JCTC Journal of Chemical Theory and Computation

Efficient Parallel Decomposition of Dynamical Sampling in Glass-Forming Materials Based on an "On the Fly" Definition of Metabasins

Dimitrios G. Tsalikis,[†] Nikolaos Lempesis,[†] Georgios C. Boulougouris,*[,†,‡,§,∥] and Doros N. Theodorou*[,†]

School of Chemical Engineering, National Technical University of Athens, Zografou Campus, GR-15780 Athens, Greece, Engineering Informatics and Telecommunications, University of Western Macedonia, Konstantinou Karamanli 55, GR-50100 Kozani, Greece, Department of Chemical Engineering, University of Patras, GR-26500 Patras, Greece, and Scienomics SARL, 17, Square Edouard VII, 75009 Paris

Received August 12, 2009

Abstract: In this work, we propose a highly parallelizable sampling scheme designed for atomistic simulations of glassy materials in the vicinity of the glass-transition temperature T_g, based on the idea of inherent structures (IS). Glassy dynamics is envisioned as a combination of two types of motions: (a) an "in basin" vibrational motion in the vicinity of a potential energy minimum (IS), and (b) transitions from one basin to another. In order to perform efficient dynamical sampling in the vicinity of T_g, we propose an "on the fly" definition of metabasins (i.e., collections of basins communicating via fast transitions in which the system spends a sufficient time before moving on to a neighboring collection). Our criterion for defining metabasins is based on the rate of identification of new basins in the course of a canonical molecular dynamics (MD) run. In order to compute individual rate constants between basins and metabasins, we propose to follow a swarm of microcanonical MD trajectories initiated at phase-space points sampled by a canonical MD run that is artificially trapped within a metabasin. The execution time required by this highly parallelizable scheme is reduced dramatically, since no information exchange takes place between the microcanonical trajectories. Results from our parallel methodology are compared against results from artificially trapped canonical MD runs, in terms of the evaluated rate constants, and found to be in very good agreement. Parallel simulations have been conducted on up to 250 processors, achieving almost linear scaling. The validity of our definition of metabasins is confirmed by analysis of the resulting network of basins.

Introduction

Glassy materials have assumed an important role in our life and consequently attract the interest of the scientific community in both applied and basic research. Over the years, glasses have been categorized according to various criteria: (a) based on the temperature dependence of dynamic viscosity,[1] into strong and fragile glass-forming liquids, and (b) based on the intramolecular interactions responsible for dynamical entrapment,[2] into repulsive and attractive glasses. Repulsive glasses are usually observed at high densities, where repulsive interactions become dominant, whereas attractive glasses appear due to a strong short-range attractive interaction.[3]

* Corresponding authors. E-mail: gboulougouris@uowm.gr (G.C.B.) and doros@central.ntua.gr (D.N.T.). Fax: +30 210 772 3112.

[†] National Technical University of Athens.

[‡] University of Western Macedonia.

[§] University of Patras.

[∥] Scienomics SARL.

10.1021/ct9004245
Published on Web 03/19/2010

The research work described in this article is focused on studying the dynamics of glass-forming materials. Unfortunately, the broad range of time scales for molecular motion present in glassy systems poses severe limitations for molecular simulation in the vicinity of and below the glass-transition temperature T_g. Any discrete numerical solution of the time evolution equations of a microscopic model is bound by the time step of the discretization, which has to be smaller than the characteristic time of the fastest process present, thereby limiting our ability to track the time evolution out to the desired longest time scales. Therefore, brute force MD simulations are doomed to describe only a very short part of the spectrum of time scales characterizing motion in a glassy system. In order to address this problem, we will try to elucidate the dynamics of glassy materials in terms of their potential energy landscape.

One of the most important features of the potential energy landscape is the local minima of the energy, or "inherent structures" (ISs),[4,5] around which the system is expected to spend most of its time trapped, at least at low temperatures. Throughout this article, we will use the term "basin" to denote the set of configuration space points from which a steepest descent construction in the potential energy leads to a given IS.[4] The entire multidimensional configuration space can be tessellated into basins.

As in all molecular simulation methods, the size of the system is part of the simulation conditions and the macroscopic behavior can only be obtained as a thermodynamic limit. In the inherent structure approach, the condition that will minimize system size effects is the independence of cooperatively rearranging domains (in three-dimensional space) during the individual transitions from one basin to another. Assuming that the simulation system is sufficiently large, extensivity will result in independent transitions being executed by its various parts. For example, if one considers a model system of size double that of the original simulation system, one will observe elementary transitions, each of which involves a molecular rearrangement in only one of the two halves of the augmented system. Each of these rearrangements would have been detected as a single transition in an analysis of individual configurations of the original model system. Since each of the two sets of rearrangements (one set for each half of the doubled system) involves a different set of degrees of freedom, and the two sets of degrees of freedom are practically uncoupled, the rate constants for the individual rearrangements will be the same in the doubled system, as computed for the original system. Thus, the descriptions of dynamics on the basis of the original system and of the doubled system are equivalent, provided the original system is sufficiently large, in comparison to the size of "cooperatively rearranging domains". In this respect, basins of the original system would continue being relevant to the dynamics of the doubled system as well.

A prerequisite, of course, for this picture of extensivity to hold is that both the small and the large system configurations have been sampled from a distribution that is representative of the real glass under a given formation history. If this is the case, an inherent structure of the large model system will be essentially a combination of mutually independent inherent structures of smaller subsystems, into which the large system can be spatially decomposed. Consequently, the partition function of the large system trapped in the vicinity of one of its inherent structures will be essentially a product of partition functions of the subsystems, each trapped in the vicinity of its own inherent structure. It is this factorization that leads to rate constants for subsystem rearrangements being the same as computed for the individual subsystems and for the large system. This has been properly demonstrated in the work of Doliwa and Heuer,[6] where they examined finite-size effects in the same model. In their work they conclude that "a system of $N = 130$ particles behaves basically as two noninteracting systems of half the size."

Below T_g, the importance of ISs and basins to simulations of a variety of condensed matter systems has been extensively explored.[7−21] The "inherent structure picture" has been used as a tool to investigate and characterize the dynamics and the thermodynamics of atomistic systems in terms of their "landscape." Some of the most widely used concepts are the disconnectivity graphs[8−11,17,18,21] and the configurational entropy.[22−28] Another popular approach is the numerical validation, via simulations, of the theoretical predictions of the mode coupling theory and its extensions.[29−32] For extensive reviews on the subject, we refer the reader to the work of Heuer,[33] Sciortino,[22] Debenedetti, and Stillinger.[26]

It is worth noting that disconnectivity graphs have been used to visualize the energy landscape of the same model glass-former system studied here.[17−19] These studies have been conducted from both a thermodynamic perspective via the use of parallel tempering sampling[19] and from a dynamical perspective via extensive analysis of the local connectivity and the presence of "metabasins" in terms of the cage-breaking[17] process relevant to many transitions at the atomic level.[18]

Mapping long atomistic dynamics onto a discrete network of states has been used to analyze the folding pathways in protein-folding simulations.[34−41]

Most of the attempts described above are related with analyzing simulation results in terms of the IS picture. On the other hand, there are attempts to use IS ideas as a basis for accelerating the dynamical sampling. In view of the vast size and extensive nature of the potential energy landscape, it is crucial to sample the dynamical evolution between potential energy minima in an efficient manner. This has been demonstrated to be possible via the discrete path sampling (DPS) algorithm of Wales,[11,21] which is able to evaluate the most probable path of first-order transitions between potential energy minima linking two regions of the potential energy landscape, upon the assumption of an intermediate set of potential energy minima acting as the "activated" state.

One alternative attempt is the dynamical integration over a Markovian web (DIMW) methodology, developed by Boulougouris and Theodorou[13] in their effort to simulate the dynamics of an atomistic model of glassy atactic polystyrene over more than 10 orders of magnitude on the time scale at temperatures far below T_g. DIMW is also an alternative to kinetic Monte Carlo sampling, which enables the direct evaluation of the time-dependent probability of occupancy of states (here, basins) for a system undergoing successive

transitions with first-order kinetics between basins in a landscape of infinite extent.

Although the DPS[11,21] and DIMW[13] algorithms can, in principle, be combined, their "philosophy" can be seen as complementary. DIMW creates a network of states through a breadth-first search that invokes no assumptions about ending and intermediate states, tracking a diffusion-like process in configuration space; DPS, on the other hand, may be seen as a depth-first search toward a known part of the landscape, upon the, most of the times logical, assumption of a rate-controlling intermediate region. Another important difference between the two methods is that the DIMW method aims at retrieving all dynamical information up to a certain time, that is the whole relaxation spectrum accessible to the potential energy landscape dynamics, whereas DPS, to our understanding, needs to include as target regions the ones that will potentially be relevant to a relaxation mechanism. The exhaustive dynamical information provided by DIMW may be of great importance. As discussed below, Boulougouris and Theodorou have recently shown[42] that knowledge of the spectrum of relaxation times, along with the eigenvectors of the dynamical transition matrix provided by the DIMW method, can be used in order to compute the time autocorrelation function and the dynamical relaxation spectrum of any observable; in this "EROPHILE" approach, a relaxation mode can be identified as a Euclidean vector in state probability space that decorrelates in a single exponential way. Thanks to the EROPHILE approach and the DIMW method, the creation of a complete basis set is guaranteed. DIMW can also be viewed as an extension to kinetic Monte Carlo of the integration over a Markovian web (IMW) method,[43] which allows the enrichment of ensemble averages and the combination of multiple integration levels. Based on DIMW, it is possible to construct an ever-expanding network of known or "explored" states, bounded by a set of "boundary" states, starting from an initial (small) set of states. For each explored state, all relevant transitions connecting it with its neighboring states have been located, and the corresponding transition rate constants computed by atomistic infrequent-event analysis. Boundary states are connected to explored states, but are not yet explored themselves. The time-dependent probability distribution among explored states is determined via analytical solution of the master equation for the "explored" states under absorbing boundary conditions for the "boundary" states. The set of explored states is expanded systematically whenever necessary, through a stochastic scheme that each time selects to include in the set of explored states a boundary state, according to the probability flux to that state. In their work Boulougouris and Theodorou[13] used multidimensional transition-state theory, within the harmonic approximation, in combination with a saddle-point search method[44−47] (specifically, the dimer method[46]) in order to locate and evaluate transitions and rate constants out of the states being explored on the fly. We note that the "dimer method" is actually an alternative implementation of hybrid eigenvector following, first described in ref 44.

