

The activation–relaxation technique: an efficient algorithm for sampling energy landscapes

G.T. Barkema^{a,*}, Normand Mousseau^b

^a *Theoretical Physics, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands*

^b *Department of Physics and Astronomy, Ohio University Athens, OH 45701, USA*

Received 15 February 2000

Abstract

Activated processes, i.e., rare events requiring thermal fluctuations many times larger than the average thermal energy, play a central role in controlling the relaxation and diffusion mechanisms of disordered materials such as amorphous and glassy solids, polymers and bio-molecules. As the time scales involved are much longer than those associated with thermal vibrations, these processes cannot be studied efficiently with standard real-space methods such as molecular dynamics (MD). They can be investigated much more efficiently by working in the potential energy space. Instead of defining moves in terms of atomic displacements, the activation–relaxation technique (ART) follows paths directly in the energy landscape, from local minima to adjacent saddle points, giving full freedom for the system to create events of any complexity. In this paper, we review the technique in detail and present some recent applications to amorphous semiconductors and glasses. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

At low enough temperatures, the dynamics of most systems is controlled by activated processes. These processes are described by the need of crossing energy barriers much larger than the average thermal energy of the system. Because of Arrhenius' law, this means that the typical time scale τ_A associated with these processes grows exponentially as a function of activation energy E_A , $\tau_A \sim \exp(E_A/k_B T)$. Even at room temperature, this time scale can often reach seconds and more, well beyond the picoseconds regime associated with atomic vibrations.

The microscopic origin of these mechanisms is difficult to study experimentally because of the need for fairly precise local probes that are not always available. Numerically, the time scale involved is simply prohibitive for standard simulation methods such as molecular dynamics (MD) and real-space Monte Carlo (MC). In the former case, solving the atomistic equations of motion implies integration time steps of the order of a fraction of a picosecond, which effectively limits the total simulation time to at best a tenth of a microsecond, much too short for many activated processes of interest. MC algorithms, that do not suffer inherently from the same restrictions, are clearly advantaged here. Real-space MC algorithms, which define atomic or specific localized moves, are however unable to produce efficiently the large collective motions associated with

* Corresponding author.

E-mail addresses: barkema@phys.uu.nl (G.T. Barkema), mousseau@helios.phy.ohiou.edu (N. Mousseau).

realistic activated processes in disordered materials. The sequence of small moves required to bring a configuration to a local saddle point is just too unlikely, in terms of a sequence of independent moves, to occur frequently. Real-space MC is therefore unable to reconstruct the rare events dominating long-time dynamics.

MC moves, however, are not restricted to real space. Activated processes are much better described on the configurational energy surface, where they correspond to minimum-energy trajectories connecting two adjacent minima. Using this definition, the activation–relaxation technique (ART) produces events directly in the configurational energy landscape, avoiding the difficult description of complex real-space atomic moves which occur in parallel. The ART [1,2] is essentially a two-step process: a configuration is first brought from a local minimum to an adjacent saddle point (activation); it is then relaxed from this saddle point to a new minimum (relaxation). These moves occur on the $3N$ -dimensional energy surface, allowing events of any complexity.

In this paper, we first present a general discussion of moves in the energy landscapes. We then detail the current implementation of the algorithm and present applications of ART to amorphous silicon, amorphous gallium arsenide and silica glass.

2. Energy landscapes

The dynamics of most systems, below the melting point, consists of long periods of oscillation around a given minimum with rare jumps from one basin of attraction to another. As mentioned above, the rate at which these events occur is to a first approximation controlled by the height of the activation barrier, the highest point of the trajectory followed by the configuration. It is then possible to reduce this process to a sequence of minimum-saddle point-minimum trajectories, neglecting safely the vibrational part of the dynamics. This effectively replaces the continuous energy landscape by a discrete network of minima, connected via tra-

jectories going through adjacent saddle points, and provides the minimal description required for any realistic dynamics. The entropic part of the potential, which determines the pre-factor in the transition-state theory, is here completely neglected. In principle, these contributions could also be incorporated in the algorithm, although the additional theoretical and numerical efforts required are large and have not been considered yet. Moreover, although possibly affecting the sequence of events, the discrete approximation discussed above should be sufficient to describe the structure and the nature of activated relaxation and diffusion.

