

Regular article

Towards an order- N DFT method

C. Fonseca Guerra¹, J.G. Snijders^{2,1}, G. te Velde¹, E.J. Baerends¹

¹Afdeling Theoretische Chemie, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

²Laboratorium voor Chemische Fysica, Rijksuniversiteit Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received: 30 March 1998 / Accepted: 6 July 1998 / Published online: 15 September 1998

Abstract. One of the most important steps in a Kohn-Sham (KS) type density functional theory calculation is the construction of the matrix of the KS operator (the “Fock” matrix). It is desirable to develop an algorithm for this step that scales linearly with system size. We discuss attempts to achieve linear scaling for the calculation of the matrix elements of the exchange-correlation and Coulomb potentials within a particular implementation (the Amsterdam density functional, ADF, code) of the KS method. In the ADF scheme the matrix elements are completely determined by 3D numerical integration, the value of the potentials in each grid point being determined with the help of an auxiliary function representation of the electronic density. Nearly linear scaling for building the total Fock matrix is demonstrated for systems of intermediate size (in the order of 1000 atoms). For larger systems further development is desirable for the treatment of the Coulomb potential.

Key words: Linear scaling – Density functional theory – ADF program

1 Introduction

There is currently much interest in the development of $O(N)$ methods (linear scaling of computation time with system size) for both Hartree-Fock (HF) and Kohn-Sham (KS) calculations. In contrast to most other current schemes for KS calculations, the Amsterdam density functional code [1] (ADF) uses 3D numerical integration for the evaluation of the “Fock” matrix elements throughout (in spite of some historical and logical inconsistency, we denote the matrix of the KS one-electron operator simply as Fock matrix). The use of 3D numerical integration also for the elements of the

electronic Coulomb potential is particular to the ADF code. It is the purpose of this contribution to discuss the way in which linear scaling can be achieved in setting up the Fock matrix in this method. In this introduction we will first briefly consider the scaling of various steps in the ADF method and compare this with alternative density functional theory (DFT) methods. In order to achieve linear scaling we need to define a finite spatial region over which a basis function can be considered to be nonzero. A definition of the spatial cut-off for the basis functions that takes into account the effect of the cut-off on the matrix elements to be constructed will be discussed in Sect. 2. Section 3 treats the linearization of the density fitting, Sect. 4 the linearization of the evaluation of the exchange-correlation and Coulomb potentials in the grid points, Sect. 5 the building of the Fock matrix and Sects. 6 and 7 offer results and conclusions respectively.

The matrix elements of the exchange-correlation (XC) potential and the Coulomb potential are both evaluated by numerical integration, using P integration points

$$G_{\mu\nu} = \langle \chi_{\mu} | \hat{V}_{xc} + \hat{V}_{Coul} | \chi_{\nu} \rangle \approx \sum_{k=1}^P \chi_{\mu}^*(\mathbf{r}_k) (V_{xc}(\mathbf{r}_k) + V_{Coul}(\mathbf{r}_k)) \chi_{\nu}(\mathbf{r}_k) w_k \quad (1.1)$$

Using 3D numerical integration in this way has the advantage that it is possible to use almost any type of basis function, although there may be some restrictions arising from the required evaluation of $V_{Coul}(\mathbf{r}_k)$ at the grid points (the Poisson solver may or may not use the basis functions). STOs have been chosen as a convenient and physically well-motivated basis. Apart from enabling the one-electron wavefunctions and the total density to have the correct behaviour in the tail and at the nucleus, they are also efficient since their number can typically be a factor of 3 lower than the number of Gaussians. Another choice that has been made is numerical atomic orbitals (NAOs, also called single-site orbitals, SSOs) [2–6] With N basis functions, there is an NP step in the calculation for the evaluation of the basis

functions at all points, assuming for the time being that no cut-offs are applied. The value of the XC potential at each point simply follows, with the current approximate functionals, from the density and its gradient at that point. The density can be calculated either through the density matrix, yielding an N^2P step, or in terms of the occupied orbitals, yielding an N_eNP step [7]. Here N_e is the number of electrons. Of course not every size parameter is equal, in particular for large molecules one usually has

$$P \gg N' > N > N_e \quad (1.2)$$

where N' is the number of auxiliary functions to be discussed below. However, each of these parameters scales linearly with system size, so the sum of their exponents can be used roughly as the scaling order and the density evaluation appears to be an $O(N^3)$ step. In fact, a set of atom-centred auxiliary functions (density fitting basis) was introduced [1], which allowed the density to be written as a linear expansion in N' functions,

$$\rho(\mathbf{r}) \rightarrow \tilde{\rho}(\mathbf{r}) = \sum_A \sum_{i \in A} a_i f_i(\mathbf{r}) \quad (1.3)$$

making evaluation of the density and of V_{xc} an $N'P$, i.e. an $O(N^2)$ step. The determination of the fit coefficients a_i was also implemented as an $O(N^2)$ step (for more about the fitting procedure, see Sect. 3). The Coulomb potential at each point is calculated from the functions f^c , obtained as Coulomb integrals from the fit functions f ,

$$\begin{aligned} f_i^c(\mathbf{r}_k) &= \int \frac{f_i(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_k|} d\mathbf{r} \\ V_{\text{Coul}}(\mathbf{r}_k) &\rightarrow \tilde{V}(\mathbf{r}_k) = \int \frac{\tilde{\rho}(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_k|} d\mathbf{r} \\ &= \sum_A \sum_{i \in A} a_i f_i^c(\mathbf{r}_k) \end{aligned} \quad (1.4)$$

This makes the evaluation of the Coulomb potential at the grid points also an $N'P$ [$O(N^2)$] step, though with a much larger prefactor than the density and V_{xc} evaluation. Finally, setting up the G matrix according to Eq. (1.1), when the values of the potentials at all points are available, as well as the values of the basis functions at all points, is an N^2P step, i.e. $O(N^3)$, as is the subsequent diagonalization. The scalings of the computation steps in the original ADF implementation [1] are summarized below:

Fit coeff. determination	Potentials in grid points	Functions in grid points	Fock matrix set-up	Diagonalization
$O(N^2)$	$aN'P + bN'P$	NP	N^2P	$O(N^3)$

For many systems, in particular systems incorporating transition metal atoms (transition-metal complexes or clusters with a limited number of metal atoms) which were the prime application targets, the dominant steps were the fit coefficient determination and the Fock matrix set-up, so the scaling was typically between N^2 and N^3 . This constituted a significant improvement over the then N^4 scaling of ab initio HF calculations.

Before discussing how the various steps in the calculations can be scaled down to $O(N)$ by appropriate cut-off procedures, we comment briefly on the difference between the ADF scheme and the now widespread use of auxiliary functions in Gaussian-based DFT codes. Sambe and Felton [8] adopted the density fitting procedure of ADF in order to scale down the computation of the Coulomb operator matrix elements, but tried to avoid 3D numerical integration as much as possible (which, however, is not completely possible). They noted that the density fitting could also be used to scale down the calculation of the matrix elements of the Coulomb operator $\hat{J} = V_{\text{Coul}}(\mathbf{r})$ from $O(N^4)$ to $O(N^3)$ when using analytical integral evaluation for the $1/r_{12}$ integrals:

$$J_{\mu\nu} = \langle \chi_\mu | V_{\text{Coul}} | \chi_\nu \rangle = \sum_i a_i \int \chi_\mu(\mathbf{r}_1) \chi_\nu(\mathbf{r}_1) \frac{1}{r_{12}} f_i(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (1.5)$$

This, of course, is not tied in any way to DFT, but is applicable in HF calculations as well. The requirement of efficient analytical integral evaluation implied, however, that STOs had to be abandoned and Gaussian basis sets had to be used instead. This approach is rather efficient if the Coulomb operator alone is considered: the N^2N' process of Sambe and Felton, which does not have a high prefactor since the Coulomb integrals of Eq. (1.5) are easy with Gaussian functions, compares favourably to the $N'P + N^2P$ process of ADF ($N'P$ for the potential evaluation on the grid; N^2P for the matrix set-up), since P is typically high. However, since the matrix elements of the exchange-correlation potential are already evaluated by numerical integration (not by Sambe and Felton, but in many DFT codes today), the added cost of including the Coulomb potential in Eq. (1.1) is only the $N'P$ evaluation of $V_{\text{Coul}}(\mathbf{r}_k)$ on the grid. This is nevertheless an important part of the calculation, since the prefactor is relatively high (the evaluation of $f^c(r_k)$ is expensive). It is also difficult to scale down to $O(N)$ due to the long-range nature of the Coulomb potential. The use of auxiliary function sets for the Coulomb problem, sometimes denoted the RI- J technique, is currently receiving renewed interest, cf. Kendall and Früchtland [9] and Eichkorn et al. [10].

