Structural Relaxation Made Simple

Erik Bitzek,[1] Pekka Koskinen,[2,3] Franz Gähler,[4] Michael Moseler,[2,3,5] and Peter Gumbsch[1,2]

[1]*Institut für Zuverlässigkeit von Bauteilen und Systemen (izbs), Universität Karlsruhe (TH), Kaiserstr.12, 76131 Karlsruhe, Germany*
[2]*Fraunhofer-Institut für Werkstoffmechanik IWM, Wöhlerstr. 11, 79108 Freiburg, Germany*
[3]*Faculty of Physics, University of Freiburg, Hermann-Herder-Str. 3, 79104 Freiburg, Germany*
[4]*Institut für Theoretische und Angewandte Physik (ITAP), Universität Stuttgart, Pfaffenwaldring 57/VI, 70569 Stuttgart, Germany*
[5]*Freiburg Materials Research Center (FMF), University of Freiburg, Stefan-Meier-Str. 21, D-79104 Freiburg, Germany*

We introduce a simple atomic structure optimization algorithm which is significantly faster than standard implementations of the conjugate gradient method and which is competitive with more sophisticated quasi-Newton schemes typically used in *ab initio* calculations. It is based on conventional molecular dynamics with additional velocity modifications and adaptive time steps. The efficiency, robustness and versatility of the method is illustrated using a variety of test cases, including typical systems encountered in nano-scale science, solid state physics, materials research and biochemistry.

PACS numbers: 61.46.-w,02.60.Pn,02.70.Ns

Finding mechanically stable equilibrium configurations of atomistic systems is one of the most common tasks in computational materials science, solid state physics, chemistry and biology. This corresponds to finding the (nearest) structures with minimum potential energy for an atomic system, starting from a given initial configuration. To solve this task a variety of well established optimization methods, like steepest descent (SD), conjugate gradient (CG), Newton-Raphson (NR), Quasi-Newton (QN) or Truncated-Newton (TN) methods can be used [1–4]. Also variants of molecular dynamics (MD) methods which systematically remove kinetic energy from the system are commonly applied for minimization purposes [5–8]. Such local 'quenching' is also applied in many algorithms for global minimization [9–11]. Interestingly, relaxation based on MD has been thought to be good for practical realization, but not very competitive with the afore mentioned sophisticated algorithms. For this reason MD-quenching algorithms have been traditionally introduced as by-products of secondary importance in regular articles [5–7], never receiving the attention and effort they actually deserve.

Here we introduce and discuss a simple and yet powerful MD scheme for structural relaxation. Consider a *blind skier* searching for the fastest way to the bottom of a valley in an unknown mountain range described by the potential energy landscape $E(\mathbf{x})$ with $\mathbf{x} = (x_1, x_2)$. Assuming that the skier is able to retard and steer we would recommend him to use the following equation of motion:

$$\dot{\mathbf{v}}(t) = \mathbf{F}(t)/m - \gamma(t)|\mathbf{v}(t)|(\hat{\mathbf{v}}(t) - \hat{\mathbf{F}}(t)) \;, \qquad (1)$$

with the mass m, the velocity $\mathbf{v} = \dot{\mathbf{x}}$, the force $\mathbf{F} = -\nabla E(\mathbf{x})$, and hat for a unit vector. Beyond the classical equation of motion based exclusively on the force acting on a particle, the skier may make use of additional acceleration in the direction of the force or the velocity vectors via the function $\gamma(t)$. In order to avoid uphill motion the skier should simply stop as soon as the power $P(t) = \mathbf{F}(t) \cdot \mathbf{v}(t)$ becomes negative. In the case $P(t) > 0$, we would recommend a continuous moderate steering in the downhill direction, realized by a small and positive $\gamma(t)$.

We show in this letter that eq. (1) brings the skier surprisingly fast to the desired destination. A properly discretized version of eq. (1) in combination with an adaptive time step results in an extremely simple and robust minimization scheme for multidimensional functions $E(x_1, \ldots x_M)$ which is nevertheless competitive with the above mentioned sophisticated optimizers. Fig. 1 shows that our skier easily keeps up with powerful standard schemes like the conjugate gradient and the QN Broyden-Fletcher-Goldfarb-Shanno (BFGS) scheme [1, 2] in a two-dimensional spiral potential landscape. In calculations on a broad range of test systems the new algorithm was always surprisingly fast and could be used with great ease for systems with millions of degrees of freedom and with very demanding convergence criteria. Contrary to the conventional schemes the new algorithm relies on inertia and, consequently, this novel method was dubbed *FIRE* for *Fast Inertial Relaxation Engine*.

