Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Journal of Computational Physics 191 (2003) 130–146

JOURNAL OF COMPUTATIONAL PHYSICS

www.elsevier.com/locate/jcp

Improved algorithm for geometry optimisation using damped molecular dynamics

M.I.J. Probert *

Department of Physics, University of York, Heslington, York YO10 5DD, UK

Received 23 August 2002; received in revised form 27 May 2003; accepted 10 June 2003

Abstract

We present a new algorithm for optimising the configuration of a system of particles using damped molecular dynamics. This new algorithm has a much faster rate of convergence to the ground-state structure than other MD-based schemes. We also show how it is possible to use this algorithm to select dynamically the optimum time step or to precondition the masses for the modified dynamics, so as to produce additional significant efficiency gains especially as convergence is approached. The performance of the new algorithm is illustrated with several model systems and also by relaxing a defect in bulk silicon. The defect calculation was also relaxed using a standard BFGS algorithm and the performance was found to be comparable.

Keywords: 65B99 Numerical analysis: acceleration of convergence; 70-08 Mechanics of particles and systems: computational methods

1. Introduction

It is a common requirement to want to optimise the configuration of a system of particles so as to minimise the total energy of the system. The simplest (and least efficient) approach of all is that of steepest-descent dynamics. This can be viewed as a quasi-Newtonian dynamics using only first-order time derivatives and gradients of the energy functional. More efficient schemes, such as conjugate gradients and BFGS methods [1], use knowledge of higher-order derivatives of the energy functional (either explicitly or implicitly). However, it may not always be possible to use these schemes if the cost of calculating higher-order derivatives is too high.

Other considerations arise in the case of ab initio calculations. For example, in the Car–Parrinello approach [2], a fictitious dynamics of the electrons is coupled to the dynamics of the ions to generate the ground state electronic and ionic configuration simultaneously. In this approach it is therefore not possible to minimise the ionic configuration using a non-dynamic scheme. An alternative ab initio approach is to use

---

* Fax: +44-1904-432214.
E-mail address: mijp1@york.ac.uk (M.I.J. Probert).

0021-9991/$ - see front matter © 2003 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0021-9991(03)00308-5

direct minimisation of the energy functional, such as the conjugate-gradient method [3,4], whereupon it is possible to use BFGS-type methods to minimise the ionic configuration. However, even within the direct minimisation approaches to the electronic problem, it is often better if the ionic configuration evolves smoothly in time (which is generally not the case in BFGS-type methods) as then techniques such as "wavefunction-extrapolation" [5] can be used to accelerate the electronic convergence at each ionic step. Therefore even in these schemes there is still a place for an efficient dynamics-based scheme to optimise the geometry of the ions.

Damped molecular dynamics, using a second-order equation of motion, would seem to be the natural solution to this problem. This approach would also make it simple to impose geometric constraints and crystal symmetry as appropriate. It could also be extended to finite temperature, using a simulated annealing technique [6], to avoid any problems in the geometry converging to a local and not the global minimum.

In [7], an algorithm for damped second-order molecular dynamics for minimising the electronic energy in the context of Car–Parrinello simulations was presented, which was shown to be comparable in efficiency to conjugate-gradient methods. However, there is a difficulty in choosing an appropriate damping rate and time step. The suggestion made in [7] was that the time step should be determined by the highest frequency component of the fictitious electron dynamics, and that the damping rate should be estimated from the analysis of a number of steepest descent steps.

In this work we shall present a new approach to the problem which has marked efficiency gains over those previously proposed, particularly for the optimisation of the ionic configuration but which may also prove useful for the electronic minimisation problem as well. We also introduce two extensions that have additional efficiency gains and which can be applied to both the existing and this new scheme, namely the dynamic selection of the optimal time step or preconditioning of the ionic masses.

2. Method

In order to converge to the correct optimal geometry, we must start with an initial configuration of ions that is within the basin of attraction of the global minimum. If this is not the case, then this approach can be generalised by the addition of thermal effects, to produce a scheme similar to simulated annealing. If we are within the appropriate basin, then to a first approximation we can consider the dynamics to be that of damped simple harmonic oscillators. The effect of anharmonic terms will be considered later.

2.1. First- vs. second-order dynamics

The equation of motion for first-order dynamics is simply

$$\mathbf{F} = -\nabla U(\mathbf{r}) = m\dot{\mathbf{r}}, \tag{1}$$

which is the basis of steepest-descent dynamics – i.e., moving downhill in the direction of the direction of the energy gradient. More conventionally, the equation of motion for second-order dynamics is

$$\mathbf{F} = -\nabla U(\mathbf{r}) = m\ddot{\mathbf{r}}. \tag{2}$$

If we assume that the current configuration $\mathbf{r}_i$ is "close" to the minimum energy configuration $\mathbf{r}_0$, then

$$\mathbf{r}(t) \simeq \mathbf{r}_0 + \delta\mathbf{r}(t), \tag{3}$$

and we can similarly expand the potential energy function as a quadratic function

$$U(\delta r) \simeq \tfrac{1}{2}m\omega^2(\delta r)^2, \tag{4}$$

whereupon we can solve analytically the first-order equation of motion as

$$\delta \mathbf{r}(t) \sim \mathrm{e}^{-\omega^2 t}, \tag{5}$$

and the corresponding second-order equation of motion as

$$\delta \mathbf{r}(t) \sim \mathrm{e}^{-\mathrm{i}\omega t}. \tag{6}$$

We therefore see that the first-order dynamics corresponds to a decay towards the ground state with a characteristic decay rate $\sim 1/\omega^2$, whereas the second-order dynamics corresponds to small oscillations around the ground state configuration with a characteristic frequency ω. In a numerical implementation therefore, first-order dynamics requires a maximum integration time step $\sim 1/\omega^2$ and second-order dynamics requires a maximum integration time step $\sim 1/\omega$. If there are multiple modes present, then the maximum time step allowed is determined by the fastest mode (smallest period).

2.2. Coupled modes damping

In order to convert second-order dynamics into a minimisation scheme, we must introduce some damping. For simplicity, we will first consider the case of a single particle moving in a one-dimensional harmonic oscillator potential centred on the origin $U(x) = \frac{1}{2}kx^2$

$$F_x = m\ddot{x} = -kx - \alpha\dot{x}, \tag{7}$$

whereupon it is trivial to show that in order to minimise the energy in the least amount of time, the theoretical optimal choice for α corresponds to "critical damping", i.e.,

$$\alpha^{\mathrm{c}} = \sqrt{4mk}. \tag{8}$$

These equations can be cast into a simpler form, by defining $\gamma = \alpha/2m$ which we will call the damping parameter. Critical damping then corresponds to $\gamma^{\mathrm{c}} = \omega_0$, where ω_0 is the natural frequency of the oscillator, under-damping corresponds to $\gamma < \gamma^{\mathrm{c}}$, etc. This is illustrated in Fig. 1.

