The Journal of Chemical Physics

Funnel hopping Monte Carlo: An efficient method to overcome broken ergodicity

Cite as: J. Chem. Phys. **152**, 164106 (2020); https://doi.org/10.1063/5.0004106
Submitted: 07 February 2020 . Accepted: 06 April 2020 . Published Online: 23 April 2020

Jonas A. Finkler, and Stefan Goedecker

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Funnel hopping Monte Carlo: An efficient method to overcome broken ergodicity

Cite as: J. Chem. Phys. 152, 164106 (2020); doi: 10.1063/5.0004106
Submitted: 7 February 2020 • Accepted: 6 April 2020 •
Published Online: 23 April 2020

Jonas A. Finkler[a] and Stefan Goedecker[b]

AFFILIATIONS

Department of Physics, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland

[a]Author to whom correspondence should be addressed: jonas.finkler@unibas.ch
[b]Electronic mail: stefan.goedecker@unibas.ch

ABSTRACT

Monte Carlo simulations are a powerful tool to investigate the thermodynamic properties of atomic systems. In practice, however, sampling of the complete configuration space is often hindered by high energy barriers between different regions of configuration space, which can make ergodic sampling completely infeasible within accessible simulation times. Although several extensions to the conventional Monte Carlo scheme have been developed, which enable the treatment of such systems, these extensions often entail substantial computational cost or rely on the harmonic approximation. In this work, we propose an exact method called Funnel Hopping Monte Carlo (FHMC) that is inspired by the ideas of smart darting but is more efficient. Gaussian mixtures are used to approximate the Boltzmann distribution around local energy minima, which are then used to propose high quality Monte Carlo moves that enable the Monte Carlo simulation to directly jump between different funnels. We demonstrate the method's performance on the example of the 38 as well as the 75 atom Lennard-Jones clusters, which are well known for their double funnel energy landscapes that prevent ergodic sampling with conventional Monte Carlo simulations. By integrating FHMC into the parallel tempering scheme, we were able to reduce the number of steps required significantly until convergence of the simulation.

I. INTRODUCTION

In the investigation of molecules and materials, the observed quantities are frequently not the result of the instantaneous state of the system but rather thermodynamic averages over an ensemble consisting of many configurations. The theoretical calculation of such thermodynamic quantities is not a trivial task because a large variety of different configurations as well as anharmonic effects have to be taken into account. Analytical solution does therefore in general not exist, and simulation methods are required. Some of the first computer simulations in condensed matter already addressed these kinds of problems using Molecular Dynamics (MD), and this method is still nowadays widely used in many fields and, in particular, in biological simulations. The harmonic approximation also plays an important role as a starting point for numerous simulation methods such as the harmonic superposition approximation.[1,2] The applicability of this method is, however, very limited because it does not account for anharmonic effects. Monte Carlo (MC) methods represent another pillar for such calculations.

The increasing speed of modern computers has enabled the application of MC methods to larger and more complex systems, making Monte Carlo simulations one of the most popular and widely used tools in the field of statistical mechanics. Nevertheless, most Monte Carlo simulations are based on force fields since performing them at the more accurate density functional level would be too expensive.

A Monte Carlo simulation consists of a random walk over configuration space that generates random samples from a target distribution. These samples can then be used to calculate expectation values over the target distribution. Starting from an initial configuration, consecutive configurations are selected by repeated application of a proposal and an acceptance/rejection step. The new proposed configuration $\mathbf{r}'$ is then accepted or rejected with probability α according to the Metropolis–Hastings criterion,[3] which is given in the following equation:

$$\alpha(\mathbf{r} \to \mathbf{r}') = \min\left(1, \frac{P(\mathbf{r}')}{P(\mathbf{r})} \frac{g(\mathbf{r}|\mathbf{r}')}{g(\mathbf{r}'|\mathbf{r})}\right). \quad (1)$$

In the above equation, $P(\mathbf{r})$ represents the target distribution from which one wants to generate samples, while $g(\mathbf{r}'|\mathbf{r})$ represents the probability to propose a move from $\mathbf{r}$ to $\mathbf{r}'$.

Throughout the whole text, boldface characters will be used to represent a whole configuration with its $3N$ coordinates ($\mathbf{r} = (\vec{r}_{1,x}, \vec{r}_{1,y}, \vec{r}_{1,z}, \vec{r}_{2,x}, \ldots, \vec{r}_{N,z})^\top$). If the same character is used in non-boldface with a vector arrow on top, the 3 coordinates of a single atom belonging to the same configuration are meant. $\vec{r}_i$ represents, therefore, the x, y, and z coordinates of the ith atom of configuration $\mathbf{r}$.

In this work, the distribution of interest is always the Boltzmann distribution, which is given by the following equation:

$$P(r) = \frac{1}{Z(T)} \exp\left(\frac{-E(\mathbf{r})}{k_B T}\right). \tag{2}$$

Here, k_B is the Boltzmann constant, T is temperature, $E(\mathbf{r})$ is the energy of configuration $\mathbf{r}$, and $Z(T)$ is the partition function.

The Markovian nature of the Monte Carlo method dictates that the proposal of the next configuration must only depend on the current state of the simulation. Usually, a new configuration is proposed by adding small random atomic displacements to the current configuration. In the case of rejection, the old configuration has to be included into the average again. It is important to note here that a trade-off has to be made in the design of the proposal step. If the displacements are chosen too big, the proposed configuration will almost always be very high in energy, resulting in a small Boltzmann probability, which will prevent the move from being accepted. If, on the other hand, the steps are chosen to be small, many of them will be accepted, but the correlation time between subsequent samples will be increased, reducing the efficiency of the simulation. There is therefore an inherent trade-off between the size of the moves and the resulting acceptance rate.

A very useful way to characterize an energy landscape is based on its local minima. By assigning every configuration to the local minimum that one obtains by performing a local energy minimization, the energy landscape can be partitioned into so-called catchment basins. Usually, these catchment basins are arranged in a cascading manner. The basins of very low energy minima are surrounded by basins with increasing energy. In many systems, several of these cascades, also called funnels, exist. Understanding the energy landscape in terms of its funnels, basins, and how they are connected can provide great insight into the dynamics of many systems.[4,5]

In systems where multiple funnels are present, the energy barriers between these funnels are much higher than the barriers between local minima in the same funnel. This can pose a great problem to Monte Carlo simulations because crossings of the high inter-funnel barriers occur rarely and may not be observed during the available simulation time. This is known as the problem of broken ergodicity. Ergodicity means that the simulation must be able to reach any point of the configuration space where the probability is non-zero. In theory, this property is satisfied for the Boltzmann distribution as its probability is non-zero everywhere. In practice, however, we are limited by computational power to some finite amount of Monte Carlo steps.

As many systems that have high energy barriers exist, solving the problem of broken ergodicity has been of high interest and a plethora of different algorithms has been proposed. Some of these methods aim to overcome energy barriers by modifying the energy landscape using a biasing potential. In umbrella sampling,[6] the potential is added before the simulation, while the metadynamics technique[7] dynamically adds biasing potentials during the simulation to avoid already visited regions of configuration space. However, both methods require collective variables for biasing. Other methods such as multicanonical sampling[8] or Wang–Landau sampling[9] focus on sampling the inverse of the density of states to obtain a flat energy histogram. Multicanonical sampling suffers from the drawback that the density of states has to be known *a priori*. This problem is overcome by the Wang–Landau method, which computes the density of states dynamically during the simulation. Despite its elegance and broad applicability, the Wang–Landau method can be difficult to apply to systems with continuous degrees of freedom and requires knowledge about the range of accessible energies. If multiple funnels with high barriers are present, the Wang–Landau method can fail to achieve adequate transition rates between the funnels as only local moves are used. In particular, the method was not able to converge to the correct result when applied to the 31 and 38 atom Lennard-Jones clusters and satisfactory results were achieved only after an order parameter was used to construct a two dimensional density of states.[10] Finding suitable collective variables or order parameters for these methods is often a challenging task, requiring in depth knowledge of the studied system, and might not always be possible.

