

## Methods for constructing Stewart atoms

A.T.B. Gilbert<sup>a</sup>, A.M. Lee<sup>b</sup>, P.M.W. Gill<sup>a,\*</sup>

<sup>a</sup>*School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

<sup>b</sup>*Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK*

### Abstract

We compare four methods for generating Stewart atoms, the spherically-symmetric nuclear-centred functions whose sum best fits a given electron density. We find that projecting a molecular density onto an atom-centred basis is a more subtle and difficult problem than is generally recognized and we conclude that new approaches, based on integral equations, may be more satisfactory than traditional projection methods. © 2000 Elsevier Science B.V. All rights reserved.

*Keywords:* Stewart atoms; Integral equations; Density models

### 1. Introduction

It is frequently beneficial to approximate a complicated function by a simple one. A well-chosen model can be advantageous both conceptually and computationally, shedding light on the most important features of a system and, at the same time, offering an efficient route to their calculation. It is therefore not surprising to find that a wide variety of models have been devised to treat the inherent complexities of molecular quantum mechanics. For example, the Hartree–Fock model approximates the molecular wavefunction  $\Psi$  by a relatively simple determinant and, in doing so, replaces an impossibly difficult computational problem by a tractable one and establishes the orbital picture at the centre of modern chemical thought.

More recently, the Hohenberg–Kohn theorem [1] has shown that the molecular electron density  $\rho(\mathbf{r})$  offers an even simpler paradigm for understanding electronic structure and a correspondingly efficient

framework for calculation. Density functional theory (DFT) is now widely embraced and many groups are working to extend its range of applicability to excited states and large systems.

Our interest lies in the next level of approximation, the construction of a model density  $\hat{\rho}(\mathbf{r})$  from a given density  $\rho(\mathbf{r})$ . This can be done in many ways but we see particular promise in the Stewart decomposition [2,3]. Given a density  $\rho(\mathbf{r})$  for a system with nuclei at  $\mathbf{A}_j$ ,  $j = 1, \dots, N$ , we seek the  $N$  Stewart atoms  $\sigma_j(r)$  whose sum, the Stewart density

$$\hat{\rho}(\mathbf{r}) = \sum_j \sigma_j(|\mathbf{r} - \mathbf{A}_j|) \quad (1)$$

best fits  $\rho(\mathbf{r})$  in a least-squares sense. The most obvious fit arises from minimizing

$$Z_0^\delta = \langle \rho - \hat{\rho} | \rho - \hat{\rho} \rangle \quad (2)$$

but one could also consider minimizing the more general least-squares residual

$$Z_0^\vartheta = \langle \rho - \hat{\rho} | \vartheta | \rho - \hat{\rho} \rangle \quad (3)$$

where  $\vartheta(r_{12})$  is a positive-definite two-electron

\* Corresponding author. Tel.: +44-115-951-3559; fax: +44-115-951-3562.

E-mail address: Peter.Gill@nottingham.ac.uk (P.M.W. Gill).

operator. There are a number of physically reasonable choices for  $\vartheta(r_{12})$  of which  $\delta(r_{12})/(4\pi r_{12}^2)$ ,  $r_{12}^{-1}$  and  $-r_{12}$  are the most important. These correspond [4] to fitting the density, the electric field modulus and the potential, respectively, of  $\hat{\rho}(\mathbf{r})$  to  $\rho(\mathbf{r})$ . However, Stewart has proven [2] the surprising result that the  $\sigma_j$  are, in fact, independent of  $\vartheta(r_{12})$  and that the choice of operator is thus immaterial. For this reason, we drop the superscript in  $Z_0^\vartheta$  henceforth. Stewart has also shown that  $\hat{\rho}(\mathbf{r})$  exactly reproduces some of the low-order multipole moments of  $\rho(\mathbf{r})$ , in particular its charge and dipole moments.

An approximate, but conceptually straightforward, method for generating Stewart atoms is to expand  $\hat{\rho}(\mathbf{r})$  in an auxiliary set of nuclear-centred radial basis functions, i.e.

$$|\hat{\rho}\rangle = c_p |\phi_p\rangle \quad (4)$$

and directly minimize Eq. (3). The expansion coefficients satisfy the matrix equation

$$\mathbf{S}\mathbf{c} = \mathbf{b} \quad (5)$$

$$S_{pq} = \langle \phi_p | \vartheta | \phi_q \rangle \quad (6)$$

$$b_p = \langle \phi_p | \vartheta | \rho \rangle \quad (7)$$

and this can be solved using standard techniques of linear algebra. Unfortunately, these approximate Stewart atoms *do* depend on the operator  $\vartheta(r_{12})$  used in Eqs. (6) and (7) and a choice must be made. This question has been considered by Dunlap et al. [5] and Eichhorn et al. [6] and these workers have recommended the use of the Coulomb operator.

We discuss the use of Hermite and Gaussian auxiliary basis functions in Sections 2 and 3, respectively. We will revisit the question of the “best” operator  $\vartheta(r_{12})$  in Section 6 but we use the overlap operator unless otherwise noted.

Following Stewart, our previous work [3] used the convolution theorem to move the minimization problem into Fourier space where the objective function becomes

$$Z_0 = \int F_\vartheta(x) |F_\rho(\mathbf{x}) - \sum F_j(x) e^{-i\mathbf{A}_j \cdot \mathbf{x}}|^2 d\mathbf{x} \quad (8)$$

where  $F_\vartheta$ ,  $F_\rho$  and  $F_j$  are the Fourier transforms of  $\vartheta(r_{12})$ ,  $\rho(\mathbf{r})$  and  $\sigma_j(r)$ , respectively.