The most important aspect of using damped molecular dynamics as an algorithm for geometry optimisation is therefore the correct choice of γ. It was shown in [7] that a simple method for estimating γ^{c} was to perform first-order dynamics ("steepest descents") which can be achieved in the conventional second-order dynamics scheme by setting $\gamma^{\mathrm{SD}} = 1/\Delta t$, where Δt is the time step used. If we assume that we are sufficiently near the minimum for the first-order dynamics to give an exponential decay of the total energy E_n for three successive steps ($n = 1, 2, 3$), then it can be easily shown that

$$\gamma^{\mathrm{c}} = \frac{1}{\Delta t}\sqrt{\frac{1}{2}\ln\left(\frac{E_1 - E_2}{E_2 - E_3}\right)} \tag{9}$$

(where the $\sqrt{\{\}}$ arises from converting the estimator for γ^{c} from first- to second-order). We can use this estimate with the same time step for optimally damped second-order dynamics.

This is the approach that was used successfully in [7] for the electronic problem and which we hereafter refer to as "coupled modes damping". The key feature of this algorithm is that it damps all modes uniformly with a damping given by the critical damping for the dominant mode in the decay of the total energy of the system, which is typically the fastest mode present. Hence it is relatively cheap in terms of memory costs when applied to systems with a large number of degrees of freedom, such as in the electronic problem.

As the simulation progresses the fast modes will be damped first, whereupon the damping for the other modes will be too small, and the second-order dynamics will then be in the under-damped regime which will

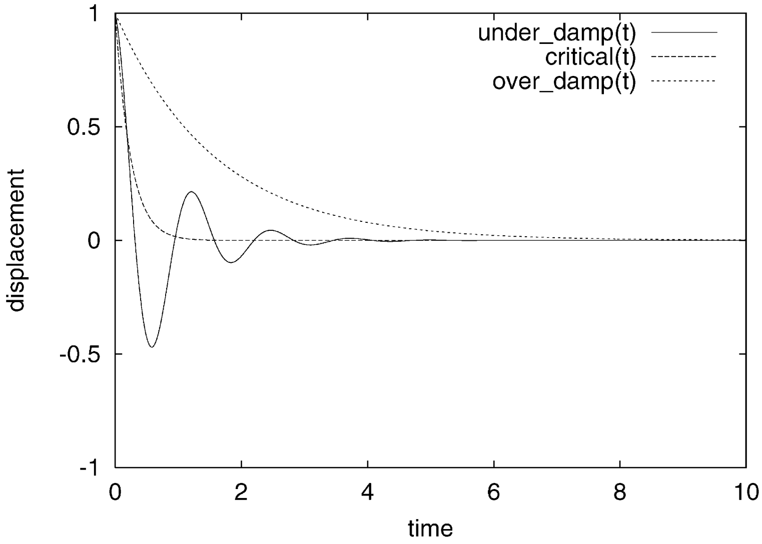

Fig. 1. Solution to the damped SHO in different damping regimes ($\omega = 5$, $\gamma_{\text{under}} = \omega/4$, $\gamma^{c} = \omega$, $\gamma_{\text{over}} = 4\omega$).

obviously be inefficient. This can be partly redressed by periodically switching over to first-order dynamics and resampling γ. However, there is a trade-off in efficiency between using the optimal value of γ in second-order dynamics, and the cost of three steps of first-order dynamics necessary to calculate it.

There are many different integration schemes that may be used to generate the dynamics. In common with many other MD practitioners we use the Velocity Verlet scheme [8] as it is both simple to implement and stable in practice (for instance, it can be shown to be symplectic [9]). The introduction of velocity-dependent forces does however require a minor amendment to the standard Velocity Verlet algorithm, as in the standard algorithm the updated velocities depend on the forces at the updated coordinates. The original Velocity Verlet algorithm can be written as:

$$\begin{aligned} \mathbf{r}_+ &= \mathbf{r}_0 + \mathbf{v}_0\Delta t + \frac{\mathbf{f}_0\Delta t^2}{2m}, \\ \mathbf{v}_+ &= \mathbf{v}_0 + \left(\frac{\mathbf{f}_0 + \mathbf{f}_+}{2}\right)\frac{\Delta t}{m}, \end{aligned} \tag{10}$$

where $\mathbf{r}_0$ are the positions at time t and $\mathbf{r}_+$ are the positions at time $t + \Delta t$, etc. Note that the forces can be saved from one time step to the next, and so there is only one force evaluation per time step required.

The introduction of damping and velocity-dependent forces results in:

$$\begin{aligned} \mathbf{r}_+ &= \mathbf{r}_0 + \mathbf{v}_0\Delta t + \frac{\widetilde{\mathbf{f}}_0\Delta t^2}{2m}, \\ \mathbf{v}_{1/2} &= \mathbf{v}_0 + \frac{\widetilde{\mathbf{f}}_0\Delta t}{2m}, \\ \widetilde{\mathbf{f}}_+ &= \left(\frac{1}{1+\gamma}\right)\left(\mathbf{f}_+ - \gamma\mathbf{v}_{1/2}\frac{2m}{\Delta t}\right), \\ \mathbf{v}_+ &= \mathbf{v}_{1/2} + \frac{\widetilde{\mathbf{f}}_+\Delta t}{2m}, \end{aligned} \tag{11}$$

where $\mathbf{f}_+$ is the conventional force calculated at $\mathbf{r}_+$ and $\widetilde{\mathbf{f}}_+$ is the modified force including the effects of damping and velocity-dependent terms. This new algorithm clearly reduces to standard Velocity Verlet in the limit $\gamma \to 0$ and has no significant additional cost.

2.3. Independent modes damping

Instead of damping all modes uniformly, the new algorithm we propose damps each mode independently of all others. We start by introducing a damping parameter, γ_i, for each degree of freedom

$$F_i = -kr_i - 2m_i\gamma_i\dot{r}_i. \tag{12}$$

We also note that in second-order dynamics we must have access to the force on each degree of freedom, and that this force will decay exponentially to zero as we approach the equilibrium configuration. Therefore it is possible to estimate the damping parameter for each degree of freedom of the second-order dynamics using only two steps of first-order dynamics (steepest descents with $\gamma^{\mathrm{SD}} = 1/\Delta t$)

$$\gamma_i^{\mathrm{c}}\Delta t = \sqrt{\ln\left(\frac{F_i^1}{F_i^2}\right)}, \tag{13}$$

where F_i^n is the ith component of the force for the nth step of steepest descents, etc.

