ELSEVIER

Computational Materials Science 27 (2003) 437–445

COMPUTATIONAL
MATERIALS
SCIENCE

www.elsevier.com/locate/commatsci

Efficient linear scaling geometry optimization and transition-state search for direct wavefunction optimization schemes in density functional theory using a plane-wave basis

Salomon R. Billeter *, Alessandro Curioni, Wanda Andreoni

IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

Received 2 October 2002; accepted 24 November 2002

Abstract

Two linear scaling schemes for the search of stationary points on the nuclear potential energy surface have been developed and implemented for density functional theory programs using plane waves: a geometry optimizer based on the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method and a linear scaling method for transition-state search based on the microiterative scheme using the partitioned rational function optimizer (P-RFO) and L-BFGS. These optimizers are written with parallelized execution in mind. It is shown that the electronic wavefunction does not need to be fully optimized in the earlier stages of geometry optimization. The reasons for the robustness and good performance of the proposed schemes are identified. Test calculations are presented that use our implementation in the CPMD code.

PACS: 68.65.−k

Keywords: Geometry optimization; Transition state; Stationary points; Linear scaling methods; Density functional theory; Car–Parrinello

1. Introduction

Direct wavefunction optimization subject to orthogonality constraints [1,2] in density functional theory (DFT) is a very efficient alternative to the diagonalization of the Kohn–Sham (KS) Hamiltonian for total energy calculations or optimizations of nuclear geometries. It can be used particularly easily (and is routinely used) with programs employing the Car–Parrinello (CP) scheme [3] for ab initio molecular dynamics (MD) simulations using DFT owing to the availability of not only the forces acting on the nuclei, $-\partial E^{\mathrm{tot}}/\partial X_\alpha$, but also those acting on the KS orbitals, $-\delta E^{\mathrm{tot}}/\delta\psi_j^*$. Setting the kinetic energy of the CP

* Corresponding author. Tel.: +411-728-8643; fax: +411-724-8952.

E-mail address: srb@zurich.ibm.com (S.R. Billeter).

0927-0256/03/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0927-0256(03)00043-0

Lagrangean to zero corresponds to this class of schemes for direct wavefunction optimization subject to orthogonality constraints. These schemes are especially efficient and robust if the direct inversion in iterative subspace (DIIS) extrapolation method [4,5] is used.

Such a wavefunction optimization scheme for DFT is not only particularly efficient, but also the CPU and memory requirements scale much better with the system size than they do in other approaches because the costly diagonalization of the KS matrix is replaced by an optimization of the wavefunction with respect to the KS total energy. The resulting ground-state electron-density distribution as a function of the nuclear coordinates defines a potential energy surface (PES), which in turn can be used for geometry optimization.

The family of iterative second-order optimizers based on the Newton–Raphson (NR) step (see e.g. [6]), has been very successful in finding stationary points on such potential energy surfaces (PES) in few iterations. The required inverse of the Hessian matrix is either directly calculated or approximated using the gradients of the preceding iterations.

The program CPMD [7–9] is a very well optimized and parallelized DFT code using a plane-wave basis, pseudopotentials for core electrons, and also offering direct wavefunction optimization and the CP scheme for ab initio MD. For large but tractable systems, DFT codes using plane waves typically spend most of the CPU time required for the calculation of the electronic interactions in fast Fourier Transform (FFT) routines, scaling as $M \cdot N \log N$ with the number of plane waves N and the number of KS states M, and in the calculation of the non-local part of pseudopotentials, scaling with a very low prefactor as M^2N with system size. Moreover, the model for the distribution of data chosen in CPMD keeps the inter-process communication overhead low in the case of parallelized execution. However, the geometry optimization itself increasingly becomes a bottleneck for the reasons explained below.

Three flaws still need to be addressed so that the algorithms most commonly used for geometry optimization in DFT programs using plane waves become practical for systems containing more than a few hundred atoms. First, a search for transition states (TS) is only possible if the Hessian matrix of the starting geometry has exactly one negative eigenvalue, and the costs for the diagonalization of the full Hessian increase with the system size cubed. Second, especially the NR algorithms converging in the fewest cycles often need to be restarted manually after a not a priori predictable number of cycles to achieve good performance or even to converge. Third, the CPU costs for the geometry optimization itself become considerable for systems containing more than a few hundred atoms. As most of the CPU time is usually spent calculating the interactions between electrons, the CPMD program does the manipulations of the ionic coordinates on one processor only. With growing speed of processors and communication between processors and therefore larger tractable systems, the routines for geometry optimization have become a bottleneck. In principle, this bottleneck could be removed by parallelizing the geometry optimization routines, but the two other problems would remain unaddressed, and such a parallelized geometry optimization scheme would not be portable between different DFT programs.