Approaching T_g from below, the number of "relevant" minima and saddle points increases dramatically. As a consequence, the computational cost for saddle-point calculations becomes prohibitively high. In order to overcome this obstacle, we have investigated the role of ISs in the vitrification process of glass-forming materials using a simple methodology, which is based on a combination of MD and potential energy minimization and on an extension of hazard plot analysis. This approach[48] showed that the dynamical transitions between basins can be described by a first-order kinetic scheme. More precisely, it was shown[49] that it is possible to reconstruct completely the dynamics of the atomistic system at a finite temperature, below T_g, based on the first-order kinetic network of interbasin transitions. This reconstruction corresponds to a "lifting"[50] of the coarse-grained Poisson process model of a succession of interbasin transitions to the detailed atomistic level. The excellent agreement obtained with full atomistic MD for temperatures around and below the glass transition temperature showed that an approach based on infrequent, uncorrelated transitions between basins is able to reproduce the full dynamics of the atomistic glassy system, where the Poisson approximation is valid.

The IS approach offers an additional advantage. The slow dynamics of the glassy systems described by the analytical solution for the time-dependent probability of occupancy of discrete states (basins around ISs) can be used for the identification of the molecular mechanisms that govern the dynamics. For this purpose, Boulougouris and Theodorou[42] developed a statistical mechanical-geometric formulation (EROPHILE) that expresses both state probabilities and all observables in the *same* Euclidean space, spanned by the eigenvectors of the symmetrized time evolution operator. EROPHILE is a general framework for computing the equilibrium and nonequilibrium behavior of systems evolving through a succession of transitions between discrete states. It provides a geometric representation of relaxation modes in a dual representation: (a) a mode corresponds to a perturbation from the equilibrium probability distribution among states that decays with time along a single exponential, and, most importantly, (b) a mode is identified with a linear combination of observables that, upon *any* perturbation from equilibrium, will return to equilibrium in a single exponentially decaying fashion. By applying EROPHILE to an atomistic model of a-PS, Boulougouris and Theodorou provided a molecular mechanism for the delta relaxation of a-PS, that of a net rotation of a single phenyl group around its stem.

As described above, well below T_g, the IS picture, coupled with infrequent event analysis, has been proven to be a reliable computational tool. Well above T_g classical MD simulation is, in most cases, an efficient strategy. The vicinity of T_g, however, still remains a complex problem. Addressing this problem is very important since, as one expects, the temperature region in the vicinity of T_g determines the quality,hence, the properties of the glass that we will obtain when we cool our glass-forming material far below T_g.[49] This happens for the simple reason that, in the temperature region far below T_g, the system practically "freezes" in the neighborhood of configurational space that it sampled just before the temperature dropped.

A necessary step for coarse graining the dynamics into the IS picture is the evaluation of rate constants for basin-to-basin transitions. This can be done with a variety of methods.[44−47,51−55] In the past we have used two distinct approaches for the rate calculations: a saddle-point search in combination with Fukui's intrinsic reaction coordinate (IRC) construction[56] and a harmonic approximation,[13] and MD simulations[48,49] in combination with hazard plot analysis.[57,58] The applicability of each approach depends on its computational demands. For temperatures far below T_g, an approach based on MD would suffer, since the system remains trapped in the vicinity of a handful of basins and does not escape even for times so long as to be inaccessible by classical MD, while a saddle-point search/IRC will show a much weaker dependence on barrier height and, therefore, will be preferable. On the other hand, for temperatures above T_g, saddle-point search suffers from the tremendous multitude of basins (several thousands in the course of nanoseconds for the model system sizes considered here) that need to be sampled, while brute force MD is expected to perform more efficiently. For the temperature range that is of primary interest here, in the vicinity of T_g, both methods suffer. The large number of visited basins makes the saddle-point search method computationally unaffordable, while classical MD must be pushed to its limits. The objective of this paper is to develop an efficient sampling method for this temperature range by achieving maximum parallelization. In a continuation of this work, we will show that it is even possible to accelerate the MD and sample rate constants over a very broad window of time scales with practically the same cost in real-time computation.

Theory

In previous work,[48,49] the dynamics of a glassy material has been described by mapping onto a sequence of transitions between few basins, each basin constructed around an IS. As in that work, the system under study here was a mixture of Lennard-Jones (LJ) spheres that has been used widely to model glassy materials.[24,59,60] The mixture, initially proposed by Kob et al.,[60] consists of two different types of atoms, A and B, with atomic fractions 80% in A and 20% in B. The parameters of the model have been selected[60] in such a way that demixing is suppressed in order to suppress nucleation. Despite the fact that A atoms are larger than B atoms, they are assumed to have the same mass $m_A = m_B = 6.634 \times 10^{-26}$ kg. The LJ interaction parameters are $\varepsilon_{AA} = 1.65678 \times 10^{-21}$J, $\sigma_{AA} = 3.4 \times 10^{-10}$m, $\varepsilon_{BB} = 0.82839 \times 10^{-21}$J, $\sigma_{BB} = 2.992 \times 10^{-10}$m, $\varepsilon_{AB} = 2.48517 \times 10^{-21}$J, and $\sigma_{AB} = 2.72 \times 10^{-10}$m. The unit for reducing time is selected[59,60] as $[m_A\sigma_{AA}^2/(48\ \varepsilon_{AA})]^{1/2} = 3.10 \times 10^{-13}$ s, and the unit for temperature is $\varepsilon_{AA}/k_B = 120$ K. If the above LJ interaction parameters are reduced[61] by the values of the A−A interaction parameters, they read:[60] $\varepsilon_{AA} = 1.0$, $\sigma_{AA} = 1.0$, $\varepsilon_{BB} = 0.5$, $\sigma_{BB} = 0.88$, $\varepsilon_{AB} = 1.5$, and $\sigma_{AB} = 0.8$. In all calculations reported here, the molecular density of the system will be $1.1908\ \sigma_{AA}^{-3}$.

For this system in the supercooled state, Kob[59] and Shell et al.[24] have performed extensive studies, on the basis of which the mode coupling critical temperature T_c is reported as 0.435 in reduced units (∼52.2 K).[59] For the same system, the glass transition temperature has been predicted[24] to be $T_g = 0.32$, that is to say, roughly equal to 38.4 K.

Most previous studies have focused on the region above T_g close to the mode coupling[29−32] temperature T_c, where the system starts to deviate from ergodic sampling according to the mode coupling theory. It was shown that the number of basins visited per unit time by a $N = 641$ particle system depends strongly on the temperature of the system. For temperatures far below T_g, the system remains trapped in the vicinity of a handful of basins, even for times significantly longer than those accessible by conventional MD simulations (microseconds). As one increases the temperature approaching T_g, the number of basins sampled by traditional MD increases, and the temperatures close to and above T_g, it grows to several hundreds. For temperatures well above T_g and T_c, MD sampling is sufficient to capture the basin-to-basin dynamics and reproduce the cage effect[59] and the process of atomic diffusion at long times.[49] Furthermore, it has been shown[48,49] that, for temperatures up to T_g, using artificially trapped MD simulations within each one of the visited basins and determining the transition rates out of the basins, it is possible fully to describe the system's atomistic dynamics in terms of both the coarse-grained motion from basin to basin and the intrabasin motion. The efficiency of such a procedure depends on the relation of the accessible simulation time to the time that the system needs to reach the basin boundary.

In this work, we introduce a self-consistent methodology that allows optimal use of MD over a wide range of temperatures. Here, by the term "optimal use", we refer to the ability of the method to automatically tune the length of MD trajectories used in order to sample inter- and intrametabasin transitions in an uncoupled fashion. For short times and low temperatures, the transitions between individual basins are rare events, while at higher temperatures, close to T_g, traditional MD can sample several basin-to-basin transitions, but the rare event is now the transition between collections of basins. Figure 1 shows results from a simulation in the NVT ensemble, which started from an equilibrated melt configuration (at 55K), under constant temperature close to T_g ($T = 37$K). As one can see clearly in the inset, the system moves between three groups of basins (MB1, MB2, and MB3), where the transitions between groups are significantly slower in comparison to transitions between basins belonging to the same group. Following Heuer et al.,[62−64] we will refer to a collection of basins connected to each other through fast transitions as a "metabasin" (MB). Note that the reason why different MBs can be visually identified in Figure 1 is the "irreversible" nature (overall downhill direction of energy change) of the cooling process. Our target is to create a general autotuned method that enables the identification of a collection of basins and its characterization as a MB and that allows calculation of the transition rates from MB minima toward basins lying outside the MB boundary.

In the literature, several definitions have been proposed for the identification of a MB. Heuer[65] proposed an algorithm based on the IS trajectory, which can be summarized in the

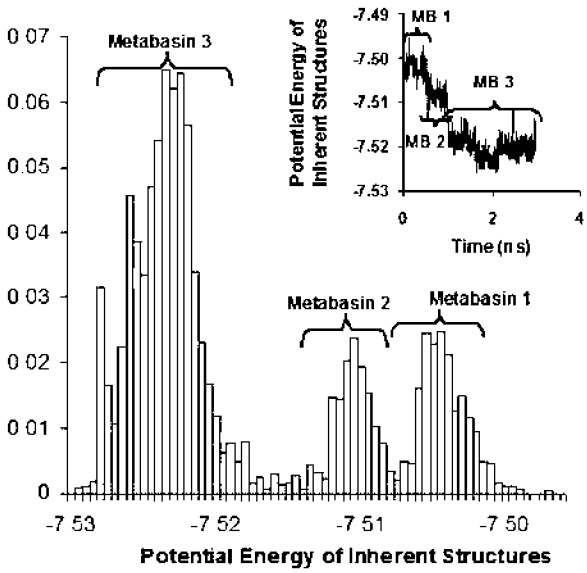

Figure 1. Distribution of the potential energies of inherent structures visited upon cooling the model system from $T =$ 55 to 37 K at a rate of 6 K/ns. In the inset to the diagram, the time evolution of the IS trajectory is given. The simulation time was 3 ns. Three metabasins can be identified visually.

following steps: (a) determine the time regions between the first and the last occurrence for each IS; and (b) group into a MB all basins for which there is an overlap in the corresponding time regions beyond a predefined time scale set to discard recrossing phenomena. Within this approach, the whole trajectory can be regarded, a posteriori, as a succession of different MBs. An alternative definition, independent of a specific trajectory, has been proposed[66,67] by Mauro and Loucks. Starting from rates between inherent structures, subsets are identified based on whether equilibration can be achieved within a prefixed time. As compared to the previous definitions, this allows one in some limit (no unbalanced transition rates)[66] to perform a partitioning of the configuration space into MBs, where the relaxation times within a MB are short compared to that of an observation time scale (prefixed time). However, in practice, many details of the potential energy landscape have to be discarded for the application of this approach. More details regarding MBs in glass-forming systems can be found in the review article of Heuer.[68] Recently,[18] the existence of MBs has been correlated with specific changes in the configuration space governed by the potential energy landscape, more precisely with the extent of cage-breaking (i.e., the molecular mechanism where the first neighbors of individual atoms are changing).