One of the most popular approaches nowadays is presumably parallel tempering (PT). With parallel tempering, multiple simulations with different temperatures are run in parallel. Configurations are then exchanged between simulations with neighboring temperatures. At higher temperatures, the simulations are able to cross the barriers. By the exchange step, the information about configurations on the other side of the barriers is then propagated down to the low temperature simulations. While this method will give exact results, it has the obvious disadvantage of the additional computational cost for the higher temperature simulations. Because the Boltzmann distributions at the different temperatures are required to have significant overlap, the individual temperatures cannot be spaced too far apart and a large number of simulations at different temperatures can be necessary. This is especially the case when the temperature of interest is low compared to the temperature required for the crossing of the highest energy barriers. Then, a large number of replicas have to be inserted between the temperature of interest and the temperature required for the crossing of the barrier. Hence, the propagation of configurations between those temperature ranges can become quite inefficient. Due to the large volume of available configuration space at high temperatures, the rate at which a Monte Carlo simulation jumps between different funnels is further reduced with higher temperatures, diminishing the efficiency of parallel tempering simulations. It is therefore advantageous to combine parallel tempering with other methods that enable jumping between different funnels at lower temperatures.[11,12]

In contrast to Monte Carlo simulations, some modern optimization methods such as minima hopping[13] are much more efficient at exploring the energy landscape as they employ mechanisms to avoid getting trapped in a single funnel. Integrating these schemes

into a sampling procedure, however, is not a trivial matter. The difficulty arises from the need to preserve the detailed balance condition during the Monte Carlo simulation. This difficulty was overcome by Andricioaei, Straub, and Voter[14] who came up with a clever way of using a set of local minima obtained prior to the Monte Carlo simulation to construct moves that directly connect the low energy regions around the local minima.

A different approach was taken by Sharapov, Meluzzi, and Mandelshtam[15] who used a set of local minima to construct an auxiliary harmonic superposition system, which was then coupled to the Monte Carlo simulation. Many more methods exist today that take advantage of information about the local minima.[16–20]

Although the number of local minima increases exponentially with the system size[21,22] so that in most cases it is not possible to obtain a complete set of all local minima, it is generally sufficient to only use a small subset of low energy local minima for these methods. If low energy minima from all major funnels are included, the newly defined move helps the Monte Carlo simulation to cross the higher energy barriers between the funnels, while the lower energy barriers in the system can be overcome by regular Monte Carlo moves.

Even for small model systems interacting with the Lennard-Jones potential, which is computationally very cheap to evaluate, the application of Monte Carlo simulations has only become tractable by using improved sampling methods.[23] Although more accurate force fields such as machine learning force fields[24] are becoming available, the high number of energy and force evaluations required by Monte Carlo simulations still limits their applicability.

Efficient sampling methods will therefore play an important role in enabling the application of Monte Carlo simulations to larger systems with computationally more demanding and accurate energy calculations.

In this work, we propose a novel method called Funnel Hopping Monte Carlo (FHMC) that also uses information about the energy landscape generated prior to the simulation to introduce a new kind of move into the Monte Carlo procedure. The way these moves are constructed is similar to the moves of the auxiliary harmonic superposition system.[15] The main difference is that FHMC uses Gaussian mixtures instead of relying on the harmonic approximation to obtain a single Gaussian. This allows us to overcome the extremely low acceptance rates, which are due to the harmonic approximation failing to describe the Boltzmann distribution accurately. The Gaussian mixtures are directly fit to samples that are generated from the Boltzmann distribution and therefore do not rely on any approximation or assumption.

II. METHOD

A. Smart darting

The main idea behind smart darting is to construct new Monte Carlo moves using information about the local minima of the energy landscape. In the original approach,[14] this is achieved in the following way. Given a set of M local minima $\{\mathbf{R}_i\}_{i=1,2,\ldots,M}$, the so-called darting vectors $\mathbf{D}_{ij}$ are defined as the pairwise differences between $\mathbf{R}_i$,

$$\mathbf{D}_{ij} = \mathbf{R}_j - \mathbf{R}_i \quad | \quad i \neq j. \tag{3}$$

Additionally, epsilon regions are placed around each local minimum. A configuration is considered to be inside such a region if the Euclidean distance to the local minimum $\mathbf{R}_i$ is smaller than ε,

$$\|\mathbf{R}_i - \mathbf{r}_i\| < \varepsilon. \tag{4}$$

ε should be chosen such that none of the regions overlap.

The darting moves will then replace a certain fraction of the standard Monte Carlo moves. In each iteration of the algorithm, a random number is drawn to decide which kind of move will be performed. In the case that a darting move is chosen, it will first be checked if the current configuration $\mathbf{r}$ is inside one of the epsilon regions. If not, the move is considered rejected. If $\mathbf{r}$ lies inside one of the epsilon regions corresponding to minimum $\mathbf{R}_i$, a darting move will be proposed. First, a darting vector starting at minimum i is chosen randomly. A new configuration is then proposed by adding this darting vector to the current configuration,

$$\mathbf{r}' = \mathbf{r} + \mathbf{D}_{ij}. \tag{5}$$

The final step is then the acceptance or rejection of the proposed configuration by the standard Metropolis criterion [Eq. (1)].

While this method of proposing darting moves works well, in some cases, there are several shortcomings. The first problem arises when we try to apply the method to systems that are invariant under rotations and translations such as clusters. Defining the epsilon regions for these systems is not trivial as each minimum of the system does not correspond to a single point in our coordinate space but rather a hypersurface, induced by rotation and translation of the atoms. Although the translational ambiguity can be removed by fixing the center of mass, the solution to the rotational problem is not so obvious. Things get even worse when we start considering systems that have multiple atoms of the same kind. Because these atoms are indistinguishable, the system will be invariant under permutation of these atoms.

This means that in our current Cartesian coordinate space, there are $2N!/h_\alpha$ hypersurfaces that correspond to the same configuration, with N being the number of atoms in our system (if all of them are equivalent) and h_α is the point group order of the configuration.

Having whole hypersurfaces that correspond to one single configuration makes it impossible to define the epsilon regions as described above. In addition, the darting vectors can only be defined using a single orientation and permutation of the system. As soon as atoms exchange places or the system rotates, these darting vectors would not connect the epsilon regions any more.

Another problem with smart darting is that it uses spherical epsilon regions. In reality, the regions of low energy, and therefore high probability, which should be targeted by the darting moves, are often rather ellipsoidal due to the presence of soft and hard modes. Because the axes of these ellipsoidal regions are, in general, not parallel, darting by addition of a darting vector will often miss the low energy regions around the local minima. We therefore set out to develop a method that is able to directly target these high probability regions, without making any prior assumptions about their shape.

B. Eckart space and the RMSD

To implement such a method, it is necessary that we are able to identify similar configurations to decide if a configuration is inside

one of the high probability regions. As systems consisting of identical atoms are invariant under rotations and translations, it is not trivial to identify equivalent configurations in the $3N$ dimensional coordinate space. To assign a set of coordinates to a given configuration $\mathbf{r}$ that is invariant under rotation and permutation, we first select a reference configuration $\mathbf{R}$, which will be one of the local minima in our case. We then determine the optimal rotation $\mathcal{R}$ and permutation $\mathcal{P}$ of $\mathbf{r}$ so that the root mean squared deviation (RMSD) to $\mathbf{R}$ is minimal. The algorithm used to minimize the RMSD is described in Sec. II C.

