

The effect of grid quality and weight derivatives in density functional calculations

Jon Baker, Jan Andzelm, Andrew Scheiner, and Bernard Delley

Citation: *The Journal of Chemical Physics* **101**, 8894 (1994); doi: 10.1063/1.468081

View online: <http://dx.doi.org/10.1063/1.468081>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jcp/101/10?ver=pdfcov>

Published by the [AIP Publishing](#)

Articles you may be interested in

[Grid-free density functional calculations on periodic systems](#)

J. Chem. Phys. **127**, 114108 (2007); 10.1063/1.2768358

[The effect of grid quality and weight derivatives in density functional calculations of harmonic vibrational frequencies](#)

J. Chem. Phys. **119**, 12763 (2003); 10.1063/1.1627291

[Constraints on density functionals and density functional derivatives](#)

J. Chem. Phys. **110**, 1873 (1999); 10.1063/1.477871

[Weighted density functionals for ferroelectric materials](#)

AIP Conf. Proc. **436**, 251 (1998); 10.1063/1.56276

[Improved radial grids for quadrature in molecular densityfunctional calculations](#)

J. Chem. Phys. **104**, 9848 (1996); 10.1063/1.471749



The effect of grid quality and weight derivatives in density functional calculations

Jon Baker, Jan Andzelm, and Andrew Scheiner
Biosym Technologies, Inc., 9685 Scranton Road, San Diego, California 92121

Bernard Delley
Paul Scherrer Institut Zurich, Badenerstrasse 569, CH-8048 Zurich, Switzerland

(Received 24 June 1994; accepted 2 August 1994)

Full density functional geometry optimizations on hydrogen peroxide and heptane/dimethyl pentane using six different numerical grids are presented. The grids vary in quality and gradients are calculated (1) assuming a fixed grid and no weight derivatives, and (2) with full allowance for a "moving" atom-centered grid and inclusion of the weight derivatives. The results clearly demonstrate that accurate energies and geometries can be obtained with around 3500 points per atom for medium-sized systems (up to say 30 atoms) *without* the necessity of including the weight derivatives. The latter only begin to influence the results for grids which are of insufficient quality to guarantee reliable values in any case.

I. INTRODUCTION

The Kohn–Sham version of density functional theory (DFT)¹ is becoming increasingly popular in computational quantum chemistry and a number of groups are actively involved in this area.^{2–10} Popular *ab initio* programs, e.g., GAUSSIAN 92,¹¹ CADPAC¹² are now including DFT capability and specialized DFT packages such as DMOL,¹³ DGAUSS,¹⁴ and DEMON¹⁵ have become well established.

The self-consistent Kohn–Sham procedure¹ can be formulated so that it resembles standard Hartree–Fock theory⁹ with a Fock matrix given by

$$\mathbf{F} = \mathbf{H} + \mathbf{J} + \mathbf{K}^{\text{xc}} \quad (1)$$

where \mathbf{H} and \mathbf{J} are the usual one-electron and Coulomb matrices and \mathbf{K}^{xc} is the DFT exchange–correlation matrix (which replaces the usual exchange-only \mathbf{K} matrix in Hartree–Fock theory).

A general density functional can be expressed as a function of the alpha and beta electron densities and their gradient invariants,⁹

$$F = F[\rho_\alpha(r), \rho_\beta(r), \gamma_{\alpha\alpha}(r), \gamma_{\alpha\beta}(r), \gamma_{\beta\beta}(r)], \quad (2)$$

where

$$\begin{aligned} \gamma_{\alpha\alpha}(r) &= \nabla \rho_\alpha(r) \cdot \nabla \rho_\alpha(r), \\ \gamma_{\alpha\beta}(r) &= \nabla \rho_\alpha(r) \cdot \nabla \rho_\beta(r), \\ \gamma_{\beta\beta}(r) &= \nabla \rho_\beta(r) \cdot \nabla \rho_\beta(r). \end{aligned} \quad (3)$$

The exchange–correlation energy is then given by the integral

$$E^{\text{xc}} = \int F[\rho_\alpha(r), \rho_\beta(r), \dots] dr. \quad (4)$$

Although the various DFT codes^{11–15} have different methodologies for solution of Eq. (1), and different ways of dealing with the various terms, they all calculate the exchange–correlation energy, Eq. (4), and the elements of the \mathbf{K}^{xc} matrix by numerical integration, the standard technique being to partition the multicenter molecular integrals—such as Eq.

(4)—into single-center atomic integrals which are then evaluated numerically on an atomic grid. Thus E^{xc} is given by

$$E^{\text{xc}} = \sum_A \sum_i \omega_{Ai} F(\rho_\alpha, \rho_\beta, \gamma_{\alpha\alpha}, \gamma_{\alpha\beta}, \gamma_{\beta\beta}; r_{Ai}), \quad (5)$$

where the first summation is over all the atoms in the system and the second is over the numerical quadrature grid points associated with that atom. The ω_{Ai} are quadrature weights and $r_{Ai} = \mathbf{R}_A + \mathbf{r}_i$, where \mathbf{R}_A is the position of atom A and \mathbf{r}_i is the location of grid point i on a spherical grid centered on atom A .

Most current grid designs stem from procedures popularized in this context by Becke.¹⁶ Essentially the single-center integrations are separated into radial and angular parts, with spherical surfaces defined at each point on a radial grid on which to carry out the angular integrations. The angular part corresponds to quadrature on the surface of a sphere and several formulas are available to do these integrations, the most commonly used are probably those due to Lebedev.¹⁷ A grid of degree l will exactly integrate any spherical harmonic function of degree l or less and grids exist for orders l up to 35 and higher. The $l=35$ grid has 434 grid points.¹⁸ Lebedev quadratures have relatively straightforward point distributions of octahedral symmetry.