Another obstacle that one has to face when dealing with efficient sampling of basins and MBs is the existence of high-lying basins at the outskirts of the "most probable" basins that the system may visit. The reverse rate constant for leaving these basins to go back to the "probable" basins and MBs is much higher than the forward rate constant, and therefore, the actual probability of being in these peripheral basins is very small. On the one hand, one cannot discard such high-energy basins, since they may constitute passages to a different part of the landscape that may also be very "probable", once it has been reached. On the other hand, one does not want to spend equal computational effort exploring the probable and improbable parts of the landscape. On the contrary, one would like to distribute the computational effort according to the probability of observing the system in each part of the landscape. Thus, besides the definition of MBs, an additional goal of our work is to implement the methodology in such a way that exploring states which do not belong to an important MB (i.e., which are occasionally visited by the system but are very quickly abandoned, as the system returns to a dominant MB) does not consume disproportionally large computational time.

A key feature of our methodology is the use of MD simulation itself in order to automatically tune the "dynamically accessible" part of the landscape under any conditions of temperature and density, given the available simulation time. In practice, we track the rate of exploring new minima by proper bookkeeping of the minima that have already been visited by our initial canonical MD run. When, for a given time interval, a plateau in the plot of the number of identified inherent structures versus simulation time is observed, which implies that the system configuration circulates within a confined collection of basins, we consider that the MB consists of the basins identified up to that point. In this way, we accomplish to group into a MB all the minima that are accessible from a starting minimum for a specific time window and to discriminate them from all other minima, for which sufficient sampling will require more computational effort. If our methodology is applied for low temperatures far below T_g, the MB is determined by a handful of minima or by even a single minimum (in the limiting case).

By construction, transitions from one MB toward its neighboring MBs will occur at a significantly longer time compared to the inner basin-to-basin transitions and to the simulation time used to define the MB. Therefore, the efficient sampling of transitions between MBs is, at least, an order of magnitude more demanding than sampling the inner MB. To accomplish such vigorous sampling, we developed an approach that allows the distribution of load within a parallel procedure that demands the same computational cost as the corresponding conventional MD run, but the results are obtained on a real-time scale more than two orders of magnitude faster. We propose a highly parallelizable scheme to achieve an efficient sampling of the MB dynamics. In this scheme, a long canonical MD trajectory is considered equivalent to a collection of microcanonical trajectories initiated at phase-space points sampled by a relatively short canonical MD simulation entrapped within the MB. Each microcanonical MD trajectory is terminated as soon as it exits the MB. The evaluation of the rate constants is based on hazard plot analysis of either the time difference between exiting and entering a basin, within the MB, or the measuring of the time that it takes the system to leave a basin (or the whole MB), given an equilibrated initialization within the same basin (or MB).

Hazard plot analysis is based on an evaluation of the cumulative hazard. The hazard rate, $h(t)$, is defined such that $h(t)dt$ is the probability that a system, which has survived a time t since its last transition, will undergo a transition at a time between t and dt. The cumulative hazard is defined as

$H(t) = \int_0^t h(t')dt'$. The probability that a transition occurs in time less than t since the last transition is $P(t) = 1 - \exp[-H(t)]$. For a Poisson process, the hazard rate is constant $h(t) = \lambda$, the cumulative hazard is $H(t) = \lambda t$, and the probability is

$$P(t) = 1 - \exp[-\lambda t] \tag{1}$$

In our case of a Poisson process, the rate λ can be extracted as the slope of a plot of the cumulative hazard H versus the residence time t at long times, when the effect of recrossing events has subsided, or as the negative slope of a plot of the quantity $\ln(1 - P^{\text{cum}}(t))$ versus the residence time. The last expression is based on solving eq 1 for λt and replacing $P(t)$ with its corresponding estimate $P^{\text{cum}}(t)$. The cumulative probability $P^{\text{cum}}(t_k)$ at a specific time t_k can be determined as the ratio of the number of transitions that occurred with residence time up to t_k divided by the total number of transitions. For a set of microcanonical trajectories, initiated at phase-space points sampled according to the Boltzmann weight that corresponds to the canonical ensemble, it is possible to group the transitions under the approximation that the system is at "*local*" *equilibrium.*

Consider that a set of transitions out of a given state with residence times less than or equal to t_k is observed at m energy levels (E_1, E_2, ..., E_m). We denote by n_{E_i} the number of transitions that occurred in time less than t_k under energy E_i. The number of transitions k that occurred in time up to t_k is $k = n_{E_1} + n_{E_2} + ... + n_{E_m}$. The total number of transitions out of the considered state is equal to n.

We can now estimate the cumulative probability as

$$P^{\text{cum}}_{\text{NVT}}(t_k) = \frac{k}{n} = \frac{n_{E_1} + n_{E_2} + ... + n_{E_m}}{n} = \frac{n_{E_1}}{n} + \frac{n_{E_2}}{n} + ... + \frac{n_{E_m}}{n} \tag{2}$$

Let u_{E_i} be the number of transitions (at any residence time) out of the considered state observed under constant energy E_i. Then we can write eq 2 as

$$\begin{aligned} P^{\text{cum}}_{\text{NVT}}(t_k) &= \frac{n_{E_1}}{n} + \frac{n_{E_2}}{n} + ... + \frac{n_{E_m}}{n} = \frac{n_{E_1}}{n}\frac{u_{E_1}}{u_{E_1}} + \\ &\quad \frac{n_{E_2}}{n}\frac{u_{E_2}}{u_{E_2}} + ... + \frac{n_{E_m}}{n}\frac{u_{E_m}}{u_{E_m}}, \text{ or} \\ P^{\text{cum}}_{\text{NVT}}(t_k) &= \frac{n_{E_1}}{u_{E_1}}\frac{u_{E_1}}{n} + \frac{n_{E_2}}{u_{E_2}}\frac{u_{E_2}}{n} + ... + \frac{n_{E_m}}{u_{E_m}}\frac{u_{E_m}}{n} \end{aligned} \tag{3}$$

In our scheme, when the system is at local equilibrium, the terms u_{E_i}/n will approximate the probabilities $p(E,T)$ to observe, in a canonical simulation under constant temperature T, the system at energy E_i.

Therefore, we can transform eq 3:

$$\hat{P}^{\text{cum}}_{\text{NVT}}(t_k) = \frac{n_{E_1}}{u_{E_1}}p^{\text{est}}(E_1,T) + \frac{n_{E_2}}{u_{E_2}}p^{\text{est}}(E_2,T) + ... + \frac{n_{E_m}}{u_{E_m}}p^{\text{est}}(E_m,T) \tag{4}$$

The term n_{E_i}/u_{E_i} is the ratio of the number of transitions that occurred with residence time less than or equal to t_k under constant energy E_i during the simulation to the total number of transitions (at any residence time) under the same energy E_i. This ratio corresponds to the probability of observing a transition, at time less than or equal to t_k, if the simulation occurred under constant energy E_i in the microcanonical ensemble. We replace in eq 4 all the terms n_{E_i}/u_{E_i} with $P^{\text{cum}}_{NVE}(t_k)$:

$$\begin{aligned} P^{\text{cum}}_{\text{NVT}}(t_k) &= P^{\text{cum}}_{\text{NVE}_1}(t_k)p^{\text{est}}(E_1,T) + P^{\text{cum}}_{\text{NVE}_2}(t_k)p^{\text{est}}(E_2,T) + ... + \\ &\quad P^{\text{cum}}_{\text{NVE}_m}(t_k)p^{\text{est}}(E_m,T), \text{ or} \\ P^{\text{cum}}_{\text{NVT}}(t_k) &= \int P^{\text{cum}}_{\text{NVE}}(t_k)p^{\text{est}}(E,T)\mathrm{d}E \end{aligned} \tag{5}$$

where $\int p^{\text{est}}(E,T)\mathrm{d}E = 1$. We have now expressed the probability to observe a transition in time $\leq t_k$ at a specific temperature T as an ensemble average over microcanonical trajectories initiated at phase-space points sampled in the course of an equilibrium canonical simulation. The initial points of each microcanonical trajectory are sampled based on the canonical ensemble in our algorithm.

Alternatively, one can derive eq 5 based on a superposition ansatz for the residence time distribution. Consider a system evolving along an NVT MD trajectory. We focus on transitions of the system into and out of a given state (basin or MB). The NVT trajectory will be assumed long enough to achieve local equilibration within a confined region of configuration space (MB or group of MBs, respectively) in which the system is temporarily trapped and which contains the considered state. In practice, the NVT MD trajectory is generated by coupling the system with a heat bath (e.g., through an extended ensemble technique). The time constant governing exchange of energy between the system and the heat bath must be long in comparison to the mean residence time in the state on which we focus; otherwise, our observations will be perturbed by interactions with the heat bath and will not reflect the true dynamics dictated by the potential energy hypersurface and the masses of system particles. Typically, each transition into and out of the state will primarily involve a relatively small subset of degrees of freedom of the system. The energy associated with this subset does fluctuate at a faster rate than the total energy of the system. Under these conditions, the total energy E of the system between entry and immediately following the exit from the considered state will remain practically constant. By definition, then, the residence time distribution $\mathrm{d}P/\mathrm{d}t|_{\text{NVT}}$ determined in the course of a long NVT MD trajectory that allows the system to go in and out of the considered state can be related to the residence time distribution $\mathrm{d}P/\mathrm{d}t|_{\text{NVE}}$ that would be observed in the course of NVE MD trajectories conducted at energy levels E as

$$\left.\frac{dP}{dt}\right|_{NVT} = \int \left.\frac{dP}{dt}\right|_{NVE} p(E,T)dE \quad (6)$$

Integrating with respect to time, this gives

$$P(t)|_{NVT} = \int P(t)|_{NVE} p(E,T)dE \quad (7)$$

In eqs 6 and 7, $p(E,T)$ is the probability of observing the system at energy E. The latter probability, however, is, by construction of the considered long, locally equilibrated NVT trajectory, proportional to the Boltzmann factor of the energy, retrieving eq 5.

Thus, we have expressed the cumulative probability to have undergone a transition at time t at a specific temperature T as an ensemble average of the corresponding cumulative probabilities calculated along microcanonical (NVE) trajectories. The initial phase-space points of the microcanonical trajectories are sampled by a canonical ensemble NVT MD simulation that has achieved local equilibration among a group of states to which the considered state belongs. The proposed approach can be envisioned as the reconstruction of an ensemble of NVT trajectories from a weighted ensemble of NVE trajectories. The correct dynamics can be sampled via either an NVE simulation or an NVT simulation in the limit where the thermostat interacts weakly with the system, in a way that does not perturb the system's time correlation functions. Nevertheless, our approach of using a swarm of NVE trajectories has a significant computational advantage; in the traditional NVT ensemble, once the interaction with the thermostat is weakened, the necessary time for thermal equilibration increases, whereas in our case, this is overcome by the proper weighting of each dynamical path. Despite the advantages of using NVE trajectories discussed above, we have actually tested whether it is possible to use NVT trajectories instead of NVE, and we have shown that, in practice, there is no significant difference.