The RMSD is defined as follows:

$$\mathrm{RMSD}(\mathbf{r},\mathbf{R}) = \sqrt{\frac{\sum_{i=1}^{N} \|\vec{r}_i - \vec{R}_i\|^2}{N}}. \tag{6}$$

It should be noted here that superimposing the centers of mass always results in the minimal RMSD with respect to translation. We therefore assume without loss of generality that the center of mass for all configurations is set to the coordinate origin.

It can be shown that the RMSD between two configurations is minimal if the so-called Eckart conditions[25] are met.[26] The Eckart conditions are the following:

$$\sum_{i=1}^{N} \vec{r}_i - \vec{R}_i = \vec{0}, \tag{7}$$

$$\sum_{i=1}^{N} \vec{r}_i \times \vec{R}_i = \vec{0}. \tag{8}$$

We now define the displacement $\mathbf{d}$ as the difference between the aligned structure and the reference,

$$\vec{d}_i = \vec{r}_i - \vec{R}_i. \tag{9}$$

With these, we can now write the Eckart conditions as follows:

$$\sum_{i=1}^{N} \vec{d}_i = \vec{0}, \tag{10}$$

$$\sum_{i=1}^{N} \vec{d}_i \times \vec{R}_i = \vec{0}. \tag{11}$$

From these six linear equations, it follows that all displacement vectors $\mathbf{d}$, obtained from a minimal RMSD alignment, are orthogonal to the following six vectors:

$$\mathbf{V}^1 = \begin{pmatrix} 1\\0\\0\\1\\0\\0\\1\\\vdots \end{pmatrix}, \mathbf{V}^2 = \begin{pmatrix} 0\\1\\0\\0\\1\\0\\0\\\vdots \end{pmatrix}, \mathbf{V}^3 = \begin{pmatrix} 0\\0\\1\\0\\0\\1\\0\\\vdots \end{pmatrix},$$

$$\mathbf{V}^4 = \begin{pmatrix} 0\\\vec{R}_{1,z}\\-\vec{R}_{1,y}\\0\\\vec{R}_{2,z}\\-\vec{R}_{2,y}\\0\\\vdots \end{pmatrix}, \mathbf{V}^5 = \begin{pmatrix} -\vec{R}_{1,z}\\0\\\vec{R}_{1,x}\\-\vec{R}_{2,z}\\0\\\vec{R}_{2,x}\\\vec{R}_{3,z}\\\vdots \end{pmatrix}, \mathbf{V}^6 = \begin{pmatrix} \vec{R}_{1,y}\\-\vec{R}_{1,x}\\0\\\vec{R}_{2,y}\\-\vec{R}_{2,x}\\0\\\vec{R}_{3,y}\\\vdots \end{pmatrix}. \tag{12}$$

Here, the vectors $\mathbf{V}^1$, $\mathbf{V}^2$, and $\mathbf{V}^3$ are obtained from Eq. (10), and $\mathbf{V}^4$, $\mathbf{V}^5$, and $\mathbf{V}^6$ are obtained from Eq. (11).

We now construct $3N-6$ basis vectors $\mathbf{B}^i$ that are orthogonal to each other as well as to the vectors $\mathbf{V}^j$,

$$\|\mathbf{B}^i\| = 1, \tag{13}$$

$$\mathbf{B}^i \cdot \mathbf{B}^j = 0 \quad \forall \quad i \neq j, \tag{14}$$

$$\mathbf{B}^i \cdot \mathbf{V}^j = 0. \tag{15}$$

$\mathbf{B}^i$ can be obtained using an orthogonalization algorithm such as the modified Gram–Schmidt process.

Using $\mathbf{B}^i$ as a basis, we can remove 6 coordinates from our displacement vectors $\mathbf{d}$. These six coordinates become redundant because we fixed the rotation and translation of the configuration. This allows us to assign a unique set of $3N-6$ coordinates to every configuration.

The $3N$ dimensional vector $\mathbf{d}$ is transformed to the $3N-6$ dimensional vector $\mathbf{d}'$, using the basis spanned by $\mathbf{B}^i$, as follows:

$$\mathbf{d}'_i = \mathbf{d} \cdot \mathbf{B}^i \quad | \quad i = 1 \ldots 3N-6. \tag{16}$$

Here, $\mathbf{d}'_i$ denotes the ith component of vector $\mathbf{d}'$.

To obtain the original configuration, $\mathbf{d}'$ is simply transformed back to the $3N$ dimensional space and added to the reference configuration $\mathbf{R}$,

$$\mathbf{r} = \mathbf{R} + \mathbf{d} = \mathbf{R} + \sum_{i=1}^{3N-6} \mathbf{d}'_i \, \mathbf{B}^i. \tag{17}$$

C. Minimizing the root mean squared deviation

In systems of distinguishable atoms, the RMSD can be considered a function of the rotation of the system as the optimal translation can be found trivially by superimposing the mean atom positions of the two systems. In systems consisting of indistinguishable atoms, however, we are confronted with some kind of chicken and egg problem as the optimal rotation on the one hand depends on the permutation indicating which atoms from each systems we pair together, while the optimal permutation on the other hand depends on the rotation. Each problem by itself can be solved by known algorithms. To find the optimal rotation to a given permutation, we can use an algorithm based on quaternions.[27–29] To determine the optimal permutation for a given rotation, we can use the Hungarian algorithm[30] or the shortest augmenting path algorithm.[31] To

solve the combined problems, we use both algorithms in alternation until a converged solution is found. As each of the two algorithms will only decrease the RMSD, repeated application of them will lead to a local optimum of the RMSD. To find the globally optimal RMSD, we initialized the local optimization with different initial rotations. These rotations were represented by unit quaternions. Because quaternions with opposite sign represent the same rotation, each rotation has to be represented by a pair of points on opposite sides of a four dimensional unit hypersphere. To distribute these rotations as uniformly as possible over the hypersphere of all rotations, we put a charge on each point and minimized the Coulomb energy using a simple gradient descent algorithm with the additional constraints that the points be on the unit sphere. These uniformly distributed rotations increase the chances of finding the globally minimal RMSD within a limited number of steps significantly.

To test our algorithm, we generated random configurations with an RMSD of 0.1 to the local minimum of the 38 atom Lennard-Jones cluster with the third lowest energy. We chose this configuration because it is the lowest local minimum, which has no rotational symmetry. The RMSD of 0.1 was chosen because it is large enough so that the alignment is not trivial but small enough to ensure no other permutation than the original one can result in a smaller RMSD. In our experiments, we found that by using 400 evenly spread initial rotations, the globally minimal RMSD solution was found in 100% of the 10 000 test alignments performed. If we used random initial rotations instead, only 85% of the alignments succeeded. Even when using 600 random initial rotations, the minimal RMSD solution was only found in 95% of the attempts.

D. Funnel hopping Monte Carlo

Using the methods described in Sec. II B, we are able to uniquely map structures into a $3N-6$ dimensional coordinate space. This capability is the foundation of our novel algorithm called Funnel Hopping Monte Carlo (FHMC) as it allows us to generate Monte Carlo moves that directly target regions of low energy.

By using some metric, which may be the RMSD or fingerprints,[32] we assign each point in the configuration space to its nearest minimum. Thus, each minimum is assigned a part of the configuration space. In our implementation, we used fingerprints because they are computationally cheaper. For each minimum $\mathbf{R}_i$, we will then define a probability distribution $q_i(\mathbf{r})$, which will live in the $3N-6$ dimensional fixed frame coordinate space, and sample the low energy region around this minimum. These q_i should cover the high probability regions as exhaustively as possible. This can be done, for example, by using the harmonic approximation that would result in an algorithm similar to the one proposed by Sharapov, Meluzzi, and Mandelshtam[15] or by a Gaussian mixture as we will propose in Sec. II E. It is important to note here that these distributions do not carry any physical meaning. How well these resemble the Boltzmann distribution does not influence the accuracy of the final algorithm as a detailed balance is always satisfied. $q_i(\mathbf{r})$'s just allow the funnel hopping Monte Carlo algorithm to propose better moves that are more likely to be accepted, which results in a more efficient sampling.