Minimization of Eq. (8) then leads to a matrix

equation for the unknown  $F_j$

$$\mathbf{J}\mathbf{F} = \mathbf{P} \quad (9)$$

$$J_{ij}(x) = j_0(R_{ij}x) \quad (10)$$

$$P_j(x) = \int \rho(\mathbf{r}) j_0(|\mathbf{r} - \mathbf{A}_j|x) d\mathbf{r} \quad (11)$$

where  $j_0(t) = \sin(t)/t$  and  $R_{ij} = |\mathbf{A}_i - \mathbf{A}_j|$  and Cramer’s Rule yields the formal solution

$$\sigma_j(r) = \frac{1}{2\pi^2} \sum_k \int_0^\infty \frac{\hat{J}_{jk} P_k}{|\mathbf{J}|} x^2 j_0(rx) dx \quad (12)$$

where  $\hat{\mathbf{J}}$  is the matrix of cofactors of  $\mathbf{J}$ . However, despite many attempts, even the simplest such integrals have always proven an insuperable obstacle for us [3]. We consider a possible way forward in Section 4.

However, it is possible to avoid both auxiliary basis sets and Fourier transforms by transforming the Stewart Eqs. (9) into real space. This leads to a system of integral equations and Section 5 discusses one way in which these may be solved. Section 6 then compares the performance of the four methods on a simple model of the  $\text{H}_2^+$  molecule and Section 7 discusses results for the  $\text{H}_2$  and  $\text{F}_2$  molecules. Atomic units are used throughout.

## 2. Expansion in a Hermite basis

There are several reasonable bases in which one could expand the Stewart atoms. For example, the Hermite functions

$$\phi_p(r) = \frac{\pi^{-3/4}}{\sqrt{2^{p+1}p!}} \frac{H_p(r)}{r} e^{-r^2/2} \quad (13)$$

( $p$  odd) provide, for the overlap operator, an orthonormal basis of even functions on each centre. They are proportional to their Fourier transforms

$$\int_0^\infty 4\pi r^2 \phi_p(r) j_0(kr) dr = (-1)^{(p-1)/2} (2\pi)^{3/2} \phi_p(k) \quad (14)$$

and their integrals over all space are given by

$$\int_0^\infty 4\pi r^2 \phi_p(r) dr = \sqrt{8\pi}^{3/4} \frac{p!!}{\sqrt{p!}} \quad (15)$$

If  $\phi_p$  and  $\phi_q$  are separated by  $R$ , their overlap integral

Table 1  
Even-tempering parameters for H and N

n	Hydrogen		Nitrogen	
	$\alpha$	$\beta$	$\alpha$	$\beta$
15	0.10	1.39	0.17	1.70
20	0.09	1.29	0.14	1.51
25	0.08	1.23	0.12	1.40

(by the convolution theorem) is

$$S_{pq} = \frac{(-1)^{(p+q)/2-1}}{\sqrt{\pi 2^p p! 2^q q!}} \int_{-\infty}^{+\infty} H_p(x) H_q(x) \frac{\sin(Rx)}{Rx} e^{-x^2} dx$$

$$= \frac{2 e^{-R^2/4}}{\sqrt{2^p p! 2^q q!}} Y_{pq} \quad (16)$$

where we have introduced the polynomials

$$Y_{pq} = \frac{(-1)^{(p+q)/2-1} e^{R^2/4}}{2\sqrt{\pi}} \int_{-\infty}^{+\infty} (H_p(x) H_q(x) - H_p(0) H_q(0)) \frac{\sin(Rx)}{Rx} e^{-x^2} dx \quad (17)$$

The term  $H_p(0)H_q(0)$  in Eq. (17) vanishes when  $p$  and  $q$  are odd and is added for convenience. The  $Y_{pq}$  vanish if  $p + q$  is odd and the members between  $Y_{00}$  and  $Y_{44}$  can be tabulated as

$$\begin{bmatrix} 0 & 0 & 1 & 0 & R^2 + 6 \\ 0 & 1 & 0 & R^2 & 0 \\ 1 & 0 & R^2 - 2 & 0 & R^4 - 6R^2 + 12 \\ 0 & R^2 & 0 & R^4 - 8R^2 + 24 & 0 \\ R^2 + 6 & 0 & R^4 - 6R^2 + 12 & 0 & R^6 - 18R^4 + 108R^2 - 120 \end{bmatrix} \quad (18)$$

Explicit formulae for the higher  $Y_{pq}$  are complicated and unsuitable for numerical evaluation. However, recurrence properties of the Hermite polynomials yield the useful identity

$$Y_{p+1,q} - 2pY_{p-1,q} = Y_{p,q+1} - 2qY_{p,q-1} \quad (19)$$

which, starting from the trivial values

$$Y_{0,0} = 0 \quad (20)$$

$$Y_{p,1} = Y_{1,p} = \begin{cases} R^{p-1} & p \text{ odd} \\ 0 & p \text{ even} \end{cases} \quad (21)$$

offers an efficient route to the recursive generation of as many  $Y_{pq}$  as one requires. We note that some care is required, because Eq. (19) becomes numerically unstable for large  $p$  and  $q$ .

The elements of the  $\mathbf{b}$  vector can also be computed using the Convolution Theorem. The fundamental integral, from which all others can be derived [7], is the overlap of the Hermite function Eq. (13) with a Gaussian function at a distance  $R$ , and is given by

$$b_p = \frac{(-1)^{(p-1)/2}}{\pi^{5/4} R \sqrt{2^p p!}} \int_0^\infty H_p(x) \times \exp \left[ -x^2 \left( \frac{1}{2} + \frac{1}{4\zeta} \right) \right] \sin(Rx) dx \quad (22)$$

where  $\zeta$  is the Gaussian exponent. This integral is easily solved in closed form [8].