In this paper we will address the reduction of the various steps in the Fock-matrix set-up to linear scaling. It is to be noted that a variety of methods are currently being explored to achieve linear scaling for the matrix elements of the Coulomb potential [11–24]. There are other steps in the calculation which we do not discuss in this paper but which have already received considerable attention elsewhere, such as the scaling of the point generation itself [25] and, of course, the diagonalization [26–35].

2 Definition of cut-offs for matrix elements

For each function it is natural to define a radius outside which it is negligible, and to determine the cut-off radius of each atom as the maximum of the radii of its

functions. In the literature [7, 25] the function radius λ_μ has usually been determined from the function values by the condition $|\chi_\mu(r)| < \tau$ for $|r| > \lambda_\mu$. This condition does not take into account the fact that cutting at a given function value is not quite the same for weakly decaying functions (small exponent) as for sharply decaying functions. A (much) larger contribution to, for instance, the normalization integral would be neglected in this way for a (very) weakly decaying function. In order to have approximately the same relative error in matrix elements of both diffuse and contracted functions we determine the function radius according to a somewhat different algorithm, without of course any change in the underlying idea.

The relative weight of the radial part of a basis or fit function beyond a certain cut-off radius r_0 is calculated as the ratio between the integral of the tail of the function beyond the cut-off point r_0 and the total radial integral,

$$w(\chi; r_0) = \frac{\int_{r_0}^{\infty} r^2 \chi(r) dr}{\int_0^{\infty} r^2 \chi(r) dr} = \frac{\int_0^{\infty} r^{n+1} e^{-\alpha r} dr}{(n+1)!/\alpha^{n+2}} \quad (2.1)$$

Here n is the main quantum number, i.e. the radial part of χ is $r^{n-1} e^{-\alpha r}$. Since we wish to have a very efficient algorithm, in view of the frequent tests on these weights, we avoid the time-consuming integration over the partial region (r_0, ∞) in the numerator by using a simple exponential function to approximate the tail of the function $r^2 \chi$,

$$g(r) = C e^{-\beta r} \quad (2.2)$$

The parameters β and C are determined by

$$g(r_0) = r_0^2 \chi(r_0) \text{ and } g'(r_0) = \frac{d}{dr} r^2 \chi|_{r=r_0} \quad (2.3)$$

At large cut-off points the function g simulates the tail of $r^2 \chi$ almost perfectly. Equations (1.1) and (1.3) lead to the following expression for the relative weight of the tail

$$w(\chi; r_0) = \frac{\alpha^{n+2} r_0^{n+1} e^{-\alpha r_0}}{(n+1)! (\alpha - (n+1)/r_0)} \quad (2.4)$$

If we had used the normalization integral rather than the direct radial integral over (the radial part of) the function in Eq. (2.1), approximately the square of the present expression would have resulted (apart from a factor of the order of magnitude 1) for the contribution of the region beyond $r = r_0$ to the normalization integral. Therefore the typical choice for the weight of 0.1% ($w = 10^{-3}$) implies neglecting a contribution 10^{-6} to the normalization integral. The weight of the neglected tails, from which the cut-off radii r_0 (denoted λ_μ) are determined, is of course chosen depending on the desired maximum ‘‘cut-off error’’. The latter has to be in keeping with the precision of the numerical integration. An atomic radius can be determined as $\lambda_A = \max_{\mu \in A}(\lambda_\mu)$. All

matrix elements of a pair AB can be excluded if $R_{AB} > \lambda_A + \lambda_B$. For a given atom A we have a neighbourhood of atoms B which are such (radius λ_B larger than $R_{AB} - \lambda_A$) that the pairs AB have to be treated. This will always be a small number and for large molecules will not increase with system size. Parts of the code with

an outer and inner loop over the atoms [$O(N^2)$] transform into a single loop over the atoms with, for each atom, a small inner loop over the neighbouring atoms [$O(N)$]. As an additional refinement, for an atom pair AB that has to be considered, the function pairs may be checked and excluded if $R_{AB} > \lambda_\mu + \lambda_\nu$. This will not change the $O(N)$ scaling but will improve the prefactor.

For matrix elements that have to be evaluated according to the chosen criterion one can still restrict the number of sample points in the numerical integration by using essentially the same criterion to neglect distant points (see Sect. 4).

3 Linearization of the density fitting

The linearization of the density fitting is trivial once an important point is recognized concerning the implementation of the density fitting in the ADF program. To make this clear we briefly review the density fitting as introduced in [1].

The fit coefficients for a density $\rho(\mathbf{r}), \rho(\mathbf{r}) \rightarrow \tilde{\rho}(\mathbf{r}) = \sum a_i f(\mathbf{r})$ are determined from a least-squares fitting procedure, minimizing the deviation D between true and fitted density subject to the condition that charge is conserved

$$D = \int (\rho - \tilde{\rho})^2 d\tau; \quad \int \tilde{\rho} d\tau = \int \rho d\tau = N \quad (3.1)$$

Introducing the Lagrangian multiplier 2λ one obtains the equation

$$\mathbf{S}\mathbf{a} = \mathbf{t} + \lambda \mathbf{n} \quad (3.2)$$

where \mathbf{a} is a column vector with the fit coefficients, \mathbf{S} is the matrix of fit function overlap integrals, \mathbf{n} is a column with the integrals $n_i = \int f_i d\mathbf{r}$ and \mathbf{t} is a column vector collecting the overlaps $t_i = \int \rho(\mathbf{r}) f_i(\mathbf{r}) d\mathbf{r}$. Solving for the Lagrange multiplier gives explicitly

$$\begin{aligned} \mathbf{a} &= \mathbf{S}^{-1} \mathbf{t} + \lambda \mathbf{S}^{-1} \mathbf{n} \\ &= \mathbf{S}^{-1} \mathbf{t} + \frac{N - \mathbf{n}^\dagger \mathbf{S}^{-1} \mathbf{t}}{\mathbf{n}^\dagger \mathbf{S}^{-1} \mathbf{n}} \mathbf{S}^{-1} \mathbf{n} \end{aligned} \quad (3.3)$$

Dunlap et al. [36] have changed the deviation to be minimized from the least-squares deviation in the density to the least-squares deviation in the Coulomb self repulsion of $\rho - \tilde{\rho}$. This allows the electronic Coulomb energy, evaluated from the fitted density, to approach the true Coulomb energy from above (‘‘variational fitting for the Coulomb term in the energy’’).

Carrying out the fitting process as outlined here for the whole density at once will be an $O(N^3)$ process in view of the linear system of Eq. (3.2). Considering possibilities to make the fitting scale more favourable, we first note that apart from the charge conservation condition showing up in the second (Lagrange multiplier) term of the equation for \mathbf{a} , the determination of the fit coefficients is a linear process in the fitted object, the density $\rho(\mathbf{r})$: it is possible to write ρ as a sum of parts ($\rho = \sum_i \rho^i$ say) and determine the coefficients as the sum

of the coefficients arising from fitting the parts separately, provided all functions are used for each ρ^i term.

It is natural to express the charge density $\rho(\mathbf{r})$ as the sum of one- and two-centre charge distributions ρ^{AB} :

$$\rho = \sum_{\mu, \nu} P_{\mu\nu} \chi_{\mu} \chi_{\nu} = \sum_{A \leq B} \sum_{\mu \in A} \sum_{\nu \in B} P_{\mu\nu} \chi_{\mu} \chi_{\nu} = \sum_{A \leq B} \rho^{AB} \quad (3.4)$$

The use of all fit functions for the fit of each ρ^{AB} term implies that apart from functions at A and B , also functions at neighbouring centres, C say, would have to be used. The ADF implementation [1], however, carries out the fitting for pair densities ρ^{AB} separately and restricts the fitting of ρ^{AB} to fit functions on A and B only, making the fitting an $O(N^2)$ process. This policy does not seem to have been followed by either Sambe and Felton [8] or Dunlap et al. [36] or any of the more recent implementations of the use of auxiliary functions, except for the recent work of Gallant and St-Amant [37], who also partition the total density into a sum of subsystem densities. We briefly review the arguments in favour of atomic pair fitting:

1. The process becomes $O(N^2)$. Moreover, the fitting of each pair density ρ^{AB} with functions on A and B only implies that the dimension of each linear system, Eq. (3.3) remains modest and is independent of system size.

2. The possible benefit in fitting ρ^{AB} that could be derived from functions on other atoms is heavily system dependent. Since the presence of atoms C close to ρ^{AB} is accidental, the quality of the fit could vary from one system to another depending on the accidental surrounding of a pair AB by other atoms. A reliable fit basis will have to be able to fit a density ρ^{AB} accurately, with approximately the same precision, under all circumstances, also in cases where other atoms are remote or not present at all, for instance in the diatomic molecule AB .