The numerical treatment of eq. (1) is quite simple. Any common MD integrator can be used for the propagation according to the conservative force term on the right hand side of eq. (1). The resulting trajectory is continuously re-adjusted by two kinds of velocity modifications: a) the above mentioned immediate stop upon uphill motion ($\mathbf{v} \to 0$) and b) a simple mixing of the global ($3N_{atoms}$ dimensional) velocity and force vectors $\mathbf{v} \to (1-\alpha)\mathbf{v} + \alpha\hat{\mathbf{F}}|\mathbf{v}|$, which results from an Euler-discretization of the last term in eq. (1) and from the introduction of Δt with $\alpha = \gamma\Delta t$. Both Δt and α are

treated as adaptive quantities.

Explicitly, the FIRE algorithm uses the following propagation rules (given initial values for Δt, $\alpha = \alpha_{start}$ and for the global vectors $\mathbf{x}$, $\mathbf{F}$ and $\mathbf{v} = 0$):

MD: calculate $\mathbf{x}$, $\mathbf{F} = -\nabla E(\mathbf{x})$, and $\mathbf{v}$ using any common MD integrator; check for convergence

F1: calculate $P = \mathbf{F} \cdot \mathbf{v}$

F2: set $\mathbf{v} \to (1-\alpha) \cdot \mathbf{v} + \alpha \cdot \hat{\mathbf{F}} \cdot |\mathbf{v}|$

F3: if the number of steps since condition (F4) was true is larger than N_{min}, increase the time step $\Delta t \to \min(\Delta t \cdot f_{inc}, \Delta t_{max})$ and decrease $\alpha \to \alpha \cdot f_\alpha$.

F4: if $P \leq 0$, decrease time step $\Delta t \to \Delta t \cdot f_{dec}$, freeze the system $\mathbf{v} \to 0$ and set α back to α_{start}.

F5: return to MD

Contrary to MD, for relaxation the accurate calculation of the atomic trajectories is not necessary. The adaptive time step allows FIRE to increase Δt until either the largest stable time step Δt_{max} is reached, or an energy minimum along the current direction of motion ($P < 0$) is encountered. In the latter case the system is instantly frozen ($\mathbf{v} \to 0$) and the time step is substantially reduced in order to resolve the trajectory of the system close to a minimum or turning point. A short 'latency' time of N_{min} MD steps before accelerating the dynamics of the system is important for the stability and robustness of FIRE.

Most of the parameters introduced above are not directly linked to the physics of the system. Like the five parameters of the standard conjugate gradient implementation, their choice is guided by experience. For all systems under study, the following parameters yielded a fast and robust behavior: $N_{min} = 5$, $f_{inc} = 1.1$, $f_{dec} = 0.5$, $\alpha_{start} = 0.1$ and $f_\alpha = 0.99$.

The maximal time step Δt_{max} is the most important adjustable parameter of the method. From a typical MD simulation time step Δt_{MD} one can obtain an initial rough estimate of $\Delta t_{max} \sim 10 \cdot \Delta t_{MD}$. The initial time step can be rather freely chosen, e.g. $\Delta t \sim \Delta t_{MD}$.

One feature of special attention is the global nature of the algorithm, which assumes that all degrees of freedom are *comparable* to each other. All the velocities should have the same scale, which for hetero-nuclear systems can be roughly achieved by setting all the atom masses equal. Note that this rescales also the time step.

To demonstrate the performance of FIRE, we compare it to two commonly used relaxation methods: BFGS and the Polak-Ribiere version of CG. BFGS is popular in quantum mechanical calculations of small systems where the system evaluation at a new location is the most time consuming part [3, 4]. However, the storage and arithmetic cost of this method can become prohibitive for large systems. For large problems with computationally inexpensive atomic interaction models CG is widely used [3, 12] [19].