Finally, we note that the canonical functions (13) can be generalized by appropriately scaling its gaussian and the Hermite polynomials to form

$$\phi_p^\gamma(r) = \gamma^{3/2} \phi_p(\gamma r). \quad (23)$$

### 3. Expansion in a Gaussian basis

Auxiliary basis sets employing Gaussian functions

$$\phi_p(r) = (\zeta_p/\pi)^{3/2} \exp(-\zeta_p r^2) \quad (24)$$

are widely used because of their computational

Table 2  
Coefficients  $c_{mp}$  for the (3, $p$ ) Fourier approximations

$p$	$c_{3,p}$
0	3.0265106465
1	-0.9409081880
2	0.3394151003
3	-0.1202380179
4	0.0395452865
5	-0.01111605698
6	0.0023707571
7	0.0000323428

simplicity. Their Fourier transforms

$$\int_0^\infty 4\pi r^2 \phi_p(r) j_0(kr) dr = \exp\left(-\frac{k^2}{4\zeta_p}\right) \quad (25)$$

are simple and their integrals over all space are unity.

We use even-tempered basis sets optimized for hydrogen and nitrogen. The exponents

$$\zeta_p = \alpha\beta^{p-1}, \quad p = 1, 2, \dots, n \quad (26)$$

are given in Table 1 for  $n = 15, 20$  and  $25$ . The  $\alpha$  and  $\beta$  parameters were chosen so that the Coulomb energy of the STO-3G atom is reasonably approximated [9], the geometric mean of the exponents is close to the nuclear charge, and the  $\alpha$  values decrease with increasing  $n$ . For calculations with other first-row atoms, we use the nitrogen parameters.

If  $\phi_p$  and  $\phi_q$  are separated by  $R$ , their overlap integral is given by

$$S_{pq} = (z/\pi)^{3/2} \exp(-zR^2) \left[ z = \frac{\zeta_p \zeta_q}{\zeta_p + \zeta_q} \right] \quad (27)$$

Because the overlap operator is very short-ranged, the  $\mathbf{S}$  matrix becomes sparse in large systems and this admits the possibility of using special techniques for solving Eq. (5) efficiently. However, if the auxiliary basis is near-complete, some of the exponents  $\zeta_p$  will be small and  $\mathbf{S}$  becomes almost singular. This causes significant numerical problems and it is essential to employ Singular Value Decomposition (SVD) to remove the offending subspace. All of the results that we present used SVD with singular values determined by machine precision.

If  $\rho(\mathbf{r})$  consists of Gaussian functions, the elements of  $\mathbf{b}$  are analogous to Eq. (27).

#### 4. Expansion of the Fourier solution

The Stewart–Fourier integral (12) proves intractable even for a simple systems. For a diatomic with bond length  $R$ , the first Stewart atom is given by

$$\sigma_1(r) = \frac{1}{2R^3\pi^2} \int_0^\infty \frac{P_1(x/R) - j_0(x)P_2(x/R)}{1 - j_0^2(x)} x^2 j_0\left(\frac{rx}{R}\right) dx \quad (28)$$

and the transcendental denominator  $1 - j_0^2(x)$  thwarts the analytic evaluation of this integral. However, the identity

$$\frac{x^2}{1 - j_0^2(x)} \equiv \left[ x^2 \sum_{k=0}^m j_0^{2k}(x) \right] + \frac{\sin^2(x) j_0^{2m}(x)}{1 - j_0^2(x)} \quad (29)$$

breaks the integral into a sum of tractable integrals and a small correction. The last term in Eq. (29) is a bell-shaped positive function with an oscillatory tail. As  $m$  increases, the frequency of the oscillations increases but their amplitudes rapidly decrease. For a given  $m$ , this term may be expanded in an auxiliary basis of even Hermite functions and truncation of the expansion at order  $n$  yields the ( $m, n$ ) approximation

$$\begin{aligned} \sigma_1(r) = & \frac{1}{2\pi^2 R^3} \sum_{k=0}^m \int_0^\infty x^2 j_0^{2k}(x) \\ & \times \left[ P_1\left(\frac{x}{R}\right) - j_0(x)P_2\left(\frac{x}{R}\right) \right] j_0\left(\frac{rx}{R}\right) dx \\ & + \frac{1}{2\pi^2 R^3} \sum_{p=0}^n \int_0^\infty c_{mp} e^{-x^2/2} H_{2p}(x) \\ & \times \left[ P_1\left(\frac{x}{R}\right) - j_0(x)P_2\left(\frac{x}{R}\right) \right] j_0\left(\frac{rx}{R}\right) dx \quad (30) \end{aligned}$$

to the Stewart atom  $\sigma_1(r)$ .

If  $\rho(\mathbf{r})$  is composed of Gaussian functions, all of the integrals in Eq. (30) can be found in closed form. The expansion coefficients  $c_{mp}$  must be determined numerically, but are valid for all diatomics because  $R$  does not occur in Eq. (29). The values required for the (3, $p$ ) approximations ( $p \leq 7$ ) are listed in Table 2. Unfortunately, it is difficult to generalize this method to polyatomic systems.

