

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,*

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,*

Universität Stuttgart, Pfaffenwaldring 57/VI, 70569 Stuttgart, Germany

⁵*Freiburg Materials Research Center, University of Freiburg,*

Stefan-Meier-Str. 21, D-79104 Freiburg, Germany

We introduce a simple local atomic structure optimization algorithm which is significantly faster than standard implementations of the conjugate gradient method and often 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 surprising efficiency and especially the robustness and versatility of the method is illustrated using a variety of test cases from nano-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) atomic structures with minimum potential energy, starting from a given initial configuration. To solve this task a variety of well established optimization methods, like steepest descent, conjugate gradient (CG), Newton-Raphson, quasi-Newton or truncated-Newton methods are available [1–4]. The current state-of-the-art methods are mostly based on some approximate representation for the Hessian matrix to determine line search directions. Also variants of molecular dynamics (MD) methods which systematically remove kinetic energy from the system are commonly applied for minimization purposes [5–7]. Such local ‘quenching’ is important also in many global minimization algorithms [8, 9]. Interestingly, relaxation methods based on MD has been thought to be good for practical realization, but not very competitive with the afore mentioned sophisticated algorithms, and for this reason they have often been introduced as by-products of secondary importance in regular articles [5–7], not receiving the attention they deserve.

Here we introduce a simple, 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:

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

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. We recommend as strategy that the skier introduces acceleration in a direction that is ‘steeper’ than the current direction of motion

via the function $\gamma(t)$ if the power $P(t) = \mathbf{F}(t) \cdot \mathbf{v}(t)$ is positive, and in order to avoid uphill motion he simply stops as soon as the power becomes negative. On the other hand, $\gamma(t)$ should not be too large, because the current velocities carry information about the reasonable ‘average’ descent direction and energy scale.

We show in this Letter that eq. (1) brings the skier surprisingly fast to the desired destination. A discretized version of eq. (1) in combination with an adaptive time step results in a minimization scheme for multidimensional functions $E(x_1, \dots, x_M)$ which is competitive in speed with the above mentioned sophisticated optimizers, but has also other important features as we shall demonstrate. Contrary to the conventional schemes the new algorithm relies on inertia and, consequently, this novel method was dubbed *FIRE* for *Fast Inertial Relaxation Engine*. Fig. 1 shows that FIRE easily keeps up with powerful standard schemes like CG and the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) scheme [10] in a two-dimensional spiral potential. In calculations on a broad range of realistic test systems the new algorithm was surprisingly fast and could be used with great ease for systems with millions of degrees of freedom.

The numerical treatment of the algorithm is simple. Any common MD integrator can be used as the basis for the propagation due to the conservative forces. The MD trajectory is continuously re-adjusted by two kinds of velocity modifications: a) the above mentioned immediate stop upon uphill motion and b) a simple mixing of the global ($3N_{atoms}$ dimensional) velocity and force vectors $\mathbf{v} \rightarrow (1 - \alpha)\mathbf{v} + \alpha\hat{\mathbf{F}}|\mathbf{v}|$ resulting from an Euler-discretization of the last term in eq. (1) with time step Δt and $\alpha = \gamma\Delta t$. Both Δt and α are treated as dynamically adaptive quantities.

Explicitly, the FIRE algorithm uses the following prop-

agation rules (given initial values for Δt , $\alpha = \alpha_{start}$ and for the global vectors \mathbf{x} 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 $\mathbf{v} \rightarrow (1 - \alpha) \cdot \mathbf{v} + \alpha \cdot \hat{\mathbf{F}} \cdot |\mathbf{v}|$

F3: if $P > 0$ and the number of steps since P was negative is larger than N_{min} , increase the time step $\Delta t \rightarrow \min(\Delta t \cdot f_{inc}, \Delta t_{max})$ and decrease $\alpha \rightarrow \alpha \cdot f_{\alpha}$

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

F5: return to MD

In relaxation an accurate calculation of the atomic trajectories is not necessary, and 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 ($\mathbf{v} \rightarrow 0$) and the time step is substantially reduced in order to have a smooth restart. A short ‘latency’ time of N_{min} MD steps before accelerating the dynamics is important for the stability of the algorithm.

Most of the parameters introduced above are not sensitive to different systems. 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 $f_{\alpha} = 0.99$ [20]. Thus the only adjustable parameter of the method is the maximum time step Δt_{max} . From a typical MD simulation time step Δt_{MD} one can obtain an initial rough estimate of $\Delta t_{max} \sim 10 \cdot \Delta t_{MD}$.

