Self-learning metabasin escape algorithm for supercooled liquids

Penghui Cao, Minghai Li, Ravi J. Heugle, Harold S. Park, and Xi Lin

Phys. Rev. E **86**, 016710 — Published 24 July 2012

DOI: 10.1103/PhysRevE.86.016710

Submitted to PRE on July 6, 2012

Self-learning metabasin escape algorithm for supercooled liquids

Penghui Cao, Minghai Li, Ravi J. Heugle, Harold S. Park, and Xi Lin[§]

Department of Mechanical Engineering and Division of Materials Science and Engineering, Boston University, Boston, Massachusetts 02215, USA

[§]To whom all correspondence should be addressed. Email: linx@bu.edu.

Abstract

A generic history-penalized metabasin escape algorithm that contains no predetermined parameters is presented in this work. The spatial location and volume of imposed penalty functions in the configurational space are determined in self-learning processes as the 3*N*-dimensional potential energy surface is sampled. The computational efficiency is demonstrated using a binary Lennard-Jones liquid supercooled below the glass transition temperature, which shows an $O(10^3)$ reduction in the quadratic scaling coefficient of the overall computational cost as compared to the previous algorithm implementation. Furthermore, the metabasin sizes of supercooled liquids are obtained as a natural consequence of determining the self-learned penalty function width distributions. In the case of a bulk binary Lennard-Jones liquid at a fixed density of 1.2, typical metabasins are found to contain about 148 particles while having a correlation length of 3.09 when the system temperature drops below the glass transition temperature.

I. Introduction

Despite the tremendous amount of scientific research effort in the last three decades, the glass transition has remained one of the major unresolved problems in condensed matter theory[1,2,3]. This is mainly because critical dynamical properties, in particular the shear viscosity η of supercooled liquids which increases more than 17 orders of magnitude upon the glass transition[4], are not directly accessible by most glass transition theories. For example, the random first-order transition theory (RFOT)[5,6] expresses the free energy of a supercooled liquid F as a functional of its inhomogeneous density field $\rho(\mathbf{r})$ and seeks the nontrivial solutions to $\frac{\delta F[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} = 0$. Using a simple hard-sphere liquid model and an approximated density functional[7], Singh *et al.* was able to identify a metastable inhomogeneous glassy state when the system density is higher than a threshold value[8]. Kirkpatrick and Wolynes [5] argued that such a metastable inhomogeneous state corresponds to the non-decaying plateau of the density correlation function $\lim_{t\to\infty}\langle \delta\rho(\mathbf{r},t)\delta\rho(\mathbf{r}',0)\rangle \neq 0$ predicted by the mode coupling theory (MCT)[9,10]. More explicitly, Kirkpatrick and Thirumalai proved that the governing equations for the spin fluctuation in the mean-field p-spin (p> 2) models[11] are identical to the MCT equations [10], such that the non-decaying correlation function plateau value in MCT acts as an order parameter, the so-called Edwards-Anderson parameter[12], signaling the onset of metastable glassy states when the temperature drops below T_{MCT} [3,13]. Within the mean-field scope of these theories, however, the lifetime of the predicted metastable states is infinite and therefore the corresponding relaxation time diverges at T_{MCT}. Such a divergence is unphysical since the structural relaxation time $\tau = \eta/G_\infty$ measured by both

experiments and computations only increases for 4 orders of magnitude at T_{MCT} [3], where G_∞ is the instantaneous shear modulus [14].

To go beyond the mean-field approximation, and more importantly to address the connection between thermodynamic properties and viscosity, one often assumes the existence of an ideal glass transition at finite temperature, called the Kauzmann temperature T_K [15], at which the configurational entropy of metastable glassy states vanishes. Using phenomenological arguments that a supercooled liquid is composed of independent droplets of linear size ξ [16], one is able to make a qualitative estimation of the free energy and configurational entropy of these droplets as a function of ξ and argue that the thermodynamic temperature T_K is the same as the dynamical Vogel-Fucker-Tamman (VFT) temperature T_0 at which the extrapolated structural relaxation time diverges [3,6]. Since the static structural factor and dynamical density correlation functions are insufficient to describe supercooled liquids beyond the mean-field approximation, higher-order density correlation functions, such as the four-point and three-point correlation functions [12,17], were introduced to address the so-called dynamical heterogeneity [3,18]. Incorporating these complicated higher-order correlation functions into the glass transition theory implies that the density may not be an efficient, although systematic, probe for the structural relaxation events in supercooled liquids. It seems much more desirable to avoid these higher-order density correlation functions by directly expressing viscosity, the most critical dynamical property of supercooled liquids, as the stress tensor correlation function time integral [19].

Evaluating such a time integral near the glass transition temperature is impossible without efficient potential energy surface (PES) exploration algorithms. Since typical structural relaxation timescales are far beyond the timescales that are accessible in molecular dynamics (MD) simulations, one often resorts to the transition state theory to estimate the rates of these rare events [2,20]. Existing methods are able to follow a specific transition from a given initial state to a final state which is either given or unknown [21,22]. For example, the blue moon method [23] and meta-dynamics [24] require a pre-determined subspace of order parameters; the string method [25], including the nudged elastic band method [26], requires *a priori* knowledge of both the initial and final states; both the hyper-dynamics [21] and dimer methods, when coupled to the kinetic Monte Carlo method [27], spend most of their time re-crossing small barriers in a hierarchy of sequential cooperative events. Therefore, none of these methods have been successfully applied to study the complex potential energy landscapes in a vitrified fluid which is known to exhibit heterogeneous structures of metastable energy basins [2].

Kushima *et al.* [28] have recently developed a metadynamics-based [24] approach, called autonomous basin climbing (ABC), to explore the 3*N*-Dimensional (3*N*-D) PES of supercooled liquids and fed the obtained trajectories to a closed form of the Green-Kubo viscosity time integral [28,29]. Using the ABC algorithm, Kushima *et al.* were able to compute the entire viscosity spectrum over thirty decades [28,29] and provided the first topological connectivity of the 3*N*-D energy basins with clear distinctions between fragile behaviors and strong ones [30,31]. Their results further suggested that supercooled liquids

should not be categorized as strong or fragile liquids; instead, all liquids have characteristic fragile-to-strong crossover temperatures [30,31].