3. It is not possible to take the accidental presence of atoms C into account when systematically developing sets of fit functions. For fit functions on A this is possible to a certain extent for directly bonded atoms B , since the variation in the number and type of chemical bonds formed by an atom A is limited (nearest neighbour distances for instance fall in a certain range). Fit function sets for an atom A can therefore be generated that can cope with the presence of any ‘‘normal’’ two-centre bond with an atom B . The presence of ‘‘third parties’’ C is not certain and can therefore not be exploited in a generation scheme. It is possible that the fit sets generated for reliable pair fits are actually larger than would be necessary in some particular case – the ADF fit sets are indeed relatively large – but this is not a problem since the linear system, Eq. (3.3), always remains small.

4. The pair fitting scheme can be trivially scaled down to $O(N)$ (see below).

Given a pair-wise fitting scheme, linear scaling can simply be achieved by including only atom pairs for which the sum of the radii, determined as described in Sect. 2, is larger than the distance between the atoms. Furthermore, a smaller prefactor can be obtained by eliminating the pairs of basis functions for which the

sum of the radii is smaller than the distance between their atoms. We note that the pair-wise fitting has also proven valuable during the parallelization of the ADF program. As long as the number of nodes is smaller than the number of atom pairs, distribution of the one- and two-centre densities over the nodes of a parallel machine, with proper account of their computational ‘‘weights’’ to achieve load balancing, yields a perfectly scaling parallel implementation [38]. In the present case of course, only the pairs that are not excluded are to be distributed.

4 Linearization of manipulations involving grid points: Coulomb and XC potential evaluation

The calculation of the XC potential will already be linear when we apply cut-offs to the evaluation of the values of fit functions at the grid points (and derivatives) to obtain the density and XC potential. Since it is the value of a fit function at the grid point that enters the density evaluation, cut-offs can be based on a threshold for the function value: $f_i(\mathbf{r})$ is considered negligible outside a radius r_0 at which the radial part of f_i is equal to a given small threshold, $f_i(\mathbf{r}_0)/Z_{lm}(\hat{\mathbf{r}}_0) = \tau$ (Z_{lm} is a real spherical harmonic). We use the weight criterion, Eq. (1.4), since the density is actually used for the XC potential, which enters the spatial integrals making up the Fock matrix elements. The XC potential evaluation is much cheaper than the evaluation of the Coulomb potential, but in our scheme the calculation of the two potentials is closely related. For the distance cut-offs in the Coulomb potential we will follow the same strategy as implemented in the ADF-BAND code for infinite periodic systems [5]. The Coulomb potential in a certain integration point can be written as a sum over the atoms:

$$\begin{aligned} V_C(\mathbf{r}_k) &= \sum_A V_C^A(\mathbf{r}_k) \quad \text{with} \quad V_C^A(\mathbf{r}_k) \\ &= \sum_{i \in A} a_i \int \frac{f_i(\mathbf{r}')}{|\mathbf{r}_k - \mathbf{r}'|} d\mathbf{r}' \end{aligned} \quad (4.1)$$

Applying the expansion of $|\mathbf{r} - \mathbf{r}'|$ in spherical harmonics and using the exponential form of the STO fit functions gives us the following spherical harmonics expansion for the Coulomb potential from each atom:

$$\begin{aligned} V_C^A(\mathbf{r}_k) &= \sum_l \sum_{m=-l}^{m=l} \frac{4\pi}{2l+1} Z_{lm}(\hat{\mathbf{R}}_{kA}) I_{lm}^A(R_{kA}) \\ I_{lm}^A(R_{kA}) &= \sum_{i \in A} \delta(l, l_i) \delta(m, m_i) a_i I(n_i, l_i, \alpha_i; R_{kA}) \end{aligned} \quad (4.2)$$

Here \mathbf{R}_{kA} is the distance vector from nucleus A to the point \mathbf{r}_k , and I_{lm}^A is the radial part of the (lm) term in the spherical harmonics expansion. The function I is obtained from incomplete Gamma functions. Going back to the algorithms used to evaluate incomplete Gamma functions, it is apparent that this function can be written as the sum of a multipolar and an exponentially decaying part:

$$I(n_i, l_i, \alpha_i; R_{kA}) = \frac{1}{(R_{kA})^{l_i+1}} \frac{(n_i + l_i + 1)!}{(\alpha_i)^{n_i+l_i+2}} \quad (4.3)$$

$$+ e^{-\alpha_i R_{kA}} J(n_i, l_i, \alpha_i; R_{kA})$$

The function J consists of a power series in R_{kA} , with n_i as highest power. This is just one factor R_{kA} more than in the value of the STO fit function itself, which of course has a simple $r^{n_i-1}e^{-\alpha_i r}$ radial behaviour. For the sake of clarity one cut-off radius for both the exponential part of the Coulomb potential and the function value for the density evaluation is used, where the $r^{n_i}e^{-\alpha_i r}$ radial behaviour of the former is to be taken into account when determining the cut-off threshold.

The multipolar part of the Coulomb potential is long-range, which hampers achieving linear scaling. It is advantageous to first compute the ‘‘strength’’ M_{lm}^A of an atomic multipolar term of nucleus A as

$$M_{lm}^A = \sum_{i \in A} \delta(l, l_i) \delta(m, m_i) a_i \frac{(n_i + l + 1)!}{(\alpha_i)^{n_i+l+2}} \quad (4.4)$$

so that the multipolar term becomes

$$Q_{lm}(R_{kA}) = \frac{M_{lm}^A}{(R_{kA})^{l+1}} \quad (4.5)$$

The multipolar strength can be calculated once for all atoms (at a certain SCF cycle), and then the possibility to neglect higher-order multipoles can, for a particular block of grid points, be based on the strength M_{lm}^A as well as the distances R_{kA} . For the size of systems we have been considering, it has never been possible to neglect also the $l = 0$ terms, not even from the most distant atoms, so there has always remained an $O(N^2)$ term in the Coulomb potential evaluation. It has been possible to neglect higher multipole terms, in particular above $l = 1$. We will investigate below to what extent the $O(N^2)$ scaling of the $l = 0$ potential terms, which only require a small fraction of the total time in the calculations without cut-offs, affects the overall scaling in the systems considered (of the order of a several hundreds of atoms). Obviously, the structure of the whole scheme is suitable for fast multipole techniques [11–13, 23], as are now being applied in Gaussian-based codes [5, 11–15, 17]. Nevertheless, it appears that such techniques would only pay off, and are actually only needed in the present scheme, when the size of systems becomes one order of magnitude larger.

5 Linearization of manipulations involving grid points: function evaluation and Fock matrix set-up

Pérez-Jordá and Yang [7] have developed an $O(N)$ method for the evaluation of the density at the grid points, and of matrices such as the overlap matrix, based on the construction of the sets $S(\mathbf{r}_k)$ of basis functions that have nonnegligible values at a given grid point \mathbf{r}_k (we ignore here the partition function p_x that also plays a role in the development of [7]):

$$S(\mathbf{r}_k) : \{ \chi_\mu \mid \chi_\mu(\mathbf{r}_k) > \tau \} \quad (5.1)$$

They determine the set of relevant basis functions for each grid point with an efficient tree algorithm. However, ADF uses (as do most DFT implementations, see TURBOMOLE [39] and GAUSSIAN 94 [40]) blocks of points for efficiency reasons, so that vectorization in inner loops over grid points can be used. Moreover, as explained in Sect. 2, we do not base the cut-off decisions on a radius for each basis function that is determined by a fixed threshold for the function value. We wish to base the cut-off decisions on an estimate of the percentage of the matrix elements that will be neglected. Therefore the set of functions that has to be taken into account for a given block of points B is based on whether or not all the points of the block are located in the region of space beyond the radius r_0 , which according to Sect. 2 contributes less than a given percentage [w of Eq. (2.4)] to the radial integral over the whole space. We define the tail function $T(\chi_\mu; \mathbf{r}_k)$ for function χ_μ with respect to a point \mathbf{r}_k as the weight in the sense of Sect. 2 of the radial integral of χ_μ beyond a radius which is the distance from the centre of the function to the point \mathbf{r}_k , or

$$T(\chi_\mu^A; \mathbf{r}_k) = w(\chi_\mu; |\mathbf{r}_k - \mathbf{R}_A|) \quad (5.2)$$

In case a block of points B is specified ($T(\chi_\mu; B)$), the radius is to be taken as the distance from the centre of the function to the nearest point of the block. The use of the tail function allows a more even-handed treatment of diffuse and contracted functions than the uniform cutting of basis functions at a given value. We now define the set of functions $S(B)$ that belongs to a block of points B as follows:

χ_μ (on atom A) belongs to $S(B)$ if there exists a χ_ν (on atom C) such that both the following conditions are satisfied:

$$\text{a) } R_{AC} < \lambda_\mu + \lambda_\nu; \quad \text{b) } T(\chi_\mu; B) * T(\chi_\nu; B) > \text{threshold} \quad (5.3)$$

Atoms A and C are chosen from the set of atoms that possess functions with a tail $T(\chi_\mu; B)$ over the block larger than a given threshold. The condition represented in Eq. (5.3a) then means that the function χ_μ is sufficiently close to some other function χ_ν to have a non-negligible matrix element with it. The second condition checks if the contribution of the block B to the numerical integration that makes up the $\mu\nu$ matrix element would indeed be non-negligible.