It might be expected that “independent modes damping” will only improve over “coupled modes damping” if there is a clear decoupling of the various modes of the system. However, it will be shown in Section 3 that this is not the case in general, and that often “independent modes damping” is considerably more efficient than “coupled modes damping”.

2.4. Optimal time step

Having calculated an estimate of the optimal damping parameter(s) using either coupled- or independent-modes damping, we may now proceed to use this information to estimate the optimum time step for the second-order molecular dynamics. If we are considering the optimisation of the atomic configuration and not the coupled atomic-electronic system, then it would be natural to choose a time step determined by the characteristic time scale of the atomic motion(s). For undamped dynamics, it would be appropriate to consider a characteristic period τ of the atomic motion and typically $\Delta t \sim \tau/20$ is chosen in order to accurately integrate the atomic motion. But in damped dynamics, our interest is not so much in accurate integration of a trajectory, but in the accurate determination of the structure that minimises the forces and so a much larger time step can be used. We therefore propose that the information gained from the steepest descents steps (for either the coupled-modes or independent-modes damping schemes) be used to calculate the optimal time step as

$$\Delta t = \frac{\tau_{\min}}{N}, \tag{14}$$

where for most problems $N \sim 7$–8 is optimal and

$$\tau_{\min} = \begin{cases} \frac{2\pi}{\gamma^{\mathrm{c}}} & \text{coupled modes,} \\ \min\left\{\frac{2\pi}{\gamma_i^{\mathrm{c}}}\right\} & \text{independent modes,} \end{cases} \tag{15}$$

where min{ } is the minimum value over all degrees of freedom i in the independent-modes approach. This “optimal time step” can therefore be calculated at the same time as the damping factors.

Inevitably, as convergence is approached exponentially the damped dynamics will slow down. This can be offset to some extent by periodically resampling the damping parameter(s) and recalculating the optimal time step using the appropriate number of steepest descent steps. Obviously there is a tradeoff in efficiency between using a sub-optimal damping parameter(s) with second-order dynamics, and the less efficient first-order dynamics steps required in the recalculation. We find that recalculating the parameters every 30 steps is often a reasonable compromise.

As the faster modes are damped out, the time step chosen will progressively increase, so that as convergence is approached very large time steps may be used. Of course, as the forces and velocities are very small, the net result is still a very small movement of the atoms. However, this does require the use of a stable integration algorithm, and it may be appropriate to impose an upper limit to the time step for numerical stability reasons. We have found the modified Velocity Verlet algorithm (Eq. (11)) to be stable and sufficiently accurate for this purpose. As demonstrated in Section 3, this automatic time step recalculation has resulted in some very large time steps being successfully used near to convergence.

2.5. Mass preconditioning

Whilst the methods presented above, with the automatic time step selection, will be useful in many classical MD situations, and in some ab initio schemes, it might not always be appropriate to choose the time step according to the ionic dynamics. For example, within the Car–Parrinello coupled electronic-ionic approach, the time step has to be chosen according to the electron dynamics. Also, in Car–Parrinello dynamics there is a degree of freedom for each wavefunction coefficient in whatever basis is used – hence it might be too computationally expensive to have independent-modes damping. We therefore need a different variant in this case.

In damped molecular dynamics, the dynamics are merely an artifice to finding the ground state configuration – hence it is apparent that the masses of the particles are irrelevant. Knowing how to calculate a simple estimate of the period of each mode, it then becomes possible to reassign the masses of all the particles in the dynamics such that all the renormalised modes are now at the same period, where that period can be chosen as appropriate to the time step used. Therefore we can use a single damping factor at a given time step to generate optimally damped motion. We call this new extension, which can be applied to either the coupled-modes or the independent-modes damping schemes, the "mass preconditioning" scheme.

Considering the case of a single, one-dimensional harmonic oscillator as in Eq. (7), we see that it is the spring constant $k = m\omega^2$ that is the key feature that represents the underlying physics of the situation and so should be kept constant. Therefore we precondition the mass so as to maintain the original value of k using

$$m_{\text{new}} = m_{\text{old}}\left(\frac{\gamma^{\text{c}} N \Delta t}{2\pi}\right), \tag{16}$$

and therefore $\tau_{\text{effective}} = N\Delta t$.

This can be simply generalised to the many particle case. We choose the masses so that all the modes are at the same period – that of the fastest/dominant mode (depending on whether using independent or coupled mode schemes to estimate the optimal damping) – which tends to reduce the mass of heavy particles. It is possible that there might be stability problems if the new masses are set to too small a value, and so the lowest mass used needs to be bounded from below. For coupled modes damping we have a single damping factor γ^{c} and so this scheme corresponds to a simple mass rescaling, whereas for the independent-modes algorithm we are effectively compressing the vibrational spectrum.

Ideally, for the independent-modes algorithm, we should generate the full mass tensor [10,11] and diagonalise this to get the normal modes and proceed from there, but this is too expensive to calculate in practice. Instead, recognising that the trace of the mass tensor is invariant under unitary transformations, we approximate the mass of the jth particle using a simple average over the different components of $\gamma^{c}_{3j\to 3j+3}$ in Eq. (16). Once the masses have been appropriately renormalised, all the damping factors become $\gamma^{c} = 2\pi/N$ for all degrees of freedom – even in the case of independent-modes damping.

This consequently makes this approach feasible even within the Car–Parrinello ab initio scheme although in this case it would be best to reassign the masses of the electrons and ions separately, such that the coupled system contains a distinct electron mode and a distinct ionic mode, so as to maintain adiabaticity.

2.6. General considerations

One obvious question that should be asked is "how generally applicable is this algorithm?" The above derivation was based upon a simple harmonic oscillator, and so it will clearly apply to all systems if the initial configuration is sufficiently near the minimum where the quadratic approximation is valid.