Table 3  
Stewart function  $f(r)$  for model  $H_2^+$  and errors ( $\times 10^6$ ) of various methods

$r$	$f(r)$	Hermite		Gaussian		Fourier <sup>a</sup>		Simpson	
		$n = 10$	$n = 20$	$n = 15$	$n = 20$	$n = 4$	$n = 5$	$n = 8$	$n = 16$
0.25	0.0154642	+6.7	+0.5	+4.3	+0.6	+0.8	-0.1	+1.1	+0.1
0.50	0.0284867	-1.1	-0.6	-4.2	+0.3	+1.2	-0.2	+1.8	+0.1
0.75	0.0367769	-6.7	+0.2	+1.6	-0.7	+1.3	-0.2	+1.7	+0.1
1.00	0.0387705	+1.9	+0.3	-0.7	+0.2	+1.0	0.0	+0.9	+0.1
1.25	0.0344114	+6.7	-0.6	+1.3	+0.4	+0.2	0.0	0.0	0.0
1.50	0.0254354	-2.7	+0.5	-3.5	-0.7	-0.6	+0.1	-0.7	0.0
1.75	0.0147369	-7.1	0.0	+6.7	+0.6	-1.3	+0.1	-1.1	-0.1
2.00	0.0051586	+2.9	-0.5	-7.1	-0.3	-1.6	+0.2	-1.2	-0.1
3.00	-0.0043633	+1.7	+0.7	-9.3	-0.4	+1.3	-0.2	+0.4	0.0
4.00	0.0009188	-7.0	-0.2	-4.2	-0.7	+1.0	+0.2	+0.4	0.0
5.00	0.0001562	-12.9	-0.7	+14.2	-2.4	-4.6	+0.1	-0.5	0.0
6.00	-0.0001484	+25.8	-0.3	-8.9	+2.4	+1.9	-0.9	+0.2	0.0
7.00	0.0000327	-40.7	-0.5	-18.5	-4.1	+13.4	+0.8	0.0	0.0

<sup>a</sup> Using  $m = 3$  in Eq. (30).

## 5. Simpson approximation

The Stewart decomposition is a spherical projection of  $\rho(\mathbf{r})$  about the nuclei and, thus,

$$S_a \tilde{\rho} = S_a \rho \quad (31)$$

where  $S_a$  is the spherical projector about the  $a$ th nucleus. It therefore follows that

$$\sigma_a + \sum_{b \neq a} S_a \sigma_b = S_a \rho \quad (32)$$

If the Stewart atom  $\sigma_b$  is even, it can be shown that its projection onto another nucleus is

$$S_a \sigma_b = \frac{1}{2R_{ab}r} \int_{r-R_{ab}}^{r+R_{ab}} t \sigma_b(t) dt \quad (33)$$

Defining Stewart functions  $f_a(r)$ , local averages  $\langle f \rangle_R$  and the density projections  $g_a(r)$

$$f_a(r) = r \sigma_a(r) \quad (34)$$

$$\langle f \rangle_R = \frac{1}{2R} \int_{r-R}^{r+R} f(t) dt \quad (35)$$

$$g_a(r) = r S_a \rho \quad (36)$$

then leads immediately to the simultaneous integral equations

$$f_a(r) + \sum_{b \neq a} \langle f_b \rangle_{R_{ab}} = g_a(r). \quad (37)$$

Although equivalent to the Fourier framework that Stewart [2] and we [3] have used, the real-space formulation (37) of Stewart theory is simpler and more physically appealing. The  $g_a(r)$  are straightforward to calculate, especially if  $\rho(\mathbf{r})$  consists of Gaussian functions, and there are several approaches for treating linear integral equations of the form (37). We will discuss some of these elsewhere [10] but, in this paper, approximate the local averages using quadrature formulae and thereby discretize the problem.

The Extended Simpson Rule approximates the local average Eq. (35) by

$$\langle f \rangle_R \approx \frac{1}{6n} [f_0 + 4(f_1 + f_3 + \dots + f_{2n-1}) + 2(f_2 + f_4 + \dots + f_{2n-2}) + f_{2n}] \quad (38)$$

$$f_k = f(r + (k - n)h) \quad (39)$$

$$h = R/n \quad (40)$$

and introduces an error [11]

$$\text{Error} = \frac{nh^5}{90} f^{(4)}(\xi) \quad [r - R \leq \xi \leq r + R] \quad (41)$$

which is acceptable if the step size  $h$  is small and the function  $f(r)$  does not vary rapidly. Furthermore, doubling  $n$  (and thereby halving  $h$ ) reduces this error by an order of magnitude.



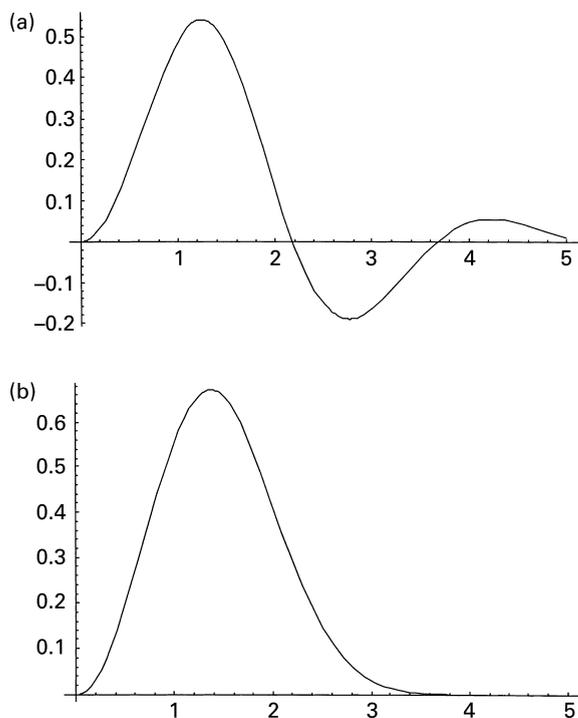


Fig. 1. (a)  $4\pi r f(r)$  and (b)  $4\pi r g(r)$  in model  $H_2^+$ .

approximation would degrade rapidly for larger values of  $r$ .

Of course, we could obtain more accurate expansions by employing a basis in which the Hermite polynomials are scaled for the specific problem under consideration. However, we have not exploited this possibility here because we are interested to now how well the unmodified basis functions (13) perform.