Stratman et al. [25] have pointed out that the points in the block should be localized in space in order to limit the set of relevant basis functions. As a matter of fact, the Voronoy polyhedron scheme of point generation in ADF already generates points in spatial cells (small spheres around an atom, or wedges of the atom’s Voronoy cell) [41, 42].

As pointed out in [7, 25] the creation of the sets $S(B)$ of basis (and fit) functions belonging to block B with just the condition expressed in Eq. (5.1) is an $O(N^2)$ step, the number of checks required being in principle (number of points) \times (number of functions). In our case, where pair tests are being done for each block, cf. Eq. (5.3), it is in principle an $O(N^3)$ step. However, with present target sizes of molecules of up to 1000 atoms, it is still very

cheap. It is organized efficiently in a tree structure by first, for a given block of points, excluding (most) atoms on account of their distance from the block, and considering only pairs $\mu\nu$ belonging to the (small) set of atoms “belonging” to the block.

Once the sets of functions $S(B)$ belonging to the blocks B have been determined, the calculation of the function values is evidently $O(N)$. The calculation of matrices of operators like the identity (overlap matrix) or $V_N(\mathbf{r})$, $V_{\text{Coul}}(\mathbf{r})$, $V_{\text{xc}}(\mathbf{r})$ is now also automatically linear, if the values of the potentials in the points are known (see previous section), since for each block of points B a limited number of basis functions has to be processed, which is independent of system size. Let us denote the total operator value by $\Omega(\mathbf{r})$ (operators like the kinetic energy $\hat{T} = \sum_i -\frac{1}{2}\nabla^2(i)$ can be treated in a completely analogous way), then for the matrix of the operator $\hat{\Omega}$, and also for the density, if it is calculated from the density matrix, we have the following double loop over the basis functions:

{Construct contribution of block B to matrix elements;
construct values of density in points of B }

For $\mu \in S(B)$

for all $\mathbf{r}_k \in B : \xi_\mu(\mathbf{r}_k) = \Omega(\mathbf{r}_k)\chi_\mu(\mathbf{r}_k)$

For $v \in S(B)$

if $R_{AC} < \lambda_\mu + \lambda_\nu$ then

if $T(\chi_\mu; B) * T(\chi_\nu; B) > \text{threshold}$ then

{add contribution of B to $\Omega_{\mu\nu}$ }

$$\Omega_{\mu\nu} = \Omega_{\mu\nu} + \sum_{\mathbf{r}_k \in B} \chi_\nu(\mathbf{r}_k) * \xi_\mu(\mathbf{r}_k) \quad (\text{operator matrix})$$

{calculate density in points of B }

for all $\mathbf{r}_k \in B : \rho(\mathbf{r}_k) = \rho(\mathbf{r}_k) + P_{\mu\nu}\chi_\mu(\mathbf{r}_k)\chi_\nu(\mathbf{r}_k) \quad (\text{density})$

End v

End μ

We have included here the density evaluation from the density matrix to show that it follows essentially the same route, although in our implementation the density is directly evaluated from the fit functions during the self-consistent field (SCF). Note that the use of the tail function $T(\chi_\mu; B)$ does not change the scaling, which would also become linear if only tests on a cut-off radius were performed. However, it allows an additional refinement by which the contracted or diffuse nature of basis functions, which is important for the contribution of block B to the matrix elements, is properly accounted for.

6 Results and discussion

This section contains an overview of the effect of the implementation of cut-offs in various stages of the calculation of the (electronic potentials part of the) Fock matrix, including graphs with timings for the separate routines of the original code of ADF and the adapted

code including the cut-offs. As a test case we chose alkanes with the following geometry: the carbon atoms lie in one plane and form a fully extended zig-zag chain; the hydrogen atoms are below and above the plane. As a second test case we took hypothetical two-dimensional systems, namely hexagonal graphitic sheets, which are terminated by nitrogens. Figure 1 shows the largest one. The fitting of our timing curves to a power function is done with the Levenberg-Marquardt algorithm [43]. For the investigation of the computational efficiency in a three-dimensional system we have chosen the $\text{Pt}(\text{P}(\text{Ph})_3)_3\text{CO}$ complex.

6.1 Density fit

Figures 2 and 3 display the scaling behaviour of the two routines involved in the fitting of the density. Prior to the SCF cycles for each pair of atoms the first routine (FITINT) calculates the elements of the matrices \mathbf{S} and \mathbf{n} , and the individual integrals $t_{\mu\nu,i} = \langle \chi_\mu^A \chi_\nu^B f_i^{A \text{ or } B} \rangle$, which are contracted on each cycle with the density matrix elements $P_{\mu\nu}$ to generate the \mathbf{t} vector [cf. Eqs. (3.2) and (3.3)], and writes these integrals to file (timing in Fig. 2). On every SCF cycle the second routine (RHOFIH) reads the matrix elements from file and performs the fitting of the present density by solving Eq. (3.3) for each pair of atoms (timing in Fig. 3). Both figures show the scaling with and without the application of cut-offs to the basis functions. The use of cut-offs implies that we are only fitting atom-pair densities of “neighbouring” atoms. The calculation of the fit integrals, which is dominated by the integrals $t_{\mu\nu,i} = \langle \chi_\mu^A \chi_\nu^B f_i^{A \text{ or } B} \rangle$, meets our expectations in that Fig. 2 shows a decrease of an almost quadratic scaling to an almost linear scaling. The small deviations from quadratic and linear scaling respectively are due to

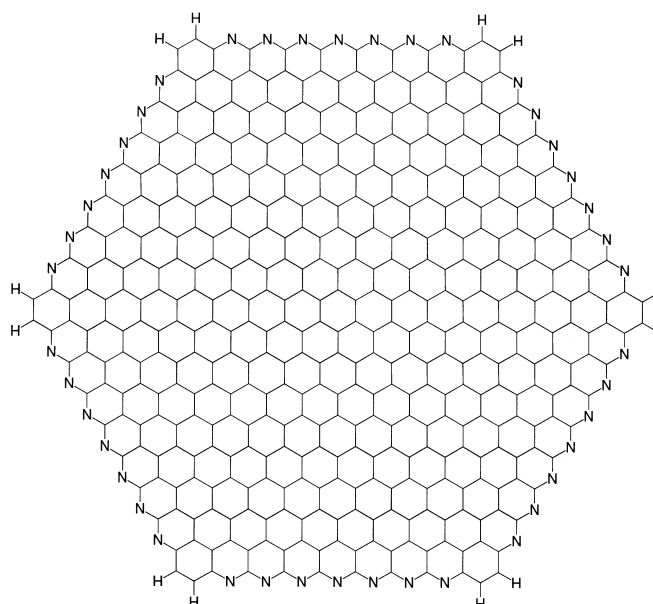


Fig. 1. The largest planar aromatic system used for the timing results

Fig. 2. The scaling behaviour of the routine (FITINT) that calculates and writes the fit integrals to file, with and without the use of cut-offs, for a series of alkane chains with the indicated total numbers of atoms