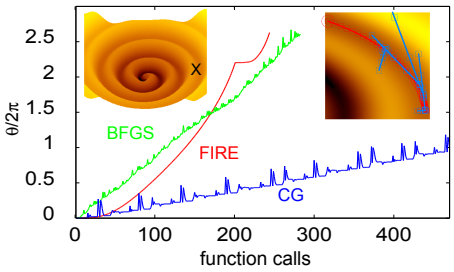

FIG. 1: Performance of FIRE in a spiral-shaped two-dimensional potential energy function (see left inset for the landscape, X is the starting point). Shown is the evolution of the azimuthal angle θ versus the function calls of FIRE (red solid line), CG (blue dotted line) and BFGS (green dashed line). Since FIRE needs some time to accelerate, it is slow at the beginning, but catches up quickly with BFGS at later times. Note that CG does not converge within 500 function calls. This is due to expensive line searches as displayed in the right inset showing a part of the trajectory of FIRE (red) and CG (blue).

We use the CG from *Numerical Recipes* [2] and BFGS from the IMSL library. Although there are more specialized implementations of CG [13] and QN methods available (see [4] and references therein), these algorithms are widely used and well documented, making them ideal reference methods. In the comparisons the root-mean-square (RMS) of the global force $F_{RMS} = |\mathbf{F}|/(3N)^{\frac{1}{2}}$ (force norm) is taken as a representative measure for the degree of relaxation, but also further conditions for minimum, such as the maximum force components, are tested. The number of 'function calls' n is a generic notation for separate points $\mathbf{x}$ where either energy, force, or both are evaluated.

As a first demonstration, Fig. 1 shows FIRE, CG and BFGS optimizations of a function $E(x_1, x_2) = \sin(\pi r + \theta/2) + r^2/10$, modeling a curved relaxation pathway. In atomic systems, curved relaxation paths are the result of the usually highly corrugated, intricate potential energy surfaces. In this model, FIRE ($m_i = 1$ and $\Delta t_{max} = 0.30$) reaches the minimum ($E < -0.99$) with 244 gradient calls. BFGS needs 283 energy and 215 gradient evaluations, compared to 1071 energy and 82 gradient calls for CG with a total of 82 line minimizations. In terms of required function evaluations FIRE compares well with BFGS, without having similar computational overhead or memory requirements. The line searches of the conjugate gradient are clearly inefficient to follow the curved path of the spiral. MD type scheme gives in this case a smooth down-hill trajectory.

Next we apply FIRE to the biomolecule fenretinide, which is used as a cancer drug (see inset in Fig. 2 for

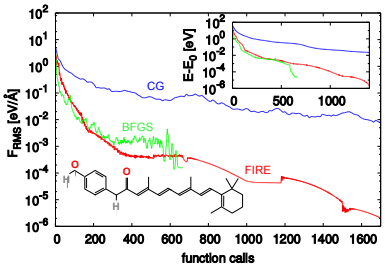

FIG. 2: Relaxation of fenretinide (Lewis structure is shown in the lower inset) modeled with density-functional based tight-binding. The force norm as a function of the number of function evaluations is shown for FIRE (red solid line), CG (blue dotted line) and BFGS (green dashed line). The upper inset shows the evolution of the total energy E with respect to the equilibrium value E_0.

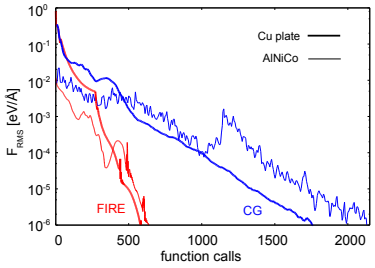

FIG. 3: Relaxation of the AlNiCo quasicrystal and the hot Cu plate. The force norm as a function of the number of function evaluations is shown for FIRE (red solid line) and CG (blue dotted line).

the structure). The atomic interaction was modeled with a density-functional based tight-binding method (see ref. [14] and references therein). The starting configuration was created by slightly bending and twisting the carbon chain. This setup is particularly challenging, since the work for the straightening and unwinding of the long chain has to be done by single carbon-carbon bonds. In FIRE all the masses were set to one atomic mass unit, $\Delta t_{max} = 1$ fs and the initial time step 0.8 fs.