It is possible to further separate the angular integration into two parts and use the more familiar one-dimensional Gauss–Legendre formalism which is simpler and more flexible than the two-dimensional quadrature scheme. Although more grid points are required for a given accuracy, Handy and co-workers recommend this approach¹⁹ and it is often used by other groups for benchmark calculations.

With the exchange–correlation energy given by Eq. (5) its derivative with respect to nuclear displacement is formally^{9,20}

$$\nabla_B E^{\text{xc}} = \sum_A \sum_i [\omega_{Ai} \nabla_B F(r_{Bi}) + \nabla_B \omega_{Ai} F(r_{Bi})]. \quad (6)$$

There are two features of Eq. (6) to which we draw explicit attention. The first is that the gradient of E^{xc} clearly has *two* components, and since the quadrature weights depend on the nuclear configuration, the second term—involving the weight derivatives—is *not* zero and should be included for a proper treatment of the energy gradient. However, the general consensus appears to be that its contribution is likely to be small and should tend to zero by taking a large enough number of grid points. There is currently some dispute amongst DFT practitioners as to how important the weight derivative term is; for example, Delley comments that leaving out the weight derivatives “results in a small residual (e.g., 10^{-3} a.u.) at the energy minimum” (the grid used is not stated)²⁰ which he apparently considers acceptable whereas Johnson *et al.* “feel it is too large to be ignored.”⁹ We will demonstrate in this article that although inclusion of the weight derivatives does have an effect, this only becomes significant for grids which are too poor to guarantee reliable results in any case.

The second point is more subtle and although it has been fully appreciated and commented on in the excellent papers of Johnson *et al.*⁹ and Handy and co-workers,¹⁰ it appears to have been completely overlooked by the earlier practitioners and is certainly a major factor contributing to the “noise” in gradients calculated with DMOL, DEMON and early versions of DGAUSS. For this reason we dwell upon it in some detail.

To simplify the discussion, we consider a local density functional, i.e., one dependent on the density only and not on any density derivatives, for a closed-shell system for which the gradient of the exchange-correlation energy—the first term in Eq. (6)—can be written⁹

$$\nabla_B E^{\text{xc}} = -2 \sum_{\mu} \sum_{\nu} P_{\mu\nu} \sum_A \sum_i \times \omega_{Ai} \frac{\partial F}{\partial \rho}(r_{Ai}) \phi_{\nu}(r_{Ai}) \nabla \phi_{\mu}(r_{Ai}), \quad (7)$$

where the restricted sum over μ runs only over the basis functions centered on atom B . P is the closed-shell density matrix, and $\phi_{\nu}(r_{Ai})$ refers to the value of basis function ϕ_{ν} at grid point \mathbf{r}_{Ai} .

One way of thinking of the numerical integration involved in evaluating expressions such as Eq. (7) is to consider atoms embedded in a *fixed* Cartesian grid. If atom B is displaced (in order to calculate its gradient) then the distance between B and *all* points in the grid will change and so there will be a contribution to the gradient on atom B from the derivatives of those basis functions centered on B —since B is moving—interacting with the basis functions centered on *all* other atoms (including B itself) evaluated at every grid point. We can write this as

$$\nabla_B E^{\text{xc}} = \sum_j \sum_A \nabla \Phi_B(j) \Phi_A(j), \quad (8)$$

where we have “ignored” all the density terms so as to focus exclusively on the quantities of interest, namely, basis functions and their derivatives evaluated over the grid points. Equation (8) should be read as the summation over *all* grid

points j of *all* basis functions centered on *all* atoms A with the derivatives of those basis functions centered on B . This represents the exchange-correlation energy contribution to the gradient with respect to displacement of atom B within a *fixed* grid.

The problem with the above analysis of course is that in practice the numerical grid used to evaluate the exchange-correlation energy is *not* fixed as the total grid is a superposition of individual atom-centered grids and hence *those grid points formerly associated with atom B will “move with” a displacement of its nucleus.*

Returning to our previous argument this means that if atom B is displaced then the distance between B and all points on *other* atomic grids will change but the distance between nucleus B and points on its *own* grid will remain unchanged and thus will *not* contribute to the gradient on B . However, since the grid centered on B “moves with” B , then the distance between all other atoms in the system and those grid points associated with B *will* change and there will be a contribution to B ’s gradient from this source. Moving the grid on B while keeping all other atoms fixed is formally equivalent to keeping the grid points on B fixed and moving the other atoms and so the gradient on B will involve the derivatives of all basis functions centered on the *other* atoms evaluated over B ’s grid points (and contributing with the opposite sign).

For our atom-centered grid then, Eq. (8) must be replaced by

$$\begin{aligned} \nabla_B E^{\text{xc}} = & \sum_{j \in B} \sum_A \nabla \Phi_B(j) \Phi_A(j) \\ & - \sum_{j \in B} \sum_A \sum_{C \neq B} \nabla \Phi_C(j) \Phi_A(j). \end{aligned} \quad (9)$$

Thus we see that a given grid point j *not* part of atom B ’s grid will contribute to the gradient on B in exactly the same way as for a fixed grid [Eq. (8); the first term of Eq. (9)], whereas if j *is* on B it contributes to B ’s gradient in a much more complicated way [the second term of Eq. (9)]. However, there is in fact no need to calculate this additional term once it is realized that the sum of the contributions evaluated at point j to the gradients on *all* atoms *except* B (if j is on B) is *exactly* equivalent to its contribution to the gradient on B with the opposite sign. Hence for all grid points j on atom B ’s grid, their contributions to the gradient on all atoms *except* B can be calculated, and their contributions to B ’s gradient obtained by invoking translational invariance ($\sum_A \nabla_A = 0$).