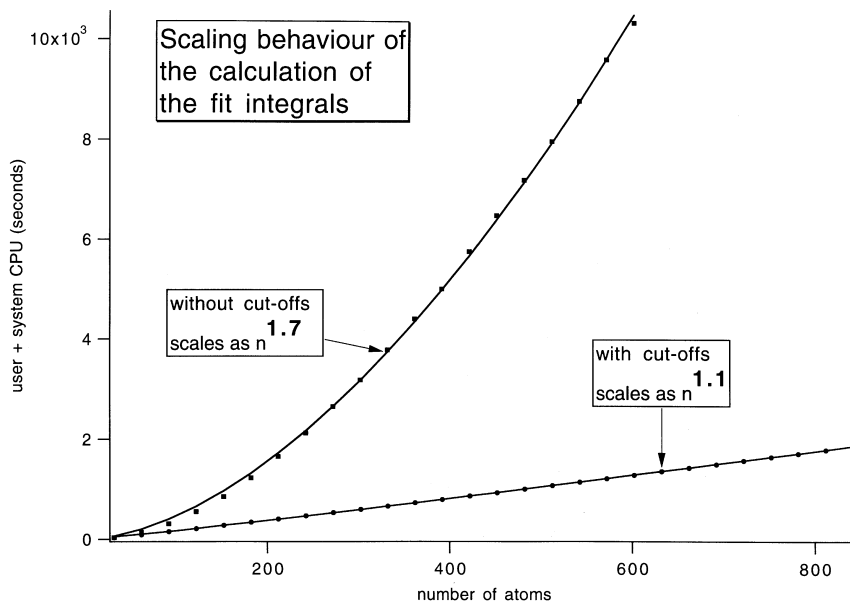
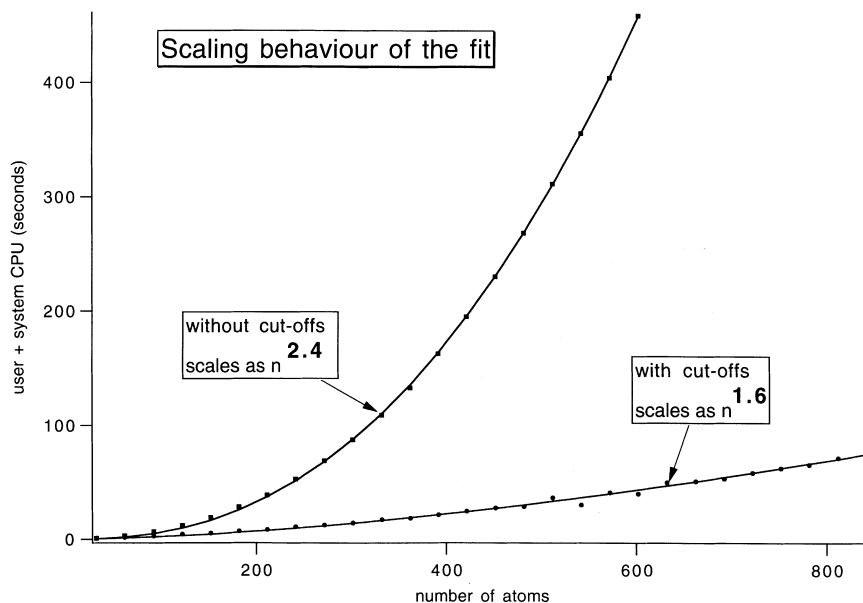


Fig. 3. For the alkane chains, the scaling behaviour of the routine (RHOFIH) that reads the fit integrals from file and solves the least-squares equation for the fit coefficients



special circumstances. The scaling lower than 2.0 in the old code arises from atom pairs with large distance being cheaper with the current integral routines. The small deviation from linear scaling of the calculation with cut-offs may be due to I/O effects perturbing the timing, since we could determine this to be also the origin of the scaling of the routine RHOFIH (Fig. 3) not being as predicted. The scaling of RHOFIH without cut-offs, Eq. (2.4), as well as with cut-offs, Eq. (1.6), is higher than predicted. Analysis of this routine showed that the discrepancy between the actual scaling and the predicted one is caused by an I/O bottleneck. The routine is not computationally intensive but spends the largest part of its time in I/O manipulations, viz. the reading of the fit integrals (compare the absolute times in Figs. 2 and 3). For the smaller molecules the file with the fit integrals will fit into memory; however, for the larger molecules

this becomes impossible, as the size of this file scales linearly with the number of atoms. At certain molecular sizes this causes a tremendous growth of paging which perturbs the timings. The same will happen in the FITINT routine (Fig. 2) and may become visible for the smaller computation time in the calculation with cut-offs.

6.2 The Coulomb potential

In Fig. 4 the scaling of the calculation of the Coulomb and the XC potential is shown with and without cut-offs. The reduction in the scaling power is rather small, although even this implies a reduction in computation time by a factor of 10 for the 600-atom system. We examine these timings more closely to understand the

small decrease in the scaling power. In Fig. 5 the different components of the calculation of the Coulomb and XC potential are presented. First, the density in the grid points $\rho(\mathbf{r}_k)$ is evaluated as written in Eq. (1.3). Due to the cut-offs for each grid point only fit functions in its neighbourhood are used. This results in a perfectly linear scaling of this step. The evaluation of the XC potential is also linear, because it only manipulates the density in the grid points. With $V_{\text{Coulomb}}^{\text{sr}}$ we denote the calculation of the short-range (exponential) part of the Coulomb potential, Eq. (4.3). For this component of the Coulomb potential the same cut-off radii are applied as for the density. Therefore this component of the potential also scales linearly. For the evaluation of the multipolar parts $Q_{lm}^A(R_{kA}) = M_{lm}^A / (R_{kA})^{l+1}$ of the Coulomb potential, $1/R_{kA}$ has to be calculated for each grid point \mathbf{r}_k with respect to all atoms in the system.

Fig. 4. For the alkane chains, the scaling behaviour of the routine that calculates the Coulomb and the XC potential in the grid points

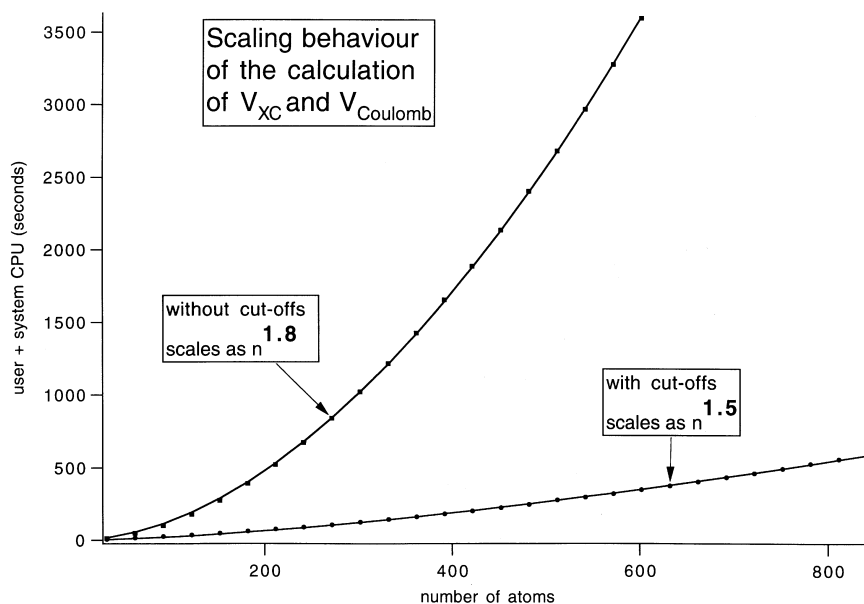
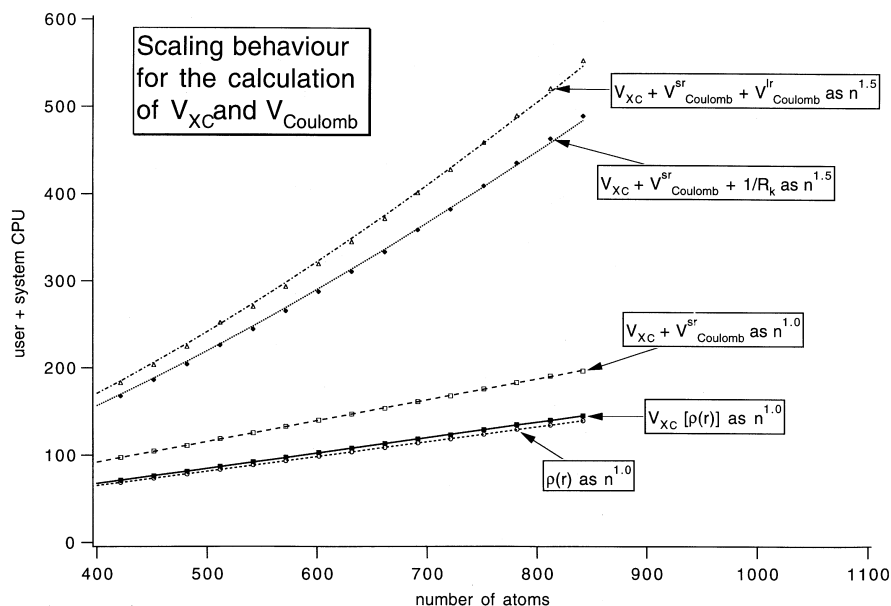


Fig. 5. Timing and scaling behaviour for the various steps in the calculation of the Coulomb and the XC potentials in the grid points when cut-offs are used (alkane chains). $V_{\text{Coulomb}}^{\text{sr}}$ is the short range (exponential) part of the Coulomb potential; $V_{\text{Coul}}^{\text{lr}}$ is the long range (multipolar) part



A few computational operations are needed to determine $1/R_{kA}$: first, $(R_{kA})^2$ from $x_{kA}^2 + y_{kA}^2 + z_{kA}^2$, then R_{kA} from $\sqrt{(R_{kA}^2)}$ and finally $1/R_{kA}$. We show explicitly just how much time the simple evaluation of the $1/R_{kA}$ values takes by adding the computation time to that of V_{XC} and $V_{\text{Coulomb}}^{\text{sr}}$. This step of course scales quadratically, and since it takes an amount of time that is comparable to the previous steps, its addition changes the scaling from linear to about 1.5 for systems up to approximately 1000 atoms. The relatively large computation time of this simple step can be attributed to the evaluation of the square root being extremely expensive. The last curve shows the scaling for the complete evaluation of the Coulomb and the XC potential. Comparing the last two curves we see that once the $1/R_{kA}$ values are known, only little time is needed to

evaluate the full multipolar contributions of Eq. (4.5). Since the evaluation of the long-range part of the Coulomb potential is a quadratically scaling step in our present implementation, its computation time would ultimately start to dominate. For the size of systems we are interested in, the evaluation of the long-range part, although algorithmically very simple, takes an amount of time that is comparable to (even somewhat larger than) the sum of the exchange-correlation potential and the short-range part of the Coulomb potential. Since several of the other computation steps can be scaled down very successfully (see the comparison for total timings below), the computation time for the long-range Coulomb potential changes from a modest fraction of the time without cut-offs to a significant fraction after application of cut-offs. The computational expense is

still acceptable, but it is one of the prime targets for future efforts towards computational speed-up.