What then are the effects of anharmonicity? If we consider a simple quartic perturbation to the harmonic potential, $U(x) = \alpha x^2 + \beta x^4$, with $\beta > 0$ to ensure only a single global minimum, then it is straightforwards to see that the effect of this perturbation is to increase the local curvature and hence increase the oscillation frequency, ω w.r.t. the harmonic oscillator ω_0. The "independent-modes damping" algorithm will therefore underestimate the damping, and so the damped dynamics will be in the under-damped regime. This will tend to produce oscillatory convergence which would be less than ideal. However, this can be easily detected as a reversal in the sign of F_i between successive steps and eliminated by quenching the velocity associated with that degree of freedom. This procedure is stable and efficient and is very useful in such situations – in fact, we find it a beneficial addition to the basic procedure and use it always.

What then may effect the stability of the algorithm? The "optimal time step" algorithm may sometimes overestimate the optimal time step that should be used (for instance the anharmonic case above). For this reason we need to impose a limit on the maximum possible allowed time step. We note that in chemical systems, the triple bond in N_2 is the strongest naturally occurring bond, and has an oscillation period of approximately 14.1 fs. So denoting the mass of a single nitrogen atom by $m(\mathrm{N})$, we use the following simple, conservative heuristic for choosing the time step that scales with the typical mass of the atoms in the system:

$$\Delta t^{\max} = \frac{\tau^{\max}}{N} = \frac{14.1}{N}\sqrt{\frac{\min\{m_i\}}{m(\mathrm{N})}}\ \mathrm{fs}, \tag{17}$$

which ensures stability. Similar considerations apply to the "mass preconditioning" calculation – the worst effects of anharmonicity are to underestimate the optimal mass and so the choice of mass should be bounded from below (e.g., taking the minimum mass as the mass of a single hydrogen atom). Similar "step-length" limiting considerations are often used to ensure stability in minimisation algorithms.

More serious is the question of how far from equilibrium can the initial configuration be and still have the algorithm converge (even if not at the optimum rate)? It can be easily shown that the damping parameter(s) (and hence time step or masses) will be very poorly estimated if there is a point of inflection in the potential between the initial configuration and the equilibrium one. This will reduce the efficiency of the convergence but it should still converge reliably, particularly with sensible limits to the time step/mass as discussed above. If we consider a Lennard–Jones pair-potential for a molecular system then there is a single point of inflection at a bond extension of approximately 10% of the equilibrium length. This corresponds to a gross distortion of the original molecule and is unlikely to be a problem in practice. Periodically resampling γ^{c} (and possibly assigning a new time step or masses) will be of substantial benefit in this situation.

Another question that needs to be considered is whether or not resampling the damping to accelerate the convergence of slower modes at long times will undo the damped modes that have already been converged at short times. As long as the total energy is converging monotonically, this will not be a problem as the fast modes generally have the largest contribution to the total energy. This may be a problem if the dynamics are in the under-damped region, whereupon the convergence may become oscillatory. Quenching is once again a useful technique in this case. Therefore, we advocate quenching and periodic resampling together with the stability heuristics above for a general purpose minimisation code.

3. Examples

In this section, the various algorithms discussed above will be tested with a variety of systems, which serve to illustrate some key points of the method. In Sections 3.1–3.3, we consider some simple "ball-and-springs" models which illustrate the effect of periodic resampling, and lead to some general conclusions about the efficacy of the different methods. Note that all the different methods have essentially the same computational cost (namely one energy and one force evaluation per step) and so we can compare efficiency by the total number of steps required to minimise the energy. The highly coupled models are simplistic representations of bulk solid calculations, whereas the uncoupled model is representative of a molecular calculation.

In Section 3.4 we compare the algorithms to an existing minimisation algorithm (BFGS) for a more complex situation – that of an ab initio calculation of an unrelaxed silicon vacancy in bulk silicon. Here, the comparison is based upon elapsed time, as the BFGS routine may take additional "line minimisation" steps within each basic iteration, and so the cost per iteration may vary due to a varying number of energy/force evaluations. We expect the BFGS routine to require less steps to reach the ground state than damped MD, but there is an additional ab initio speedup that can be used with damped MD – wavefunction extrapolation – which enables it to have a consistently faster time per step. Hence the net result should be of considerable interest.

3.1. Coupled tetrahedron

The first test system is shown in Fig. 2. It consists of a highly coupled system – a tetrahedron with unit masses at each vertex, joined by harmonic springs along the edges. Each spring has a natural length $l_0 = 0.5$ and the initial configuration is that all springs are stretched to length $l = 1.0$.

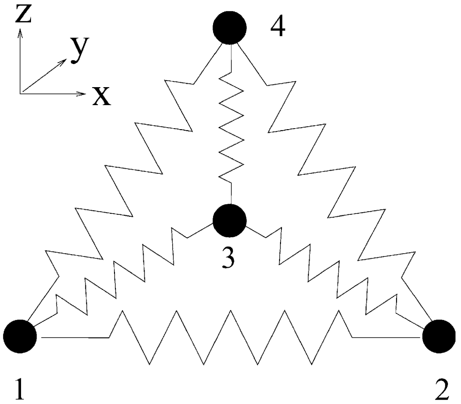

Fig. 2. Coupled tetrahedron model. Each vertex contains a point mass and each edge is an harmonic spring.

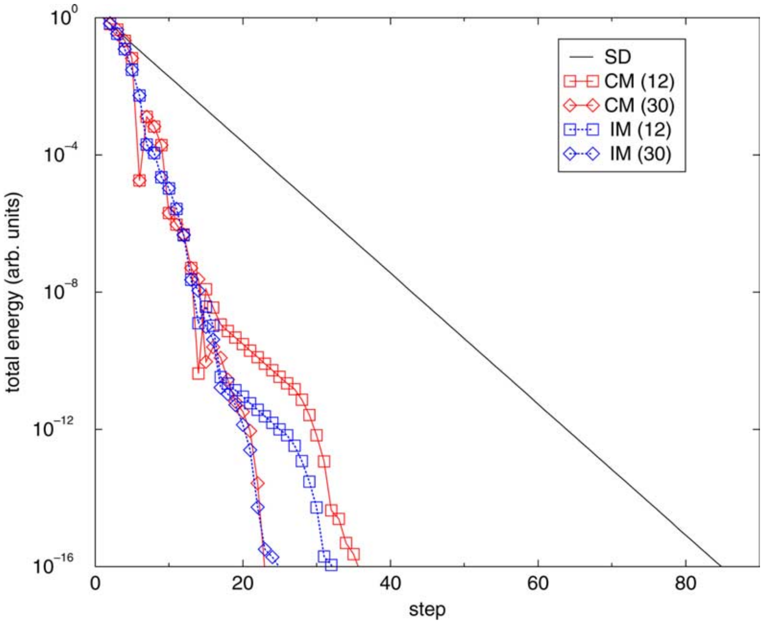

Fig. 3. Convergence of the total energy for the coupled tetrahedron model with uniform springs. The (12) and (20) labels indicate the damping resampling period.