This has previously been pointed out by Johnson *et al.*⁹ and it is clear from the arguments presented above that the first term in the expression for the exchange-correlation gradient is exactly translationally invariant at *each* grid point. A similar analysis (see later) can be applied to the second term in Eq. (6)—the weight derivatives—and both terms individually possess point-by-point translational invariance. Johnson has recently taken this a step further and has presented a scheme which guarantees rotational invariance as well.²¹

Although we may have laboured the point we feel that it is an important one which, despite a previous airing,⁹ is still not fully appreciated throughout the DFT community. Early DFT programs calculated gradients with both neglect of the weight derivative term *and* the assumption of a fixed grid, and although this was reasonable for a first pass, it is clearly incorrect and—as we have indicated earlier—is one of the main reasons for the so-called “numerical noise” in some DFT gradients and also for the fact that in such codes the gradient zero does *not* correspond to the true energy minimum.

In this paper we present full DFT geometry optimizations for hydrogen peroxide and heptane/dimethyl pentane (concentrating on the energy difference between these two isomers) using six different grids varying in quality from poor to essentially “saturated.” Calculations are performed with (1) gradients calculated assuming a fixed grid and no weight derivatives (“old” gradients); and (2) gradients handled properly with a “moving” grid and inclusion of the weight derivative term (“new” gradients). Our aim here is twofold: First we wish to determine what grid quality, i.e., number and distribution of grid points, is necessary for reliable results which in this context we consider to be energies converged to within ± 0.2 kcal/mol, bond lengths to within ± 0.001 Å, bond angles to within $\pm 0.2^\circ$, and dihedral angles to within $\pm 1^\circ$. Second is to assess the importance of treating the gradient correctly, i.e., under what circumstances does using incorrect gradients significantly affect the results? Pople and co-workers have addressed grid standardization in a recent publication.²²

In the following section (Sec. II) we present details of our DFT methodology with emphasis on grid construction and definition of the quadrature weights; in Sec. III we present our results including details of the convergence pattern during the optimizations and selected geometrical parameters. Section IV comprises a summary and conclusions.

II. METHODOLOGY

Our approach to DFT closely parallels that of the Pople⁹ and Handy¹⁰ groups, in that we treat all one-electron and the two-electron Coulomb terms in Eq. (1) exactly and only the exchange-correlation term is treated numerically.

As discussed in the previous section, the numerical integration is carried out over a series of spherical atom-centered grids. We have based our grid on the scheme used in DMOL which was developed by Delley.²³ In this scheme the angular quadrature weights are derived from superpositions of spherical atomic densities for all the atoms that make up the system being studied,

$$W_{tA}(i) = \rho_A(i)^n / \sum_B \rho_B(i)^n. \quad (10)$$

Here $W_{tA}(i)$ is the weight at the i th grid point centered on atom A and $\rho_A(i)$ is the atomic density at that point from atom A . The denominator on the right-hand side of Eq. (2) is the sum of the individual density powers from all the atoms (including A) in the molecule at point i . Clearly

$\sum W_{tA}(i) = 1$. Note that n can in principle be any power (indeed, one is not limited merely to a polynomial); for this work we have simply taken $n=2$. The “best” functional form to use for the weights is currently being investigated and will be the subject of a future publication.

In our version of the DMOL scheme UHF wave functions (with fractional orbital occupancies if necessary to ensure spherical symmetry) are calculated for each different atom in the system and the converged MOs are saved and then reused to calculate the angular quadrature weights for the molecular grid. The atomic solutions can either be found for the same density functional and basis set used in the molecular calculation or a “standard” set of UHF orbitals can be calculated once and for all with a given functional and basis and this set adopted for all molecular calculations regardless of the actual basis and functional used.

With quadrature weights defined according to Eq. (10), their derivative with respect to atomic displacement is

$$\nabla_B W_{tA}(i) = \frac{-\rho_{A(i)}^n n \rho_{B(i)}^{n-1} \nabla_B \rho_{B(i)}}{[\sum \rho_{c(i)}^n]^2} + \frac{n \rho_{A(i)}^{n-1} \nabla_B \rho_{A(i)}}{\sum \rho_{c(i)}^n}. \quad (11)$$

The second term on the right-hand side of Eq. (11) is zero unless $B=A$.

Equation (11) is the formal derivative of Eq. (10) assuming a *fixed* grid, and in an exactly analogous manner to Eq. (7) only those terms in Eq. (11) for which $A \neq B$ are actually calculated, the $A=B$ term, i.e., the weight derivative for an atom at a point on its own atomic grid, is obtained by translational invariance. Using Eq. (11), the full weight derivative term in Eq. (6) is thus straightforward to calculate.