II. Self-learning Metabasin Escape Algorithm

The essential idea of the original ABC algorithm is to avoid sampling along the 3*N*-D PES as in MD simulations, but rather to utilize the extra dimension in energy so that metabasin escaping events can be identified and compared in the (3*N*+1)-D space. Their basic algorithm can be summarized as follows. For a given *N*-particle system with potential energy $E(\mathbf{r})$ where $\mathbf{r}$ specifies its 3*N*-D atomic configuration, exploration of the PES is first guided by the force $\mathbf{F} = -\partial E / \partial \mathbf{r}$. However, if the system follows only the force, it ends up at a local minimum and gets trapped. Therefore, penalty functions are imposed to assist the system in escaping from the local energy basin in a similar manner as metadynamics [24]. Explicitly, they first located a local minimum at $\mathbf{s}_1$ via the steepest decent quench [32] and then added a Gaussian penalty function $\phi_1(\mathbf{r}-\mathbf{s}_1 \mid \mathbf{w}, h) = h \exp\left[-\frac{(\mathbf{r}-\mathbf{s}_1)^2}{2\mathbf{w}^2}\right]$ around $\mathbf{s}_1$ in the (3*N*+1)-D space where *h* is the 1D penalty height and $\mathbf{w}$ is the 3*N*-D penalty half-width. After each penalty function addition, an energy minimum search was performed on the augmented energy surface $\Psi^p(\mathbf{r}) = E(\mathbf{r}) + \sum_{i=1}^{p} \phi_i(\mathbf{r})$, which is the sum of the original potential energy and all the previously imposed penalty potentials. Using this approach, one does not need to specify the softest eigenmode searching direction as in the hyperdynamics [21] or dimer methods [27], or to restrict the searching subspace as in metadynamics [24]. By repeating the alternating sequence of penalty function addition and

augmented energy relaxation, the system was activated to fill up the local minimum basin and escaped through the lowest saddle point. By keeping all the penalty functions imposed during the simulation, they eliminated frequent re-crossing of small barriers, which is a significant advantage of such history-penalized methods [24,33].

In spite of the demonstrated successes in supercooled liquids [28-31] and also in a few other extremely slow dynamical processes such as creep [34] and void nucleation and growth [35], the original ABC algorithm developed by Kushima *et al*. [28] has two notable shortcomings. First, since the ABC method, like classical MD simulations, requires only calculations of the energy and forces, its computational expense should in principle be comparable to MD so that large atomistic systems containing 10^5 particles or more can be routinely modeled. Unfortunately, the current ABC algorithm suffers greatly from the increasing number of imposed penalty functions and therefore the extra computational load over a regular MD simulation increases dramatically as the simulation progresses. This has been limiting typical ABC trajectories implemented in various applications[28-31,34,35] to a few thousand local minima and saddle points for a system containing a few hundred atoms in 3D. Although small binary Lennard-Jones (b-LJ) systems containing as few as 150 particles[36] are sufficiently enough to obtain the inherent structures and diffusivity above T_g, the ABC trajectories of these small systems starting from deeply supercooled configurations do not fully relax to similar initial energies[28,29] . This is because the activated particles in a typical metabasin, which contains a few hundred b_LJ particles (see below), are inevitably overlapped with their periodic boundary condition (PBC) images. As to be demonstrated below, such overlapping-metabasin trajectories near and

below T_g can be symmetrically reduced to independent activation processes as the system size increases so that the correlation length distributions of an individual metabasin can be measured.

The second inevitable problem of the ABC implementation concerns the subtle choice of the penalty function parameters $\{h, \mathbf{w}\}$. In the glass transition case, Kushima *et al.* [28] followed the Lindemann criterion [2] to choose a constant Gaussian penalty half-width $\sqrt{2}|\mathbf{w}| = 0.447$ for the supercooled binary Lennard Jones (b-LJ) liquid [37]. However, for arbitrarily given materials systems, it is unclear whether such constant parameter values would still be relevant, or even exist. Moreover, since complex systems that exhibit slow dynamics naturally have a hierarchical distribution of metabasins with different sizes and energy depths, it is unlikely that the relaxation events can all be faithfully captured using the same constant parameters.

This work intends to address these two seemingly different problems, i.e. the bottleneck in the computational efficiency and the subtle issue of choosing the constant parameter values, under a unified scheme. The essential idea is to allow the system to learn by itself how the penalty functions should be applied as the trajectory evolves without any pre-specified conditions. Moreover, as a natural consequence of these self-learning strategies, the metabasin correlation lengths of supercooled liquids will be identified. Explicitly:

1) We start from a given initial configuration, find the corresponding local minimum on the PES at $\mathbf{s}_1$, and then apply the first penalty function with a small random half-width and height since there is no history yet. Or, one may follow a short

MD trajectory to generate the penalty function as in metadynamics [24], which is equivalent to small random penalty functions since atoms in supercooled liquids are essentially immobile within the accessible MD timescale. We refer to this as the random rule. The other rule, to be specified below, is the self-learning rule after new local PES minima being identified. The random rule is set to be the current rule.

2) We check whether the current penalty function overlaps with any of the previous ones according to the criterion specified below in Eq. (1). Namely, for any two penalty functions $\phi(\mathbf{r}-\mathbf{s}_i \,|\, |\mathbf{w}_i| < w_{\max}, h_i < h_{\max})$ and $\phi(\mathbf{r}-\mathbf{s}_j \,|\, |\mathbf{w}_j| < w_{\max}, h_j < h_{\max})$, if the *3N*-D distance between their centers is smaller than the sum of their half-widths,

$$2|\mathbf{s}_i - \mathbf{s}_j| \le |\mathbf{w}_i| + |\mathbf{w}_j|, \tag{1}$$

these two closely positioned penalty functions will be replaced by a combined penalty function $\phi(\mathbf{r}-\mathbf{s}_n \,|\, \mathbf{w}_n, h_n)$ with

$$h_n = h_i + h_j \tag{2}$$

and

$$|\mathbf{w}_n| = \max\left[\frac{|\mathbf{s}_i - \mathbf{s}_j| + |\mathbf{w}_i| + |\mathbf{w}_j|}{2}, |\mathbf{w}_i|, |\mathbf{w}_j|\right], \tag{3}$$

while the center of the new penalty function will be shifted to

$$\mathbf{s}_n = \rho_i \mathbf{s}_i + \rho_j \mathbf{s}_j \tag{4}$$

where the weights

$$\rho_i = \frac{|\mathbf{w}_i| h_i}{|\mathbf{w}_i| h_i + |\mathbf{w}_j| h_j} \tag{5}$$

and $\rho_j = 1 - \rho_i$. One may refer to Figs. 1(a-c) for three illustrative examples and detailed discussions below. We then label the combined penalty function as the current penalty function and the total number of penalty functions is reduced by one. Step 2) is repeated until the current penalty function does not overlap with any of the previous ones.