Molecular Simulation approach. The first step in our methodology is to determine "on the fly" the local potential energy landscape that constitutes the MB, using a small-duration canonical MD simulation. Along an atomistic NVT MD simulation, at regular time intervals, the potential energy was minimized with the method of conjugate gradients[69] in order to identify the ISs of the MB. For the identification of ISs, Stillinger[4] proposed the use of the steepest descent method. In this work, we have chosen to use the method of conjugate gradients, which leads to the same IS as steepest descent in the overwhelming majority of cases but is significantly faster than steepest descent. To each one of the visited ISs we attribute an identity, storing its potential energy and its configuration. At this stage, we can have an estimate of the rate constants based on our previous work:[48,49] For each IS, we may choose to collect the transition times toward neighboring basins and their corresponding conditional probabilities and to compare the results of this stage with the final result of the proposed parallel scheme. The criterion we use to ascertain that the current MB has been sufficiently explored is based on the rate of identification of new (not already visited) basins of the potential energy and reflects the achievement of local equilibrium within the MB. Under these conditions, the system configuration circulates within a confined collection of basins, and a plateau in a plot of the number of identified ISs versus simulation time is observed. We consider that our MB consists of the number of basins identified up to that point. In Figure 2, we present a plot of the number of identified ISs versus the simulation time for a specific MB comprised of 290 minima.

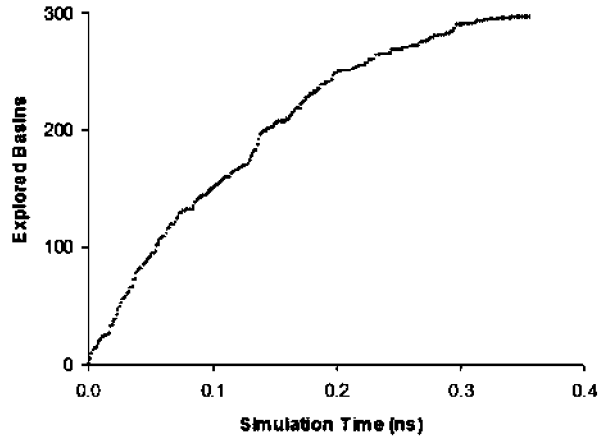

Figure 2. Number of explored minima as a function of time at $T = 37$ K. When a plateau is observed for a prefixed time interval, the explored minima are considered to belong to the same MB.

This approach can easily be combined with the DIMW methodology for creating an ever-expanding network of MBs based on the following steps: (a) Define a MB based on short NVT MD runs, as described above; (b) Evaluate rate constants with the proposed methodology for all interbasin transitions within the MB and for basin-to-basin transitions terminating outside the MB; (c) Select an unexplored basin that lies outside of, but is connected to, the current (explored) MB based on the DIMW methodology; this is a starting point for identifying an *additional* MB; and (d) Loop back to step (a) to conduct an NVT MD out of the selected basin, but now, if the MD run reaches an explored basin of an explored MB, then go to step (b) and choose to leave the explored MB from a new unexplored basin (which can be one of the basins that are grouped as the additional MB). Continue the MD run until the rate of finding new MBs drops below a preset value. Note that, in this way, every new MB is not independent of the previous ones, but the union of identified MBs defines a set of basins in which the system will spend "sufficient" time before exploring new basins.

In order to evaluate the rate constants necessary for step (b), we proceed in the following manner: We produce an equilibrium sampling of phase-space points within the basins belonging to the MB via the execution of artificially trapped long canonical MD simulations. The artificial entrapment is implemented using reflective conditions at the MB boundaries. That is, once the system exits the MB, we invert the momenta (of the atoms and thermostat) stepping the system backward, returning it to the MB, following the procedure introduced in our previous work, where the inversion of the momentum was used to trap the system in individual basins.[48,49] The inversion of momenta is followed by an appropriate Gaussian randomization (that preserves the canonical distribution) of the momenta. This is used to perturb the system away from the original trajectory and

assist the chaotic character of the system dynamics in sampling nearby trajectories. The new atomic momenta correspond to the imposed temperature of the canonical simulation, via the equipartition theorem, and sum to a total momentum of zero, as in the initial state. The duration of the artificially trapped simulation is ten times longer than that of the canonical MD that was used for the identification of the MB (step a). During the artificially trapped canonical simulation we store phase-space configurations of the system at constant time intervals. At the same time intervals, we minimize the potential energy of the system under constant volume and, thereby, make an assignment of sampled phase-space points to basins. Thereby, we can ensure that the stored phase-space point belongs to the sampled MB or record it as belonging to another MB. Within the simulation, we identify and store all transitions between basins belonging to the sampled MB and between basins of the sampled MB and unexplored basins lying outside the MB. It should be noted that the idea of a trapped simulation has been inspired by the novel methods pioneered by Voter[54,55] for studying the dynamics of rare events. Our aim is to combine these novel ideas with hazard plot analysis[58] for the calculation of rate constants via generation of a simple ensemble of MD trajectories in a high-dimensional energy landscape.

At this point, we propose a highly parallelizable scheme to achieve an efficient sampling of the MB dynamics *by considering the canonical molecular dynamics trajectory as equivalent to a collection of microcanonical trajectories* at total energies that have been chosen according to a Boltzmann weight corresponding to the canonical ensemble. To ensure that the initial states have been chosen with the appropriate Boltzmann weight, we confirm that the total energy distribution of the selected phase-state points provides a good estimate of the total energy distribution of the long canonical simulation along which they were sampled.

Out of each initial phase-space point, stored in the course of the "locally equilibrated" trapped long canonical MD simulation, we start two microcanonical (NVE) MD trajectories, one forward and one backward (by reversing the momentum of each particle). We continue our conjugate gradient potential energy minimizations at regular time intervals along each of these microcanonical trajectories, and we identify the minima based on the energy and on an Euclidean distance in configurational space (where atoms are considered distinguishable; swapping the identities of two particles results in a different basin). Each microcanonical MD simulation is terminated once the system visits an "unexplored" basin that does not belong to the current MB. We collect all the transitions between basins belonging to the sampled MB and between MB basins and unexplored basins lying outside the MB and perform hazard plot analysis on them as we have proposed for the basin-to-basin case.[48,49] We then determine the rate constant for each transition by ensemble averaging over the entire swarm of microcanonical trajectories according to eq 5.

This scheme of conducting a swarm of microcanonical MD simulations can be highly parallelized, since it does not require any communication between the simulations out of the different phase-space points. All necessary steps for the implementation of the parallel scheme are shown in Figure 3. In Figure 4, we provide a schematic representation of a configuration space, depicting the basic steps of the proposed methodology.

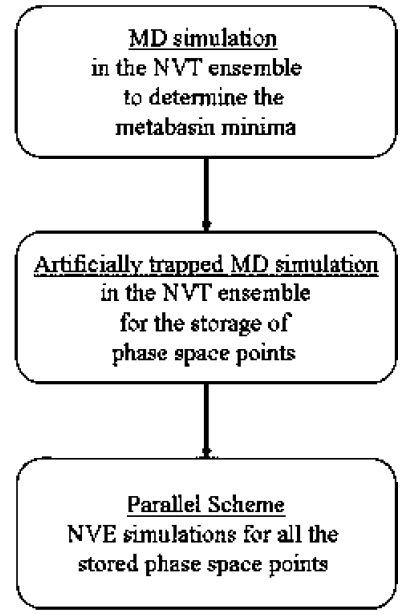

Figure 3. Flow of calculations according to the proposed parallel scheme.

Parallel Implementation. The parallel computational work was conducted at the Supercomputing Center CINECA in Bologna, Italy. The simulations were performed on the IBM BCX/5120 cluster, which is mainly used for massively parallel applications and special high-end projects. The cluster consists of 2180 nodes, where each node is supplied with 2 Opteron dual core processors at 2.6 GHz. The nodes communicate via infiniband (5Gb/s) network. More information about the cluster can be found at http://www.cineca.it.

The implementation of the parallel scheme has employed a noncommercial MD code developed by the authors, using the Message Passing Inteface (MPI) library for distribution of the computational load. As we described above, we need to perform two microcanonical MD runs for each one of the stored phase-space points. Since these MD runs are completely independent, we use MPI to distribute the initial phase-space points among the available processors. Each processor performs a set of MD trajectories that start from the phase-space points assigned to it and end once the system has reached the "boundaries" of the MB. Energy minimization is performed, with the method of conjugate gradients,[69] at regular time intervals of the order of 0.1 ps. Each microcanonical MD simulation is completed when the system comes out of the MB, that is when, after the minimization procedure, the system is found in an "unexplored" IS that does not belong to the current MB. Thus, by construction, the simulation times of the independent microcanonical trajectories vary, and so does their corresponding execution time.

In order to share the computational cost into the parallel procedure, we developed two implementations: equally and unequally distributed configurations. In the first implementation, before the parallel simulation starts, we assign to the processors equal numbers of initial phase-space points. Each processor knows the number and identity of the phase-space

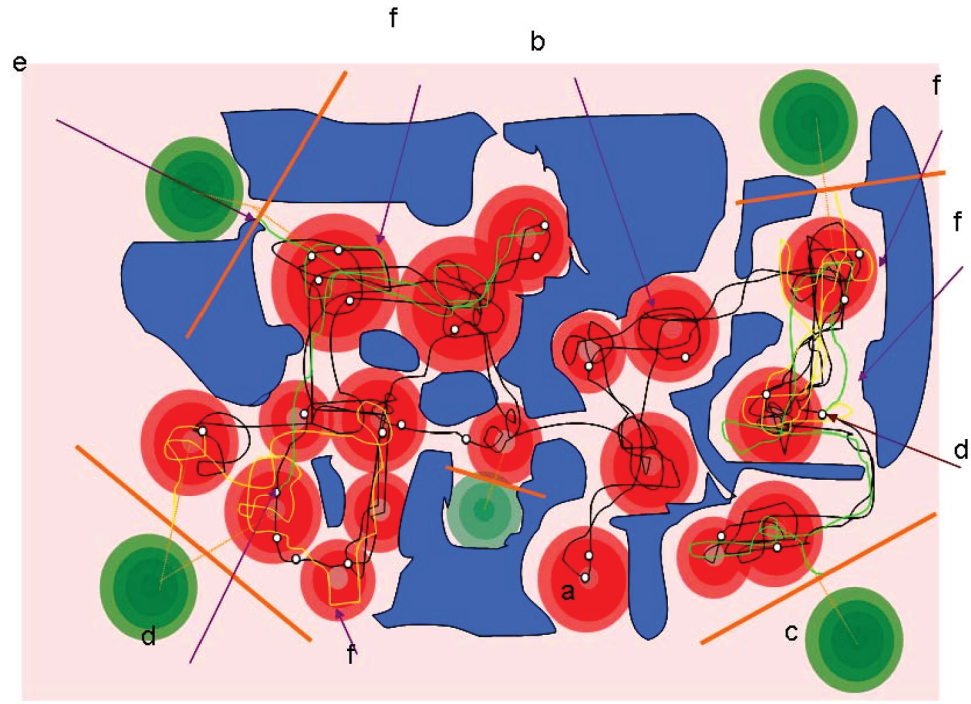

Figure 4. A two-dimensional (2D) cartoon representation of a configuration space depicting the basic steps of the proposed methodology: (a) (Red regions) representation of inherent structures/basins which constitute the sampled MB. (b) (Black line) configuration space projection of the NVT MD trajectory used to define the MB. (c) (Green regions) neighboring basins, which are not part of the MB. (d) (Small white circles) configuration space projections of points in phase space sampled during an NVT MD trajectory (represented again by the black line) entrapped within the MB; these points are employed as the starting points for the swarm of NVE MD trajectories used to evaluate the rate constants. (e) Ending points of the NVE MD trajectories, i.e., points at which these trajectories leave the MB. (f) Two NVE MD trajectories (yellow and green) started from two of the stored phase-space points. For reasons of clarity, only one NVT MD trajectory is drawn here (black line). In reality we use two, one for the initialization of the MB and one for the collection of the NVE MD starting points. In blue, we depict the inaccessible (high potential energy) part of the configuration space; in reality, this constitutes a very large fraction of the configuration space.