Recently, an implicit NR optimization method [10] has been developed for DFT programs using plane waves. It employs density functional perturbation theory to approximate the effect of the Hessian on a given atomic displacement, and inverts the Hessian using a conjugate gradient method. This method has been successfully applied to a number of molecules, has been shown to work more reliably than the previous optimizers, and can be extended to TS search. However, for the optimization of the linear response wavefunction, considerable additional effort is required.

In this paper, we chose an alternative approach and implemented the algorithms for linear scaling geometry optimization and TS search described in [11,12], optimized for best performance and reliability with a wavefunction determined using direct wavefunction optimization. As the information about the Hessian is derived from the history of the optimization rather than from a linear response wavefunction, these optimization schemes also allow geometry optimization and TS search with a mixed quantum mechanics/molecular mechanics (QM/MM) Hamiltonian.

2. Methods

The algorithms for linear scaling geometry optimization and TS search presented in [11] have been optimized for the specific requirements of DFT calculations using plane waves and implemented in the framework of the CPMD code [9]. The theory, especially for Sections 2.1 and 2.2, is given in that article and the references therein. Here, we merely summarize the methods briefly and provide the specific details for the plane-wave DFT implementation. The geometry optimizer presented in [11] can perform the optimizations in both Cartesian and hybrid delocalized internal coordinates. Because the benefit of using delocalized internal coordinates is smaller for many systems typically treated with a plane-wave code, such coordinates are not yet implemented here. Hence, Cartesian coordinates are used throughout this paper.

2.1. Geometry optimization

For linear scaling geometry optimization, the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) scheme [13,14] has been implemented. Instead of calculating, updating, and inverting the Hessian matrix, the effect of the inverse Hessian applied to a gradient is extrapolated from a limited number n_r of previous geometry steps and gradients. If the initial Hessian is diagonal and the number of steps is smaller than n_r, the L-BFGS step is equivalent to the NR step with a Hessian matrix approximated by the BFGS formula [13]. Note that only values smaller than the number of nuclear degrees of freedom are meaningful for n_r, and that with a fixed n_r, the algorithm scales linearly with system size, both in terms of CPU and memory requirements.

The trust radius algorithm in [11,15] based on the Wolfe conditions for sufficient decrease and curvature (see e.g. Ref. [6]) has been implemented with a few changes for best performance using DFT direct wavefunction optimization schemes. The maximum trust radius $T_{\max}$ and initial value T_{ini} are set to 0.5 atomic units. If a step with a trust radius below $T_{\min}$ [1] still does not decrease the energy, the geometry optimization (both trust radius and L-BFGS history) is reset. This could occur if wavefunctions of the previous steps have not been accurate enough, if the harmonic approximation was not appropriate for the part of the PES searched, or if the geometry is very close to a stationary point. If the subsequent step will be even smaller ($< T_{\text{end}}$ [1]), a stationary point must be very close because the search direction corresponds to the gradient, and the optimization is stopped.

A critical feature of this trust radius algorithm is that geometry steps can be rejected. If a geometry step is rejected, the initial guess of the electronic wavefunction is taken from the result of the last successful step rather than from the most recent one. [2]

For comparison, a line-search scheme has been implemented, working together with any external line-search algorithm with a sufficient decrease criterion. For consistency reasons, the initial guess of the wavefunction must be the same for all line-search cycles for the same search direction. The search in one direction is stopped if either more than n_{ls} [1] cycles would be required to satisfy the sufficient decrease criterion, or if the predicted step or its smallest deviation from any other attempted step is smaller than $T_{\min}$. [1] In this case, the step length that so far has yielded the lowest energy is taken.