The initial test consists of setting all spring constants to $k = 1.0$ and then running the three basic algorithms – steepest descents (SD), coupled modes (CM) and independent modes (IM) – with different rates of resampling the damping factors. The results are shown in Fig. 3. This clearly shows the effect of the resampling in periodically updating the optimal damping factors as the minimisation progresses. In all subsequent tests, the damping factors will be resampled every 30 steps. In this simple test there is little to distinguish the performance of CM and IM methods, but both are clearly superior to SD.

A slightly more taxing version of the tetrahedron test is to modify the spring constants along each edge, such that the spring constant between vertices is as given in Table 1.

With this modest spread in spring constants, it would now be expected that CM would outperform IM. The results are shown in Fig. 4 which also includes the performance of the automatic time step (“CM + dt” and “IM + dt”) and mass preconditioning (“CM + m” and “IM + m”) versions of both algorithms. It can be seen that both these modifications result in considerable performance benefits over the basic algorithms, with a net order-of-magnitude increase in convergence rate over SD. There is very little difference between CM + dt, CM + m, IM + dt and IM + m in this test.

Table 1
Spring constants used in the coupled tetrahedron model with non-uniform springs

Start vertex	Finish vertex	Spring constant
1	2	12
1	3	13
1	4	14
2	3	23
2	4	24
3	4	34

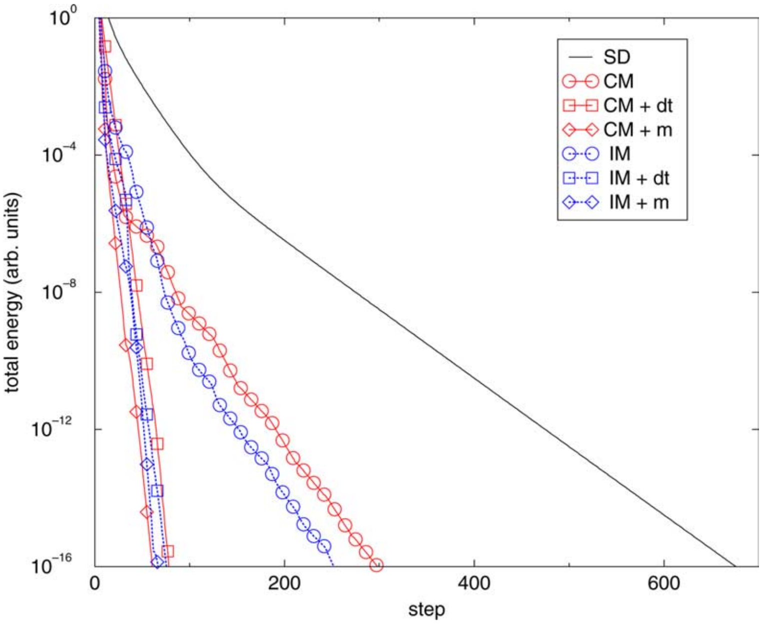

Fig. 4. Convergence of the total energy for the coupled tetrahedron model with non-uniform springs.

In order to better understand the effect of the mass preconditioning, it is instructive to calculate the vibrational density of states ($\mathrm{VDOS}(\omega) = \sum_i m_i |v_i(\omega)|^2$), using either the original or the preconditioned masses, in an undamped MD calculation. The result is shown in Fig. 5 where it can be seen that the preconditioning has lead to a considerable compression of the VDOS, from a broad spread of frequencies in the range $30 \leqslant \omega \leqslant 70$ to a single peak in the range $30 \leqslant \omega \leqslant 50$. This clearly shows the benefits of mass preconditioning on compressing the vibrational spectrum and hence reducing the spread in modes and increasing the efficiency of mode-based damping.

3.2. Coupled cube

The second test system is shown in Fig. 6. It also consists of a highly coupled system – a cube with unit masses at each vertex, joined by harmonic springs along the edges with different spring constants as in Table 2. Each spring has a natural length $l_0 = 0.5$ and the initial configuration is that all springs are stretched to length $l = 1.0$.

The performance of the different algorithms on this test system is shown in Fig. 7. This shows that as expected, for such a strongly coupled system, that the basic IM method is inferior to CM. Turning on mass preconditioning, however, makes a significant improvement in both cases, with CM + m being better than IM + m as expected. The VDOS shown in Fig. 8 helps to explain how even in the case of IM damping with such a strongly coupled system the mass preconditioning scheme can produce a significant improvement. There is a marked compression of the VDOS spectrum around the central peak, with a little additional weight appearing at higher frequencies.

However, IM + dt and CM + dt have the best performance of all and are essentially indistinguishable. The initial time step, based upon the spring constants and unit masses, was set at $\Delta t \sim 0.0355$. This was progressively increased as convergence proceeded as in Table 3 which shows the new time step at the end of each successive resampling period (three steps of SD for CM and two steps of SD for IM repeated every 30

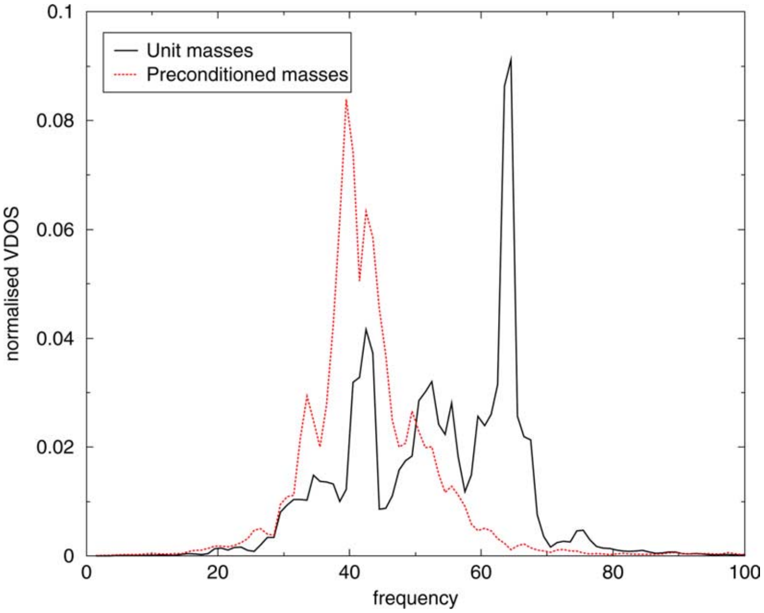

Fig. 5. Vibrational density of states for coupled tetrahedron model with non-uniform springs. The masses are either the original unit masses or those calculated by the mass preconditioning scheme using the IM algorithm.