To propose a Funnel Hopping move, we first determine the minimum $\mathbf{R}_i$ that is closest to the current configuration. We then randomly choose one of the other minima and draw a configuration from the corresponding q_j. The choice of the target minimum can be done completely random or one can include a transition matrix T, with T_{ij} being the probability to choose minimum j when the current configuration is closest to minimum i. Such a transition matrix can be used, for example, to avoid proposing moves to minima that are too different in energy. The proposed move is then accepted with probability α according to the Metropolis criterion,

$$\alpha(\mathbf{r}\to\mathbf{r}') = \min\left(1, \exp\left(-\frac{E(\mathbf{r}')-E(\mathbf{r})}{k_BT}\right)\frac{q_i(\mathbf{r})}{q_j(\mathbf{r}')}\frac{T_{ji}}{T_{ij}}\frac{h_{\alpha i}}{h_{\alpha j}}\right). \quad (18)$$

Here, $h_{\alpha i}$ is the point group order of the ith minimum. If a minimum has a rotational symmetry, $h_{\alpha i}$ alignments with the same RMSD exist. All these alignments will result in different coordinates if transformed to the basis vectors $\mathbf{B}^i$ while they describe exactly the same configuration. Therefore, $h_{\alpha i}$ is as likely to pick a configuration as $q_i(\mathbf{r})$ indicates because $h_{\alpha i}$ points in the space of the Hessian basis exist that correspond to that configuration and are equally likely.

The distributions q_i play two important roles. First, we can see from the above equation that the acceptance probability is proportional to $q_i(\mathbf{r})$, which means that the better the q_i cover the high probability regions, the higher is the expected acceptance rate. The other function of q_i is that they are used to generate the proposed configurations. Again, one can see that if q_i cover the high probability regions well, we will propose configurations with a low energy, which will result in a high acceptance probability.

Although the Gaussian mixtures are usually quite localized, they do in principle have infinite support. This means that it is possible that the proposed configuration $\mathbf{r}'$ lies outside of the part of the configuration space that is assigned to the minimum j. This would result in a move where a detailed balance is not satisfied as the inverse move is not possible. Rejecting moves to configurations outside the region of the configuration space assigned to minimum j ensures that the detailed balance condition is met and no errors are introduced.

E. Gaussian mixtures

A rather straight forward approach to define q_i is to use the harmonic approximation of the energy. As the harmonic approximation is a quadratic function, the Boltzmann distribution of this energy will be a Gaussian distribution of the following form:

$$q_i^{h.a.}(\mathbf{r}) = \frac{1}{\sqrt{(2\pi k_BT)^{3N-6}\mathrm{Det}\left(\mathrm{H}^{-1}\right)}}\exp\left[-\frac{\mathbf{r}^\top\mathrm{H}\mathbf{r}}{2k_BT}\right]. \quad (19)$$

In this equation, H represents the Hessian matrix of the energy, transformed to the basis spanned by $\mathbf{B}^i$. It should be noted that at every local minimum of the energy, the Hessian matrix will have six eigenvalues that are zero. These corresponding eigenvectors coincide with the $\mathbf{V}^i$ defined above. The Hessian matrix is therefore not singular in the basis spanned by $\mathbf{B}^i$.

Although this approximation becomes exact in the limit of the temperature going to zero, we found that at finite temperatures, acceptance rates of our algorithm were very low using the harmonic

approximation. For the 38 and 75 atom Lennard-Jones clusters, the acceptances rates were around 0.2% and 0.04%, respectively. Similar behavior was also observed in the context of the auxiliary harmonic superposition system.[12] Using the harmonic approximation in funnel hopping Monte Carlo results in an algorithm that is very similar to the auxiliary harmonic superposition system. We therefore include calculations using the harmonic approximation in Sec. III for comparison.

To overcome the deficiencies of the harmonic approximation, we propose a different approach to find suitable q_i by using a mixture of Gaussians, which is defined as follows:

$$q_i^{g.m.}(\mathbf{r}) = \sum_{k=1}^{m} a_i^k \mathcal{N}_i^k(\mathbf{r}) \quad \Big| \quad \sum_{k=1}^{m} a_i^k = 1 \text{ and } a_i^k \geq 0 \; \forall k. \tag{20}$$

Here, $\mathcal{N}_i^k$ represent normalized Gaussians defined by means μ_i^k and covariance matrices Σ_i^k. Once a_i^k and $\mathcal{N}_i^k$ are determined, we can generate samples from the Gaussian mixture by picking a random k with probability a_i^k and then drawing a random sample from $\mathcal{N}_i^k$. To generate samples from $\mathcal{N}_i^k$, we first generate a set of random numbers drawn from a standard-normal distribution using the Box–Müller algorithm. We then use the Cholesky decomposition of Σ_i^k as well as μ_i^k to transform the random numbers to the desired distribution.[33]

The parameters a_i^k, μ_i^k, and Σ_i^k are determined by fitting the Gaussian mixture to samples drawn from the Boltzmann distribution using the expectation–maximization (EM) algorithm.[34–36]

If only a single Gaussian is fit, the resulting method is equivalent to the principle mode analysis method.[37] Alternatively, the self-consistent phonon method[38,39] could also be used to fit a single Gaussian distribution.[40] As in our method, the required number of energy and force evaluations was dominated by Monte Carlo sampling and not by the construction of the Gaussian mixtures; an improved efficiency in this part is not of great advantage. What matters is the improved acceptance rates that can be achieved with fits that use multiple Gaussians.

We also developed a modified version of the EM algorithm, which takes advantage of the high symmetry present in many low energy configurations. By constraining the Gaussian mixture to have the same symmetry as the local minima, the number of free parameters can be reduced, which leads to a better quality of the fit. An outline of the modified algorithm is given in the Appendix.

For each of the local minima that are included, samples are generated and the Gaussian mixtures are fit individually. The samples are collected from a standard Monte Carlo run initialized at the local minimum $\mathbf{R}_i$ after a short equilibration phase. During the Monte Carlo run, we repeatedly check if the current configuration is still inside the region assigned to the local minimum $\mathbf{R}_i$. If the region was left, the simulation is reinitialized at R_i and some equilibration steps are performed. This ensures that the samples are all drawn from a single peak in the Boltzmann distribution that belongs to the corresponding minimum.

For our Lennard-Jones clusters, we used a local geometry optimization to check if the Monte Carlo walker has left the catchment basin of the local minimum. In our experiments, this required approximately 100 energy and force evaluations per sample that we collected. For the Lennard-Jones clusters, these energy and force evaluations are extremely cheap. If a more expensive method, such as density functional theory (DFT), would be used, one would have to resort to fingerprints or RMSD calculations to check if the catchment basin has been left. We therefore decided not to include these energy and force evaluations into the final results.

III. APPLICATION

We tested our algorithm on clusters consisting of 38 (LJ_{38}) and 75 (LJ_{75}) atoms interacting with the Lennard-Jones potential, which is as follows:

$$E_{LJ} = \sum_{i<j}^{N} 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right]. \tag{21}$$

Here, r_{ij} are the pairwise distances between atoms i and j. During each step of the simulation, we decided randomly with a ten percent probability to perform a Funnel Hopping move. All other moves were performed using a Hamiltonian Monte Carlo (HMC) approach also known as hybrid Monte Carlo.[41,42]

To avoid evaporation events, where atoms detach from the cluster, a confining potential was added to the systems. We used the same soft potential as used by Nigra, Freeman, and Doll,[11] which is given as follows:

$$V(\mathbf{r}) = \sum_{i=1}^{N} \epsilon \left(\frac{\| \vec{r}_i - \vec{r}_{cm} \|}{r_c} \right)^{20}, \tag{22}$$

with $\vec{r}_{cm}$ being the center of mass and r_c being the radius of the confining potential. We experimentally found $r_c = 3.5\sigma$ to be a good choice for LJ_{38} and $r_c = 4\sigma$ for LJ_{75}. A good choice of r_c has to prevent atoms from escaping without influencing the energy of the cluster too much. A soft potential was used because the derivatives/forces were needed for the Hamiltonian dynamics.