3) We then minimize the augmented potential energy $\Psi^p(\mathbf{r}) = E(\mathbf{r}) + \sum_{i=1}^{p} \phi_i(\mathbf{r})$ to locate the next local minimum at $\mathbf{s}_{p+1}$ on this augmented energy surface.

4) We check whether $\mathbf{s}_{p+1}$ is a true local minimum on the original PES, which satisfies the following two criteria: vanishing force and non-zero distance away from all the previous true local minima.

   a. If yes, we backtrack along the last minimization path to locate the saddle point configuration $\mathbf{s}_{\text{sad}}$. We then update the current penalty function selection rule using the displacement and the energy difference between the new local minimum and the corresponding saddle point as

$$\left|\mathbf{w}_{p+1}\right| = \frac{2}{3}\left|\mathbf{s}_{\text{sad}} - \mathbf{s}_{p+1}\right| \tag{6}$$

   and

$$h_{p+1} = E(\mathbf{s}_{\text{sad}}) - E(\mathbf{s}_{p+1}). \tag{7}$$

   One may refer to Figs. 1(d-f) for the illustrative examples and detailed discussions below.

   b. If no, we keep the current penalty function selection rule.

In either case, we apply a new penalty function $\phi(\mathbf{r} - \mathbf{s}_{p+1} \mid \mathbf{w}_{p+1}, h_{p+1})$ using the current rule and the total number of penalty functions is increased by one.

5) Repeat Steps 2) to 4) until a sufficiently large configurational space has been sampled.

Compared to the original ABC approach [28], the new algorithm outlined above introduces two designed self-learning strategies. First, the new algorithm constantly searches for redundant penalty functions and replaces them with more effective ones. Such a combination strategy is crucial to maintain a minimum amount of penalty functions while effectively occupying the penalized configurational subspace. A few scenarios occur frequently in the metabasin filling processes, which we now discuss in detail.

One special case illustrated in Figure 1(a) concerns two sequential local minima $\mathbf{s}_p$ and $\mathbf{s}_{p+1}$ on the augmented energy surfaces $\Psi^p(\mathbf{r}-\mathbf{s}_p)$ and $\Psi^{p+1}(\mathbf{r}-\mathbf{s}_{p+1})$ respectively are extremely close to each other. This often occurs right after a true local minimum on the PES is identified but the magnitude of the local curvature of the applied penalty function $\phi_p"(\mathbf{r}-\mathbf{s}_p)$ (green curve) is smaller than $E"(\mathbf{r}-\mathbf{s}_p)$. In this case, the computational efficiency critically depends on the local curvature of the augmented energy surface at $\mathbf{s}_p$. When two virtually identical penalty functions (green and blue curves) are caught in sequence, the rules specified in Eqs. (2) and (3) delete both of them and create a new penalty function (red curve) with a doubled height and the same half-width so that the local curvature of the penalty function doubles. This combination process repeats until the augmented energy surface curves downwards, pushing the system away from the trapped configuration.

Figure 1(b) illustrates the case of maximal separation between two penalty functions with identical half-widths. As specified in Eq. (1), combination occurs only when the center of one penalty function (blue curve) lies within a half-width distance away from the other center (green curve). After combination, the new penalty function (red curve) is centered halfway between the two original centers as specified in Eqs. (4-5) with a doubled height and 3/2-times half-width as specified in Eqs. (2-3). Such a 3/2 pre-factor justifies Eq. (6), which will be discussed in detail below.

The third case concerns the combination of one large penalty function and one small one. As shown in Figure 1(c), the small penalty function is essentially absorbed by the large one. Finally, we note that in our simulations the half-width upper limit $w_{\max}$ is set to be half of the simulation box and the penalty height maximum $h_{\max}$ is set to be one order of magnitude higher than the highest energy barrier. There are no combinations allowed for penalty functions that are larger than these limits to avoid unphysical basin filling processes.

The second self-learning strategy concerns the initialization of new penalty function parameters after a true local minimum on the original PES is identified, by measuring the displacement and energy difference between the true local minimum and the corresponding saddle point. Figure 1(d) plots the self-learned penalty function $\phi_{p+1}(\mathbf{r})$ with a half-width of $|\mathbf{w}_{p+1}| = |\mathbf{s}_{\text{sad}} - \mathbf{s}_{p+1}|$ and a height of $h_{p+1} = E(\mathbf{s}_{\text{sad}}) - E(\mathbf{s}_{p+1})$ on top of the original PES $E(\mathbf{r})$. The augmented energy surface Ψ^p is plotted in Figure 1(e), showing a small undesirable dip at $\mathbf{s}_{p+1}$. To enforce fast relaxations away from $\mathbf{s}_{p+1}$, one needs to

lower the local curvature at $\mathbf{s}_{p+1}$. This can be done by either increasing the height or decreasing the half-width of the applied penalty function, where the latter is preferred for better accuracy. If one assumes deep energy basins of a supercooled liquid follow the same sinusoid shape in the vicinity of a local minimum as in crystalline materials [38], it is straightforward to prove that the augmented energy surface will curve down with a critical width $|\mathbf{w}_{p+1}| = 0.9|\mathbf{s}_{\text{sad}} - \mathbf{s}_{p+1}|$ using the quartic penalty function to be discussed below. Eq. (6) takes a smaller pre-factor of $|\mathbf{w}_{p+1}| = \frac{2}{3}|\mathbf{s}_{\text{sad}} - \mathbf{s}_{p+1}|$ to further decrease the local curvature of $\Psi^{p+1}(\mathbf{r} - \mathbf{s}_{p+1})$ as plotted in Figure 1(f), and also to match the combination rule of Eq. (4) such that penalty functions with the targeted half-width $|\mathbf{w}_{p+1}| = |\mathbf{s}_{\text{sad}} - \mathbf{s}_{p+1}|$ may be restored after combination.