## 6.2. Gaussian expansion

The matrix equation that results when this Stewart problem is expanded in the even-tempered Gaussian basis (24) for hydrogen (see Table 1) is dense but quite well conditioned. Consequently, it is straightforward to solve for the expansion coefficients and the fifth and sixth columns of Table 3 show the errors incurred if  $\sigma(r)$  is expanded in bases of 15 and 20 gaussians, respectively.

The  $n = 15$  approximation is comparable to the  $n = 10$  Hermite approximation, although it seems to deteriorate slightly more slowly at large  $r$  values. Their similarity can probably be traced to the fact

that the most diffuse gaussian ( $\zeta_1 = 0.1$ ) has roughly the same extent as the most diffuse Hermite function  $\phi_{19}$ .

The  $n = 20$  approximation is comparable to the  $n = 20$  Hermite approximation for  $r \leq 4$  but deteriorates more quickly at larger  $r$ . As before, this can be explained by observing that the most diffuse gaussian ( $\zeta_1 = 0.09$ ) is significantly more compact than the Hermite  $\phi_{39}$ .

Like the Hermite results above, these Gaussian expansions were obtained by using the overlap operator  $\vartheta_O(r_{12}) = \delta(r_{12})/(4\pi r_{12}^2)$  to compute the matrix elements Eqs. (6) and (7). This operator ignores all contributions to the integrand of Eq. (3) except those where  $r_{12} = 0$ . The unit operator  $\vartheta_U(r_{12}) \equiv 1$ , which attaches equal importance to all values of  $r_{12}$ , lies at the other extreme. The Coulomb operator  $\vartheta_C(r_{12}) = r_{12}^{-1}$ , which gives reduced but significant weighting to contributions with large  $r_{12}$  lies between them and, as we noted in Section 1, has been advocated by workers such as Dunlap et al. [5] and Eichhorn et al. [6]. We have also made use [9,12] of the ‘‘short-range Coulomb’’ operator  $\vartheta_S(r_{12}) = r_{12}^{-1} \operatorname{erfc}(\omega r_{12})$  and the ‘‘long-range Coulomb’’ operator  $\vartheta_L(r_{12}) = r_{12}^{-1} \operatorname{erf}(\omega r_{12})$ , where  $0 < \omega < \infty$ . It is thus natural to ask whether the Gaussian results in Table 3 can be improved by abandoning  $\vartheta_O$  in favour of one of these alternatives.

Table 4 shows the errors that result from expanding

Table 4

Stewart function  $f(r)$  for model  $H_2^+$  and errors ( $\times 10^6$ ) of various operators (using Gaussian expansion and the  $n = 20$  even-tempered basis; using  $\omega = 0.5$  for the  $\vartheta_S$  and  $\vartheta_L$  operators; and the fitting equations for the unit operator  $\vartheta_U$  cannot be solved)

$r$	$f(r)$	$\vartheta_O$	$\vartheta_S$	$\vartheta_C$	$\vartheta_L$
0.25	0.0154642	+0.6	+10.4	-15.5	+2670000
0.50	0.0284867	+0.3	+2.9	-2.9	-2179000
0.75	0.0367769	-0.7	-11.5	+16.7	+191100
1.00	0.0387705	+0.2	+6.5	-12.3	+449700
1.25	0.0344114	+0.4	+0.2	+5.6	-24460
1.50	0.0254354	-0.7	-2.6	-4.4	-102100
1.75	0.0147369	+0.6	+1.0	+7.3	-27510
2.00	0.0051586	-0.3	+2.9	-11.6	+15480
3.00	-0.0043633	-0.4	-7.1	+8.7	-3499
4.00	0.0009188	-0.7	-6.0	+3.2	+688
5.00	0.0001562	-2.4	-6.1	-9.7	-189
6.00	-0.0001484	+2.4	+7.6	+9.2	+64
7.00	0.0000327	-4.1	-7.7	-10.8	-19

$\sigma(r)$  in the  $n = 20$  Gaussian basis using the  $\vartheta_O$ ,  $\vartheta_S$ ,  $\vartheta_C$  and  $\vartheta_L$  operators. The performance of  $\vartheta_O$  is clearly superior to that of  $\vartheta_S$  and  $\vartheta_C$ ; the  $\vartheta_L$  operator is very unsatisfactory; the  $\vartheta_U$  operator yields a singular  $\mathbf{S}$  matrix and is entirely incapable of discriminating between different basis functions. In fact, the ranking of the operators correlates exactly with the extent to which they are short-ranged

$$\vartheta_U < \vartheta_L < \vartheta_C < \vartheta_S < \vartheta_O \quad (49)$$

and leaves little doubt that the overlap operator is the best one to use for this problem.

### 6.3. Fourier expansion

The methods discussed above expand the Stewart atom in a basis of simple functions. Although this approach is reasonable, the fact that the exact  $f(r)$  is sinusoidal with exponential decay suggests that neither Hermite nor Gaussian bases should be optimal for this problem. One could explore alternative bases but we have not done so here.

Instead of expanding the Stewart atom itself, we now expand its Fourier transform  $F(x)$ . It is easy to show that the elements of the  $\mathbf{P}$  vector Eq. (11) are

$$P_1(x) = P_2(x) = j_0(x) e^{-x^2/4} \quad (50)$$

and substituting this into Eq. (30) yields integrals that are tedious but evaluable in closed form. Following preliminary investigations, we chose to fix  $m = 3$  and explore a range of  $n$  values. The errors that arise for  $n = 4$  and  $n = 5$  are listed in columns 7 and 8 of Table 3.

The Eqs. (3) and (4) approximation yields almost six decimal places for  $r \leq 4$  and is almost as accurate as the  $n = 20$  Hermite or Gaussian approximations over this range. It deteriorates for larger  $r$  values but remains respectable even at  $r = 7$ .