In crystalline materials, it is often possible to construct the discrete network using symmetry arguments. For example, the trajectory of point defects in c-Si such as simple vacancies can easily be reconstructed from the symmetries of the diamond structure. For higher precision, one can refine the path using, say, *ab initio* methods. This approach breaks down rapidly with the increased complexity of multicomponent alloys and anisotropic crystals; even taking into account the symmetries of the problem, it becomes very difficult to identify all the possible diffusion and relaxation mechanisms. The task is completely hopeless in the case of disordered materials, where each atom sees a different environment; direct sampling using, for example, the ART, is the only realistic approach.

Working directly on the potential energy surface, ART represents a change of paradigm over standard real-space algorithms. It is almost insensitive to the complexity of the moves and the height of the barriers, and can therefore provide an almost exhaustive list of realistic mechanisms. In ART, the atomistic details of the collective process are not used in the input of events but appear rather as the output of the simulation. As the configuration, as a whole, is pushed towards a saddle point, real-space rearrangements are made in response to the topography of the problem; real-space moves are not defined *a priori* but constructed by the system itself. This is particularly appreciated in systems for which a complete enumeration of the realistic microscopic mechanisms is not available.

3. Algorithm and implementation

Although almost symmetric, the two steps of the ART – the activation to a nearby saddle point and the relaxation to a new minimum – are formally very different. The relaxation to a local minimum is a rather straightforward process. Many algorithms have been proposed over the years to solve this local optimization problem. Some special care has to be taken here because the minimization is always started near a saddle point. Because the force is almost zero, most unbounded algorithms will tend to make very large jumps, risking to bring the configuration into a different basin of attraction. It is therefore recommended to start any relaxation process with a few bounded steepest-descent steps before starting any higher-level optimization procedure. For example, our implementation of the conjugate-gradient algorithm imposes a maximal allowed displacement at each line of minimization. If this displacement is reached, the configuration is considered to be still far away from the minimum, and we erase the memory of previous conjugate directions.

If the minimization procedure is rather standard, this is not the case for the search for saddle points, which forms really the core of the method. There is no optimization technique available for converging to a nearby saddle point. Formally, the only way to find all saddle points around a minimum is to map the whole potential energy surface around that minimum. This “solution” is clearly unacceptable for systems of more than a few dimensions, where the amount of information required to triangulate a surface far exceeds the capacities of the best computers available. One, therefore, has to be satisfied with approximate methods.

In the case where both the initial and final minima are known, a number of algorithms exist that search for the best connecting paths. This type of problem is central for the identification of reactivity paths in chemical physics [3,4]. Most of the methods developed there, however, are not easily applicable to situations where only one local minimum is known.

Another solution might be to perform a random search around a minimum in the hope of

identifying saddle points. Again, this approach, coupled with highly damped Newtonian motion, had some success in low-dimensional systems but fails for systems with more than a few tens of degrees of freedom. As the phase space dimensionality increases, the saddle points occupy a rapidly decreasing fraction of it.

In order to find a satisfactory algorithm, it is necessary to look at the structure of lowest-energy paths connecting saddle points and local minima. These are steepest descent trajectories following the bottom of a valley from a saddle point down to the minimum. It is of course possible to reverse the path, and follow a *steepest ascent* from the minimum to the saddle point. This trajectory is however highly unstable and any small deviation from it will send the configuration to a local maximum. A similar instability is seen with Newtonian trajectories, although here local maxima are avoided because of energy conservation.

It is therefore important to stabilize the walk at the bottom of the valley. Analyzing in detail the best algorithms for constraining this walk, Cerjan and Miller [5] proposed the vector-following method. This algorithm uses local knowledge of the force and the Hessian matrix to constrain the path, by minimizing the energy in all eigendirections but the softest one. Although this method has a high probability of finding saddle points, it requires the diagonalization and inversion of the Hessian matrix, an $O(N^3)$ computation, and is therefore limited to fairly small systems of at most a hundred degrees of freedom. This is a major limitation, especially if one is interested in bulk or biological systems, where the number of atoms can easily reach a thousand and more. It is therefore necessary to find an algorithm scaling as much as possible linearly with the number of degrees of freedom which nevertheless can follow the minimum-energy path as closely as possible.