For quenching the molecule down to $F_{RMS} < 5 \cdot 10^{-}3$ eV/Å CG required $n = 2097$ function calls [20], BFGS required $n = 171$ and FIRE $n = 161$. Afterwards the relaxation was carried further, until the numerical limit of the respective methods was reached. The final force norm was for CG $1.7 \cdot 10^{-3}$ eV/Å ($n = 3644$) and for BFGS $1.6 \cdot 10^{-4}$ eV/Å ($n = 667$). FIRE, however, was able to reach a force norm as small as $1.0 \cdot 10^{-15}$ eV/Å ($n = 5534$). The final structures produced by CG and BFGS were 3 meV and 0.02 meV higher in energy than the final structure of FIRE, respectively. Note, however, that the performance of BFGS in the parabolic region close to the minimum (inset of Fig. 2) is very good. Analysis of the trajectory of relaxation shows that, indeed, due to its global nature and inertia, FIRE is efficient in getting the chain straightened in its overall form. For CG the straightening process was crowded with inefficient line search directions, indicating the *high curvature* of the minimization pathway.

The tight-binding model was also applied for quenching of a $(5, 5)$-nanotube with fullerene endcaps (160 atoms) thermalized to 1000 K, with the criterion $F_{RMS} < 5 \cdot 10^{-3}$ eV/Å. In this case the performance of BFGS is better ($n = 61$) compared to FIRE ($n = 102$) or CG ($n = 532$), due to a fairly regular and parabolic energy landscape [21].

The relaxation behavior of two further test systems are shown in Fig. 3. The first system is an approximant to a decagonal AlNiCo quasicrystal [15, 16]. Periodic boundary conditions (PBC) with a fixed box size were applied. This system has a complex ground state with shallow minima for many atoms, requiring thus a rather small convergence criterion for the force. The starting configuration for the other test system in Fig. 3 is taken from the equilibration of a free-standing thin copper film at $T = 1000$ K and therefore shows relatively high forces on the atoms. This problem combines the relaxation of the long range thermal expansion with the relaxation of local displacements. The atomic interactions for both systems were modeled with embedded atom method (EAM) potentials [16, 17]. In both examples CG requires approximately three times more function evaluations as compared to FIRE even though the relaxed structures differ only by 10^{-10} and 10^{-13} eV/atom, respectively.

Further tests were conducted on a broad range of different systems. Amongst others were other quasi-crystalline samples, the relaxation of a crack and of vacancies in systems of different size. Similar to the crack, the relaxation of a vacancy in large boxes under PBC requires the adaption of the atomic positions to a long range strain field, however in this case at very low driving forces. The large number of atoms with low forces translates to a low starting F_{RMS}, therefore the smallness of the maximum force component experienced by an atom is used as convergence criterion. The crack, on the other hand, is characterized by large strains. For this example, the initial positions of the atoms around the crack and at the fixed borders were determined by the anisotropic linear elastic solution for a sharp crack at the Griffith load in Ni (for details see [18]). This system shows that despite its dynamic nature FIRE can be used to determine the mechanical stability of structures very close to their critical load.

The function calls necessary to relax the different test systems up to a given convergence criterion with CG and FIRE are summarized in Tab. I. In all cases FIRE is significantly faster than CG while leading to basically the same final structures. A speed-up of the calculations by a factor of 3-6 was typically achieved by using FIRE instead of CG. FIRE scales well with the system size, calculations

with up to 38 million atoms have been performed without problems.

One important advantage of FIRE is that it is only based on gradient information. Namely, the standard conjugate gradient algorithm relies on the direct minimization of the potential energy, and numerical instabilities, especially in *ab-initio* methods, may cause inaccurate determination of the local minimum. Achieving very small gradient norms can be especially important for subsequent normal-mode analysis [4] or for the study of mechanical stability of systems under load. Furthermore, the capability to work with very small forces enables FIRE to move through very flat valleys or saddle point configurations, where energy based methods experience problems. For example, the quenching of fenretinide in Fig. 2 even to $F_{RMS} = 10^{-3}$ eV/Å, was not possible with given CG implementation [2] simply because the energy variations during one line search were similar to the numerical accuracy ($\sim 10^{-12}$ eV) of the total energy.

Since FIRE requires only the first derivatives of the target function, it can easily be adapted to various other minimization problems. By now, FIRE has been applied successfully also to a handful of other general multidimensional minimization problems, where preliminary results show it to be much more efficient than previously used common algorithms [22]. Furthermore, even constrained minimization can be performed by using the standard constrained MD methods (e.g. by setting $\alpha = 0$).