points for which it will execute microcanonical simulations. This simplified parallel implementation was constructed conveniently, since the required programming cost is small. We proceeded to develop a more complex implementation, that of unequally distributed configurations, which aims at equalizing the distribution of computational load. Within this implementation, from the N processors used, one (e.g., node0) is dedicated to "dealing" initial phase-space points out to the remaining $N - 1$ processors. Initially, node0 distributes, via the MPI library, one phase-space point to each processor. As soon as a processor (from the set of $N - 1$) completes the simulation of its assigned phase-space point, it communicates, via the MPI library, with node0, and a new phase-space point is assigned to it, until all phase-space points finish their microcanonical simulation.

Results and Discussion

In order to validate the proposed scheme we compare, for a specific well-sampled basin of the identified MB, the transition rate obtained for exiting it from our proposed parallel methodology and from the artificially trapped canonical MD simulation in the MB. As one can clearly see in Figure 5, the resulting rates (long-time slopes of the cumulative hazard with respect to residence time) are in very good agreement. The comparison of Figure 5 has been performed for all the basins that constitute the MB.

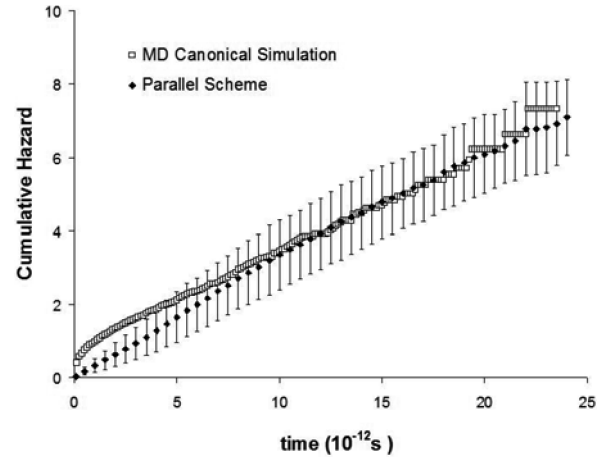

Figure 5. Comparison between the results obtained from an artificially trapped trajectory in the canonical ensemble and the proposed parallel scheme for the calculation of the cumulative hazard to exit a specific basin of configuration space at temperature $T = 37$ K.

One of the advantages of using the hazard plot analysis is that, on top of the evaluation of the rate constant, the method validates whether the process is first order or not, since the linearity in the hazard plot is equivalent to an exponential distribution of the associated residence times. As has been described in our previous work, our hazard plot analysis has been designed to evaluate the sum of rates out of a basin or

Table 1. Comparison of the Number of Saddle Points (Corresponding to Transitions between the MB minima) Identified with MD and with the Proposed Parallel Methodology[a]

	saddle points	saddle points per execution time (h^{-1})
MD	3910	326
parallel scheme	24 271	1867

[a] Both applied within a MB. The MB consists of 290 minima.

Table 2. Comparison of the Execution Time, the CPU Cost and the Simulation Time between MD and the Proposed Parallel Methodology[a]

	execution time (h)	CPU cost (h)	simulation time (s)
MD	12	12	3×10^{-9}
parallel scheme	12 + 1	250	7.7×10^{-8}

[a] The CPU cost for the parallel scheme is given by the product of the execution time of the "slowest" processor and the number of processors used.

a metabasin. Under the assumption of a Poisson process, each rate is the sum of the individual transition rates to any other basin or MB, through single or multiple routes. As has been described in our previews work,[48,49] basin-to-basin transitions are clearly of first order at low temperatures, whereas at temperatures higher than T_g, one has to look for MB-to-MB transitions to recover a clear first-order character.

Note that we have used two different but equivalent methods to determine the rate constants from our microcanonical MD trajectories based on hazard plot analysis. First, we have used the traditional idea proposed by Helfand[58] of analyzing the residence time within each of our discrete states (basins), i.e., the difference between the exit and entrance times. On the other hand, we also chose to analyze (again via hazard plot analysis) the ensemble of times that it takes to exit a basin (or a MB) when the initial phase-space point has been chosen according to the local equilibrium conditions within the state. We use the second hazard plot approach when we calculate the rate constants out of both the basin of the initial "stored" phase-space point and the MB itself toward the basins of neighboring MBs. Whereas, we use the first hazard plot approach when we calculate the rate constants between all other basins that we encounter after the initial one, until we reach the final basin of the trajectory. Our hazard plot calculations are performed along the canonical or the swarm of microcanonical MD trajectories.

In Table 1, we can see that using classical MD we observe only 3910 transitions between basins in the MB, while using the proposed parallel methodology this number becomes six times larger (24 271). In Table 2, we present a comparison of the computational cost and the simulation time between the classical MD and the parallel method. The simulation time of our parallel MD scheme is 20 times longer that that of the corresponding artificially trapped MD, and the total computational cost of the parallel scheme is also around 20 times larger.

To quantify the parallelizability of this scheme, we studied the speedup factor (S_p) and the parallel efficiency (E_p) as functions of the number of processors used. The speedup

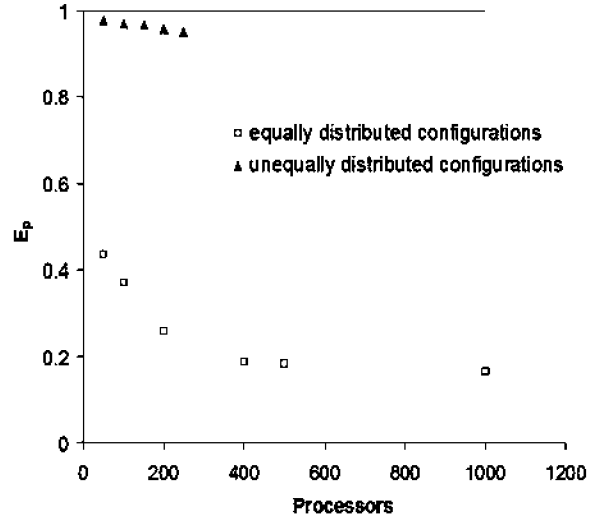

Figure 6. Dependence of the efficiency factor (E_p) on the number of processors used for the implementation of equally distributed configurations (□) and for the implementation of unequally distributed configurations (▲).

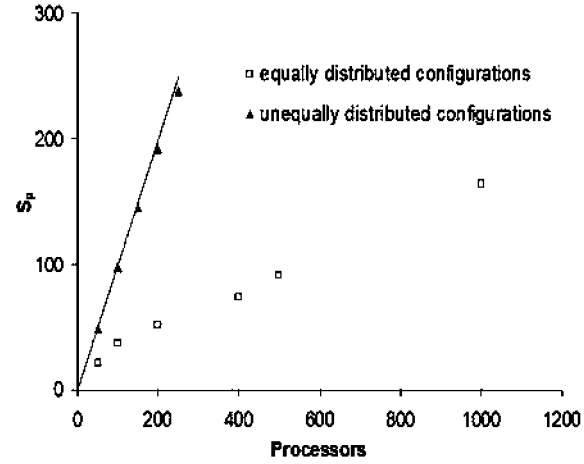

Figure 7. Dependence of the speedup factor (S_p) on the number of processors used for the implementation of equally distributed configurations (□) and for the implementation of unequally distributed configurations (▲).

factor is defined by the following equation: $S_p = T_1/T_p$, where p is the number of processors used, T_1 is the execution time of the sequential run, and T_p is the execution time of the parallel scheme with p processors. The parallel efficiency is defined as: $E_p = S_p/p$.[70,71]

Results for the equally distributed configurations are presented in Figures 6 and 7 (□). As one can clearly see in these figures, this implementation for our proposed parallel application manages to reduce the execution time of our simulations even using up to 1000 processors. The speedup factor of this implementation indicates that using 1000 processors, with computational cost approximately six times larger than the corresponding sequential simulation ($E_{1000} \approx 0.164 \approx 1/6.11$), we shorten the execution time by a factor of 163 ($S_{1000} \approx 163$). The parallel efficiency obtained from this implementation is significantly lower than the optimum value that can be achieved by our parallel scheme. Since it does not demand any communication between the microca-

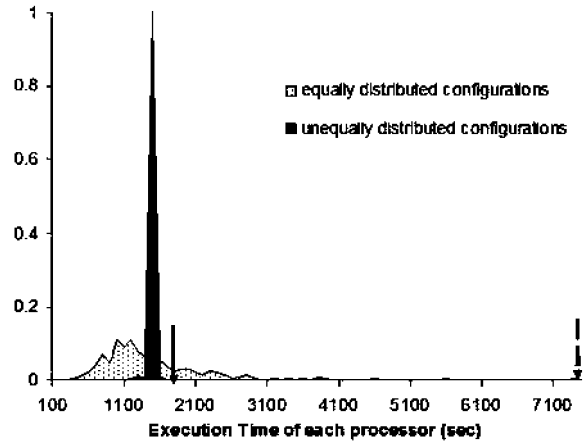

Figure 8. Representation of the distribution of the execution times of all processors for a simulation on 250 processors using the implementations of equally and unequally distributed configurations. The broken and solid arrows point at the maximum execution time for the first and second implementations, respectively.