6.3 The Fock matrix

Besides the potential in the integration points we need the values of the basis functions in the integration points to set up the Fock matrix. Figure 6 shows the two curves, with and without cut-offs, for the evaluation of $\chi(\mathbf{r}_k)$. The original code of the routine scales quadratically due to the loop over the functions and the loop over the integration points. The adapted code scales perfectly linearly as expected.

Having all ingredients available (function and potential values in the grid points) the final set-up of the Fock matrix is shown in Fig. 7. A double loop over the

Fig. 6. The scaling behaviour (alkane chains) of the routine that calculates the values of the basis functions in the grid points, with and without the use of cut-offs

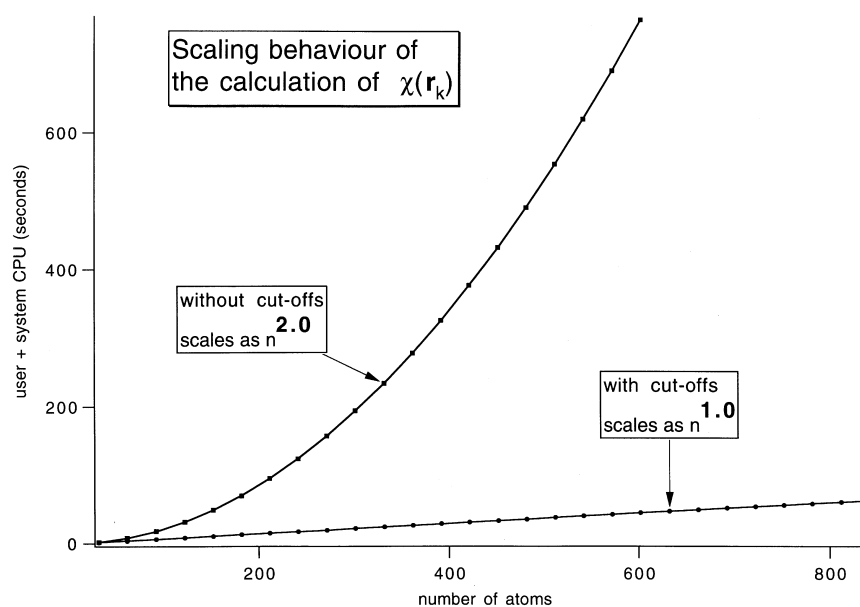
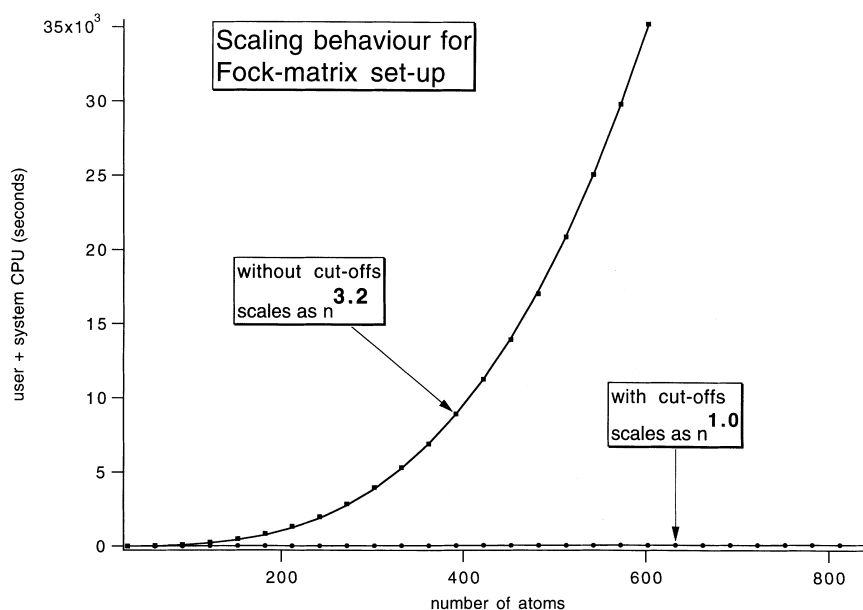


Fig. 7. The scaling behaviour (alkane chains) for the final set-up of the Fock matrix, when potentials and basis function values are known, with and without the use of cut-offs



atoms and a loop over the integration points in the original code give a cubic scaling. The adapted code shows exactly linear scaling with system size. Dramatic improvement of the computational efficiency results.

6.4 Aromatic systems

The applicability of our method to two-dimensional systems is demonstrated with the results of the aromatic systems. We do not show the individual routines but the timing is shown for the sum of the two routines involved in the density fit (Fig. 8) and the calculation of the total Fock matrix (Fig. 9), including the calculation of function values, of the potentials, and of the matrix elements of the Fock matrix with both the kinetic energy and the electronic potentials as operators. In Fig. 8 the scaling behaviour with cut-offs is dominated by the

evaluation of the fit integrals. The aromatic systems are not yet large enough to enter the size range where the cut-offs cause the number of pairs to be handled to increase linearly (note that in one dimension even the largest aromatic system has the size of only a C_{35} alkane chain). The number of pairs still increases like Eq. (1.3), explaining the observed scaling. The poorer scaling of RHOFIH, Eq. (1.7), is not visible due to the computation time being more than an order of magnitude less than that of FITINT. In Fig. 9 the scaling achieved in the calculation of the Fock matrix with cut-offs applied is an average over the various components of the calculation. The quadratic scaling of the long-range Coulomb part – still relatively cheap at this system size – is not yet important, the 1.3 scaling has its origin in the increase of the number of pairs, as in the case of the fit integrals.

Fig. 8. The combined scaling behaviour in the planar aromatic systems of the two routines (FITINT and RHOFIH) that are involved in the fitting of the density, with and without the use of cut-offs (one call of each routine)

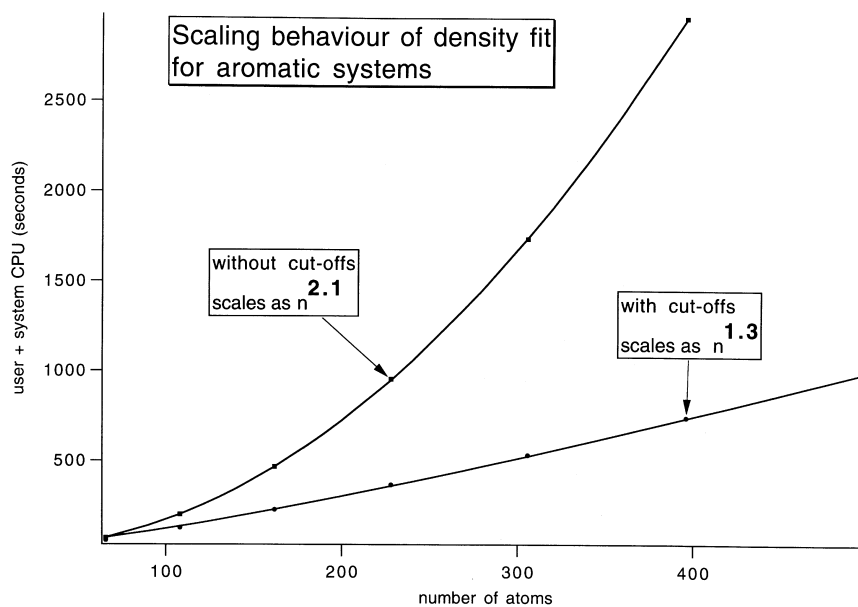
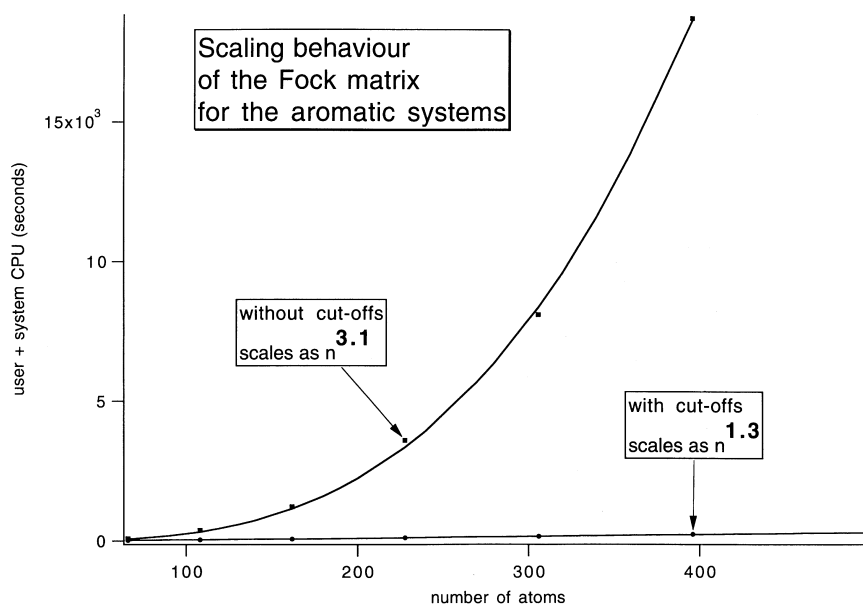