As  $n$  is increased, the accuracy improves rapidly. There is an improvement of almost an order of magnitude in moving from (3) and (4) to the (3) and (5) approximation and, moreover, the (3) and (6) approximation agrees with the exact  $f(r)$  for all of the decimal places given in Table 3.

This approach converges more quickly than the Hermite or Gaussian methods, implying that our expansion basis for  $F(x)$  is superior to those we have used for  $\sigma(r)$ . The main drawback of the

approach for this problem is that the integrals required in Eq. (30) are messy.

### 6.4. Simpson approximation

Our fourth method, which is based on the real-space integral equation formulation of Stewart theory, does not expand the Stewart atom in a basis and is thus free of the bias that any such basis presents. Its weakness, however, is the error Eq. (41) that is introduced when the integral (35) is replaced by the Simpson approximation (38).

Substituting Eq. (43) into Eq. (37), we find that the Stewart function  $f(r)$  satisfies

$$f + \langle f \rangle_2 = g \quad (51)$$

which is a Fredholm integral equation with kernel of convolution type. The right-hand side is

$$g(r) = \frac{1}{4} [\rho(r-1) - \rho(r+1)] \quad (52)$$

and  $4\pi r g(r)$  is plotted in Fig. 1b. We have solved Eq. (51) and find that its solution kernel is a surprisingly complicated function. We will discuss these results in greater detail elsewhere [10]. However, application of the Simpson approximation yields the matrix equation

$$(\mathbf{I} + \mathbf{A})\mathbf{f} = \mathbf{g} \quad (53)$$

$$\mathbf{f} = [f(h), f(2h), \dots, f(mh)]^T \quad (54)$$

$$\mathbf{g} = [g(h), g(2h), \dots, g(mh)]^T \quad (55)$$

We have  $nh = R = 2$  from Eq. (40) and, following preliminary investigations, we chose  $mh = 10$ . We explored a range of  $n$  and some of our results are summarized in Table 3.

Although one might have thought that a stepsize of  $h = 0.25$  is fairly crude, the  $n = 8$  approximation is surprisingly accurate, yielding small errors in the sixth decimal place and, unlike previous methods, not deteriorating at larger  $r$ . Reducing the step size to  $h = 0.125$  improves the accuracy by approximately an order of magnitude, as anticipated from Eq. (41) and the  $n = 16$  errors are never more than  $10^{-7}$  and usually less. The results found for  $n = 32$  agree to seven decimal places with our best Fourier values and are shown in the second column.

## 7. Other examples

The foregoing Section contains a detailed discussion of results for the simple  $H_2^+$  ion. However, since the electron density was a single gaussian centred at the bond midpoint, it is reasonable to ask whether our conclusions would also apply to complex molecular densities. Although a full examination of this question is beyond the scope of this preliminary paper, we have obtained results for the  $H_2$  and  $F_2$  molecules and these shed some light on the matter.

### 7.1. $H_2$ molecule

Unlike our single-gaussian  $H_2^+$  density, the HF/STO-3G density for the  $H_2$  molecule ( $R = 1.39304175$ ) is a qualitatively correct approximation to the experimental electron density and it is of interest to discover whether or not the HF/STO-3G Stewart atoms for  $H_2$  possess the slow, oscillatory decay seen in Fig. 1a.

It is helpful to recall that the HF/STO-3G density is a sum of 21 gaussians. Six of these are centred on one nucleus, six on the other, and the remaining nine at various points along the bond axis. The six centred on each nucleus contribute directly to their respective Stewart atoms and collectively account for 1.20 of the electrons. Our main interest therefore lies in the

Stewart decomposition of the bond-centred gaussians that bear the other 0.80 electrons. We have applied the Hermite, Gaussian and Simpson (but not Fourier) schemes to the problem and tabulations of the resulting  $f(r)$  for various  $n$  are summarized in Table 5.

Comparison of the Hermite columns of Table 3 with those of Table 5 reveals that the method is significantly less accurate for  $H_2$  than it was for our single-gaussian  $H_2^+$  density. Even using  $n = 35$ , the Hermite results for  $H_2$  are correct to only three or four significant figures and the errors are an order of magnitude larger for  $n = 30$ . We attribute this impaired performance to exponent inhomogeneity: whereas the exponent in the  $H_2^+$  density Eq. (43) is 1.0, the gaussians in the  $H_2$  density have exponents ranging from 0.34 to 6.9 and thus place much heavier demands on the canonical Hermite functions (13). The use of basis functions that all share the same exponent appears to be a fundamental weakness and, although a scaled basis would probably perform somewhat better, the essential problem would remain.

In contrast, our Gaussian basis sets enjoy a wide spread of exponents and the Gaussian columns of Table 5 tell a much happier story than the Hermite ones. The errors for  $H_2$  are at least as small as those for  $H_2^+$ , reflecting the fact that the even-tempered exponents are ideally suited for use with the STO-3G orbital basis. Although the numerical problems mentioned

Table 5  
Stewart function  $f(r)$  for HF/STO-3G  $H_2$  and errors ( $\times 10^6$ ) of various methods