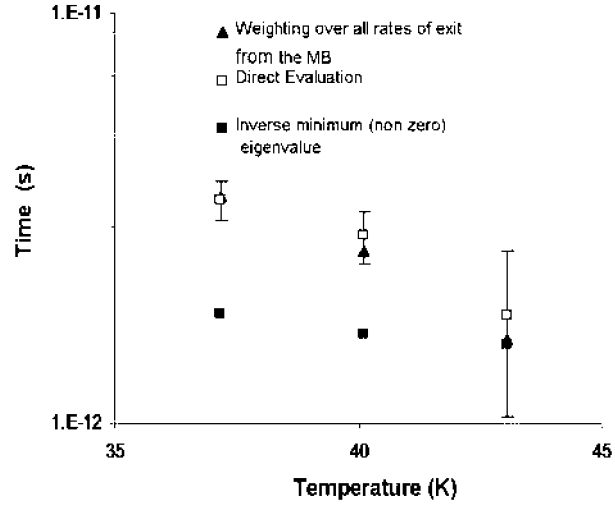

Figure 9. Comparison of the mean residence time in a MB that has been identified at 37 K, calculated by summing all rates for exiting the MB from any one of its basins, each weighted with the equilibrium probability of the corresponding basin (▲) and by hazard plot analysis over the artificially trapped microcanonical MD trajectories (□). The inverse minimum (nonzero) absolute eigenvalue of the rate-constant matrix for the specific MB is also shown, for comparison (■). The comparison has been performed for three temperatures (the basins constituting the MB are the same in all temperatures; what is changed is the temperature of the artificially trapped MD simulations).

nonical simulations, the proposed parallel scheme should allow higher values for the speedup factor (approaching linear dependence on the number of processors used − linear scaling) and for the parallel efficiency (approaching unity). The implementation of the equally distributed configurations suffers from load-balancing problems, due to the heterogeneity of the total simulation times of each processor. The distribution of execution times of the various processors for this implementation is presented in Figure 8 (gray area). Since completion of our parallel simulation demands that the simulation on the "slowest" processor be finished (computational cost is charged according to this rule in the supercomputing centers), the execution time of the overall simulation is determined by the execution time of the "slowest" processor. In order to solve this load-balancing problem, we developed the implementation of unequally distributed configurations. The dependence of the parallel efficiency and the speedup factor on the number of processors used, for this implementation, is presented in Figures 6 and 7, respectively. The distribution of execution times for a simulation on 250 processors, using the implementation of unequally distributed configurations, is presented in Figure 8. Using this implementation, as we can clearly see in Figure 6, we overcome the load-balancing problem plaguing the implementation of equally distributed configurations. As a result, we achieve an almost linear scaling for the speedup factor, even if we use a large number of processors ($S_{250} \approx 238$) and reduce the computational cost very significantly, close to the computational cost of the sequential simulation ($E_{250} \approx 0.92$). A great advantage of this implementation is that almost linear scaling is achieved by our parallel scheme, even if we use a large number of processors (250), independently of the simulated system size. For high-performance parallelization of MD simulations to be realized by domain decomposition into a large number of processors, most simulation software requires very big systems (of the order of 1×10^5 atoms) to be simulated. The implementation we propose here achieves its high performance even for very small systems, such as the system of Lennard-Jones spheres studied here (641 atoms).

The self-consistent methodology proposed for the definition of the MB aims at an automated selection of potential energy basins, among which local equilibration can be assumed to be established over the (long) time scales of interest. Equivalently, this can be translated into the requirement that the mean time spent in each of these minima is significantly longer than the time required for the achievement of local equilibration. For a system whose dynamics can be envisioned as a succession of transitions, with kinetics described by a first-order law, the response to any perturbation from the equilibrium probability distribution among states (in our case, basins) can be resolved into modes, each mode corresponding to a projection on an eigenvector of the matrix of transition-rate constants describing the dynamics of the system, appropriately symmetrized.[9,42,72,73] The matrix of transition-rate constants $\mathbf{K}$ is defined as follows:

$$K_{ij} = k_{j\to i} \forall i \neq j, \quad K_{ii} = -\sum_j k_{i\to j} \tag{8}$$

where $k_{i\to j}$ is the rate constant for the elementary transition from basins i to j.

As in the "EROPHILE" approach, the N-dimensional vector $\mathbf{P}(t)$ of state probabilities for observing the system at time t in each one of the N distinct states $\mathbf{P}(t) \equiv (P_1(t), ..., P_i(t), ..., P_N(t))$ is transformed into a reduced vector $\tilde{\mathbf{P}}(t)$ with elements $\tilde{P}_i = P_i/\sqrt{P_i(\infty)}$. The elements of the transition-rate constant matrix are correspondingly transformed as $\tilde{K}_{ij} = K_{ij}\sqrt{P_j(\infty)}/\sqrt{P_i(\infty)}$. Under the condition of microscopic

reversibility (detailed balance) on the rate constants ($k_{i\to j}P_i(\infty) = k_{j\to i}P_j(\infty)$) the matrix $\tilde{\mathbf{K}}$ is symmetric and similar to the transition-rate constant matrix $\mathbf{K}$. Since they are similar matrices, they have the same eigenvalues, which have to be all real due to the symmetry of the $\tilde{\mathbf{K}}$ matrix. On the other hand, the form of $\mathbf{K}$ guarantees that there is at least one 0 eigenvalue and that all other eigenvalues are negative. We denote these eigenvalues by $\lambda_0 = 0 \geq \lambda_1 \geq, ..., \geq \lambda_{N-1}$ and symbolize by $\tilde{\mathbf{u}}_n = (\tilde{u}_{1n}, \tilde{u}_{2n}, ..., \tilde{u}_{in}, ..., \tilde{u}_{Nn})$, the eigenvector of $\tilde{\mathbf{K}}$ corresponding to eigenvalue λ_n, $0 \leq n \leq N-1$. The solution of the time-dependent state probabilities can be written as

$$P_i(t) = \sum_{n=0}^{N-1}\sum_{j=1}^{N}\frac{\sqrt{P_i(\infty)}}{\sqrt{P_j(\infty)}}\tilde{u}_{i,n}\tilde{u}_{j,n}e^{\lambda_n t}P_j(0), \text{ or}$$

$$\tilde{P}_i(t) = \sum_{n=0}^{N-1}\sum_{j=1}^{N}\tilde{u}_{i,n}\tilde{u}_{j,n}e^{\lambda_n t}\tilde{P}_j(0) \qquad (9)$$

or, in vector form:

$$\tilde{\mathbf{P}}(t) = \sum_{n=0}^{N-1}[\tilde{\mathbf{u}}_n\cdot\tilde{\mathbf{P}}(0)]\mathrm{e}^{\lambda_n t}\tilde{\mathbf{u}}_n = \tilde{\mathbf{P}}(\infty) + \sum_{n=1}^{N-1}[\tilde{\mathbf{u}}_n\cdot\tilde{\mathbf{P}}(0)]\mathrm{e}^{\lambda_n t}\tilde{\mathbf{u}}_n \qquad (10)$$

The eigenvectors $\tilde{\mathbf{u}}_n$ form an orthonormal basis set: $\tilde{\mathbf{u}}_m \cdot \tilde{\mathbf{u}}_n = \delta_{mn}$, $0 \leq m, n \leq N-1$, with δ_{mn} being the Kronecker delta. They also satisfy $\sum_{n=0}^{N-1}\tilde{u}_{i,n}\tilde{u}_{j,n} = \delta_{ij}$.

Whereas eq 9 has been proposed in the past[9,73] in order to describe the time evolution of the state probabilities, in the work of Boulougouris and Theodorou, referred to as the "EROPHILE" approach,[42] the Euclidean orthonormal basis set created by the eigenvectors $\tilde{\mathbf{u}}_n$ is used for the first time, to our knowledge, to describe not only the state probabilities but also any real observables (i.e., their time-dependent averages and auto- and cross-correlations). For any observable A, it is possible to perform a transformation in the "EROPHILE" space, creating a Euclidean vector with components $\tilde{A}_i = A_i\sqrt{P_i(\infty)}$, where A_i is the value of the observable in state i of the system. The Euclidean vector $\tilde{\mathbf{A}}(t)$ can then be expressed in the same basis set as the state probability vector $\tilde{\mathbf{P}}(t)$.

EROPHILE is able to identify a relaxation mode either as a redistribution of the state probabilities in such a way that the return to equilibrium will occur along a single exponentially decaying function or, equivalently, as an observable for which the autocorrelation function will decay as a single exponential. From eq 10 it becomes obvious that each projection on every one of the eigenvectors evolves independently and, as the system approaches equilibrium, all mode contributions except the one corresponding to the 0 eigenvalue tend to 0 in an exponentially decaying fashion ($e^{\lambda_n t}$). Therefore, for times longer than the inverse minimum absolute value (nonzero) eigenvalue of the rate constant matrix, any perturbation of the system, no matter how big or improbable, will have been damped, and the system will have attained local equilibrium. To judge whether or not the system has achieved local equilibrium before leaving a given set of states, one has to compare the negative inverse of the smallest in absolute value nonzero eigenvalue of the transition rate matrix, $-1/\lambda_1$, with the average time it takes to leave the given set of states. In our case, we validate our "on the fly" identification of MBs by performing this comparison of $-1/\lambda_1$ against the mean residence time in the MB, as obtained from hazard plot analysis of our MD trajectories. As is shown in Figure 9, the former is three times smaller than the latter at 37 K and remains smaller at 40 and 43 K. An additional strong indication that local equilibration has been achieved within the MB is the equality between the values predicted for the mean residence time in the MB, as calculated directly via hazard plot analysis of the artificially trapped microcanonical MD trajectories and as estimated via summation of the rates for exiting the MB from any basin belonging to it, each weighted by the equilibrium probability of the basin, assuming local equilibrium (see Figure 9):

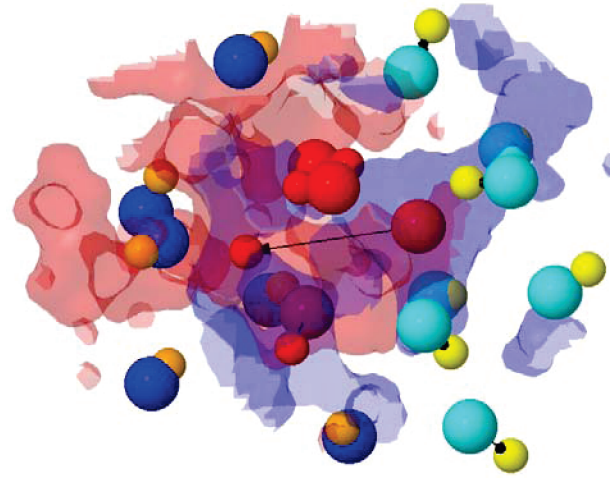

Figure 10. Schematic representation of a cage-breaking event in a single jump from one potential energy minimum to a neighboring one. The positions of the atoms that participate in the transition are plotted with different sizes (initial: big; final: small), and vectors are drawn to indicate their displacements accompanying the transition. With red color we represent the atoms that remain first neighbors to the central atom experiencing the cage-breaking event, which is also shown in red. Cyan represents atoms that used to be first neighbors of the central atom but cease being so after the transition; their new positions are shown in yellow. Dark blue represents atoms which were not first neighbors of the central atom but come into its first coordination shell after the transition; their new positions are shown in orange. The blue surface depicts the volume accessible to the central molecule initially and the red finally, illustrating the cage change accompanying the transition.

$$k_{\mathrm{MB}\to\mathrm{out}} = \sum_i p_i^{\mathrm{eq}}k_{i\to j},\ i \in \mathrm{MB},\ j \notin \mathrm{MB} \qquad (11)$$

$$\langle t\rangle_{\mathrm{MB}\to\mathrm{out}} = 1/k_{\mathrm{MB}\to\mathrm{out}} \qquad (12)$$

Note that p_i^{eq} values are normalized to 1 in this calculation.