The radial grid on each atom has a maximum extent of 12 bohr with points distributed logarithmically at distances r from the nucleus according to

$$r = RFac \times \log\{1 - [i/(1 + NRad)]^2\} \quad (12a)$$

with

$$RFac = 12 / \log\{1 - [NRad/(1 + NRad)]^2\} \quad (12b)$$

and $NRad$, the maximum number of radial points, given by

$$NRad = SP \times 14 (\text{Atomic number} + 2)^{1/3}. \quad (12c)$$

SP is a scaling factor (currently $SP=1.2$). In Eq. (12a) i runs from 1 to $NRad$. Use of Eq. (12c) gives 24, 33, 34, 36, and 44 radial points for H, C, N, O, and S, respectively. The radial weights are given by

$$Wr = \frac{2r^2 i [1/(1 + NRad)]^2 \times RFac}{1 - [i/(1 + NRad)]^2}. \quad (12d)$$

For the angular integrations we use Lebedev grids of degree 5, 7, 11, 17, 23, 29, or 35 which have 14, 26, 50, 110, 194, 302, and 435 points, respectively. As mentioned in the Introduction we use various different grids and these are distinguished by the maximum Lebedev degree allowed on each sphere. Our best grid, grid 1, has maximum degree 35; grids 2–5, in decreasing order of accuracy, have maximum degree 29, 23, 17, and 11, respectively. We emphasise that, for all grids, both the radial extent and angular degree given are the *maximum* allowed. The DMOL grids are *adaptive* and this

TABLE I. Characteristics of the numerical grids used in this work. (See text for more details.)

Grid	Maximum degree	Maximum points per shell	Average number points per atom
big	35	435 ^a	13 000
1	35	435	4 100
2	29	303	3 500
3	23	194	2 700
4	17	110	1 800
5	11	50	1 000

^a435 angular points in each shell.

maximum is only utilized if deemed absolutely necessary. It is well known that the closer one gets to the nucleus the more spherical the electron density which thus becomes amenable to treatment by progressively less sophisticated angular grids. For example, Gill *et al.*²² divide the radial grid into five different regions, innermost to outermost, in which they employ angular grids with 6, 38, 86, 194, and 86 points, respectively. This strategy, which they term “grid pruning,” has also been used by Handy and co-workers.¹⁹

In the adaptive scheme, one calculates the total electron density over each sphere—as a superposition of the individual atomic densities—using the lowest order angular grid available (which is our case is of degree 5). The grid is then progressively improved and a new value for the density over the sphere calculated using the improved grid; this procedure continues until either the maximum allowed angular degree is reached or the difference in the calculated density between the current and previous grid is less than some tolerance TolANG (we use a value of 10^{-8}). The radial grid is terminated when the total density over the last sphere calculated is below a second tolerance TolRAD (value used in this work is 10^{-20}). This procedure, which has been used in DMOL for several years, can be considered as taking the “grid pruning” strategy to its limit in which each sphere is potentially a single region.

In addition to the five grids defined above we use a sixth grid with up to 60 radial points and our maximum of 435 angular points in *each* shell. We consider this grid to be essentially “saturated” and it is used as a benchmark to gauge the quality of our adaptive grids. The pertinent characteristics of the grids we have used in this paper are summarized in Table I.

III. RESULTS

In this section we report our results on the total energy and geometry of hydrogen peroxide and the energy difference between heptane and dimethyl pentane using each of the six grids outlined in the previous section, employing both “old” and “new” gradients (as defined in the Introduction). We have used the Vosko, Wilk, and Nusair (VWN)²⁴ local functional together with the STO-3G basis for our calculations on heptane/dimethyl pentane and the Becke, Lee, Yang, and Parr (BLYP)^{25,26} nonlocal functional with the 6-31G* basis for hydrogen peroxide. The atomic densities required to

set up the grids were actually calculated using the 6-31G** basis (UHF wave functions precalculated with this basis form our current “standard” for determining the quadrature weights). Since we are concerned here with how the geometrical parameters and energetics change with the different grids, the actual values obtained are only of secondary importance.

Before commenting on our results it is important to ask what behavior would we hope to see during the course of a geometry optimization? For a typical minimization we would expect smooth convergence with both energy and gradient dropping on every cycle and with the final converged energy being lower than all other energies calculated during the optimization. An occasional rise in energy due to a poor step or towards the end of the optimization for an especially floppy molecule is acceptable but the final geometry should still show the lowest energy. This type of behavior is precisely what one observes during standard *ab initio* computations and although DFT has a large numerical component we would hope to observe a similar convergence pattern if the numerics are being treated adequately.

We first report our results for hydrogen peroxide. Table II shows the convergence pattern for the optimizations with all six grids (all optimizations were of course started from the same initial geometry). Also shown are the maximum loss of translational invariance in the Cartesian gradient vector for the “old” gradients (“new” gradients are of course translationally invariant), the number of grid points and the total number of electrons (obtained by integrating the density over the grid) at the final converged geometry (these quantities are in fact more-or-less geometry independent). The geometrical parameters at convergence are given in Table III. The optimizations were carried out with the EF algorithm²⁷ using the OPTIMIZE package.²⁸ Convergence criteria were the default values of 0.0003 a.u. on the maximum gradient component and *either* 0.0003 a.u. on the maximum displacement *or* an energy change from the previous cycle of less than 0.000 001 hartree. Energies and gradients were calculated using our DFT-modified version of TURBOMOLE.²⁹

The total number of electrons as determined by integrating the final density over the grid and the loss of translational invariance (for the “old” gradients) can be taken as a measure of grid quality. Both these measures show the decreasing quality of the grids, with a particularly marked decline between grids 2 and 3.

Looking at the energy at the starting geometry (which is the same in all cases), this is identical to four decimal places for all grids except the last (grid 5); thus based on an energy criterion alone, grids 1–4 would all seem to be reliable. Examining the convergence pattern and the final energy for the “new” gradients, we see excellent behavior in all cases, with smooth convergence, energy decreasing on every cycle and the final energy being the lowest obtained, i.e., exactly what we would hope to see. Additionally, the final converged energy is the same to within 0.0003 hartree, which satisfies our criterion for reliable results (± 0.2 kcal/mol) for all grids except grid 5, and even here the error is only 0.000 32 hartree. Again, this would strongly suggest that grids 1–4 are all acceptable.