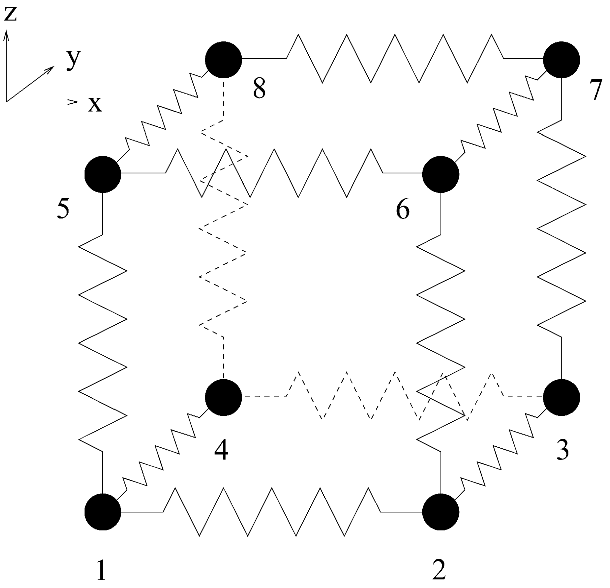

Fig. 6. Coupled cube model. Each vertex contains a point mass and each edge is an harmonic spring.

steps). The IM time step is conservatively set (based on the fastest of all modes sampled), whereas the CM time step is that of the dominant mode.

This (and the previous example) illustrates the validity and usefulness of the damped MD approach to finding the ground state geometry of bulk systems.

Table 2
Spring constants used in the coupled cube model with non-uniform springs

Start vertex	Finish vertex	Spring constant
1	2	12
1	3	13
1	4	14
2	5	25
2	6	26
3	5	35
3	7	37
4	6	46
4	7	47
5	7	58
6	8	68
7	8	78

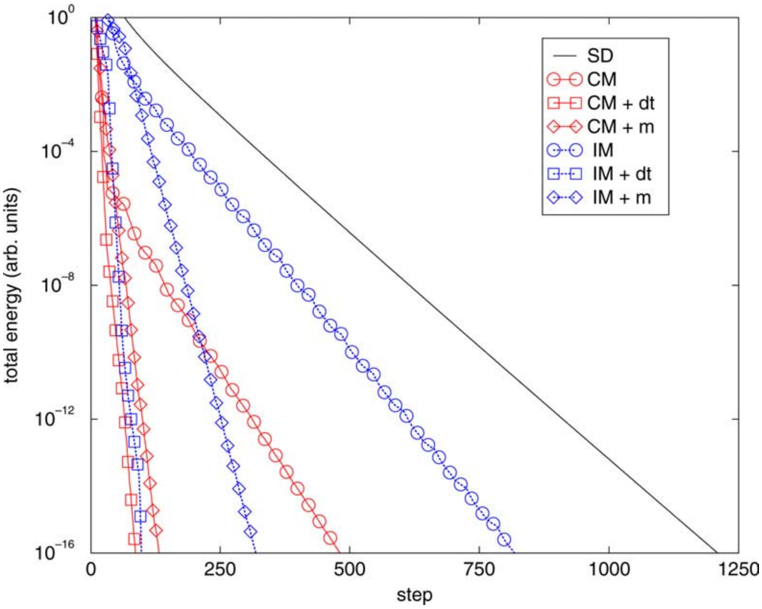

Fig. 7. Convergence of the total energy for the coupled cube model with non-uniform springs.

3.3. Uncoupled cube

Having considered some simple highly coupled model systems, in which it is expected that CM would be the superior algorithm, we now consider the opposite extreme, as shown in Fig. 9. It consists of unit masses at each vertex of a unit cube centred on the origin, each vertex being joined to the origin by harmonic springs. Each spring has a natural length $l_0 = 0.5$ and spring constants as given in Table 4.

The performance of the different algorithms on this highly uncoupled system is seen in Fig. 10. It is now apparent that SD and CM have severe difficulties as the convergence is dominated by the fastest mode and there is an order-of-magnitude spread in mode frequencies present. IM is an order-of-magnitude faster than

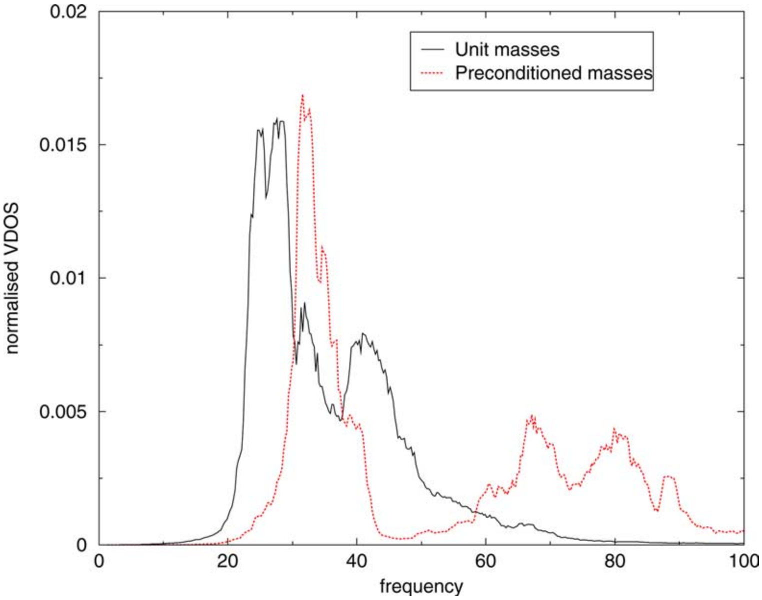

Fig. 8. Vibrational density of states for coupled cube model with non-uniform springs. The masses are either the original unit masses or those calculated by the mass preconditioning scheme using the IM algorithm.

Table 3
Time step used at different stages during the convergence of the coupled cube model with non-uniform springs

Resample	Δt_{CM}	Δt_{IM}
1	0.127	0.081
2	0.127	0.121
3	0.127	0.101
4		0.127

SD, and whilst mass preconditioning brings some gains, the best performance is to be had by IM + dt which is four times faster still.

This illustrates the validity and usefulness of the damped MD approach to finding the ground state geometry of molecular systems.