In conclusion we have presented an extremely useful method for the relaxation of atomic structures. Compared to other relaxation algorithms FIRE is embarrassingly simple (around 10 additional lines of code to any MD implementation), has practically no computational overhead and very low memory requirements. Tests on different systems show that for large scale simulations the method is usually significantly faster than the standard implementations of the commonly used conjugate gradient methods. It can furthermore compete with BFGS, which due to its memory requirements is mostly used only for small systems. The robustness, simplicity and lack of many adjustable parameters recommend FIRE as a versatile alternative to non-inertial relaxation methods.

TABLE I: Number of function calls required by FIRE and CG to reach convergence for the relaxation of different test systems. The used criteria were $F_{RMS} \leq 10^{-3}$ eV/Å (10^{-6} eV/Å) for all systems except for the large ones with a vacancy, were the maximum force component on an atom had to be used: $\mathbf{f}_{i_{x,y,z}} \leq 10^{-3}$ eV/Å (10^{-5} eV/Å).

system	N	FIRE	CG
AlNiCo	3360	136 (639)	661 (2131)
crack in Ni	4815	61 (207)	174 (764)
hot Cu plate	16200	299 (585)	545 (1767)
vacancy in Cu	107998	43 (132)	58 (329)
vacancy in Cu	1492991	43 (118)	59 (358)

The research is supported by the Fraunhofer MAVO for Multiscale Materials Modeling (MMM). One of us (PK) acknowledges the Academy of Finland for financial support.

[1] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer Series in Operations Research (Springer, New York, 1999).
[2] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1997).
[3] A. R. Leach, *Molecular Modelling: principles and applications* (Prentice Hall, Harrlow , UK, 2001), 2nd ed.
[4] T. Schlick, *Molecular modeling and simulation: an interdisciplinary guide*, vol. 21 of *Interdisciplinary Applied Mathematics* (Springer-Verlag, New York, 2002).
[5] J. R. Beeler, *Radiation Effects Computer Experiments* (North-Holland, Amsterdam, 1983), p. 27.
[6] H. Jónsson, G. Mills, and W. Jacobsen, *Classical and Quantum Dynamics in Condensed Phase Simulations* (World Scientific, 1998), p. 385.
[7] J. H. A. Hagelaar, E. Bitzek, C. F. J. Flipse, and P. Gumbsch, Phys. Rev. B **73**, 045425 (2006).
[8] S. A. Trygubenko and D. J. Wales, J. Chem. Phys. **120**, 2082 (2004).
[9] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).
[10] I. Rata, A. A. Shvartsburg, M. Horoi, T. Frauenheim, K. W. M. Siu, and K. A. Jackson, Phys. Rev. Lett. **85**, 546 (2000).
[11] K. A. Jackson, M. Horoi, I. Chaudhuri, T. Frauenheim, and A. A. Shvartsburg, Phys. Rev. Lett. **200493**, 013401 (2004).
[12] D. C. Liu and J. Nocedal, Math. Prog. **45**, 503 (1989).
[13] K. Nordlund, P. Partyka, R. S. Averback, I. K. Robinson, and P. Ehrhart, J. appl. Phys. **88**, 2278 (2000).
[14] P. Koskinen, H. Häkkinen, G. Seifert, S. Sanna, T. Frauenheim, and M. Moseler, NJP **8**, 9 (2006).
[15] S. Hocker, F. Gähler, and P. Brommer, Phil. Mag. **86**, 1051 (2006).
[16] P. Brommer and F. Gähler, Phil. Mag. **86**, 753 (2006).
[17] Y. Mishin, M. J. Mehl, D. A. Papaconstantopoulos, A. F. Voter, and J. D. Kress, Phys. Rev. B **63**, 224106 (2001).
[18] P. Gumbsch, J. Mater. Res. **10**, 2897 (1995).
[19] L-BFGS [1, 12] circumvents the memory limitations of BFGS, however it still requires substantial computational overhead and does not always attain the performance of the original BFGS [12].
[20] More sophisticated CG from IMSL library took $n = 306$ calls.
[21] L-BFGS [12] took $n = 394$ and $n = 49$ for the relaxations of fenretinide and nanotube to $F_{RMS} < 5 \cdot 10^{-3}$ eV/Å, respectively.
[22] M.Moseler, P.Gumbsch, E.Bitzek, P.Koskinen, Method for calculating a local extremum of a multidimensional function, pending US patent application