PHYSICAL REVIEW B, VOLUME 64, 161102(R)

Linear scaling relaxation of the atomic positions in nanostructures

Stefan Goedecker, Frédéric Lançon, and Thierry Deutsch
Département de Recherche Fondamentale sur la Matière Condensée, SP2M/NM, CEA-Grenoble, 38054 Grenoble cedex 9, France

(Received 27 April 2001; revised manuscript received 9 July 2001; published 28 September 2001)

We present a method to determine the equilibrium geometry of large atomistic systems with linear scaling. It is based on a separate treatment of long and short wavelength components of the forces. While the rapidly varying part is handled by conventional methods, the treatment of the slowly varying part is based on elasticity theory. As illustrated by numerical examples containing up to a million atoms this method allows an efficient relaxation of large nanostructures.

DOI: 10.1103/PhysRevB.64.161102 PACS number(s): 68.65.−k

Considerable effort has recently been devoted to the development of various linear scaling algorithms. Only with these types of algorithms it is possible to perform atomistic simulations of large systems containing many atoms. Due to these efforts it is nowadays possible to calculate the total energy of the system and the forces acting on the atoms with linear scaling under many circumstances. Linear scaling can obviously been obtained if short-range inter-atomic potentials are used. With the help of sophisticated algorithms such as the fast multipole methods and particle mesh methods it is also possible to obtain linear scaling for inter-atomic potentials that include long-range electrostatic interactions. Finally linear scaling can be obtained if the forces are calculated quantum mechanically by using so called O(N) electronic structure methods.[1]

In spite of this important algorithmic progress concerning length scaling problems various time scaling problems persist. Ordinary molecular dynamics simulations can only cover relatively short-time intervals. Another related scaling problem is encountered in geometry optimizations of large atomistic systems. Geometry optimizations require the minimization of the total energy with respect to the atomic positions. The number of iterations required by standard minimization methods such as the conjugate gradient method increases with system size, destroying thus linear scaling even if the forces are calculated with such a scaling.

Even though the problem of the increasing number of iterations in geometry optimizations has been observed by many workers in the field, it has, to the best of our knowledge, not been analyzed up to now. We will therefore start with a detailed examination of the effect and relate it to well known facts about the convergence rate of iterative methods before presenting our solution on how to overcome it.

Figure 1 shows a linear chain, where the atoms are connected by elastic springs. The upper panel shows the equilibrium configuration, the lower one a configuration where the right half is shifted to the left, compressing the spring in the middle. Let us now consider what happens if we use the lower configuration as the starting point for a geometry optimization using any standard method such as the conjugate gradient or a quasi-Newton method. Because only the forces acting on the two atoms neighboring the compressed spring are nonvanishing, we will relax only these two atoms in the middle during the first iteration. During the subsequent iterations, nonvanishing forces will appear on all the atoms whose neighbors have been moved in previous iterations and hence they will be moved as well. Consequently it takes $n/2$ iterations for a chain of length n to propagate the perturbation from the compressed spring in the middle to the end. It follows that one needs at least of the order of n iterations to find the equilibrium configuration.

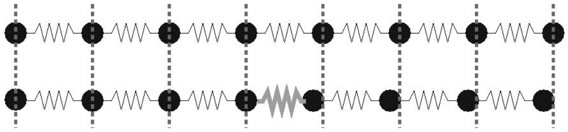

FIG. 1. A linear chain in equilibrium (top) and in a special out-of-equilibrium configuration (bottom).

One can recover the basic result of this intuitive derivation from a mathematical analysis. Writing down the elastic energy of a periodic three-dimensional model solid connected by perfect springs one finds that the eigenvalues λ_i of the Hessian matrix are the squares of the phonon frequencies ω_i. Mathematical theorems about the convergence of iterative methods[2] relate the number of iterations to the condition number κ, which is the ratio between the largest and the smallest eigenvalue of the Hessian matrix. Considering the case of large condition numbers κ and assuming that in the displacements that would bring the initial nonequilibrium configuration back into the equilibrium configuration all wavelengths are roughly equally represented, the number of iterations n_{CG} for the conjugate gradient method is given by

$$n_{it}^{CG} \propto \sqrt{\kappa}|\log(\epsilon)| = \left(\frac{\lambda_{max}}{\lambda_{min}}\right)^{1/2}|\log(\epsilon)| = \frac{\omega_{max}}{\omega_{min}}|\log(\epsilon)| \tag{1}$$