III. Results and Discussions

To test the performance of the self-learning energy basin filling algorithm, a standard supercooled b-LJ liquid at a constant reduced density of 1.2 and a 4:1 ratio for the A:B particles was used [37]. The b-LJ model used the LJ potential $V_{\alpha\beta}(\mathbf{r}) = 4\varepsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{\mathbf{r}}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{\mathbf{r}}\right)^{6}\right]$ with ε_{AA}= 1.0, σ_{AA}= 1.0, ε_{AB} =1.5, σ_{AB}= 0.8, ε_{BB}= 0.5, and σ_{BB}= 0.88, truncated and shifted at cutoff distances of $2.5\sigma_{\alpha\beta}$ [37]. All the former and subsequent potential related parameters and values are given in the reduced LJ units. Supercooled liquids were prepared via slow annealing from T= 2.0 to 10^{-5} at five different constant cooling rates 1×10^{-4}, 8×10^{-5}, 1×10^{-5}, 4×10^{-6}, and 4×10^{-7} [39]. Independent

supercooled configurations were collected, among which the highest and lowest energies are -7.632 and -7.716 per particle, respectively. Since the average energy per particle -7.653 (±0.017) corresponds to $1.09T_{\mathrm{MCT}}$ based on the monotonic temperature to averaged local minimum energy $T \to \overline{E}(T)$ mapping [28,39], these independent supercooled configurations effectively cover the temperature region from T_{MCT} and above to T_{g} and below. Here the mode coupling temperature $T_{\mathrm{MCT}} = 0.435$ [37] and $T_{\mathrm{g}} \approx 0.37$ [28], since the latter used slightly different b-LJ potential radius cutoffs [28].

Quartic penalty functions $\phi(\mathbf{r}-\mathbf{s} \mid h, \mathbf{w}) = h\left[1-\left(\frac{\mathbf{r}-\mathbf{s}}{\mathbf{w}}\right)^2\right]^2$ were used in this study to replace the Gaussian functions used in the original ABC algorithm [24,28], because the former have naturally vanishing energy and forces at $|\mathbf{r}-\mathbf{s}| = |\mathbf{w}|$ so that the radial truncations are no longer required.

The overall computational efficiency of the self-learning algorithm is summarized in Fig. 2, with direct comparisons to the original ABC results. Fig. 2(a) shows the total computational time cost t as a function of the number of local minima identified N_{s} for a b-LJ liquid with N= 256 particles. In 50 hours on a single CPU, the self-learning algorithm found an average of 758 (±49) new local minima, as compared to an average of 13 (±4) using the original ABC method. Since independent penalty functions were all kept throughout the entire simulation to avoid barrier re-crossing events, the computational time spent in finding the next local minimum should scale proportionally with the total number of penalty functions applied. This implies that the total

computational time cost t should scale quadratically with the total number of local minima N_s. A quadratic fit to the 16 independent trajectories (grey symbols, 6 for ABC and 10 for self-learning) shown in Fig. 2(a) gave $\frac{t}{N_s^2} = 9.6\times10^{-2}$ and 8.5×10^{-5} hours for the original ABC and new self-learning algorithms, respectively. Such an $O(10^3)$ reduction in the quadratic scaling coefficient of the overall computational cost found for the N= 256 case was consistent for smaller and larger system sizes, as shown in Fig. 2(b). In particular, for the b-LJ liquid with N= 10^3 particles, only one or two local minimum were found in 50 hours using the original ABC algorithm such that a meaningful quadratic fit similar to Fig. 2(a) could not be performed.

In addition to the significantly larger number of local minima that can be identified (Fig. 2), the computational efficiency enhancement should also give insight into the configurational subspace that has been visited. This is particularly necessary for supercooled liquids, because finding a large number of local minima within a small configurational subspace does not necessarily lead to any metabasin escaping events. Therefore, we plot in Fig. 3 the topological connectivity trees of all the local minima and saddle points constructed from two representative (out of sixteen in total) basin-filling trajectories shown in Fig. 2(a), with their energy values scaled against the supercooling trajectories shown in Fig. 3(a). Specifically, Fig. 3(b) is for ABC and Fig. 3(c) is for self-learning, where both started from the same supercooled configuration with an energy of -7.716 per particle, obtained from a slow annealing at a constant cooling rate of 4×10^{-7} as shown in Fig. 3a. Within 50 hours using one CPU, the original ABC trajectory visited a PES subspace covering the maximal saddle point energy of -7.702 per particle and the

maximal mean square displacement of 0.006 per particle. In contrast, the self-learning algorithm reached much further, finding the maximal saddle point energy of -7.435 per particle, which is 20 times higher in the energy per particle, and the maximal mean square displacement of 6.135 per particle, which is 1021 times larger in distance per particle. Therefore, through the self-learning strategies outlined above, the new algorithm not only avoids using pre-determined constant ABC parameters, but also explored a significantly larger subspace of the PES than the original ABC algorithm by climbing over much higher saddle points (Fig. 3) and reaching out to many more local minima (Fig. 2).

With such a dramatically enhanced computational efficiency, we can probe a few critical structural relaxation processes occurring in supercooled liquids, in particular the metabasin correlation length distribution when a bulk supercooled liquid approaches the glass transition temperature. While results obtained via using the original ABC method did quantitatively predict the structural relaxation time scales over 30 orders of magnitude[28,30], it is unclear how the dynamic correlation lengths would vary as supercooled liquids approach their glass transition temperatures. Using a generalized point-to-set correlation method, Kob *et al.* was able to obtain the dynamic correlation length for a quasi-hard sphere system down to the mode coupling temperature T_{MCT} [40], but not significantly below due to the timescale limitation in MD simulations.

Fig. 4(b) shows the penalty function half-width distributions (red symbols) obtained from the same 10 independent self-learning trajectories of Fig. 2(a) for the N= 256 case.

Consistent agreement is found among these converged distribution functions obtained from completely independent trajectories. We recall that a metabasin can be thought of as a subspace of the PES in which a supercooled liquid will reside for a long time, but will likely not revisit it after escape events. Within the context of the self-learning algorithm, when the same configurational subspace is visited over and over again, it triggers the self-learned combination condition as specified by Eq. (1) so that the widths of the penalty functions continue to grow. However, these self-learned combination processes will cease as soon as the system escapes from the current metabasin, leaving behind the penalty functions whose widths represent the actual size of the metabasin. In other words, the self-learned penalty function width distributions shown in Fig. 4(b) should directly correspond to the size distributions of metabasins.