We do so by introducing a modified force vector $\vec{G}$ which is obtained by inverting the component of the potential force *parallel* to the displacement from the current local minimum while keeping the $3N - 1$ other components unchanged:

$$\vec{G} = \vec{F} - (1 + \alpha)(\vec{F} \cdot \hat{r})\hat{r}, \quad (1)$$

where $\vec{F}$ is the $3N$ -dimensional force calculated from the interaction potential, $\hat{r}$ a unit vector pointing back to the position of the current local minimum, and α is a control parameter usually set to $0.15/r$.

Following the modified force keeps the configuration at a minimum along all directions but one, while pushing upward in the direction perpendicular to the displacement from the minimum. If the trajectory to the saddle point were in straight line, the above expression would be exactly following the bottom of the valley. This is generally not the case and the path oscillates around the ideal trajectory. Because of this approximation, the iterative process brings the configuration in the vicinity of the saddle point but not exactly on it. Forces do not vanish and we need another criterion for identifying when we are close to the saddle point: the iteration stops when the force parallel to the displacement, $\vec{F} \cdot \hat{r}$, changes sign, indicating that the configuration has just passed the saddle point and gone into a new basin of attraction.

As the force vectors span the whole space, the search for saddle points does not impose a predefined type of real-space deformation; in contrast to real-space methods, it can involve any number of atoms, from 1 to 100, crossing barriers of many eV, without any additional numerical cost. We note also that the degrees of freedom do not have to be limited to atomic coordinates but can involve a volume coordinate and other responses to external constraints.

In the close vicinity of the local minimum, the potential energy landscape is harmonic. Applying the modified force from anywhere in the basin sends the configuration in one of the two softest directions, associated to the lowest eigenmode. These directions correspond generally to some breathing-like modes, with large displacement and high energy barriers and are, therefore, of relatively little interest. In order to reach a wider range of saddle points, it is necessary to first leave the harmonic basin before following the modified force.

One could obviously make a large random displacement from the minimum and hope to be outside that region. In very compact environments (such as metallic glasses) or fairly elastic ones (like silica glasses), however, it is very difficult to escape

the harmonic basin by simply jumping in a random direction since if the displacement is not orthogonal to the softest mode, it will align with this mode exponentially rapidly. We therefore tend to get better results by eliminating the softest directions. This is achieved by a series of small moves following the opposite direction of the force. Such a procedure is essentially equivalent to applying the Hessian matrix many times to some initial vector, projecting out the lowest energy spectrum of the Hessian matrix.

We consider that the harmonic basin has been left once the size of the force pointing back to the minimum as compared to the total force perpendicular to that direction is of the same order of magnitude. The exact numerical value for their ratio depends on each problem and needs to be chosen large enough that the harmonic basin is left behind, but not so large that a ridge could be missed, leading the configuration directly in a new harmonic basin.

4. Applications

ART has been applied to a wide variety of problems. They can be classified in two broad categories: (1) structural optimization and (2) identification and classification of diffusion and relaxation mechanisms. The first category is concerned with producing low-energy structures. In this case, the procedure used is irrelevant, only the final product matters. In the second category, the processes themselves are at the center of interest. This requires more care in the selection of the interactions and details of simulation.

4.1. Ideal amorphous state for the Stillinger–Weber potential

The applicability of the Stillinger–Weber (SW) potential [6] to the problem of amorphous silicon has long been questioned. Because of the short time scale associated with MD, however, it had not been possible to identify clearly the ideal amorphous state associated with this potential. With ART, it is possible to circumvent the time limitations and provide a clear answer to this

question [1]. This is achieved following a three-step procedure:

1. Using the standard SW potential and ART, we relax a randomly packed configuration of atoms. The radial distribution function (RDF) of this 1000-atom configuration is plotted as curve (b) in Fig. 1. It can be compared with that of a Wooten–Winer–Weaire model (WWW) [7], which yields excellent agreement with experiment. We see that the second-neighbor peak has a liquid-like structure on its left side.