where ω_{min} is the smallest nonzero phonon frequency and ϵ the required precision. The highest phonon frequency ω_{max} arising either from an optic or acoustic branch is nearly independent of system size, whereas the acoustic phonon branch goes to zero at $\mathbf{k}=0$ linearly. The grid of allowed wave vectors $\mathbf{k}$ in the Brillouin zone becomes finer and finer as the periodic cell of the solid grows. The smallest phonon frequency ω_{min} is related to the grid point that is closest to the origin and is proportional to $1/L_{max}$, where L_{max} is the largest linear dimension of the system. The number of iterations is therefore given by

$$n_{it}^{CG} \propto \frac{L_{max}}{a}|\log(\epsilon)|,$$

where a is a typical inter-atomic distance. For the steepest descent method the convergence is slower:

$$n_{it}^{SD} \propto \frac{\lambda_{max}}{\lambda_{min}} |\log(\epsilon)| = \frac{\omega_{max}^2}{\omega_{min}^2} |\log(\epsilon)|. \tag{2}$$

Even though these results were derived for periodic solids numerical simulations clearly show that they also hold true for regular nonperiodic structures such as linear polymers, two-dimensional sheets and bulklike clusters.

It is interesting to relate the number of conjugate gradient iterations to the number of atoms N in the system. For a one-dimensional system we find [Eq. (1)] that n_{it}^{CG} is proportional to the number of atoms N, in the two-dimensional case $n_{it}^{CG} \propto N^{1/2}$ and in the three-dimensional case $n_{it}^{CG} \propto N^{1/3}$ assuming that all the side lengths are comparable. The possible gains by using a linear scaling algorithm are thus largest in the one-dimensional case.

In chemistry, geometry optimizations are usually done in internal coordinates pioneered by Pulay and co-workers (see Refs. 3 and 4, and references therein) instead of Cartesian coordinates. An extension of this approach to periodic systems has also been investigated.[5] The internal coordinates typically consist of bond stretching, bending, and torsion coordinates. Their construction is not easy for complicated geometries and transformations between the Cartesian and internal coordinates are required in each iteration step. These transformations are costly and have a cubic scaling when done by conventional linear algebra methods. Recently ways have, however, been found to do these transformations with linear scaling.[3] For the example in Fig. 1 there would be no dependence of the number of iterations on the size of the system if the optimization was done in internal coordinates (just bond lengths in this case). This property is also satisfied for simple polymers, but is lost when one has molecules with a higher dimensional character or bulklike structures.

Linear scaling is generally achieved by treating different length and time scales in appropriate ways. This will also be the guiding principle in this work. As we have seen in the preceding discussion the slowdown of the convergence rate is due to the long-wavelength acoustic phonons. Hence the basic idea is to treat high and low frequencies in a different way. The reduction of the high frequency components of the force during a geometry optimization is satisfactory with the standard methods such as conjugate gradient and steepest descent. To reduce the low frequency components we will use elasticity theory. This is motivated by the fact that ordinary materials have continuumlike behavior on length scales of just a few inter-atomic distances. This divide and conquer approach in frequency is illustrated in Fig. 2.

Elasticity theory, discretized on a grid of M grid points, gives the total elastic energy E_{el} of a system as a quadratic form of the $3M$ displacements u_i from the equilibrium configuration:

$$E_{el} = \frac{1}{2} \mathbf{u}^T \mathbf{A} \mathbf{u}.$$

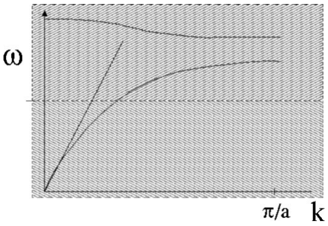

FIG. 2. Illustration of the subdivision into a high frequency part treated by conventional methods and a high frequency part described by elasticity theory. The straight line is the ‘‘phonon’’ dispersion relation of an elastic solid in the continuum limit.

To build up A one needs the elastic constants of the material.[6] Differentiating with respect to $\mathbf{u}$ and using the fact that the force $\mathbf{f}$ is just the negative of the derivative of the energy with respect to the atomic positions we obtain

$$-\mathbf{A}\mathbf{u} = \mathbf{f}. \tag{3}$$

Equation (3) is the central equation for the relaxation steps of the minimization iteration that are based on elasticity theory. Given a continuous force field $\mathbf{u}(\mathbf{r})$ discretized on the M grid points it allows us to calculate the discretized displacement field that would bring the solid into its equilibrium configuration. An ideal harmonic solid can be relaxed with one single step in this way. If we apply this step, as will be explained below, to our real solid it drastically brings down the limiting low frequency components of the force.

Equation (3) cannot be applied in a straightforward way to an atomic system. The derivative of a total energy expression or an inter-atomic potential gives only the forces acting on the atoms but not a discretized continuous force field. In the same way we need the atomic displacements for the atomic moves and not a discretized displacement field. Methods to transform between atoms and grid points are well known from the various particle mesh methods.[7] Following these ideas we map the atomic forces onto the grid by smearing them out with the function $(1-x^2/8)^4(1-y^2/8)^4(1-z^2/8)^4$ (the unit length is the grid spacing). The transformation from the grid onto the atoms is done by cubic interpolation.

