Wilson, S., 1988, J. Phys. B, 21: 1969.

Lee, Y. S., and McLean, A. D., 1982, J. Chem. Phys., 76: 735.

Mark, F., Lischka, H., and Rosicky, F., 1980, Chem. Phys. Lett., 71: 507.

Matsuoka, O., and Okada, S., 1989, Chem. Phys. Lett., 155: 547.

Partridge, H., 1989, J. Chem. Phys., 90: 1043.

Pitzer, K. S., 1979, Acc. Chem. Res., 12: 271.

Pyykkö, P., 1978, Adv. Quantum Chem., 11: 353.

Pyykkö, P., and Desclaux, J.-P., 1979, Acc. Chem. Res., 12: 276.

Pyykkö, P., 1986, Lect. Notes Chem. 41.

Pyykkö, P., 1988, Chem. Rev., 88: 563.

Quiney, H. M., 1988, in: "Methods in Computational Chemistry, vol. 2," S. Wilson, ed., Plenum, New York.

Quiney, H. M., Grant, I. P. and Wilson, S., 1989, J. Phys. B, 22: L19.

Schwarz, W. H. E., van Wezenbeek, E. M., Baerends, E. J., and Snijders, J. G., 1989, J. Phys. B, 22: 1515.

Stanton, R. E., and Havriliak, S., 1984, J. Chem. Phys., 81: 1910.

Sundholm, D., Pyykkö, P., and Laaksonen, L., 1987, <u>Physica Scripta</u>, 36: 400.

Sundholm, D., 1988, <u>Chem. Phys. Lett.</u>, 149: 251.

Visscher, L., Aerts, P. J. C., and Visser, O., 1990, this volume.

Visser, O., Aerts, P. J. C., Hegarty, D., and Nieuwpoort, W. C., 1987, <u>Chem. Phys. Lett.</u>, 134: 34.