R	$f(r)^a$	Hermite <sup>b</sup>		Gaussian		Simpson	
		$n = 30$	$n = 35$	$n = 15$	$n = 20$	$n = 16$	$n = 32$
0.25	0.0690514	-17.4	+4.7	+1.1	+0.2	-18.1	-2.0
0.50	0.0850815	+34.8	-7.3	-1.2	+0.1	+3.0	+0.4
0.75	0.0710581	-50.9	+7.2	+0.7	-0.2	+9.7	+0.4
1.00	0.0540489	+63.6	-5.0	-0.1	+0.1	+0.2	0.0
1.25	0.0383139	-70.6	+1.5	0.0	+0.1	-1.3	-0.1
1.50	0.0241631	+70.6	+2.3	-0.7	-0.2	0.0	0.0
1.75	0.0135719	-63.4	-6.1	+2.3	+0.3	+0.5	0.0
2.00	0.0071588	+49.8	+9.5	-3.2	-0.2	+0.2	0.0
3.00	0.0012417	-24.7	+17.5	-0.4	-0.1	-0.1	0.0
4.00	0.0000901	-12.3	+10.3	-0.5	+0.8	0.0	0.0
5.00	0.0000087	+22.5	-8.6	-1.5	-0.3	0.0	0.0
6.00	-0.0000001	-21.5	-9.8	+1.1	-0.1	0.0	0.0
7.00	0.0000000	+12.3	-10.2	-0.7	-0.1	0.0	0.0

<sup>a</sup> Using Simpson with  $n = 128$ .

<sup>b</sup> Using unscaled ( $\gamma = 1.0$ ) basis functions.

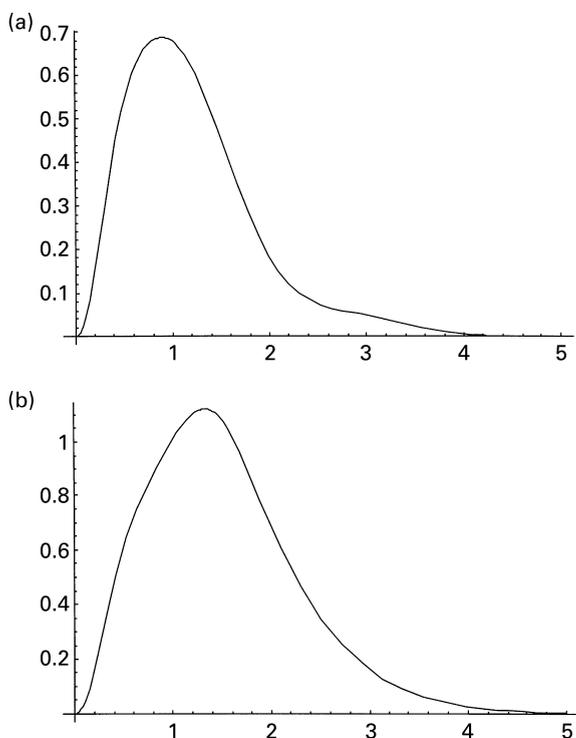


Fig. 2. (a)  $4\pi r f(r)$  and (b)  $4\pi r g(r)$  in  $H_2$ .

in Section 3 impede significant improvement beyond the errors in Table 5, the results for  $n = 20$  are already satisfactory.

Finally, we turn to the Simpson errors in Table 5. The curvature of the STO-3G density is large in the neighbourhood of each nucleus and the Stewart atoms inherit this feature. This leads to a degradation in the accuracy of the Simpson approximation (38) for small  $r$  values and explains why the largest errors occur there. The  $n = 16$  approximation is accurate beyond  $r = 2$  and the  $n = 32$  and  $n = 64$  results yield a satisfactory treatment of even the inner region. The  $n = 128$  results have converged to seven decimal places and are listed as the  $f(r)$  values.

Graphs of  $4\pi r f(r)$  and  $4\pi r g(r)$  are shown in Fig. 2. The second of these, the spherical projection of  $\rho(\mathbf{r})$  about one of the nuclei, clearly shows the bimodal character of the density. The first reveals that the Stewart atoms in  $H_2$  decay rapidly and without visible oscillations. It is interesting to observe that the oscillations of Fig. 1a are replaced by a plateau in Fig. 2a. If the Stewart atom does indeed have nodes, as Table 5 suggests somewhat unconvincingly, they are of little significance.

## 7.2. $F_2$ molecule

As a final example, we have sought to extract the

Table 6  
Stewart function  $f(r)$  for HF/STO-3G  $F_2$  and errors ( $\times 10^6$ ) of various methods

$r$	$f(r)^a$	Hermite <sup>b</sup>		Gaussian		Simpson	
		$n = 25$	$n = 35$	$n = 20$	$n = 25$	$n = 64$	$n = 128$
0.25	1.617512	+36108	-652	+97	+13	+473	+57
0.50	0.779175	+7928	+10766	-292	-81	-45	-3
0.75	0.729563	-22186	+4182	+50	-123	-6	0
1.00	0.506212	-15862	-1700	+1867	-107	-5	0
1.25	0.275214	-7080	+263	-2549	+322	-4	0
1.50	0.126623	+5295	+1868	+2296	-417	-5	0
1.75	0.049803	-40562	-2412	-293	+728	-5	0
2.00	0.011419	-11416	-4313	-3342	-1151	-4	0
3.00	0.011986	-11986	-11986	-350	+1712	-6	0
4.00	-0.000644	+644	+644	-1568	+50	+3	0
5.00	0.001349	+1349	+1349	+1024	-34	+6	0
6.00	0.000243	-243	-243	+48	464	+1	0
7.00	0.000156	-156	-156	-54	+550	+1	0

<sup>a</sup> Using Simpson with  $n = 256$ .

<sup>b</sup> Using scaled ( $\gamma = 6.5$ ) basis functions.

Stewart atoms from the HF/STO-3G density for the  $F_2$  molecule. This is a homonuclear diatomic, like  $H_2$ , but possesses both core and  $p$  electrons and a bond length ( $R = 2.6845617$ ) that is twice as large as that in  $H_2$ .