As discussed above, the metabasin activation processes in small systems such as N= 256 or less are strongly coupled to their PBC images. Consequently, the activated particles during the metabasin escape processes do not automatically relax to deeply supercooled states, so that these self-learning trajectories consist of many states of high energy. These undesirable overlapping activation processes can be systematically reduced by increasing the system size N, as shown by the connectivity trees in Fig. 4(b) for N= 500 and in Fig. 4(c) for N= 1000. In particular, the self-learning trajectories for the N= 1000 case explore deeply supercooled states with occasional activation processes associated with individual metabasin escape events.

The gradual decoupling of overlapped metabasins can also been seen in Fig. 4(d), where the penalty function half-width distribution obtained from the N= 500 (yellow symbols) and 10^3 cases (green symbols) is shifted towards larger values as compared to the N= 256 case (green symbols). In particular, the peak of the penalty function half-width distribution profile, namely the most probable metabasin size ξ, is increased from 1.88 (± 0.17) for N= 256, 2.20 (± 0.14) for N= 500, to 2.35 (±0.13) for N= 1000. Fig. 4(e) summarizes the most probable metabasin correlation length, the peak of the penalty function width distribution profile of Fig. 4(d), as a function of N.

To obtain the scaling rule for the metabasin correlation length shown in Fig. 4(e), it is important to note a few characteristic features of supercooled liquids revealed by using our history-penalized basin filling approach that are fundamentally different from the RFOT results [6]. First, the fragility of the b-LJ liquids is a natural consequence of the fact that after being trapped in a deeper energy basin, the system requires a higher activation energy to escape. Namely, there are no small activation energy events available in the connectivity tree structures of Figs. 4(a-c) that can connect a deep energy basin to another deep energy basin of similar depth. In contrast, the connectivity tree structures of SiO_2, which is known to be a "strong" liquid [4], are replete with escape mechanisms that have similarly small activation energies [30]. Such a direct energy landscape origin of fragility does not require the configurational entropy cost assumption as in the Adam-Gibbs theory [41] and in the RFOT [42], so that the configurational entropy crisis at the Kauzmann temperature T_K does not have to be the ultimate destiny for supercooled liquids.

Second, from any supercooled liquid basin, no matter how deep it is, the activation energy required to cross the bottleneck barrier to the rest of the PES is a finite value. From the perspective of the transition state theory for rare events, a finite activation energy corresponds to a finite relaxation time, with the only exception being the crystalline state where there are no other symmetry-distinct states that are degenerate in energy. Therefore, divergence in the structural relaxation time of supercooled liquids cannot occur at non-zero thermodynamic Kauzmann temperature T_K or non-zero dynamic VFT temperature T_0. Namely, the ideal glass transition should not exist at a finite temperature. Instead the diverging VFT extrapolation of the structural relaxation time, the so-called fragile behavior, is avoided by a natural crossover to the strong behavior of constant activation energy, since all energy basins have finite depths. In other words, the apparent configurational entropy crisis only exists within the metabasin, but not after metabasin escape events. From such an energy landscape perspective, the fragile-to-strong crossover is inevitable for all supercooled liquids.

Third, a finite relaxation time corresponds to a finite correlation length, so that the collective motion during any metabasin relaxation process must be a truly localized event. Namely, a metastable basin possesses an exponentially decaying tail in real space.

When the metabasin correlation length shown in Fig. 4(e) is fitted using an exponentially decaying function (solid line), an asymptotic correlation length of $\xi = 3.09$ is found. Therefore, in a macroscopic sample of the b-LJ liquid, the collective motion of particles in a typical metabasin relaxation event has a finite correlation length of 3.09, which

corresponds to a volume of 148 b-LJ particles. Using the $T \rightarrow \overline{E}(T)$ mapping[28,39] of Fig. 3(a), we conclude that a typical metabasin relaxation event consists of the collective motion of 148 particles when the bulk b-LJ liquid reaches below the glass transition temperature T_g. Such a metabasin correlation length is of the same order as the phenomenoglical $\xi \approx 4.5$ [42] or 5.8 [6] by the RFOT, the assumed value of about 100 particles for 31 different types of metallic glasses [43], the computed value of about 140 particles in a MD simulation using an embedded atom method force field [44], and the measured 3±1 nm for poly(vinyl acetate) at 10 K above T_g by nuclear magnetic resonance [45]. We note that although our computed metabasin correlation length is the same order of magnitude as the value phenomenoglically predicted by the RFOT, the former is a finite number and the latter diverges at a non-zero Kauzmann temperature T_K.

IV. Summary

In summary, a generic self-learning algorithm is presented in this work that is capable of capturing escape events from metabasins in supercooled liquids. The computational cost of the original ABC algorithm is significantly reduced via the new self-learning strategies developed in this work for the following two reasons. First, the computational load decreases because the total number of penalty functions is reduced through penalty function combinations. Second, self-learned combinations create flexibility in the penalty function widths, which then naturally self-adapt to the underlying metabasin energy landscape. The resulting variation in the penalty function widths represents the actual size distribution of metabasins in supercooled liquids. As a generic approach involving

only energy and force calculations, this self-learning algorithm may offer a new efficient computational tool for attacking many phenomena of long-standing scientific interest involving extremely slow dynamics in condensed matters.

Acknowledgements. The authors would like to acknowledge thoughtful discussions with Sidney Yip and Akihiro Kushima. This work was partially supported by Honda R&D Co., Ltd, NSF under the grant No. CMMI-1036460, and NSF-XSEDE under Grand No. DMR-0900073. PC gratefully acknowledges the support of a Dean's Fellowship from Boston University.

References:

[1] P. W. Anderson, Science **267** (5204), 1609 (1995); C. A. Angell, Science **267**, 1924 (1995).

[2] F. H. Stillinger, Science **267**, 1935 (1995).

[3] L. Berthier and G. Biroli, Reviews of Modern Physics **83** (2), 587 (2011).

[4] C. A. Angell, Journal of Physics and Chemistry of Solids **49** (8), 863 (1988).

[5] T. R. Kirkpatrick and P. G. Wolynes, Physical Review A **35** (7), 3072 (1987).

[6] V. Lubchenko and P. G. Wolynes, Annual Review of Physical Chemistry **58**, 235 (2007).

[7] T. V. Ramakrishnan and M. Yussouff, Physical Review B **8** (5), 2775 (1979).