As mentioned above, this model system has been thoroughly studied in the past[17−20] and has provided very useful insights into the molecular motion relevant to relaxation in the vicinity of the glass transition, namely the "cage-breaking" process. More precisely, the change in the number of first neighbors accompanying a transition in the potential

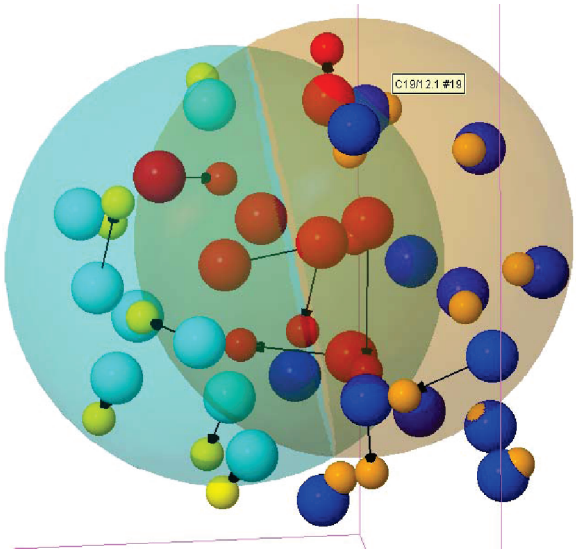

Figure 11. Schematic representation of a more complex relaxation event that takes place in a single jump from one potential energy minimum to a neighboring one, belonging to a different MB. Coloring is as in Figure 10. We have also drawn cyan and light-brown spheres representing the first coordination shell, centered at the initial and final positions of the atom with the largest displacement. In this complex elementary move, atoms look like they are moving in a concerted way, exchanging their positions in a dance-like fashion.

energy landscape has been thoroughy examined in the work of Souza and Wales.[17] In our work, we also see cage-breaking events, as depicted in Figure 10, where a "central" atom jumps to a new cage after a single transition. On the other hand, when we investigate MB-to-MB transitions, we do not only see an enhancement of this process but also observe more complex relaxation mechanisms (Figure 11), wherein a number of atoms take each other's positions moving in a, more or less, stringlike fashion, as if they were "dancing" in accordance with the "stringlike cooperative motion" demonstrated by the work[74] of Donati et al. In their work, by analyzing the van Hove correlation function produced via MD simulation for the same model system, Donati et al. showed that there is a fraction of mobile particles that at a characteristic time replace each other, executing a stringlike cooperative motion. Our result suggests that such a motion can be seen as a basin-to-basin transition intimately linked with the MB-to-MB transitions. We plan to investigate these more complex motions further in the future, since they require a great deal of cooperation between the atoms involved but may entail a less unfavorable energy barrier than the single cage-breaking events which move a molecule from one cage to another, bypassing (or pushing out) its first neighbors.

Conclusions

We have developed an automated self-consistent method which can operate on the fly within a molecular dynamics (MD) simulation, allowing the identification of collections of basins and their characterization as metabasins (MBs). The criterion used to define MBs from short MD runs is based on the rate of identification of new, not already visited, basins. In practice, when for a given time interval we observe a plateau in the number of identified inherent structures (ISs) versus simulation time, implying that the system configuration circulates within a confined collection of basins, we consider that a MB has been identified, consisting of the basins visited up to that point. The proposed approach gives the ability to calculate the presence of a MB on the fly, the "minute" the system is trapped in a part of its configuration space; it does not require a postprocessing of the dynamics after the visit of several (at least two) MBs, as some previously proposed methods do.

The identification of a MB is followed by a calculation of the individual rate constants governing transitions between the basins constituting the MB and transitions toward basins that do not belong to the current MB. The computational cost for this calculation, which demands minimization of the potential energy for basin identification at regular time intervals, is significantly high, so we proceeded to develop a methodology that overcomes this obstacle. Our methodology distributes the vast computational cost associated with this calculation into a series of small duration-independent microcanonical MD simulations, conducted in parallel. Initial phase-space points for these microcanonical simulations are taken from a canonical MD simulation trapped within the MB. The execution time of the parallel microcanonical simulations is reduced dramatically, since our methodology does not require any information exchange between the simulations. Our results from the parallel methodology were compared against results from long artificially trapped canonical MD simulations and found to be in very good agreement. By implementing a scheme of unequally distributed configurations, wherein the parallel microcanonical MD runs are assigned to available processors in a manner that ensures good balancing of the computational load, we were able to achieve almost linear scaling ($E_{250} \approx 0.92$) on up to 250 processors, at a total computational cost similar to the cost of the corresponding sequential simulation. Additional advantages of our parallel methodology are that its applicability is independent of the system size (its high parallelization speedup and efficiency can be achieved even with very small systems, contrary to what happens with domain decomposition), and independent of the cluster architecture (shared memory or not). Using the proposed parallel methodology, we have examined the validity of our definition of the MB from the point of view of achievement of local equilibration among the basins that constitute the MB. To do so, we have compared the mean residence time between entry to and exit from the identified MB, calculated in two independent ways, with the time for decay of any perturbation away from local equilibrium within the MB. We observed that the mean residence time is significantly longer, indicating that local equilibrium has been achieved and that the MB has been successfully defined. The proposed parallel methodology distributes the vast computational cost of our calculation into practically independent runs, making it ideal as a backfill job on computing clusters. On the other hand, using a "workload management system" in scheduling the independent molecular dynamics runs has proved es-

sential to efficiency. Beyond the usual benefits of parallel implementation, the Poisson character of our process (exponential distribution of the residence time) causes the individual independent runs to have widely varying computational cost.

This work is designed to extend the sampling ability of traditional MD simulation by utilizing an exremely efficient parallel approach. The approach is developed specifically to overcome some of the most vicious obstacles in the simulation of glassy systems, by turning them into an advantage. For example, the separation of time scales between intra- and intermetabasin transitions, which "immobilizes" traditional MD sampling, is now turned into an advantage, allowing for automated definition of fast and slow processes relative to the MD sampling ability. The novelty of the proposed approach lies in its design to overcome specific problems in simulating glass-forming systems. Furthermore, this is one of only few successful attempts to parallelize with high efficiency the calculation of dynamical properties. Last but not least, the use of the idea of a swarm of NVE trajectories to estimate the distribution of residence times in the NVT ensemble, and from that the rate constants, is, to our knowledge, novel and far from trivial.

The design of the algorithm aims at, as simple as possible, an implementation on the top of any existing MD package. The necessary tools are an MD simulator with the ability to perform minimizations; reflection and randomization of momentum[48,49] at specific intervals, depending on the result of the minimization; and a simple book-keeping procedure for visited minima. Furthermore, in all probability the proposed algorithm will be integrated into a general-purpose simulation package (probably as a tool in the MAPS program[75] of Scienomics SARL) based on a open-source MD platform.

Acknowledgment. This paper is part of the 03ED375 research project, implemented within the framework of the "Reinforcement Programme of Human Research Manpower" (PENED) and cofinanced by National and Community Funds (20% from the Greek Ministry of Development-General Secretariat of Research and Technology and 80% from E.U.-European Social Fund). Computational work was carried out under the HPC-EUROPA project (RII3-CT-2003-506079), with the support of the European Community—Research Infrastructure Action under the FP6 "Structuring the European Research Area" Program. The authors would like to thank Dr. Loukas Peristeras for his contribution to the development of the parallel code.

References

(1) Angell, C. A. Structural instability and relaxation in liquid and glassy phases near the fragile liquid limit. *J. Non-Cryst. Solids* **1988**, *102*, 205–221.

(2) Dawson, K. A.; Foffi, G.; Sciortino, F.; Tartaglia, P.; Zaccarelli, E. Mode-coupling theory of colloids with short-range attractions. *J. Phys.: Condens. Matter* **2001**, *13*, 9113.

(3) Boulougouris, G. C.; Frenkel, D. Novel Monte Carlo scheme for systems with short-ranged interactions. *J. Chem. Phys.* **2005**, *122*, 244106.

(4) Stillinger, F. H.; Weber, T. A. Hidden structure in liquids. *Phys. Rev. A: At., Mol., Opt. Phys.* **1982**, *25*, 978–989.

(5) Theodorou, D. N.; Suter, U. W. Detailed molecular structure of a vinyl polymer glass. *Macromolecules* **1985**, *18*, 1467–1478.

(6) Doliwa, B.; Heuer, A. Finite-size effects in a supercooled liquid. *J. Phys.: Condens. Matter.* **2003**, *15*, S849–S858.

(7) Theodorou, D. N. In *Principles of molecular simulation of gas transport in polymers*; Yampolskii, Y., Pinnau, I., Freeman, B. D., Eds. John Wiley: Hoboken, NJ, 2006; pp 47−92.

(8) Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal energy landscapes. *Nature* **1998**, *394*, 758–760.

(9) Becker, O.; Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.* **1997**, *106*, 1495–1517.

(10) Wales, D. J.; Doye, J. P. K.; Miller, M. A.; Mortenson, P. N.; Walsh, T. R. Energy landscapes: from clusters to biomolecules. *Adv. Chem. Phys.* **2000**, 115.

(11) Wales, D. J. Discrete path sampling. *Mol. Phys.* **2002**, *100*, 3285–3306.

(12) Wales, D. J. Calculating rate constants and committor probabilities for transition networks by graph transformation. *J. Chem. Phys.* **2009**, *130*, 204111.

(13) Boulougouris, G. C.; Theodorou, D. N. Dynamical integration of a Markovian web: A first passage time approach. *J. Chem. Phys.* **2007**, *127*, 084903.

(14) Jain, T. S.; de Pablo, J. J. Investigation of Transition States in Bulk and Freestanding Film Polymer Glasses. *Phys. Rev. Lett.* **2004**, *92*, 155505.

(15) Riggleman, R. A.; Douglas, J. F.; de Pablo, J. J. Characterization of the potential energy landscape of an antiplasticized polymer. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2007**, *76*, 011504.