We compared the results of our method to that of parallel tempering.[43–45] We used a geometric distribution of the temperatures as proposed by Kofke.[46] In our simulation, swaps were attempted every ten Monte Carlo steps between adjacent temperatures. They were accepted with a rate of 16–19% for LJ_{38} and between 13% and 17% for LJ_{75}, which is close to the optimal acceptance rate of 20% proposed by Rathore, Chopra, and de Pablo.[47] Swaps were performed in an alternating manner between pairs of subsequent temperatures, e.g., after the first ten steps, swaps were attempted between pairs 1–2, 3–4, 5–6, . . ., and then after ten more steps, pairs 2–3, 4–5, 6–7, etc., were used.

The heat capacity was calculated using the following equation:

$$C_V(T) = \frac{3}{2} + \frac{1}{NT^2} \left(\langle E^2 \rangle_T - \langle E \rangle_T^2 \right), \tag{23}$$

with $\langle \cdot \rangle_T$ representing the expectation values over the Boltzmann distribution at temperature T.

To obtain smooth plots of the heat capacity, we used the reweighting scheme proposed by Sharapov and Mandelshtam[12] to interpolate between the different temperatures.

A. Lennard-Jones 38

The most studied Lennard-Jones cluster is presumably the one consisting of 38 atoms (LJ_{38}), which is known for its two funnel

energy landscape that almost completely prevents ergodic sampling using conventional Monte Carlo methods.

One funnel ends in the global minimum, which is a face-centered-cubic truncated octahedral structure. The other funnel ends in the second lowest minimum, which is an incomplete Mackay icosahedron. These two funnels are separated by a high energy barrier with a transition state energy of 4.219ϵ above the ground state energy,[48] which is almost impossible to overcome at low temperatures.

Gaussian mixtures were fit for the ten lowest local minima (stereoisomers were counted as one) using 2×10^5 samples. The acceptance rates achieved are shown in Fig. 1.

As shown in Fig. 1, going from 5 to 10, Gaussians did not increase the acceptance rate. We suspect that this is due to the number of samples not being sufficient for the high numbers of parameters that have to be fitted. In this case, we have 5995 free parameters per Gaussian: $n(n+1)/2$ from the covariance matrix, n from the mean, and one a_i^k with n being the dimensionality of the Gaussian, which is $3 \times 38 - 6$ in this case.

To further increase the quality of the fit without having to generate more samples, we developed a method to incorporate the high symmetry of the local minima into the fitting procedure. In the Appendix, an outline of the algorithm is shown. The acceptance rate for the Monte Carlo run using a symmetric Gaussian mixture is labeled *sym*-4 in Fig. 1. The Gaussian mixture for this fit consists of four Gaussians per symmetry of the local minima.

This fit was then used to calculate the heat capacity of LJ_{38} using FHMC in combination with parallel tempering. The result is shown in Fig. 2 where it is compared to a reference calculation using parallel tempering with 10^9 steps. In our implementation of both the simple parallel tempering and the FHMC method, one step of the Hamiltonian Monte Carlo algorithm required 25 energy and force evaluations. The curve using both methods in combination was obtained after 10^7 steps. While this result is in agreement with the reference, the result obtained with parallel tempering alone using the same number of steps is clearly not converged.

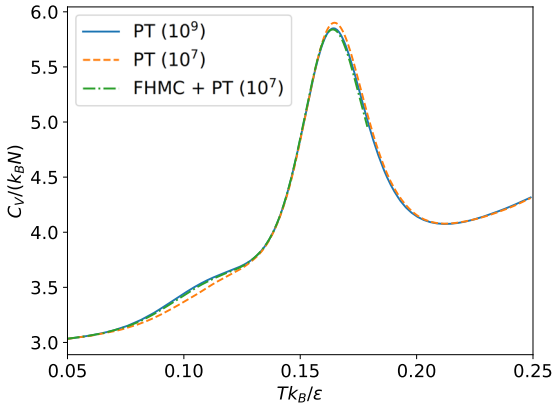

FIG. 2. Heat capacity of LJ_{38} calculated with our method compared to the result obtained using parallel tempering after different numbers of steps.

To assess the convergence properties of our method, we repeated the calculation of the heat capacity ten times with both methods individually and combined using 10^7 steps. We then calculated the root mean squared error (RMSE) with respect to the reference obtained with 10^9 parallel tempering steps. The resulting RMSEs are shown in Fig. 3.

The results show that our method alone outperforms parallel tempering at the lower temperature range up to $T = 0.11\epsilon/k_B$. In this range, the number of accessible minima is low enough so that they are well covered by the darting sites. In this special case, funnel hopping Monte Carlo can be used to perform ergodic sampling using only one simulation at a single temperature, reducing the computational effort by several orders of magnitude compared to parallel tempering simulations, where a whole range of temperatures has to be simulated. We found that if we included only the lowest

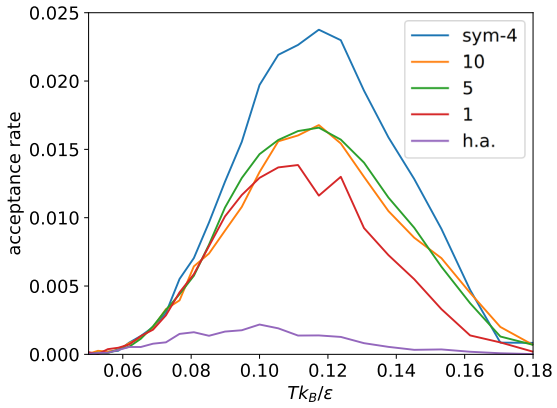

FIG. 1. Acceptance rate of funnel hopping moves in LJ_{38} plotted against temperature. The numbers represent the number of Gaussians used in the Gaussian mixtures. For the line labeled h.a., the harmonic approximation was used, and for the line labeled sym-4, an extended version of the EM algorithm was used to fit a symmetric Gaussian mixture with four Gaussians per symmetry.

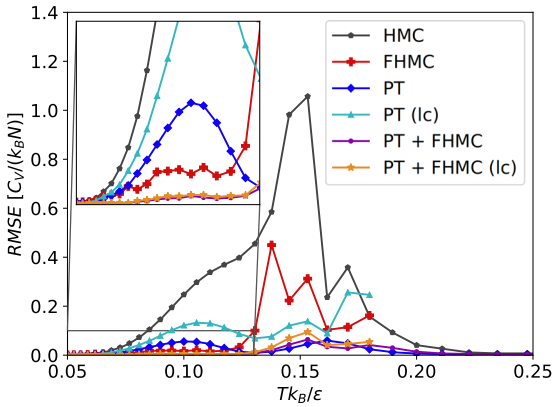

FIG. 3. Root mean squared error of the heat capacity of LJ_{38} calculated with Hamiltonian Monte Carlo (HMC), parallel tempering (PT), Funnel Hopping Monte Carlo (FHMC), and PT and FHMC methods in combination once with a regular and once with a lower maximum temperature (lc).

minimum in each funnel, the FHMC calculation did not converge within 10^7 steps. If combined with parallel tempering, however, the convergence was only slightly slower than with 10 minima.