[8] Y. Singh, J. P. Stoessel, and P. G. Wolynes, Physical Review Letters **54** (10), 1059 (1985).

[9] U. Bengtzelius, W. Gotze, and A. Sjolander, Journal of Physics C: Solid State Physics **17**, 5915 (1984); S. P. Das, Reviews of Modern Physics **76**, 785 (2004).

[10] E. Leutheusser, Physical Review A **29** (5), 9 (1984).

[11] J. Gross and M. Mezard, Nuclear Phyiscs B **240**, 431 (1984).

[12] S. F. Edwards and P. W. Anderson, Journal of Physics F: Metal Physics **5**, 965 (1975).

[13] T. R. Kirkpatrick and D. Thirumalai, Physical Review Letters **58** (20), 2091 (1987).

[14] J. C. Dyre, Reviews of Modern Physics **78**, 953 (2006).

[15] A. W. Kauzmann, Chemical Review **43**, 219 (1948).

[16] T. R. Kirkpatrick, D. Thirumalai, and P. G. Wolynes, Physical Review A **40** (2), 1045 (1989).

[17] L. Berthier, G. Biroli, J.-P. Bouchaud, L. Cipelletti, D. E. Masri, D. L'Hôte, F. Ladieu, and M. Pierno, Science **310**, 1797 (2005).

[18] M. D. Ediger, Annual Review of Physical Chemistry **51**, 99 (2000).

[19] R. Zwanzig, *Nonequilirium statistical mechanics*. (Oxford University Press, Oxford, 2001).

[20] A. Heuer, Journal of Physics: Condensed Matter **20**, 373101 (2008); D. J. Wales and H. A. Scheraga, Science **285** (5432), 1368 (1999); D. J. Wales, International Reviews in Physical Chemistry **25** (1-2), 237 (2006).

[21] A. F. Voter, Physical Review Letters **78** (20), 3908 (1997).

[22] R. Elber and M. Karplus, Chemical Physics Letters **139** (5), 375 (1987); G. Henkelman, G. Johannesson, and H. Jónsson, in *Progress on Theoretical Chemistry*, edited by S. D. Schwartz (Kluwar Academic Publishers, 2000), pp. 269.

[23] E. A. Carter, G. Ciccottia, J. T. Hynesa, and R. Kapralb, Chemical Physics Letters **156**, 123 (1989).

[24] A. Laio and M. Parrinello, Proceedings of the National Academy of Sciences of the United States of America **99** (20), 12562 (2002).

[25] W. E, W. Ren, and E. Vanden-Eijnden, Physical Review B **66**, 052301 (2002).

[26] H. Jonsson, G. Mills, and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne (World Scientific, 1998), pp. 385.
[27] G. Henkelman and H. Jónsson, Journal of Chemical Physics **111** (15), 7010 (1999).
[28] A. Kushima, X. Lin, J. Li, J. Eapen, J. C. Mauro, X.-F. Qian, D. Phong, and S. Yip, Journal of Chemical Physics **130**, 224504 (2009).
[29] J. Li, A. Kushima, J. Eapen, X. Lin, X. Qian, J. Mauro, P. Diep, and S. Yip, PLoS One **6** (3), e17909 (2011).
[30] A. Kushima, X. Lin, J. Li, X.-F. Qian, J. Eapen, J. C. Mauro, D. Phong, and S. Yip, Journal of Chemical Physics **131**, 164505 (2009).
[31] A. Kushima, X. Lin, and S. Yip, Journal of Physics: Condensed Matter **21**, 504104 (2009).
[32] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[33] F. Wang and D. P. Landau, Physical Review Letters **86** (10), 2050 (2001).
[34] T. T. Lau, A. Kushima, and S. Yip, Physical Review Letters **104**, 175501 (2010).
[35] Y. Fan, A. Kushima, S. Yip, and B. Yildiz, Physical Review Letters **106**, 125501 (2011).
[36] W. Kob, Journal of Physics: Condensed Matter **11**, R85 (1999); G. A. Appignanesi, J. A. R. Fris, R. A. Montani, and K. W., Physical Review Letters **96**, 4 (2006).
[37] W. Kob and H. C. Andersen, Physical Review E **51** (5), 4626 (1995).
[38] R. Peierls, Proc. R. Soc. London **52**, 34 (1940).
[39] S. Sastry, P. G. Debenedetti, and F. H. Stillinger, Nature **393**, 554 (1998).
[40] W. Kob, S. Roldan-Vargas, and L. Berthier, Nature Physics **8**, 164 (2012).
[41] G. Adam and J. H. Gibbs, Journal of Chemical Physics **43** (1), 139 (1965).
[42] X. Xia and P. G. Wolynes, Proceedings of National Academy of Sciences of the United States **97** (7), 2990 (1999).
[43] W. L. Johnson and K. Samwer, Physical Review Letters **95**, 4 (2005).
[44] S. G. Mayr, Physical Review Letters **97**, 4 (2006).
[45] U. Tacht, M. Wilhelm, A. Heuer, H. Feng, K. Schmidt-Rohr, and H. W. Spiess, Physical Review Letters **81** (13), 2727 (1998).