Just by invoking elasticity theory we have actually not yet really solved the scaling problem. If we were to solve Eq. (3) by a conjugate gradient method the number of iterations would as well increase with respect to the number of grid points. In this case the method might still be useful since under most circumstances a conjugate gradient step for Eq. (3) will be significantly less expensive than for the atomic system, but the overall scaling would not be linear. Fortunately recent developments in the field of multi-grid methods[8] allow us to solve sparse linear systems of the type of Eq. (3) with linear scaling and with small prefactors. In the case of periodic systems one can also use fast Fourier transform techniques that exhibit a nearly linear $M\log(M)$ scaling. In summary, if we can calculate the forces with linear scaling and solve Eq. (3) with linear scaling using the

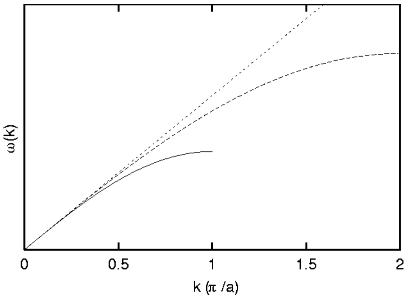

FIG. 3. The ''phonon'' dispersion relation of two grids used for the solution of Eq. (3). The wave vector *k* is measured in units of π/a, where *a* is the lattice constant of the physical system. The solid line corresponds to a grid whose spacing is equal to the inter-atomic distance *a*, the dashed line to the case where the grid spacing is half of the inter-atomic distance, and the dotted line is the continuum limit.

above mentioned methods overall linear scaling can be obtained for the geometry optimization problem in Cartesian coordinates.

The complete algorithm consists of a iteration of the following steps.

(1) Perform n_c conventional minimization sweeps. The choice of n_c is discussed below.

(2) Perform one step where the solid is described by elasticity theory. (a) Calculate the forces acting on the atoms; (b) transfer the atomic forces onto a computational grid; (c) solve Eq. (3) with a linear scaling method to obtain the displacements; (d) evaluate the numerical displacement field at the atomic positions to get the atomic displacements; (e) move the atoms along these displacement directions. This can either be done using a fixed step size or by a line minimization along these directions. We choose the latter variant.

Continuum elasticity theory considers the limit where the ratio between the wavelength of any perturbation is much larger than the inter-atomic spacing. In this limit the phonon dispersion relation is a straight line as shown in Fig. 3. The point at which the phonon dispersion relation of the true solid starts to deviate from a straight line tells us the length scale at which the solid can already be considered as a continuous elastic medium. Once one has a discretized version of elasticity the ''phonon'' dispersion relations are not any more straight lines but the phonon dispersion relation of a model harmonic solid. The size of their Brioullin zone depends on the density of the numerical grid as shown in Figs. 2 and 3.

The number of conventional sweeps n_c in our method depends on the size of the frequency range that the conventional part needs to cover. This range depends both on the physical properties of the system and on the density of the numerical elasticity grid. Obviously the range will be larger if there are high optical phonon bands. It will also be significantly larger if the number of grid points is less than the number of atoms since in this case the highest frequency that can be reached by the elastic model is small. According to our experience it is best to choose the number of grid points *M* to be in the interval $[N\colon 2^d N]$, where *d* is the dimension. Increasing *M* within this range reduces the number of iterations but each iteration becomes more expensive since the solution of Eq. (3) takes longer. Increasing *M* further does not lead to a significant additional reduction of the number of iterations. Choosing *M* smaller than *N* is not reasonable since for such values the number of iteration increases strongly and the effort for the solution of Eq. (3) is completely negligible. We found that for grid densities in the recommended interval the best values for n_c were between 2 and 4. For silicon systems conjugate gradient sweeps were slightly more efficient than steepest descent sweeps.

TABLE I. Number of force evaluations n_f and CPU time *T* in seconds for the conjugate gradient (CG) and the linear scaling (SC) method for a divacancy in silicon.

Number of atoms	n_f (CG)	n_f (LS)	*T* (CG)	*T* (LS)
510	102	106	0.41	0.50
998	124	106	0.90	0.93
1726	146	109	1.7	1.6
4094	184	115	5.1	4.2
13822	260	115	24.0	14.0
110592	502	115	373.0	135.0
884734	934	117	5586.0	1147.0

Even though each of the two steps in our method targets a specific frequency range a clearcut separation is not possible. The conjugate gradient step will slightly dampen the low frequency component. The elasticity step will introduce some additional short wavelength errors since the ''phonon'' dispersion relation for the elastic grid model deviates from the phonon dispersion relation of the material in the high frequency region. These errors are however very small since these components were already significantly decimated by the preceding conventional steps and they are therefore immediately annihilated by the following conjugate gradient steps.