TABLE II. Convergence pattern for BLYP/6-31G* geometry optimizations of hydrogen peroxide with the six numerical grids defined in Table I. Optimizations were carried out both with (new) and without (old) inclusion of the quadrature weight derivatives in the DFT gradient.

New Grad.	Max. Grad.	Old Grad.	Max. Grad.
Grid Big (51 212 points; 17.999 989 electrons over grid)			
-151.511 670 022	0.016 142		
.512 236 185	0.007 528		
.512 407 101	0.002 926		
.512 435 126	0.001 174		
.512 437 290	0.000 166		
.512 437 448	0.000 161		
Grid 1 (19 772 points; 18.000 007 electrons over grid)			
-151.511 671 005	0.016 134	-151.511 671 005	0.016 101
.512 238 593	0.007 518	.512 237 672	0.007 511
.512 408 999	0.002 932	.512 408 311	0.002 921
.512 437 146	0.001 174	.512 437 148	0.001 179
.512 439 290	0.000 147	.512 439 573	0.000 185
.512 439 528	0.000 150	.512 439 736	0.001 179
maximum loss of translational invariance at convergence: 2.2×10^{-5}			
Grid 2 (15 908 points; 18.000 008 electrons over grid)			
-151.511 679 189	0.016 129	-151.511 679 189	0.016 113
.512 244 275	0.007 507	.512 243 354	0.007 509
.512 414 738	0.002 923	.512 413 793	0.002 920
.512 442 784	0.001 162	.512 442 754	0.001 181
.512 444 861	0.000 129	.512 445 149	0.000 165
.512 444 958	0.000 124	.512 445 274	0.000 159
maximum loss of translational invariance at convergence: 2.6×10^{-5}			
Grid 3 (11 636 points; 18.000 057 electrons over grid)			
-151.511 702 609	0.016 118	-151.511 702 609	0.016 118
.512 271 692	0.007 448	.512 270 200	0.007 499
.512 440 994	0.002 895	.512 439 098	0.002 918
.512 392 300	0.000 903	.512 466 330	0.001 197
.512 471 244	0.000 354	.512 467 517	0.000 314
.512 471 674	0.000 506	.512 466 680	0.000 309
.512 472 085	0.000 144	.512 460 184	0.000 896
		.512 448 355	0.001 081
		.512 421 425	0.000 674
		.512 428 359	0.000 297
		.512 428 268	0.000 023
maximum loss of translational invariance at convergence: 4.0×10^{-4}			
Grid 4 (7 652 points; 18.000 360 electrons over grid)			
-151.511 708 029	0.016 011	-151.511 708 029	0.016 268
.512 194 548	0.007 702	.512 263 549	0.007 469
.512 451 688	0.003 292	.512 427 197	0.002 983
.512 485 151	0.001 409	.512 447 465	0.001 356
.512 490 645	0.000 583	.512 437 911	0.001 245
.512 493 336	0.000 513	.512 402 519	0.001 639
.512 500 549	0.001 204	.512 318 847	0.002 780
.512 507 487	0.000 932	.512 269 895	0.001 992
.512 507 804	0.000 291	.512 267 652	0.000 474
		.512 272 299	0.000 013
maximum loss of translational invariance at convergence: 4.8×10^{-4}			
Grid 5 (4 268 points; 18.000 393 electrons over grid)			
-151.510 563 405	0.014 943	-151.510 563 405	0.016 255
.511 276 604	0.006 808	.511 012 443	0.007 223
.511 898 834	0.006 750	.511 107 258	0.002 671
.512 251 146	0.007 483	.511 071 737	0.001 250
.512 520 031	0.007 170	.511 008 230	0.001 232
.512 719 345	0.003 202	.510 841 593	0.001 593
.512 756 004	0.001 547	.510 579 082	0.002 494
.512 760 861	0.000 245	.510 458 809	0.001 655
.512 761 012	0.000 028	.510 443 430	0.000 281
		.510 450 526	0.000 010
maximum loss of translational invariance at convergence: 1.1×10^{-3}			

TABLE III. Selected geometrical parameters (angstroms and degrees) for hydrogen peroxide calculated with the BLYP functional and the 6-31G* Gaussian basis for the six numerical grids defined in Table I. Geometries were optimized both with (new) and without (old) inclusion of the quadrature weight derivatives in the DFT gradient (see text for more details).

Grid		rOO	rOH	aHOO	oHOOH
big	new	1.494	0.985	98.4	121.9
1	new	1.494	0.985	98.4	122.0
1	old	1.494	0.985	98.4	121.9
2	new	1.494	0.985	98.4	122.0
2	old	1.494	0.985	98.4	121.9
3	new	1.495	0.985	98.4	122.8
3	old	1.493	0.986	98.6	118.3
4	new	1.496	0.985	98.3	126.2
4	old	1.492	0.986	98.7	113.9
5	new	1.502	0.983	97.7	142.8
5	old	1.493	0.986	98.4	114.3

The picture with the “old” gradients is somewhat different. Although grids 1 and 2 show good convergence behavior and give essentially the same results as the corresponding optimizations with the “new” gradients, there are obvious problems with grid 3 which worsen significantly with grids 4 and 5. For grid 3, although the lowest gradient is attained after 11 optimization cycles, the lowest energy actually occurs on cycle 5; thereafter the energy actually increases at every cycle (with a very slight decrease at the end of the optimization). This is a clear manifestation of the commonly observed phenomenon in many early DFT calculations of the gradient zero not corresponding to the energy minimum due to the gradient not being calculated properly, as outlined in the Introduction.