Figures

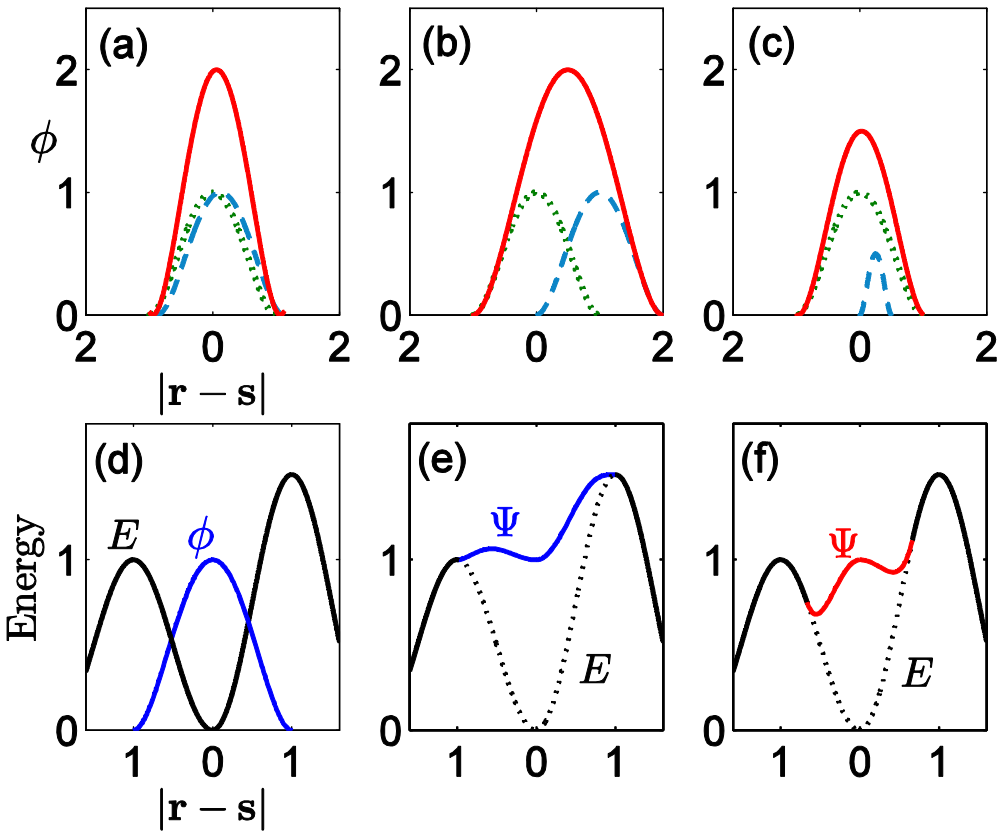

Figure 1: (Color online) Self-learning strategies on the (a-c) combination and (d-e) initial choice of penalty functions. (a) Two sequential fully overlapped penalty functions (green dotted and blue dashed) give a strong indication of inefficient sampling. The combined penalty function (red solid) doubles the local curvature to assist the system in moving away from the stuck configuration. (b) Combination of two penalty functions (green dotted and blue dashed) at the maximal separation. The new penalty function (red solid) has a half-width that is 3/2 times of the original values. (c) Combination between two penalty functions of different sizes. (d) The original energy E and the penalty function ϕ with self-learned half-width $|\mathbf{w}| = |\mathbf{s}_{\text{sad}} - \mathbf{s}|$ and height $h = E(\mathbf{s}_{\text{sad}}) - E(\mathbf{s})$. (e) Their

augmented energy Ψ still has a dip at the original local minimum. (f) The augmented energy Ψ using a smaller half-width $|\mathbf{w}| = \frac{2}{3}|\mathbf{s}_{\text{sad}} - \mathbf{s}|$ curves downwards, which is desirable for fast relaxations.

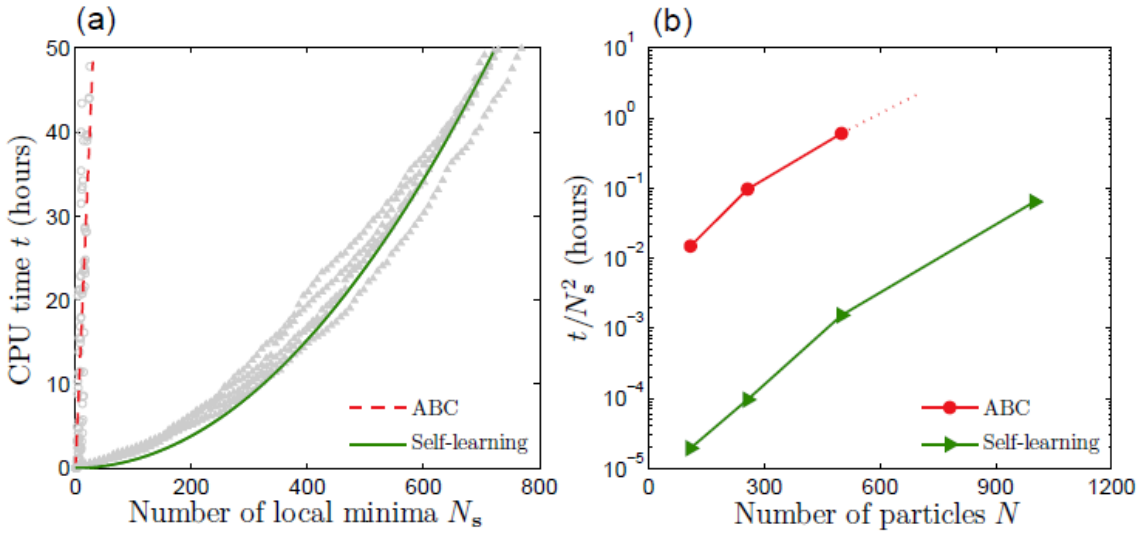

Figure 2: (Color online) Computational cost reduction via using the new self-learning algorithm, as compared to using the original ABC approach. (a) For a b-LJ liquid containing N= 256 particles, the total computational time cost t is plotted against the total number of local minima for each approach. Grey symbols are raw data, shown as circles for 6 independent ABC runs and triangles for 10 independent self-learning runs. Both sets of raw data were separately fitted by quadratic functions, red dashed line for the ABC and green solid line for the self-learning algorithms. (b) The quadratic scaling coefficients $\frac{t}{N_s^2}$ are plotted as a function of the system size for N= 108, 256, 500, and 1000 particles. Note that for N= 1000, the original ABC algorithm fails to generate enough data for meaningful quadratic fits. All computations in this work are performed on one dual-core AMD Opteron processor at 2.2 GHz.