2.2. Transition-state search

The partitioned rational function optimizer [16] (P-RFO) is widely used for TS searches. It maximizes the energy in the direction of one eigenmode of the Hessian while minimizing it in all other directions. Unlike the NR or rational function (RFO) optimizers, there are no restrictions in the choice of the mode along which the energy is

[1] Reasonable values are: $T_{\max} = T_{\text{ini}} = 0.5$ Bohr, $T_{\min} = 10^{-5}$ Bohr, $T_{\text{end}} = 10^{-7}$ Bohr, $n_{\text{ls}} = 4$.

[2] The optimizer only schedules the old wavefunction to be restored because geometry optimization is done on one processor but the wavefunction can be distributed to overcome memory limitations and for computational efficiency.

maximized, specifically, any eigenvalue may be negative or positive, and the optimization will quite likely terminate in a stationary point with exactly one negative eigenvalue which belongs to the mode to be maximized. This makes the P-RFO algorithm particularly suitable for mode following. There is, however, a severe drawback it shares with all algorithms acting in the eigenspace of the Hessian: the effort for the diagonalization of the Hessian scales with the system size cubed, and the calculation of the full Hessian is very costly.

The linear-scaling microiterative TS search algorithm has been described in [11,12]. The system is split into a small reaction core and a large environment, which is relaxed to a very small gradient,

$$\max_{\alpha \in \mathrm{env}} \left| \frac{\partial E^{\mathrm{tot}}}{\partial X_\alpha} \right| \leqslant F_{\mathrm{max,env}}, \tag{1}$$

before each step within the reaction core towards the TS. This variational separation minimizes the interdependence between core and environment.

Before the first P-RFO step in the reaction core (after the first relaxation of the environment with fixed core), the finite-difference Hessian matrix of the core degrees of freedom is calculated,

$$H_{\alpha\beta} = \frac{1}{4\varepsilon_h} \left[\frac{\partial E^{\mathrm{tot}}}{\partial X_\alpha} \bigg|_{X+\varepsilon_h \mathbf{e}_\beta} - \frac{\partial E^{\mathrm{tot}}}{\partial X_\alpha} \bigg|_{X-\varepsilon_h \mathbf{e}_\beta} + \frac{\partial E^{\mathrm{tot}}}{\partial X_\beta} \bigg|_{X+\varepsilon_h \mathbf{e}_\alpha} - \frac{\partial E^{\mathrm{tot}}}{\partial X_\beta} \bigg|_{X-\varepsilon_h \mathbf{e}_\alpha} \right], \tag{2}$$

for which a finite-difference displacement of ε_h of 0.005 atomic units along the unit vectors $\mathbf{e}_\alpha$ turned out to be a useful compromise between accuracy and required number of wavefunction optimization cycles. Such a small displacement is possible because the initial guess of the wavefunction can be taken from the wavefunction optimized at geometry X rather than from the one at the corresponding preceding geometry. We will call this matrix the partial Hessian. If necessary, it can be periodically recalculated. Note that neither translational nor rotational degrees of freedom are projected out of this Hessian, because, due to the environment, the core system is no longer invariant under these transformations.

If the current partial Hessian has been calculated for a previous P-RFO step, it is updated using the Powell formula [17] from the last partial Hessian, the last P-RFO step, and from the difference between the current gradient (after the last step with relaxed environment) and the gradient before the last P-RFO step. This update with relaxed environment is consistent with the energy and gradient of the core, the environment being subject to constraint (1), but it is not fully consistent with the initial partial Hessian (2) which is calculated for the relaxed geometry **X** but using unrelaxed trial moves $\varepsilon_h \mathbf{e}_\alpha$. However, this approach works well in practice, and saves many cycles during the costly calculation of the initial partial Hessian.

Unlike for geometry optimization, the partial Hessian is not required to be positive definite, and the trust radius algorithm in Section 2.1 cannot be applied. Therefore, the algorithms for both step acceptance and dynamic trust radius are based on energy prediction and Hessian mode overlap [11,12] for the P-RFO step.

Assuming moderate couplings between the reaction core and its environment, the partial Hessian obtained from microiterative TS search can be used for a vibrational analysis of the subsystem of interest. Therefore, a block-diagonal Hessian of the full system, consisting of the partial Hessian and either a unit matrix or a matrix constructed using empirical parametrizations such as DISCO [18] or Schlegel [19], is built and diagonalized.

2.3. Adaptive tolerance

Many DFT programs apply the same criteria for the convergence (self-consistency) of the electronic wavefunction throughout a geometry optimization. We address two problems associated with this.