Special attention needs to be paid to the global nature of the algorithm, which assumes that *all degrees of freedom are comparable*. All the velocities should be on the same scale, which for hetero-nuclear systems can be roughly achieved by setting all the atom masses equal.

To demonstrate the performance of FIRE, we compare it to two relaxation methods: the Polak-Ribière version of CG with the popular *Numerical Recipes* implementation [2], and the limited memory version of BFGS (L-BFGS) [1, 10]. These are widely used methods in large systems where storage and arithmetic costs are an issue [21]. Although there are more specialized implementations available for these methods [4, 11], the popularity and good documentations make these algorithms ideal reference methods [22]. 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 in most cases. The number of ‘function calls’ is a generic notation for the number of separate points \mathbf{x} where either energy, force, or both are evaluated.

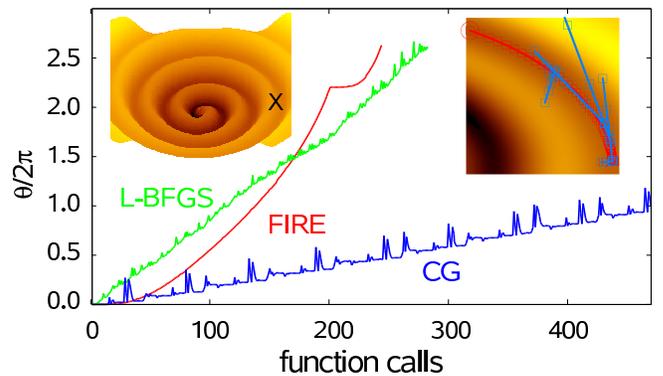


FIG. 1: Optimization of a spiral-shaped potential energy function (see left inset, X is the starting point). Shown is the evolution of the azimuthal angle θ versus the number of function calls of FIRE, CG and L-BFGS. FIRE is slower at the beginning, but catches up quickly with L-BFGS as the curvature increases, with CG not converging within 500 function calls due to inefficient line searches as displayed in the right inset showing a part of the trajectory of FIRE (red) and CG (blue).

As a first demonstration, Fig. 1 shows FIRE, CG and L-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 (masses = 1 and $\Delta t_{max} = 0.3$) with its smooth down-hill trajectory reaches the minimum ($E < -0.99$) first, followed soon by L-BFGS. However, according to wall clock FIRE is almost three times faster than L-BFGS due to small computational overhead. CG is slow for the curved, non-harmonic function due to inefficient line searches as shown in Fig. 1.

The first real test system is the biomolecule fenretinide, which is used as a cancer drug (see upper inset in Fig. 2(a)). The atomic interaction was modeled using density-functional based tight-binding (DFTB) method (see ref. [12] and references therein). The starting configuration was created by twisting the carbon chain along the chain axis. This is a challenging set-up since the unwinding is done by the torsional force of single carbon-carbon bonds. In FIRE all the masses were set to one atomic mass unit and $\Delta t_{max} = 1$ fs.

Fig. 2(a) shows that in terms of F_{RMS} FIRE is always ahead of L-BFGS even though the energies go nearly parallel. CG is much slower and shows almost no decrease either in F_{RMS} or energy beyond a certain point. Analysis of the relaxation trajectory shows that especially for CG the straightening process is crowded with inefficient line search directions, indicating the *high curvature* of the minimization pathway. It is especially interesting to note that FIRE becomes overall faster relative to both CG and L-BFGS when the initial twist angle is increased, result-

ing in a larger distance from the minimum and larger anharmonicity of the energy landscape.

The above relaxation was repeated using SCC-DFTB [12], which requires self-consistent charge distribution and often pertains intrinsic (random) errors in the potential energy due to imperfect convergence of the electronic degrees of freedom. Both CG and L-BFGS optimizations did not converge due to their sensitivity to errors in energy. Even though forces are also erroneous, FIRE was able to optimize the structure because the inertia manages to *smoothen out the errors*. This is an important feature, since this situation is frequent in ab initio calculations.

DFTB was also applied for quenching of a metallic Na_{71}^- nanocluster, where the initial structure was a sphere cut from a perfect lattice. Fig. 2(b) shows that FIRE first relaxes the symmetric structure but is then able to break the symmetry of the cluster and finally finds a new low-energy structure. L-BFGS and CG relax the symmetric structure but are not able to break the symmetry of the cluster, which requires to go through a shallow valley with very small forces which an energy-based optimization method does not achieve.

One larger test system is shown in Fig. 2(c), which is an approximant to a decagonal AlNiCo quasicrystal [13, 14] with 3360 atoms in periodic boundary conditions (PBC) and a fixed box size. The atomic interaction was modeled with embedded atom method potential [14, 15]. Although this is a seemingly ideal problem for the classical algorithms since no large conformation changes occur, FIRE is surprisingly competitive to L-BFGS and roughly three times faster than CG.