2. We relax another randomly packed configuration of atoms with ART, this time, however, using a modified version of the SW potential which leads to a good agreement with the WWW RDF. This is achieved by increasing the three-body term of the SW potential by 50% (curve (a) in Fig. 1).

3. The configuration prepared in 2 is relaxed again with ART using the *standard* SW potential.

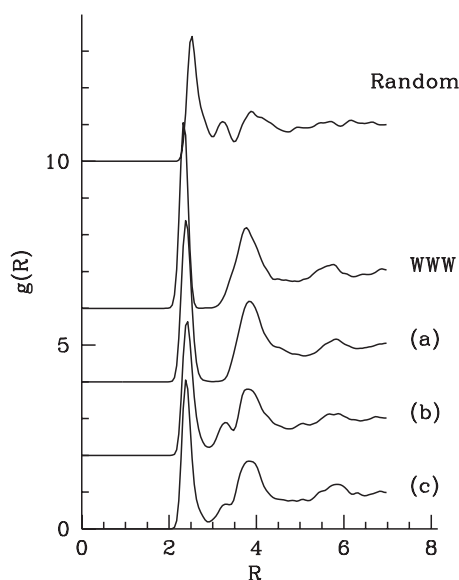


Fig. 1. Radial distribution functions (RDFs) for amorphous silicon produced by various means. The top two curves are shown for comparison with the others, and show the RDF obtained after only a local energy minimization under the SW potential, and the RDF obtained with the WWW method. The next curves are obtained as follows: (a) a-Si obtained with ART using a modified SW potential and starting from a randomly packed configuration; (b) a-Si obtained with ART using a standard SW potential and starting from a randomly packed configuration; (c) a-Si obtained with ART using a standard SW potential but starting from configuration (a).

As shown in Fig. 1 (curve (c)), the configuration evolves to a state similar to that of curve (b). We conclude from this that the ideal disordered configuration associated with the standard SW potential does not correspond to a-Si but rather to a mixture of the amorphous and liquid states.

4.2. Chemical ordering in amorphous GaAs

The microscopic details of the structure of amorphous binary semiconductors are rather difficult to establish. This is particularly so for GaAs where the small size difference between the two species prohibits the identification of the nearest-neighbors with local probes such as EXAFS. It is therefore not possible to establish whether or not binaries have the same topology as elemental amorphous semiconductors. In weakly ionic materials, the chemical cost associated with creating homopolar (wrong) bonds is in competition with the elastic strain resulting from the removal of odd-membered rings and could therefore lead to some frustrated network.

For simplicity, continuous random networks without a constraint on the ring distribution are referred to as Polk networks [8] while those with only even-membered rings, such as fully chemically ordered amorphous materials, are called Connell–Temkin networks [9].

As the weak ionicity of Ga and As imposes long relaxation times to reach chemical equilibrium, MD cannot be used directly to probe local ordering. The ART again, is ideal to address this problem because it is not as sensitive as MD to the slowing down of the dynamics.

As there exists no satisfactory empirical potential for GaAs, we use a modified SW potential along with an additional repulsive potential describing the homopolar energetical costs:

$$V_{\text{rep}} = \sum_{\langle ij \rangle} \epsilon A_{ij} \left[1 + \cos \left(\pi \frac{r_{ij}}{s_{ij}} \right) \right] \quad (2)$$

with $A_{ij} = 1.2$ for like pairs and zero otherwise, ϵ is the energy scale of the SW potential, and $s_{ij} = 3.6$ Å, the cut-off of this function.

Two networks are created, one using the modified SW potential discussed above, and leading to

a Polk-type network, and a second using this additional repulsive potential and leading to a Connell–Temkin-type CRN. In order to allow meaningful comparison, both networks are further relaxed using realistic tight-binding interactions for Si [10] and GaAs [11], leading to four different final configurations. Overall, the structural properties of these samples are excellent. The TB-relaxed GaAs Connell–Temkin model yields an average coordination of 3.95, with a single 5-fold coordinated atom and 3.9% wrong (homopolar) bonds. The corresponding TB-relaxed Si Polk network, has coordination of 3.98, with three-5-fold coordinated atoms and (optimally decorated with Ga and As) 14.2 % wrong bonds [12].