First, these criteria cause unnecessary computational expenses. In order to obtain forces on the nuclei that are accurate enough for the final stages of the optimization, relatively tight criteria must be applied for the wavefunction. This leads to unnecessarily accurate gradients at the beginning of the optimization and thus to excessive computational effort. For the direct wavefunction optimization schemes, such a wavefunction convergence criterion for the geometry optimization step s is

$$g_{\max,s} = \max_{i,\mathbf{G}} |g_{i,s}(\mathbf{G})| \leqslant g_{\max}, \tag{3}$$

$$g_{i,s}(\mathbf{G}) = \langle \mathbf{G}|\hat{H}_s^{\mathrm{KS}}|i,s\rangle - \sum_j^{\mathrm{occ}} \langle \mathbf{G}|j,s\rangle\langle j,s|\hat{H}_s^{\mathrm{KS}}|i,s\rangle, \tag{4}$$

where $\hat{H}_s^{\mathrm{KS}}$ is the KS operator of geometry step s, $\langle \mathbf{G}|i,s\rangle$ is the expansion coefficient of KS state i of step s to plane wave $\mathbf{G}$, and $g_{\max}$ is the criterion for the wavefunction gradient. In Eq. (4), $\hat{H}_s^{\mathrm{KS}}|i\rangle$ is a convenient expression for the gradient of the total KS energy with respect to state $|i\rangle$. Eq. (3) is now replaced by the gradient adaptive tolerance criterion,

$$g_{\max,s} \leqslant \max(g_{\max}, c_{\mathrm{gr,ad}} \cdot \min_{s'<s} F_{\max,s'}), \tag{5}$$

$$F_{\max,s'} = \max_{\alpha} \left| \frac{\partial E_{s'}^{\mathrm{tot}}}{\partial X_\alpha} \right|, \tag{6}$$

where E^{tot} is the KS energy, X_α are the nuclear coordinates, and $c_{\mathrm{gr,ad}}$ is the adaptive gradient tolerance criterion. For good performance, taking the minimum over all previous steps s' in Eq. (5) turned out to be very important to ensure consistency because the maximum component of the gradient does not decrease monotonically. Criterion (5) is most important at the beginning of an optimization.

Second, proper assessment of the quality of an optimization step by the progress of the target function of the optimization, in our case the KS energy E^{tot}, helps to improve both the stability and the performance of the optimizer. Therefore, the absolute difference between the KS energies of two wavefunction optimization cycles w and $w-1$ at geometry $X_{\alpha,s}$ must be much smaller than the change between the two geometry optimization steps $s-1$ and $s-2$,

$$|E_{s,w}^{\mathrm{tot}} - E_{s,w-1}^{\mathrm{tot}}| \leqslant \max\left(\Delta E_{\min}, c_{\mathrm{en,ad}} \cdot |E_{s-1}^{\mathrm{tot}} - E_{s-2}^{\mathrm{tot}}|\right) \quad \text{if } s > 1. \tag{7}$$

The constant $c_{\mathrm{en,ad}}$ is the energy adaptive tolerance criterion, and the threshold $\Delta E_{\min}$ helps bridging geometry steps with very small effect.

For the linear scaling optimizers, which can reject steps, only the geometries resulting from accepted steps are considered for Eqs. (5) and (7). Note that, if applied, the adaptive energy criterion Eq. (7) needs to be satisfied together with either Eq. (3) or (5). Reasonable values for parameters $c_{\mathrm{gr,ad}}$, $c_{\mathrm{en,ad}}$, and $\Delta E_{\min}$ turned out to be 0.02, 0.05, and 10^{-7} hartrees, respectively. The usual tolerance of 10^{-5} atomic units is suggested for $g_{\max}$.

The microiterative optimization scheme requires special care with adaptive tolerance criteria of the electronic wavefunction. For relaxation of the environment, both the gradient and the energy adaptive tolerance criteria can be applied, but only the force components acting on the environment degrees of freedom are considered for Eq. (6). During the calculation of the partial Hessian and the P-RFO steps, the gradient adaptive tolerance criterion Eq. (5) is not applicable and replaced by the standard criterion Eq. (3) for the wavefunction optimization, but the adaptive energy tolerance criterion Eq. (7), using the energy difference of the last accepted optimization step within the environment, was able to reduce the number of required P-RFO steps considerably. Note that this energy difference is usually very small and ensures a very accurate wavefunction for the critical steps in the reaction core.