Table I shows numerical results for the relaxation of a divacancy in silicon crystals of varying cell size modeled with the environment dependent interatomic potential (EDIP) inter-atomic potential.[9] The norm of the forces $\sqrt{\Sigma_i f_i}$ was reduced down to 10^{-8} resulting in total energies that were converged to machine precision. The initial configuration was the perfect silicon crystal with two neighboring atoms removed. We compare both the number of force evaluations and the CPU time of the conjugate gradient method with our method. All line minimizations were done by minimizing the projected force and under the assumption that the energy function is quadratic. This results in just two force evaluations per conjugate gradient step and turned out to be the most efficient conjugate gradient implementation. The elastic equations of an isotropic homogeneous medium were solved using Fourier techniques. The two elastic constants needed in this context were calculated numerically by applying a compression and a shear to the system.

Taking the CPU time as the criterion, the crossover-point is at around 1000 atoms for EDIP inter-atomic potential that is fast to evaluate. If we would use another total energy scheme where the calculation of the forces is more costly, the number of force evaluations would be the best criterion and the crossover point goes down to 600 atoms roughly.

The divacany example allows the examination of the scaling behavior over a large range of system sizes, since the behavior of the divacancy remains qualitatively the same. In many other nanosystems the behavior can fundamentally change with size due, for instance, to size dependent surface reconstructions. This can then lead to highly irregular trends in the number of iterations. For these reasons we have chosen the divacany example even though it is a particularly difficult example to demonstrate the advantages of the linear scaling method over the conjugate gradient method. It is fully three dimensional and the importance of the long wavelength perturbations induced by the di-vacancy decreases with system size, leading to an increase of the number of conjugate gradient iterations that is slightly slower than predicted by Eq. (1).

We have also applied the method to more complicated systems such as incommensurate interfaces in silicon and clusters of silicon. In all these cases we obtained cross over points of 1000 atoms or below. As expected the gains were particularly large for systems with two-dimensional character. Our previous analysis of the number of iterations for geometry optimization was based on the assumption that we are in a region where the energy functional has nearly quadratic behavior. Our numerical experiments indicate however that our linear scaling method also offers advantages in the case where one starts the geometry optimization from a point that is far away from such a region. In this case multiple minima are frequently encountered. We found that our linear scaling method is more likely to find lower energy local minima than the conjugate gradient method.

We have presented a method that allows us to relax very large systems with linear scaling, removing thus an important bottleneck in atomistic simulations. This method in combination with inter-atomic potentials and O(N) electronic structure methods allows efficient geometry optimizations for large solid state materials and nanostructures. It is expected that it can be extended to large molecular and biological systems for which sufficiently simple elasticity models exist.

S. G. thanks M. Brezina, K. Nemeth P. Pulay, and G. Scuseria for interesting discussions.

---

[1] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999); P. Ordejon, Comput. Mater. Sci. **12**, 157 (1998); G. Scuseria, J. Phys. Chem. **103**, 4782 (1999).

[2] Y. Saad, *Iterative Methods for Sparse Linear Systems* (PWS Publishing Company, Boston, 1996).

[3] K. Nemeth, O. Coulaud, G. Monard, and G. Angyan, J. Chem. Phys. **113**, 5598 (2000).

[4] B. Paizs, G. Fogarasi, and P. Pulay, J. Chem. Phys. **109**, 6571 (1998); J. Baker, A. Kessi, and B. Delley, *ibid.* **105**, 192 (1996).

[5] K. Kudin, G. Scuseria, and B. Schlegel, J. Chem. Phys. **114**, 2919 (2001).

[6] E. M. Lifshitz and L. D. Landau, *Theory of Elasticity* (Butterworth-Heinemann, Oxford, 1995).

[7] H.G. Petersen, J. Chem. Phys. **103**, 3668 (1995).

[8] M. Brezina, A. Cleary, R. Falgout, V. Henson, J. Jones, T. Manteuffel, S. McCormick, and J. Ruge, SIAM. J. Sci. Comp. **22**, 1570 (2001). P. Vanek, M. Brezina, and J. Mandel, Numer. Math. **88**, 559 (2001).

[9] M.Z. Bazant and E. Kaxiras, Phys. Rev. Lett. **77**, 4370 (1996); M.Z. Bazant, E. Kaxiras, and J.F. Justo, Phys. Rev. B **56**, 8542 (1997); J.F. Justo, M.Z. Bazant, E. Kaxiras, V.V. Bulatov, and S. Yip, *ibid.* **58**, 2539 (1998).