Fig. 9. The combined scaling behaviour of all routines that are used to set-up the Fock matrix: the calculation of the Coulomb and the XC potential in the grid points, the evaluation of the basis functions in the grid points, the initialization of the Fock matrix with the kinetic energy part, and the final set-up of the Fock matrix. The timings are shown with and without the use of cut-offs



6.5 Relative error

Our objective throughout this project has been to improve the scaling with system size, while retaining sufficient precision in the results. It is natural to gauge the required precision against the precision to which the numerical integration is typically set. A key quantity is the bond energy, i.e. the energy of the molecule minus the sum of the energies of the atoms. With increasing system size, the *relative* error in this quantity is required to stay constant (approximately), implying a constant absolute error in the bonding energy *per bond*, which of course also stays roughly constant irrespective of system size. The numerical integration has the required property of yielding a constant relative error independent of system size; its magnitude being determined of course by the setting of the accuracy parameter for the numerical integration. The cut-off parameters should be set in such a way that if the numerical integration affords a bonding energy with a relative error of only 0.1% say, the calculation done with the cut-off radii should also return bonding energies with a relative error not larger than 0.1%.

The setting of the cut-off parameters has been tested for the following molecules as test cases: $C_{50}H_{102}$, $C_{100}H_{202}$ and $Pt(P(Ph)_3)_3CO$. Each cut-off parameter was investigated separately to prevent spurious results from accidental cancellation of errors and also with a high numerical precision (relative error less than $10^{-4}\%$) to avoid numerical noise. Our goal has been to obtain a setting for each cut-off parameter that gives a relative error in the bonding energy smaller than 0.1%. At the same time we have tried to achieve stability in the computational process in the sense that not just the final bonding energy is stable to the required precision but also the SCF cycling proceeds in the same way, so that the same number of SCF cycles is needed to reach convergence. The latter requirement resulted for some of the cut-offs in values that lead to a relative error in the bonding energy much less than 0.1%.

The cut-off used in the fitting of the density is determined by the radius of a basis function as calculated from the weight parameter w for neglecting the tail of the basis functions given in Eq. (2.4). The radii are used to decide which pairs of basis functions have negligible overlap. The neglecting of basis function pairs $\chi_\mu\chi_\nu$ in the fitting of the density leads to loss of charge (the contribution $P_{\mu\nu}S_{\mu\nu}$ is neglected) and to a relative error in the bonding energy. A good choice for the cut-off criterion w turned out to be 0.05%. This gives a radius of 5.8 Å for the carbon atom. The loss of charge is almost negligible: less than $10^{-7}\%$. Although this is a small charge deficit, it turned out to be important to rescale the fit coefficients so the exact charge is retained. This rescaling of the number of electrons in the fit density proved to be particularly beneficial for stability in the (number of) SCF cycles. The relative error in the bonding energy due to this cut-off in the density fit is very small: less than $10^{-6}\%$.

The calculation of the Coulomb and XC potential in the grid points uses two cut-off criteria. One is for the cut-off of the tails of the fit functions, which is used for

both the density evaluation and the short-range Coulomb potential. The weight function w of Eq. (2.4) is used as a criterion to determine the function radius. The other cut-off is for the multipolar part of the Coulomb potential, Eq. (4.5). The influence of the two cut-off parameters on the relative error of the bonding energy and the convergence of the SCF was investigated independently. The cut-off criterion of the fit functions was set at 0.005%, which implies a maximum radius of carbon of 7.4 Å. The relative error in the bond energy was about 0.01% for $Pt(P(Ph)_3)_3CO$, but for the alkanes it was much smaller: about $10^{-5}\%$. The cut-off for the multipolar part was set to 10^{-5} a.u., i.e. a multipole term $M_{lm}^A/(R_{kA})^{l+1}$ is evaluated in all the points of a block B if $M_{lm}^A/([R_{kA}]^{l+1})$ in the point \mathbf{r}_k of the block nearest to atom A is larger than 10^{-5} a.u. In this case the error in the energy was less than $10^{-5}\%$.

For the building of the Fock matrix two cut-off criteria are used: (1) the threshold that determines if two basis functions have a sufficiently large overlap and therefore a Fock matrix element, analogous to the (independent) threshold used to decide which basis function pairs can be neglected in the density fit; (2) the tail-function indicating the weight of a basis function over a block of points, which determines if the points of a block have to be used for the numerical integration of a Fock matrix element, Eq. (5.3). We will refer to the former cut-off as the overlap threshold. Again the two cut-off criteria are investigated separately. We investigated the error for the building of the Fock matrix by comparing the matrix of the unit operator, calculated numerically with either one of the cut-offs applied, to the analytically calculated “exact” overlap matrix. The radii of the basis functions for the overlap threshold were determined with w [Eq. (2.4)] set to 0.05%. The maximum radius of the carbon atom in the building of the Fock matrix then equals 7.4 Å. The maximum absolute error in any overlap matrix element is less than 10^{-4} for this cut-off. To determine if a block B has to be used for a $\langle\chi_\mu|\hat{h}_s|\chi_\nu\rangle$ matrix element of the Fock matrix, the threshold of Eq. (5.3b) was set to 10^{-4} , which also resulted in a maximum absolute error of 10^{-4} .

We have finally ascertained that with the above mentioned setting of the cut-off criteria the relative error in the bond energy does not increase with the system size. Figure 10 displays the relative error of the bond energy with a precision of 10^{-6} ($10^{-4}\%$) in the numerical integrations of the Fock matrix elements, as well as with a numerical precision of 10^{-3} . A numerical precision of 10^{-6} assures that we do not have errors due to numerical integration and cut-offs that accidentally cancel each other. The results with numerical precision of 10^{-3} are displayed because this precision was used for the timings. We see that the cut-off criteria satisfy our requirement of a relative error less than 0.1% and of a stable relative error, independent of system size.

7 Summary

In Tables 1–3 we present some overall timings for the old and new codes. As a genuinely three-dimensional

Fig. 10. The relative error in the bond energy (total energy minus sum of atomic energies) for the alkanes. The cases with a numerical integration precision of 10^{-3} and 10^{-6} are shown

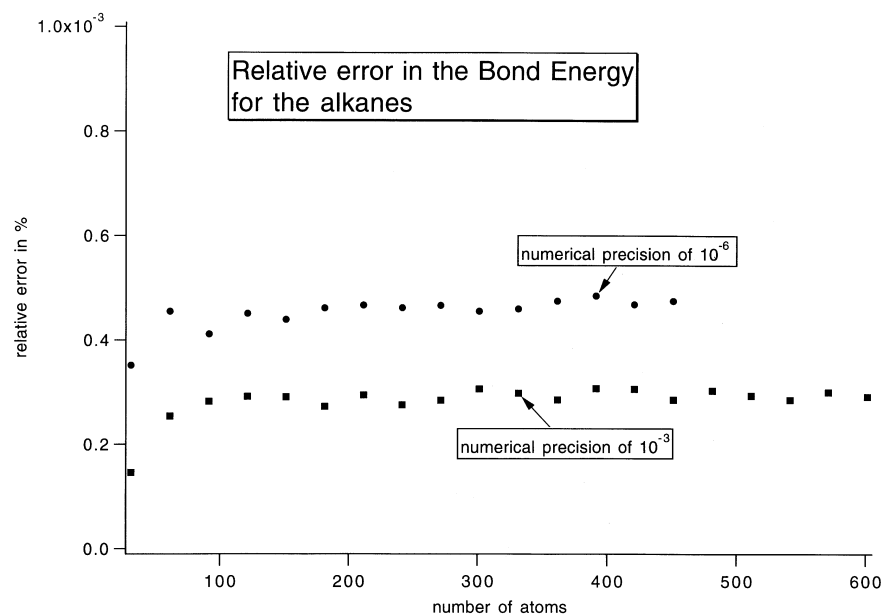


Table 1. Timing results (times in seconds) for Pt(P(Ph)₃)₃CO

	Old	New	Speed-up
Calculation of fit integrals	2892.1	2582.7	1.1
Calculation of kinetic energy matrix	4089.5	759.9	5.4
Calculation of V_{Coulomb} and V_{XC}	926.1	266.3	3.5
Calculation of $\chi(r_k)$	162.3	90.5	1.8
Set-up of Fock matrix	2292.4	416.1	5.5
Determination of fit coefficients	144.7	95.5	1.5
Diagonalization	0.8		