2.4. Test calculations

All calculations were performed using a modified version of CPMD [9]. Unless mentioned, the convergence criteria for the gradient of the ions and the wavefunction were 10^{-4} and 10^{-5} atomic units, respectively, and the density cutoff for the calculation of the gradient corrections was 10^{-5} atomic units. Test systems were the Diels–Alder reaction between ethylene and butadiene, and three silicon clusters.

For the Diels–Alder test system, the Becke88 [20]/Perdew86 [21] exchange-correlation functional (BP) has been used. A cubic cell having an edge of 28.3 Bohr has been used with a 40-Rydberg cutoff for the plane-wave basis. The reaction core for the TS search consisted of the four carbon atoms between which σ bonds are formed.

The three clusters contained 307, 513, and 747 silicon atoms, respectively, and the valences at the edges were saturated with hydrogen. The clusters were geometry optimized using the Perdew–Burke–Enzerhof [22,23] exchange-correlation functional (PBE) and cells of 59-, 68-, and 74-Bohr edge, respectively, with a cutoff of 10 Rydberg for the plane-wave basis.

3. Results

3.1. Geometry optimization

Fig. 1 and Table 1 summarize the effects of several optimization parameters on the geometry optimization performance in a relatively small test system, the Diels–Alder complex of ethylene and butadiene.

Fig. 1 compares the linear scaling geometry optimizer with some algorithms implemented in CPMD [9].

All currently used geometry optimizers were outperformed by the linear scaling optimizer in all cases. Both second-order optimizers with BFGS update of the Hessian performed well initially, but they took a wrong route after an overshoot due to a missing criterion for step rejection and to a fixed rather than a dynamically adjusted trust radius. Whereas DIIS extrapolation of the nuclear displacements was unable to recover from the wrong route, the NR optimizer recovered after some time. Conjugate gradient optimization was successful, too. Its implementation in CPMD uses a line-search algorithm. Consequently, no wrong route was taken, and the spikes typical for line search are observable in Fig. 1, too. As expected, the slope of the conjugate gradient optimizer is smaller than those of the second-order methods.

Fig. 1 also compares the L-BFGS optimization algorithms with trust radius and line search to each other. Initially, line search is able to reduce the number of geometry steps required to reach a given value for the total energy, as to be expected. However, the number of energy and gradient evaluations is not much reduced because the larger geometry steps require more cycles for the wavefunction to converge. In the later stages of the optimization, the benefits from the optimum step lengths using line search are more than compensated by the additional effort to find them. The small spikes stem from rejected attempts to take too large a step when searching.