The HF/STO-3G density of  $F_2$  consists of 465 gaussians, with exponents ranging from 1.0 to 333. Of these, 120 are centred on each nucleus and the remaining 225 at points along the bond axis. The former hold 17.76 of the electrons and the latter the other 0.24 electrons. As before, we have applied the Hermite, Gaussian and Simpson schemes to this Stewart problem and tabulations of the resulting  $f(r)$  for various  $n$  are summarized in Table 6.

We were unable to obtain satisfactory results using canonical Hermite functions (13) and so resorted to the scaled functions (23). Exploratory fitting to an F atom led us to select  $\gamma = 6.5$  but there were indications that this is a strained compromise between the requirements of the core and valence electrons. The Hermite columns of Table 6 show that even the scaled basis functions yield poor approximations to the Stewart functions and it is difficult to obtain more than two significant figures. Beyond  $r = 2$ , the Hermite models are useless.

The Gaussian results are superior to the Hermite ones—presumably because the range of exponents is greater—but they are far from satisfactory nonetheless. With  $n = 25$  functions the errors are reasonable at small  $r$  but grow significantly near to the interesting third peak at  $r = 2.8$  in Fig. 3a. This feature, which arises from the corresponding spike in Fig. 3b, allows each Stewart atom to contribute to the description of the core around the other nucleus. However, both the Hermite and Gaussian models struggle to capture this phenomenon and it is evidently difficult for expansion functions around one nucleus to treat a compact region of high density around the other nucleus.

Although the presence of the core electrons means that we must use larger  $n$  values than were required for the  $H_2^+$  and  $H_2$  cases, the Simpson method is clearly the best of the three for this molecule. As we observed for  $H_2$ , the method is least satisfactory for small  $r$  values where the curvature of the density is large but performs well for moderate and large  $r$ . The  $f(r)$  values in the second column of Table 6, and the plot of  $4\pi r f(r)$  shown in Fig. 3a, were obtained using

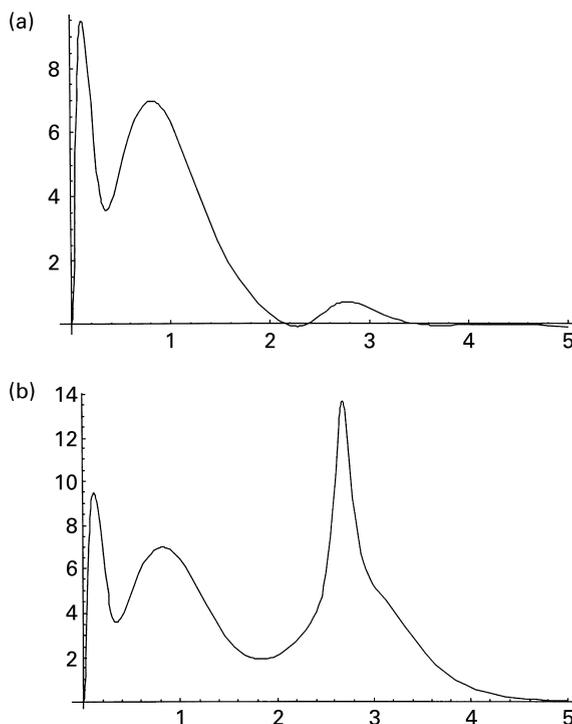


Fig. 3. (a)  $4\pi r f(r)$  and (b)  $4\pi r g(r)$  in  $F_2$ .

Simpson with  $n = 256$ . The Stewart atom seems to have nodes but its negative regions are scarcely significant.

## 8. Conclusions

This paper is a preliminary exploration of some of the methods that can be used to construct Stewart atoms from molecular electron densities. These are useful both in their own right [2,3] and because they illuminate the more general problem [5,6] of projecting a density onto an atom-centred auxiliary basis set. We find that such projection problems are more subtle and difficult than is often recognized because the exact projected densities are often surprisingly complicated and consequently awkward for standard auxiliary bases to model. Traditional expansion techniques are plagued by numerical difficulties and our results suggest that new approaches based on integral equations may be more satisfactory both conceptually and computationally.

We are investigating this proposal and will present a detailed analysis elsewhere [10].

### Acknowledgements

We thank Stephen Taylor (Auckland) and Nikhil Nair and Simon Andrews (Cambridge) for a number of stimulating discussions about Eq. (51). This work was partly supported by the UK Engineering and Physical Sciences Research Council (Grants GR/M09483 and GR/M63430).

### References

- [1] P. Hohenberg, W. Kohn, Phys. Rev. 136 (1964) B864.
- [2] R.F. Stewart, Isr. J. Chem. 16 (1977) 124.
- [3] P.M.W. Gill, J. Phys. Chem. 100 (1996) 15421.
- [4] P.M.W. Gill, B.G. Johnson, J.A. Pople, S.W. Taylor, J. Chem. Phys. 96 (1992) 7178.
- [5] B.I. Dunlap, J.W.D. Connolly, J.R. Sabin, J. Chem. Phys. 71 (1979) 3396 (see also page 4993).
- [6] K. Eichhorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, Chem. Phys. Lett. 240 (1995) 283.
- [7] P.M.W. Gill, Adv. Quant. Chem. 25 (1994) 141.
- [8] I.S. Gradshteyn, I.M. Ryzhik, Table of integrals, series, and products, Academic, New York, 1994.
- [9] A.M. Lee, P.M.W. Gill, Chem. Phys. Lett. 286 (1998) 226.
- [10] A.T.B. Gilbert, S.W. Taylor, P.M.W. Gill, in preparation.
- [11] M. Abramowitz, I.A. Stegun (Eds.), Handbook of Mathematical Functions Dover, New York, 1972.
- [12] J.P. Dombroski, S.W. Taylor, P.M.W. Gill, J. Phys. Chem. 100 (1996) 6272.