Table 2. Timing results (times in seconds) for C₃₅H₇₂

	Old	New	Speed-up
Calculation of fit integrals	420.5	179.6	2.3
Calculation of kinetic energy matrix	343.2	28.7	12.0
Calculation of V_{Coulomb} and V_{XC}	138.8	31.3	4.4
Calculation of $\chi(r_k)$	24.7	7.3	3.4
Set-up of Fock matrix	170.4	11.5	14.8
Determination of fit coefficients	9.4	3.6	2.6
Diagonalization	8.0		

Table 3. Timing results (times in seconds) for C₂₀₀H₄₀₂

	Old	New	Speed-up
Calculation of fit integrals	10342.2	1280.8	8.1
Calculation of kinetic energy matrix	62777.0	205.0	306.2
Calculation of V_{Coulomb} and V_{XC}	3602.8	352.9	10.2
Calculation of $\chi(r_k)$	765.5	44.5	17.2
Set-up of Fock matrix	35277.0	74.8	471.6
Determination of fit coefficients	459.0	39.6	11.6
Diagonalization	53.5		

system Pt(P(Ph)₃)₃CO is included, where three P(Ph)₃ ligands are directly coordinated to the Pt atom, as is the CO ligand. The number of atoms (105) is equal to that in

the chain-like alkane C₃₅H₇₂. The 602 atom alkane C₂₀₀H₄₀₂ is more typical for the size of system for which the present development is intended. The numbers of STO basis functions are 685, 459 and 2604 for these three systems, respectively.

In the old situation, the calculation of the kinetic energy matrix, which is presented separately since it is carried out once, prior to the SCF cycles, and the setting up of the Fock matrix of the electronic potentials, are the most time-consuming parts of the calculation for Pt(P(Ph)₃)₃CO and in particular for C₂₀₀H₄₀₂. They are also the steps for which (by far) the largest speed-up factors are achieved. Although the total speed-up is already quite significant for the 105-atom systems, the real benefit of the better scaling is, of course, most apparent in the largest system. In all systems the relative weights of computational burden shift from the matrix evaluation to the calculation of the fit integrals. This is notably the case for the Pt complex, which is related to the large number of basis functions on the Pt atom. It should be realized, however, that the calculation of the fit integrals is only executed once, as is the calculation of the kinetic energy matrix, whereas the other steps are repeated on each cycle (in the so-called direct-SCF mode). Incidentally we note that the diagonalization times are still fairly modest and that the diagonalization is not yet a bottleneck, in spite of its N^3 scaling.

We conclude that significant advances have been made towards an $O(N)$ DFT method. The calculation of the fit integrals is still relatively time-consuming but has been scaled down to linear scaling. The calculation of the Coulomb and XC potentials is, in the alkane chains, 4–6 times less expensive, but since its scaling has not been reduced to linear, it will eventually dominate and should be the next target for algorithmic development.

Acknowledgements. The authors wish to thank the National Computing Facilities (NCF) of the Netherlands Foundation for Scientific Research (NWO) for the granting of computer time.

References

1. Baerends EJ, Ellis DE, Ros P (1973) *Chem Phys* 2: 41
2. Ellis DE, Adachi H, Averill FW (1976) *Surf Sci* 58: 497
3. Rosén A, Ellis DE, Adachi H, Averill FW (1976) *J Chem Phys* 65: 3629
4. Delley B, Ellis DE (1982) *J Chem Phys* 76: 1949
5. te Velde G, Baerends EJ (1991) *Phys Rev B* 44: 7888
6. Delley B (1990) *J Chem Phys* 92: 508
7. Pérez-Jordá JM, Yang W (1995) *Chem Phys Lett* 241: 469
8. Sambe H, Felton RH (1975) *J Chem Phys* 62: 1122
9. Kendall RA, Früchtl HA (1997) *Theor Chim Acta* 97: 158
10. Eichkorn K, Weigend F, Treutler O, Ahlrichs R (1997) *Theor Chim Acta* 97: 119
11. Rohklin V (1985) *J Comput Phys* 60: 187
12. Greengard L, Rohklin V (1987) *J Comput Phys* 73: 325
13. Greengard L, Rohklin V (1991) *Commun Pure Appl Math* 44: 419
14. White CA, Johnson BG, Gill PMW, Head-Gordon M (1994) *Chem Phys Lett* 230: 8
15. Strain MC, Scuseria GE, Frisch MJ (1996) *Science* 51: 271
16. Kudin KN, Scuseria GE (1998) *Chem Phys Lett* 283: 61
17. Kutteh R, Aprà E, Nichols J (1995) *Chem Phys Lett* 238: 173
18. Adamson RD, Dombroski JP, Gill PMW (1996) *Chem Phys Lett* 254: 329
19. Dombroski JP, Taylor SW, Gill PMW (1996) *J Phys Chem* 100: 6272
20. Gill PMW (1997) *Chem Phys Lett* 270: 193
21. Goedecker S, Ivanov OV (1998) *Sol Stat Comm* 105: 665
22. Pérez-Jordá JM, Yang W (1997) *J Chem Phys* 107: 1218
23. Challacombe M, Scgwegler E, Almlöf J (1996) *J Chem Phys* 104: 4685
24. Challacombe M, Schwegler E (1997) *J Chem Phys* 106: 5526
25. Stratman RE, Scuseria GE, Frisch MJ (1996) *Chem Phys Lett* 257: 213
26. Li XP, Nunes RW, Vanderbilt D (1993) *Phys Rev B* 47: 1089
27. Galli G, Parrinello M (1992) *Phys Rev Lett* 69: 3547
28. Hernández E, Gillan MJ (1995) *Phys Rev B* 51: 10157
29. Hernández E, Gillan MJ, Goringe CM (1996) *Phys Rev B* 53: 7147
30. Ordejón P, Drabold DA, Martin RM, Grumbach MP (1995) *Phys Rev B* 51: 1456
31. Sánchez-Portal D, Ordejón P, Artacho E, Soler JM (1997) *Int J Quantum Chem* 65: 453
32. Stewart JP (1996) *Int J Quantum Chem* 58: 133
33. Xu CH, Scuseria GE (1996) *Chem Phys Lett* 262: 219
34. Millam JM, Scuseria GE (1997) *J Chem Phys* 106: 5569
35. Daniels AD, Millam JM, Scuseria GE (1997) *J Chem Phys* 107: 425
36. Dunlap BI, Connolly JWD, Sabin JR (1979) *J Chem Phys* 71: 3396
37. Gallant RT, St-Amant A (1996) *Chem Phys Lett* 256: 569
38. Fonseca Guerra C, Visser O, Snijders JG, te Velde G, Baerends EJ (1995) In: Clementi E, Corongiu G (eds) *Methods and techniques for computational chemistry*. STEF, Cagliari, p 305
39. Treutler O, Ahlrichs R (1995) *J Chem Phys* 102: 346
40. Frisch MJ, Trucks GW, Schlegel HB, Gill PMW, Johnson BG, Robb MA, Cheeseman JR, Keith T, Petersson GA, Montgomery JA, Raghavachari K, Al-Laham MA, Zakrzewski VG, Ortiz JV, Foresman JB, Cioslowski J, Stefanov BB, Nanayakkara A, Challacombe M, Peng CY, Ayala PY, Chen W, Wong MW, Andres JL, Replogle ES, Gomperts G, Martin RL, Fox DJ, Binkley JS, Defrees DJ, Baker J, Stewart JP, Head-Gordon M, Gonzalez C, Pople JA (1995) *GAUSSIAN 94*, Gaussian, Pittsburgh, Pa
41. Boerrigter PM, te Velde G, Baerends EJ (1988) *Int J Quantum Chem* 33: 87
42. te Velde G, Baerends EJ (1992) *J Comput Phys* 99: 84
43. Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) *Numerical recipes*. Cambridge University Press, Cambridge