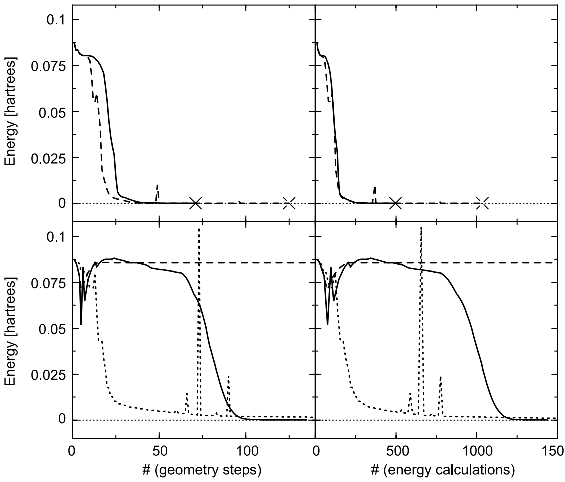

Fig. 1. Geometry optimization of the Diels–Alder complex of ethylene and butadiene to the products using the BP exchange-correlation functional [20,21]. The total energy is shown as function of the number of geometry steps performed (left column) and as function of the number of energy/gradient evaluations (right column). The optimization started at the AM1 TS of the reaction. The initial maximum component and norm of the energy gradient on ions were 3.9×10^{-2} and 1.3×10^{-2} atomic units, respectively. Top row: optimization using the linear scaling optimizer, with trust radius (solid line), and with line search (dashed line). Bottom row: optimization using the NR/BFGS (Newton–Raphson with approximate Hessian (Broyden–Fletcher–Goldfarb–Shanno update), solid line), the DIIS/BFGS (Direct inversion in iterative subspace with approximate Hessian, dashed line), and the conjugate gradient (short-dashed lines) optimizers. The energies are given relative to the lowest value obtained (−40.016140 hartrees). The dotted line indicates the lowest energy reached (adjusted to zero). The crosses in the top-row panels indicate where the optimization converged. The convergence criterion for the maximum component of the gradient on ions was 10^{-4} atomic units. The regular convergence criterion for the wavefunction was 10^{-5} atomic units in all cases. For the linear scaling optimizer, the adaptive tolerance criteria (see Section 2.3) of 0.02 for the gradient and 0.05 for the energy change have been applied. See also Table 1.

The trends in Table 1 are consistent with Fig. 1. Geometry optimization to the products represents an "easy" problem, whereas optimization

Table 1

Geometry optimization[a] of the Diels–Alder complex of ethylene and butadiene to the products[b] (PS) and the reactants[c] (RS) using the Becke88 [20]/Perdew86 [21] (BP) exchange-correlation functional

Geometry	Optimizer	Criteria		Energy	Steps		Steps (ene)	
		$c_{\mathrm{gr,ad}}$	$c_{\mathrm{en,ad}}$		wf	geo	wf	geo
PS	L-BFGS/TR	0.02	0.05	−40.016128	**498**	71	327	50
PS	L-BFGS/LS	0.02	0.05	−40.016125	1035	125	296	40
PS	L-BFGS/TR	0.01	–	**−40.016140**	770	96	**283**	32
PS	L-BFGS/TR	–	–	−40.016077	507	**62**	375	**31**
PS	NR/BFGS	–	–	−40.015998	1435	135	n/r	
PS	DIIS/BFGS	–	–	(−39.930444)	>2500	>472	n/r	
PS	CGRAD	–	–	(−40.015720)	>2500	>403	n/r	
RS	L-BFGS/TR	0.02	0.05	**−39.953912**	1477	152	**1084**	**109**
RS	L-BFGS/LS	0.02	0.05	−39.953562	1423	**144**	n/r	
RS	L-BFGS/TR	0.01	–	−39.953848	1870	232	1591	155
RS	L-BFGS/TR	–	–	−39.953052	**1266**	206	n/r	
RS	NR/BFGS	–	–	(−39.952271)	>2500	>504	n/r	
RS	DIIS/BFGS	–	–	(−39.930504)	>2500	>261	n/r	
RS	CGRAD	–	–	(−39.947667)	>2500	>383	n/r	

The final energy in atomic units is given for some geometry optimizers and wavefunction convergence criteria. For the optimization algorithms, see Section 2.1. "L-BFGS/TR" and "L-BFGS/LS" denote the L-BFGS optimizer with trust radius and line search, respectively. The regular convergence criterion for the wavefunction was 10^{-5} atomic units in all cases. Where given, the gradient ($c_{\mathrm{gr,ad}}$) and energy ($c_{\mathrm{en,ad}}$) adaptive tolerance criteria have been applied, see Section 2.3. Also given are the number of single-point energy calculations (columns "wf") and geometry steps (columns "geo") required to reach convergence[a] (columns "Steps"), and the respective numbers of steps required to reach a given value of the total energy. The energies to be reached were −40.016 atomic units for the "PS" optimizations and −39.9538 atomic units for the "RS" optimizations. [columns "Steps (ene)"]. See also Fig. 1 and Table 3.

[a] The convergence criterion for the maximum component of the gradient on ions was 10^{-4} atomic units throughout.

[b] The optimization started at the AM1 TS of the reaction, the initial maximum component and norm of the energy gradient on ions were 3.9×10^{-2} and 1.3×10^{-2} atomic units, respectively.

[c] The optimization started very near the BP TS. The initial maximum component and norm of the energy gradient on ions were 3.0×10^{-3} and 1.1×10^{-3} atomic units, respectively. Note that the small initial gradient is due to the closeness to the TS rather than to the reactant state.

to the reactants is "difficult" because the starting geometry for the latter was near a saddle point in the potential energy landscape, and the potential energy surface is flat near the reactants. Table 1 demonstrates that different algorithms and options lead to different optimized energies even for such a small system, and several criteria need to be applied to determine which one works best. The linear scaling optimizer with adaptive tolerance for both gradient and energy change is the method of choice considering robustness, fast convergence to a small gradient and to a low energy simultaneously, both in terms of the number of energy/gradient evaluations required and in terms of the number of geometry steps. For the reasons mentioned above, line search did not save optimization cycles.