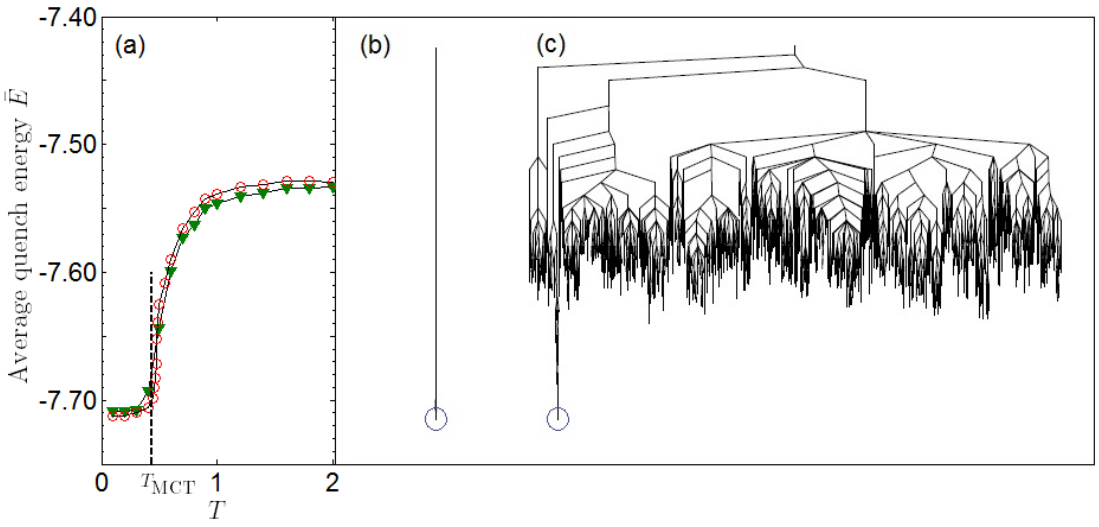

Figure 3: (Color online) (a) Average quench energy per particle $\overline{E}(T)$ as a function of temperature T obtained from slow annealing at a constant quenching rate of 4×10^{-7}. This defines the $T \rightarrow \overline{E}(T)$ mapping [28,39] at the given quenching rate, essentially identical for N= 256 (red circles) and N= 1000 (green triangles). Topological connectivity tree structures consisting of the PES local minima and saddle points for an N= 256 b-LJ liquid, obtained from two basin filling trajectories starting from the same initial deeply trapped potential energy state (blue circles), one using (b) the original ABC algorithm and the other using (c) the new self-learning algorithm. The lower end point of each vertical line, or a leaf, represents an independent local minimum. The connecting point between any two local minima represents the saddle point of the minimum activation energy between them.

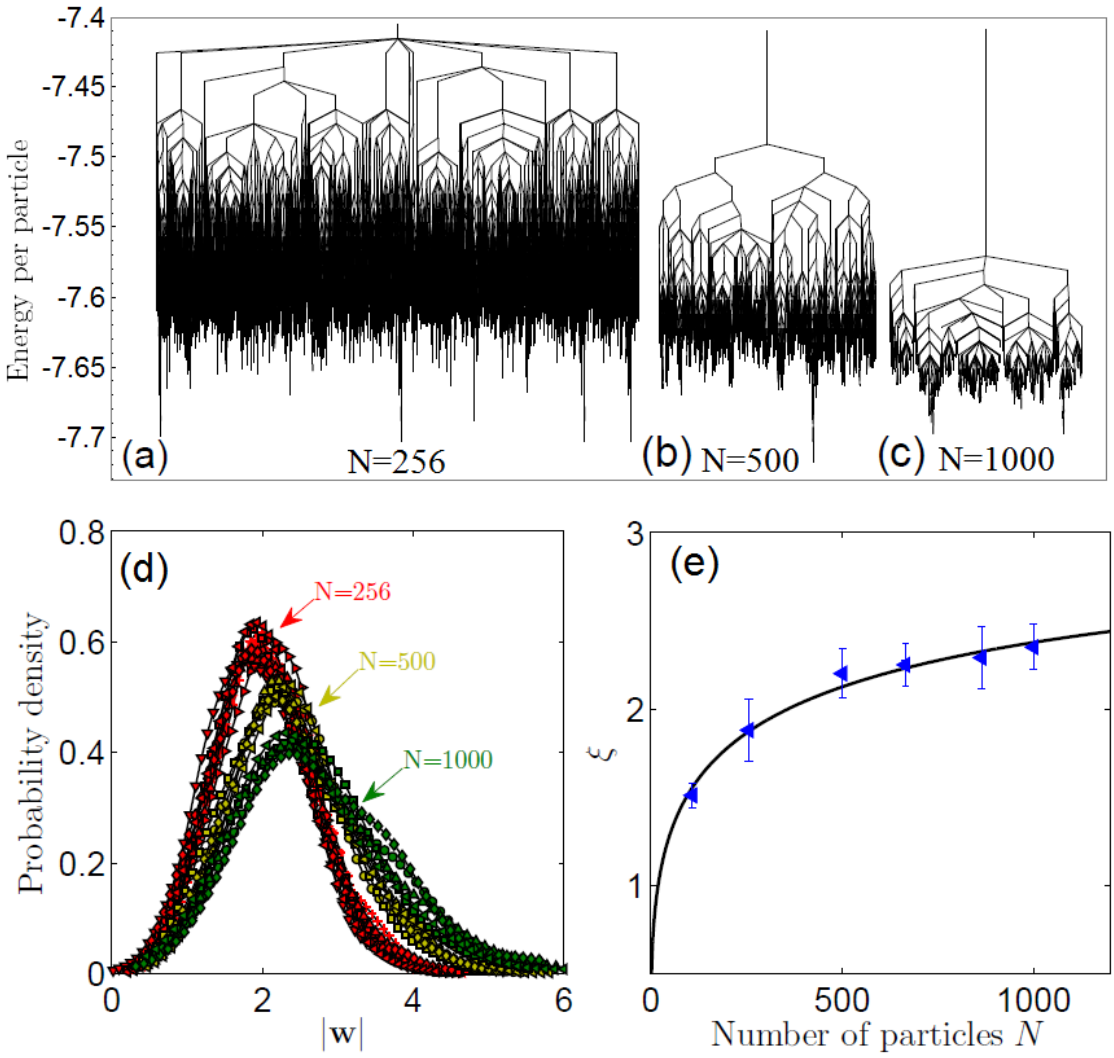

Figure 4: (Color online) Topological connectivity tree structures (a-c) and distribution of the self-learned penalty function half-widths |**w**| (d) for supercooled b-LJ liquids with N= 256, 500, and 1000 particles. There are 8287, 2107, and 627 local minima, respectively. As N gets larger, metabasins are decoupled more from their PBC images, and eventually self-learning trajectories probe purely independent metabasin escape events. (e) The most probable metabasin sizes, measured by the peaks of the half-width distribution profiles shown in (d), as a function of the system size N. The solid line is an

exponentially decaying function of $\xi(N) = 3.09 \times \left[1 - \exp\left(-0.147 N^{\frac{1}{3}} \right) \right]$. The asymptotic metabasin size $\xi(N \to \infty) = 3.09$ corresponds to a volume of 148 b-LJ particles.