With grid 3 the final “converged” energy is only some 0.000 04 hartree higher than the lowest energy attained during the optimization; however for grid 4 this difference has increased to 0.000 18 hartree and for grid 5 to 0.000 66 hartree. The “converged” energy for grid 5 is in fact *higher* than the starting energy! Clearly not calculating the gradient correctly has severe repercussions for the poorer grids, and for the “old” gradients only grids 1 and 2 seem to provide an acceptable performance.

Considering the energetics and convergence pattern, then the conclusion thus far is that grids that are obviously unacceptable—have too few quadrature points—for use with “old” gradients can be “cleaned up” and made functional if the gradient is calculated correctly. However, if we examine the converged geometries (Table III), this picture changes. Grids 1 and 2 give identical final geometrical parameters—to each other and to the “saturated” grid—regardless of how the gradient is calculated. Deviations first start to appear with grid 3 and thereafter worsen (not unexpectedly) as grid quality decreases. Especially noticeable is the change in the HOOH dihedral angle. The important point to note is that geometrical deviations are unacceptable for grids 3–5 even with the “new” gradients; in fact if anything deviations are even worse with the “new” than with the “old” gradients.

Thus our final conclusion is that *only grids with a suffi-*

cient number of quadrature points provide acceptable results and if grid quality is good then *it is not necessary to include quadrature weight derivatives in the gradient* although this is of course preferable for consistency and final accuracy. In fact, if one looks solely at the energetics, then results with a proper gradient treatment can be misleading. The performance of grid 5 with the “new” gradients is excellent; only when the final geometrical parameters are examined does one realize how bad this grid actually is.

The above conclusions are reinforced by our calculations on dimethyl pentane/heptane (Table IV). Heptane is a rigid molecule with many of its geometrical parameters fixed by symmetry (C_{2v}); dimethyl pentane on the other hand is a somewhat extended, floppy molecule with no symmetry which is expected to be a fairly severe test of our numerical grids.

Overall, the convergence pattern shows similar characteristics as for hydrogen peroxide, with all grids converging smoothly with the “new” gradients (there is some “noise” for dimethyl pentane, probably due to the lack of rotational invariance in our gradient treatment, but this is very minor and only affects the energy in the sixth decimal place) and some obvious inconsistencies with the “old” gradients, with energies clearly rising at the end of the optimization with grids 4 and 5. Both the absolute and relative energies are in essential agreement with grids 1 and 2, regardless of how the gradients are calculated; thereafter there is a marked decline in quality and by the time we reach grid 5 the energy ordering between the two isomers has actually reversed (from -0.55 to $+1.14$ kcal/mol with the “new” gradients and $+1.83$ kcal/mol with the “old”).

Another feature that should be noted, for both hydrogen peroxide and dimethyl pentane (and which is in fact general), is that as grid quality worsens the number of optimization cycles required to attain convergence increases. This tends to counteract any reduction in cpu time expected due to the reduced number of grid points, even if the results were any good, which they are not. The whole grid argument becomes moot for very large systems in any case, since for the DFT formalism used in this work the entire calculation will then be dominated by the time taken to calculate and treat the Coulomb integrals.

We end with a couple of technical points. The first is that if the grid is treated as “moving with the atom” then—at least with the method of forming the quadrature weights we have adopted—weight derivatives must be included to obtain anywhere near correct gradients; the gradients are just plain wrong if this term is left out. Secondly, it should be noted that there are basically two formalisms for representing the so-called nonlocal functionals (i.e., functionals which involve density derivatives in their definition, such as BLYP)—spectral^{9,30} and nonspectral. The latter formalism involves density second derivatives in the functional definition (see e.g. Ref. 26); the former does not. Due to subtle effects involving integration by parts combined with numerical quadrature,¹⁰ it turns out that correct gradients *cannot* be obtained in the manner described in this article using a moving grid for the nonspectral formalism; a fixed grid must be assumed when calculating nonspectral gradients. Our current

TABLE IV. Convergence pattern for VWN/STO-3G geometry optimizations of heptane and dimethylpentane with the six numerical grids defined in Table I. Optimizations were carried out both with (new) and without (old) inclusion of the quadrature weight derivatives in the DFT gradient.