3.4. Silicon vacancy

Having considered various model systems, of varying complexity, lying at either extreme of uncoupled or very tightly coupled, we now consider the performance in a "real-world" calculation. The various algorithms were implemented on top of the existing MD routines in the ab initio plane-wave DFT code CASTEP v4.2 [12]. This code uses a conjugate-gradient approach to the electronic minimisation problem, and so is capable of using a BFGS-type minimiser for the ions which had already been implemented a number of years earlier. Whilst it is to be expected that the damped MD would take more steps to converge a given problem than the BFGS minimiser, there are additional speed-ups such as wavefunction extrapolation that can be used with MD and not with BFGS. Hence the most reasonable comparison between the two approaches must be based upon comparing wall-clock time for equivalent electronic calculations on the same computer and not the number of minimisation steps taken.

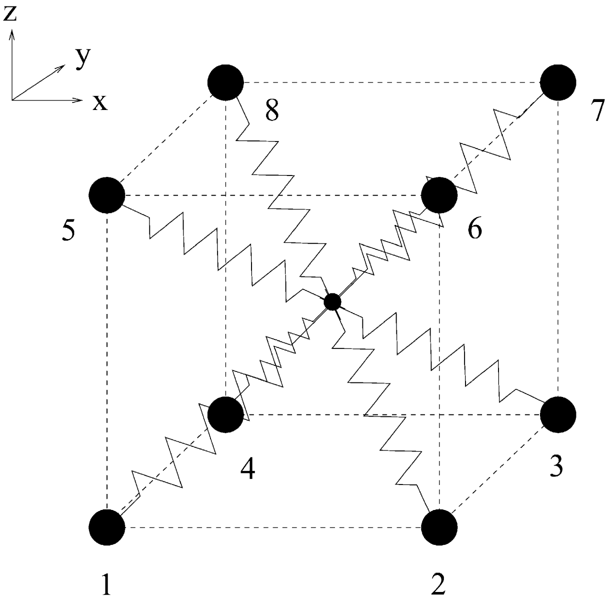

Fig. 9. Uncoupled cube model. Each vertex contains a point mass and is joined to the centre using an harmonic spring.

Table 4
Spring constants used in the uncoupled cube model with non-uniform springs

Vertex	Spring constant
1	10
2	20
3	30
4	40
5	50
6	60
7	70
8	80

The actual calculation undertaken was the relaxation of the neutral silicon vacancy. Whilst this might not be the hardest test case, it is an interesting one as previous ab initio studies have suggested that the energy surface around the minimum is quite flat. Hence it is a useful test of the reliability and accuracy of this approach. All calculations used the same electronic parameters (a well-tested ultrasoft pseudopotential for silicon with 120 eV plane-wave cutoff energy, the Perdew-Wang 1991 approximation to the exchange-correlation functional, and a $2 \times 2 \times 2$ Monkhorst-Pack grid for Brillouin zone sampling) and were continued until the same simultaneous convergence parameters were reached for the ionic configuration. The choice of convergence parameters were those commonly used in high quality calculations of defects (e.g. [13]) – namely that the energy difference per iteration be less than 0.00001 eV/atom, the RMS force be less than 0.001 eV/Å, and the RMS displacement be less than 0.0001 Å per iteration. The calculation contained 31 silicon atoms, initially situated at the perfect lattice sites around the vacancy in a 32-atom BCC supercell. Further details of the calculation are reported elsewhere [14].

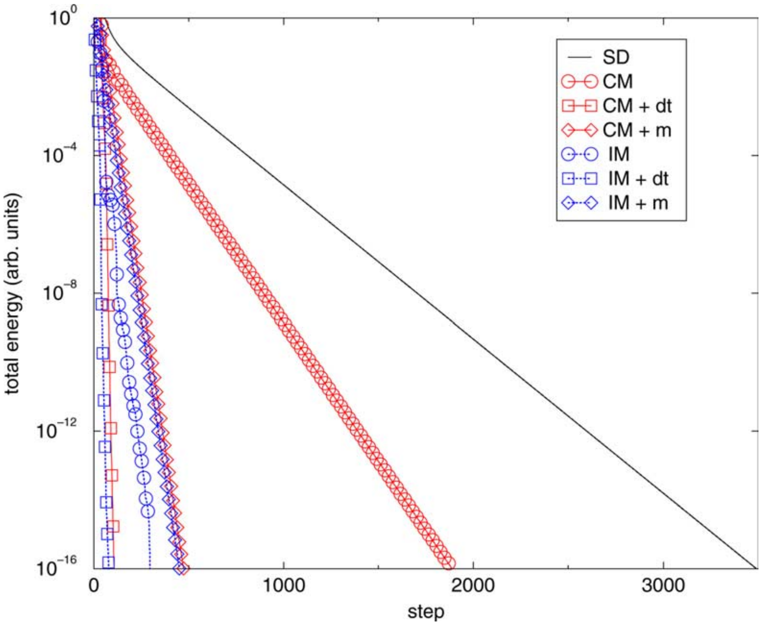

Fig. 10. Convergence of the total energy for the uncoupled cube model with non-uniform springs.

The convergence of the total energy vs. wall-clock time on a Silicon Graphics Octane (with a 300-MHz R12000 MIPS CPU) can be seen in Fig. 11. All calculations, using the same convergence criteria, converged to the same final structure to a remarkably high degree, which illustrates the accuracy and reliability of the

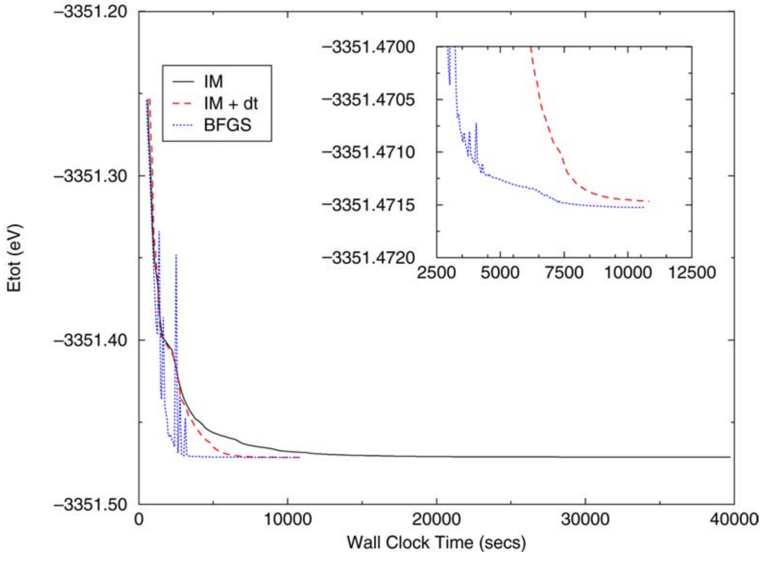

Fig. 11. Convergence of the total energy for the ab initio silicon vacancy calculation in a 31-atom BCC supercell. The inset shows a magnified portion of the final part of the ionic relaxation, showing the marginal difference between the performance of the BFGS and IM + dt minimisation strategies.