In order to test the performance of the different geometry optimizers on larger systems, the geometry of three silicon clusters consisting of up to 1051 atoms has been optimized (see Table 2). In addition to the trends already observed in Fig. 1 and Table 1, also the percentage of CPU time spent in the routines for geometry optimization is negligible for the linear scaling optimizers whereas it is significant for larger systems for the currently implemented second-order optimizers (they were the most time-consuming subroutines in these calculations), despite the small number of DIIS steps used for extrapolation (5 for the two smaller clusters, 2 for the largest cluster). This shows that the upper limit of sizes handled by these optimizers is reached at around 1000 atoms, and the linear scaling optimizer described in this article is needed for these and larger systems.

Table 2
Geometry optimization of large systems with the L-BFGS and the DIIS/BFGS optimizer

Size	Optimizer	Processes	Energy	Steps		CPU time
				wf	geo	
307 Si + 172 H	L-BFGS/TR	64	**−1297.8437**	310	47	**<0.3**%
307 Si + 172 H	DIIS/BFGS	52	−1297.8415	394	96	15.5%
513 Si + 252 H	L-BFGS/TR	128	**−2149.4113**	100[a]	12[a]	**<0.3**%
513 Si + 252 H	DIIS/BFGS	128	−2149.3993	87	13	21.8%
747 Si + 304 H	L-BFGS/TR	128	**−3095.5003**	1002	98	**<0.1**%
747 Si + 304 H	DIIS/BFGS	128	−3095.3474	>1002[b]	>92[b]	19.6%

Three silicon clusters have been optimized using the PBE gradient correction [22,23]. For the column labels, see Table 1. Also given is the percentage of the total CPU time spent in geometry optimization routines for a given number of parallel processes on an IBM SP3.

[a] The L-BFGS optimizer required 61 energy evaluations and 7 geometry steps to obtain an energy below −2149.4 atomic units.

[b] The DIIS/BFGS optimization was stopped after the number of single-point energy and gradient calculations the L-BFGS optimization took to reach convergence. The smallest maximum component and norm of the energy gradient on ions reached up to that point were 2.8×10^{-3} and 8.4×10^{-4} atomic units, respectively.

3.2. Transition-state search

The practical feasibility of the P-RFO algorithm [16] and the microiterative scheme for TS search [11,12] has been shown in the literature. Table 3 tests the suitability and performance of the algorithm in DFT. Although the reaction core of only four carbon atoms is very strongly coupled to its environment, the microiterative TS search from two different starting geometries converged to the same geometry with one negative eigenvalue of the Hessian. Note that the evaluation of the initial Hessian alone would have required 96 geometry steps if the core consisted of all atoms. For good variational separation between core and environment, a very small gradient on the ions is required (originally, one third of the convergence criteria of the reaction core [11,12]). For the Diels–Alder re-

Table 3
TS search of the Diels–Alder reaction between ethylene and butadiene using the BP exchange-correlation functional [20,21] from two different starting conditions ("Geometry")

Geometry	Criteria		Barrier	Steps		
	$F_{\mathrm{max,env}}$	$c_{\mathrm{en,ad}}$		wf	geo[a]	core[b]
TS1[c]	5×10^{-4}	–	0.023423	344	43	12
TS1[c]	10×10^{-4}	–	Wrong mode		–	–
TS1[c]	5×10^{-4}	0.05	0.023418	350	44	12
TS1[c]	10×10^{-4}	0.05	0.023480	491	33	16
TS2[d]	5×10^{-4}	–	0.023486	571	78	13
TS2[d]	10×10^{-4}	–	0.023486	562	77	13
TS2[d]	5×10^{-4}	0.05	0.023471	589	77	13
TS2[d]	10×10^{-4}	0.05	0.023453	546	67	11

See Tables 1 and 2 for the column labels. The convergence criterion "env" denotes the maximum component of the gradient on the ions of the environment until a step within the reaction core is done. The convergence criterion for the maximum component of the gradient on ions in the reaction core was 5×10^{-4} atomic units.