At higher temperatures, additional minima become relevant that are not included into the FHMC scheme and have to be reached by standard Monte Carlo moves, which can slow down convergence.[44]

One major drawback of parallel tempering is that a large range of temperatures has to be simulated with the maximum temperature being high enough so that the highest energy barriers can be crossed by the Monte Carlo simulation. In our experiments, we chose a maximum temperature of $0.4\epsilon/k_B$ for parallel tempering, while the maximum temperature for the FHMC simulations can be chosen arbitrarily because each simulation is performed independently of the others. For our FHMC simulations, we chose a maximum temperature of $0.18\epsilon/k_B$. Parallel tempering simulations with this maximum temperature did not converge to the correct result. When we combine parallel tempering with our method, convergence could be achieved. In this case, the Funnel Hopping moves allow the simulation to cross the highest barriers, while parallel tempering enables the crossing of the lower barriers between basins within a funnel that are not included into the FHMC scheme. Using both methods in combination allows us therefore to use a significantly lower cutoff temperature than with standalone parallel tempering. It combines the best of both methods by using parallel tempering to skip barriers inside funnels and funnel hopping Monte Carlo to move between different funnels, leading to improved sampling capabilities across the whole temperature range, which is shown in Fig. 3.

B. Lennard-Jones 75

As a final test, we applied funnel hopping Monte Carlo to an even more challenging system, namely, the 75 atom Lennard-Jones cluster (LJ_{75}). Similar to the 38 atom cluster, its energy landscape also consists of two major funnels, one ending in the global optimum, which has a decahedral structure, and the other one ending in the second lowest local minimum, which has an icosahedral structure. These two minima are separated by a barrier that lies 8.69ϵ above the ground state energy. This barrier is over 3ϵ higher than any other barrier between the 250 lowest minima.[22] Unlike in the case of LJ_{38}, the peak in the heat capacity caused by the solid–solid transition is well separated from the melting peak.

The very high barrier between the two funnels makes ergodic sampling of this system particularly difficult. It seems that parallel tempering alone is not enough to obtain converged results for LJ_{75}.[11,12] Our own calculation using parallel tempering with the Hamiltonian Monte Carlo approach did not converge after 5×10^8 steps per temperature (5×10^{11} energy and force evaluations in total). Because of the high energy barrier between the two funnels, transitions are limited to the very high temperature range of the parallel tempering simulation. At these temperatures, the accessible configuration space is extremely large causing the transition between the funnels to be particularly rare.

By combining parallel tempering with Funnel Hopping Monte Carlo, transitions between the two funnels become possible already at low temperatures.

We used funnel hopping Monte Carlo in combination with parallel tempering to calculate the heat capacity of LJ_{75}. The two lowest minima were included into the FHMC scheme to facilitate the crossing of the high inter funnel barrier. We used our version of the EM algorithm to fit Gaussian mixtures of three Gaussians per symmetry using 2×10^5 samples per local minimum.

FHMC moves were included with a probability of 0.1 up to a temperature of $0.119\epsilon/k_B$ above which the acceptance rate of the moves decays to almost zero. A total of 40 parallel tempering replicas were used, which were run in parallel each on one central processing unit (CPU) core. The lower 20 of the replicas included FHMC moves. Because we included two local minima into the FHMC scheme, 40 Gaussian mixtures were fit. The resulting acceptance rates are shown in Fig. 4. The fitted Gaussian mixtures outperform the harmonic approximation in terms of the acceptance rate of the proposed moves by about a factor of 20.

Parallel tempering swaps were again included after every 10 steps. Samples were collected after an equilibration period of 10^5 steps.

The obtained heat capacity after 1.4×10^7 steps is shown in Fig. 5. We identified the peak of the heat capacity corresponding to the solid–solid transition at a temperature of $0.085\epsilon/k_B$. This is slightly higher than the result of $0.083\epsilon/k_B$ reported by Sharapov and Mandelshtam.[12] To explain this minor discrepancy, we ran several simulations, initialized with the second lowest instead of the lowest minimum, with a larger confining radius as well as with a longer equilibration period. However, the results of all these calculations gave the same numerical value for the peak.

In Fig. 6, the peak corresponding to the low temperature solid–solid transition is shown again and compared to the results obtained with our method and with a run, where the harmonic approximation was used instead of fitted Gaussians, both after 10^5 and 10^6 steps. After 10^6 steps, the FHMC calculation is converged to the final result after 1.4×10^7 steps up to a very high precision, while the result from the harmonic approximation is still significantly shifted toward the right. After only 10^5 steps, the FHMC calculation is already converged to a result where the heat capacity peak is in good qualitative

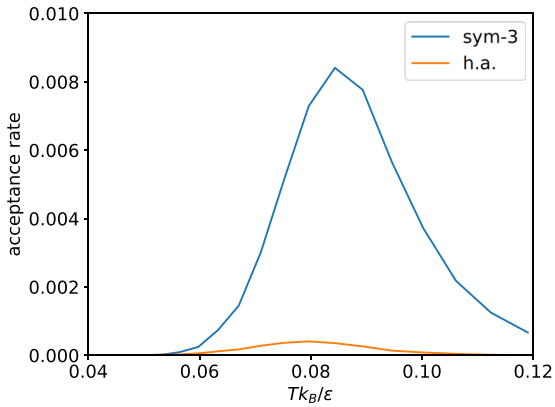

FIG. 4. Acceptance rate of funnel hopping moves in LJ_{75} plotted against temperature. The result obtained using the Gaussian mixtures, consisting of 3 Gaussians per symmetry, is labeled *sym*-3, while the result obtained using the harmonic approximation is labeled *h.a.*

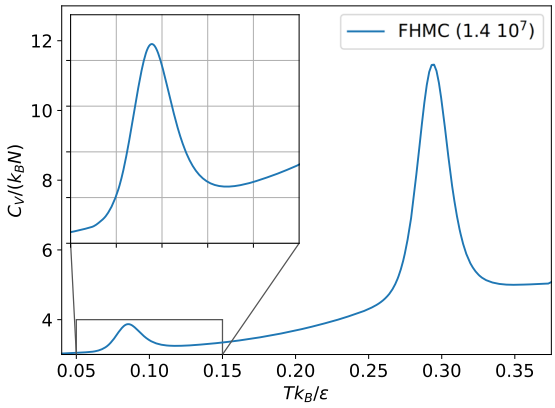

FIG. 5. Heat capacity of LJ_{75} calculated with funnel hopping Monte Carlo.

agreement with the converged result. These results clearly indicate that the rate at which the simulation jumps between the two funnels is the limiting factor for the convergence of the Monte Carlo simulation.

Hence, using Funnel Hopping Monte Carlo, we were able to obtain a converged result after only 10^6 steps (3.25×10^7 energy and force evaluations per temperature or 1.3×10^9 in total, including sample generation for the Gaussian mixtures as well as the equilibration part). This is almost 100 times less than the 3×10^9 energy evaluations per temperature reported by Sharapov, Meluzzi, and Mandelshtam,[15] where an auxiliary harmonic superposition system was used, and more than 100 times less than the 4×10^{11} energy evaluations in total reported by Martiniani *et al.*,[20] where the approximate SENS method was employed. The basin-sampling method[19] used 0.27×10^9 energy evaluations per replica, resulting in a total of 8.64×10^9 energy evaluations for all 32 replicas combined, not including the energy evaluations required for the minimizations in the final BS phase. This is significantly more than in the FHMC method.

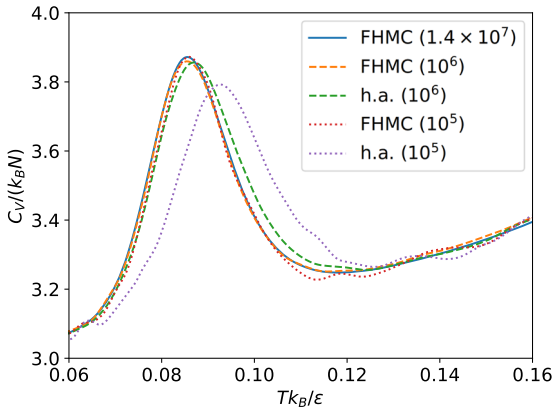

FIG. 6. Low temperature heat capacity peak of LJ_{75} calculated with our method using a Gaussian mixture and the harmonic approximation. The number given in the legend indicates the number of HMC steps during which the samples were collected (1 step = 25 energy and force evaluations per temperature).