damped MD approach. The convergence rate of the basic IM calculation obviously slows down as equilibrium is approached, but IM + dt does not. As a test of the stability of the method, no upper limit was imposed on the time step allowed, and by the end of the calculation $\Delta t \sim 25$ fs was being safely used (although the actual motion of the ions was by this stage very slight). The net result is that the execution time for IM + dt (10833 s) was hardly distinguishable from that of the BFGS calculation (10621 s), as can be seen in the inset to Fig. 11 which shows the final approach to equilibrium in more detail.

This result may seem surprising – certainly, the expectation of the author had been that BFGS would be an order of magnitude faster. However, closer investigation revealed several clues (the upwards spikes in the total energy and the slowing down as convergence was approached) as to what was going on with the BFGS minimizer. The lack of a rigorous line minimisation due to the ab initio cost involved resulted in a general degradation of the Hessian with time. This explains the occasional upwards step (a wrong step direction was chosen before the routine back-tracked). Simple descriptions of the BFGS method [1] suggest that there should be an acceleration in the convergence rate as the minimum is approached and the quadratic regime is entered but this does not seem to be happening here. It would appear that the routine is either struggling to find a step direction or to set an appropriate step length after a large number of BFGS steps. One possibility (given the age of the code) is that there is no periodic resetting of the Hessian as is nowadays often recommended [15] to combat error accumulation. Such limitations of the BFGS implementation are quite common in ab initio codes due to the high cost of the energy/force evaluation.

It can be seen that the damped MD routine does not suffer from these deficiencies, and that with the periodic resampling and automatic time step selection its efficiency accelerates as convergence is approached. There is also no difficulty with “poisoning” of the step direction as this is determined by the second-order dynamics. Whilst the performance difference might be greater on other problems (and indeed on this problem with a different implementation of BFGS), there will always be situations for which BFGS will struggle to converge, whereupon having a second independent minimiser will always be a useful fall-back option.

4. Conclusions

We have introduced a new scheme, which we call “independent modes” (IM) for calculating the optimal damping parameters for geometry optimisation using damped MD. We have also introduced two additional refinements, namely automatic time step calculation and mass preconditioning, which can be applied to this scheme or to the “coupled modes” scheme introduced in [7]. These refinements result in significant speed advantages.

We have shown how these different schemes perform for a variety of model systems, comprising different mixtures of springs and masses. We also consider the performance on a “real world” calculation, namely the lattice relaxation around a silicon vacancy using an ab initio DFT treatment of the electrons and direct minimisation of the energy functional. This latter calculation showed that using IM + dt as the minimisation strategy results in a scheme that (for this system at least) is comparable in execution time to BFGS. This is partly due to the reduction in ab initio computation by using wavefunction extrapolation and partly due to the inexact BFGS implementation commonly found in ab initio codes (due to the high cost of rigorous line minimisation steps). We therefore propose that these schemes should be of widespread interest as they are simple to implement and efficient in practice, and can be used in conjunction with many different types of ab initio calculation, including Car–Parrinello and direct minimisation DFT codes. In more approximate schemes, where the BFGS implementation can afford to be more rigorous and the benefits of wavefunction extrapolation are lost, it is to be expected that this new scheme will not be as competitive.

As a general rule, we find that the independent-modes scheme with automatic time step calculation (“IM + dt”) to be fast and robust, and is our preferred choice for direct minimisation ab initio codes, whereas the independent-modes scheme with mass preconditioning scheme ("IM + m") should be used for Car–Parrinello codes.

Acknowledgements

This work was funded by a Lloyds of London Tercentenary Fellowship.

References

[1] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipies (FORTRAN), first ed., Cambridge University Press, Cambridge, 1989.
[2] R. Car, M. Parrinello, Unified approach for molecular-dynamics and density-functional theory, Phys. Rev. Lett. 55 (1985) 2471–2474.
[3] M.P. Teter, M.C. Payne, D.C. Allan, Solution of Schrödinger equation for large systems, Phys. Rev. B 40 (1989) 12255–12263.
[4] M.J. Gillan, Calculation of the vacancy formation energy in aluminum, J. Phys.-Condens. Matter 1 (1989) 689–711.
[5] T.A. Arias, M.C. Payne, J.D. Joannopoulos, Ab initio molecular-dynamics – analytically continued energy functionals and insights into iterative solutions, Phys. Rev. Lett. 69 (1992) 1077–1080.
[6] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 671–680.
[7] F. Tassone, F. Mauri, R. Car, Acceleration schemes for ab initio molecular-dynamics simulations and electronic-structure calculations, Phys. Rev. B 50 (1994) 10561–10573.
[8] W.C. Swope, H.C. Andersen, P.H. Berens, K.R. Wilson, A computer-simulation method for the calculation of equilibrium-constants for the formation of physical clusters of molecules – application to small water clusters, J. Chem. Phys. 76 (1982) 637–649.
[9] D. Frenkel, B. Smit, Understanding Molecular Simulations, Academic Press, San Diego, 1996 (Chapter 4).
[10] C.H. Bennett, Mass tensor molecular dynamics, J. Comput. Phys. 19 (1975) 267.
[11] D.B. Nicolaides, Why the Car–Parrinello method is too good to leave to the electronic structure theorists, CCP5 Newslett. 45 (1996) 52–60.
[12] Molecular Simulations Inc., CASTEP 4.2 Academic version, licensed under the UKCP-MSI Agreement, 1999.
[13] M.J. Puska, S. Poykko, M. Pesola, R.M. Nieminen, Convergence of supercell calculations for point defects in semiconductors: vacancy in silicon, Phys. Rev. B 58 (1998) 1318–1325.
[14] M.I.J. Probert, M.C. Payne, Improving the convergence of defect calculations in supercells – an ab initio study of the neutral silicon vacancy, Phys. Rev. B 67 (2003) 75204–75215.
[15] M.A. Bhatti, Practical Optimization Methods, Springer, New York, 1998 (Chapter 5).