New gradients		Old gradients	
Grid big (266 910 points; 57.999 980 electrons over grid)			
heptane	dimethylpentane		
-270.449 038 099	-270.458 663 497		
.460 512 218	.461 943 681		
.461 230 875	.462 211 941		
.461 552 153	.462 283 061		
.461 557 154	.462 348 083		
.461 557 968	.462 407 647		
	.462 426 850		
	.462 429 357		
	.462 429 844		
Energy difference:	-0.000 872		
	-0.55 kcal/mol		
Grid 1 (93 738 points; 57.999 932 electrons over grid)			
heptane	dimethylpentane	heptane	dimethylpentane
-270.449 039 362	-270.458 684 720	-270.449 039 362	-270.458 684 720
.460 499 990	.461 946 302	.460 504 744	.461 943 152
.461 213 368	.462 211 563	.461 267 548	.462 215 126
.461 541 569	.462 345 239	.461 536 790	.462 280 446
.461 546 681	.462 347 896	.461 545 880	.462 369 073
.461 547 516	.462 412 557	.461 547 282	.462 417 163
	.462 431 650	.461 547 259	.462 431 073
	.462 429 356		.462 429 467
	.462 432 614		.462 429 387
	.462 431 126		
	.462 430 879		
maximum loss of translational invariance at convergence: 5.2×10^{-5}			
Energy difference:	-0.000 883		-0.000 882
	-0.55 kcal/mol		-0.55 kcal/mol
Grid 2 (75 654 points; 58.000 193 electrons over grid)			
heptane	dimethylpentane	heptane	dimethylpentane
-270.449 008 496	-270.458 706 340	-270.449 008 496	-270.458 706 340
.460 428 372	.461 991 408	.460 445 047	.461 990 687
.461 158 851	.462 266 670	.461 171 673	.462 255 074
.461 489 836	.462 316 657	.461 488 348	.462 318 832
.461 495 189	.462 382 394	.461 494 289	.462 434 693
.461 496 068	.462 445 118	.461 495 850	.462 433 917
	.462 464 475		.462 459 250
	.462 467 475		.462 469 198
	.462 469 430		.462 470 127
	.462 459 242		
	.462 475 678		
	.462 475 317		
maximum loss of translational invariance at convergence: 7.0×10^{-5}			
Energy difference:	-0.000 979		-0.000 974
	-0.61 kcal/mol		-0.61 kcal/mol
Grid 3 (58 458 points; 58.000 730 electrons over grid)			
heptane	dimethylpentane	heptane	dimethylpentane
-270.450 017 283	-270.458 587 663	-270.450 017 283	-270.458 587 663
.461 274 252	.461 993 268	.461 312 686	.461 947 932
.461 965 444	.462 258 956	.461 944 244	.462 253 887
.462 288 609	.462 322 398	.462 287 867	.462 331 407
.462 292 842	.462 389 961	.462 290 071	.462 347 541
.462 293 531	.462 386 373	.462 288 974	.462 346 257
	.462 405 723		.462 335 802
	.462 420 051		.462 330 444
	.462 485 979		.462 335 919
	.462 480 475		.462 350 018
	.462 473 087		.462 373 092
	.462 520 405		.462 379 375
	.462 567 176		.462 390 012
	.462 591 157		.462 387 900
	.462 467 028		.462 393 439
	.462 579 019		.462 410 678
	.462 593 986		.462 420 657
	.462 599 429		.462 422 285
	.462 605 183		.462 422 360
	.462 619 516		
	.462 611 759		
	.462 615 607		
	.462 635 554		
	.462 638 427		

TABLE IV. (*Continued.*)

New gradients		Old gradients	
maximum loss of translational invariance at convergence: 2.7×10^{-4}			
Energy difference:	-0.000 345		-0.000 133
	-0.22 kcal/mol		-0.08 kcal/mol
Grid 4 (40 158 points; 58.001 408 electrons over grid)			
heptane	dimethylpentane	heptane	dimethylpentane
-270.448 377 611	-270.458 192 002	-270.448 377 611	-270.458 192 002
.461 298 245	.461 527 723	.460 642 676	.461 405 042
.461 501 508	.462 326 747	.461 395 515	.461 865 223
.461 822 992	.462 660 241	.461 722 255	.462 310 893
.461 829 801	.462 994 660	.461 738 007	.462 841 729
.461 830 732	.463 104 635	.461 744 639	.462 850 180
	.463 125 114		.462 891 100
	.463 133 701		.462 927 344
	.463 144 014		.462 924 173
	.463 154 446		.462 917 938
	.463 161 229		.462 901 116
	.463 177 546		.462 831 891
	.463 198 803		.462 735 427
	.463 225 368		.462 787 129
	.463 249 557		.462 807 289
	.463 258 000		.462 806 406
	.463 259 669		.462 808 083
	.463 262 342		.462 819 522
	.463 270 651		.462 834 650
	.463 281 939		.462 839 049
	.463 293 057		.462 840 126
	.463 303 108		.462 840 913
	.463 306 592		
	.463 307 549		
maximum loss of translational invariance at convergence: 7.5×10^{-4}			
Energy difference:	-0.001 478		-0.001 105
	-0.93 kcal/mol		-0.69 kcal/mol
Grid 5 (21 870 points; 58.007 903 electrons over grid)			
heptane	dimethylpentane	heptane	dimethylpentane
-270.452 815 141	-270.457 809 085	-270.452 815 141	-270.457 809 085
.463 944 653	.462 001 754	.464 121 568	.460 286 425
.464 848 088	.462 473 639	.463 016 832	.459 779 325
.465 234 836	.462 668 984	.463 834 470	.461 473 524
.465 229 779	.462 832 980	.463 811 737	.461 075 612
.465 250 319	.462 897 502	.463 716 135	.461 136 805
.465 250 671	.462 940 780	.463 752 741	.461 197 754
	.462 970 450		.461 214 758
	.463 031 795		.461 048 892
	.463 108 480		.461 059 429
	.463 195 097		.460 952 980
	.463 296 408		.460 969 514
	.463 315 337		.460 989 118
	.463 325 714		.460 975 563
	.463 336 309		.460 922 134
	.463 380 057		.460 875 826
	.463 367 364		.460 852 633
	.463 384 366		.460 832 997
	.463 402 409		.460 838 941
	.463 413 688		.460 838 988
	.463 427 231		
	.463 431 304		
	.463 431 814		
	.463 431 946		
maximum loss of translational invariance at convergence: 5.0×10^{-3}			
Energy difference:	+0.001 819		+0.002 923
	1.14 kcal/mol		1.83 kcal/mol

DFT code employs *both* formalisms, handling the gradients appropriately in each case.