IV. CONCLUSION

With funnel hopping Monte Carlo, we developed a new tool to overcome broken ergodicity by exploiting precomputed knowledge about the energy landscape into the Monte Carlo simulation. Our method generates an accurate approximation to the Boltzmann distribution even for anharmonic systems. This allows us to propose good moves between different funnels that have a high chance of being accepted by the Monte Carlo algorithm. Using Gaussian mixtures allows for a systematic improvement of the proposed moves by increasing the number of samples used for fitting and the number of Gaussians in the Gaussian mixture. With our newly developed variant of the expectation–maximization algorithm, we are able to take advantage of the high symmetry present in many local minima, which results in an even better fit of the Gaussian mixtures. With our fits, we were able to achieve acceptance rates about 20 times higher than with the harmonic approximation. We observed that the convergence of the Monte Carlo simulation is limited by the rate at which the simulation is able to transition between the different funnels and is therefore directly dependent on the acceptance rate of the inter funnel moves.

If the temperature of interest is low enough so that only a limited number of basins are accessible and if it is possible to include all of them into the algorithm, funnel hopping Monte Carlo can be performed at a single temperature, whereas parallel tempering requires many auxiliary simulations at higher temperatures.

We also showed that by combining our method with the parallel tempering scheme, the maximum simulation temperature can be significantly reduced, which allows us to avoid unnecessary calculations, resulting in a reduced computational cost. In addition, the convergence of the simulation is sped up massively as the funnel hopping Monte Carlo moves help the simulation to cross the highest barriers between different funnels very efficiently.

Using funnel hopping Monte Carlo, we were able to obtain the heat capacity of the 75 atom Lennard-Jones cluster, a notoriously difficult system, known to suffer from a particularly strong broken ergodicity. Nevertheless, using only 1.3×10^9 energy and force evaluations in total, we could obtain the converged results. This number of evaluations is significantly less than the number required by the existing state of the art methods. We also observed that the results were already in good qualitative agreement after only about 10^8 energy and force evaluations.

ACKNOWLEDGMENTS

This work was partly supported by the NCCR MARVEL and funded by the SNF. J.A.F. acknowledges support from the SNF within a joint DFG/SNF project. Computational resources provided by the Swiss National Supercomputing Centre (CSCS) in Lugano under the Project No. s963 are gratefully acknowledged. Calculations were also performed at sciCORE (http://scicore.unibas.ch/), the scientific computing center at the University of Basel.

APPENDIX: FITTING SYMMETRIC GAUSSIAN MIXTURES

Low energy configurations of clusters often exhibit a high degree of symmetry. This is especially the case for the Lennard-Jones

38 cluster, where the ground state has 24 rotational symmetries as well as an inversion symmetry, resulting in a total of 48 symmetries.

These symmetries will also be present in the Boltzmann distribution, which we approximate using the Gaussian mixtures. By constraining the Gaussian mixtures to have the same symmetries as the local minima, the number of free parameters can be reduced, which results in an increased quality of fit with the same number of training samples used. We therefore developed the following variant of the expectation–maximization algorithm.

In a first step, we determine all rotation and inversion symmetries of the configuration. For that, the configuration is first rotated randomly, and then the alignment algorithm described in Sec. II C is used to align the rotated structure to the original configuration. All distinct assignments with an RMSD of zero correspond to a symmetry operation. The procedure is then repeated with the structure being inverted so that all symmetries that include an inversion can be found. Alternatively, the symmetries can also be detected using a more efficient code, such as libmsym.[49]

We then define a symmetric Gaussian mixture by replicating a normal Gaussian mixture for each of the symmetry operations,

$$q^{sym}(\mathbf{r}) = \sum_{j=1}^{N_{sym}} \sum_{k=1}^{m} a^k \mathcal{N}_j^k(\mathbf{r}). \tag{A1}$$

Here, N_{sym} is the number of symmetries, m is the number of Gaussians per symmetry, and $\mathcal{N}_j^k$ is the kth Gaussian under the jth symmetry transformation. Similar to the non-symmetric Gaussian mixture, a_k's are weights for the individual Gaussians. Because each Gaussian is replicated N_{sym} times, a_k have to sum up to $1/N_{sym}$.

Each symmetry operation consists of a rotation represented by a rotation matrix R, a permutation represented by a permutation matrix P, and optionally an inversion. To apply a symmetry transformation to the $3N-6$ dimensional vectors, we first have to transform them back to the $3N$ dimensional space. Then, the rotation, permutation, and inversion are applied before transforming back to the $3N-6$ dimensional coordinates. Combining all these operations yields the matrix M,

$$\mathrm{M} = \mathrm{B}^\top \mathrm{PQIB}, \tag{A2}$$

with B being a $3N \times 3N-6$ matrix with its columns consisting of the $3N-6$ basis vectors $\mathbf{B}^i$, Q being a block diagonal matrix with the rotation matrix R repeated N times along its diagonal, and I being the identity matrix $\mathbb{1}$ or $-\mathbb{1}$ if an inversion is applied. The Gaussian $\mathcal{N}_j^k$ is hence defined by the mean $\mu_j^k = \mathrm{M}_j\mu^k$ and the covariance $\Sigma_j^k = \mathrm{M}_j\Sigma^k\mathrm{M}j^\top$. The symmetric Gaussian mixture model is therefore parametrized by m weights a^k, m mean vectors, and m covariance matrices.

To fit this symmetric Gaussian mixture, we modified the original expectation–maximization algorithm in the following way. During the expectation part of each iteration, we first construct the full symmetric Gaussian mixture as is given by Eq. (A1). We then calculate the weights y_i^{jk} for each sample x_i in the same way as in the original algorithm,

$$y_i^{jk} = \frac{a^k \mathcal{N}_j^k(x_i)}{\sum_{j=1}^{N_{sym}} \sum_{k=1}^{m} a^k \mathcal{N}_j^k(x_i)}. \tag{A3}$$

For the parameter estimation in the maximization step of the algorithm, we apply the inverse symmetry transformations $\mathrm{M}^\top$ to the samples. The weight calculated for sample i with Gaussian k transformed with symmetry j is now used on the sample transformed with $\mathrm{M}_j^\top$ to estimate the parameters μ^k and Σ^k,

$$\mu^k = \frac{\sum_{j=1}^{N_{sym}} \sum_{i=1}^{N} y_i^{jk} \mathrm{M}_j^\top x_i}{\sum_{ij} y_i^{jk}}, \tag{A4}$$

$$\Sigma^k = \frac{1}{\sum_{ij} y_i^{jk}} \sum_{j=1}^{N_{sym}} \sum_{i=1}^{N} y_i^{jk} \left(\mathrm{M}_j^\top x_i - \mu^k\right)\left(\mathrm{M}_j^\top x_i - \mu^k\right)^\top. \tag{A5}$$

As in the original version of the algorithm, the expectation and maximization steps are repeated alternately until the convergence is achieved.

With this modified version of the expectation–maximization algorithm, we were able to achieve significantly better fits and hence higher performance of our funnel hopping Monte Carlo algorithm whenever symmetries were present in any of the local minima used.