We compare the energies of these four networks in Table 1 and extract the following conclusions: (1) For elemental networks there appears to be no additional strain associated with the constraint to only allow even-membered rings. Polk’s predominance is therefore only due to entropic contributions. (2) GaAs clearly prefers a Connell–Temkin-like network and its topology is therefore qualitatively different from that of elemental amorphous semiconductors. (3) Although not discussed here, Ref. [14] also shows that it is almost impossible to measure indirectly the proportion of wrong bonds in a-GaAs.

4.3. Relaxation mechanisms in a-Si and g-SiO₂

ART events are created following closely the physical activation paths; this allowed for the first time to identify and classify microscopic relaxation and diffusion mechanisms in disordered materials. We have recently generated more than 8000 events

from three independent runs on 1000-atom samples of a-Si [15,16]. We find a continuous distribution for activation barriers, extending from 0 to 15 eV, with a peak at 4 eV. These events involve, on average, the displacement of about 40 atoms, although in most cases, only a few bonds are broken and created. A typical event, showing only atoms rearranging their topology and their near-neighbors, is shown in Fig. 2.

Events can be classified in terms of topological rearrangements of the network between the initial and final minimum. *Perfect events* see 4-fold atoms exchanging bonds but keeping their total coordination; in *conserved events*, coordination defects diffuse around, while the overall coordination is preserved; *annealing events*, finally, involve the creation or the annihilation of defects. Each of these categories can be further broken down into topological subclasses. Fig. 2 shows the three most common types of *perfect* rearrangements. These form the basis of most events in a-Si.

A similar study in silica glass has revealed a completely different dynamics for this material [17]. In particular, because of the need to maintain chemical ordering, the perfect mechanisms of a-Si do not have a direct counterpart. Detailed analysis of these two models is currently in progress.

5. Conclusion

The activation relaxation technique provides a generic approach for exploring the potential energy surface of complex systems such as amorphous and glassy materials as well as polymers and bio-molecules. Working directly in the configurational energy space, ART does not impose any pre-defined atomic moves; real-space rearrangement of any size can take place. This makes ART a unique tool for identifying the microscopic mechanisms responsible for relaxation and diffusion in these materials.

ART has already produced significant results both as an optimization and a dynamical tool. In particular, we have been able to provide the first analysis of relaxation and diffusion mechanisms taking place below melting in a-Si and g-SiO₂. Its

Table 1
Tight-binding energy (eV/atom) of the Polk and Connell–Temkin networks when decorated with Si or GaAs^a

Network	TB parameters		
	Si	GaAs	GaAs (SL)
Polk	–13.172	–13.450	
CT	–13.163	–13.561	–13.450
Crystal	–13.389	–13.802	–13.802

^aOther results are from the TB–MD simulations of Seong and Lewis (SL), Ref. [13].