IV. CONCLUSIONS

The results we have presented in this work—and many more which we have not reported here—clearly demonstrate that chemically accurate energies and geometries can be obtained from DFT calculations using carefully selected grids of only moderate size (around 3500 points per atom) for medium-sized systems (say up to 30 atoms). Furthermore, essentially identical results are obtained with these grids *whether or not quadrature weight derivatives are included in the DFT energy gradient*. Weight derivatives only begin to influence the results for grids which are of insufficient quality, e.g., grids 4 and 5 of this work, to guarantee reliable values in any case.

Although treating the gradient properly clearly improves the convergence pattern and essentially guarantees that the gradient zero coincides with the energy minimum, this is in fact almost purely cosmetic if used on grids with an insufficient number of quadrature points. *You cannot compensate for a poor quality grid by including weight derivatives.*

Two of the grids used here (grids 1 and 2) would appear to be sufficiently reliable for routine application and we have adopted grid 2 as our “standard grid.” This has around 3500 points per atom. In line with our previous remarks we do not consider it necessary to include the weight derivatives in calculations using this grid; however since we have coded exactly translationally invariant gradients with full inclusion of derivative weights, then we of course use them whenever applicable. The ability to calculate accurate gradients is of course very important if stationary points need to be located with high precision. For instance, when computing harmonic vibrational frequencies, tight convergence is required if one wishes to obtain “zero” frequencies for the translational and rotational modes. Furthermore, very good gradients are needed if anharmonic or vibration-rotation constants are being calculated through energy third and fourth derivatives.

We end on a note of caution. Clearly any grid error is going to increase as one treats larger and larger systems. For the same level of accuracy it is necessary to use a better quality grid with more grid points as the system size increases. Calculations with grids of insufficient quality—for example, grid 5 which has around 1000 points per atom—are likely to give even more unreliable results on larger systems than those reported in this work. If you are contemplating calculations on systems which you suspect are beyond the size for which your grid was originally calibrated, it is probably a good idea to deliberately *exclude* the weight deriva-

tives and use “old” gradients; any systematic increase in energy as the gradient decreases is a clear sign that you do not have enough grid points.

ACKNOWLEDGMENT

This work was supported in part by a grant to Biosym Technologies from the National Institute of Standards and Technology/Advanced Technology Program.

- ¹P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964); W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).
- ²T. Ziegler, Chem. Rev. **91**, 651 (1991), and references therein.
- ³*Density Functional Methods in Chemistry*, edited by J. Labanowski and J. Andzelm (Springer, New York, 1991).
- ⁴R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford University, New York, 1989).
- ⁵B. Delley, J. Chem. Phys. **92**, 508 (1990).
- ⁶A. D. Becke, J. Chem. Phys. **96**, 2155 (1992); **97**, 9173 (1992); **98**, 5648 (1993).
- ⁷J. Andzelm and E. Wimmer, J. Chem. Phys. **96**, 1280 (1992).
- ⁸A. St-Amant and D. R. Salahub, Chem. Phys. Lett. **169**, 387 (1990).
- ⁹B. G. Johnson, P. M. W. Gill, and J. A. Pople, J. Chem. Phys. **98**, 5612 (1993).
- ¹⁰N. C. Handy, D. J. Tozer, G. J. Laming, C. W. Murray, and R. D. Amos, Isr. J. Chem. **33**, 331 (1993).
- ¹¹M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, GAUSSIAN 92 (Gaussian Inc., Pittsburgh, PA, 1991).
- ¹²R. D. Amos, I. L. Alberts, J. S. Andrews, S. M. Colwell, N. C. Handy, D. Jayatilaka, P. J. Knowles, R. Kobayashi, N. Koga, K. E. Laidig, P. E. Malsen, C. W. Murray, J. E. Rice, J. Sanz, E. D. Simandiras, A. J. Stone, and M-D Su, CADPACS (Cambridge, 1992).
- ¹³DMOL, v.2.3.5, Biosym Technologies, San Diego, CA, 1993.
- ¹⁴DGAUSS, Cray Research Inc., Eagan, MN.
- ¹⁵DEMON, University of Montreal, Quebec, Canada.
- ¹⁶A. D. Becke, J. Chem. Phys. **88**, 2547 (1988).
- ¹⁷V. I. Lebedev, Zh. Vychisl. Mat. Fiz. **15**, 48 (1975); **16**, 293 (1976).
- ¹⁸DMOL User Guide, v.2.3.5 (1993) Section 2, p. 6, Biosym Technologies, San Diego, CA.
- ¹⁹C. W. Murray, N. C. Handy, and G. J. Laming, Mol. Phys. **78**, 997 (1993).
- ²⁰B. Delley, J. Chem. Phys. **94**, 7245 (1991).
- ²¹B. G. Johnson, P. M. W. Gill, and J. A. Pople, Chem. Phys. Lett. **220**, 377 (1994).
- ²²P. M. W. Gill, B. G. Johnson, and J. A. Pople, Chem. Phys. Lett. **209**, 506 (1993).
- ²³B. Delley, in *Modern Density Functional Theory: A Tool for Chemistry*, edited by P. Politzer and J. M. Seminario (Elsevier, New York, 1994).
- ²⁴S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
- ²⁵A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
- ²⁶C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
- ²⁷J. Baker, J. Comp. Chem. **7**, 385 (1986).
- ²⁸J. Baker and D. Bergeron, J. Comp. Chem. **14**, 1339 (1993).
- ²⁹R. Ahlrichs, M. Bär, M. Ehrig, M. Häser, H. Horn, and C. Kölmel, TURBOMOLE, v.2.3, Biosym Technologies, San Diego, California, 1993.
- ³⁰K. Kobayashi, N. Kurita, H. Kumahara, and K. Tago, Phys. Rev. A **43**, 5810 (1991).