Further performance tests were conducted on a broad range of different systems such as the relaxation of a crack and a vacancy, and the quenching of a hot copper film. The crack system is set up according to the anisotropic linear elastic solution for a sharp crack at the Griffith load in Ni (for details see [16]). It requires the relaxation of large strains at the crack tip in their interaction with the long range strain field of the crack. Similarly the vacancy system requires the relaxation of very large forces around the vacancy and the adjustment of the long range stress field. Finally, the quenching of a free-standing thin copper film from 1000 K temperature combines the relaxation of the long range thermal expansion with the relaxation of local displacements. The results for these systems are compiled in Tab. I. In performance FIRE is in all cases between CG and L-BFGS, being faster than CG by a factor of 3-6 and becoming more efficient especially for small convergence criteria.

The determination of saddle points, transition states or critical points constitutes another class of typical relaxation problems. As a well documented example [17], we have chosen the determination of the Peierls stress τ_P of an edge dislocation in Al ($2.0 \text{ MPa} < \tau_P < 2.2 \text{ MPa}$). Details of the set-up and the boundary conditions are

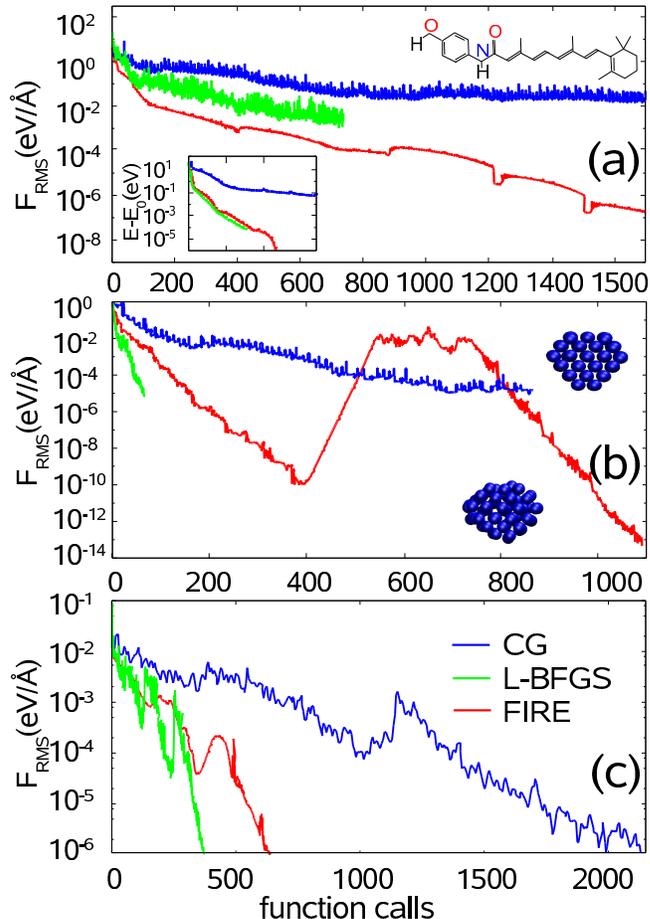


FIG. 2: a) Relaxation of fenretinide (Lewis structure shown in the upper inset) modeled with density-functional based tight-binding. The force norm as a function of the number of function evaluations is shown for FIRE, CG and L-BFGS (The color coding is shown in c). The lower inset shows the evolution of the total energy E above the equilibrium value E_0 . b) Relaxation of metallic Na_{71}^- nanocluster. CG and L-BFGS remain in the symmetric structure (upper inset), while FIRE finds a lower energy symmetry-broken structure (lower inset). c) Relaxation of the AlNiCo quasicrystal.

given in [18]. Starting from a fully relaxed stable structure at $\tau = 1.8 \text{ Mpa}$ the systems were loaded to $\tau =$

TABLE I: Number of function calls required by FIRE, CG and L-BFGS to reach convergence for the relaxation of different test systems. The used criteria were $F_{RMS} \leq 10^{-3} \text{ eV/\AA}$ (10^{-6} eV/\AA) (for vacancy also the maximum force component was used: $f_i \leq 10^{-3} \text{ eV/\AA}$ (10^{-5} eV/\AA)).

system	N_{atoms}	FIRE	CG	L-BFGS
AlNiCo	3360	136 (639)	661 (2131)	98 (350)
crack in Ni	4815	61 (207)	174 (764)	20 (118)
hot Cu plate	16200	299 (585)	545 (1767)	61 (217)
vacancy in Cu	107998	43 (132)	58 (329)	9 (55)
vacancy in Cu	1492991	43 (118)	59 (358)	11 (-)