[a] Number of P-RFO, L-BFGS, and finite-difference Hessian steps.

[b] Number of P-RFO steps alone.

[c] This initial geometry has a maximum component of the gradient on the ions of the environment: 4.8×10^{-4} atomic units. The finite-difference Hessian has been calculated already.

[d] This initial geometry has a maximum component of the gradient on the ions of the environment: 1.6×10^{-3} atomic units. The finite-difference Hessian is not yet available.

action in Table 3, the relatively loose criterion of 5×10^{-4} atomic units was sufficient.

4. Conclusions and outlook

This article presented methods for geometry optimization and TS search suitable for molecular and nanoscale systems treated with DFT codes using plane waves and direct wavefunction optimization. These procedures can run without user intervention for the entire optimization, can be resumed without loss of information, and the costs of the geometry optimization itself are negligible compared with those of the evaluation of energy and forces for all system sizes considered. Among the reasons for the good performance and robustness, an adequate trust radius algorithm and the ability to reject geometry steps have been identified to be the most important. Additionally, it was possible to improve the stability and performance of the geometry optimizers by dynamically adjusting the convergence criteria for the electronic wavefunction based on the progress of the geometry optimization and by restoring the guess of the electronic wavefunction to be optimized from a step earlier than the last step under given circumstances.

The test systems contained 4 to 1051 atoms. The geometry optimizers presented in this paper will be particularly useful for very large systems such as macromolecular and nanoscale systems, including systems described by a QM/MM Hamiltonian for which the computational costs of the conventional algorithms are prohibitive.

References

[1] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, J.D. Joannopoulos, Rev. Mod. Phys. 64 (1992) 1045.

[2] M.C. Payne, J.D. Joannopoulos, D.C. Allan, M.P. Teter, D.H. Vanderbilt, Phys. Rev. Lett. 56 (1986) 2656.

[3] R. Car, M. Parrinello, Phys. Rev. Lett. 55 (1985) 2471.

[4] J. Hutter, H.P. Lüthi, M. Parrinello, Comput. Mater. Sci. 2 (1994) 244.

[5] P. Pulay, Chem. Phys. Lett. 73 (1980) 393.

[6] R. Fletcher, Practical Methods of Optimization, Wiley, Chichester, 1980.

[7] D. Marx, J. Hutter, in: J. Grotendorst (Ed.), Modern Methods and Algorithms of Quantum Chemistry, NIC, vol. 1, John von Neumann Institute for Computing, Jülich, Germany, 2000, pp. 301–449.

[8] W. Andreoni, A. Curioni, Parallel Computing 26 (2000) 819.

[9] CPMD (Copyright IBM Corporation 1990, 1997, 2001 and MPI für Festkörperforschung, Stuttgart, Germany, 1997). See http://www.cpmd.org.

[10] F. Filippone, S. Meloni, M. Parrinello, J. Chem. Phys. 115 (2001) 636.

[11] S.R. Billeter, A.J. Turner, W. Thiel, Phys. Chem. Chem. Phys. 2 (2000) 2177.

[12] A.J. Turner, V. Moliner, I.H. Williams, Phys. Chem. Chem. Phys. 1 (1999) 1323.

[13] D.C. Liu, J. Nocedal, Math. Prog. 45 (1989) 503.

[14] J. Nocedal, Math. Comput. 35 (1980) 773.

[15] S.R. Billeter, A.J. Turner, W. Thiel. http://www.rsc.org/suppdata/cp/a9/a909486e/suppl.txt.

[16] A. Banerjee, N. Adams, J. Simons, R. Shepard, J. Phys. Chem. 89 (1985) 52.

[17] M.J.D. Powell, Math. Program. 1 (1971) 26.

[18] T.H. Fischer, J. Almlöf, J. Phys. Chem. 96 (1992) 9768.

[19] H.B. Schlegel, Theo. Chim. Acta 66 (1984) 333.

[20] A.D. Becke, Phys. Rev. A 38 (1988) 3098.

[21] J.P. Perdew, Phys. Rev. B 33 (1986) 8822.

[22] J.P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. 77 (1996) 3865.

[23] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, C. Fiolhais, Phys. Rev. B 46 (1992) 6671.