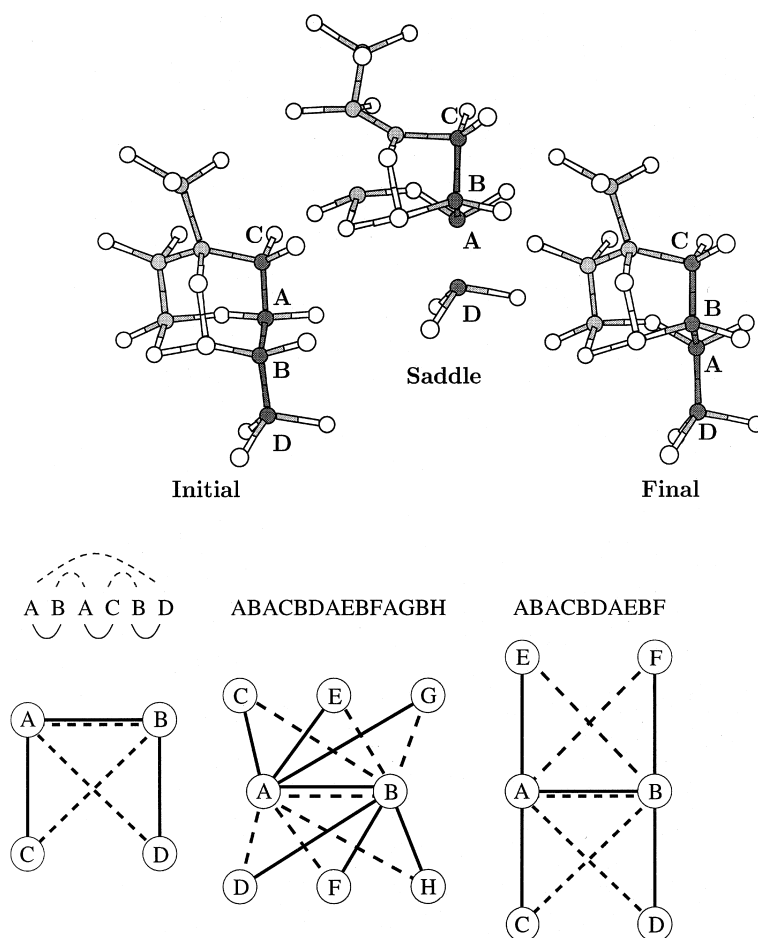


Fig. 2. Top: An ABACBD event in amorphous silicon. From left to right, the initial, saddle point, and final configurations are shown. The dark atoms are those changing coordination. The activation barrier and asymmetry for this event are 3.9 and 0.9 eV, respectively. Bottom: The three commonest types of perfect mechanisms in amorphous silicon, according to the topological classification in [15,16]; continuous lines are broken bonds, dashed lines created bonds, and double lines conserved bonds. ABACBD corresponds to the mechanism proposed by Wooten, Winer and Weaire for the construction of their Sillium model.

application to the investigation of the potential energy surfaces of other disordered systems like alloys and proteins looks promising.

Acknowledgements

Part of the work presented in this review was performed in collaboration with Laurent J. Lewis and Simon W. de Leeuw. NM wishes to acknowledge partial support by the NSF under Grant No. DMR 9805848.

References

- [1] G.T. Barkema, N. Mousseau, Phys. Rev. Lett. 77 (1996) 4358.
- [2] N. Mousseau, G.T. Barkema, Phys. Rev. E 57 (1998) 2419.
- [3] R.S. Berry, H.L. Davis, T.L. Beck, Chem. Phys. Lett. 147 (1988) 13.
- [4] V. Ivanova, E.A. Carter, J. Chem. Phys. 98 (1993) 6377.
- [5] C.J. Cerjan, W.H. Miller, J. Chem. Phys. 75 (1981) 2800.
- [6] F.H. Stillinger, T.A. Weber, Phys. Rev. B 31 (1985) 5262.
- [7] F. Wooten, K. Winer, D. Weaire, Phys. Rev. Lett. 54 (1985) 1392.
- [8] D.E. Polk, J. Non-Cryst. Solids 5 (1971) 365.
- [9] G.A.N. Connell, R.J. Temkin, Phys. Rev. B 9 (1974) 5323.

- [10] L. Goodwin, A.J. Skinner, D.G. Pettifor, *Europhys. Lett.* 9 (1989) 701.
- [11] C. Molteni, L. Colombo, L. Miglio, *J. Phys.: Condens. Matter* 6 (1994) 5243.
- [12] N. Mousseau, L.J. Lewis, *Phys. Rev. Lett.* 78 (1997) 1484.
- [13] H. Seong, L.J. Lewis, *Phys. Rev. B* 53 (1996) 4408.
- [14] N. Mousseau, L.J. Lewis, *Phys. Rev. B* 56 (1997) 9461.
- [15] G.T. Barkema, N. Mousseau, *Phys. Rev. Lett.* 81 (1998) 1865.
- [16] N. Mousseau, G.T. Barkema, *Phys. Rev. B* 61 (2000) 1898.
- [17] N. Mousseau, G.T. Barkema, S.W. de Leeuw, *J. Chem. Phys.* 112 (2000) 964.