2.0 MPa and 2.2 MPa. FIRE performed as expected. Relaxation at 2.0 MPa led to a stable configuration at the initial position. At 2.2 MPa, the dislocation started to move in the pseudo-dynamics of FIRE. In contrast, L-BFGS provided no indication for instability. The dislocation moved for a significant distance ($\sim 10 \text{ \AA}$) at both stresses until the optimization stopped since no further minimization was possible. The direct applicability of FIRE to such critical point analysis, which apparently is not possible with L-BFGS and other energy minimizing algorithms is due to its strict adherence to minimizing forces. This has to be regarded as a big advantage of the algorithm and indication of its wide applicability.

Apart from the performance, FIRE thus incorporates features making it a very versatile optimizer. As already mentioned, the method is stable with respect to random errors in the potential energy, it is well suited for problems near critical points and it can be used with very small convergence criteria with the force scale decreasing nearly exponentially. In addition to that, compared to other relaxation algorithms, FIRE is extremely simple with around 10 additional lines of code to any MD implementation. Moreover, it has nearly no computational overhead, has very low memory requirements and scales well with the system size; calculations with up to 38 million atoms have been performed without problems. FIRE is also well suited for other optimization tasks, such as constrained minimization (using the existing standard constrained MD methods e.g. by setting $\alpha = 0$), parallel computing, and transition state calculations such as the nudged elastic band method, where MD-type methods are often used [19]. Since the initial guess for the maximum time step δt_{max} is easy to make, and usually one makes several optimizations for similar systems, the parametrization does not require more attention than in other methods.

Finally, 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 often more efficient than previously used common algorithms [23].

In conclusion we have presented an extremely simple, universally applicable robust new algorithm for the relaxation of atomic structures. Tests on different systems show that for large scale simulations the method is significantly faster than commonly used CG and competes even with L-BFGS. FIRE can be generally recommended as a versatile alternative to non-inertial atomistic relaxation methods. It is also ideally suited for the study of the mechanical stability of systems for the determination of transition states, where competing methods often fail.

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

* Both authors contributed equally to this work.

† Electronic address: Michael.Moseler@iwmm.fraunhofer.de

- [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, Harlow, 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] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).
- [9] I. Rata, A. A. Shvartsburg, M. Horoi, T. Frauenheim, K. W. M. Siu, and K. A. Jackson, Phys. Rev. Lett. **85**, 546 (2000).
- [10] D. C. Liu and J. Nocedal, Math. Prog. **45**, 503 (1989).
- [11] K. Nordlund, P. Partyka, R. S. Averback, I. K. Robinson, and P. Ehrhart, J. appl. Phys. **88**, 2278 (2000).
- [12] P. Koskinen, H. Häkkinen, G. Seifert, S. Sanna, T. Frauenheim, and M. Moseler, NJP **8**, 9 (2006).
- [13] S. Hocker, F. Gähler, and P. Brommer, Phil. Mag. **86**, 1051 (2006).
- [14] P. Brommer and F. Gähler, Phil. Mag. **86**, 753 (2006).
- [15] Y. Mishin, M. J. Mehl, D. A. Papaconstantopoulos, A. F. Voter, and J. D. Kress, Phys. Rev. B **63**, 224106 (2001).
- [16] P. Gumbsch, J. Mater. Res. **10**, 2897 (1995).
- [17] D. L. Olmsted, K. Y. Hardikar, and R. Phillips, **9**, 215 (2001).
- [18] E. Bitzek and P. Gumbsch, Mat. Sci. Eng. A **400-401**, 40 (2005).
- [19] G. Henkelman and H. Jonsson, J. Chem. Phys. **113**, 9978 (2000).
- [20] Qualitatively they should satisfy: N_{min} larger than one (at least a few smooth steps after freezing); f_{inc} larger than but near to one (avoid too fast acceleration); f_{dec} smaller than one but much larger than zero (avoid too heavy slowing down), α_{start} larger than, but near to zero (avoid too heavy damping); f_{α} smaller than, but near to one (mixing is efficient some time after restart).
- [21] Memory requirements are roughly $3M$, $4M$ and $14M$, for FIRE, CG and L-BFGS(default), respectively, where M is the number of variables.
- [22] We tested also BFGS from the IMSL library, but for the test systems the overall performance did not differ much from L-BFGS.
- [23] M. Moseler, P. Gumbsch, E. Bitzek, P. Koskinen, Method for calculating a local extremum of a multidimensional function, pending US patent application.