REFERENCES

[1] F. H. Stillinger and T. A. Weber, "Hidden structure in liquids," Phys. Rev. A **25**, 978–989 (1982).
[2] F. Calvo, J. Doye, and D. Wales, "Equilibrium properties of clusters in the harmonic superposition approximation," Chem. Phys. Lett. **366**, 176–183 (2002).
[3] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," Biometrika **57**, 97–109 (1970).
[4] D. J. Wales, "Decoding the energy landscape: Extracting structure, dynamics and thermodynamics," Philos. Trans. R. Soc., A **370**, 2877–2899 (2012).
[5] D. J. Wales, "Energy landscapes: Some new horizons," Curr. Opin. Struct. Biol. **20**, 3–10 (2010).
[6] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," J. Comput. Phys. **23**, 187–199 (1977).
[7] A. Laio and M. Parrinello, "Escaping free-energy minima," Proc. Natl. Acad. Sci. U. S. A. **99**, 12562–12566 (2002).
[8] B. A. Berg and T. Neuhaus, "Multicanonical algorithms for first order phase transitions," Phys. Lett. B **267**, 249–253 (1991).
[9] F. Wang and D. P. Landau, "Efficient, multiple-range random walk algorithm to calculate the density of states," Phys. Rev. Lett. **86**, 2050–2053 (2001).
[10] P. Poulain, F. Calvo, R. Antoine, M. Broyer, and P. Dugourd, "Performances of Wang-Landau algorithms for continuous systems," Phys. Rev. E **73**, 056704 (2006).
[11] P. Nigra, D. L. Freeman, and J. D. Doll, "Combining smart darting with parallel tempering using Eckart space: Application to Lennard-Jones clusters," J. Chem. Phys. **122**, 114113 (2005).
[12] V. A. Sharapov and V. A. Mandelshtam, "Solid–solid structural transformations in Lennard-Jones clusters: Accurate simulations versus the harmonic superposition approximation," J. Phys. Chem. A **111**, 10284–10291 (2007).
[13] S. Goedecker, "Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems," J. Chem. Phys. **120**, 9911–9917 (2004).
[14] I. Andricioaei, J. E. Straub, and A. F. Voter, "Smart darting Monte Carlo," J. Chem. Phys. **114**, 6994–7000 (2001).
[15] V. A. Sharapov, D. Meluzzi, and V. A. Mandelshtam, "Low-temperature structural transitions: Circumventing the broken-ergodicity problem," Phys. Rev. Lett. **98**, 105701 (2007).
[16] A. Uhlherr and D. N. Theodorou, "Accelerating molecular simulations by reversible mapping between local minima," J. Chem. Phys. **125**, 084107 (2006).

[17]F. Noé, S. Olsson, J. Köhler, and H. Wu, "Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning," Science **365**, eaaw1147 (2019).
[18]T. V. Bogdan, D. J. Wales, and F. Calvo, "Equilibrium thermodynamics from basin-sampling," J. Chem. Phys. **124**, 044102 (2006).
[19]D. J. Wales, "Surveying a complex potential energy landscape: Overcoming broken ergodicity using basin-sampling," Chem. Phys. Lett. **584**, 1–9 (2013).
[20]S. Martiniani, J. D. Stevenson, D. J. Wales, and D. Frenkel, "Superposition enhanced nested sampling," Phys. Rev. X **4**, 031034 (2014).
[21]F. H. Stillinger, "Exponential multiplicity of inherent structures," Phys. Rev. E **59**, 48–51 (1999).
[22]J. P. K. Doye, M. A. Miller, and D. J. Wales, "Evolution of the potential energy surface with size for Lennard-Jones clusters," J. Chem. Phys. **111**, 8417–8428 (1999).
[23]P. A. Frantsuzov and V. A. Mandelshtam, "Size-temperature phase diagram for small Lennard-Jones clusters," Phys. Rev. E **72**, 037102 (2005).
[24]J. Behler, "First principles neural network potentials for reactive simulations of large molecular and condensed systems," Angew. Chem., Int. Ed. **56**, 12828–12840 (2017).
[25]C. Eckart, "Some studies concerning rotating axes and polyatomic molecules," Phys. Rev. **47**, 552–558 (1935).
[26]K. N. Kudin and A. Y. Dymarsky, "Eckart axis conditions and the minimization of the root-mean-square deviation: Two closely related problems," J. Chem. Phys. **122**, 224105 (2005).
[27]S. K. Kearsley, "On the orthogonal transformation used for structural comparisons," Acta Crystallogr., Sect. A: Found. Crystallogr. **45**, 208–210 (1989).
[28]E. A. Coutsias, C. Seok, and K. A. Dill, "Using quaternions to calculate RMSD," J. Comput. Chem. **25**, 1849–1857 (2004).
[29]S. V. Krasnoshchekov, E. V. Isayeva, and N. F. Stepanov, "Determination of the Eckart molecule-fixed frame by use of the apparatus of quaternion algebra," J. Chem. Phys. **140**, 154104 (2014).
[30]H. W. Kuhn, "The Hungarian method for the assignment problem," Naval Res. Logist. Q. **2**, 83–97 (1955).
[31]R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," Computing **38**, 325–340 (1987).
[32]B. Schaefer and S. Goedecker, "Computationally efficient characterization of potential energy surfaces based on fingerprint distances," J. Chem. Phys. **145**, 034101 (2016).
[33]J. E. Gentle, *Computational Statistics*, Statistics and Computing (Springer, Dordrecht, New York, 2009), pp. 315–316, oCLC: ocn437081973.
[34]A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. Royal Stat. Soc.: Ser. B **39**, 1–38 (1977).
[35]J. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," Technical Report ICSI-TR-97-021, University of Berkeley, 1997.
[36]M. R. Gupta and Y. Chen, "Theory and use of the EM algorithm," Found. Trends Signal Process. **4**, 223–296 (2010).
[37]B. R. Brooks, D. Janežič, and M. Karplus, "Harmonic analysis of large systems. I. Methodology," J. Comput. Chem. **16**, 1522–1542 (1995).
[38]T. R. Koehler, "Theory of the self-consistent harmonic approximation with application to solid neon," Phys. Rev. Lett. **17**, 89 (1966).
[39]N. Gillis, N. Werthamer, and T. Koehler, "Properties of crystalline argon and neon in the self-consistent phonon approximation," Phys. Rev. **165**, 951 (1968).
[40]I. Georgescu and V. A. Mandelshtam, "Self-consistent phonons revisited. I. The role of thermal versus quantum fluctuations on structural transitions in large Lennard-Jones clusters," J. Chem. Phys. **137**, 144106 (2012).
[41]S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," Phys. Lett. B **195**, 216–222 (1987).
[42]R. M. Neal, "MCMC using Hamiltonian dynamics," *Handbook of Markov Chain Monte Carlo*, Vol. 2, edited by S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng (CRC Press, Boca Raton, FL, USA, 2011), pp. 113–162.
[43]R. H. Swendsen and J.-S. Wang, "Replica Monte Carlo simulation of spin-glasses," Phys. Rev. Lett. **57**, 2607–2609 (1986).
[44]J. P. Neirotti, F. Calvo, D. L. Freeman, and J. D. Doll, "Phase changes in 38-atom Lennard-Jones clusters. I. A parallel tempering study in the canonical ensemble," J. Chem. Phys. **112**, 10340–10349 (2000).
[45]D. J. Earl and M. W. Deem, "Parallel tempering: Theory, applications, and new perspectives," Phys. Chem. Chem. Phys. **7**, 3910 (2005).
[46]D. A. Kofke, "On the acceptance probability of replica-exchange Monte Carlo trials," J. Chem. Phys. **117**, 6911–6914 (2002).
[47]N. Rathore, M. Chopra, and J. J. de Pablo, "Optimal allocation of replicas in parallel tempering simulations," J. Chem. Phys. **122**, 024111 (2005).
[48]J. P. K. Doye, M. A. Miller, and D. J. Wales, "The double-funnel energy landscape of the 38-atom Lennard-Jones cluster," J. Chem. Phys. **110**, 6896–6906 (1999).
[49]M. Johansson and V. Veryazov, "Automatic procedure for generating symmetry adapted wavefunctions," J. Cheminf. **9**, 8 (2017).