(16) Papakonstantopoulos, G. J.; Riggleman, R. A.; Barrat, J. L.; de Pablo, J. J. Molecular plasticity of polymeric glasses in the elastic regime. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2008**, *77*, 041502.

(17) Souza, V. K. d.; Wales, D. J. Connectivity in the potential energy landscape for binary Lennard-Jones systems. *J. Chem. Phys.* **2009**, *130*, 194508.

(18) Souza, V. K. d.; Wales, D. J. Energy landscapes for diffusion: Analysis of cage-breaking processes. *J. Chem. Phys.* **2008**, *129*, 164507.

(19) Calvo, F.; Bogdan, T. V.; Souza, V. K. d.; Wales, D. J. Equilibrium density of states and thermodynamic properties of a model glass former. *J. Chem. Phys.* **2007**, *127*, 044508.

(20) Middleton, T. F.; Wales, D. J. Comparison of kinetic Monte Carlo and molecular dynamics simulations of diffusion in a model glass former. *J. Chem. Phys.* **2004**, *120*, 8134–8143.

(21) Wales, D. Some further applications of discrete path sampling to cluster isomerization. *Mol. Phys.* **2004**, *102*, 891–908.

(22) Sciortino, F. Potential energy landscape description of supercooled liquids and glasses. *J. Stat. Mech.: Theory Exp.* **2005**, P05015.

(23) La Nave, E.; Sastry, S.; Sciortino, F. Relation between local diffusivity and local inherent structures in the Kob-Andersen Lennard-Jones model. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2006**, *74*, 050501.

(24) Shell, M. S.; Debenedetti, P. G.; Panagiotopoulos, A. Z. A conformal solution theory for the energy landscape and glass transition of mixtures. *Fluid Phase Equilib.* **2006**, *241*, 147–154.

(25) Debenedetti, P. G.; Stillinger, F. H.; Shell, M. S. Model Energy Landscapes. *J. Phys. Chem. B* **2003**, *107*, 14434–14442.

(26) Debenedetti, P. G.; Stillinger, F. H. Supercooled liquids and the glass transition. *Nature* **2001**, *410*, 259–267.

(27) Crisanti, A.; Ritort, F. Inherent Structures, Configurational Entropy and Slow Glassy Dynamics. *J. Phys. Condens. Matter* **2002**, *14*, 1381–1395.

(28) Chowdhary, J.; Keyes, T. Energy Landscapes Composed of Continuous Intertwining Equipotential Ribbons. *J. Phys. Chem. B* **2004**, *108*, 19786–19798.

(29) Götze, W.; Sjögren, L. Relaxation processes in supercooled liquids. *Rep. Prog. Phys.* **1992**, *55*, 241–376.

(30) Götze, W.; Sjögren, L. The glass transition singularity. *Z. Phys. B: Condens. Matter* **1987**, *65*, 415.

(31) Götze, W. Aspects of structural glass transition. In *Liquids, Freezing and Glass Transition*; Hansen, J.-P., Levesque, D., Zinn-Justin, J., Eds. Elsevier Science Publishers: Amsterdam, The Netherlands, 1991; Vol. I, pp 287−503.

(32) Kob, W. Computer simulations of supercooled liquids and glasses. *J. Phys. Condens. Matter* **1999**, *11*, R85–R115.

(33) Heuer, A. Exploring the potential energy landscape of glass-forming systems: from inherent structures via metabasins to macroscopic transport. *J. Phys. Condens. Matter* **2008**, *20*, 373101.

(34) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1−39). *J. Am. Chem. Soc.* **2010**, *10*, 1021.

(35) Noe, F.; Schutte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19011–19016.

(36) Noe, F. Probability distributions of molecular observables computed from Markov models. *J. Chem. Phys.* **2008**, *128*, 13.

(37) Snow, C. D.; Nguyen, N.; Pande, V. S.; Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **2002**, *420*, 102–106.

(38) Krivov, S. V.; Karplus, M. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766–14770.

(39) Elmer, S. P.; Park, S.; Pande, V. S. Foldamer dynamics expressed via Markov state models. II. State space decomposition. *J. Chem. Phys.* **2005**, *123*, 7.

(40) Swope, W. C.; Pitera, J. W.; Suits, F. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B.* **2004**, *108*, 6571–6581.

(41) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and beta-hairpin peptide. *J. Phys. Chem. B* **2004**, *108*, 6582–6594.

(42) Bolougouris, G. C.; Theodorou, D. N. Probing subglass relaxation in polymers via a geometric representation of probabilities, observables, and relaxation modes for discrete stochastic systems. *J. Chem. Phys.* **2009**, *130*, 044905–7.

(43) Bolougouris, G. C.; Frenkel, D. Monte Carlo Sampling of a Markov Web. *J. Chem. Theory Comput.* **2005**, *1*, 389–393.

(44) Munro, L. J.; Wales, D. J. Defect migration in crystalline silicon. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 3969–3980.

(45) Munro, L. J.; Wales, D. J. Rearrangements of bulk face-centred cubic nickel modelled by a Sutton-Chen potential. *Faraday Discuss.* **1997**, 409–423.

(46) Henkelman, G.; Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys.* **1999**, *111*, 7010–7022.

(47) Barkema, G. T.; Mousseau, N. Event-Based Relaxation of Continuous Disordered Systems. *Phys. Rev. Lett.* **1996**, *77*, 4358.

(48) Tsalikis, D. G.; Lempesis, N.; Bolougouris, G. C.; Theodorou, D. N. On the role of inherent structures in glass-forming materials: I. The vitrification process. *J. Phys. Chem. B* **2008**, *112*, 10619–10627.

(49) Tsalikis, D. G.; Lempesis, N.; Bolougouris, G. C.; Theodorou, D. N. On the role of inherent structures in glass-forming materials: II. Reconstruction of the mean square displacement by rigorous lifting of the inherent structure dynamics. *J. Phys. Chem. B* **2008**, *112*, 10628–10636.

(50) Makeev, A. G.; Maroudas, D.; Panagiotopoulos, A. Z.; Kevrekidis, I. G. Coarse bifurcation analysis of kinetic Monte Carlo simulations: a lattice gas model with lateral interactions. *J. Chem. Phys.* **2002**, *117*, 8229–8240.

(51) Bolhuis, P. G.; Dellago, C.; Chandler, D. Sampling ensembles of deterministic transition pathways. *Faraday Discuss.* **1998**, *110*, 421–436.

(52) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.

(53) Voter, A. F.; Doll, J. D. Dynamical corrections to transition-state theory for multistate systems, Surface self-diffusion in the rare-event regime. *J. Chem. Phys.* **1985**, *82*, 80–92.

(54) Voter, A. Parallel replica method for dynamics of infrequent events. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1998**, *57*, R13985–R13988.

(55) Voter, A. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* **1997**, *78*, 3908–3911.

(56) Fukui, K. The Path of Chemical Reactions. The IRC Approach. *Acc. Chem. Res.* **1981**, *14*, 363–368.

(57) Helfand, E.; Wasserman, Z. R.; Weber, T. A. Brownian dynamics study of polymer conformational transitions. *Macromolecules* **1980**, *13*, 526–533.

(58) Helfand, E. Brownian dynamics study of transitions in a polymer chain ofbistable oscillators. *J. Chem. Phys.* **1978**, *69*, 1010–1018.

(59) Kob, W.; Andersen, H. C. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1995**, *51*, 4626.

(60) Kob, W.; Andersen, H. C. Scaling behavior in the beta relaxation regime of a supercooled Lennard-Jones mixture. *Phys. Rev. Lett.* **1994**, *73*, 1376–1379.

(61) Allen, M. P.; Tildesley, D. J. Advanced Simulation techniques. In *Computer simulation of liquids*, 1sted.; Clarendon Press: Oxford, U.K., 1987; pp 212−219.

(62) Appignanesi, G. A.; Rodríguez, Fris J. A.; Montani, R. A.; Kob, W. Democratic Particle Motion for Metabasin Transitions in Simple Glass Formers. *Phys. Rev. Lett.* **2006**, *96*, 057801–057804.

(63) Doliwa, B.; Heuer, A. Hopping in a supercooled Lennard-Jones liquid: Metabasins, waiting time distribution, and diffusion. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2003**, *67*, 030501.

(64) Denny, R. A.; Reichman, D. R.; Bouchaud, J. P. Trap models and slow dynamics in supercooled liquids. *Phys. Rev. Lett.* **2003**, *90*, 025503.

(65) Buechner, S.; Heuer, A. Metastable states as a key to the dynamics of glass forming Liquids. *Phys. Rev. Lett.* **2000**, *84*, 2168–2171.

(66) Mauro, J. C.; Loucks, R. J.; Gupta, P. K. Metabasin approach for computing the master equation dynamics of systems with broken ergodicity. *J. Phys. Chem. A.* **2007**, *111*, 7957–7965.

(67) Mauro, J. C.; Loucks, R. J. Selenium glass transition: A model based on the enthalpy landscape approach and nonequilibrium statistical mechanics. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2007**, *76*, 174202.

(68) Heuer, A. Exploring the potential energy landscape of glass-forming systems: from inherent structures via metabasins to macroscopic transport. *J. Phys.: Condens. Matter* **2008**, *20*, 373101.

(69) Press, H. W.; Teukolsky, A. S.; Vettering, T. W.; Flannery, P. B. *Numerical Recipes: The Art of Scientific ComputingNumerical Recipes: The Art of Scientific Computing*, 3rd ed.; Cambridge University Press: Cambridge, U.K., 2007.

(70) Grama, A.; Gupta, A.; Karypis, G.; Kumar, V. *Introduction to Parallel Computing - Design and Analysis of AlgorithmsIntroduction to Parallel Computing - Design and Analysis of Algorithms*. 2nd ed.; The Benjamin/Cummings Publishing Company: Redwood City, CA, 2003.

(71) Fox, G. C.; Johnson, M. A.; Lyzenga, G. A.; Otto, S. W.; Salmon, J. K.; Walker, D. W. Solving problems on concurrent processors. In *General techniques and regular problems*, Prentice-Hall, Inc.: Upper Saddle River, NJ, 1988; Vol. 1, p 592.

(72) Berry, R. S.; Breitengraserkunz, R. Topography and dynamics of multidimensional interatomic potential surfaces. *Phys. Rev. Lett.* **1995**, *74*, 3951–3954.

(73) Berry, R.; Wales, D. Freezing, melting, spinodals, and clusters. *Phys. Rev. Lett.* 1989.

(74) Donati, C.; Douglas, J. F.; Kob, W.; Plimpton, S. J.; Poole, P. H.; Glotzer, S. C. Stringlike cooperative motion in a supercooled liquid. *Phys. Rev. Lett.* **1998**, *80*, 2338–2341.

(75) *Materials and Processes Simulations - MAPS*, 3.1 ed.; Scienomics: Paris, 2009.

CT9004245